

dr bob's elementary differential geometry

a slightly different approach
based on elementary undergraduate linear algebra,
multivariable calculus and differential equations

by bob jantzen

(Robert T. Jantzen)

Department of Mathematics and Statistics

Villanova University

Copyright 2007, 2008, 2013

<http://www.homepage.villanova.edu/robert.jantzen/notes/diffgeom/>

original 2007 source typeset by Hans Kuo, Taiwan

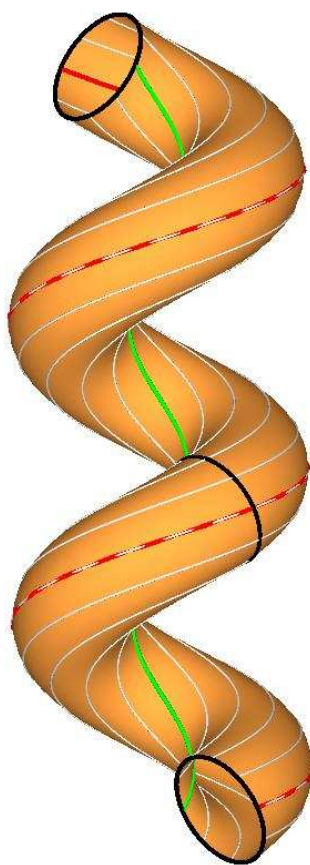
in progress: version: February 1, 2015

Abstract

There are lots of books on differential geometry, including at the introductory level. Why yet another one by an author who doesn't seem to take himself that seriously and occasionally refers to himself in the third person? This one is a bit different than all the rest. dr bob loves this stuff, but how to teach it to students at his own (not elite) university in order to have a little more fun at work than usual? This unique approach may not work for everyone, but it attempts to explain the nuts and bolts of how a few basically simple ideas taken seriously underlie the whole mess of formulas and concepts, without worrying about technicalities like "manifolds," "coordinate coverings" and "differentiability," which only serve to put off students at the first pass through this scenery. It is also presented with an eye towards being able to understand the key concepts needed for the mathematical side of modern physical theories, while still providing the tools that underlie the classical theory of curves and surfaces in space. Examples of curves and surfaces in 2- and 3-dimensional spacetimes have been incorporated as examples, with an Appendix presenting a review of the elementary special relativity (hyperbolic geometry, directly analogous to trigonometry) needed to make sense of them. The continuing theme of symmetry groups and their implications for geometry have also now been woven into the narrative, which is somewhat uncommon for expositions of differential geometry, but essential to a proper understanding of the implications of the subject for applications to physical theories. Finally, economy of explanation and derivations is abandoned for lengthy explanations and detailed derivations to try to make the material as readable as possible without forcing the reader to think about missing steps unless intended to be an explicit exercise or digest new ideas that are presented too concisely.

100 years of general relativity.

Differential geometry got an enormous boost from Albert Einstein and his general relativity theory of 1915, a theory for which the then new tensor analysis tools developed by Ricci and Levi-Civita based on the work of Gauss and Riemann and a handful of others provided the key to its development. The crucial idea of parallel transport did not come until 1917 when Levi-Civita, motivated by general relativity, thought it up, and soon after Cartan created the modern theory of connections on fiber bundles. Indeed a lot of attention was focused on differential geometry because of its obvious importance for general relativity in those early days, but in fact the rest of the century revealed that geometrization was the key to understanding the fundamental forces and their unification. It is therefore entirely reasonable that finally in the twenty first century, the subject might be taught at an elementary level acknowledging this aspect of its history.



cavatappo 2.0: the fixed radius tubular surface of a helix
(and mathematical pasta shape)

©2012 drbobenterprises.com

Acknowledgments

The patience of my wife and life partner Ani must be acknowledged in dealing with the obessional tinkering of the author with the ideas and problems of this text and the supporting Maple worksheets. Fritz Hartmann has to remembered for allowing me to let me make my first overly ambitious attempt at a course on this topic when I had just arrived at Villanova University in my first year ever of teaching in George Orwell's year 1984. Some time later in 1991 I was able to give it the sophomore try with 6 trusting students who probably left the course wondering what they had just done, but the opportunity allowed me to write up the first version of the handwritten notes from which this book eventually sprang to life. Without Hans Kuo, there never would have been a book, since starting to LaTeX such a project from scratch would never have occurred to me, and he gifted me the first draft taken from my scanned handwritten notes I had posted on the internet. These were then seriously developed with the addition of problems and technology graphics and more text in an offering of the course in 2008, after which fours years passed before I had the time to return to the project. Cole Johnston must be credited with pushing me to offer the course again in 2013, which coincided with an awakening of my own interest in surfaces motivated by the opportunity to give a popular talk on differential geometry and relativity in which the surface geometry of corkscrew pasta played a starring role in conveying not only the visual ideas of metric geometry, but tied these mathematical abstractions to the Italian mathematicians who were instrumental in developing the tools for Einstein's general theory of relativity. Remo Ruffini gets credit for drawing me into his obsession with the early work on electromagnetic mass and general relativity done by Enrico Fermi, where the corkscrew pasta surface in a 3-dimensional Minkowski spacetime describes the equator of a classical spherical electron in a circular orbit, inspiring me to incorporate special relativity into the examples and problems of the text. Eduard Bachmakov gave me the missing computer expertise I needed to finally fix my outstanding problem for the coding of my numbered exercises and activate complete hyperlinking of all cross-references in the exported PDF document that tremendously increased the usability of that electronic platform.

Finally without LaTeX such a self-produced book would not have been possible, nor would I have been able to create the graphics illustrations without Maple, nor back up the calculations which make this subject come alive without its computer algebra engine.

Contents

Abstract	2
Acknowledgments	4
Table of Contents	4
Preface	6
I ALGEBRA	8
0 Introduction: motivating index algebra	9
Geometry?	14
1 Foundations of tensor algebra	19
1.1 Index conventions	20
1.2 A vector space V	21
Elementary linear algebra: solving systems of linear equations	27
Elementary linear algebra: the eigenvalue problem and linear transformations	31
1.3 The dual space V^*	34
1.4 Linear transformations of a vector space V into itself (and tensors)	51
Concrete examples of linear transformations	51
Linear transformations and tensors	54
More than 2 indices: general tensors	60
1.5 Linear transformations of V into itself and a change of basis	67
Matrix form of the “transformation law” for $\binom{1}{1}$ -tensors	72
Matrices of symmetric $\binom{0}{2}$ -tensors	73
A 3-index example	74
1.6 Linear transformations between V and V^*	82
Invertible maps between V and V^*	83
Inner products	83
Index shifting with an inner product	92
Index shifting conventions	94
Partial evaluation of a tensor and index shifting	98
Contraction of tensors	98
Geometric interpretation of index shifting	99

	Cute fact (an aside for your reading pleasure): geometric interpretation of index lowering on vectors	115
1.7	Matrix groups	121
	Transformation groups	129
	Non-Abelian groups	131
	Representations of Lie groups	140
2	Symmetry properties of tensors	147
2.1	Measure motivation and determinants	148
2.2	Tensor symmetry properties	153
2.3	Epsilons and deltas	158
2.4	Antisymmetric tensors	171
2.5	Symmetric tensors and multivariable Taylor series	174
	Symmetric tensors and multipole moments in physics	176
	Moments of inertia?	178
3	Time out	183
3.1	Whoa! Review of what we've done so far	184
4	Antisymmetric tensors, subspaces and measure	197
4.1	Determinants gone wild	198
4.2	The wedge product	204
4.3	Subspace orientation and a new star duality	211
	Rescaled inner product for antisymmetric tensors	218
	The unit n -form on an oriented vector space with inner product	220
	The metric dual	222
	Inner product, duality and wedge product relations	226
	Determining subspaces	235
4.4	Wedge and star duality on R^n in practice	240
4.5	Matrix generators of the generalized orthogonal matrix groups	248
	Hermitian and unitary matrices	254
II	CALCULUS	264
5	From multivariable calculus to the foundation of differential geometry	265
5.1	The tangent space in multivariable calculus	266
5.2	More motivation for the re-interpretation of the tangent space	276
5.3	Flow lines of vector fields	281
5.4	Frames and dual frames and Lie brackets	293
5.5	Non-Cartesian coordinates on \mathbb{R}^n (polar coordinates in \mathbb{R}^2)	301
5.6	Cylindrical and spherical coordinates on R^3	311
5.7	Cylindrical coordinate frames	315
5.8	Spherical coordinate frames	323

5.9	Lie brackets and noncoordinate frames	333
6	Covariant derivatives	341
6.1	Covariant derivatives on R^n with Euclidean metric	342
6.2	Notation for covariant derivatives	345
6.3	Covariant differentiation and the general linear group	354
6.4	Covariant constant tensor fields	362
6.5	The clever way of evaluating the components of the covariant derivative	365
	Key to Riemannian geometry and surface geometry in R^3	366
6.6	Noncoordinate frames	369
6.7	Geometric interpretation of the Lie bracket	374
	Lie brackets and transformation groups	375
6.8	Isometry groups and Killing vector fields	381
6.9	Noncoordinate frames and $SO(3, \mathbb{R})$	397
7	More on covariant derivatives	399
7.1	Gradient, curl and divergence	400
7.2	Second covariant derivatives and the Laplacian	405
7.3	Spherical coordinate orthonormal frame	413
7.4	Rotations and derivatives	419
8	Parallel transport and geodesics	425
8.1	Covariant differentiation along a curve and parallel transport	427
8.2	Parallel transport within coordinate surfaces in space	438
8.3	Geodesics	444
	Conserved momentum, symmetries and Killing vector fields	448
8.4	Surfaces of revolution	450
8.5	Parametrized curves as “motion of point particles” and the geodesic motion approach	462
8.6	The Euclidean plane and the Kepler problem	469
	Kepler’s problem	473
8.7	2-spheres, pseudospheres and other conics of revolution	480
	spheres	481
	Minkowski geometry	489
	unit spacelike pseudosphere	490
	unit timelike pseudosphere	491
8.8	The torus	496
8.9	Geodesics as extremal curves: a peek at the calculus of variations	506
	The boundary value problem for geodesics	516
8.10	The rigid body example and $SO(3, \mathbb{R})$	517
8.11	The screw-symmetric helical tube	519
8.12	The Schwarzschild equatorial plane geometry	530

9	Intrinsic curvature	537
9.1	Calculating the curvature tensor and its symmetries	539
9.2	Interpretation of curvature	551
9.3	The limiting loop parallel transport calculation of curvature	556
	The limiting loop parallel transport frame curvature calculation	559
	The symmetry of the covariant derivative	560
9.4	Positive curvature focusing of geodesics and negative curvature defocusing	563
10	Extrinsic curvature	573
10.1	The extrinsic curvature tensor	574
10.2	Spheres, cylinders and cones: some useful concrete examples	582
	Cones: a useful cautionary example	584
10.3	Extrinsic curvature as a quadratic approximation	586
10.4	Total curvature: intrinsic plus extrinsic curvature	590
10.5	Tube/tubular surfaces	600
10.6	Surface geodesics studied from the outside	605
11	Differential forms: integration and differentiation	613
11.1	Changing the variable in a single variable integral	614
11.2	Changing variables in multivariable integrals	616
11.3	Parametrized p -surfaces and pushing forward the coordinate grid and tangent vectors	620
11.4	Pulling back functions, covariant tensors and differential forms	623
11.5	Changing the parametrization	628
11.6	Integration and a metric	632
	Curves in a 3-dimensional flat space	632
	Surfaces in a 3-dimensional flat space	632
	Integrating over a p -surface in an n -dimensional space	635
11.7	The exterior derivative d	637
11.8	The exterior derivative and a metric	650
	star, sharp, flat, d and ∇	650
	grad, div and curl	653
	the codifferential δ	658
	Commutative diagrams?	661
11.9	Induced orientation on a boundary	664
	First coordinate adapted to boundary	664
	Induced orientation for any coordinate ordering	667
11.10	Stokes' theorem	670
	The case $p = 2$ in \mathbb{R}^3	670
	The case $p = 3$ in \mathbb{R}^3	672
	Idea of the proof of Stokes' Theorem	673
11.11	Worked examples of Stokes' theorem and Gauss's law for \mathbb{R}^3	676
	The ordinary Stokes' theorem in \mathbb{R}^3	676

A Gauss's law problem	680
11.12 Examples in \mathbb{R}^4 and \mathbb{M}^4	683
3-spheres, 3-cylinders and 2-cylinders in \mathbb{R}^4	683
3-cylinders in \mathbb{R}^4	688
2-cylinders in \mathbb{R}^4	688
pseudospheres in \mathbb{M}^4	688
cylinders in \mathbb{M}^4	688
constant inertial time hypersurface	688
12 Wrapping things up	689
12.1 Final remarks	690
1991	690
2013	690
12.2 MATH 5600 Spring 1991 Differential Geometry: Take Home Final	691
Appendices	699
A From trigonometry to hyperbolic functions and hyperbolic geometry	701
B Hyperbolic geometry and special relativity	713
C Curves in 3-space	727
The Euclidean helix	727
Circles to pseudo-circles: hyperbolas	735
The Lorentz helix	739
D Surfaces in 3-space	747
E Visualizing vector space duality in the vector space R^3	755
III Supplementary materials	763
Solutions to Exercises	765
Chapter 0	767
Chapter 1	768
Chapter 2	777
Chapter 3	780
Chapter 4	781
Chapter 5	785
Chapter 6	794
Chapter 7	804
Chapter 8	806
Chapter 9	808
Chapter 10	810

Chapter 11	811
Chapter 12: final exam worked	813
Maple worksheets	827

Preface

This book began as a set of handwritten notes from a course given at Villanova University in the spring semester of 1991 that were scanned and posted on the web in 2006 at <http://www34.homepage.villanova.edu/robert.jantzen/notes/dg1991/> and were converted to a L^AT_EX compuscript and completely revised in 2007–2008 with the help of Hans Kuo of Taiwan through a serendipitous internet collaboration and chance second offering of the course to actual students in the spring semester of 2008, offering the opportunity for serious revision with feedback. Life then intervened and the necessary cleanup operations to put this into a finished form were delayed indefinitely.

Most undergraduate courses on differential geometry are leftovers from the early part of the last century, focusing on curves and surfaces in space, which is not very useful for the most important application of the twentieth century: general relativity and field theory in theoretical physics. Most mathematicians who teach such courses are not well versed in physics, so perhaps this is a natural consequence of the distancing of mathematics from physics, two fields which developed together in creating these ideas from Newton to Einstein and beyond. The idea of these notes is to develop the essential tools of modern differential geometry while bypassing more abstract notions like manifolds, which although important for global questions, are not essential for local differential geometry and therefore need not steal precious time from a first course aimed at undergraduates. On the other hand physicists interested in getting students to the heart of general relativity under time constraints often neglect the mathematical structure that makes tensor analysis more digestible when recast in a more modern light. (One of these shortcuts I think is particularly regrettable is to bypass the understanding of linearity embodied in the concept of the dual space to a vector space by using reciprocal bases to evaluate components along a basis of a vector space. See Appendix E.) Since this is not the primary objective of these notes, we can take a compromise path which tries to give a better view of the overall mathematical structure that will enable interested students to explore applications on their own.

Part 1 (Algebra) develops the vector space structure of \mathbb{R}^n and its dual space of real-valued linear functions, and builds the tools of tensor algebra on that structure, getting the index manipulation part of tensor analysis out of the way first. Part 2 (Calculus) then develops \mathbb{R}^n as a manifold first analyzed in Cartesian coordinates, beginning by redefining the tangent space of multivariable calculus to be the space of directional derivatives at a point, so that all of the tools of Part 1 then can be applied pointwise to the tangent space. Non-Cartesian coordinates and the Euclidean metric are then used as a shortcut to what would be the consideration of more general manifolds with Riemannian metrics in a more ambitious course, followed by the covariant derivative and parallel transport, leading naturally into curvature. The exterior derivative and integration of differential forms is the final topic, showing how conventional vector analysis fits into a more elegant unified framework. Flat Minkowski spacetime geometry is woven into the story together with its symmetry groups, and a few curved space examples from general relativity help drive home the point of truly curved spaces.

The theme of Part 1 is that one needs to distinguish the linearity properties from the inner product (“metric”) properties of elementary linear algebra. The inner product geometry

governs lengths and angles, and the determinant then enables one to extend the linear measure of length to area and volume in the plane or 3-dimensional space, and to p -dimensional objects in \mathbb{R}^n . The determinant also tests linear independence of a set of vectors and hence is key to characterizing subspaces independent of the particular set of vectors we use to describe them while assigning an actual measure to the p -parallelepipeds formed by a particular set, once an inner product sets the length scale for orthogonal directions. By appreciating the details of these basic notions in the setting of \mathbb{R}^n , one is ready for the tools needed point by point in the tangent spaces to \mathbb{R}^n , once one understands the relationship between each tangent space and the simpler enveloping space. Along the way we discover how basic notions about matrices and vectors and their algebra resurface in so many ways in the tensor algebra needed to do basic differential geometry.

This book is not for everyone. It is verbose, trying to explain in much detail how everything works, with lots of examples interwoven into the discussion. It is aimed at those students who only have the limited foundation of multivariable calculus (see Appendix C for curves and D for surfaces), linear algebra and differential equations, and tries to avoid abstractions. No inverse function theorem remarks here, for example.

In the spring of 2013, I had a second opportunity to go further with this project by incorporating the mathematics of special relativity into the applications since clearly relativity is a more interesting application than surfaces in space which are the prime target of the usual differential geometry offerings. This in turn led to extending the existing material naturally to include continuous symmetry groups, the missing component of these notes until then. I decided to start the course with a simple multivariable calculus calculation which evaluates the dominant contribution to the geodetic precession effect measured by the GP-B satellite experiment in recent years, and follow with a crash course in hyperbolic geometry (see Appendix A) that is always skipped in our calculus offerings, connecting it up with special relativity (see Appendix B) which would then be woven into the main text in parallel with the more familiar Euclidean geometry associated with the dot product. During the fall of 2012 I tried to think of interesting ways to incorporate relativity into the applications at an elementary level, and having gotten excited about the surface geometry of screw-symmetric surfaces in modeling pasta and circularly orbiting particles, added some more appendices reviewing the basics of special relativity and reviewing curves and surfaces from multivariable calculus. Only at the end of this upgrade will it be clear whether this burst of enthusiasm was successful in exciting the students.

Part 3 is indispensable to students trying to self-study using this book as well as to those rare exceptions who might be in an actual course using it, since it is an ambitious text and includes many options explored in exercises that might appeal to particular interests of the reader that time won't permit discussion of in a course setting. These supplementary materials begin with a complete solution manual electronically linked back to the exercises of the main text in the PDF version of the book for ready access. Many of the exercise solutions refer to Maple worksheets freely available on dr bob's website while it exists, so an index of these follows the solution manual.

Part I

ALGEBRA

Chapter 0

Introduction: motivating index algebra

Elementary linear algebra is the mathematics of linearity, whose basic objects are 1- and 2-dimensional arrays of numbers, which can be visualized as at most 2-dimensional rectangular arrangements of those numbers on sheets of paper or computer screens. Arrays of numbers of dimension d can be described as sets that can be put into a 1-1 correspondence with regular rectangular grids of points in \mathbb{R}^d whose coordinates are integers, used as index labels:

$$\begin{array}{ll} \{a_i | i = 1, \dots, n\} & \text{1-}d \text{ array : } n \text{ entries} \\ \{a_{ij} | i = 1, \dots, n_1, j = 1, \dots, n_2\} & \text{2-}d \text{ array : } n_1 n_2 \text{ entries} \\ \{a_{ijk} | i = 1, \dots, n_1, j = 1, \dots, n_2, k = 1, \dots, n_3\} & \text{3-}d \text{ array : } n_1 n_2 n_3 \text{ entries} \end{array}$$

1-dimensional arrays (vectors) and 2-dimensional arrays (matrices), coupled with the basic operation of matrix multiplication, itself an organized way of performing dot products of two sets of vectors, combine into a powerful machine for linear computation. When working with arrays of specific dimensions (3 component vectors, 2×3 matrices, etc.), one can avoid index notation and the sigma summation symbol $\sum_{i=1}^n$ after using it perhaps to define the basic operation of dot products for vectors of arbitrary dimension, but to discuss theory for indeterminant dimensions (n -component vectors, $m \times n$ matrices), index notation is necessary. However, index “positioning” (distinguishing subscript and superscript indices) is not essential and rarely used, especially by mathematicians. Going beyond 2-dimensional arrays to d -dimensional arrays for $d > 2$, the arena of “tensors”, index notation and index positioning are instead both essential to an efficient computational language.

Suppose we start with 3-vectors to illustrate the basic idea. (We will sometimes use an over arrow symbol to signal a vector in \mathbb{R}^n for emphasis, but not always.) The dot product between two vectors is symmetric in the two factors

$$\begin{aligned} \vec{a} &= \langle a_1, a_2, a_3 \rangle, \quad \vec{b} = \langle b_1, b_2, b_3 \rangle \\ \vec{a} \cdot \vec{b} &= a_1 b_1 + a_2 b_2 + a_3 b_3 = \sum_{i=1}^3 a_i b_i = \vec{b} \cdot \vec{a}, \end{aligned}$$

but using it to describe a linear function in \mathbb{R}^3 , a basic asymmetry is introduced

$$f_{\vec{a}}(\vec{x}) = \vec{a} \cdot \vec{x} = a_1x_1 + a_2x_2 + a_3x_3 = \sum_{i=1}^3 a_i x_i .$$

The left factor is a constant vector of “coefficients”, while the right factor is the vector of “variables” and this choice of left and right is arbitrary but convenient, although some mathematicians like to reverse it for some reason. To reflect this distinction, we introduce superscripts (up position) to denote the variable indices and subscripts (down position) to denote the coefficient indices, and then agree to sum over the understood 3 values of the index range for any repeated such pair of indices (one up, one down)

$$f_{\vec{a}}(\vec{x}) = a_1x^1 + a_2x^2 + a_3x^3 = \sum_{i=1}^3 a_i x^i = a_i x^i .$$

The last convention, called the Einstein summation convention, turns out to be an extremely convenient and powerful shorthand, which in this example, streamlines the notation for taking a “linear combination of variables,” namely the sum of the matched products of corresponding coefficients and variables.

This index positioning notation encodes the distinction between rows and columns in the matrix notation for a linear transformation. We will represent a matrix (a_{ij}) representing a linear transformation instead as (a^i_j) with row indices (left) associated with superscripts, and column indices (right) with subscripts. A single row matrix or column matrix is used to denote respectively a “coefficient” vector and a “variable” vector

$$(a_1 \quad a_2 \quad a_3) , \quad \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} ,$$

where the entries of a single row matrix are labeled by the column index (down), and the entries of a single column matrix are labeled by the row index (up).

The matrix product of a row matrix on the left by a column matrix on the **right** re-interprets the dot product between two vectors as the way to combine a row vector (left factor) of coefficients with a column vector (right factor) of variables to produce a single number, the value of a linear function of the variables

$$(a_1 \quad a_2 \quad a_3) \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = a_1x^1 + a_2x^2 + a_3x^3 = \vec{a} \cdot \vec{x} .$$

If we agree to use an underlined kernel symbol \underline{x} for a column vector, and the transpose \underline{a}^T for a row vector, where the transpose simply interchanges rows and columns of a matrix, this can be represented as $\underline{a}^T \underline{x} = \vec{a} \cdot \vec{x}$. Since many geometric objects also have component matrices, it will be useful to link them together by using the same kernel symbol and underlining the matrix symbol to distinguish it from the object from which the components are taken.

Extending the matrix product to more than one row in the left factor is the second step in defining a general matrix product, leading to a column vector result

$$\begin{pmatrix} a^1_1 & a^1_2 & a^1_3 \\ a^2_1 & a^2_2 & a^2_3 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} \underline{a}^{1T} \\ \underline{a}^{2T} \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} \vec{a}^1 \cdot \vec{x} \\ \vec{a}^2 \cdot \vec{x} \end{pmatrix} = \begin{pmatrix} a^1_i x^i \\ a^2_i x^i \end{pmatrix}.$$

Thinking of the coefficient matrix as a 1-dimensional vertical array of row vectors (the first right hand side of this sequence of equations), one gets a corresponding array of numbers (a column) as the result, consisting of the corresponding dot products of the rows with the single column. Denoting the left matrix factor by \underline{A} , then the product column matrix has entries

$$[\underline{A} \underline{x}]^i = \sum_{k=1}^3 a^i_k x^k = a^i_k x^k, \quad 1 \leq i \leq 2.$$

Finally, adding more columns to the right factor in the matrix product, we generate corresponding columns in the matrix product, with the resulting array of numbers representing all possible dot products between the row vectors on the left and the column vectors on the right, labeled by the same row and column indices as the factor vectors from which they come

$$\begin{pmatrix} a^1_1 & a^1_2 & a^1_3 \\ a^2_1 & a^2_2 & a^2_3 \end{pmatrix} \begin{pmatrix} x^1_1 & x^1_2 \\ x^2_1 & x^2_2 \\ x^3_1 & x^3_2 \end{pmatrix} = \begin{pmatrix} \underline{a}^{1T} \\ \underline{a}^{2T} \end{pmatrix} (\underline{x}_1 \quad \underline{x}_2) = \begin{pmatrix} \vec{a}^1 \cdot \vec{x}_1 & \vec{a}^1 \cdot \vec{x}_2 \\ \vec{a}^2 \cdot \vec{x}_1 & \vec{a}^2 \cdot \vec{x}_2 \end{pmatrix}.$$

Denoting the new left matrix factor again by \underline{A} and the right matrix factor by \underline{X} , then the product matrix has entries (row index left up, column index right down)

$$[\underline{A} \underline{X}]^i_j = \sum_{k=1}^3 a^i_k x^k_j = a^i_k x^k_j, \quad 1 \leq i \leq 2, \quad 1 \leq j \leq 2,$$

where the sum over three entries (representing the dot product) is implied by our summation convention in the second equality, and the row and column indices here go from 1 to 2 to label the entries of the 2 rows and 2 columns of the product matrix. Thus matrix multiplication in this example is just an organized way of displaying all such dot products of two ordered sets of vectors in an array where the rows of the left factor in the matrix product correspond to the coefficient vectors in the left set and the columns in the right factor in the matrix product correspond to the variable vectors in the right set. The dot product itself in this context of matrix multiplication is representing the natural evaluation of linear functions (left row) on vectors (right column). No geometry (lengths and angles in Euclidean geometry) is implied in this context, only linearity and the process of linear combination.

The matrix product of a matrix with a single column vector can be reinterpreted in terms of the more general concept of a vector-valued linear function of vectors, namely a linear combination of vectors, in which case the right factor column vector entries play the role of

coefficients. In this case the left factor matrix must be thought of as a horizontal array of column vectors

$$\begin{aligned} (\underline{v}_1 \quad \underline{v}_2 \quad \underline{v}_3) \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} &= \begin{pmatrix} v^1_1 & v^1_2 & v^1_3 \\ v^2_1 & v^2_2 & v^2_3 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} v^1_1 x^1 + v^1_2 x^2 + v^1_3 x^3 \\ v^2_1 x^1 + v^2_2 x^2 + v^2_3 x^3 \end{pmatrix} \\ &= x^1 \begin{pmatrix} v^1_1 \\ v^2_1 \end{pmatrix} + x^2 \begin{pmatrix} v^1_2 \\ v^2_2 \end{pmatrix} + x^3 \begin{pmatrix} v^1_3 \\ v^2_3 \end{pmatrix} = x^1 \underline{v}_1 + x^2 \underline{v}_2 + x^3 \underline{v}_3 = x^i \underline{v}_i. \end{aligned}$$

Thus in this case the summed-over index pair performs a linear combination of the columns of the left factor of the matrix product, whose coefficients are the entries of the right column matrix factor. This interpretation extends to more columns in the right matrix factor, leading to a matrix product consisting of the same number of columns, each of which represents a linear combination of the column vectors of the left factor matrix. In this case the coefficient indices are superscripts since the labels of the vectors being combined linearly are subscripts, but the one up, one down repeated index summation is still consistent. Note that when the left factor matrix is not square (in this example, a 2×3 matrix multiplied by a 3×1 matrix), one is dealing with coefficient vectors \underline{v}_i and vectors \underline{x} of different dimensions, in this example combining three 2-component vectors by linear combination.

If we call our basic column vectors just vectors (contravariant vectors, indices up) and call row vectors “covectors” (covariant vectors, indices down), then combining them with the matrix product represents the evaluation operation for linear functions, and implies no geometry in the sense of lengths and angles usually associated with the dot product, although one can easily carry over this interpretation. In this example \mathbb{R}^3 is our basic vector space consisting of all possible ordered triplets of real numbers, and the space of all linear functions on it is equivalent to another copy of \mathbb{R}^3 , the space of all coefficient vectors. The space of linear functions on a vector space is called the dual space, and given a basis of the original vector space, expressing linear functions with respect to this basis leads to a component representation in terms of their matrix of coefficients as above.

It is this basic foundation of a vector space and its dual, together with the natural evaluation represented by matrix multiplication in component language, reflected in superscript and subscript index positioning respectively associated with column vectors and row vectors, that is used to go beyond elementary linear algebra to the algebra of tensors, or d -dimensional arrays for any positive integer d . Index positioning together with the Einstein summation convention is essential in letting the notation itself directly carry the information about its role in this scheme of linear mathematics extended beyond the elementary level.

Combining this linear algebra structure with multivariable calculus leads to differential geometry. Consider \mathbb{R}^3 with the usual Cartesian coordinates x^1, x^2, x^3 thought of as functions on this space. The differential of any function on this space can be expressed in terms of partial derivatives by the formula

$$df = \frac{\partial f}{\partial x^1} dx^1 + \frac{\partial f}{\partial x^2} dx^2 + \frac{\partial f}{\partial x^3} dx^3 = \partial_i f dx^i = f_{,i} dx^i$$

using first the abbreviation $\partial_i = \partial/\partial x^i$ for the partial derivative operator and then the abbreviation $f_{,i}$ for the corresponding partial derivatives of the function f . At each point of \mathbb{R}^3 , the

differentials df and dx^i play the role of linear functions on the tangent space. The differential of f acts on a tangent vector \vec{v} at a given point by evaluation to form the directional derivative along the vector

$$D_{\vec{v}}f = \frac{\partial f}{\partial x^1}v^1 + \frac{\partial f}{\partial x^2}v^2 + \frac{\partial f}{\partial x^3}v^3 = \frac{\partial f}{\partial x^i}v^i,$$

so that the coefficients of this linear function of a tangent vector \vec{v} at a given point are the values of the partial derivative functions there, and hence have indices down compared to the up indices of the tangent vector itself, which belongs to the tangent space, the fundamental vector space describing the differential geometry near each point of the whole space. In the linear function notation, the application of the linear function df to the vector \vec{v} gives the same result

$$df(\vec{v}) = \frac{\partial f}{\partial x^i}v^i.$$

If $\partial f/\partial x^i$ are therefore the components of a covector, and v^i the components of a vector in the tangent space, what is the basis of the tangent space, analogous to the natural (ordered) basis $\{e_1, e_2, e_3\} = \{\langle 1, 0, 0 \rangle, \langle 0, 1, 0 \rangle, \langle 0, 0, 1 \rangle\}$ of \mathbb{R}^3 thought of as a vector space in our previous discussion? In other words how do we express a tangent vector in the abstract form like in the naive \mathbb{R}^3 discussion where $\vec{x} = \langle x^1, x^2, x^3 \rangle = x^i e_i$ is expressed as a linear combination of the standard basis vectors $\{e_i\} = \{\langle 1, 0, 0 \rangle, \langle 0, 1, 0 \rangle, \langle 0, 0, 1 \rangle\}$ usually denoted by $\mathbf{i}, \mathbf{j}, \mathbf{k}$? This question will be answered in the following notes, making the link between old fashioned tensor analysis and modern differential geometry.

One last remark about matrix notation is needed. We adopt here the notational conventions of the computer algebra system Maple for matrices and vectors. A vector $\langle u^1, u^2 \rangle$ will be interpreted as a column matrix in matrix expressions

$$\underline{u} = \langle u^1, u^2 \rangle = \begin{pmatrix} u^1 \\ u^2 \end{pmatrix}$$

while its transpose will be denoted by

$$\underline{u}^T = \langle u^1 | u^2 \rangle = (u^1 \quad u^2).$$

In other words within triangle bracket delimiters, a comma will represent a vertical separator in a list, while a vertical line will represent a horizontal separator in a list. A matrix can then be represented as a vertical list of rows or as a horizontal list of columns, as in

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \langle \langle a | b \rangle, \langle c | d \rangle \rangle = \langle \langle a, c \rangle | \langle b, d \rangle \rangle.$$

Finally if \underline{A} is a matrix, we will not use a lowercase letter a^i_j for its entries but retain the same symbol: $\underline{A} = (A^i_j)$.

Since the matrix notation and matrix multiplication which suppresses all indices and the summation is so efficient, it is important to be able to translate between the summed indexed

notation to the corresponding index-free matrix symbols. In the usual language, matrix multiplication the i th row and j th column entry of the product matrix is

$$[\underline{A} \underline{B}]_{ij} = \sum_{k=1}^n A_{ik} B_{kj},$$

while in our streamlined notation when these represent linear transformations of the vectors space into itself, this becomes

$$[\underline{A} \underline{B}]^i_j = A^i_k B^k_j.$$

However, as we will see all other index position combinations are possible with corresponding different meanings. In our application of the matrix product to matrices with indices in various up/down positions, the left index will always be the row index and the right index the column index and to translate from indexed notation to symbolic matrices we always have to use the above correspondence independent of the index up or down position: only left-right position counts. Thus to translate an expression like $M_{ij} B^i_m B^j_n$ we need to first rearrange the factors to $B^i_m M_{ij} B^j_n$ and then recognize that the second summed index j is in the right adjacent pair of positions for interpretation of matrix multiplication, but the first summed index i is in the row instead of column position so the transpose is required to place it adjacent to the middle matrix factor

$$(B^i_m M_{ij} B^j_n) = ([\underline{B}^T \underline{M} \underline{B}]_{mn}) = \underline{B}^T \underline{M} \underline{B}.$$

Geometry?

Finally it is important to give some sense of what all this index business is needed for, connecting up with what has already been encountered in undergraduate multivariable calculus. In particular, what does all of this have to do with geometry? First, the term differential geometry refers to the study of the differential structure of “manifolds” which encompass not only the familiar straight line, flat plane and flat space of multivariable calculus but more complicated “curved spaces” like circles, spheres and cylinders or other conic-section related surfaces or more general surfaces in space as well as higher dimensional examples like the 3-sphere within 4-dimensional Euclidean space, a space which is often encountered in popular discussions of closed universe models. Second, geometry we first encounter in the context of lengths of line segments or angles between them (high school geometry class!). This idea of lengths and angles underlies the foundations of Riemannian (or pseudo-Riemannian) geometry in which the differential structure of differential geometry is given an additional “metric structure” that we associate with the distance formula and the dot product in multivariable calculus. To explain in more detail, we need more preliminary tools, but we can begin with an example that gives us a preview of what a metric is.

When we study arclengths of curves, we easily accept the expression for the square of the differential of arclength for a differential displacement in the plane from a point (x, y) to $(x + dx, y + dy)$ as the Pythagorean relation for right triangles

$$ds^2 = dx^2 + dy^2 = \begin{pmatrix} dx & dy \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix}, \quad (1)$$

which is very useful for integrating up finite arclengths along curves once the curve is parametrized. We also learn to re-express this formula passing from the standard Cartesian coordinates to polar coordinates in the plane by way of the coordinate transformation

$$(x, y) = (r \sin \theta, r \cos \theta) \quad (2)$$

whose differential

$$\begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} \cos \theta dr - r \sin \theta d\theta \\ \sin \theta dr + r \cos \theta d\theta \end{pmatrix} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} \begin{pmatrix} dr \\ d\theta \end{pmatrix} \equiv \underline{J} \begin{pmatrix} dr \\ d\theta \end{pmatrix}, \quad (3)$$

which defines the so-called Jacobian matrix \underline{J} of partial derivatives of the old coordinates with respect to the new ones. It is then a simple matter to substitute these relations into the square of the differential arclength, and expand and simplify the result using the fundamental trigonometric identity

$$ds^2 = (\cos \theta dr - r \sin \theta d\theta)^2 + (\sin \theta dr + r \cos \theta d\theta)^2 = \dots = dr^2 + r^2 d\theta^2. \quad (4)$$

In terms of the matrix representation of this same calculation, we have instead

$$ds^2 = \begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} dx \\ dy \end{pmatrix} = \left(\underline{J} \begin{pmatrix} dr \\ d\theta \end{pmatrix} \right)^T \left(\underline{J} \begin{pmatrix} dr \\ d\theta \end{pmatrix} \right) = \begin{pmatrix} dr \\ d\theta \end{pmatrix}^T \underline{J}^T \underline{J} \begin{pmatrix} dr \\ d\theta \end{pmatrix}, \quad (5)$$

where

$$\underline{J}^T \underline{J} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix} \quad (6)$$

and we have used the well known property $(\underline{A} \underline{B})^T = \underline{B}^T \underline{A}^T$ of the matrix transpose which converts columns into rows and vice versa.

The quantity ds^2 is a quadratic function of the coordinate differentials, or a “quadratic form,” usually called the “line element” in differential geometry. The diagonal matrix $g = \underline{J}^T \underline{J}$ consists of the coefficients of this quadratic form, or the components of the “metric.”

But so far no indices! For that we have to introduce superscripted coordinate variables. The old and new coordinates are

$$(x^1, x^2) = (x, y), \quad (y^1, y^2) = (r, \theta) \quad (7)$$

and we can let $i, j, k, \dots = 1, 2$. Then their differential relationship is

$$dx^i = \frac{\partial x^i}{\partial y^j} dy^j \equiv J^i_j dy^j \quad (8)$$

and

$$ds^2 = g_{ij} dy^i dy^j, \quad g_{ij} = (\underline{J}^T \underline{J})_{ij} = \sum_{m=1}^2 J^m_i J^m_j \quad (9)$$

defines the metric expressed in the new coordinates. More precisely, the components of the metric g_{ij} are the entries of the matrix of the quadratic form represented by the “line element”

ds^2 . The only way to get rid of the summation symbol here is to introduce the unit matrix with both indices down to be able to use the summation convention

$$\underline{I} = (\delta_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then we can write

$$g_{ij} = J^m_i J^n_j \delta_{mn} = J^m_i \delta_{mn} J^n_j = (\underline{J}^T \underline{I} \underline{J})_{ij}.$$

In fact since lowered indices are associated with coefficients of linear forms, it makes sense that the coefficients of a bilinear quadratic form $g_{ij}X^iY^j$ also have lowered indices.

In my institution the section of the textbook on changes of variables and the Jacobian matrix is omitted from the multivariable calculus syllabus, but as you can see, it easily determines the differential arclength in non-Cartesian coordinates as in this toy calculation. It also determines the correction factor for the differential of area in the plane in polar coordinates $dA = r dr d\theta$ through its determinant $|\underline{J}| = r$, which we instead alternatively derive using the formula for the area of a sector of a circle. In fact while the dot product and its generalization to a metric determine lengths of line segments and arclengths of curves, the 1-dimensional measure associated with geometry, it is the determinant and its generalizations that allow this 1-dimensional measure to be extended to higher dimensional structures like parallelograms and parallelopipeds and differentials of surface area and volume in non-Cartesian coordinates in \mathbb{R}^3 . In \mathbb{R}^n with $n > 3$, there are $p = 1, 2, \dots, n-1, n$ dimensional structures and measures, all governed by the mathematics of determinants and Jacobians.

Old fashioned tensor calculus deals with understanding differential properties of spaces described by different coordinate systems, giving preference to those quantities which do not depend on the choice of coordinates and therefore have to “transform” in a certain way to guarantee that coordinate independence. The Jacobian is the matrix of the linear transformation of derivatives between different coordinate systems. Calculations with it can become tedious, as we will see when we later study the corresponding derivation for spherical coordinates in space. Fortunately computer algebra systems can now save us from a lot of the hand algebra that becomes quite cumbersome in this subject. For this reason it is important to use either Maple or Mathematica as a tool in learning the ropes of this area of mathematics as one works problems to gain familiarity with the concepts. These software products have slightly different conventions from each other and the classical notation of the discipline, so we have to be a bit flexible in dealing with notation. Modern differential geometry organizes old fashioned tensor calculus in such a way that we deal with invariant objects instead of collections of components which change with the choice of coordinate system, like the notion of a vector as an abstract arrow rather than as a list of components, but those components are still implicit in everything we do, even if we do it with a modern flair.

Exercise 0.0.1.

arclength in the plane

- Evaluate ds^2 for polar coordinates, filling in the lower dots in the above derivation.
- Evaluate g_{ij} for polar coordinates from the matrix product $\underline{J}^T \underline{J}$.

c) If you want to try something unfamiliar, evaluate ds^2 for $x = uv, y = \frac{1}{2}(u^2 - v^2)$. This too results in a sum of squares, which characterizes what are called “orthogonal coordinates” as we will learn about later. These particular orthogonal coordinates in the plane are called parabolic coordinates since the coordinate lines for both coordinates are parabolas. We will return to these much later.

■

Exercise 0.0.2.

matrix multiplication and the trace

The matrix equation defining the inverse of a matrix $\underline{A}^{-1}\underline{A} = \underline{I}$ can be written with our index conventions as $A^{-1i}_j A^j_k = \delta^i_k$, where here we need one index up and one index down on the identity matrix component representation in this context for consistency with our index conventions (the identity matrix plays a different role here than above!). The identity matrix property $\underline{I}\underline{A} = \underline{A}$ translates to $\delta^i_j A^j_k = A^i_k$. Given that the trace is defined by $\text{Tr } \underline{A} = A^i_i$, write out the matrix identity $\text{Tr } (\underline{A}^{-1}\underline{B}\underline{A}) = \text{Tr } \underline{B}$ in our index notation and use the inverse property to show why the left hand side reduces to the right hand side, thus proving the identity easily. Similarly show the more general property of the trace holds: $\text{Tr } (\underline{A}\underline{B}\underline{C}) = \text{Tr } (\underline{C}\underline{A}\underline{B})$, etc., for cyclic permutations of the factor matrices. These are simple examples of how this streamlined index notation for linear behavior easily allows one to prove identities that otherwise are not obvious.

■

Chapter 1

Foundations of tensor algebra

1.1 Index conventions

We need an efficient abbreviated notation to handle the complexity of mathematical structure before us. We will use indices of a given “type” to denote all possible values of given index ranges. By index type we mean a collection of similar letter types, like those from the beginning or middle of the Latin alphabet, or Greek letters

$$a, b, c, \dots$$

$$i, j, k, \dots$$

$$\alpha, \beta, \gamma, \dots$$

each index of which is understood to have a given common range of successive integer values. Variations of these might be barred or primed letters or capital letters. For example, suppose we are looking at linear transformations between \mathbb{R}^n and \mathbb{R}^m where $m \neq n$. We would need two different index ranges to denote vector components in the two vector spaces of different dimensions, say $i, j, k, \dots = 1, 2, \dots, n$ and $\alpha, \beta, \gamma, \dots = 1, 2, \dots, m$.

In order to introduce the so called Einstein summation convention, we agree to the following limitations on how indices may appear in formulas. A given index letter may occur only once in a given term in an expression (call this a “free index”), in which case the expression is understood to stand for the set of all such expressions for which the index assumes its allowed values, or it may occur twice but only as a superscript-subscript pair (one up, one down) which will stand for the sum over all allowed values (call this a “repeated index”). Here are some examples. If $i, j = 1, \dots, n$ then

$$A^i \longleftrightarrow n \text{ expressions : } A^1, A^2, \dots, A^n,$$

$$A^i_i \longleftrightarrow \sum_{i=1}^n A^i_i, \text{ a single expression with } n \text{ terms}$$

(this is called the trace of the matrix $\underline{A} = (A^i_j)$),

$$A^{ji}_i \longleftrightarrow \sum_{i=1}^n A^{1i}_i, \dots, \sum_{i=1}^n A^{ni}_i, n \text{ expressions each of which has } n \text{ terms in the sum,}$$

$$A_{ii} \longleftrightarrow \text{no sum, just an expression for each } i, \text{ if we want to refer to a specific diagonal component (entry) of a matrix, for example,}$$

$$A^i v_i + A^i w_i = A^i(v_i + w_i), 2 \text{ sums of } n \text{ terms each (left) or one combined sum (right).}$$

A repeated index is a “dummy index,” like the dummy variable in a definite integral $\int_a^b f(x)dx = \int_a^b f(u)du$. We can change them at will: $A^i_i = A^j_j$.

1.2 A vector space V

Let V be an n -dimensional real vector space. Elements of this space are called “vectors.” Ordinary real numbers (let \mathbb{R} denote the set of real numbers) will be called “scalars” and denoted by a, b, c, \dots , while vectors will be denoted by various symbols depending on the context: u, v, w or $u_{(1)}, u_{(2)}, \dots$, where here the parentheses indicate that the subscripts are only labeling the vectors in an ordered set of vectors, to distinguish them from component indices. Sometimes X, Y, Z, W are convenient vector symbols.

The basic structure of a real vector space is that it has two operations defined, vector addition and scalar multiplication, which can then be combined together to perform linear combinations of vectors:

- vector addition:
the sum of two vectors $u + v$ is again a vector in the space,
- scalar multiplication:
the product cu of a scalar c and a vector u is again a vector in the space, called a scalar multiple of the vector,

so that linear combinations of two or more vectors with scalar coefficients

$$au + bv$$

are defined. These operations satisfy a list of properties that we take for granted when working with sums and products of real numbers alone, i.e., the set of real numbers \mathbb{R} thought of as a 1-dimensional vector space. Every vector space has a zero vector, often denoted by the same symbol as the zero scalar: 0 . It is the additive identity and the result of multiplication by the zero scalar: $u + 0 = u = 0u$.

A basis of V , denoted by $\{e_i\}$, $i = 1, 2, \dots, n$ or just $\{e_i\}$, where it is understood that a “free index” (meaning not repeated and therefore summed over) like the i in this expression will assume all of its possible values, is a linearly independent spanning set for V

1. spanning condition:

Any vector $v \in V$ can be represented as a linear combination of the basis vectors:

$$v = \sum_{i=1}^n v^i e_i = v^i e_i$$

whose coefficients v^i are called the “components” of v with respect to $\{e_i\}$. The index i on v^i labels the components (coefficients), while the index i on e_i labels the basis vectors.

2. linear independence:

If $v^i e_i = 0$, then $v^i = 0$, (i.e., more explicitly if $v = \sum_{i=1}^n v^i e_i = 0$, then $v^i = 0$ for all $i = 1, 2, \dots, n$).

Example 1.2.1. the vector spaces \mathbb{R}^n

$V = \mathbb{R}^n = \{u = (u^1, \dots, u^n) = (u^i) \mid u^i \in \mathbb{R}\}$, the space of n -tuples of real numbers with the natural basis $e_1 = (1, 0, \dots, 0), e_2 = (0, 1, \dots, 0), \dots, e_n = (0, 0, \dots, 1)$, which we will refer to as the “standard basis” or “natural basis.” In \mathbb{R}^3 , these basis vectors are customarily denoted by $\mathbf{i}, \mathbf{j}, \mathbf{k}$, and a typical vector $a^1\mathbf{i} + a^2\mathbf{j} + a^3\mathbf{k}$ then becomes the more compact expression $a^i e_i$ if we let $\{e_1, e_2, e_3\} = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$.

When we want to distinguish the vector properties of \mathbb{R}^n from its point properties, we will emphasize the difference by using angle brackets instead of parentheses: $\langle u^1, u^2, u^3 \rangle$. In the context of matrix calculations, this representation of a vector will be understood to be a column matrix

$$\langle u^1, u^2, u^3 \rangle \sim \begin{pmatrix} u^1 \\ u^2 \\ u^3 \end{pmatrix} \equiv \underline{u}.$$

Underlining symbols will remind us that we are dealing with matrix quantities.

As a set of points, \mathbb{R}^n has a natural set of “Cartesian” coordinate functions x^i which pick out the i th entry in an n -tuple, for example on \mathbb{R}^3 : $x^1((a^1, a^2, a^3)) = a^1$, etc. These are linear functions on the space. Interpreting the points as vectors, these coordinate functions pick out the individual components of the vectors with respect to the standard basis. When thought of as column matrices, the standard basis vector e_i is the i th column vector of the identity matrix: $\underline{I} = \langle \underline{e}_1 \mid \dots \mid \underline{e}_n \rangle$.

We will also consider complex vector spaces where real numbers are replaced by complex numbers, leading to \mathbb{C}^n which has the same natural basis as \mathbb{R}^n but complex rather than vector components for elements expressed in this basis. One also consider \mathbb{C}^n to be a $(2n)$ -dimensional real vector space with basis $\{e_k, i e_k\}$. Reinterpreting a complex vector space as a real vector space with twice the dimension is a useful idea for studying certain complex matrix vector spaces.

□

Any two real n -dimensional vector spaces are “isomorphic.” This just means there is some map from one to the other, say $\Phi : V \rightarrow W$, and it does not matter whether the vector operations (vector sum and scalar multiplication, i.e., linear combination which encompasses them both) are done before or after using the map: $\Phi(au + bv) = a\Phi(u) + b\Phi(v)$. The practical implication of this rather abstract statement is that once you establish a basis in any real n -dimensional vector space V , the n -tuples of components of vectors with respect to this basis undergo the usual vector operations in \mathbb{R}^n when the vectors they represent undergo the vector operations in V . For example, the set of at most quadratic polynomial functions in a single variable $ax^2 + bx + c = a(x^2) + b(x) + c(1)$ has the natural basis $\{1, x, x^2\}$ and under linear combination of these functions, the triplet of coordinates (c, b, a) (coefficients ordered by increasing powers) undergo the corresponding linear combination as vectors in \mathbb{R}^3 . We might as well just work in \mathbb{R}^3 to visualize relationships between vectors in the original abstract space.

Exercise 1.2.1.

2×2 matrices as a vector space

The space $gl(n, \mathbb{R})$ of real $n \times n$ matrices is an n^2 -dimensional vector space isomorphic to \mathbb{R}^{n^2} . Each of n rows has n entries, so listing all the entries row after row gives a list of n^2 entries, which gives a point in \mathbb{R}^{n^2} . For example

$$\begin{pmatrix} x^1 & x^2 \\ x^3 & x^4 \end{pmatrix} \leftrightarrow \langle x^1, x^2, x^3, x^4 \rangle$$

shows the correspondence between $gl(2, \mathbb{R})$ and \mathbb{R}^4 , which identifies the natural bases of the two spaces.

The trace of a matrix is the sum of its diagonal entries, in this case $x^1 + x^4$. A tracefree matrix has zero trace, in this case : $x^1 + x^4 = 0$. This is a vector subspace since linear combinations of tracefree matrices are still tracefree (check it!). We can choose a new basis of the whole space consisting of the identity matrix and 3 linearly independent tracefree matrices, for example

$$\underline{E}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \underline{E}_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \underline{E}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \underline{E}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

using the index range $\alpha = 0, 1, 2, 3$ instead of $1, 2, 3, 4$. The latter three matrices $\{\underline{E}_k\}, k = 1, 2, 3$ form a basis of the subspace $sl(2, \mathbb{R})$ of tracefree matrices, to which we can easily specialize by ignoring the pure trace identity matrix \underline{E}_0 . “gl” stands for “general linear” while “sl” stands for “special linear” because of the relation of these spaces to the general linear group of invertible square matrices $GL(n, \mathbb{R})$ and its special linear subgroup $SL(n, \mathbb{R})$ of unit determinant matrices, which we will study later. The first three of these matrices are basis of the subspace of symmetric matrices, while the last is a basis of the subspace of antisymmetric matrices.

The new basis has interesting matrix properties. Let

$$\underline{Y} = y^0 \underline{E}_0 + y^1 \underline{E}_1 + y^2 \underline{E}_2 + y^3 \underline{E}_3.$$

a) Show that the unit determinant matrices when expressed in this basis correspond to a hyperboloid of revolution in \mathbb{R}^4

$$1 = \det \underline{Y} = (y^0)^2 + (y^1)^2 - (y^2)^2 - (y^3)^2.$$

b) Show that setting the trace of the square of the matrix equal to 1 when expressed in this basis corresponds instead to a different kind of hyperboloid

$$1 = \frac{1}{2} \text{Tr} (\underline{Y}^2) = (y^0)^2 - (y^1)^2 + (y^2)^2 + (y^3)^2.$$

c) Show that this basis is orthogonal with respect to the usual inner product on \mathbb{R}^4

$$1 = \frac{1}{2} \text{Tr} (\underline{Y}^T \underline{Y}) = (y^0)^2 + (y^1)^2 + (y^2)^2 + (y^3)^2.$$

This is equivalent to showing that $\text{Tr} \underline{E}_i^T \underline{E}_j \propto \delta_{ij}$. What are the self inner products of these matrices?

This example shows that linear isomorphisms between different vector spaces can sometimes be useful for nonlinear properties of those spaces.

**Exercise 1.2.2.**

2×2 complex matrices as a real vector space $h(2)$

Define the three tracefree Pauli matrices

$$\underline{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \underline{\sigma}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \underline{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and let $\underline{E}_0 = \underline{I}$ as in the previous problem but let $\underline{E}_a = \underline{\sigma}_a$, $a = 1, 2, 3$.

These determine a 4-dimensional real vector space $h(2)$ of complex matrices, namely all real linear combinations of these matrices

$$\underline{X} = x^i \underline{E}_i = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix}.$$

Repeat the previous exercise to see some sign changes. In particular show that

$$\det \underline{X} = - \left(-(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2 \right).$$

**Exercise 1.2.3.**

up to quadratic functions as a vector space

Real polynomial functions of a variable x which are at most of degree 2 are of the form $Q(x) = ax^2 + bx + c = c1 + bx + ax^2$, if we order them by increasing powers so that we can add higher power terms to the right to more easily generalize the degree, where the coefficient vector $\langle c, b, a \rangle$ is a vector in \mathbb{R}^3 . Linear combinations of these polynomials correspond to the same linear combinations of the corresponding coefficient vectors, so these two vector spaces are naturally isomorphic. The ordered basis $\{1, x, x^2\}$ corresponds to the natural basis $\{e_1, e_2, e_3\}$ of \mathbb{R}^3 : $c1 + bx + ax^2 \leftrightarrow ce_1 + be_2 + ae_3$.

By expanding at most quadratic polynomial functions of a variable x in a Taylor series about $x = 1$, one expresses these functions in the new basis $\{(x - 1)^p\}$, $p = 0, 1, 2$, say as $Q(x) = A(x - 1)^2 + B(x - 1) + C(1)$. Express (c, b, a) as linear functions of (C, B, A) by expanding out this latter expression. Then solve these relations for the inverse expressions, giving (C, A, B) as functions of (c, b, a) and express both relationships in matrix form, showing explicitly the coefficient matrices. Alternatively, actually evaluate (C, B, A) in terms of (c, b, a) using the Taylor series expansion technique, which will give the same result through calculus. Make a crude drawing of the three new basis vectors in \mathbb{R}^3 which correspond to the new basis functions, or use technology to draw them.



Exercise 1.2.4. **3×3 antisymmetric matrices and the cross product**

An antisymmetric matrix is a square matrix which reverses sign under the transpose operation: $\underline{A}^T = -\underline{A}$. Any 3×3 antisymmetric matrix has the form

$$\begin{aligned}\underline{A} &= \begin{pmatrix} 0 & -a^3 & a^2 \\ a^3 & 0 & -a^1 \\ -a^2 & a^1 & 0 \end{pmatrix} = a^1 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} + a^2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} + a^3 \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &\equiv a^1 \underline{L}_1 + a^2 \underline{L}_2 + a^3 \underline{L}_3.\end{aligned}$$

The space of all such matrices is a 3-dimensional vector space with basis $\{\underline{L}_i\}$ since it is defined as the span of this set of vectors (hence a subspace of the vector space of 3×3 matrices), and setting the linear combination equal to the zero matrix forces all the coefficients to be zero proving the linear independence of this set of vectors (which is therefore a linear independent spanning set).

a) Show that matrix multiplication of a vector in \mathbb{R}^3 by such a matrix \underline{A} is equivalent to taking the cross product with the corresponding vector $\underline{a} = \langle a^1, a^2, a^3 \rangle$:

$$\underline{A} \underline{b} = \underline{a} \times \underline{b}.$$

Hint: use a computer algebra system!

b) Although the result of two successive cross products $\underline{a} \times (\underline{b} \times \underline{u})$ is not equivalent to a single cross product $\underline{c} \times \underline{u}$, the difference of two such successive cross products is. Confirm the matrix product

$$\underline{A} \underline{B} - \underline{B} \underline{A} = (\underline{a} \times \underline{b})^i \underline{L}_i$$

Then by part a) it follows that

$$(\underline{A} \underline{B} - \underline{B} \underline{A}) \underline{u} = (\underline{a} \times \underline{b}) \times \underline{u},$$

c) Use the matrix distributive law on the left hand side, together with the iteration of part a) for the individual terms, to fill in the one further step which then proves the “vector identity”

$$\underline{a} \times (\underline{b} \times \underline{u}) - \underline{b} \times (\underline{a} \times \underline{u}) = (\underline{a} \times \underline{b}) \times \underline{u}.$$

Comment.

Under this relationship $\langle a^1, a^2, a^3 \rangle \mapsto a^i \underline{L}_i = “\vec{a} \times ”$, the three matrices $\underline{L}_1, \underline{L}_2, \underline{L}_3$ correspond directly to the standard basis vectors of \mathbb{R}^3 often denoted by $\hat{i}, \hat{j}, \hat{k}$, which satisfy the cyclic relations

$$\hat{i} \times \hat{j} = \hat{k}, \quad \hat{j} \times \hat{k} = \hat{i}, \quad \hat{k} \times \hat{i} = \hat{j}.$$

This means that

$$[\underline{L}_2, \underline{L}_3] = \underline{L}_1, \quad [\underline{L}_3, \underline{L}_1] = \underline{L}_2, \quad [\underline{L}_1, \underline{L}_2] = \underline{L}_3.$$

These “angular momentum” equations turn out to be fundamentally important in quantum mechanics, determining the energy levels of electrons in atoms by providing 2 quantum numbers

associated with the electron's total (magnitude of the vector) angular momentum about the nucleus and one of its components, or in electromagnetic radiation theory where these same quantities help classify the radiation itself. We will see in part 2 how they are related to angular momentum.

■

Exercise 1.2.5.

complex numbers as a 2-dimensional real vector space

The field \mathbb{C} of complex numbers is a 2-dimensional real vector space isomorphic to \mathbb{R}^2 through the isomorphism $z = x + iy \leftrightarrow (x, y)$ which associates the ordered basis $\{1, i\}$ with the standard basis $\{e_1, e_2\} = \{(1, 0), (0, 1)\}$. Addition of complex numbers corresponds to addition of the corresponding vectors. Let $\bar{z} = x - iy$ denote the complex conjugate.

a) If $a = a^1 + ia^2$ and $b = b^1 + ib^2$ and we think of vectors in the plane as sitting in \mathbb{R}^3 so we can take their cross product: $\vec{a} = \langle a^1, a^2, 0 \rangle$, $\vec{b} = \langle b^1, b^2, 0 \rangle$. Then with $\vec{e}_3 = \langle 0, 0, 1 \rangle$, show that

$$\bar{a}b = \vec{a} \cdot \vec{b} + i\vec{e}_3 \cdot (\vec{a} \times \vec{b}).$$

b) Recall the notation introduced in Chapter 0 for partial derivatives with respect to the coordinates x^i on \mathbb{R}^n :

$$\frac{\partial}{\partial x^i} = \partial_i.$$

Let $a = \partial_1 + i\partial_2 = \nabla$ and $\vec{a} = \langle \partial_1, \partial_2, 0 \rangle$. Using part a), show that if $\vec{b} = \langle b^1, b^2, 0 \rangle$ is a 3-vector valued function on the plane (i.e., b^1 and b^2 only depend on x^1 and x^2), then

$$\bar{\nabla} b = \text{div } \vec{b} + i\vec{e}_3 \cdot \text{curl } \vec{b} = \vec{\nabla} \cdot \vec{b} + i\vec{e}_2 \cdot \vec{\nabla} \times \vec{b}.$$

Complex numbers have a nice way of combining real vector operations in a much more efficient way than separate real expressions, which can be exploited in differential geometry through what are called spinors and quaternions and other complex “group representations.” For example, calculus operations acting on $e^{ix} = \cos x + i \sin x$ are much simpler than those acting on the individual trig functions which are its real and imaginary parts. This idea generalizes to differential geometry under the name of geometric algebra.

■

A p -dimensional linear subspace of a vector space V can be represented as the set of all possible linear combinations of a set of p linearly independent vectors, and such a subspace results from the solution of a set of linear homogeneous conditions on the variable components of a vector variable expressed in some basis. Thus if $\underline{x} = \langle x^1, \dots, x^n \rangle$ is the column matrix of components of an unknown vector in V with respect to a basis $\{e_i\}$, and \underline{A} is an $m \times n$ matrix of rank m (i.e., the rows are linearly independent), the solution space of $\underline{A}\underline{x} = \underline{0}$ will be a $(p = n - m)$ -dimensional subspace, since $m < n$ independent conditions on n variables leave $n - m$ variables freely specifiable. In \mathbb{R}^3 , these are the lines ($p = 1$) and planes ($p = 2$) through the origin. In higher dimensional \mathbb{R}^n spaces, the $(n - 1)$ -dimensional subspaces are called hyperplanes in analogy with the ordinary planes in the case $n = 3$, and we can refer to p -planes through the origin for the values of p between 2 and $n - 1$.

Elementary linear algebra: solving systems of linear equations

It is worth remembering the basic problem of elementary linear algebra: solving m linear equations in n “unknowns” or “variables” x^i , which is most efficiently handled with matrix notation

$$\begin{array}{rcl} A^1_1 x^1 + \cdots A^1_n x^n & = & b^1 \\ \vdots & & \vdots \\ A^m_1 x^1 + \cdots A^m_n x^n & = & b^m \end{array}, \quad A^i_j x^j = b^i, \quad \underline{A} \underline{x} = \underline{b}.$$

The interpretation of the problem requires a slight shift in emphasis to the n columns $u_{(i)} \in \mathbb{R}^m$ of the coefficient matrix by defining $u^i_{(j)} = A^i_j$ or $\underline{A} = \langle \underline{u}_{(1)} | \cdots | \underline{u}_{(n)} \rangle$. Then this is equivalent to setting a linear combination of these columns equal to the right hand side vector $\underline{b} = \langle b^1, \dots, b^m \rangle \in \mathbb{R}^m$

$$\underline{A} \underline{x} = x^1 u_{(1)} + \cdots x^n u_{(n)} = \underline{b}.$$

If $\underline{b} = 0$, the homogeneous case, this is equivalent to trying to find a linear relationship among the n column vectors, namely a linear combination of them equal to the zero vector whose coefficients are not all zero; then for each nonzero coefficient, one can solve for the vector it multiplies and express it as a linear combination of the remaining vectors in the set. When no such relationship exists among the vectors, they are called linearly independent, otherwise they are called linearly dependent. The span (set of all possible linear combinations) of the set of these column vectors is called the column space $\text{Col}(\underline{A})$ of the coefficient matrix \underline{A} . If $\underline{b} \neq 0$, then the system admits a solution only if \underline{b} belongs to the column space, and is inconsistent if not. If $\underline{b} \neq 0$ and the vectors are linearly independent, then if the solution admits a solution, it is unique. If they are not linearly independent, then the solution is not unique but involves a number of free parameters.

The solution technique is row reduction involving a sequence of elementary row operations of three types: adding a multiple of one row to another row, multiplying a row by a nonzero number, and interchanging two rows. These row operations correspond to taking new independent combinations of the equations in the system, or scaling a particular equation, or changing their order, none of which changes the solution of the system. The row reduced echelon form $\langle \underline{A}_R | \underline{b}_R \rangle$ of the augmented matrix $\langle \underline{A} | \underline{b} \rangle$ leads to an equivalent (“reduced”) system of equations $\underline{A}_R \underline{x} = \underline{b}_R$ which is easily solved. The row reduced echelon form has all the zero rows (if any) at the bottom of the matrix, the leading (first from left to right) entry of each nonzero row is 1, the columns containing those leading 1 entries (the leading columns) have zero entries above and below those leading 1 entries, and finally the pattern of leading 1 entries moves down and to the right, i.e., the leading entry of the next nonzero row is to the right of a preceding leading entry. The leading 1 entries of the matrix are also called the pivot entries, and the corresponding columns, the pivot columns. A pivot consists of the set of “add row” operations which makes the remaining entries of a pivot column zero.

The number of nonzero rows of the reduced augmented matrix is called the rank of the augmented matrix and represents the number of independent equations in the original set. The number of nonzero rows of the reduced coefficient matrix alone is called its rank: $r = \text{rank}(\underline{A}) \leq m$ and equals the number of leading 1 entries in \underline{A}_R , in turn the number of leading 1 columns of \underline{A}_R . The remaining $n - r \geq n - m$ columns are called free columns. This classification

of the columns of the reduced coefficient matrix into leading and free columns is extended to the original coefficient matrix. The associated variables of the system of linear equations then fall into two groups, the leading variables ($r \leq m$ in number) and the free variables ($n - r$ in number), since each variable corresponds to one of the columns of the coefficient matrix. Each leading variable can immediately be solved for in its corresponding reduced system equation and expressed in terms of the free variables, whose values are then not constrained and may take any real values. Setting the $n - r$ free variables equal to arbitrary parameters t^B , $B = 1, \dots, n - r$ leads to a solution in the form

$$x^i = x^i_{(\text{particular})} + t^B v^i_{(B)}$$

The “particular solution” satisfies $\underline{A} \underline{x}_{(\text{particular})} = \underline{b}$, while the remaining part is the general solution of the related homogeneous linear system for which $\underline{b} = 0$, an $(n - r)$ -dimensional subspace $\text{Null}(\underline{A})$ of \mathbb{R}^n called the null space of the matrix \underline{A} , since it consists of those vectors which are taken to zero under multiplication by that matrix.

$$\underline{A}(t^B v^i_{(B)}) = t^B (\underline{A} v^i_{(B)}) = 0.$$

This form of the solution defines a basis $\{v_{(B)}\}$ of the null space since by definition any solution of the homogeneous equations can be expressed as a linear combination of them, and if such a linear combination is zero, every parameter t^B is forced to be zero, so they are linearly independent.

This basis of coefficient vectors $\{v_{(B)}\} \in \mathbb{R}^n$ is really a basis of the space of linear relationships among the original n vectors $\{u_{(1)}, \dots, u_{(n)}\}$, each one representing the coefficients of an independent linear relationship among those vectors: $0 = A^j_i v^i_{(B)} = v^i_{(B)} u^j_{(i)}$. In fact these relationships correspond to the fact that each free column of the reduced matrix can be expressed as a linear combination of the leading columns which precede it going from left to right in the matrix, and in fact the same linear relationships apply to the original set of vectors (since the coefficients x^i of the solution space are the same!). Thus one can remove the free columns from the original set of vectors to get a basis of the column space of the matrix consisting of its r leading columns, so the dimension of the column space is the rank r of the matrix.

By introducing the row space of the coefficient matrix $\text{Row}(\underline{A}) \subset \mathbb{R}^n$ consisting of all possible linear combinations of the rows of the matrix, the row reduction process can be interpreted as finding a basis of this subspace that has a certain characteristic form: the r nonzero rows of the reduced matrix. The dimension of the row space is thus equal to the rank r of the matrix. Each equation of the original system corresponding to each (nonzero) row of the coefficient matrix separately has a solution space which represents a hyperplane in \mathbb{R}^n , namely an $(n - 1)$ -dimensional subspace. Re-interpreting the linear combination of the variables as a dot product with the row vector, in the homogeneous case, these hyperplanes consist of all vectors orthogonal to the original row vector, and the joint solution of all the equations of the system is the subspace which is orthogonal to the entire row space, namely the orthogonal complement of the row space within \mathbb{R}^n . Thus $\text{Null}(\underline{A})$ and $\text{Row}(\underline{A})$ decompose the total space \mathbb{R}^n into an orthogonal decomposition with respect to the dot product, and the solution algorithm for the homogeneous linear system provides a basis of each such subspace.

Left multiplication of \underline{A} by a row matrix of variables $\underline{y}^T = \langle y_1 | \dots | y_m \rangle$ yields a row matrix, so one can consider the transposed linear system in which that product is set equal to a constant row vector $\underline{c}^T = \langle c_1 | \dots | c_m \rangle$

$$\underline{y}^T \underline{A} = \underline{c}^T, \quad \text{or} \quad \underline{A}^T \underline{y} = \underline{c}$$

This is the linear system of equations associated with the transpose of the matrix, which interchanges rows and columns and hence the row space and column space

$$\text{Row}(\underline{A}^T) = \text{Col}(\underline{A}), \quad \text{Col}(\underline{A}^T) = \text{Row}(\underline{A}),$$

but adds one more space $\text{Null}(\underline{A}^T)$, which can be interpreted as the subspace orthogonal to $\text{Row}(\underline{A}^T) = \text{Col}(\underline{A})$, hence determining an orthogonal decomposition of \mathbb{R}^m as well.

Example 1.2.2. concrete example of row reduction and linear system solution

Here is the augmented matrix and its row reduced echelon form for 5 equations in 7 unknowns

$$\langle \underline{A} | \underline{b} \rangle = \begin{pmatrix} -1 & 2 & 4 & 11 & 0 & -4 & 1 & 16 \\ 1 & -2 & 1 & 4 & 0 & -2 & 0 & 5 \\ 0 & 0 & -4 & -12 & 0 & 2 & 4 & 12 \\ -3 & 6 & -4 & -15 & 0 & 2 & -4 & -42 \\ -4 & 8 & -1 & -7 & 0 & -1 & 3 & 7 \end{pmatrix}, \quad \langle \underline{A}_R | \underline{b}_R \rangle = \begin{pmatrix} 1 & -2 & 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 3 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and its solution

$$\underline{x} = \begin{pmatrix} 2 + 2t^1 - t^2 \\ t^1 \\ 3 - 3t^2 \\ t^2 \\ t^3 \\ 0 \\ 6 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 3 \\ 0 \\ 0 \\ 0 \\ 6 \end{pmatrix} + t^1 \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t^2 \begin{pmatrix} -1 \\ 0 \\ -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t^3 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \underline{x}_{(\text{particular})} + t^B \underline{v}_{(B)}.$$

The rank of the 5×7 coefficient matrix (and of the 5×8 augmented matrix) is $r = 4$ with 4 leading variables $\{x^1, x^3, x^6, x^7\}$ and 3 free variables $\{x^2, x^4, x^5\}$. By inspection one sees that the 2nd, 4th, and 5th columns are linear combinations of the preceding leading columns with coefficients which are exactly the entries of those columns. The same linear relationships apply to the original matrix, so columns 1,3,6,7 of the coefficient matrix $\underline{A} = \langle \underline{u}_1 | \dots | \underline{u}_7$, namely $\{\underline{u}_1, \underline{u}_3, \underline{u}_6, \underline{u}_7\}$, are a basis of the column space $\text{Col}(\underline{A}) \subset \mathbb{R}^5$. The 4 nonzero rows of the reduced coefficient matrix \underline{A}_R are a basis of the row space $\text{Row}(\underline{A}) \subset \mathbb{R}^7$. The three columns $\{\underline{v}_{(1)}, \underline{v}_{(2)}, \underline{v}_{(3)}\}$ appearing in the solution vector \underline{x} multiplied by the arbitrary parameters $\{t_1, t_2, t_3\}$ are a basis of the homogeneous solution space $\text{Null}(\underline{A}) \subset \mathbb{R}^7$. Together these 7 vectors form a basis of \mathbb{R}^7 .

One concludes that the right hand side vector $\underline{b} \in \mathbb{R}^5$ can be expressed in the form

$$\underline{b} = x^i \underline{u}_{(i)} = x^i_{(\text{particular})} \underline{u}_{(i)} + t^B \underline{v}_{(B)} \underline{u}_{(i)} = x^i_{(\text{particular})} \underline{u}_{(i)} = 2\underline{u}_{(1)} + 3\underline{u}_{(3)} + 6\underline{u}_{(7)}$$

since the homogeneous part of the solution forms the zero vector from its linear combination of the original columns. Notice that the fifth column $u_{(5)} = 0$; the zero vector makes any set of vectors trivially linearly dependent, so t^3 is a trivial parameter and $v_{(3)}$ represents that trivial linear relationship. Thus there are only two independent relationships among the 6 nonzero columns of \underline{A} .

The row space $\text{Row}(\underline{A}) = \text{Col}(\underline{A}^T)$ is a 4-dimensional subspace of \mathbb{R}^5 . If one row reduces the 7×5 transpose matrix \underline{A}^T , the 4 nonzero rows of the reduced matrix are a basis of this space, and one finds one free variable and a single basis vector

$$\langle -258, 166, -165, -96, 178 \rangle / 178$$

for the 1-dimensional subspace $\text{Null}(\underline{A}^T)$, which is the orthogonal subspace to the 4-dimensional subspace $\text{Row}(\underline{A}) \subset \mathbb{R}^5$.

Don't worry. We will not need the details of row and column spaces in what follows, so if your first introduction to linear algebra stopped short of this topic, don't despair. \square

Example 1.2.3. solving linear systems for unknown matrices

We can also consider multiple linear systems with the same coefficient matrix. For example consider the two linearly independent vectors $X_{(1)} = \langle 1, 3, 2 \rangle$, $X_{(2)} = \langle 2, 3, 1 \rangle$ which span a plane through the origin in \mathbb{R}^3 and let

$$\underline{X} = \langle X_{(1)} | X_{(2)} \rangle = \begin{pmatrix} 1 & 2 \\ 3 & 3 \\ 2 & 1 \end{pmatrix}.$$

Clearly the sum $X_{(1)} + X_{(2)} = \langle 3, 6, 3 \rangle$ and difference $X_{(2)} - X_{(1)} = \langle 1, 0, -1 \rangle$ vectors are a new basis of the same subspace (since they are not proportional) so if we try to express each of them in turn as linear combinations of the original basis vectors, we know already the unique solutions for each

$$\begin{aligned} u^1 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + u^2 \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 & 2 \\ 3 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \\ 3 \end{pmatrix}, & v^1 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + v^2 \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 & 2 \\ 3 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, & \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} &= \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \end{aligned}$$

Clearly from the definition of matrix multiplication, we can put these two linear systems together as

$$\begin{pmatrix} 1 & 2 \\ 3 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} u^1 & v^1 \\ u^2 & v^2 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 6 & 0 \\ 3 & -1 \end{pmatrix},$$

which has the form $\underline{X}\underline{Z} = \underline{Y}$ where \underline{X} is the 3×2 coefficient matrix, \underline{Y} is the 3×2 right hand side matrix, and \underline{Z} is the unknown 2×2 matrix whose columns tell us how to express

the vectors $Y_{(1)}, Y_{(2)}$ as linear combinations of the vectors $X_{(1)}, X_{(2)}$. Of course here we know the unique solution is

$$\underline{Z} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix},$$

a matrix which together with its inverse can be used to transform the components of vectors from one basis to the other.

In other words it is sometimes useful to generalize the simple linear system $\underline{A}\underline{x} = \underline{b}$ to an unknown matrix \underline{X} of more than one column $\underline{A}\underline{X} = \underline{B}$ when the right hand side matrix is more than one column

$$\underbrace{\underline{A}}_{m \times n} \underbrace{\underline{X}}_{n \times p} = \underbrace{\underline{B}}_{m \times p}.$$

□

Elementary linear algebra: the eigenvalue problem and linear transformations

The next step in elementary linear algebra is to understand how a square $n \times n$ matrix acts on \mathbb{R}^n by matrix multiplication as linear transformation of the space into itself

$$\underline{x} \rightarrow \underline{A}\underline{x}, \quad x^i \rightarrow A^i_j x^j$$

which maps each vector \underline{x} to the new location $\underline{A}\underline{x}$. Under this mapping the standard basis vectors e_i are mapped to the new vectors $\underline{A}e_i$, each of which can be expressed as a unique linear combination of the basis vectors with coefficients A^j_i , hence the index notation

$$e_i \rightarrow \underline{A}e_i = e_j A^j_i$$

which makes those coefficients for each value of i into the columns of the matrix \underline{A} . To understand how this matrix multiplication moves around the vectors in the space, one looks for special directions (“eigendirections”) along which matrix multiplication reduces to scalar multiplication, i.e., subspaces along which the direction of the new vectors remains parallel to their original directions (although they might reverse direction)

$$\underline{A}\underline{x} = \lambda \underline{x}, \quad \underline{x} \neq 0,$$

which defines a proportionality factor λ called the “eigenvalue” associated with the “eigenvector” \underline{x} , which must be nonzero to have a direction to speak about. This eigenvector condition is equivalent to

$$(\underline{A} - \lambda \underline{I})\underline{x} = \underline{A}\underline{x} - \lambda \underline{x} = 0.$$

In order for the square matrix $\underline{A} - \lambda \underline{I}$ to admit nonzero solutions it must row reduce to a matrix which has at least one free variable and hence at least one zero row, and hence zero determinant, so a necessary condition for finding an eigenvector is that the “characteristic equation” is satisfied by the eigenvalue

$$\det(\underline{A} - \lambda \underline{I}) = 0.$$

The roots of this n th degree polynomial are the eigenvalues of the matrix, and once found can be separately backsubstituted into the linear system to find the solution space which defines the corresponding eigenspace. The row reduction procedure provides a default basis of this eigenspace, i.e., a set of linearly independent eigenvectors for each eigenvalue.

It is easily shown that eigenvectors corresponding to distinct eigenvalues are linearly independent so this process leads to a basis of the subspace of \mathbb{R}^n spanned by all these eigenspace bases. If they are n in number, this is a basis of the whole space and the matrix can be diagonalized. Let $\underline{B} = \langle \underline{b}_1 | \dots | \underline{b}_n \rangle$ be the matrix whose columns are such an eigenbasis of \mathbb{R}^n , with $\underline{A} \underline{b}_i = \lambda_i \underline{b}_i$. In other words define $B^j_i = b^j_i$ as the j th component of the i th eigenvector. Then

$$\underline{A} \underline{B} = \langle \underline{A} \underline{b}_1 | \dots | \underline{A} \underline{b}_n \rangle = \langle \lambda_1 \underline{b}_1 | \dots | \lambda_n \underline{b}_n \rangle = \langle \underline{b}_1 | \dots | \underline{b}_n \rangle \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix},$$

where the latter diagonal matrix multiplies each column by its corresponding eigenvalue, so that

$$\underline{B}^{-1} \underline{A} \underline{B} = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix} \equiv \underline{A}_B$$

is a diagonal matrix whose diagonal elements are the eigenvalues listed in the same order as the corresponding eigenvectors. Thus (multiplying this equation on the left by \underline{B} and on the right by \underline{B}^{-1}) the matrix \underline{A} can be represented in the form

$$\underline{A} = \underline{B} \underline{A}_B \underline{B}^{-1}.$$

This matrix transformation has a simple interpretation in terms of a linear transformation of the Cartesian coordinates of the space, expressing the old coordinates x^i (with respect to the standard basis) as linear combinations of the new basis vectors \underline{b}_j whose coefficients are the new coordinates $x^i = y^j b^i_j = B^i_j y^j$, which takes the matrix form

$$\begin{aligned} \underline{x} &= \underline{B} \underline{y}, & x^i &= B^i_j y^j, \\ \underline{y} &= \underline{B}^{-1} \underline{x}, & y^i &= B^{-1 i}_j x^j, \end{aligned}$$

The top line expresses the old coordinates as linear functions of the new Cartesian coordinates y^i . Inverting this relationship by multiplying both sides of the top matrix equation by \underline{B}^{-1} , one arrives at the bottom line, which instead expresses the new coordinates as linear functions of the old coordinates. Then under matrix multiplication of the old coordinates by \underline{A} , namely $\underline{x} \rightarrow \underline{A} \underline{x}$, the new coordinates are mapped according to

$$\begin{aligned} x^i &\rightarrow A^i_k x^k \\ y^i &= B^{-1 i}_j x^j \rightarrow B^{-1 i}_j (A^j_k x^k) = B^{-1 i}_j A^j_k B^k_m y^m = [A_B]^i_m y^m, \end{aligned}$$

so \underline{A}_B is just the new matrix of the linear transformation with respect to the new basis of eigenvectors. In the eigenbasis, matrix multiplication is reduced to distinct scalar multiplications along each eigenvector, which may be interpreted as a contraction $0 \leq \lambda_i < 1$ or a

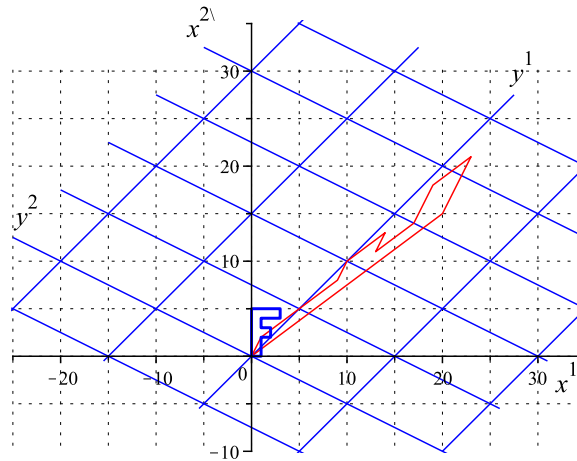


Figure 1.1: The action of a linear transformation on a figure shown with a grid adapted to the new basis of eigenvectors. Vectors are stretched by a factor 5 along the y^1 direction and reflected across that direction along the y^2 direction.

stretch $1 < \lambda_i$ (but no change if $\lambda_i = 1$) combined with a change in direction (reflection) if the eigenvalue is negative $\lambda_i < 0$. Not all square matrices can be diagonalized in this way. For example, rotations occur in the interesting case in which one cannot find enough independent (real) eigenvectors to form a complete basis, but correspond instead to complex conjugate pairs of eigenvectors.

Don't worry. We will not need to deal with the eigenvector problem in most of what follows, except in passing for symmetric matrices $\underline{A} = \underline{A}^T$ which can always be diagonalized by an orthogonal matrix \underline{B} . However, the change of basis example is fundamental to everything we will do. When eigenvector analysis is needed or helpful, computer algebra systems give us the results effortlessly.

Example 1.2.4. eigenvalues and stretch/contraction interpretation

Consider the matrix

$$\underline{A} = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix} = \underline{B} \underline{A}_B \underline{B}^{-1}, \quad \underline{A}_B = \begin{pmatrix} 5 & 0 \\ 0 & -1 \end{pmatrix}, \quad \underline{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix}, \quad \underline{B}^{-1} = \frac{1}{3} \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}.$$

Under matrix multiplication by \underline{A} , the first eigenvector $\underline{b}_1 = \langle 1, 1 \rangle$ is stretched by a factor of 5 while the second one $\underline{b}_2 = \langle -2, 1 \rangle$ is reversed in direction. As shown in Fig. 1.1, this reflects the letter F across the y^1 axis and then stretches it in the y^1 direction by a factor of 5.

□

1.3 The dual space V^*

Let V^* be the “dual space” of V , just a fancy name for the space of real-valued linear functions on V ; elements of V^* are called “covectors.” These are also referred to as “1-forms” in the same sense that one sometimes speaks of a “linear form” or a “quadratic form” on a vector space. The condition of linearity is

$$\textbf{linearity condition: } f \in V^* \longrightarrow f(au + bv) = af(u) + bf(v),$$

or in words: “the value on a linear combination = the linear combination of the values.” This easily extends to linear combinations with any number of terms; for example

$$f(v) = f\left(\sum_{i=1}^N v^i e_i\right) = \sum_{i=1}^N v^i f(e_i)$$

where the coefficients $f_i \equiv f(e_i)$ are the “components” of a covector with respect to the basis $\{e_i\}$, or in our shorthand notation

$$\begin{aligned} f(v) &= f(v^i e_i) && \text{(express in terms of basis)} \\ &= v^i f(e_i) && \text{(linearity)} \\ &= v^i f_i. && \text{(definition of components)} \end{aligned}$$

A covector f is entirely determined by its values f_i on the basis vectors, namely its components with respect to that basis.

Our linearity condition is usually presented separately as a pair of separate conditions on the two operations which define a vector space:

- sum rule: the value of the function on a sum of vectors is the sum of the values, $f(u+v) = f(u) + f(v)$,
- scalar multiple rule: the value of the function on a scalar multiple of a vector is the scalar times the value on the vector, $f(cu) = cf(u)$.

Example 1.3.1. linear homogeneous functions!

In the usual calculus notation on \mathbb{R}^3 , with Cartesian coordinates $(x^1, x^2, x^3) = (x, y, z)$, linear functions are of the form $f(x, y, z) = ax + by + cz$, but a function with an extra additive term $g(x, y, z) = ax + by + c + d$ is called “linear” as well. Only linear *homogeneous* functions (no additive term) satisfy the basic linearity property $f(au + bv) = af(u) + bf(v)$. Unless otherwise indicated, the term “linear” here will always be intended in its narrow meaning of “linear homogeneous.”

Warning:

In this example, the “variables” (x, y, z) in the defining statement $f(x, y, z) = ax + by + cz$ are simply place holders for any three real numbers in the equation, while the Cartesian coordinate functions denoted by the same symbols are instead the names of three independent (linear)

functions on the vector space whose values on any triplet of numbers are just the corresponding number from the triplet: $y(1, 2, 3) = 2$, for example. To emphasize that it is indeed a function of the vector $u = (1, 2, 3)$, we might also write this as $y(u) = y((1, 2, 3)) = 2$ or even $y(\langle 1, 2, 3 \rangle) = 2$ if we adopt the vector delimiters \langle, \rangle instead of the point delimiters $(,)$. Notation is extremely important in conveying mathematical meaning, but we only have so many symbols to go around, so flexibility in interpretation is also required. \square

The dual space V^* is itself an n -dimensional vector space, with linear combinations of covectors defined in the usual way that one can take linear combinations of any functions, i.e., in terms of values

covector addition: $(af + bg)(v) \equiv af(v) + bg(v), \quad f, g \text{ covectors, } v \text{ a vector.}$

Exercise 1.3.1.

closure of the dual space

Show that this defines a linear function $af + bg$, so that the space is closed under this linear combination operation. [All the other vector space properties of V^* are inherited from the linear structure of V .] In other words, show that if f, g are linear functions, satisfying our linearity condition, then $af + bg$ also satisfies the linearity condition for linear functions:

$$(af + bg)(c_1u + c_2v) = c_1(af + bg)(u) + c_2(af + bg)(v).$$

■

Let us produce a basis for V^* , called the dual basis $\{\omega^i\}$ or “the basis dual to $\{e_i\}$,” by defining n covectors which satisfy the following “duality relations”

$$\omega^i(e_j) = \delta^i_j \equiv \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

where the symbol δ^i_j is called the “Kronecker delta,” nothing more than a symbol for the components of the $n \times n$ identity matrix $\underline{I} = (\delta^i_j)$. We then extend them to any other vector by linearity. Then by linearity

$$\begin{aligned} \omega^i(v) &= \omega^i(v^j e_j) && \text{(expand in basis)} \\ &= v^j \omega^i(e_j) && \text{(linearity)} \\ &= v^j \delta^i_j && \text{(duality)} \\ &= v^i && \text{(Kronecker delta definition)} \end{aligned}$$

where the last equality follows since for each i , only the term with $j = i$ in the sum over j contributes to the sum. Alternatively matrix multiplication of a vector on the left by the identity matrix $\delta^i_j v^j = v^i$ does not change the vector. Thus the calculation shows that the i -th dual basis covector ω^i picks out the i -th component v^i of a vector v .

Notice that a Greek letter has been introduced for the covectors ω^i partially following a convention that distinguishes vectors and covectors using Latin and Greek letters, but this convention is obviously incompatible with our more familiar calculus notation in which f denotes a function, so we limit it to our conventional symbol for the dual basis associated with a starting basis $\{e_i\}$.

Why do the n covectors $\{\omega^i\}$ form a basis of V^* ? We can easily show that the two conditions for a basis are satisfied.

1. spanning condition:

Using linearity and the definition $f_i = f(e_i)$, this calculation shows that every linear function f can be written as a linear combination of these covectors

$$\begin{aligned}
 f(v) &= f(v^i e_i) && \text{(expand in basis)} \\
 &= v^i f(e_i) && \text{(linearity)} \\
 &= v^i f_i && \text{(definition of components)} \\
 &= v^i \delta^j_i f_j && \text{(Kronecker delta definition)} \\
 &= v^i \omega^j(e_i) f_j && \text{(dual basis definition)} \\
 &= (f_j \omega^j)(v^i e_i) && \text{(linearity)} \\
 &= (f_j \omega^j)(v). && \text{(expansion in basis, in reverse)}
 \end{aligned}$$

Thus f and $f_i \omega^i$ have the same value on every $v \in V$ so they are the same function: $f = f_i \omega^i$, where $f_i = f(e_i)$ are the “components” of f with respect to the basis $\{\omega^i\}$ of V^* also said to be the “components” of f with respect to the basis $\{e_i\}$ of V already introduced above. The index i on f_i labels the components of f , while the index i on ω^i labels the dual basis covectors.

2. linear independence:

Suppose $f_i \omega^i = 0$ is the zero covector. Then evaluating each side of this equation on e_j and using linearity

$$\begin{aligned}
 0 &= 0(e_j) && \text{(zero scalar = value of zero linear function)} \\
 &= (f_i \omega^i)(e_j) && \text{(expand zero vector in basis)} \\
 &= f_i \omega^i(e_j) && \text{(definition of linear combination function value)} \\
 &= f_i \delta^i_j && \text{(duality)} \\
 &= f_j && \text{(Kronecker delta definition)}
 \end{aligned}$$

forces all the coefficients of ω^i to vanish, i.e., no nontrivial linear combination of these covectors exists which equals the zero covector (the existence of which would be a linear relationship among them) so these covectors are linearly independent. Thus V^* is also an n -dimensional vector space. [A nontrivial linear combination has at least one nonzero coefficient. If all the coefficients are 0, the linear combination is not very interesting!]

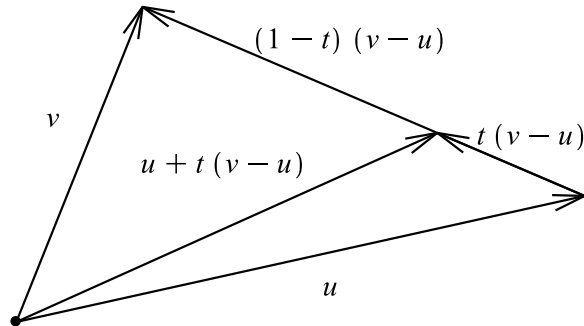


Figure 1.2: Interpolating between two points via their position vectors.

Example 1.3.2. Cartesian coordinates are the standard dual basis of \mathbb{R}^n

The familiar Cartesian coordinates on \mathbb{R}^n are defined by

$$x^i((u^1, \dots, u^n)) = u^i \quad (\text{value of } i\text{-th number in } n\text{-tuple}).$$

But this is exactly what the basis $\{\omega^i\}$ dual to the natural basis $\{e_i\}$ does—i.e., the set of Cartesian coordinates $\{x^i\}$, interpreted as linear functions on the vector space \mathbb{R}^n (why are they linear?), is the dual basis: $\omega^i = x^i$. A general linear function on \mathbb{R}^n has the familiar form $f = f_i \omega^i = f_i x^i$. The components of the linear function (covector) are just its coefficients.

If we return to \mathbb{R}^3 and calculus notation where a general linear function has the form $f = ax + by + cz$, then all we are doing is abstracting the familiar relations

$$\begin{pmatrix} \omega^1(e_1) & \omega^1(e_2) & \omega^1(e_3) \\ \omega^2(e_1) & \omega^2(e_2) & \omega^2(e_3) \\ \omega^3(e_1) & \omega^3(e_2) & \omega^3(e_3) \end{pmatrix} = \begin{pmatrix} x(1, 0, 0) & x(0, 1, 0) & x(0, 0, 1) \\ y(1, 0, 0) & y(0, 1, 0) & y(0, 0, 1) \\ z(1, 0, 0) & z(0, 1, 0) & z(0, 0, 1) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

for the values of the Cartesian coordinates on the standard basis unit vectors along the coordinate axes, making the three simple linear functions $\{x, y, z\}$ a dual basis to the standard basis. The standard basis vectors $\{e_1, e_2, e_3\}$ are often designated by $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$ with or without “hats” (the physics notation to indicate unit vectors).

Note that linearity of a function can be interpreted in terms of linear interpolation of intermediate values of the function. Given any two points u, v in \mathbb{R}^n , then the set of points $u + t(v - u) = (1 - t)u + tv$ from $t = 0$ to $t = 1$ is the directed line segment between from u to v . Then the linearity condition $f((1 - t)u + tv) = (1 - t)f(u) + tf(v)$ says that the value of the function at a certain fraction of the way from u to v is exactly that fraction of the way between the values of the function at those two points.

□

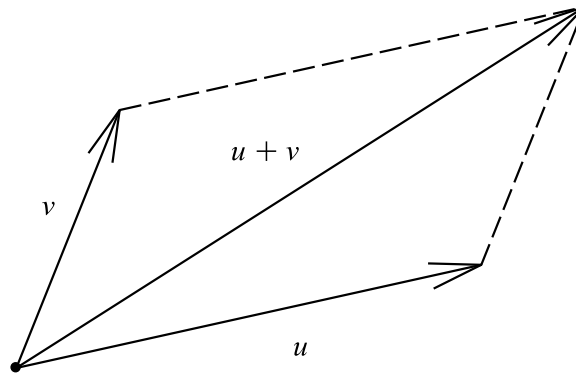


Figure 1.3: Vector addition: main diagonal of parallelogram, same as tip to tail in either order.

Vectors and vector addition are best visualized by interpreting points in \mathbb{R}^n as directed line segments from the origin (“arrows”). Fig. 1.3 illustrates the usual parallelogram addition law for vectors, which shows that putting the two vectors tip to tail in either order yields the same final point, whose position vector corresponds to the main diagonal of the parallelogram naturally formed by the two vectors by translating each to the tip of the other.

Functions can instead be visualized in terms of their level surfaces. For linear functions these level surfaces $f(x) = f_i x^i = t$ (t , a parameter) are a family of parallel hyperplanes, best represented by selecting an equally spaced set of such hyperplanes, say by choosing integer values of the parameter t . However, it is enough to graph two such level surfaces $f(x) = 0$ and $f(x) = 1$ to have a mental picture of the entire family since they completely determine the orientation and separation of all other members of this family. This pair of planes also enables one to have a geometric interpretation of covector addition on the vector space itself, like the parallelogram law for vectors. However, instead of the directed main diagonal line segment, one has the cross diagonal hyperplane for the result. By the way, a hyperplane in \mathbb{R}^n is just the solution of a single linear equation on the space (an $(n - 1)$ -dimensional set of points), generalizing the planes of \mathbb{R}^3 . In \mathbb{R}^2 these solution sets reduce to lines since $n - 1 = 1$ for $n = 2$.

Let’s look at two pairs of such hyperplanes representing f and g but “edge on,” namely in the 2-plane orthogonal to the $(n - 2)$ -plane of intersection of the two $(n - 1)$ -planes which are these hyperplanes. The intersection of two nonparallel hyperplanes, each of which represents the solution of a single linear homogeneous condition on n variables, represents the solution of two independent conditions on n variables, and hence must be an $(n - 2)$ -dimensional plane.

This is easier to see if we are more concrete. Figs 1.5 and 1.6 illustrate this in three dimensions. The first figure looking at the intersecting planes edge on down the lines of intersection is actually the two-dimensional example, where it is clear that the cross-diagonal intersection points of the two pairs of lines must both belong to the line $(f + g)(x) = 1$ on which the sum covector has the value $1 = 0 + 1 = 1 + 0$. The second line of the pair $(f + g)(x) = 0$ needed to

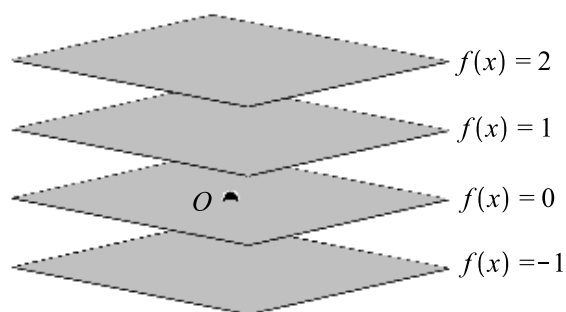


Figure 1.4: Geometric representation of a covector or 1-form f (just a linear function on \mathbb{R}^3): the representative planes of function values 0 and 1 are enough to capture its orientation and magnitude, the latter of which is directly correlated with the density of spacing of these planes and therefore inversely correlated with the distance between them. The same mental image works for the corresponding hyperplanes in \mathbb{R}^n (lines in \mathbb{R}^2).

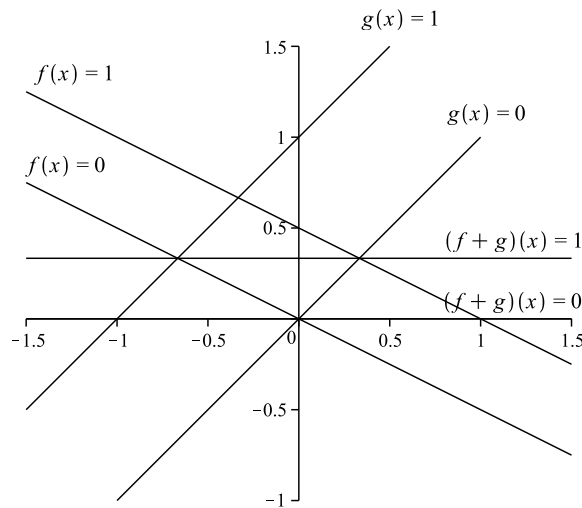


Figure 1.5: Covector addition seen edge-on in \mathbb{R}^3 . The plane $(f + g)(x) = 1$ representing the addition of two covectors is the plane through the lines of intersection of the cross-diagonal of the parallelogram formed by the intersection of the two pairs of planes when seen edge-on down the lines of intersection. Moving that plane parallel to itself until it passes through the origin gives the second plane of the pair representing the sum covector. This same diagram directly illustrates covector addition in \mathbb{R}^2 .

represent the sum covector is the parallel line passing through the origin. If we now rotate our point of view away from the edge-on orientation, we get the picture depicted in Fig. 1.6, which looks like a honeycomb of intersecting planes, with the cross-diagonal plane of intersection representing the sum covector.

Of course the dual space $(\mathbb{R}^n)^*$ is isomorphic to \mathbb{R}^n

$$f = f_i \omega^i = f_i x^i \in (\mathbb{R}^n)^* \longleftrightarrow f^b \equiv (f_i) = (f_1, \dots, f_n) \in \mathbb{R}^n,$$

where the flat symbol notation reminds us that a correspondence has been established between two different objects (effectively lowering the component index), and since $(\mathbb{R}^n)^*$ is a vector space itself, covector addition is just the usual parallelogram vector addition there. However, the above hyperplane interpretation of the dual space covector addition occurs on the original vector space!

These same pictures apply to any finite dimensional vector space. The difference in geometrical interpretation between directed line segments and directed hyperplane pairs is one reason for carefully distinguishing V from V^* by switching index positioning.

For \mathbb{R}^n the distinction between n -tuples of numbers which are vectors and (the component n -tuples of) covectors is still made using matrix notation. Vectors in \mathbb{R}^n are identified with column matrices and covectors in the dual space with row matrices

$$u = (u^1, \dots, u^n) \longleftrightarrow \begin{pmatrix} u^1 \\ \vdots \\ u^n \end{pmatrix},$$

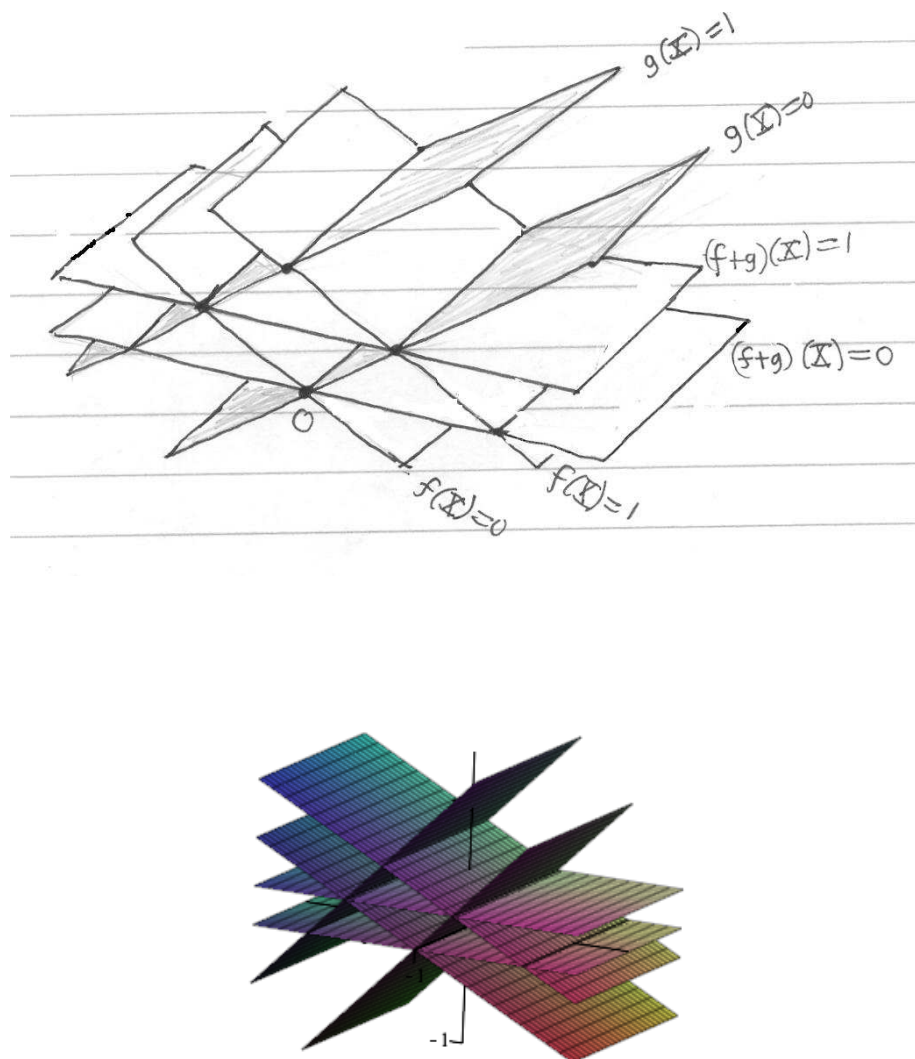


Figure 1.6: Covector addition in R^3 no longer seen edge-on. One has a honeycomb of intersecting planes, with the sum covector represented by the “cross-diagonal” plane of intersection and its parallel companion through the origin.

$$f = f_i \omega^i \longleftrightarrow f^\flat \equiv (f_1, \dots, f_n) \longleftrightarrow (f_1 \dots f_n) \text{ [no commas here]},$$

which we will sometimes designate respectively by $\langle u^1, \dots, u^n \rangle$ and $\langle f_1 | \dots | f_n \rangle$ to emphasize the vector/covector column/row matrix dual interpretation of the n -tuple of numbers. The natural evaluation of a covector on a vector then corresponds to matrix multiplication

$$f(u) = f_i u^i = (f_1 \dots f_n) \begin{pmatrix} u^1 \\ \vdots \\ u^n \end{pmatrix}.$$

This evaluation of a covector (represented by a row matrix on the left) on a vector (represented by a column matrix on the right), which is just the value of the linear function $f = f_i x^i$ at the point with position vector u , is a matrix product of two *different* objects, although it can be represented in terms of the usual dot product on \mathbb{R}^n of two vectors (*like* objects)

$$f(u) = f_i u^i = f^\sharp \cdot u \equiv \sum_{i=1}^n f^\sharp_i u^i = \underline{f}^T \underline{u},$$

but the relationship between the covector $f \in (\mathbb{R}^n)^*$ and its vector of components $f^\sharp = (f_1, \dots, f_n) \in \mathbb{R}^n$ involves *additional* mathematical structure, that of an inner product on \mathbb{R}^n , which is associated with the Euclidean geometry we all know and love. We will get to this later. In terms of matrix notation, this sharp map \sharp is just the transpose operation, which converts a covector to a vector by converting a row matrix into a column matrix.

Example 1.3.3. dual basis in the plane

Suppose we consider the basis $\underline{b}_1 = \langle 1, 1 \rangle, \underline{b}_2 = \langle -1, 2 \rangle$ of \mathbb{R}^2 interpreted as column matrices and form the 2×2 matrix using them as the columns in this order, and consider its inverse

$$\underline{B} = \langle \underline{b}_1 | \underline{b}_2 \rangle = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix}, \quad \underline{B}^{-1} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix} \equiv \begin{pmatrix} \underline{w}^{1T} \\ \underline{w}^{2T} \end{pmatrix},$$

whose two rows define the row vectors $\underline{w}^{1T} = \langle 2/3 | -1/3 \rangle, \underline{w}^{2T} = \langle 1/3 | 1/3 \rangle$, which are the coefficient vectors of two linear functions w^1, w^2 . The defining relation for the inverse matrix $\underline{I} = \underline{B}^{-1} \underline{B}$ is explicitly

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} \underline{w}^{1T} \underline{b}_1 & \underline{w}^{1T} \underline{b}_2 \\ \underline{w}^{2T} \underline{b}_1 & \underline{w}^{2T} \underline{b}_2 \end{pmatrix} = \begin{pmatrix} w^1(\underline{b}_1) & w^1(\underline{b}_2) \\ w^2(\underline{b}_1) & w^2(\underline{b}_2) \end{pmatrix},$$

which has the interpretation that the matrix of evaluations of these two linear functions on the two basis vectors is the identity matrix. In other words $\{w^1, w^2\}$ is the basis dual to $\{\underline{b}_1, \underline{b}_2\}$.

This clearly generalizes to an ordered basis of n linearly independent vectors in \mathbb{R}^n : form the basis matrix $\underline{B} = \langle \underline{b}_1 | \dots | \underline{b}_n \rangle$ containing those basis vectors as columns, then the rows of the inverse matrix \underline{B}^{-1} are the components of the corresponding dual basis.

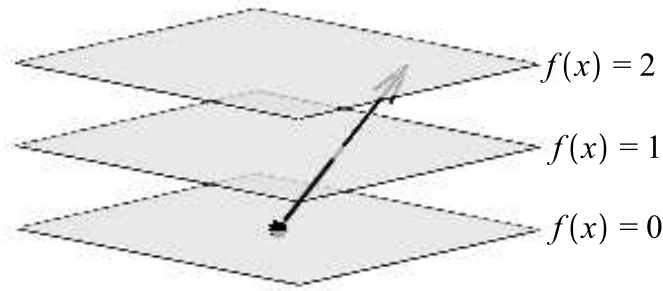


Figure 1.7: Illustration of the condition $f(u) = 2$ for the evaluation of the covector f on the vector u .

□

The evaluation of a covector on a vector also has a geometrical interpretation. If we imagine the 1-parameter family of hyperplanes $f_i x^i = f(x) = t$, then the parameter value t of the evaluation is the “number” of hyperplanes pierced by the arrow representing u , if by “number” we refer to the integer subfamily and we interpolate between them when the result is not an integer. It is exactly this natural evaluation operation which is embodied in the Einstein summation convention for repeated subscript/superscript index pairs, which requires more structure as we will see below. In the case illustrated in Fig. 1.7, the vector pierces exactly 2 of these parallel planes.

Scalar multiplication of vectors and covectors also has a geometrical interpretation. Under scalar multiplication $u \rightarrow cu$, a vector’s length is multiplied by $|c|$, with a direction reversal if $c < 0$, but under scalar multiplication of a covector $f \rightarrow cf$, the separation between the two parallel planes representing the covector is divided by $|c|$ since the hyperplane $(cf)(x) = 1$ is the plane $f(x) = 1/c$ compared to the original hyperplane $f(x) = 1$. If $c < 0$ there is also

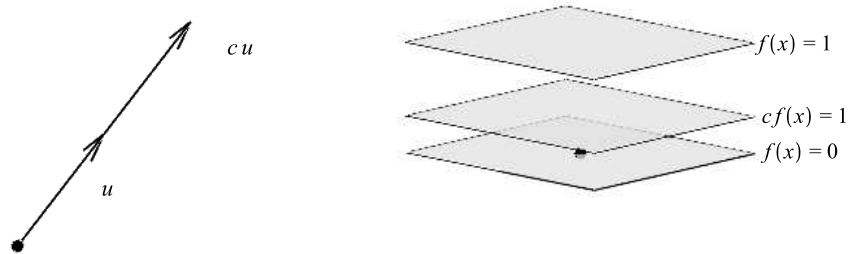


Figure 1.8: Scalar multiplication, assuming $c > 1$: the vector is stretched, the covector planes are squeezed (shorter separation but bigger density of planes in the original family of integer value planes).

a direction reversal in the sense that the hyperplane $cf(x) = 1$ is on the opposite side of the hyperplane $f(x) = 0$ from $f(x) = 1$. This increases the “number” of planes pierced by a given vector if $|c| > 1$, thus increasing the value of the covector on the vector, and decreases the number if $|c| < 1$.

Using these geometrical pictures we can give a geometric construction of the dual basis to a given basis. Suppose we have 3 linearly independent vectors $\{E_i\}$ in \mathbb{R}^3 . They form a basis. What is the parallel plane representation of the three dual basis vectors $\{W^i\}$? The dual basis is defined by the duality relations

$$\begin{cases} W^i(E_j) = 0, & i \neq j, \\ W^i(E_i) = 1, & (\text{no sum on } i). \end{cases}$$

The first (“offdiagonal”) relation says that a given dual basis vector W^i should “kill” (give zero on) the “other” ($j \neq i$) basis vectors, and hence on any linear combination of the other vectors and hence on any vector in the plane (hyperplane in \mathbb{R}^n) spanned by the other vectors. So the plane of the two vectors E_1 and E_2 is the plane $W^3(x) = 0$ from the pair used to represent W^3 , and similarly for the others. So we’ve determined the orientations of each of the dual basis covectors from the “offdiagonal” duality relations. The “magnitude” is determined by the “diagonal” relations. The “diagonal” relation $W^3(E_3) = 1$ means that the tip of E_3 lies in the plane $W^3(x) = 1$ which must be parallel to the plane $W^3(x) = 0$, completely fixing the former.

Drawing in all 3 pairs of planes makes a 3-dimensional honeycomb structure which I won’t

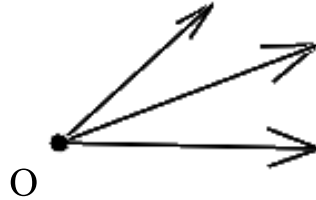


Figure 1.9: 3 independent vectors in \mathbb{R}^3 (noncoplanar): a basis.

attempt to draw. However, the 3 pairs of planes contain the 6 faces of the parallelopiped formed with the 3 basis vectors as edges from a common vertex at the origin. This “unit parallelopiped” corresponds to all points which have new coordinates $0 \leq x^{i'} \leq 1$, whose volume in the Euclidean geometry is the absolute value of the determinant of the matrix of new basis vectors: $Vol(E_1, E_2, E_3) = |\det(\underline{E}_1 | \underline{E}_2 | \underline{E}_3)|$ (the triple scalar product from multivariable calculus). Tiling \mathbb{R}^3 with this unit parallelopiped creates the unit coordinate grid associated with the new coordinates.

Notice that changing E_3 for fixed E_1, E_2 does not change the orientation of W^3 which must contain the directions of both E_1 and E_2 , only its “magnitude” or separation parameter changes in order to maintain $W^3(E_3) = 1$. On the other hand changing E_3 for fixed E_1, E_2 does change W^1 and W^2 so that their planes remain parallel to E_3 . These “complementary” effects of such a change reflect the “duality” between vectors and covectors. A vector is represented by a directed line segment (1-dimensional) while a covector is represented by a directed pair of parallel planes of dimension $3 - 1 = 2$ in \mathbb{R}^3 or hyperplanes (dimension $n - 1$) in \mathbb{R}^n . The “directed” qualifier refers to the sense in which we start at the 0-value and finish at the 1-value, although there is no particular direction from the origin (i.e., specific vector) along which we go, unless we

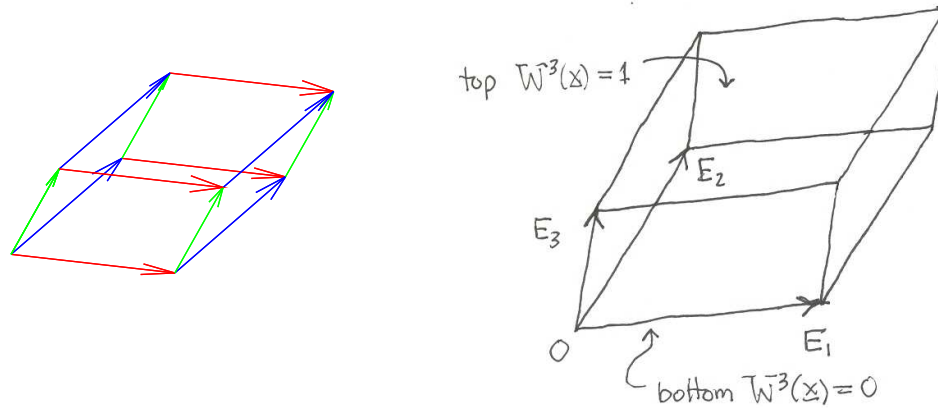


Figure 1.10: Visualizing dual basis relations in \mathbb{R}^3 . The 3 basis vectors determine a parallelogram whose 3 pairs of faces lie in the three pairs of planes $W^i(x) = 0, 1$ for $i = 1, 2, 3$. The zero value plane of W^3 contains E_1 and E_2 , and the tip of E_3 touches the unit value plane for W^3 , so the pair of planes representing W^3 contain the two faces of the parallelepiped formed by the three basis vectors which are connected at one corner by the edge E_3 .

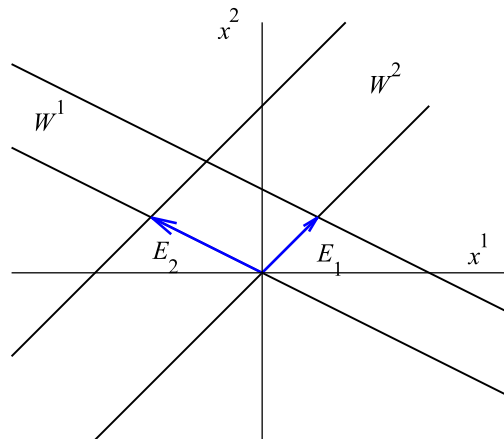


Figure 1.11: Visualizing dual basis relations in \mathbb{R}^2 . The sides of the parallelogram formed by the two basis vectors are contained in the pairs of lines $W^i(x) = 0, 1$ for $i = 1, 2$. The zero value plane of W^2 contains E_1 , and the tip of E_2 touches the unit value plane for W^2 .

introduce a Euclidean geometry, for example, which picks out the direction orthogonal to the pair of the planes.

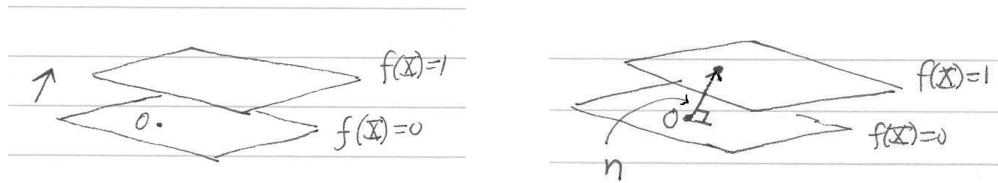


Figure 1.12: Representation of covector as pair of planes: a connecting normal vector captures its orientation information but not directly the length of the covector (instead it captures the inverse length).

Suppose n (for “normal”) is the orthogonal vector in \mathbb{R}^3 from the origin to the second plane representing the covector $f = f_i x^i$, so that it satisfies $f(n) = f_i n^i = 1$. In the Euclidean geometry n^i is proportional to f_i since index position does not matter, but that means that their lengths must be inversely proportional from the constraint that their inner product is 1. More on this later. For now it suffices to say that we can’t pick out a particular direction in which the level surfaces of a covector increase in their value without additional structure, like the dot product in our usual approach to geometry in \mathbb{R}^n .

All of this is much simpler to visualize in the plane \mathbb{R}^2 where we only have two basis vectors E_1 and E_2 , which form a parallelogram whose sides extend to represent in pairs the two dual basis covectors W^1 and W^2 . The Euclidean area of the “unit parallelogram” formed by the two basis vectors is just the length of the cross-product of the two vectors thought of as vectors in the plane x^1 - x^2 plane of \mathbb{R}^3 which in turn is just the absolute value of the determinant of the matrix containing these vectors as columns: $\text{Area}(E_1, E_2) = |E_1 \times E_2| = |\det\langle E_1 | E_2 \rangle|$.

Well, we began with a vector space V and introduced its dual space V^* which is itself a vector space in its own right and so has its own dual space $(V^*)^* = V^{**}$ of real-valued linear functions of covectors

$$F \in V^{**} \text{ means } F(af + bg) = aF(f) + bF(g) \quad (\text{linearity condition}).$$

However, unlike a relationship between a vector space and its dual which requires additional mathematical structure, there is a “natural” identification of a vector space and the dual of its dual. Of course they are all n -dimensional vector spaces and therefore isomorphic, but one has to choose a basis to establish a particular isomorphism which then depends on that choice of basis (“unnatural,” not independent of the choice of basis). But the relationship between a basis of V and a corresponding basis of V^{**} is natural, independent of basis, and is a result of the fact that a linear function value $f(u) = f_i u^i$ for fixed vector input u is a linear function of the covector f thought of as a variable. This natural pairing enables us to view this as a linear function of variable f for fixed u , i.e., switching the interpretation of variables and coefficients that led us to describe linear combinations of variables that form linear functions of those variables.

For each $v \in V$, define a $\tilde{v} \in V^{**}$ by $\tilde{v}(f) = f(v)$ for any covector f , in other words given a vector in V , we define the value of a dual space linear function on a covector to be the value of the covector on the original vector. Then this tilde map association satisfies

$$\begin{aligned} (a\tilde{u} + b\tilde{v})(f) &= a\tilde{u}(f) + b\tilde{v}(f) && \text{(def. of linear comb. of functions)} \\ &= af(u) + bf(v) && \text{(def. of tilde map)} \\ &= f(au + bv) && \text{(linearity of covector)} \\ &= \widetilde{(au + bv)}(f), && \text{(def. of tilde map)} \end{aligned}$$

and since this is true for all covectors f , then it is true of the functions themselves $\widetilde{(au + bv)} = a\tilde{u} + b\tilde{v}$ so $\sim: V \rightarrow V^{**}$ is a linear map. It is also 1-1 since if $\tilde{u} = \tilde{v}$, then $\tilde{u}(f) = \tilde{v}(f)$ so $f(u) = f(v)$ and $f(u - v) = 0$ (linearity) for *every* covector f , which can only be true if $u - v = 0$ or $u = v$. This means it is a vector space isomorphism (1-1 linear map).

So if you start with $F \in V^{**}$, then

$$\begin{aligned} F(f) &= F(f_i\omega^i) = f_iF(\omega^i) \equiv f_iF^i = f_i\delta^i_jF^j \\ &= f_i\omega^i(e_j)F^j = (f_i\omega^i)(F^je_j) = f(F^je_j), \end{aligned}$$

where $F(\omega^i) \equiv F^i$ are the “components” of F with respect to the basis $\{e_i\}$, i.e., evaluation of F on f is equivalent to evaluation of f on the vector F^ie_i . Furthermore $\{\tilde{e}_i\}$ is the basis dual to the basis $\{\omega^i\}$ of V^* , since $\tilde{e}_i(\omega^j) = \omega^j(e_i) = \delta^j_i$. We can therefore forget about V^{**} by using evaluation of covectors on vectors to produce linear functions of covectors.

Thus the natural evaluation $f(v) = f_iv^i$ can be interpreted as a linear function of v for fixed f or as a linear function of f for fixed v . This puts vectors and covectors on an equal footing with respect to evaluation, and sometimes this is made explicit (by mathematicians) by using the notation

$$f(v) = \langle f, v \rangle \quad (\text{“scalar product” of covector and vector})$$

which eliminates having to write one evaluated on the other as function notation requires. This scalar product notation also enables one to connect the transpose matrix with a transpose map:

$$f(Av) = f_i(A^i_jv^j) = (f_iA^i_j)v^j \quad \text{becomes} \quad \langle f, Av \rangle = \langle A^T f, v \rangle,$$

namely the action $u^i \rightarrow A^i_ju^j$ of a linear transformation \underline{A} on v in this expression can be transferred to the action of the transpose matrix acting by left multiplication on the components of the covector f thought of as a column matrix. In other words if \underline{f}^T is the row matrix representing the covector, then \underline{f} is the column matrix and $f_i \rightarrow f_jA^j_i$ translates in matrix notation to $\underline{f}^T \rightarrow \underline{f}^T \underline{A}$, the transpose of which is $\underline{f} \rightarrow \underline{A}^T \underline{f}$. In more transparent language, any matrix which acts on the vector space by left multiplication of component columns, also naturally acts on component covectors as rows by right multiplication, which is called the transpose map associated with the original linear transformation.

Another indication of natural versus unnatural isomorphisms is that each time we go to the dual space we interchange index positions—after two interchanges they are back in the right position so one doesn’t need additional structure to get the indices back in the “right position” $F = F^i\tilde{e}_i \longleftrightarrow F^ie_i$ as is necessary in the relationship between a vector space and its own dual.

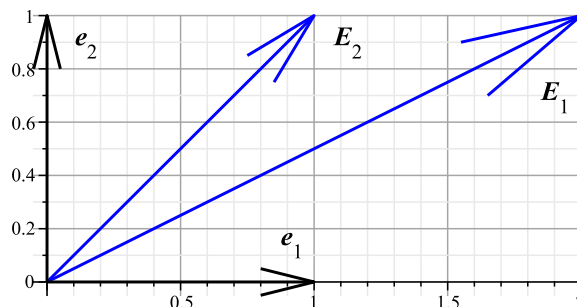


Figure 1.13: Change of basis in the plane \mathbb{R}^2 from the natural basis $\{e_1, e_2\}$ to a new basis $\{E_1, E_2\}$

Exercise 1.3.2.

change of basis in the plane

Given the new basis $E_1 = \langle 2, 1 \rangle$, $E_2 = \langle 1, 1 \rangle$ of \mathbb{R}^2 :

a) Find the dual basis $\{W^1, W^2\}$ to $\{E_1, E_2\}$ in terms of the dual basis $\{\omega^1, \omega^2\}$ to the natural basis $\{e_1, e_2\} = \{\langle 1, 0 \rangle, \langle 0, 1 \rangle\}$.

Hint: $W^1 = a\omega^1 + b\omega^2$, $W^2 = c\omega^1 + d\omega^2$, (why?)

so express the 4 conditions $W^i(E_j) = \delta^i_j$ to obtain 4 linear equations to determine the 4 constants a, b, c, d .

b) What are the new coordinates of $\langle 0, 2 \rangle$, namely $(W^1(\langle 0, 2 \rangle), W^2(\langle 0, 2 \rangle))$?

c) Plot the vectors $\langle a, b \rangle$ and $\langle c, d \rangle$ on the same axes. What do you notice about their relation to $\{E_i\}$?

d) Plot the 4 representative lines $W^1(X) = 0, 1$ and $W^2(X) = 0, 1$ and interpret the result of part b) in terms of the number of such unit tickmark lines associated with W^2 that are pierced by the vector $\langle 0, 2 \rangle$.

e) Recall Example 1.3.3. Notice that the matrix $\underline{A} = \langle \langle a|b \rangle, \langle c|d \rangle \rangle$ whose rows are the components of the new dual basis covectors is just the inverse of the matrix whose columns are the components of the new basis vectors $\langle \underline{E}_1 | \underline{E}_2 \rangle$, i.e., show that $\underline{A} = \langle \underline{E}_1 | \underline{E}_2 \rangle^{-1}$. In other words there is no need to solve equations for the dual basis since they are equivalent to finding the inverse matrix of the matrix of basis vectors themselves, which is a trivial task for computer algebra systems.

■

Affine spaces

When we consider physical problems in the plane or in space, as in calculus when we introduce a Cartesian coordinate system based on a choice of origin and a set of rectangular axes, we then make a direct mapping to \mathcal{R}^2 or \mathcal{R}^3 respectively of n -tuples of real numbers with $n = 2, 3$,

and we work in that environment using the vector space structure of position vectors to locate points and we deal with relative positions using difference vectors. In fact we really never even consider this first step and simply work with Cartesian coordinates on these spaces, but it is understood that we have the freedom to pick another origin and another set of rectangular axes to describe points in those spaces. When we think of physical space within the framework of physics, we recognize this freedom in the construction of Cartesian axes. The mathematical construct that we employ is not a vector space but in contrast, a vector space modulo a choice of origin, which is called an affine space. The difference vectors (directed line segments) between points of our affine space belong to the corresponding vector space. Although we utilize the vector space structure of these \mathcal{R}^n spaces, it should be clear that this is a convenience for not having to worry about the more precise situation of an affine space. In particular, when we deal with the flat spacetime of special relativity as a vector space with an inertial coordinate system (t, x^1, x^2, x^3) on \mathcal{R}^4 , we understand that it is really an affine space structure that we assume and not a vector space.

1.4 Linear transformations of a vector space V into itself (and tensors)

Introducing the dual space of covectors is just a way of giving explicit mathematical structure to the set of real-valued linear functions on a vector space that we usually take for granted when working with real-valued functions on \mathbb{R}^n . Generalizing linear functions from one vector argument to more than one vector argument to consider multilinear functions leads in a similar way to tensors, which are real-valued multilinear functions of a number of vector and covector arguments. Matrix multiplication of vectors in \mathbb{R}^n by a square matrix is a familiar operation that can be easily re-interpreted in this light, thus connecting the idea of linear transformations of a vector space into itself with an associated multilinear function, or tensor. Finally changing the basis of \mathbb{R}^n from the standard basis by such a linear transformation amounts to matrix multiplication by an invertible matrix, the columns of which represent the standard components of the new basis vectors, and whose inverse matrix contains the standard components of the new dual basis as its rows. We postpone a discussion of basis changes until the next section.

Concrete examples of linear transformations

Example 1.4.1. rotations of the plane

It is important to have a concrete example of a linear transformation. Consider a counter-clockwise rotation of the plane \mathbb{R}^2 by an angle θ , shown in Fig. 1.14 as a 30° angle, accomplished by matrix multiplication by the matrix \underline{R} , i.e., a starting vector X is moved to the new position $\underline{R}(X) = \underline{R}X$. Multiplying the standard basis vectors $e_1 = \langle 1, 0 \rangle, e_2 = \langle 0, 1 \rangle$ produces respectively the two columns $\underline{b}_1, \underline{b}_2$ of the rotation matrix $\underline{R} = \langle \underline{b}_1 | \underline{b}_2 \rangle$. By simple trigonometry shown in this figure, this rotation matrix is given by

$$\langle \underline{e}_1 | \underline{e}_2 \rangle = \underline{I} \rightarrow \underline{R} \langle \underline{e}_1 | \underline{e}_2 \rangle = \langle \underline{R} \underline{e}_1 | \underline{R} \underline{e}_2 \rangle = \langle \underline{b}_1 | \underline{b}_2 \rangle$$

where explicitly

$$\underline{R} = \langle \underline{b}_1 | \underline{b}_2 \rangle = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

In terms of the Cartesian coordinates this transformation of the plane into itself takes the form

$$\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \rightarrow \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} \cos \theta x^1 - \sin \theta x^2 \\ \sin \theta x^1 + \cos \theta x^2 \end{pmatrix},$$

or

$$\underline{x} \rightarrow \underline{R} \underline{x}.$$

This matrix defines a real-valued function of a covector and a vector in the obvious way

$$\underline{f}^T \underline{R} \underline{v} = (f_1 \ f_2) \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = f_1 \cos \theta v^1 - f_1 \sin \theta v^2 + f_2 \sin \theta v^1 + f_2 \cos \theta v^2,$$

which is simply the value of the linear function f on the rotated vector $\underline{R}(\underline{v})$, namely $f(\underline{R}(\underline{v}))$. This is linear in both the covector f and the vector \underline{v} , and so defines a tensor \mathcal{R} such that $\mathcal{R}(f, \underline{v}) = f(\underline{R}(\underline{v}))$.

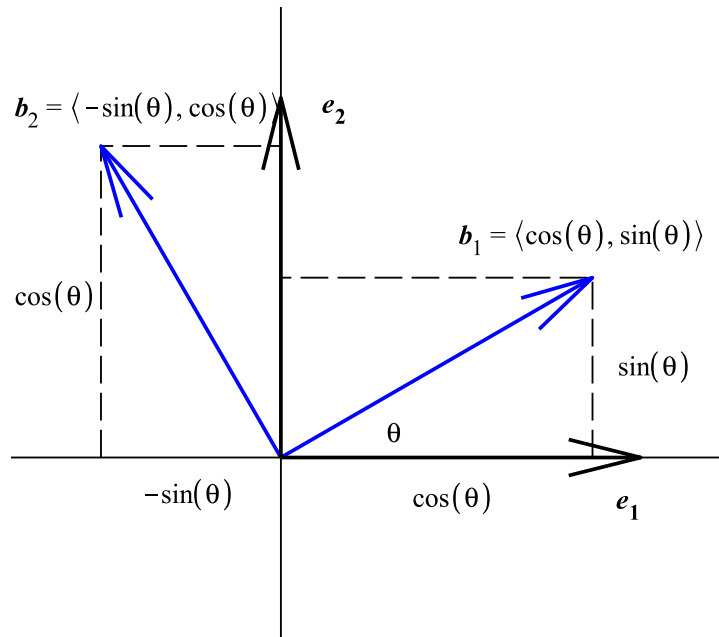


Figure 1.14: The effect of a counterclockwise rotation of the plane by the angle θ (shown here as $\pi/6$ or 30°) acting on the standard basis vectors. The rotated vectors form the columns of the rotation matrix.

□

Exercise 1.4.1.

rotations and pseudorotations in the plane

a) If we use the notation

$$\underline{R}(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix},$$

show that

$$\underline{R}(\theta_1) \underline{R}(\theta_2) = \underline{R}(\theta_1 + \theta_2),$$

and that

$$\underline{R}(\theta)^{-1} = \underline{R}(-\theta), \quad \underline{R}(0) = \underline{I}.$$

This corresponds to our intuition that a rotation of the plane by two successive rotations is a rotation by the sum of the angles of the individual rotations, that we can undo any rotation by rotating the plane by the negative of its angle, and that a rotation by a zero angle leaves the plane unchanged. Mathematically this corresponds to the fact that the set of rotations of the plane about the origin by all possible angles forms what is called an Abelian group.

A group is any set with a binary group operation (“multiplication”) associating each ordered pair of elements A, B with another member AB of the set such that the group multiplication is associative: $A(BC) = (AB)C$, there is one element I called the identity satisfying $AI = IA$, and every element A has an inverse denoted by A^{-1} such that $A^{-1}A = AA^{-1} = I$. For an Abelian (or “commutative”) group, the order of the factors does not matter: $AB = BA$; the two group elements are said to “commute” if this relation is true for a given pair. Since matrix multiplication is associative, a set of invertible matrices (matrices which have an inverse) is a matrix group under matrix multiplication if it contains the identity matrix and is closed under matrix multiplication. Recall a matrix is invertible if its determinant is nonzero, and since the determinant of a matrix product is the product of the determinants, the product matrix of two invertible matrices (nonzero determinants) also has a nonzero determinant and is therefore invertible. Groups are very important in differential geometry.

In this case it is obvious that the inverse of a rotation matrix is just its transpose

$$\underline{R}(\theta)^{-1} = \underline{R}(-\theta) = \underline{R}(\theta)^T \quad \leftrightarrow \quad \underline{R}(\theta)^T \underline{R}(\theta) = \underline{I},$$

This last condition may be interpreted as stating that the dot product of the columns of a rotation matrix are zero if the columns are distinct, and 1 if the same, so the set of columns of the matrix form a set of orthonormal vectors in the geometry of the dot product. This group is called the orthogonal group in 2 dimensions $O(2)$, understood to be real (not complex).

b) Show that the following family of hyperbolic rotations or “boost” matrices

$$\underline{B}(\beta) = \begin{pmatrix} \cosh \beta & \sinh \beta \\ \sinh \beta & \cosh \beta \end{pmatrix},$$

also forms a group by checking that $\underline{B}(\beta_1) \underline{B}(\beta_2) = \underline{B}(\beta_1 + \beta_2)$, using the corresponding identities for the hyperbolic functions. The hyperbolic angle β is also called the rapidity. This hyperbolic geometry is explored Appendix C. It is at the heart of the geometry of special relativity and Lorentz transformations.

c) Evaluate the matrix differentials $\underline{R}^{-1}d\underline{R} = d\underline{R} \underline{R}^{-1}$ and $\underline{B}^{-1}d\underline{B} = d\underline{B} \underline{B}^{-1}$. Notice that the result is an antisymmetric matrix in the first case and a symmetric matrix in the second case, but tracefree in both cases. To interpret this, consider the parametrized curve $\underline{R}(\theta) \underline{x}$ (a circle) obtained by rotating a fixed point \underline{x} in the plane, and calculate the tangent

$$\frac{d}{d\theta} (\underline{R}(\theta) \underline{x}) = \left(\frac{d\underline{R}}{d\theta}(\theta) \underline{R}(\theta)^{-1} \right) (\underline{R}(\theta) \underline{x}).$$

This means that if we rotate a point by an angle, at each new location the tangent vector is obtained from the position vector by multiplying the latter by an antisymmetric matrix. In other words for a very small rotation the tip of the position vector is rotated by multiplying by an antisymmetric matrix. In the boost case instead this role is played by a symmetric matrix. ■

Linear transformations and tensors

Remark.

The set of all invertible linear transformations of a vector space V into itself is called the general linear group $GL(V)$ for that space. $GL(n, \mathbb{R})$ is the space of invertible $n \times n$ matrices, which acts on \mathbb{R}^n as a transformation group $GL(\mathbb{R}^n)$ by matrix multiplication. Tensor transformation laws extend this action to the tensor spaces above \mathbb{R}^n , which we will understand only after we know what a tensor is. Both of the above 1-dimensional matrix groups (one free group parameter) are subgroups of $GL(2, \mathbb{R})$. In fact both are subgroups of the special linear group $SL(2, \mathbb{R})$ which consists of all unit determinant invertible matrices. \blacktriangle

Suppose $A : V \rightarrow V$ is a linear transformation of V into itself, i.e., a V -valued linear function on V , or equivalently a linear function on V with values in V . For each i , the result $A(e_i)$ is a vector with components defined by $A^j_i \equiv \omega^j(A(e_i))$ (note natural index positions up/down): $A(e_i) \equiv A^j_i e_j$. By linearity

$$A(v) = A(v^i e_i) = v^i A(e_i) = v^i (A^j_i e_j) = (A^j_i v^i) e_j \quad \text{or} \quad [A(v)]^j = A^j_i v^i.$$

The j -th component of the image vector is the j -th entry of the matrix product of the matrix $\underline{A} \equiv (A^j_i)$ (the row index j is on the left, the column index i is on the right) with the column vector $\underline{v} \equiv (v^i)$. Here the underlined symbol \underline{A} distinguishes the matrix of the linear transformation from the transformation A itself. This matrix $\underline{A} = (\omega^j(A(e_i)))$ is referred to as the “matrix of A with respect to the basis $\{e_i\}$.” Obviously if you change the basis, the matrix will change. We’ll get to that later.

Even if we are not working with \mathbb{R}^n , any choice of basis $\{e_i\}$ of V establishes an isomorphism with \mathbb{R}^n , namely the n -tuple of components of a vector with respect to this basis is a point in \mathbb{R}^n —this essentially identifies the basis $\{e_i\}$ of V with the standard basis of \mathbb{R}^n . Recall the natural correspondence of quadratic polynomials with R^3 explored in the earlier Exercise 1.2.3.

Expressing everything associated with an abstract vector space in terms of components with respect to a given basis leads us to matrix notation. Vectors in component form become column matrices, covectors become row matrices, and a linear transformation becomes a square matrix acting by matrix multiplication on the left, while natural evaluation of a covector on a vector is the matrix multiplication of the corresponding row (left) and column (right) matrices

$$\underline{v} = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}, \quad \underline{f}^T = (f_1 \cdots f_n), \quad \underline{A} = (A^j_i), \quad \rightarrow \quad \begin{pmatrix} [A(v)]^1 \\ \vdots \\ [A(v)]^n \end{pmatrix} = \underline{A} \underline{v}, \quad f(v) = \underline{f}^T \underline{v}.$$

Since $A(v)$ is another vector we can evaluate it on the covector f to get a number which has the triple matrix product representation

$$f(A(v)) = \underline{f}^T \underline{A} \underline{v}. \quad (\text{row} \times \text{square matrix} \times \text{column} = \text{scalar})$$

For every linear transformation A , this enables us to define an associated bi-linear real-valued function \mathbb{A} of a pair of arguments consisting of a covector and a vector. Bi-linear simply

means linear in each of two arguments. This bi-linear function is

$$\begin{aligned}\mathbb{A}(f, v) \equiv f(A(v)) &= (f_i \omega^i)(A^j_k v^k e_j) = f_i A^j_k v^k \omega^i(e_j) \\ &= f_i A^j_k v^k \delta^i_j = f_i A^i_k v^k,\end{aligned}$$

noting that $A(v)$ is a vector and $f(A(v))$ is a scalar (real number). For fixed f , $\mathbb{A}(f, v)$ is a real-valued linear function of v , namely the covector with components $f_i A^i_k$ (one free down index). For fixed v , it is a real-valued linear function of f , namely evaluation on the vector with components $A^i_k v^k$ (one free up index). This reinterprets the linear transformation A as a bilinear function \mathbb{A} of a covector (first argument) and a vector (second argument), i.e., a “tensor.” Note the component relation $A^i_j = \mathbb{A}(\omega^i, e_j)$. We will notationally identify A and \mathbb{A} once we are more familiar with these matters. Sometimes one writes the linear transformation as $u \rightarrow A(u) = \mathbb{A}(, u) = C_u \mathcal{A}$, namely as the tensor with only one of its two arguments evaluated, or sometimes as the “contraction” of the tensor \mathbb{A} with the vector u to indicate its natural evaluation on that argument alone.

In general a $\binom{p}{q}$ -tensor over V is simply a real-valued multilinear function of p covector arguments (listed first) and q vector arguments (listed last):

$$T(\underbrace{f, g, \dots}_p, \underbrace{v, u, \dots}_q) \in \mathbb{R}.$$

Listing all the covector arguments first and the vector arguments last is just an arbitrary choice, and later on it will be convenient to allow any ordering. By definition then, a covector is a $\binom{0}{1}$ -tensor over V (1 vector argument, no covector arguments) while a vector is a $\binom{1}{0}$ -tensor over V (1 covector argument, no vector argument) recalling that $v(f) \equiv f(v)$ (the value of a vector on a covector is the value of the covector on the vector).

Thus a linear transformation A has (naturally) a $\binom{1}{1}$ -tensor \mathbb{A} over V associated with it. Any time we have a space of linear functions over a vector space, it has a natural linear structure by defining linear combinations of functions through linear combination of values, i.e., is itself a vector space and we can look for a basis. In this case the space of bilinear real-valued functions on the Cartesian product vector space of pairs (f, v) of covectors and vectors is itself a vector space and in the same way that a basis of V determined a basis of the dual space V^* , they both together determine a basis of this latter vector space.

Let $V \otimes V^*$ denote this space of $\binom{1}{1}$ -tensors over V . The symbol \otimes is called the tensor product, explained below. The zero element of this vector space is a multilinear function

$$\begin{array}{ccccccc} 0(f, v) & = & 0 & \longleftrightarrow & 0^i_j & = & \omega^i(0(e_j)) = 0, \\ \text{zero tensor} & & \text{zero number} & & \text{zero matrix} & & \text{zero linear transformation} \end{array}$$

whose square matrix of components is the zero matrix (note $0(e_j) = 0$ is the zero vector). Another special element in this space is the evaluation tensor associated with the identity transformation $Id(v) = v$

$$EVAL(f, v) = f(v) = f_i \delta^i_j v^j \longleftrightarrow (EVAL)^i_j = \omega^i(Id(e_j)) = \omega^i(e_j) = \delta^i_j$$

whose square matrix of components is the unit matrix \underline{I} , the index symbol for which has been called the Kronecker delta. *Eval* is sometimes called the unit tensor, and the associated linear transformation of the vector space is just the identity transformation which sends each vector to itself.

To come up with a basis of $V \otimes V^*$ we need a simple definition. Given a covector and a vector we produce a $\binom{1}{1}$ -tensor by defining

$$(v \otimes f)(g, u) \equiv g(v)f(u) = (g_i v^i)(f_j u^j) = g_i(v^i f_j)u^j.$$

Thus $(v^i f_j)$ is the matrix of components of $v \otimes f$, and is also the result of evaluating this tensor on the basis vectors and dual basis covectors

$$(v \otimes f)^i_j = (v \otimes f)(\omega^i, e_j) = \omega^i(v)f(e_j) = v^i f_j.$$

The symbol \otimes is called the tensor product and only serves to hold v and f apart until they acquire arguments to be evaluated on. It simply creates a function taking 2 arguments from two functions taking single arguments. The component expression shows that $v \otimes f$ is clearly bilinear in its arguments g and u , so it is a $\binom{1}{1}$ -tensor.

In terms of the corresponding matrix notation, given a column matrix $\underline{u} = \langle u^1, \dots, u^n \rangle$ and a row matrix $\underline{f}^T = \langle f_1 | \dots | f_n \rangle$, then the tensor product corresponds exactly to the other matrix product (column times row instead of row times column)

$$(u^i f_j) = \underbrace{\underline{u} \quad \underline{f}^T}_{\substack{n \times 1 \quad 1 \times n \\ n \times n}} \quad \text{in contrast with} \quad \underbrace{\underline{f}^T \quad \underline{u}}_{\substack{1 \times n \quad n \times 1 \\ 1 \times 1}} = f(u) = f_i u^i.$$

Thus the tensor product of a vector and a covector is just an abstract way of representing the multiplication of a column vector on the left by a row vector on the right to form a square matrix, a two-index object created out of two one-index objects.

Example 1.4.2. matrix product and linear function evaluation

A concrete example can help. The matrix product on the left below is the usual order of a row on the left and a column on the right, resulting in a scalar. The rows and columns of the matrix product on the right below of a column on the left multiplying a row on the right have only 1 entry each respectively so the row-column products are simply products of those two entries.

$$\underline{f}^T \underline{u} = (f_1 \quad f_2) \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} = f_1 u^1 + f_2 u^2, \quad \underline{u} \underline{f}^T = \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} (f_1 \quad f_2) = \begin{pmatrix} f_1 u^1 & f_2 u^1 \\ f_1 u^2 & f_2 u^2 \end{pmatrix}$$

The latter matrix product is the matrix of components of the tensor product $u \otimes f$ of the vector u with the 1-form f . Its matrix product with a component vector corresponds to a linear transformation of the vector space.

□

We can use the tensor product \otimes to create a basis for $V \otimes V^*$ from a basis $\{e_i\}$ and its dual basis $\{\omega^i\}$, namely the set $\{e_j \otimes \omega^i\}$ of $n^2 = n \times n$ such tensors. By definition

$$(e_j \otimes \omega^i)(g, u) = g(e_j)\omega^i(u) = g_j u^i = u^i g_j.$$

We can use this to show the two conditions that they form a basis are satisfied:

1. **spanning set:**

$$\mathbb{A}(f, v) = \cdots = f_j A^j_k v^k = A^j_k v^k f_j = (A^j_k e_j \otimes \omega^k)(f, v)$$

since $v^k f_j = (e_j \otimes \omega^k)(f, v)$, so $\mathbb{A} = A^j_k e_j \otimes \omega^k$ holds since the two bi-linear functions have the same values on all pairs of arguments. The components of the tensor \mathbb{A} with respect to this basis are just the components of the linear transformation A with respect to $\{e_i\}$ introduced above : $A^j_k = \omega^j(A(e_k))$.

2. **linear independence:** if $A^j_k e_j \otimes \omega^k = 0$ (zero tensor) then evaluating both sides on the argument pair (ω^m, e_n) leads to

$$\begin{aligned} (A^j_k e_j \otimes \omega^k)(\omega^m, e_n) &= 0(\omega^m, e_n) = 0 \\ &= A^j_k \omega^m(e_j) \omega^k(e_n) = A^j_k \delta^m_j \delta^k_n \\ &= A^m_n, \end{aligned} \tag{1.1}$$

so since this is true for all possible values of (m, n) , all the coefficients must be zero, proving linear independence.

Thus $V \otimes V^*$ is the space of linear combinations of tensor products of vectors with covectors, explaining the notation.

Example 1.4.3. basis of the vector space of $m \times n$ matrices

We have no notation for the natural basis of the vector space $gl(n, \mathbb{R})$ of $n \times n$ matrices, namely the standard basis of the corresponding \mathbb{R}^{n^2} we get by listing the entries of the matrix row by row as a single 1-dimensional array. Let \underline{e}^j_i be the matrix whose only nonzero entry is a 1 in the i th row and j th column. Then $\underline{A} = A^i_j \underline{e}^j_i$ is how we represent the matrix in terms of its entries. The ordering of the indices on \underline{e}^j_i allows us to think of this product as having adjacent indices (the j 's) being summed over and taking the trace of the result (the i s), which are natural matrix kinds of index operations. (The equally acceptable alternative notation would be instead $\underline{A} = A^i_j \underline{e}^j_i$, but for some reason the first index ordering pleases me more for the interpretational reason I stated.) Then to the matrix \underline{A} corresponds a tensor $\mathbb{A} = A^i_j e_i \otimes \omega^j$, whose components with respect to this basis are just the corresponding entries of the matrix, so really the basis $\{e_i \otimes \omega^j\}$ of $\mathbb{R}^n \otimes \mathbb{R}^{n*}$ induced by the standard basis $\{e_i\}$ of \mathbb{R}^n corresponds exactly to the obvious basis $\{\underline{e}^j_i\}$ of the vector space of square matrices. Again we are taking familiar objects and looking deeper at their mathematical structure, which requires new notation like the tensor product to make explicit that structure.

□

Example 1.4.4. projections as linear transformations

Trying to gain intuition about linear transformations $A : V \rightarrow V$ from a vector space into itself using the rotations and boosts of the plane is a bit misleading since they only give us intuition about linear transformations which are 1-1 and do not “lose any points” as they move them around in the vector space on which they act. Such linear transformations are represented by nonsingular matrices when expressed in a basis, i.e., matrices with nonzero determinant $\det \underline{A} \neq 0$, which means that the only solution to $\underline{A}\underline{x} = \underline{0}$ is the zero solution. Those matrices with zero determinant also arise naturally.

Suppose we decompose $V = V_1 \oplus V_2$ into a direct sum of two subspaces, which simply means that any vector can be expressed uniquely as the sum of one vector in V_1 and another in V_2 . In multivariable calculus, one of the first things we do with vector algebra is project a general vector in space into a vector component along a given direction and another one orthogonal to it. If \hat{u} is a unit vector which picks out a direction in \mathbb{R}^3 , then the projections of another vector v parallel to and perpendicular to \hat{u} are

$$P_{u||}(v) = (v \cdot \hat{u}) \hat{u}, \quad P_{u\perp}(v) = v - P_{u||}(v) = v - (v \cdot \hat{u}) \hat{u},$$

If v is already along \hat{u} , then the first projection just reproduces it, while the second gives the zero vector. If v is orthogonal to \hat{u} , then the second projection just reproduces it, while the first gives the zero vector. By definition, the sum of the two projections just reproduces the original vector.

This is an example of a simple pair of projection maps P and Q which satisfy $P^2 = P$, $Q^2 = Q$, $PQ = QP = 0$ for a pair (P, Q) which projects onto a pair of subspaces in a direct sum total space

$$v \mapsto P(v) + Q(v).$$

Each acts as the identity on its corresponding subspace, and acts as the zero linear transformation on the other. This can be extended to a direct sum of any number of subspaces in an obvious way by iterating these conditions.

The vanishing of the determinant of a matrix \underline{A} is the condition that the homogeneous linear system $\underline{A}\underline{x} = \underline{0}$ has nonzero solutions. The space of solutions is called either the null space or kernel of the matrix. Row reduction of the matrix produces a basis of that subspace of \mathbb{R}^n . However, there is no natural complementary subspace to complete projection into this subspace to a pair of projections as above without additional structure. The problem is that if $\underline{A}\underline{x} \neq \underline{0}$ then one can add any element of the null space to \underline{x} and it will also satisfy the same condition of \underline{x} . But in \mathbb{R}^n we have the orthogonal complement using the dot product to pick out a representative subspace we can use to decompose any vector into an element of the null space and another subspace. This is because the condition $A^i_j x^j = 0$ means that the vector \underline{x} is orthogonal to each of the rows of the coefficient matrix in the dot product interpretation, so that the span of the rows of the matrix (called the row space) is the orthogonal complement of the null space with this natural inner product. Similarly the set of all nonzero vectors $A^i_j x^j$

for all possible x^j corresponds to what is called the range of the linear transformation, but by definition of span, this is simply the span of the set of columns of the matrix, called the column space of the matrix. This too has no natural complement without an inner product, but of course the dot product is ready to do the job. The row and column spaces of a matrix were discussed in detail in Section 1.2. \square

For every linear transformation $A : V \rightarrow V$, there is an associated linear transformation $A^T : V^* \rightarrow V^*$ called its transpose, defined by

$$(A^T)(f)(u) = f(A(u)) = f_i A^i_j u^j \rightarrow [(A^T)(f)]_j = f_i A^i_j,$$

which takes the matrix form

$$\underline{A^T}(f) = \underline{f^T} \underline{A}.$$

Thus with the row vector $\underline{f^T}$ we associate the new row vector

$$\underline{f^T} \mapsto \underline{f^T} \underline{A},$$

or equivalently taking the matrix transpose of this equation, the corresponding column vector \underline{f} is associated with the new column vector

$$\underline{f} \mapsto \underline{A^T} \underline{f}.$$

In words left multiplication of a row matrix $\underline{f^T}$ by the matrix \underline{A} of the linear transformation A is equivalent to right multiplication by the transpose matrix $\underline{A^T}$ of the corresponding transposed column matrix \underline{f} . The abstract transpose linear transformation therefore corresponds directly to the transposed matrix acting in the transposed direction by matrix multiplication. In other words, when we want to think of V^* as itself a vector space undergoing a linear transformation, we then want to think of its elements as column matrices multiplied on the left by a matrix, and this leads to the transpose matrix of the original transformation acting on V .

This transpose linear transformation corresponds to partial evaluation of the tensor \mathbb{A} in its covector argument

$$A^T : f \mapsto \mathbb{A}(f, \quad),$$

resulting in a new covector. Thus the $\binom{1}{1}$ -tensor \mathbb{A} packages both the linear transformation A and its transpose A^T in the same machine, so we can identify these particular transformations as two particular ways in which this tensor acts on both the vector space and its dual, and in fact we might as well use the same kernel symbol A for the tensor as well, letting its partial evaluation in either argument represent the two respective linear transformations.

Example 1.4.5. dot product as a tensor on \mathbb{R}^n

The dot product of two vectors

$$\text{dot}(a, b) = a \cdot b = \langle a^1, \dots, a^n \rangle \cdot \langle b^1, \dots, b^n \rangle = a^1 b^1 + \dots + a^n b^n = \sum_{i=1}^n a^i b^i = \underline{a^T} \underline{b}$$

is the simplest bilinear scalar function of a pair of vectors in \mathbb{R}^n . It is therefore a $\binom{0}{2}$ -tensor “dot.” The standard basis vectors are orthonormal so we need an appropriate symbol for their dot products, which will be interpreted as the components of the dot product tensor. Numerically the matrix of these components is the identity matrix but the index positioning must be covariant, so we introduce a covariant Kronecker delta symbol for these components

$$\text{dot}(e_i, e_j) = e_i \cdot e_j = \delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Then by bilinearity

$$\text{dot}(a, b) = a \cdot b = (a^i e_i) \cdot (b^j e_j) = a^i b^j e_i \cdot e_j = \delta_{ij} a^i b^j = \delta_{ij} \omega^i(a) \omega^j(b) = \delta_{ij} \omega^i \otimes \omega^j(a, b),$$

so this tensor is

$$\text{dot} = \delta_{ij} \omega^i \otimes \omega^j.$$

The dot product on \mathbb{R}^n is an example of an inner product on a vector space, named for its symbolic representation as a raised dot between the vector arguments. An inner product on any vector space is a “symmetric” $\binom{0}{2}$ -tensor which accepts two vector arguments in either order and produces a real number (and such that the determinant of its symmetric matrix of components is nonzero, the condition of nondegeneracy). The dot product on \mathbb{R}^n is such an inner product whose matrix of components is the identity matrix with respect to the standard basis of \mathbb{R}^n . The index positioning δ_{ij} for a $\binom{0}{2}$ tensor shows that it is fundamentally different from the identity $\binom{1}{1}$ -tensor with components δ^i_j , even though both matrices of components are the unit matrix. Section 1.6 will explore inner products on both V and its dual space V^* and their interpretation as linear transformations. □

More than 2 indices: general tensors

So we’ve taken the linear algebra of \mathbb{R}^n , as embodied in column matrices (vectors), row matrices (covectors), both with n entries, and square $n \times n$ matrices ($\binom{1}{1}$ -tensors), and generalized them into the mathematical structure of a vector space V , its dual space V^* and their tensor product space $V \otimes V^*$. This abstracts from 1 and 2 index objects associated with the elementary linear algebra of introductory courses to allow us to consider objects in an invariant way (no indices) that correspond to any number of indices on those objects. Clearly we can play the same game with any space of tensors over V with arbitrary numbers of arguments of either type.

Tensor products with more than two vector or covector factors are defined in an obvious way. For example, the tensor product of two covectors and two vectors is defined by

$$(f_{(1)} \otimes f_{(2)} \otimes u_{(1)} \otimes u_{(2)})(v_{(1)}, v_{(2)}, g_{(1)}, g_{(2)}) = f_{(1)}(v_{(1)}) f_{(2)}(v_{(2)}) u_{(1)}(g_{(1)}) u_{(2)}(g_{(2)}),$$

remembering the key identification which allows us to equate the value $u(g)$ of a vector on a covector to be the value $g(u)$ of the covector on the vector.

If T is a $\binom{p}{q}$ -tensor over V , then $T(f, g, \dots, v, u, \dots) \in \mathbb{R}$ is a scalar. Define its components with respect to $\{e_i\}$ by

$$T^{ij\dots}_{mn\dots} = T(\omega^i, \omega^j, \dots, e_m, e_n, \dots) \quad (\text{scalars}).$$

p is the number of upper indices on these components, equal to the number of covector arguments, while q is the number of lower indices, equal to the number of vector arguments, and it is convenient but not necessary to order all the covector arguments first and the vector arguments last. Next introduce the n^{p+q} basis “vectors,” i.e., $\binom{p}{q}$ -tensors

$$\underbrace{\{e_i \otimes e_j \otimes \dots\}}_{p \text{ factors}} \underbrace{\{\omega^m \otimes \omega^n \otimes \dots\}}_{q \text{ factors}}.$$

We can then expand any tensor in terms this basis

$$T = T^{ij\dots}_{mn\dots} e_i \otimes e_j \otimes \dots \otimes \omega^m \otimes \omega^n \otimes \dots.$$

This expansion follows from the multilinearity and the various definitions just as in the previous case of $\binom{1}{1}$ -tensors over V . Namely

$$\begin{aligned} T(f, g, \dots, u, v, \dots) &= T(f_i \omega^i, g_j \omega^j, \dots, u^m e_m, v^n e_n, \dots) && (\text{argument component expansion}) \\ &= f_i g_j \dots u^m v^n \dots T(\omega^i, \omega^j, \dots, e_m, e_n \dots) && (\text{multilinearity}), \\ &= T^{ij\dots}_{mn\dots} f_i g_j \dots u^m v^n \dots && (\text{definition tensor components}) \\ &= T^{ij\dots}_{mn\dots} e_i(f) e_j(g) \dots \omega^m(u) \omega^n(v) \dots && (\text{definition argument components}) \\ &= (T^{ij\dots}_{mn\dots} e_i \otimes e_j \otimes \dots \omega^m \otimes \omega^n \otimes \dots)(f, g, \dots, u, v, \dots). && (\text{definition tensor product}) \end{aligned}$$

Thus T and its expansion in parentheses in the last line have the same value on any set of arguments, so they must be the same multilinear function.

Example 1.4.6. tensor products by multiplication

The simplest tensor products are just multilinear functions of a set of vectors that result from multiplying together in a certain order linear functions of a single vector. For example, the product of the values of three linear functions of single vectors defines a multilinear function of three vectors by

$$(f \otimes g \otimes h)(u, v, w) = f(u)g(v)h(w).$$

Expressing this tensor in components leads to

$$(f \otimes g \otimes h) = (f \otimes g \otimes h)_{ijk} (\omega^i \otimes \omega^j \otimes \omega^k)$$

where

$$(f \otimes g \otimes h)_{ijk} = f_i g_j h_k.$$

In other words we have constructed a $\binom{0}{3}$ -tensor $f \otimes g \otimes h$ from the tensor product of 3 covectors f , g , and h and in terms of components in index notation, we have just multiplied their components together.

We can do the same thing with vectors instead of covectors

$$u \otimes v \otimes w = (u^i e_i) \otimes (v^j e_j) \otimes (w^k e_k) = u^i v^j w^k e_i \otimes e_j \otimes e_k, \quad (u \otimes v \otimes w)^{ijk} = u^i v^j w^k.$$

Notice that

$$(u \otimes v \otimes w)(f, g, h) = f(u)g(v)h(w) = (f \otimes g \otimes h)(u, v, w).$$

This is the same duality which allows us to think of a linear function of a single covector as a vector and vice versa, together sharing a natural pairing to produce the linear combination which is the value of the linear function. □

Example 1.4.7. determinant as a tensor

On \mathbb{R}^3 with the usual dot and cross products, introduce the $\binom{0}{3}$ -tensor D by

$$\begin{aligned} D(u, v, w) = u \cdot (v \times w) &= \det \begin{pmatrix} u^1 & u^2 & u^3 \\ v^1 & v^2 & v^3 \\ w^1 & w^2 & w^3 \end{pmatrix} \\ &= \det \begin{pmatrix} u^1 & v^1 & w^1 \\ u^2 & v^2 & w^2 \\ u^3 & v^3 & w^3 \end{pmatrix} \quad (\text{“triple scalar product”}) \\ &= \det \langle \underline{u} | \underline{v} | \underline{w} \rangle, \end{aligned}$$

where we use the property that the determinant is invariant under the transpose operation in order to keep our vectors associated with column matrices (while students usually see vectors as rows in the matrix in this context in calculus courses). This is linear in each vector argument (the determinant is a linear function of each row or column, which should be obvious from its representation in terms of the linear dot and cross product operations). It therefore has the expansion

$$D = D_{ijk} \omega^i \otimes \omega^j \otimes \omega^k,$$

where

$$D_{ijk} = D(e_i, e_j, e_k) = e_i \cdot (e_j \times e_k) = \begin{cases} 1 & \text{if } (i, j, k) \text{ even perm. of } (1, 2, 3) \\ -1 & \text{if } (i, j, k) \text{ odd perm. of } (1, 2, 3) \\ 0 & \text{otherwise} \end{cases}$$

so that

$$\begin{aligned} D &= \omega^1 \otimes \omega^2 \otimes \omega^3 + \omega^2 \otimes \omega^3 \otimes \omega^1 + \omega^3 \otimes \omega^1 \otimes \omega^2 \\ &\quad - \omega^1 \otimes \omega^3 \otimes \omega^2 - \omega^2 \otimes \omega^1 \otimes \omega^3 - \omega^3 \otimes \omega^2 \otimes \omega^1. \end{aligned}$$

This corresponds directly to the usual explicit formula

$$\det \begin{pmatrix} u^1 & u^2 & u^3 \\ v^1 & v^2 & v^3 \\ w^1 & w^2 & w^3 \end{pmatrix} = u^1 v^2 w^3 + u^2 v^3 w^1 + u^3 v^1 w^2 - u^1 v^3 w^2 - u^2 v^1 w^3 - u^3 v^2 w^1.$$

We will soon give this determinant component symbol d_{ijk} a new name ϵ_{ijk} called the Levi-Civita symbol or Levi-Civita epsilon, since it is exactly what we need to handle the cross product in \mathbb{R}^3 and the easily prove vector identities involving the dot and cross-products, while generalizing to \mathbb{R}^n to provide a terribly useful tool. In this new notation we then have

$$\det\langle u|v|w\rangle = \epsilon_{ijk}\omega^i \otimes \omega^j \otimes \omega^k(u, v, w) = \epsilon_{ijk}u^i v^j w^k.$$

□

Exercise 1.4.2.

determinant as a tensor

a) Continuing the example, convince yourself that the nonzero components of the determinant function $\epsilon_{ijk} = e_i \cdot (e_j \times e_k)$ (which correspond directly to the 3 positive and 3 negative terms in the expansion of the determinant, respectively the 3 positive cyclic permutations of 123 and the 3 negative cyclic permutations of 123) are

$$1 = \epsilon_{123} = \epsilon_{231} = \epsilon_{312} = -\epsilon_{132} = -\epsilon_{213} = -\epsilon_{321}.$$

b) Notice that if we consider the determinant function $D(c, a, b) = \epsilon_{ijk}c^i a^j b^k$ unevaluated in its first (or last) vector input slot $D(\ , a, b) = D(a, b, \)$, we get one free index in the component representation of the resulting covector $f = D(\ , a, b) = \epsilon_{ijk}a^j b^k \omega^i$. Show that this covector has components

$$\langle f_1, f_2, f_3 \rangle = \langle a^2 b^3 - a^3 b^2, a^3 b^1 - a^1 b^3, a^1 b^2 - a^2 b^1 \rangle = \langle (a \times b)^1, (a \times b)^2, (a \times b)^3 \rangle,$$

which you recognize as the same components as the cross product vector $a \times b$. To make index position work out we must introduce a Kronecker delta with both indices up to write this in index form with our index conventions

$$(a \times b)^i = \delta^{il} \epsilon_{ljk} a^j b^k.$$

Thus we must introduce additional structure to understand this last shift in index position to take a covector to the corresponding vector with the same components. Let's wait till after the next exercise to start tackling that.

■

Exercise 1.4.3.

quadruple scalar product

On \mathbb{R}^3 with the usual dot and cross products, introduce the $\binom{0}{4}$ -tensor

$$Q(u, v, w, z) = (u \times v) \cdot (w \times z)$$

called the “scalar quadruple product.” It satisfies an identity that we will prove easily in Chapter 4

$$(a \times b) \cdot (c \times d) = (a \cdot c)(b \cdot d) - (a \cdot d)(b \cdot c) = \begin{vmatrix} a \cdot c & a \cdot d \\ b \cdot c & b \cdot d \end{vmatrix}.$$

Its components in the standard basis are $Q_{ijmn} = (e_i \times e_j) \cdot (e_m \times e_n)$, from which one can immediately evaluate some of its nonzero components: $Q_{2323} = Q_{3131} = Q_{1212} = 1$. Check these values.

Notice that interchanging either (i, j) or (m, n) results in a sign change, but exchanging these pairs of indices does not

$$Q_{ijkl} = -Q_{jikl} = -Q_{ijlk} = Q_{klij}.$$

These “symmetries” are important and will be explored below. Show that this tensor satisfies one further cyclic identity (first index fixed, last 3 undergo a sum of all cyclic permutations)

$$Q_{ijkl} + Q_{iklj} + Q_{iljk} = 0.$$

■

The simplest example of a tensor created with the dot product is a covector: $f_u(v) = u \cdot v$. For each fixed u , this defines a linear function of v , i.e., a covector f_u . It is exactly this correspondence that allows one to avoid covectors in elementary linear algebra. For a general inner product on any vector space, the degeneracy condition guarantees that this map from vectors to covectors is a vector space isomorphism and hence can be used to establish an identification between the vector space and its dual space. This will prove very useful.

Remark. Tensor product and matrix multiplication

By linearity, the components of the tensor product of a vector and a covector are

$$\begin{aligned} v \otimes f &= (v^i e_i) \otimes (f_j \omega^j) && \text{(expand in bases)} \\ &= v^i f_j e_i \otimes \omega^j && \text{(factor out scalar coefficients)} \\ &\equiv (v \otimes f)^i_j e_i \otimes \omega^j && \text{(definition of components of tensor)} \\ &\rightarrow (v \otimes f)^i_j = v^i f_j \end{aligned}$$

or equivalently

$$(v \otimes f)^i_j = (v \otimes f)(\omega^i, e_j) = \omega^i(v) f(e_j) = v^i f_j.$$

With the representation in component form of a vector and a covector as column and row matrices respectively, this tensor product is exactly equivalent to matrix multiplication

$$\underline{v} \underline{f}^T = \underbrace{\begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}}_{n \times 1} \underbrace{(f_1 \cdots f_n)}_{1 \times n} = \underbrace{\begin{pmatrix} v^1 f_1 \cdots v^1 f_n \\ \vdots \\ v^n f_1 \cdots v^n f_n \end{pmatrix}}_{n \times n} = (v^i f_j),$$

(the vector \underline{v} is a column matrix, the covector \underline{f}^T is a row matrix), but in the opposite order from the evaluation of a covector on a vector, leading to a matrix rather than a scalar (number).

Thus matrix multiplication of a row matrix by a column matrix on the right represents the abstract evaluation operation of a covector on a vector or vice versa, while the matrix multiplication on the left represents the tensor product operation. In this sense the name “scalar product” for evaluation is more analogous to “tensor product” (the first produces a “scalar” or real number, the second a tensor).

Example 1.4.8. dual basis vector projections

When such a tensor product matrix product acts by matrix multiplication on a component vector on the right, it corresponds to evaluating the corresponding $\binom{1}{1}$ -tensor on its second argument

$$\underline{v} \underline{f}^T \underline{X} \leftrightarrow (v \otimes f)(\cdot, X) = e_i v^i f_j X^j = f(X) v.$$

This is exactly how the dual basis 1-forms project out the scalar components along their corresponding basis vectors $\omega^j(X) = X^j$. Multiplying the original basis vector by this scalar component yields the vector component along that basis vector $X^j e_j$ (no sum on j). The sum then recovers the original vector by adding all these separate vector components together. For a new basis $e_{i'} = e_{i'}^j e_j = B^j_{i'} e_j$, with corresponding dual basis $\omega^{i'} = \omega^{i'}_j \omega^j = B^{-1i'}_j \omega^j$, the summed tensor product

$$e_{i'} \otimes \omega^{i'} = B^j_{i'} B^{-1i'}_k e_j \otimes \omega^k = \delta^j_k e_j \otimes \omega^k = e_j \otimes \omega^j$$

has the matrix representation

$$\underline{e}_{i'} \underline{\omega}^{i'T} = \underline{B} \underline{B}^{-1} = \underline{I}$$

Each term in the matrix product sum over i' is the projection matrix which picks out the i' vector component $e_{i'} \otimes \omega^{i'}(\cdot, X) = e_{i'} \omega^{i'}(X) = X^{i'} e_{i'}$ (no sum on i') of the component vector to which it is applied by matrix multiplication, i.e., by evaluation of the corresponding tensor product on its second argument. □

Exercise 1.4.4.

transforming a tensor on \mathbb{R}^2

In Exercise 1.3.2 the dual basis $W^1 = \omega^1 - \omega^2$, $W^2 = -\omega^1 + 2\omega^2$ was found for the new basis $\{E_1, E_2\} = \{\langle 2, 1 \rangle, \langle 1, 1 \rangle\} = \{2e_1 + 1e_2, 1e_1 + 1e_2\}$ on \mathbb{R}^2 . Find the components of the $\binom{1}{1}$ -tensor T in terms of the standard basis $\{e_i\}$ and $\{\omega^i\}$ if T has the following components in terms of the basis $\{E_i\}$:

$$\begin{aligned} T(W^1, E_1) &= 1, & T(W^1, E_2) &= 2, \\ T(W^2, E_1) &= -1, & T(W^2, E_2) &= 0, \end{aligned}$$

i.e.

$$T = 1E_1 \otimes W^1 + 2E_1 \otimes W^2 - 1E_2 \otimes W^1 + 0E_2 \otimes W^2 = T^i_j e_i \otimes \omega^j.$$

Do this in two ways.

a) Just substitute into T the new basis vectors and dual basis covectors expressed in terms

of the old ones and expand out the result to identify the 4 old components as the resulting coefficients of $e_i \otimes \omega^j$.

b) Use the matrix transformation law that will be justified in the next section. With a prime introduced to distinguish components $T^{i'}_{j'}$ in the new basis $E_i = e_{i'}$ and dual basis $W^i = \omega^{i'}$ from those T^i_j in the old basis, then the following matrix product will reproduce our previous result

$$(T^{i'}_{j'}) = \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix}, \quad \underline{T}' = \underline{A} \underline{T} \underline{A}^{-1} \rightarrow \underline{T} = \underline{A}^{-1} \underline{T}' \underline{A} = \underline{B} \underline{T}' \underline{B}^{-1},$$

where the basis changing matrix and its inverse are

$$\underline{B}^{-1} = \underline{A} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = (W^i_j), \quad \underline{B} = \underline{A}^{-1} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = (E^i_j).$$

■

Remark.

Our notation is so compact that certain facts may escape us. For example

$$v \otimes f = (v^i e_i) \otimes f = v^i (e_i \otimes f) = v^i e_i \otimes f$$

is actually a distributive law for the tensor product. A simpler example shows this

$$(u + v) \otimes f = u \otimes f + v \otimes f.$$

How do we know this? Well, the only thing we know about the tensor product is how it is defined in terms of evaluation on its arguments

$$\begin{aligned} [(u + v) \otimes f](g, w) &= g(u + v)f(w) = [g(u) + g(v)]f(w) && \text{(linearity)} \\ &= g(u)f(w) + g(v)f(w) && \text{(distributive law)} \\ &= (u \otimes f)(g, w) + (v \otimes f)(g, w) && \text{(definition of } \otimes \text{)} \\ &= [u \otimes f + v \otimes f](g, w) && \text{(linearity)} \end{aligned}$$

which is “how one adds functions” to produce the sum function, namely by adding their values on arguments. But if these functions inside the square brackets on each side of the equation have the same values on all pairs of arguments, they are the same function (i.e., $\binom{1}{1}$ -tensor), namely $(u + v) \otimes f = u \otimes f + v \otimes f$.

In fact it is easy to show (exercise) that $(cv) \otimes f = c(v \otimes f)$ for any constant c , so in fact the tensor product behaves like a product should with linear combinations.

▲

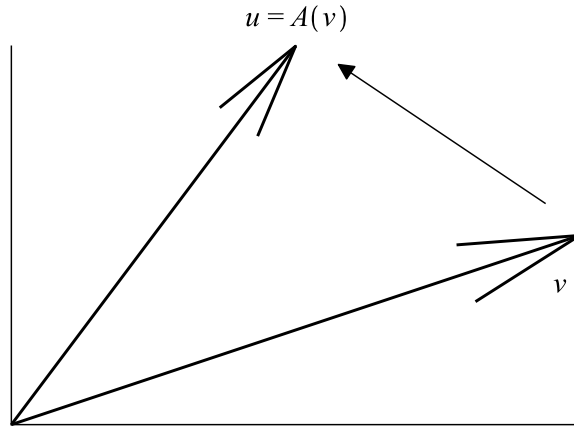


Figure 1.15: Active linear transformation of points: the point v moves to the point $u = A(v)$.

1.5 Linear transformations of V into itself and a change of basis

We have developed the description of a vector space V and the tensor spaces that are defined “over it” (first the dual space and then all of the $\binom{p}{q}$ -tensor spaces as we add more and more upper and lower indices to the component symbols) starting from some fixed basis of V which induces a basis of each of these other tensor spaces in terms of which we can express all tensors in terms of their components. However, the most interesting thing about this description is what happens to those components when we change the basis of V . This is accomplished via the space of $\binom{1}{1}$ -tensors, which we have identified with the space of linear transformations of the vector space V into itself. Partial evaluation of such a tensor on the vector argument leaves a vector value as the result—this is how one accomplishes a linear transformation: $u = u^i e_i \rightarrow A(\ , u) = A^i_j e_i \omega^j(u) = A^i_j u^j e_i$.

Suppose $A : V \rightarrow V$ is a linear transformation of V into itself. If $\{e_i\}$ is a basis of V , then the matrix of A with respect to $\{e_i\}$ is defined by

$$\underline{A} = (A^i_j), \quad A^i_j = \omega^i(A(e_j)) = i\text{-th component of } A(e_j).$$

where i (left) is the row index and j (right) the column index (first and second indices respectively, although the first is a superscript instead of the usual subscript like the second in the usual notation of elementary linear algebra). The j -th column of \underline{A} is the column matrix $\underline{A(e_j)}$ of components of the vector $A(e_j) = A^i_j e_i$ with respect to the basis, denoted by underlining

$$\underline{A} = (\underline{A(e_1)} \ \underline{A(e_2)} \ \dots \ \underline{A(e_n)}).$$

If we expand the equation

$$u = A(v) \rightarrow u^i e_i = A(v^j e_j) = v^j A(e_j) = v^j A^i_j e_i = (A^i_j v^j) e_i,$$

we get the component relation $u^i = A^i_j v^j$ or its matrix form $\underline{u} = \underline{A} \underline{v}$, where

$$\underline{u} = \begin{pmatrix} u^1 \\ \vdots \\ u^n \end{pmatrix}, \quad \underline{v} = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}$$

are the column matrices of components of u and v with respect to the basis.

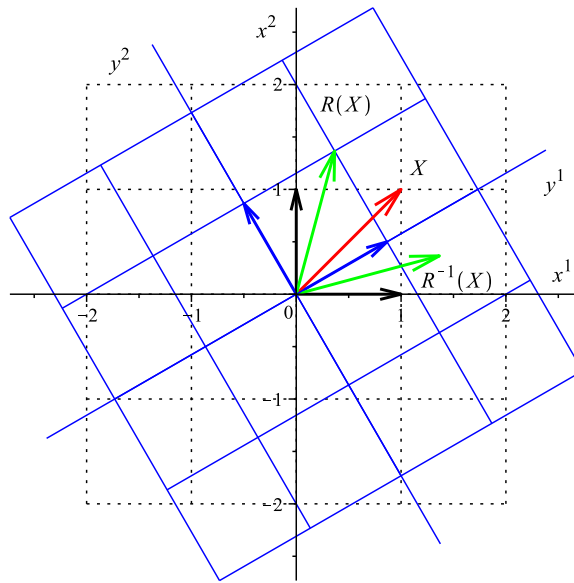


Figure 1.16: Active rotation of the plane showing the old and new bases and the old and new coordinate grids. Notice that the starting vector X has the same relationship to the new axes as the vector $R^{-1}(X)$ rotated by the inverse rotation has to the original axes. In other words matrix multiplication by the inverse matrix gives the new components of a fixed vector with respect to the new rotated basis.

We can interpret this as an “active” linear transformation of the points (vectors) of V to new points of V . We start with a vector v and end up at the new vector u as shown in Fig. 1.15. The rotation of the plane illustrated in Example 1.3.1 is a good example to keep in mind. Fig. 1.16 shows an active rotation of the standard basis and its grid by a 30° rotation.

We can also use a linear transformation to change the basis of V , provided that it is non-singular (its matrix has nonzero determinant), just the condition that the n image vectors of the original basis $\{e_i\}$ are linearly independent so they can be used as a new basis. The point of view here is that general vectors do not move, but they change their components since they are expressed in terms of a new basis which is obtained by moving the old basis by the original

active linear transformation. This mere change of coordinates is sometimes called a passive linear transformation since vectors remain fixed and are simply re-expressed in terms of a new basis which is obtained from the old basis by an active linear transformation.

If $B : V \rightarrow V$ is such a linear transformation, with matrix $\underline{B} = (B^i_j) = (\omega^i(B(e_j)))$ such that $\det \underline{B} \neq 0$, then define $e_{i'} = B(e_i) = B^j_i e_j$. As discussed above, the columns of $\underline{B} = (B(e_1) \cdots B(e_n))$ are the components of a new basis vectors with respect to the old ones: $B^i_j = \omega^i(B(e_j)) \equiv \omega^i(e_{j'})$ are the old components (i) of the j th new basis vector. Primed indices will be associated with component expressions in the new basis.

Since B is invertible, we have

$$e_i = B^{-1}(e_{i'}) = B^{-1j}_i e_{j'},$$

which states that the new components (j) of the old basis vectors (i) are the columns (i fixed, j variable) of the inverse matrix \underline{B}^{-1} . The new basis $\{e_{i'}\}$ has its own dual basis $\{\omega^{i'}\}$ satisfying $\omega^{i'}(e_{j'}) = \delta^i_j$. If we define

$$\omega^{i'} = B^{-1i}_j \omega^j,$$

which says that the rows of the inverse matrix (i fixed, j variable) are the old components of the new dual basis covectors, then

$$\begin{aligned} \omega^{i'}(e_{j'}) &= B^{-1i}_k \omega^k(B^{\ell}_j e_{\ell}) = B^{-1i}_k B^{\ell}_j \delta^k_{\ell} \\ &= B^{-1i}_k B^k_j = \delta^i_j \quad (\text{since } \underline{B}^{-1} \underline{B} = \underline{I}) \end{aligned}$$

confirms that this is the correct expression for the new dual basis.

Given any vector v , we can express it either in terms of the old basis or the new one

$$\begin{aligned} v &= v^i e_i, \quad v^i = \omega^i(v), \\ v &= v^{i'} e_{i'}, \quad v^{i'} = \omega^{i'}(v) = B^{-1i}_j \omega^j(v) = B^{-1i}_j v^j. \end{aligned}$$

In other words, if we actively transform the old basis to a new basis using the linear transformation B , the new components of any vector are related to the old components of the *same* vector by matrix multiplication by the inverse matrix \underline{B}^{-1} as is clear from the rotation example in Fig. 1.16

$$\underline{v}' = \underline{B}^{-1} \underline{v}$$

or equivalently

$$\underline{v} = \underline{B} \underline{v}'.$$

Similarly we can express any covector in terms of the old or new dual basis

$$\begin{aligned} f &= f_i \omega^i, \quad f_i = f(e_i), \\ f &= f_{i'} \omega^{i'}, \quad f_{i'} = f(e_{i'}) = f(B^j_i e_j) = B^j_i f(e_j) = f_j B^j_i, \end{aligned}$$

i.e., the covector components transform by the matrix \underline{B} but multiplying from the right if we represent covectors as row matrices

$$(f_{1'} \cdots f_{n'}) = (f_1 \cdots f_n) \underline{B} \leftrightarrow \underline{f}'^T = \underline{f}^T \underline{B}$$

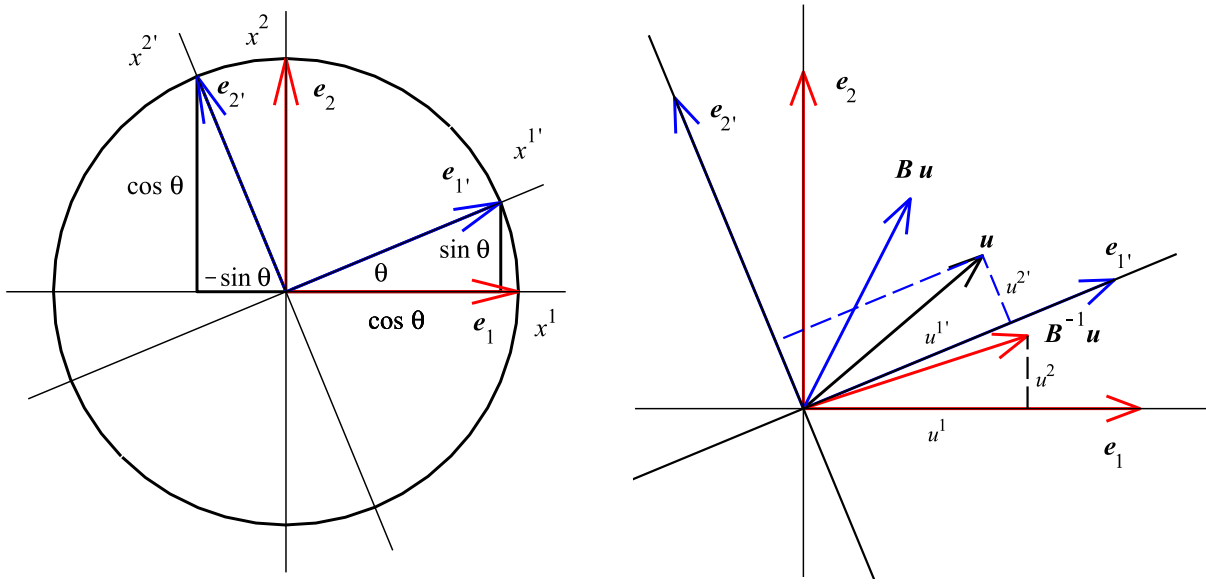


Figure 1.17: Left: the trigonometry of new basis vectors rotated by an angle θ . Right: a point u can be rotated actively by the rotation to a new position Bu in terms of components with respect to the old basis, or it can simply be re-expressed passively in terms of the new rotated basis vectors, with new components $u' = B^{-1}u$, which can be visualized by rotating u in the opposite direction by the angle θ and expressing it with respect to the original basis vectors.

or equivalently

$$\underline{f}^T = \underline{f'}^T \underline{B}^{-1},$$

where the explicit transpose makes it clear that \underline{f}^T and $\underline{f'}^T$ are row vectors, necessary to multiply the square matrix on its left. This describes a “passive” transformation of V into itself or of V^* into itself, since the points of these spaces do not change but their components do change due to the change of basis.

Changing the basis actively by a linear transformation B makes the components of vectors change by the inverse matrix \underline{B}^{-1} of B , while an active transformation of V into itself gives the components with respect to the unchanged basis of the new vectors as the matrix product by \underline{B} with the old components. The active and passive transformations go in opposite directions so to speak.

Example 1.5.1. rotation as a coordinate transformation

Consider a rotation of the plane by an angle θ , imagined as a small positive acute angle for purposes of illustration, see Fig. 1.17. The basis vector $e_1 = \langle 1, 0 \rangle$ is moved to the new basis vector $e_{1'} = \langle \cos \theta, \sin \theta \rangle$, while the basis vector $e_2 = \langle 0, 1 \rangle$ is moved to the new basis vector $e_{2'} = \langle -\sin \theta, \cos \theta \rangle$ by the basic trigonometry shown in that figure, so the matrix whose

columns are the new basis vectors is

$$\underline{B} = \langle \underline{e}_1' | \underline{e}_2' \rangle = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

Any point $\underline{u} = \langle u^1, u^2 \rangle$ in the plane is rotated to its new position $\underline{B}\underline{u}$ as shown in the figure, but we can also re-express the same original vector \underline{u} with respect to the new basis. Its new coordinates are related to its old coordinates by the inverse rotation

$$\underline{u}' = \langle u^{1'}, u^{2'} \rangle = \underline{B}^{-1}\underline{u}.$$

□

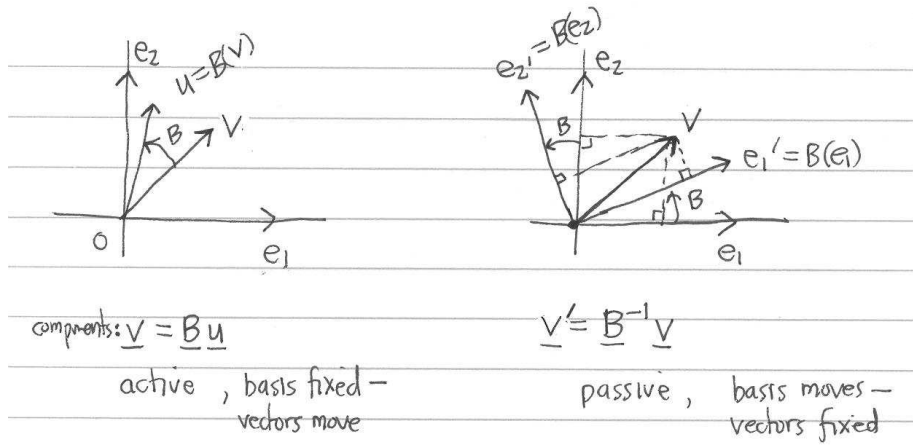


Figure 1.18: Left: active transformation, points move, basis fixed. Right: passive transformation, points fixed, basis changes.

If we are more interested in merely changing bases than in active linear transformations, we can let $A = B^{-1}$ so that the old components of vectors are multiplied by the matrix rather than the inverse matrix. Then we have

$$\begin{aligned} \omega^{i'} &= A^i_{j'} \omega^j \longrightarrow v^{i'} = A^i_{j'} v^j, \\ e_{i'} &= A^{-1j'}_i e_j \longrightarrow f_{i'} = f_j A^{-1j'}_i, \end{aligned}$$

Thus upper indices associated with vector component labels transform by the matrix \underline{A} (whose rows are the old components of the new dual basis covectors), while lower indices associated with covector component labels transform by the matrix \underline{A}^{-1} (whose columns are the old components of the new basis vectors).

In the jargon of this subject, these upper indices on components are called “contravariant” while the lower indices on components called “covariant”. Vectors and covectors themselves are sometimes called “contravariant vectors” and “covariant vectors” respectively. The above

relations between old and new components of the same object are called “transformation laws” for contravariant and covariant vector components.

By the linearity of the tensor product, these “transformation laws” can be extended to the components of any tensor. For example, suppose $L = L^i_j e_i \otimes \omega^j$ is the $\binom{1}{1}$ -tensor associated with a linear transformation $L : V \rightarrow V$, now using the same symbol for the linear transformation and the tensor. Then

$$\begin{aligned} L &= L^i_j e_i \otimes \omega^j, \quad L^i_j = L(\omega^i, e_j), \\ L &= L^{i'}_{j'} e_{i'} \otimes \omega^{j'}, \quad L^{i'}_{j'} = L(\omega^{i'}, e_{j'}) = L(A^i_k \omega^k, A^{-1\ell}_j e_\ell) = A^i_k A^{-1\ell}_j L(\omega^k, e_\ell) \\ &= A^i_k A^{-1\ell}_j L^k_\ell. \end{aligned}$$

In other words the contravariant and covariant indices each transform by the appropriate factor of A^i_j or A^{-1i}_j

$$L^{i'}_{j'} = A^i_k A^{-1\ell}_j L^k_\ell \quad \text{or inversely} \quad L^i_j = A^{-1i}_k A^\ell_j L^{k'}_{\ell'}.$$

This generalizes in an obvious way to any $\binom{p}{q}$ -tensor

$$\begin{aligned} T &= T^{i\cdots}_{j\cdots} e_i \otimes \cdots \otimes \omega^j \otimes \cdots, \quad T^{i\cdots}_{j\cdots} = T(\omega^i, \dots, e_j, \dots), \\ T &= T^{i'}_{j'\cdots} e_{i'} \otimes \cdots \otimes \omega^{j'} \otimes \cdots, \quad T^{i'}_{j'\cdots} = T(\omega^{i'}, \dots, e_{j'}, \dots) \\ &= A^i_k \cdots A^{-1\ell}_j \cdots T^{k\cdots}_{\ell\cdots}. \end{aligned}$$

It is just a simple consequence of multilinearity.

Example 1.5.2. transforming the identity tensor

We first defined the Kronecker delta just as a convenient shorthand symbol δ^i_j , but then saw it coincided with the components of the evaluation or identity tensor

$$Id = \delta^i_j e_i \otimes \omega^j = e_i \otimes \omega^i = e_1 \otimes \omega^1 + \cdots + e_n \otimes \omega^n.$$

Since this must be true in any basis, if we “transform” the Kronecker delta as the components of a $\binom{1}{1}$ -tensor, it should be left unchanged

$$\delta^{i'}_{j'} = A^i_k A^{-1\ell}_j \delta^k_\ell = A^i_k A^{-1k}_j = (\underline{A} \underline{A}^{-1})^i_j = (I)^i_j = \delta^i_j.$$

The new components do equal the old!

□

Matrix form of the “transformation law” for $\binom{1}{1}$ -tensors

The “transformation law” for the $\binom{1}{1}$ -tensor L associated with a linear transformation $L : V \rightarrow V$ is

$$L^{i'}_{j'} = A^i_k A^{-1\ell}_j L^k_\ell = A^i_k L^k_\ell A^{-1\ell}_j = [\underline{A} \underline{L} \underline{A}^{-1}]^i_j.$$

In other words we recover the matrix transformation for a linear transformation under a change of basis discussed in the eigenvector problem

$$\underline{L}' = \underline{A} \underline{L} \underline{A}^{-1}$$

which leads to the conjugation operation (just a term for sandwiching a matrix between another matrix and its inverse), except that in the eigenvector change of basis discussion, this relation was written in terms of the inverse matrix $\underline{A}^{-1} = \underline{B}$

$$\underline{L}' = \underline{B}^{-1} \underline{L} \underline{B} \leftarrow \text{columns of } \underline{B} = \text{old components of new basis vectors}$$

which corresponds to the transformation of vector components

$$\underline{v} = \underline{B} \underline{v}', \quad \underline{v}' = \underline{B}^{-1} \underline{v}.$$

Note that when one succeeds in finding a square matrix $\underline{B} = \langle \underline{b}_1 | \dots | \underline{b}_n \rangle$ of linearly independent eigenvectors of a matrix \underline{L} (namely $\underline{L} \underline{b}_i = \lambda_i \underline{b}_i$ so that $\underline{L} \underline{B} = \langle \underline{L} \underline{b}_1 | \dots | \underline{L} \underline{b}_n \rangle = \langle \lambda_1 \underline{b}_1 | \dots | \lambda_n \underline{b}_n \rangle$), then the new components of the matrix with respect to a basis consisting of those eigenvectors is diagonal

$$\underline{L}' = \underline{B}^{-1} \underline{L} \underline{B} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix},$$

with the corresponding eigenvalues λ_i along the main diagonal. This process is called the diagonalization of the matrix.

Matrices of symmetric $\binom{0}{2}$ -tensors

In elementary linear algebra, no distinction is made between matrices associated with linear transformations (namely $\binom{1}{1}$ -tensors) and matrices which are associated with bilinear functions of a pair of vectors (namely $\binom{0}{2}$ -tensors). Under general changes of basis, they transform very differently, although under orthogonal transformations normally considered in that context, the distinction disappears, as we will see later. So let's confront this difference up front by considering bilinear tensors.

A $\binom{0}{2}$ -tensor $G = G_{ij} \omega^i \otimes \omega^j$ takes 2 vector arguments $G(u, v) \in \mathbb{R}$. The transformation law is

$$G_{i'j'} = A^{-1m}{}_i A^{-1n}{}_j G_{mn} = A^{-1m}{}_i G_{mn} A^{-1n}{}_j = [(\underline{A}^{-1})^T \underline{G} \underline{A}^{-1}]_{ij} = [\underline{B}^T \underline{G} \underline{B}]_{ij}.$$

Note that the indices n are adjacent in the second equation product, but the indices m are not, which requires the transpose to get them into the proper position for the matrix product between two matrices (the right index of the left factor summed against the left index of the right factor). Although G also has a matrix representation $\underline{G} = (G_{ij})$ in component form, its matrix transformation law involves the transpose, rather than the inverse as for the $\binom{1}{1}$ -tensor.

The transpose is necessary to represent to summation of the (left upper) row index of \underline{A}^{-1} against the (lower left) row index of \underline{G} in terms of matrix multiplication.

A well known fact from linear algebra is that a symmetric matrix can be diagonalized by an orthogonal transformation of the basis. The symmetry $\underline{A}^T = \underline{A}$ for a matrix \underline{A} makes sense index-wise only for interchange of two indices at the same level: $A_{ji} = A_{ij}$ and not for the matrix of a linear transformation where A^j_i and A^i_j cannot really be compared in any meaningful way. However, if only orthogonal matrixes \underline{B} are used (the restriction to orthonormal coordinates), for which the transpose is the inverse: $\underline{B}^{-1} = \underline{B}^T$, then the matrix form of this transformation law for a $\binom{0}{2}$ -tensor is equivalent to the usual one for a $\binom{1}{1}$ -tensor of the eigenvector discussion:

$$\underline{A} \rightarrow \underline{B}^{-1} \underline{A} \underline{B}.$$

Thus if only orthonormal bases are considered, there is no difference between the transformation laws for a $\binom{1}{1}$ or $\binom{0}{2}$ tensor or even a $\binom{2}{0}$ -tensor (apart from the switch to the inverse matrix everywhere), allowing this distinction enforced by upper/lower index positions to remain hidden in those elementary discussions. But hiding this structure requires the additional operation of the dot product to identify the vector space and its dual space, a concept which is just avoided for simplicity at an elementary level.

A 3-index example

We can consider the change in components under a change of basis for any $\binom{p}{q}$ -tensor, but let's wait on that. The determinant is a useful example, however, since it is another major ingredient of our elementary linear algebra landscape.

Example 1.5.3. determinants and Levi-Civita: tensor densities

As already discussed in Exercise 1.4.7, the triple scalar product on \mathbb{R}^3 is a multilinear function on triplets of vectors, namely the determinant function, which is a $\binom{0}{3}$ -tensor when thought of as a function on its columns

$$D(u, v, w) = u \cdot (v \times w) = \det \begin{pmatrix} u^1 & v^1 & w^1 \\ u^2 & v^2 & w^2 \\ u^3 & v^3 & w^3 \end{pmatrix} = \epsilon_{ijk} u^i v^j w^k.$$

The components $D_{ijk} = \epsilon_{ijk}$ with respect to the standard basis of \mathbb{R}^3 define the Levi-Civita symbol.

Suppose we evaluate the new components of the determinant tensor under a change of basis away from the standard basis.

$$D = D_{ijk} \omega^i \otimes \omega^j \otimes \omega^k = D_{i'j'k'} \omega^{i'} \otimes \omega^{j'} \otimes \omega^{k'}.$$

Then from the rules for determinants

$$\begin{aligned}
 D_{i'j'k'} &= A^{-1m}_i A^{-1n}_j A^{-1p}_k D_{mnp} = \epsilon_{mnp} A^{-1m}_i A^{-1n}_j A^{-1p}_k \\
 &= \sum_{\sigma} (-1)^{\text{sgn } \sigma} A^{-1\sigma(1)}_i A^{-1\sigma(2)}_j A^{-1\sigma(3)}_k \\
 &= \begin{cases} \det \underline{A}^{-1}, & \text{if } (i, j, k) = (1, 2, 3) \text{ by definition,} \\ \det \underline{A}^{-1}, & \text{if } (i, j, k) = \text{positive permutation of } (1, 2, 3), \\ -\det \underline{A}^{-1}, & \text{if } (i, j, k) = \text{negative permutation of } (1, 2, 3), \\ 0 & \text{otherwise (repeated rows).} \end{cases} \\
 &= (\det \underline{A}^{-1}) \epsilon_{ijk}.
 \end{aligned}$$

In other words the new components of the determinant function differ from the old by the factor $\det \underline{A}^{-1}$. So the symbol ϵ_{ijk} *does not* define a tensor in the sense that its components in any basis have these same values, as does the Kronecker delta symbol δ^i_j which represents the components of the identity tensor. Instead it is similar to the Kronecker delta δ_{ij} or δ^{ij} which do not retain their same numerical component values under a change of basis.

Another way of starting this is that this Levi-Civita symbol ϵ_{ijk} (also sometimes called the “alternating symbol” because of its alternating signed value) defines a different tensor for each choice of basis

$$D_{(e)} = \epsilon_{ijk} \omega^i \otimes \omega^j \otimes \omega^k \neq D_{(e')} = \epsilon_{ijk} \omega^{i'} \otimes \omega^{j'} \otimes \omega^{k'}.$$

So the important lesson from this example is, if we define an object with indices not by taking components of some tensor, then it is not necessarily a tensor—but may define a different tensor in each choice of basis.

Another way of handling this particular problem with the alternating symbol is to generalize the idea of a tensor (independent of the choice of basis) to a “tensor density” which is a family of tensors, one in each choice of basis, related by a more general transformation law which not only changes the components of the tensor but also changes the tensor itself by an overall factor of some power of the determinant of the transformation.

Dividing through the above equation by the determinant factor gives

$$\epsilon_{ijk} = (\det A^{-1})^{\overbrace{-1}^{\text{“weight”}}} A^{-1m}_i A^{-1n}_j A^{-1p}_k \epsilon_{mnp},$$

where ϵ_{mnl} are the old components of an old tensor, $A^{-1m}_i A^{-1n}_j A^{-1p}_k \epsilon_{mnp}$ are the new components of the old tensor, ϵ_{ijk} are the new components of new tensor which is scaled from the old one by the factor $(\det A^{-1})^{-1}$ whose power $W = -1$ of the determinant of the inverse matrix is called the weight of the tensor density. This then becomes the transformation law for a tensor density of weight -1 , whose old components are ϵ_{mnl} and whose new components are ϵ_{ijk} , which are numerically the same.

Summarizing, the alternating symbol ϵ_{ijk} may be interpreted (by definition) as the components of an antisymmetric $\binom{0}{3}$ -tensor density of weight -1 . This tensor density has the form $\epsilon_{ijk} \omega^i \otimes \omega^j \otimes \omega^k$ in any basis, with numerically constant components, like the Kronecker delta

identity tensor. This generalizes to an alternating symbol $\epsilon_{i_1 \dots i_n}$ on any n -dimensional vector space V . We will return to this below. \square

Suppose that instead of making the matrix of the linear transformation explicit with its own kernel symbol A , we introduce the following component notation for the new basis vectors and covectors

$$e_{i'} = A^{-1j}{}_i e_j = e^j{}_{i'} e_j, \quad \omega^{i'} = A^i{}_j \omega^j = \omega^{i'}{}_j \omega^j.$$

Note that the rows of $\underline{A} = \langle (\omega^{1'})^T, \dots, (\omega^{n'})^T \rangle$ are the old components of the new dual basis covectors (recall that $(\omega^i)^T$ is a row matrix), while the columns of $\underline{A}^{-1} = \langle \underline{e}_{1'} | \dots | \underline{e}_{n'} \rangle$ are the old components of the new basis vectors.

Duality then just requires that the coefficient matrices be inverse matrices

$$\omega^{i'}(e_{j'}) = \omega^{i'}{}_k e^k{}_{j'} = \delta^i{}_j \leftrightarrow \underline{A} \underline{A}^{-1} = \underline{I}.$$

But the matrix product in the other order is also valid and implies something different

$$\underline{A}^{-1} \underline{A} = \underline{I}, \quad e^i{}_{k'} \omega^{k'}{}_j = \delta^i{}_j,$$

namely, that the identity tensor can be expressed in terms of them in this way, which then further implies

$$e_i \otimes \omega^i = \delta^i{}_j e_i \otimes \omega^j = e^i{}_{k'} \omega^{k'}{}_j e_i \otimes \omega^j = e_{k'} \otimes \omega^{k'}.$$

In other words the identity tensor has the same form in either basis.

Example 1.5.4. eigenvectors in the plane

Consider the matrix \underline{A} and its matrix of eigenvectors $\underline{B} = \langle \underline{b}_1 | \underline{b}_2 \rangle$ corresponding to the eigenvalues $(\lambda_1, \lambda_2) = (5, -1)$

$$\underline{A} = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}, \quad \underline{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix}, \quad \underline{B}^{-1} = \frac{1}{3} \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}.$$

Introducing new coordinates $\{y^1, y^2\} \equiv \{x^{1'}, x^{2'}\}$ with respect to the new basis vectors $\{\vec{b}_1, \vec{b}_2\} \equiv \{e_{1'}, e_{2'}\} = \{\langle 1, 1 \rangle, \langle -2, 1 \rangle\}$, the unit grid associated with these new coordinates for the ranges $y^1 = -2..2, y^2 = -2..2$ is illustrated in Fig. 1.19. When the two grids intersect at a grid point, one has vectors with integer coordinates in both systems that are easily read off the grid. Notice that the point $(x^1, x^2) = (4, 1)$ has coordinates $(y^1, y^2) = (2, -1)$, for example. What are the new components of the point $(0, 3)$? What are the old components of the point with $(y^1, y^2) = (2, 2)$?

If we make the coordinate transformation and its inverse explicit, one finds

$$\begin{aligned} \underline{x} &= \underline{B} \underline{y} : & \underline{y} &= \underline{B}^{-1} \underline{x} : \\ x^1 &= y^1 - 2y^2, & y^1 &= \frac{1}{3}x^1 + \frac{2}{3}x^2, \\ x^2 &= y^1 + y^2, & y^2 &= -\frac{1}{3}x^1 + \frac{1}{3}x^2. \end{aligned}$$

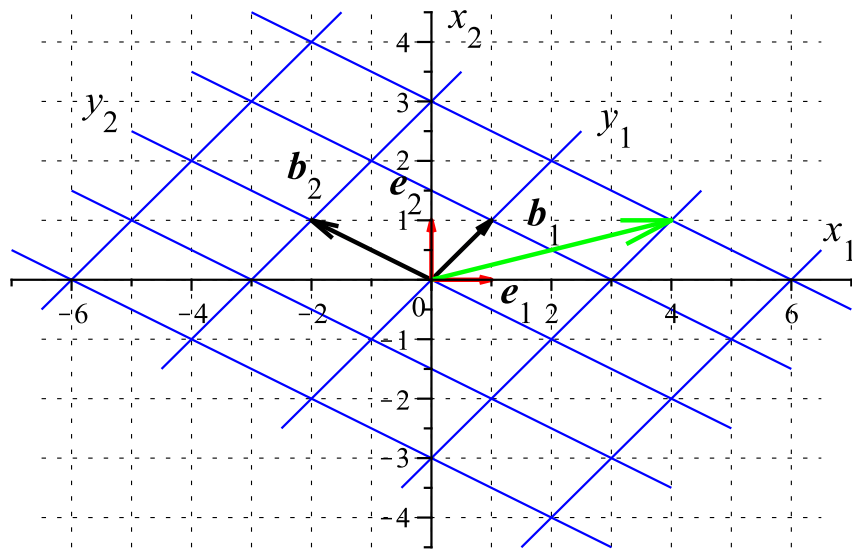


Figure 1.19: The new coordinate grid associated with a new basis of the plane is obtained as the image of the old coordinate grid under the active linear transformation $e_i \rightarrow B^j_i e_j$ under which a point with old coordinates (u^1, u^2) goes to a point with new coordinates which are the same: $B(u) = B(u^i e_i) = u^i B(e_i) = u^i e_{i'} = u^1 b_1 + u^2 b_2$. Thus to find the new coordinates of a point on the new grid, one has to apply the inverse linear transformation to its old coordinates to find the original point from which it came. Under this active deformation of the plane the basic grid square $0 \leq x^1 \leq 1, 0 \leq x^2 \leq 1$ is actively deformed into the parallelogram formed by the new basis vectors b_1 and b_2 .

The pair of lines $y^1 = 0, y^1 = 1$ represents the 1-form $\omega^{1'}$, while the pair of lines $y^2 = 0, y^2 = 1$ represents $\omega^{2'}$. These 2 pairs enclose the basic unit parallelogram of the new coordinate grid $y^1 = 0..1, y^2 = 0..1$. Note that there is a basic symmetry between old and new coordinates. \square

Remark.

Sometimes it is more useful to introduce new “kernel letters” (the letter symbols to which indices are added) instead of using primed indices as we have done here for the new coordinates. In Fig. 1.19 we also used boldface letters to distinguish the basis vectors rather than arrow notation, as well as the more common subscripted variables that are natural in Maple. The important thing is to be flexible with notation so that more than one common choice can be made depending on the circumstances. \blacktriangle

Exercise 1.5.1.

eigenvectors of a matrix of eigenvectors

In Example 1.5.4, analyze the active deformation of the plane by the linear transformation B with matrix \underline{B} by finding its eigenvalues and eigenvectors (as opposed to those of the matrix \underline{A} for which \underline{B} is a matrix of eigenvectors). Namely, solve the characteristic equation for the eigenvalues of \underline{B} , then back substitute each such eigenvalue into the linear system which produces the corresponding eigenvector. Look it up if you have forgotten this process or simply get the result directly from a computer algebra system. Plot them on the grid shown in Fig. 1.20. Do these directions for the stretch and compression make sense? \blacksquare

Exercise 1.5.2.

changing coordinates in the plane

On some unit square grid paper in a field of view $x^1 = -6..6, x^2 = -8..8$ (print a blank plot with grid lines in a computer algebra system), draw in the unit grid for $y^1 = -2..2, y^2 = -2..2$ for the new coordinates associated with following matrix and its matrix of eigenvectors

$$\underline{A} = \begin{pmatrix} 7 & -4 \\ 6 & -7 \end{pmatrix}, \quad \underline{B} = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}, \quad \underline{B}^{-1} = \frac{1}{5} \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}.$$

From the grid read off the new components of the point $(5, 5)$. What point has new coordinates $(1, -1)$? Use the matrix transformation $\underline{x} = \underline{B}\underline{y}$ with inverse $\underline{y} = \underline{B}^{-1}\underline{x}$ to confirm your graphical results. \blacksquare

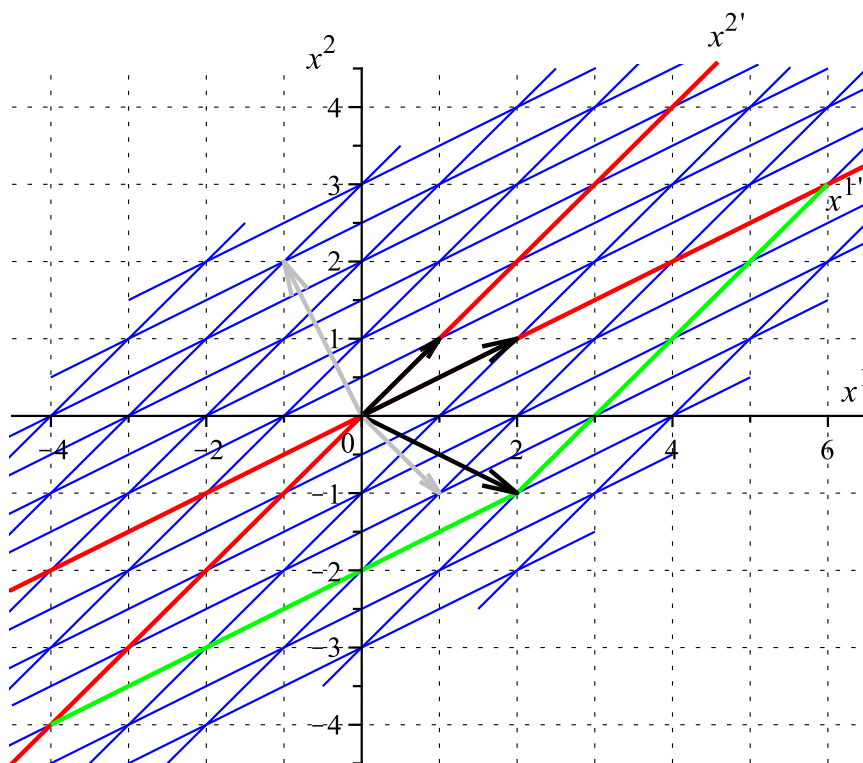


Figure 1.20: The new grid associated with the new basis vectors $\langle 2, 1\rangle, \langle 1, 1\rangle$ is obtained by a deformation of the old coordinate grid. For example the vector $\langle 2, -1\rangle$ is sent to the vector $2b_1 - b_2 = \langle 3, 1\rangle$ under this deformation. Shown also in gray are the two vectors with the same components as the new dual basis covectors $\langle 1| - 1\rangle, \langle -1| 2\rangle$, which are each perpendicular to the coordinate lines of the other coordinate. These two vectors also form a basis of the plane, called the “reciprocal basis,” and are often used to avoid mention of a distinction between vectors and covectors.

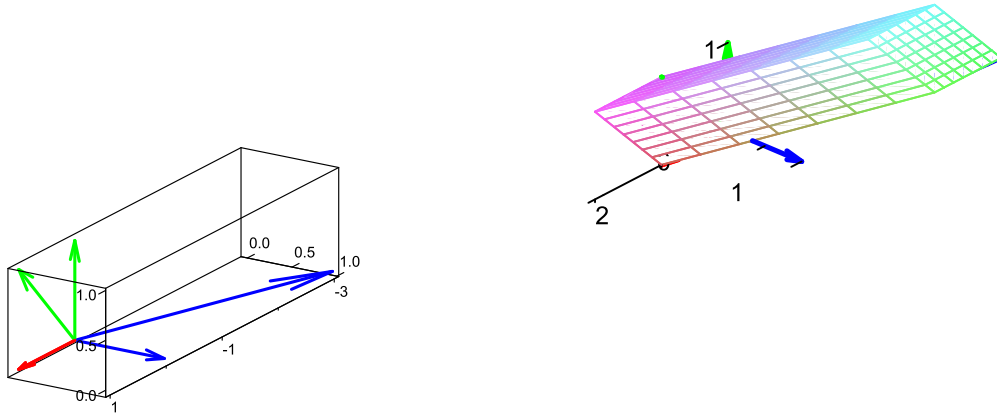


Figure 1.21: The new coordinate grid unit parallelepiped (right) associated with a new basis of space (left), shown together with the original standard basis.

Example 1.5.5. eigenvectors in \mathbb{R}^3

Consider the upper triangular matrix \underline{A} and its upper triangular matrix of eigenvectors $\underline{B} = \langle \underline{b}_1 | \underline{b}_2 | \underline{b}_3 \rangle$ corresponding to the eigenvalues $(\lambda_1, \lambda_2, \lambda_3) = (3, 1, 1)$

$$\underline{A} = \begin{pmatrix} 3 & 6 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \underline{B} = \begin{pmatrix} 1 & -3 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \underline{B}^{-1} = \begin{pmatrix} 1 & 3 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Let $\{b_i\}$ be a new basis of R^3 , with dual basis $\{\beta^i\}$. Then the old and new bases contain one element $b_1 = e_1$ in common. Since the e_1 - e_2 plane coincides with the b_1 - b_2 plane, the third dual basis covectors must be proportional $\beta^3 \propto \omega^3$ in order that the new one also have this plane as its zero value surface; in fact they agree $\beta^3 = \omega^3$ since b_3 has the same height as e_3 . Figure 1.20 shows the unit parallelepiped formed by the new basis vectors as edges bounded by the sides composed of the three pairs of representative planes of values 0 and 1 of the new dual basis covectors, i.e., corresponding to the new coordinate ranges $y^i = 0..1$. This is best viewed in a computer algebra system where the object can be rotated and viewed with some transparency.

The eigenvectors of the matrix \underline{A} enable us to interpret the linear transformation $x^i \rightarrow A^i_j x^j$. All points along the first eigenvector are scaled up by a factor of 3 (the eigenvalue), while all points in the 2-plane spanned by the remaining eigenvectors (namely the eigenspace associated with the eigenvalue 1) remain fixed. Thus any region of space will be stretched along the first eigenvector direction, but its cross-sections parallel to the second eigenspace will remain fixed in shape as they move apart from each other along the first eigenvector direction.

**Exercise 1.5.3.****changing coordinates in \mathbb{R}^3**

For the matrices of the previous example and the associated new coordinate system, what are the new coordinates of the point $\langle x^1, x^2, x^3 \rangle = \langle 1, 1, 1 \rangle$? What point has new coordinates $\langle y^1, y^2, y^3 \rangle = \langle 1, 1, 1 \rangle$?



1.6 Linear transformations between V and V^*

So far we've generalized vectors, covectors, and $\binom{1}{1}$ -tensors from column matrices, row matrices, and the square matrices of linear transformations from a vector space into itself (and contrasted the latter briefly with the symmetric matrices of bilinear forms), but have not considered the relationship of the square matrices of components of $\binom{0}{2}$ -tensors and $\binom{2}{0}$ -tensors to linear transformations. These tensors may in fact be interpreted as defining linear transformations between the vector space and its dual space going in opposite directions.

For example, suppose $\ell : V \longrightarrow V^*$ is a linear map (“ ℓ ” for “l”ower), defining a covector $\ell(v)$ for each vector v . In component form we can represent this covector by

$$[\ell(v)]_i = \ell_{ij} v^j \leftrightarrow (\underline{\ell(v)})^T = (\underline{\ell} \underline{v})^T,$$

which can then be evaluated on a vector u to obtain a scalar $\ell(v)(u)$. Using the same symbol, define the associated $\binom{0}{2}$ -tensor $\ell = \ell_{ij} \omega^i \otimes \omega^j$ by

$$\ell(u, v) \equiv \ell(v)(u) = (\ell_{ij} v^j) u^i = \ell_{ij} u^i v^j,$$

where the components of the tensor are defined as usual by

$$\ell_{ij} = \ell(e_i, e_j) = \ell(e_j)(e_i).$$

The linear map ℓ is realized by evaluating the second argument of the corresponding tensor ℓ on a vector, so that a covector remains waiting for the first argument of the tensor, using the same symbol for the linear transformation and the corresponding tensor: $\ell(\cdot, v) = \ell(v)$.

In component form the transformation is again matrix multiplication but because both matrix indices are down, one is left with a covector, requiring an additional transpose in matrix form to yield a row matrix as agreed for representing covectors. These linear transformations *LOWER* the index position.

In exactly same way a linear map $r : V^* \longrightarrow V$ (“ r ” for “r”aise), defining a vector $u = r(g)$ for each covector g , in component form

$$r^{ij} g_j = u^i,$$

has an associated $\binom{2}{0}$ -tensor $r = r^{ij} e_i \otimes e_j$

$$r(f, g) \equiv f(r(g)) = f_i (r^{ij} g_j) = r^{ij} f_i g_j,$$

where

$$r^{ij} = r(\omega^i, \omega^j) = \omega^i(r(\omega^j)).$$

In matrix form a transpose is needed to make the covector \underline{g}^T (row matrix) into a column matrix \underline{g} for matrix multiplication to produce again a column vector

$$\underline{r(g)} = \underline{r} \underline{g}.$$

This linear map is realized by evaluating the second argument of the corresponding tensor, leaving the first argument of the tensor waiting for a vector to be evaluated on. These linear transformations *RAISE* the index position.

Invertible maps between V and V^*

The images of the basis or dual basis vectors under such maps are

$$\ell(e_i) = \ell_{ij}\omega^j, \quad r(\omega^i) = r^{ij}e_j.$$

The condition that these maps be invertible is just that their corresponding matrices be invertible, i.e., have nonzero determinants so that their inverses exist: $\det(\ell_{ij}) \neq 0$, $\det(r^{ij}) \neq 0$ implies that $\underline{\ell}^{-1} \equiv (\ell^{ij})$ and $\underline{r}^{-1} \equiv (r_{ij})$ exist such that the corresponding matrix products are the identity matrix

$$\ell^{ij}\ell_{jk} = \delta^i_k = \ell_{kj}\ell^{ji}, \quad r^{ij}r_{jk} = \delta^i_k = r_{kj}r^{ji}.$$

These are the matrices of linear maps $\ell^{-1} : V^* \rightarrow V$ and $r^{-1} : V \rightarrow V^*$. Either pair, consisting of a tensor and its “inverse” tensor (characterized by having component matrices which are inverses), whether we start with ℓ or r (set $r = \ell^{-1}$ or $\ell = r^{-1}$ for example) establishes an isomorphism between the vector space and its dual.

Although one can use an arbitrary nonsingular matrix (ℓ_{ij}) and its inverse (ℓ^{ij}) to play this game, in practice only two special kinds of such matrices are used, either symmetric or antisymmetric matrices

- symmetric: $\ell_{ij} = \ell_{ji}$ or $\ell(v, u) = \ell(u, v)$,
- antisymmetric: $\ell_{ij} = -\ell_{ji}$ or $\ell(v, u) = -\ell(u, v)$.

The corresponding tensors are also called symmetric or antisymmetric. A symmetric tensor is said to define an inner product, while an antisymmetric tensor defines a symplectic form over an even dimensional vector space and is also important in spinor algebra (also involving even-dimensional spaces), both more sophisticated notions that are important in physics. Inner products are also referred to as metrics in the context of differential geometry.

Remark.

The antisymmetric case of a symplectic form describes the geometry of Hamiltonian mechanics which is a crucial part of the foundation of classical and quantum mechanics in physics. Unfortunately to appreciate this, one must have an advanced knowledge of mechanics which includes the variational approach to the equations of motion through Lagrangians and Hamiltonians. We will pass on that here. ▲

Inner products

An inner product on a vector space V is just a symmetric bilinear function of its elements, i.e., it is a symmetric $\binom{0}{2}$ -tensor over V , and hence is represented by its symmetric matrix of components with respect to any basis $\{e_i\}$ of V

$$G(u, v) = G_{ij}u^iv^j, \quad G = G_{ij}\omega^i \otimes \omega^j, \quad G_{ij} = G(e_i, e_j) = G_{ji}.$$

In the older language, an inner product is referred to as a quadratic form since the repeated evaluation on a single vector $G(u, u)$ is a quadratic function of the vector components. Any

symmetric $n \times n$ matrix \underline{G} with nonzero determinant defines such a “nondegenerate” inner product on \mathbb{R}^n , or on a general vector space V in a given basis.

For a nondegenerate inner product, the determinant of its matrix of components should be nonzero: $\det \underline{G} \neq 0$, which if true with respect to one basis, will be true in any other as we will see below. A nondegenerate inner product establishes a 1-1 map from the vector space V to its dual space which is therefore a “vector space isomorphism” (any two n -dimensional vector spaces are isomorphic since the component vector of vectors with respect to any basis acts like a vector in \mathbb{R}^n under linear operations). Since any symmetric matrix can be diagonalized, one can consider the number of positive and negative signs of the n nonzero eigenvalues of a nondegenerate inner product component matrix: this turns out to be is an invariant, i.e., independent of the basis used to express that matrix. If they are all positive (negative), the inner product is called positive-definite (negative-definite), otherwise indefinite. The number of positive signs minus the number of negative signs is called the signature of the inner product.

Given any inner product G on a vector space V we can always use the dot product notation by defining

$$G(u, v) \equiv u \cdot v.$$

The self-dot-product $G(v, v) = v \cdot v$ contains two independent pieces of information: its sign (the “type”: $+$, $-$, 0) and its absolute value. Define the magnitude or length of a vector and its sign (or “type”) by

$$\begin{aligned} \|v\| &= |v \cdot v|^{1/2} = |G(v, v)|^{1/2}, \\ \operatorname{sgn} v &= \operatorname{sgn}(v \cdot v) = \operatorname{sgn}(G(v, v)) \in \{+1, 0, -1\}. \end{aligned}$$

A vector v with $\|v\| = 1$ is called a unit vector, while a nonzero vector with $\|v\| = 0$ is called a null vector. Dividing a vector with nonzero length by that length produces a unit vector, namely a vector for which the self-inner-product is just the sign ± 1 of the vector

$$\hat{v} \equiv v/\|v\| = v/|G(v, v)|^{1/2} \rightarrow G(\hat{v}, \hat{v}) = \pm 1.$$

This process is called normalizing the vector. If a nondegenerate inner product is positive-definite or negative-definite, then any nonzero vector must have nonzero length, but in the indefinite case nonzero vectors may have zero length. Such null vectors cannot be normalized.

When the inner product of two vectors is zero $G(u, v) = 0$, the two vectors are called orthogonal. A basis $\{e_i\}$ consisting of mutually orthogonal vectors is called an orthogonal basis. A basis consisting of mutually orthogonal unit vectors is called orthonormal

$$G_{ij} = \pm \delta_{ij}.$$

The component matrix is diagonal and each diagonal entry is ± 1 . The difference $s = P - M$ (Plus/Minus) in the number of positive and negative signs is called the signature and is fixed for a given inner product (accept as a fact for now; these are just the signs of the eigenvalues of the symmetric matrix—these signs turn out to be invariant under a change of basis). A “positive-definite” inner product has all positive signs, i.e., signature $s = n$, while a “negative-definite” inner product has all negative signs, i.e., signature $s = -n$. An “indefinite” inner

product has a signature s in between these two extreme values. A “Lorentz” inner product has only one negative sign or only one positive sign (the choice depends on prejudice, motivated by convenience of competing demands) and so the absolute value of the signature is $|s| = (n-1) - 1 = n-2$. Since $n = P + M$, one gets the relation $M = (n-s)/2$. For a Lorentz inner product with only 1 negative sign, the determinant of the component matrix is always negative. For example, the standard basis of \mathbb{R}^n with its usual Euclidean inner product is orthonormal. The standard basis of \mathbb{R}^4 with the Minkowski inner product is also orthonormal (sometimes called pseudo-orthonormal). Note that in this case the product of the signs associated with the basis vectors is negative: $(-1)(+1)(+1)(+1) = -1 = \text{sgn det } \underline{G}$.

Example 1.6.1. dot product on \mathbb{R}^n

On \mathbb{R}^n with the standard basis, the dot product defines a particular positive-definite inner product

$$G(u, v) = u \cdot v = \delta_{ij} u^i v^j = u^1 v^1 + \dots + u^n v^n,$$

where

$$G_{ij} = G(e_i, e_j) = e_i \cdot e_j = \delta_{ij}.$$

Then

$$G(x, x) = \delta_{ij} x^i x^j = (x^1)^2 + \dots + (x^n)^2$$

is interpreted as the square of the distance of the point (x^1, \dots, x^n) from the origin $(0, \dots, 0)$ or the square of the length of the vector $\vec{x} = \langle x^1, \dots, x^n \rangle \geq 0$, which is always positive unless $\vec{x} = 0$. Sometimes n -dimensional Euclidean space (\mathbb{R}^n with the usual dot product) is designated by E^n or \mathbb{E}^n to emphasize its geometry.

Note that if you change from the standard basis to an arbitrary basis of \mathbb{R}^n , the components of G will change

$$G_{i'j'} = e_{i'} \cdot e_{j'} = A^{-1m}{}_i \delta_{mn} A^{-1n}{}_j, \quad \underline{G}' = \underline{A}^{-1T} \underline{I} \underline{A}^{-1} = \underline{A}^{-1T} \underline{A}^{-1}.$$

Only if the new basis is also orthonormal so that $(A^i{}_j)$ is an orthogonal matrix satisfying $\underline{A}^T \underline{A} = \underline{I} = \underline{A} \underline{A}^T$ (just the condition that the columns of the matrix are mutually orthogonal unit vectors), therefore equivalent to the condition $A^T = A^{-1}$ which in turn implies $\underline{A}^{-1T} \underline{A}^{-1} = \underline{I}$, will the new matrix of components again equal (δ_{ij}) . In other words the symbol δ_{ij} defines a different tensor in each basis unless one restricts the change of basis to only orthonormal bases corresponding to basis-changing matrices which are orthogonal. In a nonorthonormal basis, the values $G_{i'i'} \neq 1$ break the normality (unit vector) condition, while the values $G_{i'j'} \neq 0$ ($i \neq j$) break the orthogonality condition.

□

Remark.

How does one write the matrix equation $A^T = A^{-1}$ with our index conventions? Identifying the row indices on both sides, which gets switched to the right by the transpose we might write $A^j{}_i = A^{-1i}{}_j$ but that breaks our convention. The only way we can respect index position (an

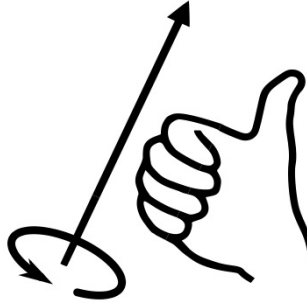


Figure 1.22: The right hand rule correlates a choice of normal direction with the sense of the rotation in the orthogonal plane in a simple way, thus characterizing a rotation in Euclidean 3-space by its unique axis of fixed points (with direction \hat{n}) and the angle θ of rotation about that axis. The corresponding matrix $R(\hat{n}, \theta)$ depends only on these independent parameters (two for the direction, one for the angle).

index must be at the same level on each side of the equation) is by introducing the identity matrix with both indices down to bring the upper indices down to the same level

$$(\underline{I}A)^T = \underline{I} \underline{A}^{-1} \leftrightarrow \delta_{jk} A^k_i = \delta_{ik} A^{-1k}_j.$$

▲

Orthogonal matrices represent familiar physical operations in ordinary Euclidean 3-space: rotations and reflections. The group of real orthogonal 3×3 matrices is called $O(3, R)$. Note that the product determinant formula $\det \underline{A} \underline{B} = \det \underline{A} \det \underline{B}$ applied to $\underline{A} \underline{A}^T = \underline{I}$ together with the identity $\det \underline{A}^T = \det \underline{A}$ shows that $\det \underline{A}^2 = 1$, so the determinant of an orthogonal matrix can only have the values ± 1 . Those with positive unit determinant form a subgroup called the special orthogonal group $SO(3, R)$ and represent rotations of space, while the remaining negative determinant orthogonal matrices also involve either space reflections or odd permutations of the axes. The latter do not form a subgroup since their products have unit positive determinants by the same product formula.

While rotations occur in a family of parallel 2-planes in general in any dimension, three dimensions are special in that there is a unique orthogonal direction to that family which we can associate with an axis of rotation, namely the line of fixed points perpendicular to the family of rotation planes. This gives us a nice physical representation of any rotation, which can be specified by the angle of rotation about this axis, together with a unit vector giving the direction of the axis. The right hand rule illustrated in Fig. 1.22 allows us to correlate the direction of the rotation within the planes with the direction of the axis in a simple way: with the thumb pointing along the chosen axis direction, the fingers curl in the direction of the rotation in the orthogonal planes in the direction of the fingertips.

Example 1.6.2. the usual dot product on \mathbb{R}^3

Sometimes it is helpful to be more concrete. If on \mathbb{R}^3 we let $G_{ij} = \delta_{ij}$ be the entries of the identity matrix

$$\underline{G} = (\delta_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

then $G = \delta_{ij}\omega^i \otimes \omega^j$ in standard Cartesian coordinates labeled (x^1, x^2, x^3) defines the usual dot product inner product which determines the flat geometry of Euclidean space

$$\begin{aligned} a \cdot b &= G(\langle a^1, a^2, a^3 \rangle, \langle b^1, b^2, b^3 \rangle) = G_{ij}a^ib^j \\ &= \langle a^1, a^2, a^3 \rangle \cdot \langle b^1, b^2, b^3 \rangle \\ &= a^1b^1 + a^2b^2 + a^3b^3 = \underline{a}^T \underline{b}. \end{aligned}$$

Rotations about the origin (linear transformations of the vector space) leave the dot product invariant. □

Example 1.6.3. new dot product

We can introduce an example of a more general inner product “ \bullet ” on \mathbb{R}^n starting from any symmetric matrix $\underline{M} = (M_{ij}) = (M_{ji}) = \underline{M}^T$ and extending the inner products by multilinearity from the standard basis vectors $\{e_i\}$ to any other vectors, namely defining $M(u, v) = M(u^ie_i, v^je_j) = u^iv^jM(e_i, e_j) = M_{ij}u^iv^j$, with $M_{ij} = M(e_i, e_j) \equiv e_i \bullet e_j$. Transforming to another basis, this matrix of components will change. However, we don’t need to consider different inner products on \mathbb{R}^n in order to have component matrices which are not the identity matrix. As soon as we choose a general basis of this space, its matrix of inner products with the usual dot product can be any symmetric matrix with nonzero determinant.

Using Cartesian coordinates x^1, x^2 on \mathbb{R}^2 , we can introduce the following new inner product of two vectors u, v :

$$\begin{aligned} \underline{M} &= \begin{pmatrix} 8 & -2 \\ -2 & 5 \end{pmatrix}, \\ u \bullet v &= M(u, v) = \underline{u}^T \underline{M} \underline{v} = \begin{pmatrix} u^1 & u^2 \end{pmatrix} \begin{pmatrix} 8 & -2 \\ -2 & 5 \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} \\ &= M_{ij}u^iv^j = 8u^1v^1 - 2(u^1v^2 + u^2v^1) + 5u^2v^2. \end{aligned}$$

The “unit circle” for this inner product would be described by the curve in the plane: $1 = x \bullet x = M(x, x) = 8(x^1)^2 - 4x^1x^2 + 5(x^2)^2$. We will see below that this is a rotated ellipse as viewed in the original Euclidean geometry of the plane. □

Example 1.6.4. the Minkowski inner product on \mathbb{R}^4

On \mathbb{R}^4 let $-\eta_{00} = \eta_{11} = \eta_{22} = \eta_{33} = 1$ and $\eta_{ij} = 0$ ($i \neq j$), so that the component matrix is

$$\underline{\eta} = (\eta_{ij}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

If we let $G_{ij} = \eta_{ij}$ then $G = \eta_{ij}\omega^i \otimes \omega^j$ in standard Cartesian coordinates labeled (x^0, x^1, x^2, x^3) defines an inner product, called the Minkowski metric. We can simply reinterpret the dot product in this context to be

$$\begin{aligned} a \cdot b &= G(\langle a^0, a^1, a^2, a^3 \rangle, \langle b^0, b^1, b^2, b^3 \rangle) = G_{ij}a^ib^j \\ &= \langle a^0, a^1, a^2, a^3 \rangle \cdot \langle b^0, b^1, b^2, b^3 \rangle \\ &= -a^0b^0 + a^1b^1 + a^2b^2 + a^3b^3 = \underline{a}^T \underline{\eta} \underline{b}. \end{aligned}$$

This inner product determines the flat geometry of spacetime in special relativity, referred to as Lorentzian as opposed to Euclidean because of the single minus sign which distinguishes time directions from spatial directions. This is associated with the clearly different nature of time compared to spatial dimensions. To emphasize this we put the time coordinate first before the space coordinates and distinguish it by using the subscript 0: $(x^0, x^1, x^2, x^3) = (t, x, y, z)$, letting $i, j, \dots = 0, 1, 2, 3$. Then

$$\eta_{ij}x^ix^j = -t^2 + x^2 + y^2 + z^2 = \begin{cases} \equiv \ell^2 > 0 & \text{spacelike} \\ = 0 & \text{lightlike or null} \\ \equiv -\tau^2 < 0 & \text{timelike} \end{cases}$$

is interpreted as the signed squared spacetime distance from the origin or the signed squared length of the spacetime position vector $\langle t, x, y, z \rangle$, often called a 4-vector since we are also interested in the usual 3 component vectors in space alone when discussing relativity. However, in contrast with the Euclidean case of the familiar dot product, here this self-inner product of a vector can be both positive and negative as well as zero even when this vector itself is nonzero. The square root of the absolute value of the self-inner product is still interpreted as the length of the vector, but the sign of the self-inner product gives additional information about the vector.

The hypersurface

$$x^2 + y^2 + z^2 = t^2 \quad \text{or} \quad s^2 = -t^2 + x^2 + y^2 + z^2 = 0$$

is actually a cone with vertex at the origin in this 4-dimensional space, called the light cone, consisting of points whose signed squared distance from the origin is 0. The interpretation of this vanishing value is that in units where the speed of light $c = 1$ is unity, then after a time interval $t > 0$ a light ray starting at the spatial origin at time 0 travels a distance t to reach a spatial point (x, y, z) which is at the distance $(x^2 + y^2 + z^2)^{1/2} = t$. In other words this metric captures the behavior of light rays by assigning zero spacetime interval to all spacetime points

which are connected to the origin by a light ray. When the spatial distance $(x^2 + y^2 + z^2)^{1/2} =$ of a spatial point is greater than t , then there has not been enough time for any light ray to reach the point, or if it is less than t , any light ray emitted at time 0 has already passed by that point. Thus the sign of the signed squared distance is associated with the causality of events in the spacetime. The point (t, x, y, z) is said to be spacelike separated from the origin $(0, 0, 0, 0)$ in the positive case $\eta_{ij}x^ix^j > 0$, lightlike separated when it is zero $\eta_{ij}x^ix^j = 0$ and timelike separated when it is negative $\eta_{ij}x^ix^j < 0$.

In some contexts the opposite overall sign is used for the Lorentz inner product: $s^2 = t^2 - x^2 - y^2 - z^2$ so that the ordinary spatial dimensions are associated with the minus signs instead of the time dimension. Since length is defined by the square root of the absolute value of the self inner product, the interpretation of the signs is a matter of choice. Extending the usual inner product on \mathbb{R}^3 by one extra dimension with a minus sign seems the most reasonable, but there are mathematical reasons for adopting the other sign convention in some quantum mechanical contexts.

\mathbb{R}^4 with this inner product is called Minkowski spacetime, sometimes designated by \mathbb{M}^4 . One can consider various dimensions for Minkowski spacetime just like Euclidean space. \mathbb{M}^n is just \mathbb{R}^n with the sign reversed on the self-inner product of the first standard basis vector. The 2-dimensional Minkowski plane \mathbb{M}^2 is useful for studying 1-dimensional motion along a single spatial dimension, while 3-dimensional Minkowski spacetime \mathbb{M}^3 is useful for studying motion in a spatial plane, like a planet orbiting a central sun, or “classical” electrons orbiting a nucleus. Minkowski and Lorentz were the two most important pioneers in advancing the mathematics underlying special relativity but it was the genius of Einstein who understood the central concept of spacetime itself. Appendix A explores Minkowski spacetime in a bit more depth. □

Exercise 1.6.1.

Euclidean inner product on $h(2)$

In Exercise 1.2.2 we introduced a set of four 2×2 matrices which are the basis of a 4-dimensional real subspace of the space of complex 2×2 matrices

$$\underline{E}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \underline{E}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \underline{E}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \underline{E}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The elements of this real vector space $h(2)$ are of the form

$$\underline{X} = x^i \underline{E}_i = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix},$$

where the coefficients $(x^i) = (x^0, x^1, x^2, x^3)$ are real.

If $\underline{X} = x^i \underline{E}_i$ and $\underline{Y} = y^i \underline{E}_i$, show that the following inner product is just the Euclidean inner product on this vector space in this basis

$$G(\underline{X}, \underline{Y}) \equiv \frac{1}{2} \text{Tr } \underline{X} \underline{Y} = x^0 y^0 + x^1 y^1 + x^2 y^2 + x^3 y^3.$$

■

Exercise 1.6.2.**two inner products on $gl(2, \mathbb{R})$**

In Exercise 1.2.1 we introduced a new basis of the set $gl(2, \mathbb{R})$ of all real 2×2 matrices adapted to its tracefree subspace $sl(2, \mathbb{R})$ as well as to its decomposition into a 3-dimensional subspace of symmetric matrices and a 1-dimensional space of antisymmetric matrices

$$\underline{E}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \underline{E}_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \underline{E}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \underline{E}_3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

so that any 2×2 matrix has the representation

$$\underline{X} = x^i \underline{E}_i = \begin{pmatrix} x^0 + x^1 & x^1 - x^2 \\ x^1 + x^2 & x^0 - x^1 \end{pmatrix}.$$

a) If $\underline{X} = x^i \underline{E}_i$ and $\underline{Y} = y^i \underline{E}_i$, show that the following inner product is just the Lorentzian inner product on this vector space in this basis

$$G(\underline{X}, \underline{Y}) \equiv \frac{1}{2} \text{Tr } \underline{X} \underline{Y} = x^0 y^0 + x^1 y^1 + x^2 y^2 - x^3 y^3.$$

The negative sign is associated with the antisymmetric subspace.

b) Show that this basis is orthogonal with respect to the Euclidean inner product

$$\text{Tr } \underline{X}^T \underline{Y} = 2(x^0 y^0 + x^1 y^1 + x^2 y^2 + x^3 y^3).$$

This inner product is the usual Euclidean one on $(\mathbb{R})^4$ interpreting $gl(2, \mathbb{R})$ as \mathbb{R}^4 by listing its entries row by row

$$\text{Tr} \begin{pmatrix} a^1 & a^2 \\ a^3 & a^4 \end{pmatrix}^T \begin{pmatrix} b^1 & b^2 \\ b^3 & b^4 \end{pmatrix} = a^1 b^1 + a^2 b^2 + a^3 b^3 + a^4 b^4.$$

■

Given a general inner product or metric G on a vector space V , invariance of this metric under a change of basis requires

$$\underline{G} = \underline{A}^{-1T} \underline{G} \underline{A}^{-1},$$

or right multiplying by \underline{A} and using properties of the transpose and inverse together with $\underline{G}^T = \underline{G}$ one transforms this into

$$\underline{G} \underline{A} = \underline{A}^{-1T} \underline{G}^T = (\underline{G} \underline{A}^{-1})^T$$

from which it follows that

$$(\underline{G} \underline{A})^T = \underline{G} \underline{A}^{-1}, \quad (\text{generalized orthogonality condition})$$

which generalizes the orthogonality condition $\underline{A}^T = \underline{A}^{-1}$ which holds when $\underline{G} = \underline{I}$, and has the index form

$$G_{kj}A^j_i = G_{ij}A^{-1j}_k.$$

In components, using the index lowering convention $A_{ij} = G_{ik}A^k_j$ to be discussed shortly, this just says that $A_{ji} = [A^{-1}]_{ij}$, or that the transpose of the index-lowered component matrix is the index-lowered form of the inverse matrix.

If the starting basis is an orthonormal basis, then this condition describes a change from one orthonormal basis to another and the matrices which accomplish this are called generalized orthogonal matrices. The corresponding generalized orthogonal matrix groups are classified by the signature of the inner product. For an inner product with P positive signs and M minus signs, this group is designated by $O(P, M)$, with $SO(P, M)$ for its special orthogonal group of unit determinant matrices. Over the complex numbers, one can always find bases which are orthonormal in the usual sense (if $G(u, u) = -1$, then $G(iu, iu) = 1$), so one does not need to include \mathbb{R} explicitly in the symbol as in $GL(n, \mathbb{R})$. For \mathbb{M}^4 , this group $O(3, 1)$ is sometimes called the pseudo-orthogonal group or Lorentz group.

Repeating the same argument made for the ordinary orthogonal matrices shows that generalized orthogonal matrices must have a determinant of value ± 1 . Using the fact that the determinant of a matrix product is the product of the determinants, and that the transpose does not change the determinant, by taking the determinant of the transformation invariance relation ($\det \underline{G} \neq 0$ by the nondegeneracy condition)

$$\underline{G} = \underline{A}^{-1T} \underline{G} \underline{A}^{-1} \rightarrow \det \underline{G} = \det(\underline{A}^{-1})^2 \det \underline{G} = \det(\underline{A})^{-2} \det \underline{G},$$

it follows that $\det \underline{A} = \pm 1$, so all of the generalized orthogonal groups differ from their special orthogonal subgroups only by reflections under which one or more of the coordinates change sign, or by odd permutations of the coordinates. Thus $O(3, \mathbb{R})$ is enlarged from the group of rotations of ordinary space $SO(3, \mathbb{R})$ by the discrete subgroup of reflections and odd permutations.

Exercise 1.6.3.

pseudo-orthogonality in the Lorentz plane

Many problems of special relativity only require one space and one time dimension, where the Minkowski metric with components $G_{11} = -\eta_{00} = 1$ and $G_{01} = 0 = G_{10}$ on \mathbb{R}^2 with standard coordinates x^0, x^1 is relevant. This metric leads to the hyperbolic geometry discussed in Appendix A. In 2-dimensional “spacetime diagrams” involving this mathematical description, the time axis is usually plotted vertically and the spatial axis horizontally.

a) Show by direct computation that the matrices

$$\underline{G} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \underline{A} = \begin{pmatrix} \cosh \alpha & \sinh \alpha \\ \sinh \alpha & \cosh \alpha \end{pmatrix}, \quad \underline{A}^{-1} = \begin{pmatrix} \cosh \alpha & -\sinh \alpha \\ -\sinh \alpha & \cosh \alpha \end{pmatrix}$$

satisfy the pseudo-orthogonality condition for a change of basis.

b) Show that the vectors $\langle 1, 1 \rangle$ and $\langle 1, -1 \rangle$ are both vectors with zero length. How do these vectors change under matrix multiplication by the matrix \underline{A} ?



Exercise 1.6.4.

Euclidean and Lorentzian dot products

a) In \mathbb{R}^n (with usual dot product), what is the value of $\text{sgn } v$ for any $v \neq 0$? (This property makes \mathbb{R}^n Euclidean.)

b) In \mathbb{R}^n (with usual dot product), if $\|v\| = 0$, then what must v be?

c) In \mathbb{R}^4 with the Minkowski inner product, what is the sign of the vectors $\langle 0, 1, -1, 1 \rangle$, $\langle 2, 1, 0, 0 \rangle$, $\langle 1, 0, 0, 1 \rangle$? What are their magnitudes? Which of these vectors can be “normalized” to unit vectors by dividing by their lengths?

d) In \mathbb{R}^4 with the Minkowski inner product, if $\|v\| = 0$, then what must v satisfy?



Index shifting with an inner product

For fixed X the usual dot product $X \cdot Y = \delta_{ij} X^i Y^j = f_X(Y)$ is a linear function of the vector Y in \mathbb{R}^n with coefficients $(f_X)_j = \delta_{ij} X^i$ numerically equal to the components X^j of the vector X in the standard basis, so it defines a linear function associated with X which we could denote suggestively by “ $X \cdot$ ”. This covector needs a better name. Since effectively the upper vector index is changed to a lower covector index through this dot product relationship, the index of X is “lowered” by this process. In music when you lower a note like B slightly, it is referred to as a flat: B^b , so we can denote the index lowered covector by X^b . Similarly raising a note slightly is called a sharp: B^\sharp , so the inverse process of raising an index from a lower covector index to an upper vector index can be indicated by a sharp: $(X^b)^\sharp = X$. Since by definition $X^b(Y) = X \cdot Y$, if this is zero, Y must be orthogonal to X , so the level surface $X^b(Y) = 0$ consists of the (hyper)plane through the origin perpendicular to the vector X . On the other hand, the parallel (hyper)plane through the tip of X corresponds to the value $X^b(X) = X \cdot X = |X|^2$, so one must divide X by $|X|^2$ to locate at its tip the parallel (hyper)plane corresponding to the value 1 to represent the covector geometrically. Fig. 1.13 illustrates this in the Euclidean plane.

Since the vector and covector have the same components in the standard basis, they have the same length, and their orientations are locked together by orthogonality, so it makes sense to think of the pair as just two different realizations of the same physical vector, which is why we retain the same kernel letter X to represent them, and only distinguish their components by the index position, which can be raised or lowered using this dot product relationship which is symbolized by the Kronecker delta of dot products of the standard basis or its dual basis: $X_i = \delta_{ij} X^j$, $X^i = \delta^{ij} X_j$. In any other basis, one must use the corresponding matrices of dot products to do this index lowering and raising, or “shifting.”

It is natural to extend this discussion to any vector space V with a nondegenerate inner product G , with inverse G^{-1} , thus satisfying

$$\begin{aligned}\underline{G}^{-1}\underline{G} &= \underline{I}, & G^{ik}G_{kj} &= \delta^i_j = G^{ik}G_{jk}, \\ \underline{G}\underline{G}^{-1} &= \underline{I}, & G_{jk}G^{ki} &= \delta^i_j = G_{jk}G^{ik}.\end{aligned}$$

Then we can introduce a streamlined notation for the related pair of maps between V and V^* , calling them \flat from V to V^* and \sharp from V^* to V

$$\begin{aligned}v^\flat(u) &\equiv G(u, v) = G_{ij}u^i v^j = (G_{ij}v^j)u^i &\rightarrow [v^\flat]_i &= G_{ij}v^j, \\ f(g^\sharp) &\equiv G^{-1}(f, g) = G^{ij}f_i g_j = f_i(G^{ij}g_j) &\rightarrow [g^\sharp]^i &= G^{ij}g_j,\end{aligned}$$

where the flat symbol \flat stands for “down”, lowering the index, and the sharp symbol \sharp stands for “up”, raising the index.

In this way using the metric and its inverse we associate a covector v^\flat with each vector v and a vector g^\sharp with each covector g . These two maps are inverses of each other

$$\begin{aligned}[(v^\flat)^\sharp]^i &= G^{ij}(v^\flat)_j = G^{ij}G_{jk}v^k = \delta^i_k v^k = v^i, \\ [(g^\sharp)^\flat]_i &= G_{ij}(g^\sharp)^j = G_{ij}G^{jk}g_k = \delta^k_i g_k = g_i.\end{aligned}$$

The inner product provides an “identification map” between a vector space and its dual. This turns out to be so useful that more shorthand notation is introduced

$$\begin{aligned}v_i &\equiv v^\flat(e_i) = G_{ij}v^j \quad (\text{“lowering the index”}), \\ g^i &\equiv \omega^i(g^\sharp) = G^{ij}g_j \quad (\text{“raising the index”}).\end{aligned}$$

In component notation we use the same letter for the corresponding covector or vector (called the kernel symbol, kernel in the sense that we add sub/superscripts to it) and just put the index in the right location, while the sharp or flat helps distinguish the two objects in index free form. One then refers to the “contravariant” ($u \sim u^i$) or “covariant” form of a vector ($u^\flat \sim u_i$), to distinguish the two, for example.

Furthermore, the inner product of a pair of vectors has the same value as the inner product of the pair of corresponding covectors

$$G^{-1}(u^\flat, v^\flat) = G^{mn}u_m v_n = G^{mn}G_{mi}u^i G_{nj}v^j = \delta^n_i u^i G_{nj}v^j = G_{ij}u^i v^j = G(u, v).$$

Thus a vector and its corresponding covector have the same self-inner product and the same length.

Example 1.6.5. linearity becomes geometry

Consider \mathbb{R}^n with the standard basis $\{e_i\}$ and the standard (dot) inner product $G_{ij} = \delta_{ij}$, $G^{ij} = \delta^{ij}$. The index shifting maps identify vectors and covectors with the same standard components

$$\begin{aligned}(V^\flat)_i &\equiv v_i = \delta_{ij}v^j \quad (\text{i.e., } v_i = v^i \text{ for each } i), \\ (f^\sharp)^i &\equiv f^i = \delta^{ij}f_j \quad (\text{i.e., } f^i = f_i \text{ for each } i).\end{aligned}$$

Thus evaluation of a covector on a vector

$$f(v) = f_i v^i = \delta_{ij} f^j v^i = f^\sharp \cdot v$$

is represented as the standard dot product of the vector with another vector whose components are the same as the covector.

In this way *linearity* is converted into *geometry* and one can ignore the distinction between the vector space and its dual and thus only use subscript indices. However, there is a catch. For everything to work, one has to use only orthonormal bases—otherwise things fall apart. (If the basis is not orthonormal, one no longer has the same components for a vector and its corresponding covector.) This turns out to be no problem for elementary linear algebra with its limited goals, but it is a problem if you want to go beyond that. \square

Example 1.6.6. \mathbb{R}^n and \mathbb{M}^n

Suppose we introduce one minus sign for the self-inner product of the first basis vector in the previous problem on n -dimensional Euclidean space \mathbb{R}^n to get Minkowski spacetime \mathbb{M}^n , with metric matrix $\underline{\eta}$ and standard coordinates $(x^0, x^1, \dots, x^{n-1})$ and now interpret the dot as the new inner product. The standard basis is still orthonormal with respect to this new “Lorentzian” inner product, namely an inner product that only has one direction that has negative self-inner products (almost, clarification later). These vector spaces are the same but we use different coordinate labels adapted to the two different standard inner products.

The dual basis is automatically orthonormal and has the same matrix of inner products as the basis vectors but we need to raise the indices as we did before with the Kronecker delta: $\omega^i \cdot \omega^j = \eta^{ij}$. If we raise an index on the covariant eta, or lower an index on the contravariant eta using the eta matrix itself, in both cases we get the mixed Kronecker delta which is the matrix of the identity tensor, or if we raise both indices on the covariant eta we get the contravariant eta, etc.

$$\eta_{ik} \eta^{kj} = \delta^j_i, \quad \eta_{mn} \eta^{mi} \eta^{nj} = \eta^{ij}.$$

Thus in the Lorentzian geometry we can think of η_{ij} , η^{ij} and δ^i_j as the components of the three possible forms of the same physical tensor, the identity tensor, as long as we are working in an orthonormal basis (where we agree to put the negative-signed basis vector first). For a general inner product we can similarly think of G_{ij} , G^{ij} and δ^i_j as the components of the various index-shifted forms of the identity tensor. \square

Index shifting conventions

In a situation where an inner product G is available and relevant to the kind of problem being described mathematically, we can extend the “index shifting” maps to any type of tensor. A $\binom{p}{q}$ -tensor is said to have “rank” $(p + q)$ and have p contravariant indices (i.e., p covector arguments) and q covariant indices (i.e., q vector arguments). For all tensors of a given total

rank, we can establish a correspondence between tensors with different “index” positions. For example, if $p + q = 2$, then we are dealing with $\binom{0}{2}$, $\binom{1}{1}$, or $\binom{2}{0}$ -tensors.

Suppose $T = T^i_j e_i \otimes \omega^j$. Then we can introduce three other tensors by

$$T^{ij} \equiv G^{ij} T^i_k, \quad T_{ij} \equiv G_{ij} T^k_j, \quad T_i^j \equiv G_{ik} T^k_\ell G^{\ell j}.$$

These are related to each other in turn by

$$T^{ij} = G^{im} G^{jn} T_{mn}, \quad T_{ij} = G_{im} G_{jn} T^{mn}, \quad \text{etc.}$$

For a given starting tensor T , we can interpret all four such related tensors as different “representations” of the same physical object, but with different index arguments. Of course this is a convenient fiction since a vector v and covector v^b have completely different geometric interpretations, but those interpretations are related to each other in an interesting way. We use the same kernel letter and let the index position distinguish between the different tensors of this family of related tensors. The last of these four tensors has the representation $T_i^j e_j \otimes \omega^i$ if we agree always to list the covector inputs of a tensor first and the vector inputs second, effectively identifying $V \otimes V^*$ and $V^* \otimes V$, but we have to suspend our convention to list contravariant indices first and covariant indices second in order to distinguish between different index positions of different arguments of the tensor. This is not a problem since index shifting turns out to be extremely useful.

For rank 3 tensors there are $2^3 = 8$ different index positions

$$\underbrace{T_{ijk}}_{\binom{0}{3}}; \underbrace{T^i_{jk}, T_i^j{}_k, T_{ij}^k}_{\binom{1}{2}}; \underbrace{T_i{}^{jk}, T^i{}_j{}^k, T^{ij}{}_k}_{\binom{2}{1}}; \underbrace{T^{ijk}}_{\binom{3}{0}},$$

while for rank 4 tensors there are $2^4 = 16$ different index positions. However, when tensors have symmetries, this number is then reduced. For example, for symmetric second rank tensors, the symmetry condition $T_{ij} = T_{ji}$ implies $T^i_j = T_j^i$.

Given any $\binom{p}{q}$ -tensor there are two special members of the family of tensors related to it by index shifting, namely the “totally covariant” form of the tensor (all indices down) and the “total contravariant” form of the tensor (all indices up)

$$T \sim T^{i\dots}_{j\dots} \rightarrow \begin{cases} T^{i\dots j\dots} \sim T^\sharp \\ T_{i\dots j\dots} \sim T^\flat \end{cases},$$

where we slide the lower covariant indices over to the right of the upper contravariant indices before raising them or lowering the upper indices. This extends the \sharp and \flat maps to arbitrary tensors, meaning respectively “raise all indices” and “lower all indices.”

For the usual dot product on \mathbb{R}^n , using the standard basis, all of these tensors have the same numerical values for corresponding components, so one can always use the totally covariant form of a tensor accepting only vector arguments to discuss elementary linear algebra. For a general inner product we can introduce the magnitude and sign of a tensor just like that of a vector in terms of the totally covariant or contravariant form. Define $\|T\| \geq 0$ and $\text{sgn } T \in \{+, 0, -\}$ by

$$\begin{aligned} (\text{sgn } T) \|T\|^2 &= G^{im} G^{jn} \dots T_{ij\dots} T_{mn\dots} = G_{im} G_{jn} \dots T^{ij\dots} T^{mn\dots} \\ &= T_{ij\dots} T^{ij\dots} \quad (\text{with } \|T\| \equiv \text{sgn } T \equiv 0 \text{ if this vanishes}). \end{aligned}$$

Since the space of $\binom{p}{q}$ -tensors for fixed p and q is itself a vector space, it can have an inner product. This defines such an inner product induced by the inner product on the underlying vector space, namely

$$T \cdot S = G_{im} G_{jn} \cdots T^{ij\cdots} S^{mn\cdots} = T^{ij\cdots} S_{ij\cdots}.$$

For tensors over \mathbb{R}^n with the usual dot product, this inner product for tensors is very simple to describe. The sign is always positive, except for the zero tensor of a given valence (specific values of p and q), and the magnitude is always positive (except for the zero tensor) and equal to the square root of the sum of the squares of all its components (just like for vectors!). For example

$$\begin{aligned} \|T\|^2 &= \delta^{im} \delta^{jn} T_{ij} T_{mn} = T_{ij} T^{ij} = \sum_{i=1}^n \sum_{j=1}^n T^{ij} T^{ij} = \sum_{i=1}^n \sum_{j=1}^n (T^{ij})^2 \\ &= \text{Tr } \underline{T}^T \underline{T}, \end{aligned}$$

since $T_{ij} = T^{ij}$ for each pair of index values (i, j) . The last line shows that this is equivalent to the transpose trace inner product of square matrices when expressed in terms of the corresponding matrix of components, already touched upon in Exercise ??, which is just the usual dot product on the space of $n \times n$ matrices when thought of as \mathbb{R}^{n^2} by listing entries row by row.

Note that the inverse tensor $G^{-1} = G^{ij} e_i \otimes e_j$ defines an inner product on the dual space V^* thought of as a vector space in its own right, and this definition of the magnitude and sign of a covector is exactly the definition we introduced above for a vector space V , except now applied to the dual space.

Exercise 1.6.5.

trace inner products of antisymmetric 3×3 matrices

On Euclidean \mathbb{R}^3 with the usual dot product, consider the mixed tensor whose matrix of components is

$$\underline{\omega} = (\omega^i_j) = \begin{pmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{pmatrix}.$$

Evaluate

$$\text{Tr } \underline{\omega}^2 = \omega^i_j \omega^j_i = -\omega_{ij} \omega^{ij}$$

in terms of the corresponding vector $\langle \omega_1, \omega_2, \omega_3 \rangle$. Note that this is a negative-definite inner product on this subspace of $gl(3, \mathbb{R})$.

By multiplying by the factor $-1/2$, we are back to the usual dot product of the corresponding vector in \mathbb{R}^3 . By inserting a transpose, we remove the minus sign and get the self-dot product as a 2 index tensor

$$\text{Tr } \underline{\omega}^T \underline{\omega} = \omega_{ij} \omega^{ij},$$

and the remaining factor of two corresponds to the overcounting by the two permutations of the indices that contribute for each unordered distinct index pair (i, j) .

**Exercise 1.6.6.****electromagnetic field matrices**

On 4-dimensional Minkowski spacetime in coordinates $(x^\alpha) = (x^0, x^1, x^2, x^3) = (x^0, x^a)$ with inner product $\eta = \text{diag}(-1, 1, 1, 1)$, index-shifting is easy. Changing the level of the 0 index changes the sign of the component, but the remaining components do not change under index shifting. Consider the matrix

$$\underline{F} = (F^\alpha{}_\beta) = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix} \equiv [[E, B]]$$

which defines a mixed second rank tensor $F = F^\alpha{}_\beta \omega^\alpha \otimes e_\alpha$ and the dual matrix that we will explain in Chapter 4

$$*\underline{F} = (*F^\alpha{}_\beta) = \begin{pmatrix} 0 & -B_1 & -B_2 & -B_3 \\ -B_1 & 0 & E_3 & -E_2 \\ -B_2 & -E_3 & 0 & E_1 \\ -B_3 & E_2 & -E_1 & 0 \end{pmatrix} = [[-B, E]].$$

This combines the electric and magnetic vector fields into a single unified electromagnetic tensor field on spacetime. [The double square bracket notation just allows us to have a way to refer to a 4×4 matrix formed out of two 3-vectors in this way.]

a) Show that the component matrices

$$(F_{ij}) = (\eta_{ik} F^k{}_j) = \underline{\eta} \underline{F}, \quad (*F_{ij}) = (\eta_{ik} *F^k{}_j) = \underline{\eta} * \underline{F}$$

of the tensors F^\flat and F^\sharp are antisymmetric matrices, which is the condition discussed in Chapter 2 (see Exercise 2.3.8) that the original matrices are tangents to curves of matrices which are orthogonal with respect to the Lorentzian inner product. This is not an accident. The mixed electromagnetic field tensor generates a pseudorotation in spacetime, called a Lorentz transformation.

b) Use a computer algebra system to evaluate the scalars

$$\begin{aligned} \text{Tr } \underline{F} &= F^\alpha{}_\alpha, & \text{Tr } *\underline{F} &= *F^\alpha{}_\alpha, \\ \text{Tr } \underline{F}^2 &= F^\alpha{}_\beta F^\beta{}_\alpha = -F_{\alpha\beta} F^{\alpha\beta}, \\ \text{Tr } *\underline{F} *\underline{F} &= *F^\alpha{}_\beta *F^\beta{}_\alpha = -*F_{\alpha\beta} *F^{\alpha\beta}, \\ \text{Tr } \underline{F} *\underline{F} &= F^\alpha{}_\beta *F^\beta{}_\alpha = -F_{\alpha\beta} *F^{\alpha\beta} = \text{Tr } *\underline{F} \underline{F}. \end{aligned}$$

c) Use a computer algebra system to evaluate the so called energy-momentum tensor associated with the electromagnetic field tensor

$$4\pi(T^\alpha{}_\beta) = \left(-F^\alpha{}_\gamma F^\gamma{}_\beta - \frac{1}{4} \delta^\alpha{}_\beta F_{\gamma\delta} F^{\gamma\delta} \right) = -\underline{F}^2 + \frac{1}{4} \underline{I} \text{Tr } \underline{F}^2,$$

and its trace

$$4\pi \operatorname{Tr} \underline{T} = 4\pi T^\alpha{}_\alpha.$$

Then evaluate the matrix of totally contravariant components of this tensor ($T^{\alpha\beta}$). Recognize the cross product in the components $T^{0a} = T^{a0}$ and the magnitudes of the electric and magnetic fields in T^{00} .

d) Use a computer algebra system or hand calculation to evaluate the matrix product $\gamma^{-1}q\underline{F}\underline{u}$ if

$$u = \gamma\langle 1, v^1, v^2, v^3 \rangle = \gamma\langle 1, \vec{v} \rangle.$$

Show that the result can be written in terms of 3-vectors as

$$\langle q\vec{E} \cdot \vec{v}, q(\vec{E} + \vec{v} \times \vec{B}) \rangle,$$

which is the right hand side of the Lorentz force law

$$m \frac{du}{dt} = \gamma^{-1} m \frac{du}{d\tau} = \gamma^{-1} q \underline{F} \underline{u}.$$

■

Partial evaluation of a tensor and index shifting

If we evaluate the inner product G only on its second argument “ $G(\ , v)$ ”, then it still needs a vector in its first argument to produce a real number. This is a linear function of that argument, i.e., defines a covector, which is exactly v^\flat . We can write suggestively $v^\flat = G(\ , v)$, for the partial evaluation of G on one argument. Similarly $f^\sharp = G^{-1}(\ , f)$.

We can partially evaluate any tensor on any number of arguments. For example, if

$$T = T_{ijk} \omega^i \otimes \omega^j \otimes \omega^k = “T(\ , \ , \)”$$

then

$$T(\ , v, \) \equiv T_{ijk} \omega^i \otimes \omega^k \omega^j(v) = T_{ijk} v^j \omega^i \otimes \omega^k$$

makes sense as a way to represent partial evaluation on a single argument. Iteration of this extends it to any number of arguments.

Contraction of tensors

For a $\binom{p}{q}$ -tensor with at least one index each type ($p \geq 1, q \geq 1$), one can select one upper index and one lower index and sum over them, reducing the number of free indices by 2 leading to a $\binom{p-1}{q-1}$ -tensor. This is called contraction of the tensor on that pair of indices of opposite valence (one up, one down!). For example, with a $\binom{1}{2}$ -tensor $T = T^i{}_{jk} e_i \otimes \omega^j \otimes \omega^k$, we get two covectors

$$T^k{}_{ki} \omega^i, \quad T^k{}_{ik} \omega^i$$

from the two possible contractions of the single contravariant index with the two covariant indices.

The previous partial evaluation $T_{ijk}v^j$ is then a special case of two consecutive operations, easiest to depict in terms of index (component) language. First the tensor product tensor $T_{ijk}v^\ell$ is formed, and then it is contracted on the index pair (j, ℓ) to yield $T_{ijk}v^j$. We also say that we are “contracting the index j of the tensor T_{ijk} with the vector v^j in index language.

This can be generalized to any subset of corresponding indices on a pair of tensors, representing the tensor product of the two tensor factors followed by contractions on all index pairs associated with this subset. For example, the component relations

$$C^i_{jk}D^{jk}_m, C^i_{jk}C^k_{mi}, R^i_{jmn}A^{mn}, R^{ij}_{mn}R^{mn}_{pq}, C^i_{jmn}\eta^{jmnk}$$

are all examples of contractions of a pair of tensors on two or three indices, leading to tensors of the type indicated by the remaining free indices. This in turn may be extended to any number of tensor factors.

Most tensors arise with natural index positions, so only certain contractions are possible, but if we have an inner product tensor with components G_{ij} and inverse G^{ij} , we can use it to shift index positions and thus contract any pair of indices on any tensors or simultaneously contract as many pairs as we wish. For example, if we have a 3 index object like C^i_{jk} we can do natural contractions with the single upper index and the two lower indices, or a metric contraction on the last two indices, which can be written in two equivalent ways

$$C^i_{ik}, C^i_{ji}, C^i_{jk}G^{jk} = C^{ik}_k = C^i_j{}^j.$$

Geometric interpretation of index shifting

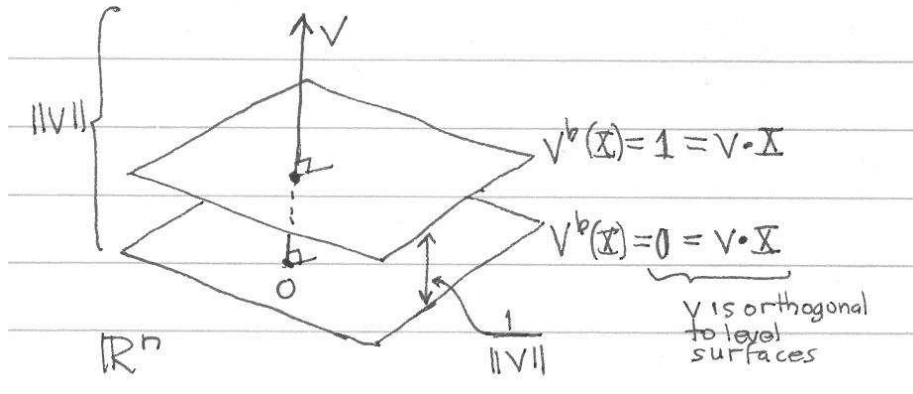


Figure 1.23: A vector v is orthogonal to the level surfaces of its corresponding covector v^b .

The relation $v^b(X) = v \cdot X = 0$ shows that the vector v is orthogonal to the level surfaces of the covector v^b . The relation

$$v^b(v) = v \cdot v = ||v||^2$$

may be interpreted as stating that the vector v pierces $\|v\|^2$ “layers” (integer valued level surfaces) of the covector v^b , and hence the unit vector $\hat{v} = v/\|v\|$ pierces $\|v\|$ layers, so each layer must have a separation of $1/\|v\|$. Thus while v has length $\|v\|$, the “layer thickness” of the pair of planes representing the covector v^b (namely the distance between the planes) is the reciprocal of the length of the vector. For vector v which is already a unit vector $\|v\| = 1$, this separation is also 1 so the vector pierces exactly one layer of its associated covector.

In \mathbb{R}^3 we first learn to write an equation for a plane in the form

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

or using the position vector notation $\vec{r} = \langle x, y, z \rangle$ and introducing the normal vector $\vec{N} = \langle a, b, c \rangle$ one has

$$\vec{N} \cdot (\vec{r} - \vec{r}_0) = 0.$$

In fact (a, b, c) are the components of the associated covector $(\vec{N})^b$, one of whose level surfaces is being described. This condition is then converted into a geometric statement about points whose difference vector from a reference point is perpendicular to the vector whose components are the same as the coefficients of the linear function (components of the covector).

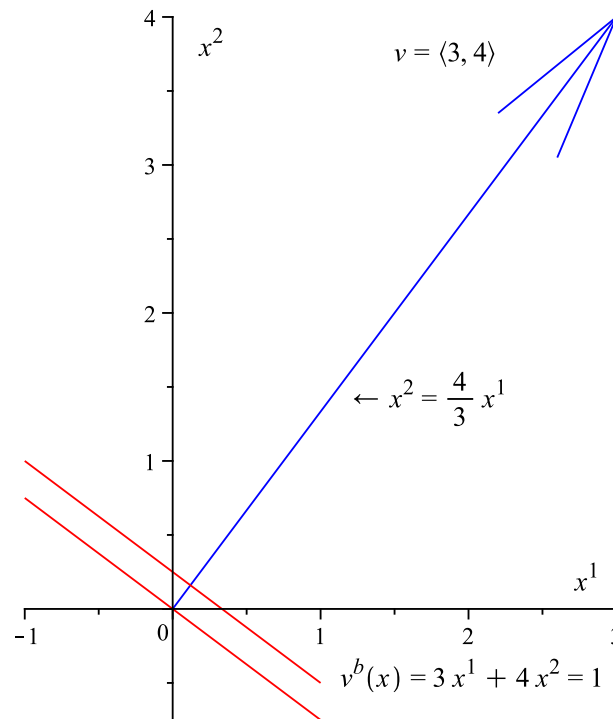


Figure 1.24: Visualizing the covector obtained from a vector by index-lowering from a vector in the plane using the usual dot product.

Exercise 1.6.7.**visualizing a covector in the plane**

Consider the vector $v = \langle 3, 4 \rangle$ of length $5 = \sqrt{3^2 + 4^2}$ in the plane. Find the coordinates of the intersection of the line through v with the unit value line associated with the covector with the same components shown in Fig. 1.24. (Note that the axis intercepts of this latter line are the reciprocals of those components.) Since the latter line has the negative reciprocal slope, it is perpendicular to the line through v , and hence the point of intersection is the closest point to the origin. Show that this distance is in fact $1/5$, the reciprocal of the length of v . The number of times this separation vector fits into v is the ratio of the length of v divided by the length of this separation vector, namely $5/(1/5) = 25$, which must be $v^b(v) = v \cdot v = ||v||^2 = 25$ which is correct. Thus the one geometric length associated with the covector is the reciprocal of the one geometric length associated with the vector, even though formally the vector and covector have the same length 5, which is the square root of the sum of their components.

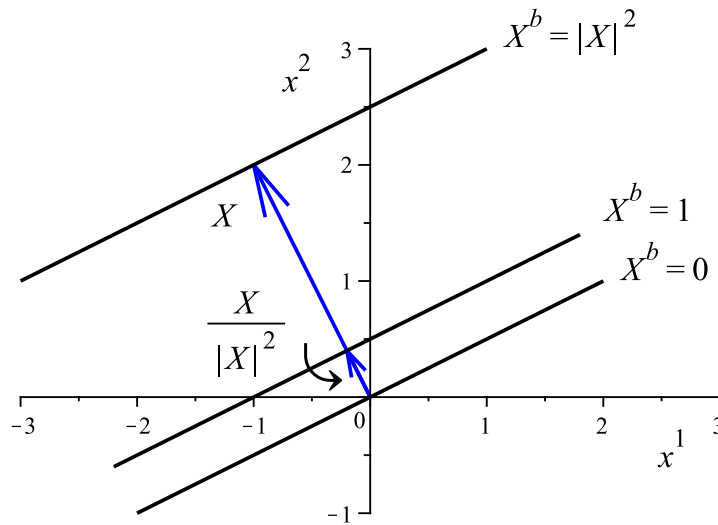


Figure 1.25: In the plane a vector $X = \langle -1, 2 \rangle$ and its corresponding covector $X^b = -\omega^1 + 2\omega^2 = -x^1 + 2x^2$. The level lines of the covector X^b are orthogonal to the vector X , while the tip of X lies in the level line $X^b = |X|^2$. For a unit vector the tip of the vector would lie in the level line corresponding to the value 1.

Example 1.6.7. non-orthonormal basis and index shifting

Consider the change of basis from Example 1.5.4, where the new basis is $e_{1'} = b_1 = \langle 1, 1 \rangle$, $e_{2'} = b_2 = \langle -2, 1 \rangle$, whose coordinate grid is shown in Fig. 1.19. The matrix of inner products of these basis vectors is

$$\underline{G}' = \begin{pmatrix} b_1 \cdot b_1 & b_1 \cdot b_2 \\ b_2 \cdot b_1 & b_2 \cdot b_2 \end{pmatrix} = \underline{B}^T \underline{I} \underline{B} = \begin{pmatrix} 1 & 1 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}$$

As explained in Example 1.3.3, the old components of the new dual basis vectors are the rows of $\underline{B}^{-1} = \langle \langle \frac{1}{3} | \frac{2}{3} \rangle, \langle -\frac{1}{2} | \frac{1}{3} \rangle \rangle$. The vector $X = \langle -3, 0 \rangle = -3e_1 + 0e_2 = -b_1 + b_2$ (i.e., has new components $(X^{1'}, X^{2'}) = (-1, 1)$) has corresponding covector $X^\flat = -3\omega^1 + 0\omega^2$ but we get the new components of this covector by matrix multiplication by \underline{G}'

$$(X_{i'})^T = (G_{i'j'} X^{j'}) = \begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -3 \\ 6 \end{pmatrix} \rightarrow X^\flat = -6\omega^{1'} + 3\omega^{2'} = -3y^1 + 6y^2,$$

To raise the indices back up we need the inverse matrix of our inner product matrix

$$\begin{aligned} \underline{G}'^{-1} &= (\underline{B}^T \underline{B})^{-1} = \underline{B}^{-1} (\underline{B}^T)^{-1} = \underline{B}^{-1} (\underline{B}^{-1})^T \\ &= \frac{1}{3} \begin{pmatrix} 1 \\ \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} = \frac{1}{9} \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} \omega'^1 \cdot \omega'^1 & \omega'^1 \cdot \omega'^2 \\ \omega'^2 \cdot \omega'^1 & \omega'^2 \cdot \omega'^2 \end{pmatrix}, \end{aligned}$$

which are just the usual dot products of the rows of $\underline{A} = \underline{B}^{-1}$. Thus raising the indices back up we do

$$(X^{i'})^T = (X_{j'} G^{j'i'}) = \begin{pmatrix} -3 & 6 \end{pmatrix} \frac{1}{9} \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} -1 & 1 \end{pmatrix}.$$

□

Remark.

Raising and lowering indices are linear maps so we can apply them to the basis vectors and covectors directly

$$\begin{aligned} u^b &= (u^i e_i)^b = (G_{ji} u^i) \omega^j && \text{(component definition)} \\ &= u^i (e_i)^b, && \text{(linearity of map),} \end{aligned}$$

from which it follows that $(e_i)^b = G_{ij} \omega^j$. A similar calculation shows that $(\omega^i)^b = G^{ij} e_j$. Conversely $\omega^i = G^{ij} (e_j)^b$. ▲

Exercise 1.6.8.**transformation of dot products**

We consider the change of basis considered in Exercise 1.3.2 and illustrated in Fig. 1.13, with the associated grid shown as well in Fig. 1.20. The inverse matrix changing the basis has as its columns the old components of the new basis vectors

$$\underline{A}^{-1} = (\underline{E}_1 | \underline{E}_2) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \leftrightarrow \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \underline{A}^{-1} \begin{pmatrix} x^{1'} \\ x^{2'} \end{pmatrix} \leftrightarrow \begin{pmatrix} x^{1'} \\ x^{2'} \end{pmatrix} = \underline{A} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}.$$

Let $e_{i'} \equiv E_i$ for the problem originally discussed above, to follow the change of basis notation. Let $G = \delta_{ij} \omega^i \otimes \omega^j$ be the standard dot product tensor.

(i) Compute $G_{i'j'} = e_{i'} \cdot e_{j'}$ directly by evaluating these dot products individually and then use the matrix transformation law to get the same result.

(ii) Compute $\underline{G}^{-1'} = (G^{i'j'})$ from its index transformation law re-expressed in matrix form.

(iii) The vector $Y = (0, 2) = -2E_1 + 4E_2 = -2e_{1'} + 4e_{2'}$ has $Y^b = 2\omega^1$. Use \underline{G}' to “lower” its indices in the new basis. Verify that the expression for Y^b in terms of $\omega^{1'}$ and $\omega^{2'}$ is $2\omega^1$. ■

Remark.

For a positive-definite inner product where $\text{sgn } u = \text{sgn } G(u, u)$ is always positive for every nonzero vector u , the determinant of the component matrix is positive $\det(G_{ij}) > 0$ since it equals the product of its eigenvalues, which in turn represent the signs of the orthonormal basis vectors in an orthonormal basis, which are all positive by definition.

Using the symmetry property of our inner product $G(X, Y) = G(Y, X)$ it follows that

$$G(X + Y, X + Y) = G(X, X) + G(Y, Y) + 2G(X, Y)$$

from which it follows that

$$G(X, Y) = \frac{1}{2}(G(X + Y, X + Y) - G(X, X) - G(Y, Y)),$$

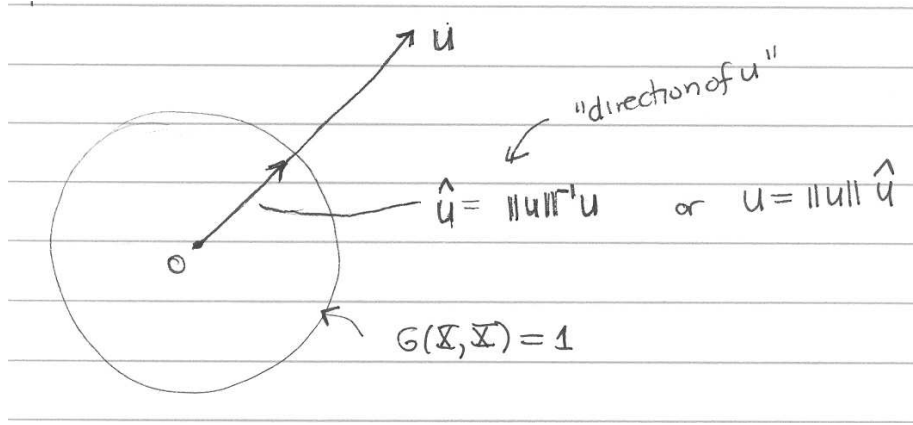


Figure 1.26: The decomposition of a vector into length (magnitude) and direction (unit vector).

so we can determine all inner product values from self-inner-product values, which explains how the unit sphere $G(X, X) = 1$ can contain all the information about the inner product, even angle information.

A given vector u whose magnitude can be represented in terms of its length or magnitude, and a unit vector $\hat{u} \equiv u/||u||$ can be defined by projecting the vector to the unit sphere by dividing the vector by its length, provided it is nonzero: $u = ||u||\hat{u}$, a process called normalization of the vector. Similarly we can evaluate the inner product of two vectors by representing each in this way, thus projecting the inner product to the unit sphere by this process of normalization by factoring out the magnitudes. Thus

$$G(X, Y) = ||X|| ||Y|| G(\hat{X}, \hat{Y})$$

or

$$X \cdot Y = ||X|| ||Y|| \hat{X} \cdot \hat{Y}$$

using the dot product notation.

For unit vectors the above relation becomes

$$G(\hat{X}, \hat{Y}) = \frac{1}{2}[G(\hat{X} + \hat{Y}, \hat{X} + \hat{Y}) - 1 - 1] = \frac{1}{2}||\hat{X} + \hat{Y}||^2 - 1 \in [-1, 1]$$

since $||\hat{X} + \hat{Y}|| \in [0, 2]$ by Euclidean geometry: the extreme values are 0 when $X = -Y$ and 2 when $X = Y$. Thus we can define the result to be the cosine of an angle between two directions

$$\cos \theta \equiv \hat{X} \cdot \hat{Y}.$$

For inner products with nonpositive sign values for some vectors, this argument must be revised. More on this later. ▲

Exercise 1.6.9.**inner products on spaces of square matrices and symmetry**

In Exercise 1.2.1, we explored two trace inner products which agreed on the subspace of symmetric matrices but had opposite signs on the subspace of antisymmetric matrices, one positive-definite and the other indefinite. This generalizes to the n^2 -dimensional vector space $gl(n, \mathbb{R})$ of $n \times n$ real matrices, which it is said to be the Lie algebra of the general linear group, but that is another story we will get to in due time.

Example 1.4.3 defined the standard basis of $gl(n, \mathbb{R})$ by $\underline{A} = A^i_j \underline{e}^j_i$ where \underline{e}^j_i is the $n \times n$ matrix with a single nonzero entry 1 in the i -th row, j -th column, and zeros elsewhere. Then listing entries by consecutive rows

$$\begin{aligned} \underline{A} = & A^1_1 \underline{e}^1_1 + A^1_2 \underline{e}^2_1 + \cdots + A^1_n \underline{e}^n_1 &= u^1 \underline{E}_1 + u^2 \underline{E}_2 + \cdots + u^n \underline{E}_n \\ & + A^2_1 \underline{e}^1_2 + A^2_2 \underline{e}^2_2 + \cdots + A^2_n \underline{e}^n_2 &+ u^{n+1} \underline{E}_{n+1} + u^{n+2} \underline{E}_{n+2} + \cdots + u^{2n} \underline{E}_{2n} \\ & \vdots & \vdots \\ & + A^n_1 \underline{e}^1_n + \cdots + A^n_n \underline{e}^n_n &+ u^{(n-1)n+1} \underline{E}_{(n-1)n+1} + \cdots + u^{n^2} \underline{E}_{n^2} \end{aligned}$$

defines an isomorphism

$$\underline{A} \in V \mapsto (u^1, \dots, u^{n^2}) \in \mathbb{R}^{n^2}$$

from the space of $n \times n$ matrices to \mathbb{R}^{n^2} , mapping this basis onto the standard basis of that space. However, the original matrix notation is more useful because of matrix multiplication.

(i) If the dual basis is defined by $\omega^i_j(\underline{e}^m_n) = \delta^i_n \delta^m_j$, how are the components A^i_j related to them?

(ii) Show that the matrix product law $\underline{e}^i_j \underline{e}^m_n = \delta^i_n \underline{e}^m_j$ for the basis matrices extends by linearity to the usual index formulas for matrix multiplication $[\underline{A} \underline{B}]^i_j = A^i_k B^k_j$.

(iii) Using the notation for trace $\text{Tr}(\underline{A}) = A^i_i$ and transpose $[\underline{A}^T]^i_j = A^j_i$, and recalling the properties (rederive them by expressing in component form!)

$$\text{Tr} \underline{A} = \text{Tr} \underline{A}^T, \quad (\underline{A} \underline{B})^T = \underline{B}^T \underline{A}^T, \quad \text{Tr} \underline{A} \underline{B} = \text{Tr} \underline{B} \underline{A},$$

define two inner products on V by

$$\begin{aligned} G(\underline{A}, \underline{B}) &= \text{Tr} \underline{A}^T \underline{B} = \text{Tr} \underline{A} \underline{B}^T = \sum_{i,j=1}^n A^i_j B^i_j, \\ \mathcal{G}(\underline{A}, \underline{B}) &= \text{Tr} \underline{A} \underline{B} = A^i_j B^j_i. \end{aligned}$$

If we write $[\underline{A}^T]^i_j = \delta_{jn} A^n_m \delta^{mi}$ in order to respect our index conventions, then

$$G(\underline{A}, \underline{B}) = [\underline{A}^T]^i_j B^j_i = \delta_{jn} \delta^{mi} A^n_m B^j_i,$$

and

$$G(\underline{A}, \underline{B}) = \delta_{jn} \delta^{mi} A^n_m B^j_i = \sum_{i,j=1}^n (A^j_i)^2 = \text{sum of squares of all entries of matrix.}$$

Thus G corresponds to the usual dot product on \mathbb{R}^{n^2} under the above correspondence. Make sure you understand this. Note that $\mathcal{G}(\underline{A}, \underline{B}) = G(\underline{A}^T, \underline{B})$.

(iv) Suppose $\underline{A} = \underline{A}^T$ is symmetric and $\underline{B} = -\underline{B}^T$ is antisymmetric. Using the Euclidean property of positive-definiteness $G(\underline{A}, \underline{A}) \geq 0$, with $G(\underline{A}, \underline{A}) = 0$ iff $\underline{A} = \underline{0}$, then

$$\mathcal{G}(\underline{A}, \underline{A}) = G(\underline{A}^T, \underline{A}) = G(\underline{A}, \underline{A}) \geq 0$$

$$\mathcal{G}(\underline{B}, \underline{B}) = G(\underline{B}^T, \underline{B}) = -G(\underline{B}, \underline{B}) \leq 0$$

shows that $\text{sgn } \underline{A} = 1$, $\text{sgn } \underline{B} = -1$ for all nonzero symmetric and antisymmetric matrices respectively. Use a similar argument to show that \underline{A} and \underline{B} are orthogonal with respect to both inner products.

(v) Is the basis $\{e_i^j\}$ of V orthogonal with respect to both inner products? Why?

(vi) The subspaces $\text{SYM}(V)$ and $\text{ALT}(V)$ of symmetric and antisymmetric matrices of V are each vector subspaces (why?), and every matrix can be written uniquely in terms of its symmetric and antisymmetric parts

$$\begin{aligned} \underline{A} &= \underbrace{\text{SYM}(\underline{A})}_{\equiv \frac{1}{2}(\underline{A} + \underline{A}^T)} + \underbrace{\text{ALT}(\underline{A})}_{\equiv \frac{1}{2}(\underline{A} - \underline{A}^T)} \\ &\equiv \frac{1}{2}(\underline{A} + \underline{A}^T) + \frac{1}{2}(\underline{A} - \underline{A}^T) \end{aligned}$$

V is said to be a “direct sum” of these two vector subspaces. Their dimensions are

$$\dim(\text{SYM}(V)) = \sum_{i=1}^n i = n(n+1)/2, \quad \dim(\text{ALT}(V)) = \left(\sum_{i=1}^n i \right) - n = n(n-1)/2.$$

Why?

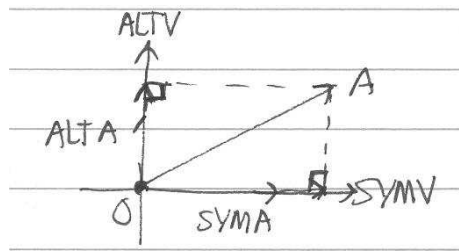


Figure 1.27: The decomposition of a matrix into its symmetric and antisymmetric parts is orthogonal with respect to either inner product.

The maps $A \mapsto \text{SYM}(A)$, $A \mapsto \text{ALT}(A)$ are projection maps associated with this direct sum. They are orthogonal with respect to both inner products, in the sense that they project onto orthogonal subspaces. [Projection maps satisfy $P^2 = P$, $Q^2 = Q$, $PQ = QP = 0$ for a pair (P, Q) which projects onto a pair of subspaces in a direct sum total space.]

(vii) Make the following definitions using an inverted caret to indicate the antisymmetric unit vectors and a rounded one for the symmetric unit vectors

$$\begin{aligned}\check{\underline{E}}^i_j &= \begin{cases} \underline{e}^i_j, & i = j, \\ 2^{-1/2}(\underline{e}^i_j + \underline{e}^j_i), & i \neq j, \end{cases} & \check{\underline{A}}^i_j &= \begin{cases} A^i_j, & i = j, \\ 2^{-1/2}(A^i_j + A^j_i), & i \neq j, \end{cases} \\ \check{\underline{E}}^i_j &= 2^{-1/2}(\underline{e}^i_j - \underline{e}^j_i), & \check{\underline{A}}^i_j &= 2^{-1/2}(A^i_j - A^j_i), \quad i \neq j.\end{aligned}$$

Then

$$\underline{A} = A^i_j \underline{e}^j_i = \sum_{i \leq j} \check{\underline{A}}^i_j \check{\underline{E}}^j_i + \sum_{i < j} \check{\underline{A}}^i_j \check{\underline{E}}^j_i$$

shows that

$$\{\check{\underline{E}}^j_i\}_{i \leq j} \cup \{\check{\underline{E}}^j_i\}_{i < j}$$

is a basis of V adapted to the “orthogonal” direct sum into symmetric and antisymmetric matrices.

Evaluate both inner products of the pairs $(\check{\underline{E}}^j_i, \check{\underline{E}}^m_n)$, $(\check{\underline{E}}^j_i, \check{\underline{E}}^m_n)$, $(\check{\underline{E}}^j_i, \check{\underline{E}}^m_n)$.

What are the lengths of these basis vectors?

What are their signs with respect to each inner product?

What kind of basis is this with respect to either inner product?

(viii) If we introduce the vector index positioning by

$$f = f^i_j \omega^j_i, \quad f^i_j = f(\underline{e}^j_i), \quad \omega^i_j(e^m_n) = \delta^i_n \delta^m_j \text{ (duality),}$$

then we can associate a vector $\underline{F} = f^i_j \underline{e}^j_i$ with each such covector. Show that

$$f(\underline{A}) = \text{Tr } \underline{F} \underline{A} = \mathcal{G}(\underline{F}, \underline{A}),$$

i.e., $\underline{F} = f^\#$ with respect to \mathcal{G} .

[Remark. If we had instead used the notation

$$f = f_i^j \omega^j_i, \quad \omega^i_j(e^m_n) = \delta^i_n \delta^m_j, \quad f_i^j = f(\underline{e}^j_i)$$

we would have found instead

$$f(\underline{A}) = \text{Tr}(\underline{F}^T \underline{A}) = G(\underline{F}, \underline{A})$$

if we let $\underline{F} = f_i^j \underline{e}^j_i$. We would have also used the alternate notation $\underline{A} = A^i_j e_i^j$ from the beginning, which would have resulted in further changes. It is important to realize that a choice of notation implies certain implicit choice not obvious at first. Even other choices $A = A_{ij} e^{ij}$ or $A = A^{ij} e_{ij}$ are possible.]

(ix) Suppose we define

$$H = H^i_j{}^m{}_n \omega^j_i \otimes \omega^m_n, \quad H^i_j{}^m{}_n = H(e^j_i, e^m_n)$$

for any $\binom{0}{2}$ -tensor over V . What are the components of G and \mathcal{G} using this notation?

(x) $\mathcal{G}(\underline{A}, \underline{B}) = \text{Tr } \underline{A} \underline{B}$ defines a $\binom{0}{2}$ -tensor. Why?

For the same reason, for each positive integer p , the following defines a $\binom{0}{p}$ -tensor over V

$$T^{(p)}(\underbrace{\underline{A}, \underline{B}, \dots, \underline{C}}_{p \text{ vector arguments}}) = \text{Tr}(\underbrace{\underline{A} \underline{B} \dots \underline{C}}_{p \text{ factors}}).$$

$T^{(1)}$ is a covector. Express it in terms of the dual basis. Note that the cyclic property of the trace $\text{Tr } \underline{A} \underline{B} \dots \underline{C} \underline{D} = \text{Tr } \underline{B} \dots \underline{C} \underline{D} \underline{A} = \dots$ implies certain symmetries of these tensors. This makes $T^{(2)} = \mathcal{G}$ symmetric.

(xi) If we define

$$D^{(p)}(\underline{A}, \underline{B}, \dots, \underline{C}) = \det(\underline{A} \underline{B} \dots \underline{C}),$$

is this a tensor? Why or why not?

(xii) Sketchy remark for your mathematical interest (just read for pleasure).

The “deWitt” inner product (Google it, or “deWitt metric”)

$$\mathcal{G}_{\text{dW}}(\underline{A}, \underline{B}) = \text{Tr } \underline{A} \underline{B} - \text{Tr } \underline{A} \text{Tr } \underline{B}$$

only differs from $\mathcal{G} = \text{Tr } \underline{A} \underline{B}$ on the symmetric matrices since antisymmetric matrices have zero trace. (Why?) The symmetric matrices themselves may be decomposed into an offdiagonal subspace (again zero trace) and a diagonal subspace, while the diagonal subspace itself can be decomposed into the tracefree subspace and the 1-dimensional “pure trace” subspace of multiples of the identity matrix

$$\begin{aligned} \underline{A} &= \underbrace{\left(\frac{1}{n} \text{Tr } \underline{A}\right) \underline{I}}_{(1) \underline{A}^{\text{trace}}} + \underbrace{\left[\underline{A} - \left(\frac{1}{n} \text{Tr } \underline{A}\right) \underline{I}\right]}_{\underline{A}^{\text{tracefree, sym}}} = \underbrace{\sum_{i=j} A^i_j \underline{e}^j_i}_{(2) \underline{A}^{\text{diagonal}}} + \underbrace{\sum_{i \neq j} A^i_j \underline{e}^j_i}_{\underline{A}^{\text{offdiagonal, sym}}} \\ &\stackrel{(3)}{=} \underbrace{\underline{A}^{\text{trace}} + \underline{A}^{\text{tracefree, diagonal}}}_{(4) \underline{A}^{\text{diagonal}}} + \underline{A}^{\text{offdiagonal, sym}}. \end{aligned}$$

Each of these three decompositions (1), (2), (3) and the restriction (4) of the tracefree decomposition (1) to the diagonal matrices are orthogonal decompositions of the subspace of symmetric matrices with respect to \mathcal{G} (which coincides with G for symmetric matrices, but differs only in sign for the antisymmetric matrices), while the symmetric and antisymmetric matrices are orthogonal with respect to both \mathcal{G} and G so it extends to an orthogonal decomposition of V itself. Anyway the new inner product \mathcal{G}_{dW} only differs from G and \mathcal{G} on the 1-dimensional subspace of pure trace matrices, which has a negative sign with respect to \mathcal{G} . (G and \mathcal{G} have all positive signs for symmetric matrices.) The basis

$$\{\underline{I}\} \cup \{\underline{e}^i_i - \frac{1}{n} \underline{I}\}_{i=1, \dots, n-1}$$

is an orthogonal basis of the diagonal subspace adapted to this pure trace/tracefree decomposition, which has only one basis vector with a negative sign. Such inner products where the

orthonormal bases have only one negative sign are called *Lorentzian* (like 4-dimensional flat Minkowski spacetime).

Without pursuing the details, you can see that just pushing on some simple familiar properties of matrices leads to an extremely rich structure complete with geometry. In fact the space of symmetric matrices with a nonzero determinant is an open subspace of the set of all symmetric matrices and may be interpreted as a “curved space” of all possible (symmetric) inner products on \mathbb{R}^n . This turns out to play a key role in the structure of the complicated nonlinear couple partial differential equations of general relativity called Einstein’s equations, since the metric (inner product on the tangent space at each spacetime point) is the field variable that must be determined by those equations. ■

If you really like mathematics, you can see that by properly recognizing mathematical structure and adapting notation to it, one can create out of nothing a beautiful area of geometry—which in fact is not just idle games playing but often has important applications in physical science. On the other hand, sweeping the structure under the rug in order to arrive immediately at calculational algorithms (as unfortunately we must in a one semester linear algebra course) completely hides this structure and the “beauty.” Our goal is simply to begin to appreciate how this can be uncovered and see how it applies to the geometry of “curved spaces,” which itself has enormous importance in the physical sciences.

Exercise 1.6.10.

projections in \mathbb{R}^3 and \mathbb{M}^4

a) Let $\underline{n} = (n^a)$ be a fixed unit vector in \mathbb{R}^3 : $\delta_{ab}n^an^b = (n^1)^2 + (n^2)^2 + (n^3)^2 = 1$, with $\underline{n}^T = (n_a)$ and $n_a = \delta_{ab}n^b = n^a$. Define the projection matrix $\underline{P}^{\parallel} = (n^an_b)$ along this direction, and the orthogonal projection $\underline{P}^{\perp} = \underline{I} - \underline{P}^{\parallel} = (\delta^a_b - n^an_b)$. Show that $\underline{P}^{\parallel}$ and \underline{P}^{\perp} separately satisfy the projection property and that their product in either order is the zero matrix. Then

$$\underline{v} = \underbrace{\underline{P}^{\parallel} \underline{v}}_{\underline{v}^{\parallel}} + \underbrace{\underline{P}^{\perp} \underline{v}}_{\underline{v}^{\perp}} = (\underline{v} \cdot \underline{n})\underline{n} + (\underline{v} - (\underline{v} \cdot \underline{n})\underline{n})$$

represents the orthogonal decomposition of any vector parallel to and perpendicular to the given direction specified by the unit vector \underline{n} .

b) Clearly these formulas apply to any \mathbb{R}^n with the usual dot product. They can be generalized to any signature inner product by including the sign of the self-dot product $n^kn_k = \pm 1$ of the unit vector specifying the direction. Show that

$$(\underline{P}^{\parallel})^i_j = \frac{n^in_j}{n^kn_k} \quad (\underline{P}^{\perp})^i_j = \delta^i_j - \frac{n^in_j}{n^kn_k}$$

are the components of mutually orthogonal projection operators.

c) For a Lorentzian spacetime timelike directions have a negative sign in these formulas. Given a unit timelike vector $u \cdot u = -1$, its parallel projection picks out the timelike part of a vector with respect to an observer whose world line is aligned with u , while its orthogonal

complement projection projects out the spacelike part belonging to what is called the “local rest space” associated with this observer. The projection matrices then become

$$(\underline{P}^{\parallel})^i_j = -u^i u_j \quad (\underline{P}^{\perp})^i_j = \delta^i_j + u^i u_j.$$

For a unit vector in \mathbb{M}^4 of the form $u = \langle \cosh \alpha, \sinh \alpha n^a \rangle$, $a = 1, 2, 3$, where $n \cdot n = \delta_{ab} n^a n^b = 1$ is a spacelike unit vector, evaluate the two projections of a general vector $X = \langle X^0, X^a \rangle$. ■

Exercise 1.6.11.

Gram Schmidt diagonalization

The orthogonal projection process with respect to a single direction can be iterated to achieve an orthogonal direct sum of 1-dimensional subspaces and an associated orthonormal basis. For any nondegenerate inner product on a vector space V , given a basis consisting of vectors with nonzero length, one can always construct an orthogonal basis with respect to that inner product by a simple algorithmic procedure called the Gram-Schmidt procedure of orthogonalization, which can then be normalized to make an orthonormal basis. This procedure depends on the order of the vectors.

One keeps the first vector in the set, and then projects the second vector orthogonally to the first vector to get an orthogonal vector to replace the second vector, but the span of the two vectors is still the same. Next one takes the third vector and projects it orthogonally to the plane of the first two vectors by removing its vector projections along each of the first two vectors already obtained to obtain a third vector orthogonal to the plane of the first two as the third vector in the new set. One continues until the last vector has been replaced in this manner. The result is a set of linearly independent orthogonal vectors since at each step we took linearly independent combinations of the previous vectors to obtain the next vector. Provided all these vectors have nonzero lengths, we can normalize them by dividing each by its length. In the indefinite-case we have some complications, but then we also get something new when we encounter null vectors in this process.

To illustrate this procedure consider the three columns of the upper triangular matrix

$$\underline{M} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \langle \underline{m}_1, \underline{m}_2, \underline{m}_3 \rangle,$$

considered as (obviously linearly independent) vectors in \mathbb{R}^3 , first ordered left to right, then ordered right to left. Their inner products are

$$(m_i \cdot m_j) = \underline{M}^T \underline{M} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}.$$

Using the notation

$$\text{proj}_u v = (v \cdot \hat{u}) \hat{u} = \frac{(v \cdot u)}{(u \cdot u)} u$$

for the projection of v along u , then $v - \text{proj}_u v$ is the projection orthogonal to u . We start by keeping $e_{1'} = m_1 = \langle 1, 0, 0 \rangle$, which is already a unit vector. Then we calculate

$$e_{2'} = m_2 - \text{proj}_{e_{1'}} m_2 = \langle 1, 1, 0 \rangle - (\langle 1, 1, 0 \rangle \cdot \langle 1, 0, 0 \rangle) \langle 1, 0, 0 \rangle = \langle 0, 1, 0 \rangle.$$

Finally we calculate

$$\begin{aligned} e_{3'} &= m_3 - \text{proj}_{e_{1'}} m_3 - \text{proj}_{e_{2'}} m_3 \\ &= \langle 1, 1, 1 \rangle - (\langle 1, 1, 1 \rangle \cdot \langle 1, 0, 0 \rangle) \langle 1, 0, 0 \rangle - (\langle 1, 1, 1 \rangle \cdot \langle 0, 1, 0 \rangle) \langle 0, 1, 0 \rangle = \langle 0, 0, 1 \rangle. \end{aligned}$$

Well, this was too simple: the vectors ended up already unit vectors, and in fact we returned to the standard orthonormal basis of \mathbb{R}^3 .

a) Now try it in the reverse order: $\langle 1, 1, 1 \rangle$, $\langle 1, 1, 0 \rangle$, $\langle 1, 0, 0 \rangle$. Then let $e_{i'} = B^j_i e_j$ be the resulting orthogonal vectors, and let $e_{i''} = P^j_i e_j$ be the resulting orthonormal vectors, expressed in terms of the standard basis. Since both $\{e_i\}$ and $\{e_{i''}\}$ are orthonormal bases, the matrix \underline{P} must be an orthogonal matrix. We already know that its columns are mutually orthogonal unit vectors. Check that its rows are also mutually orthogonal unit vectors by evaluating $\underline{P} \underline{P}^T = \underline{I}$.

b) Evaluate the relatively simple looking matrix $\underline{G} = \underline{M}^T \underline{I} \underline{M}$ of inner products of the basis m_i . What happens when you try to find the exact eigenvectors of \underline{G} with technology in order to diagonalize this matrix? You quickly see that you must numerically approximate them, and you can show that the numerical approximations to the eigenvectors are orthogonal to a high degree of approximation, the error due to the numerics in the approximation process of finite digit math (and hence these eigenvectors can be normalized to make the choice of eigenvectors orthonormal). If \underline{G} were the matrix of some other interesting quantity like a moment of inertia tensor, then we would be limited to using orthogonal transformations to diagonalize it (in order to apply laws of physics which are simple in orthonormal Cartesian coordinates) and these would be the unique principal axes associated with that tensor. However, if we are only trying to find an orthonormal basis of the space starting from the original non-orthonormal basis, then the Gram-Schmidt process applied to all six orderings of the original three vectors easily leads to orthonormal bases which not only diagonalize the matrix \underline{G} of inner products but make it equal to the identity matrix. The big difference is that the eigenvalue problem treats the matrix \underline{M} as the components of a $\binom{1}{1}$ -tensor, while the orthonormalization of the original vectors treats it as a $\binom{0}{2}$ -tensor. This tells us that in the eigenvalue problem with a symmetric matrix, there is more going on, since it requires the usual dot product to rethink it as the components of a mixed tensor and therefore of a linear transformation. In fact the symmetric moment of inertia tensor in the rigid body problem we will encounter later, is actually the matrix of a linear transformation from the angular velocity vector to the angular momentum vector. In other words a symmetric linear transformation requires both a linear transformation and an inner product to describe.

■

Fact

Any real symmetric matrix can (in principle) be diagonalized by the eigenvector approach with all real eigenvalues and orthogonal eigenvectors, which can be chosen to be normalized and

therefore orthonormal. When the eigenvalues are distinct, the diagonalizing transformation is unique up to reflections and permutations of the orthogonal axes.

Explanation

A simple derivation shows that the eigenvalues have to be real. Letting \bar{x} be the complex conjugate of an eigenvector x , and using the symmetry $A^T = A$ and reality $\bar{A} = A$ properties, one sees from the eigenvalue condition and its complex conjugate

$$\begin{aligned}\bar{x}^T [Ax = \lambda x] &\rightarrow \lambda \bar{x}^T x = \bar{x}^T Ax, \\ [A\bar{x} = \bar{\lambda}\bar{x}]^T x &\rightarrow \bar{\lambda} \bar{x}^T x = (A\bar{x})^T x = \bar{x}^T A^T x = \bar{x}^T Ax,\end{aligned}$$

and by subtraction

$$0 = (\bar{\lambda} - \lambda) \bar{x}^T x = (\bar{\lambda} - \lambda) \delta_{ij} \bar{x}^i x^j = (\bar{\lambda} - \lambda) \sum_{i=1}^n |x^i|^2,$$

which forces $\bar{\lambda} = \lambda$ since x must be a nonzero vector. Orthogonality of eigenvectors x_1, x_2 associated with distinct eigenvalues $\lambda_1 \neq \lambda_2$ is a similar short computation

$$\begin{aligned}x_2^T [Ax_1 = \lambda_1 x_1] &\rightarrow \lambda_1 x_2^T x_1 = x_2^T Ax_1, \\ [Ax_2 = \lambda_2 x_2]^T x_1 &\rightarrow \lambda_2 x_2^T x_1 = (Ax_2)^T x_1 = x_2^T A^T x_1 = x_2^T Ax_1,\end{aligned}$$

and again by subtraction

$$0 = (\lambda_1 - \lambda_2) x_2^T x_1 = (\lambda_1 - \lambda_2) x_2 \cdot x_1$$

but since the eigenvalues are distinct it follows that the dot product of the eigenvectors must be zero, i.e., the eigenvectors must be orthogonal. Here we introduced the standard dot product on \mathbb{R}^n , which defines the orthogonality properties.

For degenerate eigenvalues, one can choose an orthogonal basis of the eigenspace, so that one can get an orthogonal basis of eigenvectors, which can then be normalized to an orthonormal basis, and hence can be obtained from the standard basis by an orthogonal matrix.

If the eigenvalues of the original matrix are unique, then the orthonormal frame of eigenvectors is fixed up to reflections of each eigenvector. When an eigenvalue is repeated, its eigenspace allows any choice of orthonormal frame in that subspace.

Exercise 1.6.12.

second derivative test

Consider the function

$$f(x^1, x^2) = \frac{1}{2} (8(x^1)^2 - 4x^1 x^2 + 5(x^2)^2) = \frac{1}{2} M_{ij} x^i x^j, \quad M = \begin{pmatrix} 8 & -2 \\ -2 & 5 \end{pmatrix},$$

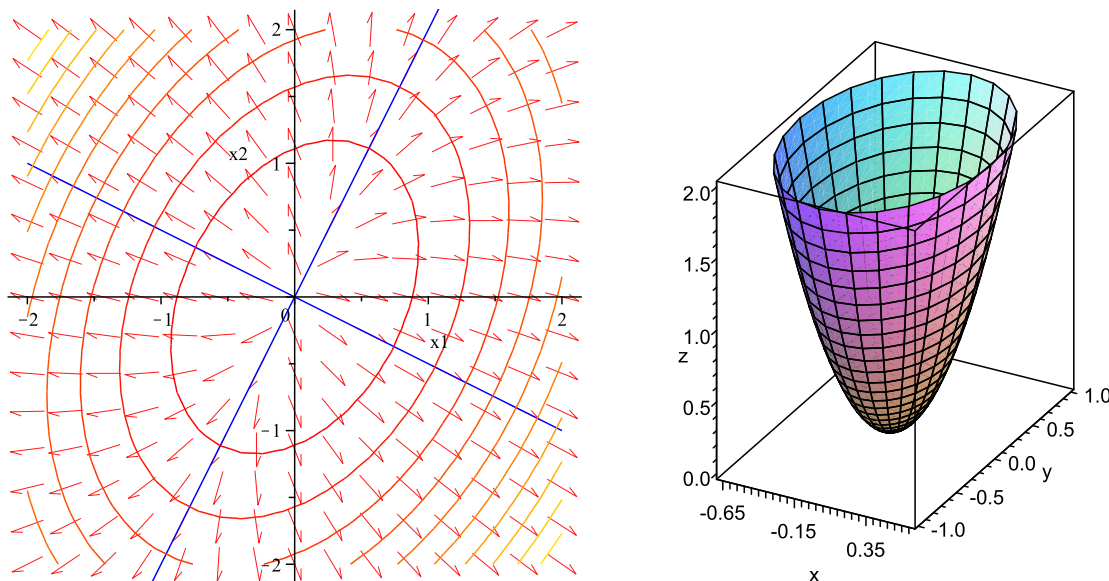


Figure 1.28: Left: the gradient of a quadratic form function is a linear vector field whose directionfield is perpendicular to the level curves of the function in the plane. Right: a plot of the graph rotated so that its principal axes (the eigenvectors of the quadratic form coefficient matrix, rotated from the Cartesian axes) are aligned with the Cartesian coordinate axes. The elliptical and parabolic cross-sections are shown through its representation as a parametrized surface.

which is a quadratic form in the two variables x^1, x^2 with an obvious critical point at the origin. Figure 1.18 shows the contour plot of this function, together with its gradient vector field, and the plot of the function with respect to rotated axes aligned with the semiaxes of its elliptical level curves.

a) Confirm that \underline{M} is the constant symmetric second derivative matrix for this function.

b) Find the eigenvalues λ_1, λ_2 and eigenvectors of this matrix, normalize the eigenvectors to obtain an orthonormal basis $\{b_1, b_2\}$ (order them so the second is obtained from the first by a 90 degree rotation in the counterclockwise direction) with associated orthogonal matrix $\underline{B} = \langle \underline{b}_1 | \underline{b}_2 \rangle$, and use this new basis to change to new orthonormal coordinates in which the second derivative matrix is diagonal. What is the angle of rotation of the axes? Re-express the function in terms of the new coordinates confirming that it takes the form

$$f(x^1, x^2) = g(x^{1'}, x^{2'}) = \frac{1}{2}[\lambda_1(x^{1'})^2 + \lambda_2(x^{2'})^2]$$

and re-evaluate the new second derivative matrix directly and also by the transformation law $\underline{M}' = \underline{B}^T \underline{M} \underline{B}$, which implies $\det \underline{M}' = (\det \underline{B})^2 \det \underline{M} = \det \underline{M}$ (since for an orthogonal matrix, $(\det \underline{B})^2 = \det(\underline{B}\underline{B}^T) = \det \underline{I} = 1$). Expressed in terms of partial derivatives in multivariable calculus notation using (x, y, z) , this says that $f_{xx}f_{yy} - f_{xy}^2 = g_{x'x'}g_{y'y'}$. In the new axes both these second partial derivatives (the eigenvalues) are positive indicating a local minimum at

the origin in each new coordinate direction, so since $g_{x'y'} = 0$, it is easy to see that this must be a local minimum in all directions.

c) Now let's simplify the problem by working in the new coordinates, dropping primes, and reverting to the multivariable calculus notation

$$g(x, y) = \frac{1}{2} (9x^2 + 4y^2) .$$

The right hand side of Figure 1.18 shows the graph of this function with respect to the rotated axes. Each vertical cross-section of the graph of g by a vertical plane through the z axis is a parabola, while each horizontal plane cross-section is an ellipse, so this is an elliptic paraboloid. We can make a clever parametrization of both the parabolas and ellipses at once by introducing deformed polar coordinates in the new axes by substituting $(x, y) = (2\rho \cos \phi, 3\rho \sin \phi)$ into g to get the graph $z = g(2\rho \cos \phi, 3\rho \sin \phi) = 36\rho^2$, so the position vector of a point on the graph of g can be represented in the form

$$\vec{r}(\rho, \phi) = \langle 2\rho \cos \phi, 3\rho \sin \phi, 18\rho^2 \rangle .$$

Notice that if we compare the horizontal part of this position vector with polar coordinates in the plane

$$(r \cos \theta, r \sin \theta) = (2\rho \cos \phi, 3\rho \sin \phi)$$

one finds

$$\frac{y}{x} = \tan \theta = \frac{3}{2} \tan \phi ,$$

so ϕ does not agree with the polar coordinate angle θ of the projection to the x - y plane (or the cylindrical coordinate angle ϕ) except where the tangent is zero or plus or minus infinity, which occurs along the x and y axes where either angle is some integer multiple of $\pi/2$.

This is an example of a parametrized surface. Varying independently the two parameters ρ and ϕ sweeps out the surface. Varying one at a time traces out the elliptical and parabolic cross-sectional curves. In particular, letting $\rho = \rho_0, \phi = t$ parametrizes the elliptical cross-sectional curves at constant $z_0 = 18\rho_0^2$, while $\rho = t, \phi = \phi_0$ parametrizes the parabolic cross-sections in the direction making an angle θ with the x axis satisfying $\tan \theta = 3/2 \tan \phi_0$.

We can use the following formula from multivariable calculus to calculate the curvature of a parametrized space curve applied to each of these parametrized curves (here the prime denotes differentiation)

$$\kappa(t) = \frac{\|\vec{r}'(t) \times \vec{r}''(t)\|}{\|\vec{r}'(t)\|^3} .$$

Use this for the ellipses first and the parabolas second. For the ellipses, plot the curvature as a function t , $0 \leq t \leq 2\pi$ for $\rho_0 = 1/3$, noting extrema at the minor and major axes of the ellipse. For the parabolas, plot the curvature as a function ϕ_0 , $0 \leq \phi_0 \leq 2\pi$ for $t = 0, 1/3$, again noting extrema at the minor and major axes of the ellipse.

d) The two tangent vectors $r'(t)$ for each such parametrized ellipse and parabola are just the partial derivatives of the position vector:

$$\vec{r}_1(\rho, \phi) = \frac{\partial \vec{r}}{\partial \rho}(\rho, \phi) , \quad \vec{r}_2(\rho, \phi) = \frac{\partial \vec{r}}{\partial \phi}(\rho, \phi) .$$

Evaluate the matrix of their inner products and show that these two tangent vectors are orthogonal only along the minor and major axes of the ellipses.

e) Evaluate the normal vector $\vec{N}(\rho, \phi) = \vec{r}_1(\rho, \phi) \times \vec{r}_2(\rho, \phi) = |\vec{N}(\rho, \phi)|\hat{n}(\rho, \phi)$ and its length $|\vec{N}(\rho, \phi)|$ and direction $\hat{n}(\rho, \phi)$. Evaluate numerically the double integral of the length for the parameter range $0 \leq \rho \leq 1/3$, $0 \leq \phi \leq 2\pi$. Later we will see that this is the surface area of this surface below the plane $z = 2$. ■

Cute fact (an aside for your reading pleasure): geometric interpretation of index lowering on vectors

The relationship between a vector and covector determined by the Euclidean metric has a cute geometric interpretation. Consider the case of \mathbb{R}^2 . The unit circle (all vectors of length 1) tells us everything we need to know about the Euclidean geometry of the metric tensor. The following identity tells us self-inner products are enough to recover all inner products

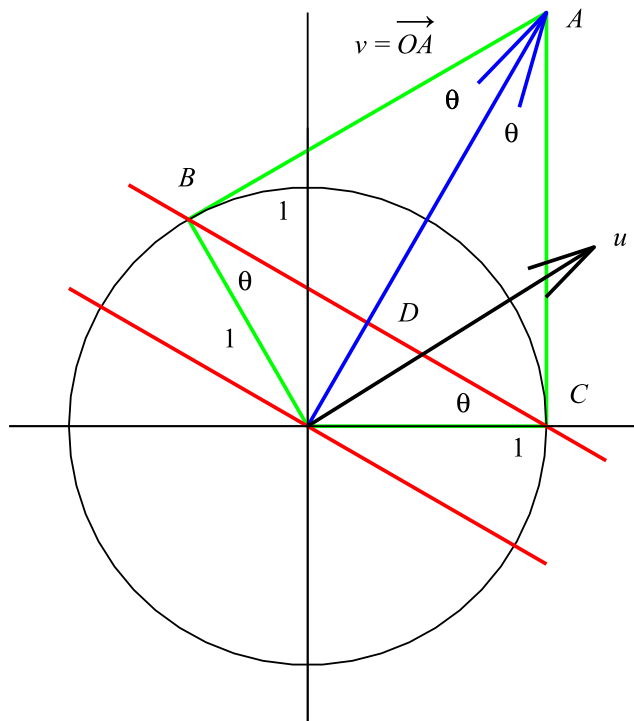
$$\begin{aligned} G(X + Y, X + Y) &= G(X, X) + G(Y, Y) + 2G(X, Y) \\ \rightarrow G(X, Y) &= \frac{1}{2}[G(X + Y, X + Y) - G(X, X) - G(Y, Y)]. \end{aligned}$$

The self-inner product is a “quadratic form” in the same language that calls a linear function (covector) a “linear form.” Thus if we know the set of all vectors with unit length, we can determine the length of all multiples of these unit vectors. The unit circle (or the unit sphere in higher dimensions) is therefore the nonlinear analogue of the line (or plane or hyperplane in higher dimensions) of vectors X satisfying $f(X) = 1$ for a covector f , which can be taken as a representative set in the vector space to visualize the quadratic form. This geometry can be extended to visualize geometrically the relationship between a covector and a vector.

Suppose $v = \overrightarrow{OA}$ is a vector with length bigger than 1 as in Fig. 1.29. Draw in the tangents AB and AC to the unit circle and connect the points of tangency B and C , letting D be the intersection of BC with OA . By symmetry the line segment OA is the angle bisector of angle BAC and the bisector of the opposite side of the isosceles triangle $\triangle ABC$, to which it is perpendicular. Note that the right triangles $\triangle ABO$ and $\triangle BDO$ are similar. Then from the right triangle $\triangle ABO$, since the hypotenuse has length $\|v\|$, one has $\sin \theta = 1/\|v\|$, and from the right triangle $\triangle BDO$, since the side has unit length, one has $\sin \theta = |OD|/1$. Equating this shows that the $|OD| = 1/\|v\|$, namely the line BC is the level curve $v^b(x) = 1$ associated with the index-lowered covector v^b with the same components as the original vector according to our general discussion above. Draw a line parallel to BC through the origin. Then these two parallel lines represent the covector $v^b = G(\cdot, v)$, since their separation is the reciprocal of the length of v , and they are orthogonal to v . If we have another vector u , then the value of the metric on the pair

$$G(u, v) = v^b(u) = v \cdot u$$

is the number of “layers” of v^b pierced by u , which is about 2.5 in the diagram. This picture can be extended to the case $\|v\| < 1$ by inversion. Thus we get a nice geometrical way to



associate v^b with v and with its evaluation on another vector u using the geometry associated with the usual dot product.

Does the same scheme work for any “positive definite” inner product on \mathbb{R}^2 ? Such an inner product has the following form

$$G = \underbrace{A}_{G_{11}} \omega^1 \otimes \omega^1 + \underbrace{B}_{G_{12} = G_{21}} (\omega^1 \otimes \omega^2 + \omega^2 \otimes \omega^1) + \underbrace{C}_{G_{22}} \omega^2 \otimes \omega^2,$$

$$A > 0, \quad C > 0, \quad AC - B^2 = \det G > 0.$$
$$1 = G(X, X) = Ax^2 + 2Bxy + Cy^2$$

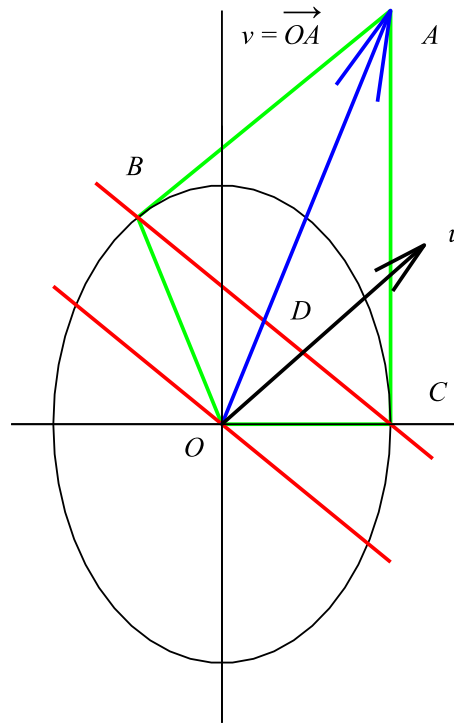


Figure 1.30: The same construction with an ellipse determines the pair of lines representing the index-lowered covector $v^b = G(\cdot, v)$ associated with a vector v in a general positive-definite inner product.

is now an ellipse centered at the origin. The condition $AC - B^2 > 0$ guarantees that this is indeed an ellipse. However, exactly the same tangent construction shown in Fig. 1.30 continues to determine the 2 lines which represent the index lowered vector v^b in terms of the original vector v . Thus the “unit circle” in the new geometry continues to contain all the geometrical information contained in the corresponding inner product. For higher dimensions the corresponding “unit sphere” or “unit hypersphere” construction of the usual dot product becomes an ellipsoidal surface for a more general positive-definite inner product, which again contains all the geometrical information necessary to determine the inner product with which it is associated. ▲

This also works with the unit “hypersphere” in \mathbb{R}^n with the usual dot product except one has a tangent “hypercone” with an $(n - 2)$ -sphere of tangency through which passes a hyperplane orthogonal to v . Together with the parallel hyperplane through the center of the hypersphere (the origin O), we get the representation of the covector v^b and its value on another vector u in terms of the number of layers pierced. Thus the unit hypersphere can “represent” an inner product, which is a symmetric positive-definite $\binom{0}{2}$ -tensor. For an indefinite inner product, spheres become “pseudo-spheres” (some kind of higher dimensional hyperbolic hypersurfaces). The “degenerate” (zero determinant component matrix) symmetric $\binom{0}{2}$ -tensors

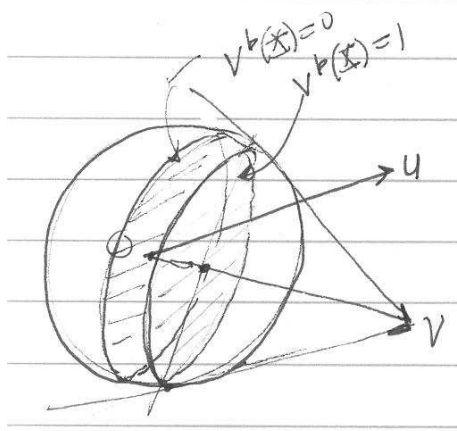


Figure 1.31: The same construction with an sphere in \mathbb{R}^3 (revolve the circle construction around the axis from the tip of the vector to the center of the circle to justify it).

have hypercylinder representations, etc. We don't need these geometric interpretations, but sometimes they can be useful, and it is important to realize that the abstraction of tensors and their mathematics is very closely connected to concrete visualizable geometry.

Exercise 1.6.13.

visualizing positive-definite inner products for the plane

The orthogonal matrix $\underline{B} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ represents an active counterclockwise rotation of the plane by 45° . Its inverse $\underline{A} = \underline{B}^{-1} = \underline{B}^T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ is the matrix of the associated coordinate transformation for the components of vectors with respect to the new basis vectors $\langle e_{1'}, e_{2'} \rangle = \langle \underline{b}_1, \underline{b}_2 \rangle = \underline{B}$ which result from the active rotation of the standard basis vectors.

The Cartesian coordinates are the standard dual basis $x = \omega^1$, $y = \omega^2$, so the change of basis

$$\begin{aligned} \omega^{i'} &= A^i_{j'} \omega^j \longleftrightarrow \begin{pmatrix} \omega^{1'} \\ \omega^{2'} \end{pmatrix} = \underline{A} \begin{pmatrix} \omega^1 \\ \omega^2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \omega^1 \\ \omega^2 \end{pmatrix} = \begin{pmatrix} (\omega^1 + \omega^2)/\sqrt{2} \\ (-\omega^1 + \omega^2)/\sqrt{2} \end{pmatrix} \\ e_{i'} &= e_j A^{-1j}_{i'} \longleftrightarrow (e_{1'} \ e_{2'}) = (e_1 \ e_2) \underline{A}^{-1} = (e_1 \ e_2) \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{(e_1 + e_2)}{\sqrt{2}} & \frac{(e_1 - e_2)}{\sqrt{2}} \end{pmatrix} \end{aligned}$$

corresponds to the Cartesian coordinate change

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \underline{A} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (x + y)/\sqrt{2} \\ (-x + y)/\sqrt{2} \end{pmatrix}.$$

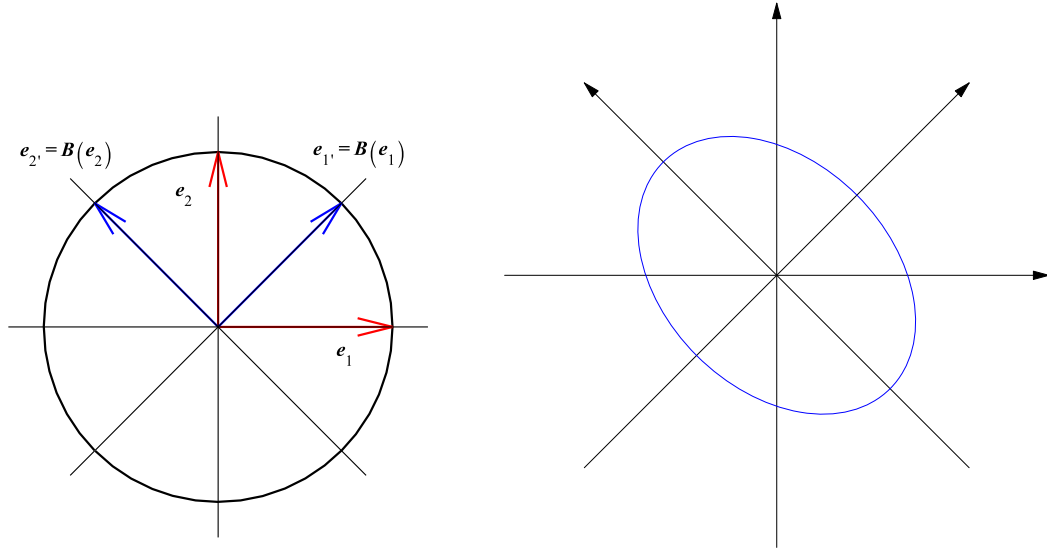


Figure 1.32: Left: a rotation of the natural basis of \mathbb{R}^2 counterclockwise by 45 degrees. Right: the principal axes of the ellipse associated with H are rotated by 45 degrees with respect to the Cartesian axes associated with the natural basis.

Consider the symmetric tensor $H = H_{ij} \omega^i \otimes \omega^j$ and the mixed tensor $L = L^i_j e_i \otimes \omega^j$ with the same matrix of components $\underline{H} = \begin{pmatrix} 3/2 & 1/2 \\ 1/2 & 3/2 \end{pmatrix} = \underline{L}$.

(i) Verify that the change of basis leads to

$$\underline{H}' = \underline{A}^{-1T} \underline{H} \underline{A}^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad \underline{L}' = \underline{A} \underline{H} \underline{A}^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix},$$

i.e., diagonalizes $\underline{H} = \underline{L}$, while not changing the Euclidean inner product G :

$$\underline{G}' = \underline{A}^{-1T} \underline{I} \underline{A}^{-1} = \underline{A} \underline{A}^{-1} = \underline{I},$$

which are consequences of the orthogonality condition $\underline{A}^{-1} = \underline{A}^T$ or equivalently $\underline{A}^{-1T} = \underline{A}$.

(ii) Compute the magnitude of H in each basis (computing the magnitude with \underline{G}).

(iii) H may itself define an inner product. Its “unit circle” is an ellipse defined by the equation

$$1 = H(\langle x, y \rangle, \langle x, y \rangle) = \frac{1}{2}(3x^2 + 3y^2 + 2xy) = 2(x')^2 + (y')^2,$$

whose semiaxes are 1 and $1/\sqrt{2}$. Note that we can introduce orthonormal coordinates with respect to this new inner product by scaling the new basis by the diagonal scaling matrix

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}, \quad \underline{H}'' = \begin{pmatrix} 1 & 0 \\ \sqrt{2} & 1 \end{pmatrix} \underline{H}' \begin{pmatrix} 1 & 0 \\ \sqrt{2} & 1 \end{pmatrix} = \underline{I},$$

which then leads to the standard equation of a unit circle in the new coordinates. This can be used to check the geometric construction that interprets the covector related to the original vector by index lowering, but with respect to the new inner product. If you had more time, maybe you would do this.

iv) Note that the final change of basis has no effect on the matrix \underline{L}' since diagonal matrices commute

$$\underline{L}'' = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \end{pmatrix}^{-1} \underline{L}' \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \end{pmatrix} = \underline{L}'.$$

Thus different geometrical interpretations of the same symmetric matrix to which we apply the eigenvector algorithm leads to different final outcomes because the matrix transforms differently because of those different geometrical interpretations, once we go beyond the orthogonal transformations which are sufficient to diagonalize it.

■

1.7 Matrix groups

Unavoidably tangled up with differential geometry is the theory of continuous groups, called Lie groups after the mathematician who made gigantic first steps in their understanding. The rotations and translations of ordinary flat space are imbedded in our thinking about geometry, and many interesting curved spaces have symmetries that are useful to understand in studying their geometry. But the general theory of such groups is another can of worms, so we need to confine our attention to the simpler subset of matrix groups and translation groups here or we will get bogged down in too many details. However, the group of linear transformations of a vector space V into itself has already been a fundamental tool needed to consider changes of bases and the resulting transformation laws for the components of tensors, which involves the general linear group $GL(n, \mathbb{R}^n)$ acting first on \mathbb{R}^n , and then on the tensor spaces over \mathbb{R}^n . This important matrix group contains subgroups which are important for geometry, so we need to pay a bit more attention to its mathematical structure.

We are already familiar with some groups and their properties in the very number system we use every day. The set of real numbers \mathbb{R} is an “Abelian” group under the operation of addition. Abelian just means that the group law which assigns a third member of the group to an ordered pair of group elements does not depend on the order of the two elements, as indeed addition works: $a + b = b + a = c$ (closure of the group operation means that the result of the group law is contained in the same set). The number zero 0 is the additive identity element of the group, so called since adding 0 to any other element does not change the element: $a + 0 = 0 + a = a$. Each element a in the group of real numbers under addition has an additive inverse $-a$ such that $a + (-a) = 0$: their sum is the additive identity. Finally addition is associative: $(a + b) + c = a + (b + c)$. Usually the group law is called a product, which is exactly what it is in the group of nonzero real numbers under multiplication, another Abelian group: $ab = ba$. Here the multiplicative identity is the number 1: $1a = a1 = a$, while the multiplicative inverse of every positive number is its reciprocal $a^{-1} = 1/a$. Of course multiplication of real numbers is associative: $(ab)c = a(bc)$.

In general a group G is a set of elements with a group product: $(a, b) \in G \times G \mapsto ab = P(a, b) \in G$, where P is some function of a and b , and which has an identity element I such that $Ia = aI = a$ for every element, and every element a has an inverse a^{-1} such that $a^{-1}a = I = aa^{-1}$, and finally it is associative: $(ab)c = a(bc)$. The set of positive numbers is clearly a subgroup of the group of nonzero numbers since the product of two positive numbers is again positive and the inverse of every positive number is positive, while the number 1 is positive, so it is a group in its own right, but a subset of the larger group of all nonzero real numbers, which has two disjoint subsets: the positive numbers and the negative numbers.

Every real vector space V is an Abelian group under vector addition. In particular \mathbb{R}^n is an n -dimensional real group. If we let one element of the group “multiply” the entire group, every point moves to a new point in general. With these Abelian groups, adding a given vector to all points in the space (thought of as position vectors) is said to “translate” all the points of space. We can think of \mathbb{R}^n as acting on itself by translation, which in fact is a symmetry of the Euclidean geometry we associate with these spaces as encoded in the usual dot product. Translation does not affect dot products. It is said to be a symmetry of the dot product or

metric. Of course another symmetry of the geometry that we take for granted in the cases $n = 2, 3$ are the rotations. The rotations together with the translations form a larger group of symmetries of the Euclidean geometry of \mathbb{R}^n . The rotations of Euclidean 3-space are the most familiar example of a matrix group, a group which acts on \mathbb{R}^3 by matrix multiplication to move the points around on spheres of constant distance from the origin.

The set of points onto which a given point is moved by matrix multiplication by all members of the group is called the orbit of the point under the action of the group. The spheres centered at the origin are the orbits of the rotation group alone acting on ordinary space. Or consider any fixed point in \mathbb{R}^n . It can be translated to any other location by a translation of all points of space by the difference vector, so the orbit of any point is the whole space. Such a group action is called transitive, otherwise intransitive, like the spherical orbits of the rotation group within \mathbb{R}^3 . Take the group consisting of the rotations about the z axis of \mathbb{R}^3 together with all translations along that axis. The orbits of any point not on the axis are cylinders: the rotation sweeps out a circle around the axis while the translations sweep out lines parallel to the axis, and together, cylinders result. The action of this 2-dimensional subgroup of the 6-dimensional Euclidean group of rotations (3-dimensions) plus translations (3 dimensions) is intransitive. The translational symmetry alone describes the homogeneity of the geometry of \mathbb{R}^3 in that any point is equivalent to any other point under the action of this group which keeps all lengths and angles between vectors at any given point unchanged. This is called a symmetry of the Euclidean geometry. Similarly the rotations about any fixed point describe the isotropy of that geometry, all directions are equivalent and rotations preserve the lengths and angles between vectors at any other point. The mathematical name for such a group is isometry group. We are interested in the isometry groups of flat space and of Lorentz Minkowski spacetimes of various dimensions.

The set of all $n \times n$ real matrices is designated by $gl(n, \mathbb{R})$. We have already seen that this is an n^2 -dimensional vector space. The subset of matrices for which the determinant is nonzero is designated by $GL(n, \mathbb{R})$, which stands for the general linear group of the given dimension. Any square matrix with nonzero determinant has an inverse, and the identity matrix is the multiplicative identity for matrix multiplication, which is associative, so this is a group, clearly of the same dimension as the set of all $n \times n$ matrices. Any subset of the group which is closed under matrix multiplication is also a group in its own right, and hence a subgroup. The matrix groups are all the possible subgroups of $GL(n, \mathbb{R})$, including $GL(n, \mathbb{R})$ itself. Since the determinant of a product is the product of the determinants, the set $SL(n, \mathbb{R})$ of unit determinant matrices is closed under matrix multiplication and hence a natural subgroup. If we have an inner product, the condition that inner products of pairs of vectors be invariant under matrix multiplication determines a subgroup, called the orthogonal group $O(n, \mathbb{R})$ for the usual dot product. This was already explored for \mathbb{R}^2 in Exercise 1.4.1 with the rotations of the Euclidean plane and the boosts of Minkowski 2-spacetime.

Like the real numbers under addition or nonzero real numbers under multiplication the rotations and boosts in 2 dimensions, these rotations and boosts are 1-parameter groups, depending on the trigonometric angle or hyperbolic angle parameter respectively which are both

additive parameters:

$$R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2), \quad B(\theta_1)B(\theta_2) = B(\theta_1 + \theta_2).$$

A fact that we will simply state is that all 1-parameter continuous groups can be “reparametrized” so that the new parameter which describes the set of group elements is additive in the group multiplication, which makes them all Abelian. For example, the positive real numbers under multiplication can be made additive in a new logarithmic parameter $t = \ln x$:

$$x_1 = e^{t_1}, \quad x_2 = e^{t_2} : \quad x_1 x_2 = e^{t_1} e^{t_2} = e^{t_1 + t_2}.$$

This exponential representation is more general and can be repeated in matrix form for the matrix groups. Exercise 1.4.1 showed that the differential of the rotation/boost matrix has the form

$$d\underline{A}(\lambda) \underline{A}(\lambda)^{-1} = \underline{K} d\lambda \leftrightarrow \frac{d\underline{A}}{d\lambda} = \underline{K} \underline{A}$$

where $\underline{A}(0) = \underline{I}$ and

$$(\lambda, \underline{A}, \underline{K}) = \left(\theta, \underline{R}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right), \quad \left(\beta, \underline{B}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right).$$

By introducing the matrix exponential defined in the same way as the ordinary exponential by its Taylor series

$$e^{\underline{P}} = \sum_{n=0}^{\infty} \frac{1}{n!} \underline{P}^n, \quad \underline{P}^0 = \underline{I},$$

so that

$$e^0 = \underline{I},$$

one sees by differentiation

$$\begin{aligned} \frac{d}{d\lambda} e^{\lambda \underline{K}} &= \frac{d}{d\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \underline{K}^n = \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \underline{K}^n = \underline{K} \left(\sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \underline{K}^{n-1} \right) \\ &= \underline{K} \left(\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \underline{K}^n \right) = \underline{K} e^{\lambda \underline{K}} \\ &= e^{\lambda \underline{K}} \underline{K} \quad (\text{order does not matter here, powers of } \underline{K}) \end{aligned}$$

and hence if \underline{C} is a constant square matrix

$$\frac{d}{d\lambda} (e^{\lambda \underline{K}} \underline{C}) = \underline{K} (e^{\lambda \underline{K}} \underline{C}).$$

This is the general solution of $d\underline{A}/d\lambda = \underline{K} \underline{A}$ and the initial condition $\underline{A}(0) = \underline{I}$ sets $\underline{C} = \underline{I}$, so we can conclude that

$$\underline{A}(\lambda) = e^{\lambda \underline{K}}$$

for the rotation/boost matrix. In fact one can easily sum the series to verify this. As we already have shown for this explicit case for the resulting 2×2 matrices of a rotation or hyperbolic rotation in the plane, this parametrization has the additive property

$$e^{\lambda_1 \underline{K}} e^{\lambda_2 \underline{K}} = e^{(\lambda_1 + \lambda_2) \underline{K}}.$$

This is indeed true for the matrix exponential of any matrix, for the same reason that it is true for the scalar exponential, following from its power series representation.

Another way to sum the exponential series is by the diagonalization technique, since it is easy to exponentiate a diagonal matrix—the result is just a new diagonal matrix whose entries are the exponentials of the original entries

$$\exp(\text{diag } \underline{K}_D) = \exp(\text{diag}(k_1, \dots, k_n)) = \text{diag}(e^{k_1}, \dots, e^{k_n}),$$

adopting the obvious notation for a diagonal matrix with entries k_i along the main diagonal and zeros elsewhere

$$\text{diag}(k_1, \dots, k_n) = \begin{pmatrix} k_1 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & k_n \end{pmatrix}.$$

Since matrix multiplication of such matrices is just scalar multiplication of the corresponding diagonal entries, powers of a diagonal matrix just raise the diagonal entries to the power, and the exponential of the matrix just reduces to the diagonal matrix of the exponentials of its diagonal entries.

Suppose $\underline{K} = \underline{B} \underline{K}_D \underline{B}^{-1}$ (equivalent to $\underline{K}_D = \underline{B}^{-1} \underline{K} \underline{B}$) diagonalizes \underline{K} in terms of such a diagonal matrix \underline{K}_D , where \underline{B} is the corresponding matrix of eigenvectors of \underline{K} . Then

$$\begin{aligned} e^{\underline{K}} = e^{\underline{B} \underline{K}_D \underline{B}^{-1}} &= \sum_{j=0}^{\infty} \frac{1}{j!} (\underline{B} \underline{K}_D \underline{B}^{-1})^j \\ &= \sum_{j=0}^{\infty} \frac{1}{j!} \underline{B} \underline{K}_D^j \underline{B}^{-1} \\ &= \underline{B} \left(\sum_{j=0}^{\infty} \frac{1}{j!} \underline{K}_D^j \right) \underline{B}^{-1} \\ &= \underline{B} \left(\sum_{j=0}^{\infty} \frac{1}{j!} \text{diag}(k_1^j, \dots, k_n^j) \right) \underline{B}^{-1} \\ &= \underline{B} \left(\text{diag} \left(\sum_{j=0}^{\infty} \frac{1}{j!} k_1^j, \dots, \sum_{j=0}^{\infty} \frac{1}{j!} k_n^j \right) \right) \underline{B}^{-1} \\ &= \underline{B} (\text{diag}(e^{k_1}, \dots, e^{k_n})) \underline{B}^{-1}. \end{aligned}$$

The crucial step in the second equality follows from the fact that the factors of \underline{B} and \underline{B}^{-1} cancel each other in powers of $\underline{B} \underline{K}_D \underline{B}^{-1}$, for example,

$$\underline{B} \underline{K}_D \underline{B}^{-1} \underline{B} \underline{K}_D \underline{B}^{-1} = \underline{B} \underline{K}_D \underline{K}_D \underline{B}^{-1} = \underline{B} \underline{K}_D^2 \underline{B}^{-1}.$$

Thus one can simply exponentiate the eigenvalues and carry out the final multiplication above to obtain the entries of the exponential matrix.

Remark.

Recall $\underline{x} = \underline{B} \underline{y}$, $\underline{y} = \underline{B}^{-1} \underline{x}$ expresses the old coordinates \underline{x} in \mathbb{R}^n in terms of the new coordinates \underline{y} in a basis of eigenvectors which form the columns of \underline{B} . The combination $\underline{K}_D \underline{y} = \underline{B}^{-1} \underline{K} \underline{B} \underline{y}$ working from right to left takes the new coordinates to the old coordinates, multiplies the old coordinates by \underline{K} , then takes the old coordinates of that new point to its new coordinates, but the final new coordinates are simply rescaled by the eigenvalues so the result has to be a diagonal matrix.

A constant coefficient linear system of differential equations, usually written

$$\frac{d\underline{x}}{dt} = \underline{K} \underline{x},$$

when transformed to the new coordinates become uncoupled first order differential equations for them

$$\frac{d\underline{y}}{dt} = \underline{K}_D \underline{y} \leftrightarrow \frac{dy^i}{dt} = k_i y^i \quad (\text{no sum on } i),$$

with exponential solutions

$$y^i = e^{tk_i} C^i \rightarrow \underline{y} = \text{diag}(e^{tk_1}, \dots, e^{tk_n}) \underline{C}, \quad \underline{C} = \underline{y}(0) = \underline{B}^{-1} \underline{x}(0).$$

Going back to the old coordinates

$$\underline{x} = \underline{B} e^{t\underline{K}_D} \underline{y}(0) = \underline{B} e^{t\underline{K}_D} \underline{B}^{-1} \underline{x}(0) = e^{t\underline{B} \underline{K}_D \underline{B}^{-1}} \underline{x}(0) = e^{t\underline{K}} \underline{x}(0).$$

Thus we were only one step from the matrix exponential (back substituting the constant vector \underline{C} in terms of the initial value of \underline{x}) when we arrived at our eigenvector solution algorithm for such systems of differential equations.

If we can find a way to directly sum the matrix exponential series in terms of scalar power series we recognize, we can avoid that technique altogether. This proves to be possible when exponentiating matrices whose index-lowered form is antisymmetric.

While we are discussing the properties of the matrix exponential via diagonalization, let's derive a very useful relation between the determinant, the trace and the exponential. Using the invariance of the trace $\text{Tr}(\underline{A} \underline{B} \underline{C}) = \text{Tr}(\underline{C} \underline{A} \underline{B})$ to remove the factors of \underline{B} and its inverse, and using the fact that the determinant of a diagonal matrix is the product of its diagonal values, we find for 2×2 matrices

$$\begin{aligned} \det e^{\underline{K}} &= \det e^{(\underline{B} \underline{K}_D \underline{B}^{-1})} = \det (\underline{B} e^{\underline{K}_D} \underline{B}^{-1}) && (\underline{B} \text{ comes out of exponential}) \\ &= \det \underline{B} \det(e^{\underline{K}_D}) \det(\underline{B}^{-1}) && (\text{determinant factors}) \\ &= \det e^{\underline{K}_D} && (\text{determinant of inverse} = \text{inverse of determinant}) \\ &= \det e^{\text{diag}(k_1, k_2)} = \det \text{diag}(e^{k_1}, e^{k_2}) = e^{k_1} e^{k_2} && (\text{product of diagonal values}) \\ &= e^{k_1 + k_2} = e^{\text{Tr} \underline{K}_D} = e^{\text{Tr}(\underline{B}^{-1} \underline{B} \underline{K}_D)} = e^{\text{Tr}(\underline{B} \underline{K}_D \underline{B}^{-1})} && (\text{cyclic property of trace}) \\ &= e^{\text{Tr} \underline{K}}. \end{aligned}$$

This clearly holds for any dimension n . ▲

Exercise 1.7.1.

2×2 matrix exponentials

- a) Show that if $\underline{K} = \langle\langle 0|\epsilon\rangle, \langle 1|0\rangle\rangle$, where $\epsilon = \pm 1$, then $\underline{K}^2 = \epsilon \underline{I}$.
- b) Use this to separate the Taylor series for the matrix exponential $e^{\lambda \underline{K}}$ into even and odd series which are the scalar coefficients of \underline{I} and \underline{K} . Recognize the Taylor series of the trigonometric/hyperbolic cosine and sine functions to recover the original definition of $\underline{R}(\lambda)$ for $\epsilon = -1$ and $\underline{B}(\lambda)$ for $\epsilon = 1$.
- c) Carry out the diagonalization approach to evaluate the boost exponential $e^{\lambda \underline{K}}$ with $\epsilon = -1$. If you feel motivated, wade through a bit more complex arithmetic to do the same for the rotation exponential, which leads to complex eigenvalues and eigenvectors. This is actually useful, since it shows how the eigenvector approach for rotations naturally leads to complex exponentials. ■

What we have also shown in these two examples of rotations and pseudo-rotations in the plane is that the tangent to these two curves $\underline{A}(\lambda)$ in the space of 2×2 matrices when they pass through the identity matrix, namely $d\underline{A}/d\lambda(0) = \underline{K}$, lie in the span of the matrix \underline{K} , each a 1-dimensional subspace of the space of 2×2 matrices. Had we parametrized these curves in other ways, the tangents at the identity matrix would have still been multiples of \underline{K} . In other words this subspace contains all possible tangents to parametrized curves in the given matrix subgroup.

Suppose we have an $n \times n$ matrix subgroup and consider a curve $\underline{A}(t)$ in the group passing through the identity matrix $\underline{I} = \underline{A}(0)$. Consider the subspace of the tangent space at the identity corresponding to the tangents to all such possible curves in the subgroup:

$$\frac{d\underline{A}}{dt}(0).$$

This subspace characterizes the group and is called the Lie algebra of the matrix group. $gl(n, \mathbb{R})$ plays this role for $GL(n, \mathbb{R})$, since the derivative of a curve of matrices with nonzero determinant is simply a matrix without any special properties. The symmetry groups of inner products are important for geometry, so we need to consider them. If we take any matrix \underline{K} and consider the curve $e^{\lambda \underline{K}}$, then under matrix multiplication we will find

$$e^{\lambda_1 \underline{K}} e^{\lambda_2 \underline{K}} = e^{(\lambda_1 + \lambda_2) \underline{K}}.$$

This closure of the matrix multiplication makes this a 1-parameter subgroup of $GL(n, \mathbb{R})$, coupled with the fact that the matrix exponential satisfies the identity (see the preceding remark discussion with $\underline{K} \rightarrow \lambda \underline{K}$)

$$\det e^{\lambda \underline{K}} = e^{\lambda \text{Tr } \underline{K}}$$

which guarantees that this curve which starts at the identity matrix remains invertible. The inverse of $e^{\lambda \underline{K}}$ is just $e^{-\lambda \underline{K}}$. In general, the matrices of a matrix group close enough to the

identity matrix can be obtained by the matrix exponential of the matrices in their Lie algebra. This identity relating the determinant and trace through the matrix exponential can be proven for diagonalizable matrices as we showed above for 2×2 matrices, and is in fact true for all matrices, but we will simply accept this latter fact without a proof. A direct consequence of this relation between the trace, the exponential and the determinant is that unit determinant matrices result from the matrix exponential of tracefree matrices, so the Lie algebra $sl(n, \mathbb{R})$ of the special linear group is just the subspace of tracefree matrices.

We are primarily interested in the matrix groups associated with invariance of inner products, the so called orthogonal groups. As described in section 1.6 the matrix of components $\underline{G} = (G_{ij})$ of an inner product on \mathbb{R}^n transforms back to itself (therefore remaining invariant) under a linear transformation of the natural basis with matrix \underline{A} according to the generalized orthogonality condition

$$\begin{aligned} (\underline{G} \underline{A})^T &= \underline{G} \underline{A}^{-1} \quad \Leftrightarrow \quad G_{kj} A^j_i = G_{ij} A^{-1j}_k \\ &\quad \Leftrightarrow \quad A_{ki} = A^{-1}_{ik} . \end{aligned}$$

This says that if we use our index-lowering convention to lower the contravariant indices on the matrix and its inverse, they are related by the ordinary transpose operation. If A belongs to a matrix group, we can infer from this by differentiation a condition on its Lie algebra. We need a preliminary result for this.

Exercise 1.7.2.

differential of a family of matrices preserving an inner product

a) Show by differentiating $\underline{A}^{-1} \underline{A} = \underline{I}$ and multiplying the result on the right by \underline{A}^{-1} that

$$d\underline{A}^{-1} = -\underline{A}^{-1} d\underline{A} \underline{A}^{-1} ,$$

or in components

$$dA^{-1i}_j = -A^{-1i}_m dA^m_n A^{-1n}_j .$$

b) Convince yourself of the steps in the following derivation, recalling that $(\underline{A} \underline{B} \underline{C})^T = \underline{C}^T \underline{B}^T \underline{A}^T$ and $(\underline{A}^T)^T = \underline{A}$

$$\begin{aligned} \underline{A}^{-1T} \underline{G} \underline{A}^{-1} &= \underline{G} \quad \rightarrow \quad d\underline{A}^{-1T} \underline{G} \underline{A}^{-1} + \underline{A}^{-1T} \underline{G} d\underline{A}^{-1} = 0 \\ &\rightarrow \quad -(\underline{A}^{-1T} d\underline{A} \underline{A}^{-1})^T \underline{G} \underline{A}^{-1} - \underline{A}^{-1T} \underline{G} (\underline{A}^{-1} d\underline{A} \underline{A}^{-1}) = 0 \\ &\rightarrow \quad \underline{A}^{-1T} d\underline{A}^T (\underline{A}^{-1} \underline{G} \underline{A}^{-1}) = -(\underline{A}^{-1T} \underline{G} \underline{A}^{-1}) d\underline{A} \underline{A}^{-1} \\ &\rightarrow \quad \underline{A}^{-1T} d\underline{A}^T \underline{G} = -\underline{G} d\underline{A} \underline{A}^{-1} \\ &\rightarrow \quad (\underline{G} d\underline{A} \underline{A}^{-1})^T = -\underline{G} d\underline{A} \underline{A}^{-1} . \end{aligned}$$

Translating this back into index notation and using the index-lowering convention on the contravariant index yields

$$[d\underline{A} \underline{A}^{-1}]_{ji} = -[d\underline{A} \underline{A}^{-1}]_{ij} .$$

This just says that the fully covariant component matrix $[d\underline{A} \underline{A}^{-1}]^b$ is antisymmetric, independent of the inner product matrix (G_{ij}) . Thus antisymmetric matrices will play a key role in inner product geometry.



If we go further with this calculation, passing from the differential to the derivative along a curve $\underline{A}(\lambda)$ through the identity $\underline{A}(0) = \underline{I}$, then we get

$$\frac{dA_{ij}}{d\lambda}(0) = -\frac{dA_{ji}}{d\lambda}(0).$$

By definition the Lie algebra of a matrix group consists of the set of all matrices which result from this tangent operation at the identity matrix, so this condition says that the Lie algebra of these generalized orthogonal matrix groups consists of matrices which are antisymmetric after the first index is lowered with the metric matrix

$$K_{ij} = -K_{ji} \quad \text{or more explicitly} \quad G_{ik}K^k{}_j = -G_{jk}K^k{}_i.$$

If we examine this condition in an orthonormal basis in which \underline{G} is diagonal and $G_{ii} = \pm 1$, then this forces \underline{K} to be an off-diagonal matrix since $i = j$ implies $K_{ii} = 0$ so $K^i{}_i = 0$ (no sum on i). If $i \neq j$ and G_{ii} and G_{jj} have the same sign, K is antisymmetric in the index pair (i, j) and the matrix $\underline{e}^j{}_i - \underline{e}^i{}_j$ generates an ordinary rotation in the x^i - x^j plane, while if G_{ii} and G_{jj} have the opposite sign, \underline{K} is symmetric in the index pair (i, j) and the matrix $\underline{e}^j{}_i + \underline{e}^i{}_j$ generates a hyperbolic rotation in the x^i - x^j plane.

This result tells us that the set of matrices whose index lowered form is antisymmetric exponentiate to orthogonal matrices with respect to the given inner product, i.e., this inner product defined antisymmetry condition on matrices defines the Lie algebra of the corresponding orthogonal group. To be more concrete, let $SO(P, M)$ be the special generalized orthogonal group of unit determinant $n \times n$ matrices which leave invariant the diagonal metric component matrix

$$\underline{G} = \text{diag}(\underbrace{-1, \dots, -1}_M, \underbrace{1, \dots, 1}_P) \quad n = P + M$$

with P positive signs and M of negative signs. We can assume $P > M$ without loss of generality since the overall sign has no influence on the corresponding matrix group, and we can order an orthonormal basis so that the negative signs are all first. The corresponding Lie algebra is denoted by $so(P, M)$. The ordinary orthogonal group $O(n, \mathbb{R})$, including its unit determinant subgroup called the special orthogonal group $SO(n, \mathbb{R})$, has the Lie algebra $so(n, \mathbb{R})$ of antisymmetric matrices. $O(n-1, 1)$ is the Lorentz group, while $SO(n-1, 1)$ is the proper Lorentz group of unit determinant Lorentz matrices.

The matrices \underline{K} for the Lie algebras of these various orthogonal groups, also called the generating matrices since they exponentiate to matrices in the group itself, are interpreted as limiting small transformations in the sense

$$\frac{d}{d\lambda} [e^{\lambda \underline{K}} \underline{x}] = \underline{K} [e^{\lambda \underline{K}} \underline{x}] \rightarrow \left. \frac{d}{d\lambda} [e^{\lambda \underline{K}} \underline{x}] \right|_{\lambda=0} = \underline{K} \underline{x}.$$

These Lie algebra matrices \underline{K} give the direction and amount by which a point \underline{x} in \mathbb{R}^n begins to move under an orthogonal transformation $e^{\lambda \underline{K}}$ associated with the given inner product. In fact

the right hand side of the vector differential equation is a column matrix linear function of the coordinates \underline{x} representing the components of a vector field whose flow lines are the solutions of the equation.

Transformation groups

Of course we are not interested in groups for their own sake here, but in how they act on other spaces (including themselves) as transformation groups. A transformation of a space into itself is a 1-1 map of the space into itself which moves its points around (without losing any, this is the 1-1 condition!). A group G is said to act on a space M if there is a map ρ from the group into the set of transformations of M into itself which respects the group multiplication law

$$x \in M \mapsto \rho_a(x), \quad \rho_a \circ \rho_b = \rho_{ab}.$$

However, the map itself ρ itself may not be 1-1 in the sense that more than one group element may be sent to the same transformation of M into itself. For example, suppose we consider the group $GL(n, \mathbb{R})$ acting on the whole real line by multiplying numbers by the matrix determinant

$$x \rightarrow \rho(\underline{A})(x) = (\det \underline{A}) x.$$

Because of the product law for determinants, this satisfies the above condition, but the image of the entire group is $GL(1, \mathbb{R})$, the 1-dimensional Abelian scale and reflection group of the real line (multiplication by positive numbers is said to scale the real line, while multiplication by -1 is a reflection). Thus we really have a map from one group into another which respects the group law, which is called a homomorphism. When it is instead 1-1, it is called an isomorphism, which means the groups are essentially the same.

Exercise 1.7.3.

linear transformations plus translations: the inhomogeneous linear group

Suppose we consider both the general linear group $GL(n, \mathbb{R})$ and the translation group acting together as a combined group on \mathbb{R}^n called the inhomogeneous general linear group $IGL(n, \mathbb{R})$

$$\underline{x} \in \mathbb{R}^n \mapsto \rho(\underline{A}, \underline{b})(\underline{x}) = \underline{A}\underline{x} + \underline{b}.$$

If we add one more row and column to the matrices of $GL(n, \mathbb{R})$ such that the bottom row is all zeros except for a 1 in the final entry, but the entries above that 1 in the last column are n arbitrary numbers, then the new matrix is also invertible and so belongs to $GL(n+1, \mathbb{R})$. Its action on $n+1$ component vectors of the form $\langle x^1, \dots, x^n, 1 \rangle$ exactly mirrors the action of the combined linear transformations and translations of \mathbb{R}^n .

Consider the case $n = 2$ for concreteness and let

$$\underline{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \underline{b} = \begin{pmatrix} b^1 \\ b^2 \end{pmatrix}, \quad (\underline{A}, \underline{b}) \mapsto \begin{pmatrix} a & b & b^1 \\ c & d & b^2 \\ 0 & 0 & 1 \end{pmatrix}.$$

a) Show that the product of two matrices of this type is again of this type. Use a computer algebra system to make this calculation less boring.

b) Show that left multiplication of $\langle x^1, x^2, 1 \rangle$ by such a matrix exactly mirrors the combined linear transformations plus translations of \mathbb{R}^2 , so the inhomogeneous general linear group on \mathbb{R}^2 is really just an isomorphic matrix subgroup of $GL(3, \mathbb{R})$. This is true in general.

c) From the matrix product of two such matrices derive the group multiplication law for the combined linear transformations and translations

$$\rho(\underline{A}_2, \underline{b}_2) \circ \rho(\underline{A}_1, \underline{b}_1) = \rho(\underline{A}_2 \underline{A}_1, \underline{A}_2 \underline{b}_1 + \underline{b}_2).$$

This is said to be a semi-direct product group, since the linear transformation matrix subgroup retains its own group law uninfluenced by the translations, but the translations are acted on by this subgroup. ■

Exercise 1.7.4.

$U(1)$, unit complex numbers

The unitary group $U(1)$ consists of complex numbers U (1×1 matrices!) such that their complex conjugate is their inverse: $\bar{U} = U^{-1}$ or $\bar{U}U = 1$, which is the “unitary condition” that U have unit magnitude. Expressing $U = x + iy$ in terms of its real and imaginary parts, then we get the equation of the unit circle in the complex plane

$$1 = (x - iy)(x + iy) = x^2 + y^2.$$

Pretending we don’t know how to parametrize the unit circle S^1 in the complex plane for a moment through its identification with the real plane, consider the following.

a) If we represent $U = e^u$ in terms of the exponential of a complex number u , then show that the unitary condition implies that u be pure imaginary: $u = i\theta$, where θ is any real number. Thus $U = e^{i\theta}$ and hence we have a canonical Abelian parametrization of this 1-dimensional (over the real numbers) group whose group manifold is S^1 which reduces the group law to addition of the group parameter

$$e^{i\theta_1} e^{i\theta_2} = e^{i(\theta_1 + \theta_2)}.$$

b) Evaluate $e^{i\theta}$ using the exponential power series by separating the even and odd power terms and recognizing the cosine and sine power series to obtain

$$e^{i\theta} = \cos \theta + i \sin \theta.$$

c) Show that if this group acts on the complex vector space \mathbb{C} (the complex plane!) by multiplication, $\Psi \rightarrow e^{i\theta}\Psi$, then the magnitude of Ψ is invariant. If $|\Psi|^2 = \bar{\Psi}\Psi \leq 1$ is interpreted as a probability, then this action of $U(1)$ leaves the probability invariant. If instead we have a function Ψ defined on \mathbb{R}^3 whose squared magnitude $|\Psi|^2$ is interpreted as a quantum field probability distribution, then $e^{i\theta}\Psi$ for any real function θ gives the same probability distribution. If only the probability matters, i.e., is measurable, then the “quantum wave function” Ψ

can undergo a unitary change at each point without affecting outcomes. This is called a gauge transformation.

d) Evaluate the real and imaginary parts of the product $e^{i\theta_0}(x+iy)$ to show that this complex rotation of the complex plane corresponds exactly to an active rotation of the corresponding real plane. If instead we use the polar representation of $x+iy = re^{i\theta}$, then $e^{i\theta_0}re^{i\theta} = re^{i(\theta+\theta_0)}$ makes this obvious. ■

Remark.

The Lie algebra of the group of positive numbers under multiplication is the set of all real numbers, related by the exponential function: $x = e^\theta$. The Lie algebra of the unitary group $U(1)$ of unit complex numbers is the set of all purely imaginary numbers, again related by the exponential: $U = e^{i\theta}$. They have the same group multiplication law, which is addition of the real parameter θ in each case, but although they are locally the same (homomorphic) near the identity $\theta = 0$, there are an infinite number of values of θ for the first group that correspond to each distinct point on the circle of the second group. We can also introduce the real group $S(2, \mathbb{R})$ of rotations in the plane, whose Lie algebra consists of the 1-dimensional vector space of all antisymmetric 2×2 matrices, in terms of which the matrix exponential allows one to represent any 2×2 orthogonal matrix. This group is isomorphic to $U(1)$ since there is a clear 1-1 relationship between the two (same group manifold S^1 which corresponds to the possible distinct rotations of the plane). The Lie algebra of a group determines its local structure near the identity, but globally one may have very different realizations of that local structure. ▲

Non-Abelian groups

The 1-parameter continuous matrix subgroups $e^{t\mathbf{K}}$ are all Abelian for fixed \mathbf{K} and variable t , with an additive law for the parameter t in the composition of two such matrices, but it is the non-Abelian nature of continuous groups of higher dimension that is the most interesting. The translations of \mathbb{R}^3 are the typical prototype of a 3-dimensional Abelian group, but the rotations are clearly not Abelian as we will see shortly in an example below.

For non-Abelian groups the order of the factors in the group multiplication law matters, so we can define three different ways in which we can use that group law to allow the group to act on itself as a transformation group

$$\begin{aligned} \underline{A} &\rightarrow L_{\underline{B}}(\underline{A}) = \underline{B}\underline{A}, & (\text{left translation by } \underline{A}) \\ \underline{A} &\rightarrow R_{\underline{B}}(\underline{A}) = \underline{A}\underline{B}, & (\text{right translation by } \underline{A}) \\ \underline{A} &\rightarrow AD_{\underline{B}}(\underline{A}) = \underline{B}\underline{A}\underline{B}^{-1} = L_{\underline{B}} \circ R_{\underline{B}^{-1}}(\underline{A}) = R_{\underline{B}^{-1}} \circ L_{\underline{B}}(\underline{A}). & (\text{conjugation by } \underline{A}) \end{aligned}$$

Under translation, the any matrix \underline{A} can be moved to any other point \underline{B} in the matrix group simply by left translating all points of the group by the matrix $\underline{B}\underline{A}^{-1}$ or right translating all points of the group by the matrix $\underline{A}^{-1}\underline{B}$ so it is a transitive action: the orbit of any point

under this action is the whole space on which the group acts (itself). A given pair of left and right translations clearly commute when performed in succession

$$R_C \circ L_B(\underline{A}) = L_B \circ R_C(\underline{A}) = \underline{B} \underline{A} \underline{C}.$$

When the pair of left and right translations are by inverse matrices, their combined action is a conjugation. The action of the group on itself by conjugation is called the adjoint action, and the identity matrix is a fixed point of this action since it does not move, so this is an intransitive action (its orbits are not the whole space) and is more like the rotation group action on \mathbb{R}^3 , which leaves the origin fixed, and the orbits of all other points are spheres. For an Abelian group the adjoint action is trivial, that is, it acts as the identity transformation since the matrix and its inverse “cancel each other out,” so every point is a fixed point. The size (dimension) of the adjoint group is in some sense a measure of how non-Abelian a group is. Its maximum dimension is the dimension of the group itself.

Exercise 1.7.5.

left and right translations and the adjoint action of a group

a) Establish the following relations

$$L_A \circ L_B = L_{AB}, \quad R_A \circ R_B = R_{BA}, \quad AD_A \circ AD_B = AD_{AB}.$$

b) Show that right translation by the inverse satisfies the group composition rule in the order required for the action of a group

$$R_{A^{-1}} \circ R_{B^{-1}} = R_{(AB)^{-1}}.$$

Since the left and right translation actions of the group onto itself are 1-1 (why?), the two group actions are isomorphic to each other. ■

Successive rotations of the unit vectors $\hat{i}, \hat{j}, \hat{k}$ illustrate the non-Abelian nature of the rotation group. Suppose we rotate \hat{j} by 90 degrees around \hat{i} to end up at \hat{k} , and then rotate \hat{k} by 90 degrees around \hat{j} to end up at \hat{i} , using the right hand rule to determine the direction of each rotation about an axis. If instead we first rotate \hat{j} around itself by 90 degrees so that it does not change, and then apply to it a rotation by 90 degrees around \hat{i} , it ends up at \hat{k} instead of at \hat{i} . Thus reversing the order of the two rotations leads to very different results.

Suppose we try this in the limit of very small rotations. A counterclockwise rotation by a small angle θ of a position vector $\vec{r} = r\hat{r}$ about an axis with direction unit vector \hat{n} moves the tip of the position vector along the cross-product direction $\hat{n} \times \hat{r}$ by approximately an amount equal to the arc $s = r\theta$, namely by the change in position $r\theta \hat{n} \times \hat{r} = \vec{\theta} \times \vec{r}$, where $\vec{\theta} = \theta \hat{n}$. Thus the effect of a limitingly small rotation on the position vector is that

$$\vec{r} \mapsto \vec{r} + \vec{\theta} \times \vec{r}.$$

We saw in Exercise 1.2.4 that the matrix form of taking a cross product with a fixed vector $\underline{\theta}$ on the left is

$$\underline{r} \mapsto \theta^i \underline{L}_i \underline{r}$$

so that under a small rotation we have

$$\underline{r} \mapsto (\underline{I} + \theta^i \underline{L}_i) \underline{r} \approx e^{\theta^i \underline{L}_i} \underline{r},$$

which is just the linear approximation to the matrix exponential power series. If we do two small rotations in succession, then

$$\underline{r} \mapsto (\underline{I} + \theta_2^i \underline{L}_i) (\underline{I} + \theta_1^i \underline{L}_i) \underline{r}.$$

If we do them in the opposite order the difference between the two to lowest order measures the failure of the two rotations to commute, and gives the extra rotation necessary to make one agree with the other. Ignoring terms higher than second order in products of the two angular vectors, one finds

$$\begin{aligned} [e^{\theta_1^i \underline{L}_i}, e^{\theta_2^j \underline{L}_j}] &\equiv e^{\theta_1^i \underline{L}_i} e^{\theta_2^j \underline{L}_j} - e^{\theta_2^j \underline{L}_j} e^{\theta_1^i \underline{L}_i} \\ &\approx (\underline{I} + \theta_1^i \underline{L}_i) (\underline{I} + \theta_2^j \underline{L}_j) - (\underline{I} + \theta_2^j \underline{L}_j) (\underline{I} + \theta_1^i \underline{L}_i) \\ &= \dots \\ &= (\theta_1^i \underline{L}_i)(\theta_2^j \underline{L}_j) - (\theta_2^j \underline{L}_j)(\theta_1^i \underline{L}_i) \\ &= [\theta_1^i \underline{L}_i, \theta_2^j \underline{L}_j] \\ &= [\vec{\theta}_1 \times \vec{\theta}_2]^k \underline{L}_k. \end{aligned}$$

where the last equality follows from that same Exercise. This is only zero if both rotations are along the same axis so the cross product vanishes. From the next to last equality, we see that the matrix commutator arises naturally. The commutator of any two matrices is simply the difference of their products in both orders: $[\underline{A}, \underline{B}] = \underline{A} \underline{B} - \underline{B} \underline{A}$.

Exercise 1.7.6.

commutators of antisymmetric 3×3 matrices

a) Fill in the dots in the previous derivation by expanding the products (maintaining the order of the matrix factors!) and simplifying.

b) If we let $\theta_1^j \delta_j^1$ for $j = 1, 2$, aligning these two vectors with the standard basis vectors so that their cross product is the third basis vector, then show that this gives the cyclic commutator relations

$$[\underline{L}_i, \underline{L}_j] = \epsilon_{ijk} \underline{L}_k \quad (\text{sum over } k).$$

We could have drawn this conclusion directly from Exercise 1.2.4. ■

Notice that the commutator of the finite rotations in the limit of small rotations equals the commutator of the corresponding matrix logarithms, the corresponding Lie algebra matrices. Thus the failure of the finite rotations to commute but which instead leads to a further rotation is quantified by the commutator of the corresponding Lie algebra matrices. The commutators of any two rotation generators is again some rotation generator. The Lie algebra is said to be closed under this operation. **One of the fundamental results of group theory is that**

if a linear subspace of matrices is closed under the matrix commutator, it is the Lie algebra of a matrix subgroup whose elements near the identity matrix may be obtained by exponentiating the elements of the Lie algebra. In general a Lie algebra is a vector space with a commutator product defined on it which has all the properties of the matrix commutator, namely antisymmetry

$$[\underline{A}, \underline{B}] = \underline{A}\underline{B} - \underline{B}\underline{A} = -(\underline{B}\underline{A} - \underline{A}\underline{B}) = -[\underline{B}, \underline{A}]$$

and the so called Jacobi identity

$$\begin{aligned} & [\underline{A}, [\underline{B}, \underline{C}]] + [\underline{B}, [\underline{C}, \underline{A}]] + [\underline{C}, [\underline{A}, \underline{B}]] \\ &= \underline{A}(\underline{B}\underline{C} - \underline{C}\underline{B}) - (\underline{B}\underline{C} - \underline{C}\underline{B})\underline{A} \\ &+ \underline{B}(\underline{C}\underline{A} - \underline{A}\underline{C}) - (\underline{C}\underline{A} - \underline{A}\underline{C})\underline{B} \\ &+ \underline{C}(\underline{A}\underline{B} - \underline{B}\underline{A}) - (\underline{A}\underline{B} - \underline{B}\underline{A})\underline{C} \\ &= \dots = \underline{0} \end{aligned}$$

since the 12 terms cancel in pairs due to the associativity of the matrix product. This square bracket operation is called the Lie bracket of the Lie algebra elements. This is our first association of the square bracket delimiters with antisymmetry—we will see this notation in a more general context in the next chapter.

Exercise 1.7.7.

commutator of small rotations

Fill in the dots in the previous derivation by expanding the products (maintaining the order of the matrix factors!) and simplifying. ■

Exercise 1.7.8.

matrix Lie algebra commutators

If we have a basis $\{\underline{E}_a\}$ ($a, b = 1 \dots r$) of an r -dimensional matrix Lie algebra, then we can expand the components of the commutator in terms of this basis since by definition the commutator is a closed operation on such a Lie algebra

$$[\underline{E}_a, \underline{E}_b] = C^c_{ab} \underline{E}_c.$$

The coefficients $C^a_{bc} = -C^a_{cb}$ are not only antisymmetric by definition but satisfy the component form of the Jacobi identity

$$C^d_{ea} C^e_{bc} + C^d_{eb} C^e_{ca} + C^d_{ec} C^e_{ab} = 0.$$

These constants are called the structure constants of the group, or the components of the structure constant tensor C of the Lie algebra, a $\binom{1}{2}$ -tensor and completely determine the

local properties of the matrix group itself near the identity matrix. Under a change of basis $\underline{E}_{a'} = \underline{E}_b A^{-1b}_c$ it transforms as

$$C^{a'}_{b'c'} = A^a_p C^p_{mn} A^{-1m}_b A^{-1n}_c.$$

a) Verify the above cyclic quadratic identity satisfied by these constants by evaluating the Jacobi identity on the triplet $\underline{E}_a, \underline{E}_b, \underline{E}_c$.

b) By defining the matrices $(\underline{k}_a)^b_c = C^b_{ac}$, show that the Jacobi identity can be rewritten in matrix form as

$$[\underline{k}_a, \underline{k}_b] = C^c_{ab} \underline{k}_c.$$

This means that the span of these matrices (they are not necessarily linearly independent, and in fact vanish for Abelian groups) generates another matrix group called the adjoint group of the original Lie algebra. Note that if the matrices \underline{k}_a are linearly independent, this says that in turn that the adjoint group of the Lie algebra adjoint group is the same matrix group, since it has the same structure constant tensor.

c) Define the linear transformation $\text{ad}(\underline{X})$ of the Lie algebra into itself by

$$\text{ad}(\underline{X}) \underline{Y} = [\underline{X}, \underline{Y}].$$

Show that the matrix of $\text{ad}(\underline{X})$ for $\underline{X} = X^a \underline{E}_a$ is

$$\text{ad}(\underline{X}) = X^a \underline{k}_a.$$

This is called the adjoint action of the matrix Lie algebra on itself.

d) Defining $\underline{Z}(\lambda) = \text{AD}(e^{\lambda \underline{X}}) \underline{Y} = e^{\lambda \underline{X}} \underline{Y} e^{-\lambda \underline{X}}$ show successively that

$$\begin{aligned} \underline{Z}'(\lambda) &= \text{AD}(e^{\lambda \underline{X}}) \text{ad}(\underline{X}) \underline{Y}, \\ \underline{Z}^{(n)}(\lambda) &= \text{AD}(e^{\lambda \underline{X}}) \text{ad}(\underline{X})^n \underline{Y}, \\ \underline{Z}^{(n)}(0) &= \text{ad}(\underline{X})^n \underline{Y}, \\ \underline{Z}(\lambda) &= \text{AD}(e^{\lambda \underline{X}}) \underline{Y} = e^{\lambda \text{ad}(\underline{X})} \underline{Y}, \end{aligned}$$

using a power series representation of the matrix-valued function.

e) Express this in terms of the basis to show that

$$\text{AD}(e^{\lambda \underline{X}}) \underline{Y} = \underline{E}_a (e^{\lambda X^a \underline{k}_a})^a_b Y^b, \quad [\text{Ad}(e^{\lambda \underline{X}})]^a_b = (e^{\lambda X^a \underline{k}_a})^a_b.$$

When the matrix group acts on its own Lie algebra by conjugation, it leads to the action of the “linear adjoint matrix group” $\text{Ad}(G)$ generated by the matrices $\{\underline{k}_a\}$ according to this result we have just demonstrated.

f) **Remark.**

Recall that the standard basis of the Lie algebra $so(3, \mathbb{R})$ is $(\underline{L}_k)^i_j = \epsilon_{ikj}$, but the commutators are $[\underline{L}_i, \underline{L}_j] = \epsilon_{ijk} \underline{L}_k$, so

$$(\underline{k}_i)^k_j = C^k_{ij} = \epsilon_{ijk} = \epsilon_{kij} = (\underline{L}_i)^k_j.$$

In other words the adjoint matrices of the rotation group are again the same matrices, so the linear adjoint group is just the same rotation group, with the identity representation.



Exercise 1.7.9.

exponential parametrizations of matrix groups

For any matrix group G one can introduce a local parametrization of the matrices which belong to the group using the matrix exponential. Near the identity matrix this exponential is a 1-1 map from the matrix Lie algebra to the group manifold within $GL(n, \mathbb{R})$. This allows a coordinate system in the Lie algebra based on a basis $\{\underline{E}_a\}$ of that vector space to be exponentiated to a coordinate system on the group to “parametrize” the elements of this matrix group by the parameters θ^a

$$\theta^a \underline{E}_a \rightarrow \exp(\theta^a \underline{E}_a).$$

In other words where this exponential map is 1-1 near the identity matrix, one can invert it with the matrix logarithm so that for a point in the matrix group, we define coordinate functions θ^a by assigning the coordinates of the logarithm of that matrix with respect to the basis of the Lie algebra, so that the “parameters” can be turned around into actual functions on that region of the matrix group. This maps the origin of the Lie algebra to the identity matrix. These are called canonical coordinates of the first kind, as opposed to those of the second kind which are a product of the individual exponentials of each of the basis matrices in turn

$$e^{\theta^1 \underline{E}_1} \dots e^{\theta^r \underline{E}_r}.$$

In Exercise 1.7.1 we saw how to sum the series representing the matrix exponential of a multiple of a single matrix using an iteration formula for the powers of that matrix, in order to evaluate an exponential parametrization for the 1-dimensional matrix groups of rotations and boosts in the plane. Now we generalize that approach to the matrix exponential of a linear combination of matrices for rotations in space.

a) For the rotation group $SO(3, \mathbb{R})$ with the standard basis of its Lie algebra of antisymmetric matrices $[\underline{L}_a]^b_c = \epsilon_{bac}$, we can easily evaluate the exponential map by first showing with a computer algebra system the following identity for a unit vector $\delta_{ab} n^a n^b = 1$

$$(n^c \underline{L}_c)^3 = -n^c \underline{L}_c.$$

Note as well that

$$(n^c \underline{L}_c)^2 = \underline{I} - (n^a n_b) = (\delta^a_b - n^a n_b),$$

where the second term is the square matrix with those components, namely $\underline{n} \underline{n}^T$, which as explored in Exercise 1.6.10 is the projection along the direction n^a (the axis of the rotation), while the whole expression is the projection into the plane perpendicular to that direction, which is the plane where the rotation takes place

$$(\delta^a_b - n^a n_b) X^b = [X - (X \cdot \hat{n}) \hat{n}]^a.$$

It acts as the identity transformation in this plane.

b) Now consider the exponential power series

$$e^{\theta^c \underline{L}_c} = e^{\theta n^c \underline{L}_c} = \underline{I} + \theta n^c \underline{L}_c + \frac{1}{2!}(\theta n^c \underline{L}_c)^2 + \frac{1}{3!}(\theta n^c \underline{L}_c)^3 + \dots$$

After the first three terms show that the above identity can simplify each of the odd terms to a multiple of $n^c \underline{L}_c$ from which the odd series of the sine function is recognizable, while the even terms after the first can be recognized as the even series of the cosine function minus its first term 1, so that one obtains the identity

$$\begin{aligned} e^{\theta n^c \underline{L}_c} &= \underline{I} + \sin \theta n^c \underline{L}_c + (\cos \theta - 1)(n^c \underline{L}_c)^2 \\ &= \underbrace{\underline{I} - (n^c \underline{L}_c)^2}_{\text{axis identity}} + \sin \theta n^c \underline{L}_c + \cos \theta \underbrace{(n^c \underline{L}_c)^2}_{\text{2-identity}}. \end{aligned}$$

The first term with an underbrace represents the identity along the axis of the rotation where vectors are not changed by the rotation, while the second underbrace term represents the identity on the plane of the rotation orthogonal to the axis.

c) Show that if $n^a = \delta^a_3$ this reduces to the usual matrix of a rotation by the angle θ in the x - y plane.

d) Check out Rodriguez' rotation formula in Wikipedia to see how this formula can be used to rotate another vector, remembering Exercise 1.2.4 which showed that $n^a \underline{L}_a x = \hat{n} \times x$. ■

Exercise 1.7.10.

rotations and rotating bodies

The rotation group becomes a bit complicated when we try to parametrize all possible rotations by successive rotations about various axes, the angles of which are called Euler angles, but there are many choices for how to do this. The most interesting application of this is the problem of the motion of a rigid body, where the orientation of a basis of orthonormal vectors fixed in the rotating body (referred to as body-fixed axes) with respect to the standard orthonormal basis of \mathbb{R}^3 (referred to as space-fixed axes) is all that is needed to describe the orientation of the body about its center of mass. If we let $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ be the standard basis of \mathbb{R}^3 , and $\{\hat{e}_{1'}, \hat{e}_{2'}, \hat{e}_{3'}\}$ the body-fixed axes, they are related to each other by a time-dependent rotation matrix whose inverse transforms the coordinates with respect to those bases

$$\begin{aligned} \vec{x} &= x^i \hat{e}_i = x^{i'} \hat{e}_{i'}, \\ x^i &= R(\theta)^i_{j'} x^{j'} = S(\theta)^{-1i}_j x^{j'}, \quad x^{i'} = S(\theta)^{i'}_j x^j = R(\theta)^{-1i'}_j x^j, \\ \hat{e}_{i'} &= \hat{e}_j R(\theta)^j_{i'} = e_{j'} S(\theta)^{-1j}_{i'}. \end{aligned}$$

The rotation matrix $\underline{R}(\theta)$ actively rotates the old space-fixed basis by a rotation about those axes to the new axes, while its inverse matrix $\underline{S}(\theta)$ is said to accomplish the corresponding “passive” coordinate transformation between the old and new coordinate systems. Depending on whether we consider $\underline{R}(\theta)$ or $\underline{S}(\theta)$ the primary rotation, we adopt the so-called “active” and

“passive” points of view on how the group of rotation matrices act on \mathbb{R}^3 . They are related to each other by the group inverse map, which leaves the identity matrix invariant, and so is some kind of group generalization of a reflection through the origin, which is exactly what happens to the logarithms of these rotation matrices in their Lie algebra: $e^{\underline{K}} \rightarrow (e^{\underline{K}})^{-1} = e^{-\underline{K}}$. If we left multiply the active group matrix by a fixed rotation $\underline{R}(\theta) \rightarrow \underline{R}(\theta_0) \underline{R}(\theta)$, then

$$\underline{S}(\theta) \rightarrow (\underline{R}(\theta_0) \underline{R}(\theta))^{-1} = \underline{R}(\theta)^{-1} \underline{R}(\theta_0)^{-1} = \underline{S}(\theta) \underline{S}(\theta_0),$$

which corresponds to right multiplication of the passive group matrix by the fixed matrix. These are called left and right translations of the group into itself, and clearly the inverse map exchanges these two kinds of point transformations of the group into itself, thus interchanging the idea of left and right.

These left and right actions are physically distinct in this problem. If we change the body-fixed axes by a constant rotation $\underline{R}(\theta_0)$, then $\underline{R}(\theta) \rightarrow \underline{R}(\theta) \underline{R}(\theta_0)$ undergoes a right translation, but if we change the space-fixed axes by a constant rotation $\underline{R}(\theta_0)$, then $\underline{R}(\theta) \rightarrow \underline{R}(\theta_0) \underline{R}(\theta)$ undergoes a right translation. Rotations of the space-fixed axes are symmetries of the space, so these lead to conserved angular momenta of a rigid body without any applied torques. For a typical spinning top with a fixed point in which gravity is an applied vertical torque, the angular momentum about the vertical axis is still conserved. Conserved momentum components in the space-fixed axes enable one to more easily solve for the actual motion (Euler’s equations).

One can represent a general rotation matrix in terms of three successive rotations

$$\underline{S}(\theta) = e^{-\theta^2 \underline{L}_3} e^{-\theta^1 \underline{L}_1} e^{-\theta^3 \underline{L}_3} \leftrightarrow \underline{R}(\theta) = \underline{S}(\theta)^{-1} = e^{\theta^3 \underline{L}_3} e^{\theta^1 \underline{L}_1} e^{\theta^2 \underline{L}_3} \equiv \underline{R}_3(\theta^3) \underline{R}_2(\theta^1) \underline{R}_1(\theta^2).$$

The active rotations of the basis are easiest to visualize

$$\hat{e}_i \rightarrow \hat{e}_j \underline{R}_1(\theta^2)^j_i \rightarrow \hat{e}_k \underline{R}_2(\theta^1)^k_j \underline{R}_1(\theta^2)^j_i \rightarrow \hat{e}_l \underline{R}_3(\theta^3)^l_k \underline{R}_2(\theta^1)^k_j \underline{R}_1(\theta^2)^j_i = \hat{e}_l \underline{R}(\theta)^l_i,$$

which shows that they multiply from right to left as they are performed successively about the space-fixed axes.

We first rotate the first two basis vectors about the z -axis by an angle θ^3 , then perform a rotation of those axes about the x -axis by an angle θ^1 , and finally rotate those three axes about z -again by the angle θ^3 . The key angle is θ^1 which tilts the original vertical axis down the polar angle from the vertical of the new third axis, which does not change when we rotate all the vectors about that z -axis, while $\theta^3 - \pi/2$ is the azimuthal angle of the direction along the new third axis, i.e., $(\theta^1, \theta^2) = (\theta, \phi)$ are the usual physicist spherical coordinate angles of the vector $\hat{e}_{3'}$, while $\theta^2 = \psi$ makes $(\theta^1, \theta^2, \theta^3) = (\theta, \psi, \phi)$ the Goldstein *Classical Mechanics* textbook choice of Euler angles parametrizing a rotation matrix.

Note that if we consider a parametrization $\underline{R}(\theta) = e^{\theta^1 \underline{L}_a} e^{\theta^2 \underline{L}_b} e^{\theta^3 \underline{L}_c}$ for a triplet (a, b, c) of indices taken from 1, 2, 3, then since no two consecutive indices can be the same without collapsing the 3-parameter family of rotations into a 2-parameter family (why?), there are $12 = 6 \cdot 2$ different possibilities for such choices of coordinates on the rotation group. When (a, b, c) all distinct, then $\{\theta^a\}$ are called canonical coordinates of the second kind on the group, while $\underline{R}(\theta) = e^{\theta^a \underline{L}_a}$ leads instead to the simpler canonical coordinates of the first kind, merely exponentiating coordinates on the matrix Lie algebra to coordinates on the matrix Lie group.

Such coordinate systems do not always reach every matrix in the group. See Wikipedia: “Euler angles” to see all the possible variations on parametrizing a rotation matrix by three successive rotations, each one about one of the coordinate axes.

a) Using a computer algebra system, evaluate the matrix product of these three rotations $\underline{R}(\theta) = e^{\theta^1 \underline{L}_a} e^{\theta^2 \underline{L}_b} e^{\theta^3 \underline{L}_c}$ and its inverse $\underline{S}(\theta)$.

b) Evaluate

$$\underline{R}^{-1} d\underline{R} = \omega^a \underline{L}_a .$$

Read off the three differential expressions from the coefficients of \underline{L}_a .

c) Repeat for

$$d\underline{R} \underline{R}^{-1} = \tilde{\omega}^a \underline{L}_a .$$

Read off the three differential expressions from the coefficients of \underline{L}_a .

d) The body-fixed coordinates of points in the rotating rigid body are constant since the axes are fixed in the body, so the space-fixed coordinates of these points undergo a rotation, and we can compute their velocities in the space-fixed or body-fixed axes.

$$\vec{x} = \underline{R} \vec{x}' \rightarrow \frac{d\vec{x}}{dt} = \frac{d}{dt} (\underline{R} \vec{x}') = \left(\frac{d}{dt} \underline{R} \right) \underline{R}^{-1} \underline{R} \vec{x}' = \left(\frac{d}{dt} \underline{R} \right) \underline{R}^{-1} \vec{x} = \frac{\tilde{\omega}^a}{dt} \underline{L}_a \vec{x} = \frac{\vec{\tilde{\omega}}}{dt} \times \vec{x}$$

or in terms of the body-fixed coordinates

$$\underline{R}^{-1} \frac{d}{dt} (\underline{R} \vec{x}') = \underline{R}^{-1} \left(\frac{d}{dt} \underline{R} \right) \vec{x}' = \frac{\omega^a}{dt} \underline{L}_a \vec{x}' = \frac{\vec{\omega}}{dt} \times \vec{x}' .$$

Thus

$$\underline{\Omega}^a = \frac{\omega^a}{dt} = \omega^a_b \frac{d\theta^b}{dt} \underline{k}_a$$

are the body-fixed components of the angular velocity of the body, while

$$\tilde{\Omega}^a = \frac{\tilde{\omega}^a}{dt} = \tilde{\omega}^a_b \frac{d\theta^b}{dt} \underline{k}_a$$

are the space-fixed components of the angular velocity of the body. ■

This detour takes us a bit astray from our primary objective: to get an overall sense of differential geometry without getting bogged down in interesting detours. The commutator on the Lie algebra is very important and completely determines the group law of composition. One can search on “Campbell Baker Hausdorff” to find more about the deviation of a non-Abelian group from the additive group law when written in exponential form. For example, for an Abelian matrix group one has simply vector addition of the logarithms, that is, vector addition of the exponential exponent matrices

$$e^{\underline{X}} e^{\underline{Y}} = e^{\underline{X} + \underline{Y}} ,$$

but a complicated formula like a series expansion shows that for a non-Abelian group, the matrix product of two matrices in the group can be written

$$e^{\underline{X}}e^{\underline{Y}} = e^{\underline{X}+\underline{Y}+\frac{1}{2}[\underline{X},\underline{Y}]+\dots},$$

where all the higher order terms are nested commutators, and hence reduce to matrices in the Lie algebra of the matrix group since the Lie algebra is closed under the commutator operation. This is more than we want to deal with at our elementary introduction. However, in order to handle these antisymmetric properties that arise in considering linear transformations which preserve inner products as well as how we extend those inner products to measure area, volume, etc., we need to develop a few more mathematical tools that make life easier when dealing with antisymmetric linear operations.

Representations of Lie groups

When a matrix group G acts as a transformation group on a vector space V rather than a more general space, life is nicer because of the additional structure of the points being moved around, provided that action respects their nature as vectors. This just means that the matrix group has to act on the vectors as linear transformations, so the action is governed by a map $\rho : G \rightarrow GL(V)$ into the general linear group of the vector space which is a homomorphism, i.e., respects the group law

$$\rho(\underline{A}) \circ \rho(\underline{B}) = \rho(\underline{A} \underline{B}).$$

For example, every matrix group acts on its own Lie algebra through conjugation which is called the adjoint representation of the group: $\underline{A} \mapsto Ad(\underline{A})$. The derivative of the map Σ at the identity matrix leads to a corresponding representation of the Lie algebra, namely a map from the Lie algebra to the Lie algebra of the group of linear transformations of the vector space on which the matrix group acts. The adjoint representation of the Lie algebra is $\underline{X} \mapsto ad(\underline{X})$.

The general linear group acts on all the tensor spaces above \mathbb{R}^n as a linear transformation group through the transformation laws for the components of tensors under a change of basis $e_i \mapsto e_{i'} = e_j A^{-1j}_i$. These are called the tensor representations. For the $\binom{p}{q}$ -tensor introduce the notation

$$\begin{aligned} T^{i\dots}_{j\dots} &\mapsto [\rho^{(p,q)}(\underline{A})T]^{i\dots}_{j\dots} \\ &= A^i_m \dots A^{-1n}_j \dots T^{m\dots}_{n\dots}. \end{aligned}$$

By construction this map $\rho^{(p,q)}$ for successive changes of basis satisfies the key composition condition that a change of basis by \underline{A} followed by a change of basis by \underline{B} is equivalent to a direct change of basis by $\underline{A} \underline{B}$.

Now consider a curve $\underline{A}(\lambda)$ through the identity matrix $\underline{A}(0) = \underline{I}$ with tangent vector

$\underline{B} = \underline{A}'(0)$ and evaluate its derivative there

$$\begin{aligned}
 \left. \frac{d}{d\lambda} \right|_{\lambda=0} [\rho^{(p,q)}(\underline{A}(\lambda))T]^{i\cdots}_{j\cdots} &= \left(\left. \frac{d}{d\lambda} A^i_m \right|_{\lambda=0} \cdots A^{-1n}_j \cdots T^{m\cdots}_{n\cdots} + \cdots \right. \\
 &\quad \left. + A^i_m \cdots \left(\left. \frac{d}{d\lambda} A^{-1n}_j \right|_{\lambda=0} \cdots T^{m\cdots}_{n\cdots} + \cdots \right) \right. \\
 &= B^i_m \cdots T^{m\cdots}_{j\cdots} + \cdots - B^n_j \cdots T^{i\cdots}_{n\cdots} - \cdots \\
 &\equiv [\sigma^{(p,q)}(\underline{B})T]^{i\cdots}_{j\cdots}.
 \end{aligned}$$

Here we have used the fact that $\underline{A}^{-1'}(0) = \underline{I}$ and that from Exercise 1.7.2 where we showed that $d\underline{A}^{-1} = -\underline{A}^{-1}d\underline{A}\underline{A}^{-1}$, it follows that $\underline{A}^{-1'}(0) = -\underline{B}$.

Exercise 1.7.11.

tensor representation Lie algebra

It seems clear that $\sigma^{(p,q)}$ is a linear map from the Lie algebra of the general linear group to the Lie algebra of the linear transformations of the space of $\binom{p}{q}$ -tensors. Convince yourself that this satisfies the Lie algebra compatibility condition

$$[\sigma^{(p,q)}(\underline{A}) \circ \sigma^{(p,q)}(\underline{B}) - \sigma^{(p,q)}(\underline{B}) \circ \sigma^{(p,q)}(\underline{A})]T = \sigma^{(p,q)}([\underline{A}, \underline{B}])T.$$

(Examine the case with only 1 contravariant index.)

■

This notation will prove useful in Part 2 when we start differentiating tensor fields in various ways. For now we can use this relationship to understand the idea of spin.

Exercise 1.7.12.

spin and rotations

Consider the rotation group $SO(3, \mathbb{R})$ and its Lie algebra $so(3, \mathbb{R})$ with basis \underline{L}_a acting on \mathbb{R}^3 and its tensor spaces. For each space of $\binom{p}{q}$ -tensors, define the linear maps (spin operator along a -axis)

$$T \mapsto \sigma^{(p,q)}(\underline{L}_a)T \equiv S_a T,$$

and their quadratic combination (total spin operator)

$$T \mapsto S^2 T \equiv \delta^{ab} S_a S_b T = \delta^{ab} \sigma^{(p,q)}(\underline{L}_a) \sigma^{(p,q)}(\underline{L}_b) T.$$

When a tensor satisfies $S^2 T = -s(s+1)T$ it is said to have spin s .

a) Consider the action of the rotation group on \mathbb{R}^3 , the so called identity representation $\sigma^{(1,0)}$ since it acts as a subgroup of the group of linear transformations of \mathbb{R}^3 into itself. The elements of \mathbb{R}^3 are contravariant vectors, or $\binom{1}{0}$ -tensors T with components T^i . The rotation group simply multiplies its components by matrix multiplication $[\rho^{(1,0)}(\underline{A})T]^i = A^i_j T^j$. Here one has $[S_a T]^i = [\sigma^{(1,0)}(\underline{L}_a)T]^i = [\underline{L}_a \underline{T}]^i$, the identity representation of its Lie algebra. Show

that $\delta^{ab} \underline{L}_a \underline{L}_b = -2\underline{I}$, so $S^2 T = -1(1+1)T$ for every vector identically, i.e., vectors have spin 1.

b) Consider the action of the rotation group on $\binom{0}{2}$ -tensors over \mathbb{R}^3 . Any such tensor has a matrix of components which can be decomposed into its symmetric and antisymmetric parts first, then its symmetric part can be decomposed into its tracefree part and its pure trace part (a multiple of the identity matrix)

$$\begin{aligned} T_{ij} &= T_{(ij)} + T_{[ij]} \\ &= \underbrace{\frac{1}{3} \delta^{mn} T_{mn} \delta_{ij}}_{T_{\text{trace}}} + \underbrace{\left[T_{(ij)} - \frac{1}{3} \delta^{mn} T_{mn} \delta_{ij} \right]}_{T_{\text{sym,tracefree}}} + \underbrace{T_{[ij]}}_{T_{\text{antisym}}} . \end{aligned}$$

Use a computer algebra system to show by explicit multiplication that

$$S^2 T_{\text{trace}} = 0, \quad S^2 T_{\text{antisym}} = -1(1+1)T, \quad S^2 T_{\text{sym,tracefree}} = -2(2+1)T,$$

or $S^2 T = -s(s+1)T$ for $s = 0, 1, 2$. These parts of the total tensor are referred to as its spin s parts. Note that any multiple of δ_{ij} , a pure trace tensor, has spin 0. An antisymmetric second rank tensor has spin 1, and in fact can be represented in terms of a vector $T_{[ij]} = \epsilon_{ijk} T^k$ as explored in Exercise 1.2.4.

Maple hint:

$$\begin{aligned} (\underline{S_i T})_{mn} &= -(S_i)^k{}_m T_{kn} - T_{mk} (S_i)^k{}_n = (-\underline{S_i}^T \underline{T} - \underline{T} \underline{S_i})_{mn} \\ &= (\underline{S_i} \underline{T} - \underline{T} \underline{S_i})_{mn} \end{aligned}$$

which is the same result as for the mixed tensor matrix $(T^m{}_n)$, since index position does not matter in an orthonormal frame. Thus one can use matrix multiplication to iterate this

$$\underline{S^2 T} = \underline{S^i S_i T} = \underline{S_1 (S_1 T)} + \underline{S_2 (S_2 T)} + \underline{S_3 (S_3 T)}$$

Then make a generic matrix: `M:=Matrix(3,3,symbol=T)`,
and define its pure trace part: `Mtrace:= $\frac{1}{3}$ Trace(M) IdentityMatrix(3,3)`,
its tracefree symmetric part: `Msymtf:= $\frac{1}{2}$ (M+Transpose(M))`,
and antisymmetric part: `Masym:= $\frac{1}{2}$ (M-Transpose(M))`.
Then evaluate S^2 Mtrace, S^2 Masym+2 Masym, and S^2 Msymtf+6 Msymtf
to show they are all zero.

c) The property $[S^2 \delta]_{ij} = 0$ is a consequence of the fact that δ_{ij} is invariant under rotations. Since index raising and lowering is also invariant under rotations, then $\delta^i{}_j$ should have $[S^2 \delta]^i{}_j = 0$ [in fact $S_i \underline{I} = \underline{S_i} \underline{I} - \underline{I} \underline{S_i} = 0$], and therefore $\delta^{ij}{}_{mn} \equiv \delta^i{}_m \delta^j{}_n - \delta^i{}_n \delta^j{}_m$ should have $S^2 \delta^{(2)} = 0$ (we will study this tensor in the next chapter). Show that applying this operator S^2 to the tensor with components ϵ_{ijk} which is invariant under rotations also yields 0. Any multiples of these tensors are said to have spin 0.

d) Define the Pauli matrices

$$\underline{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \underline{\sigma}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \underline{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The basis $\underline{S}_a = \underline{E}_a = \frac{i}{2}\underline{\sigma}_a$ of the Lie algebra $su(2)$ of the special unitary group $SU(2)$ has the same commutation relations as the basis \underline{L}_a of $SO(3, \mathbb{R})$ given in Exercise 1.7.6

$$[\underline{S}_a, \underline{S}_b] = \epsilon_{abc}\underline{S}_c \quad (\text{sum over } c).$$

Verify this with a computer algebra system. $SU(2)$ and its Lie algebra act on 2-complex-dimensional vector space \mathbb{C}^2 of pairs of complex numbers as its identity representation. Show that $\delta^{ab}\underline{S}_a\underline{S}_b = -\frac{1}{2}(\frac{1}{2} + 1)\underline{I}$. Thus the elements of \mathbb{C}^2 have total spin $\frac{1}{2}$, if we extend the notion of total spin to this group, whose adjoint representation is the rotation group. These elements are called spinors, and play a fundamental role in quantum mechanics, needed to describe the class of elementary particles called fermions, which have spin $1/2$. We will study this group in Chapter 4 when the meaning of unitary will be explained.

e) We can exponentiate the matrices of this Lie algebra as easily as we did the rotations and pseudorotations of the real plane, except compared to the real rotations of the plane, here there is an extra factor of $1/2$ that finds its way into the cosine and sine of the angle of rotation, making dramatically clear what half integer spin means compared to the rotations of a real vector (spin 1). If $\theta^i = \theta\hat{n}^i$, where \hat{n}^i is a unit vector, show that

$$(\theta^i\underline{S}_i)^2 = -\frac{\theta^2}{4}\underline{I}, \quad (\theta^i\underline{S}_i)^3 = -\frac{\theta^3}{8}2\hat{n}^i\underline{S}_i,$$

[hint: only the symmetric terms contribute to the sum: $\theta^i\theta^j\underline{S}_i\underline{S}_j = \sum_{i=1}^3(\theta^i)^2\underline{S}_i^2$, why?] so that separating the even and odd powers of the power series leads to

$$e^{\theta^i\underline{S}_i} = \cos\left(\frac{\theta}{2}\right)\underline{I} + i\sin\left(\frac{\theta}{2}\right)\underline{\sigma}_i.$$

Thus a rotation by of a vector by an angle θ corresponds to a rotation of a spinor by $\theta/2$, so a spinor has to be revolved two revolutions to return to its original state compared to a vector. When one similarly examines the action of a rotation on a tracefree symmetric tensor (spin 2) under such a rotation one instead finds cosines and sines of 2θ . This is the meaning of this positive half-integer/integer spin parameter s .

■

Remark.

The topic we have to sweep under the rug is that of group representation theory which turns out to be very interesting and underlies how we classify electron states in atoms as well as electromagnetic radiation and other states whose description involves directional information parametrized by the unit sphere. But why not a few lines explanation.

Vector spaces may be decomposed into direct sums of subspaces so that a vector can be uniquely represented as a sum of vectors, one in each of those subspaces. Consider the simplest case of two subspaces: $V = V_1 \oplus V_2$, so every vector is of the form $v = v_1 + v_2$, $v_1 \in V_1, v_2 \in V_2$. Projections are defined for each subspace: $P_1(v) = v_1$, $P_2(v) = v_2$, so clearly $P_1 \circ P_1 = P_1$, $P_2 \circ P_2 = P_2$, $P_1 \circ P_2 = P_2 \circ P_1 = 0$ (the zero vector, the only vector which belongs to both subspaces). When there is an inner product on V such that these subspaces are orthogonal,

these are called orthogonal projections, and we are familiar with these from \mathbb{R}^3 when we project vectors along a line through the origin (1-dimensional subspace) and orthogonal to it, to a plane through the origin (2-dimensional subspace). In fact we can project along all three coordinate axes to decompose a vector into 3 vector components, one along each axis. This is what we do when we express a vector in any vector space in terms of any basis: $v = v^a e_a$, decomposing it into n vectors each of which is a multiple of a basis vector. This is an orthogonal decomposition when the basis itself is orthogonal.

When a matrix group acts on a vector space through a representation, one can often decompose that vector space into a series of subspaces, each of which alone is a representation of the group, thus simplifying how one sees the action of the group on that space. When a representation can no longer be decomposed further into subspace representations, it is called “irreducible.” For the rotation group acting on $\binom{1}{1}$ -tensors over \mathbb{R}^3 , we saw that their component matrices decomposed into three separate orthogonal subspaces with respect to the trace inner product: the pure trace matrices (multiples of the identity matrix, $s = 0$), the tracefree symmetric matrices ($s = 2$), and the antisymmetric matrices ($s = 1$), which we classified through the integer s parametrizing the eigenvalues $-s(s + 1)$ of the total angular momentum matrix operator L^2 for that representation. These matrices can be further classified by the eigenvalues of the matrix L_3 in that representation, which amounts to choosing a new basis of these spaces of matrices which behave in a certain way under rotations. Each subspace of a given spin s is an irreducible representation of the rotation group.

Later, when we consider how functions on \mathbb{R}^3 behave under rotations, we could (but won’t here) do something similar on the infinite-dimensional vector space of functions, which can be decomposed into an infinite direct sum of vector subspaces of functions which are characterized by all the integer and half-integer values of the orbital angular momentum parameter ℓ corresponding to s in the finite-dimensional case. A basis can be adapted to their behavior under rotations. This leads to the spherical harmonics, which are used to describe functions on the unit sphere which behave in simple ways under rotations. These ideas are fundamental for quantum mechanics and electromagnetic theory.

▲

Exercise 1.7.13.

a) Consider the set of all matrices of the form

$$\underline{A} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = a\underline{E}_0 + b\underline{E}_1,$$

where

$$\underline{E}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \underline{E}_1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Use a computer algebra system to show that this set is closed under matrix multiplication. Show that only the zero matrix in this set is not invertible (has zero determinant), so that the 2-parameter family of nonzero matrices of this form is actually a subgroup of the general linear

group in 2 dimensions. Show that this group is abelian, i.e., the order of the factors in the matrix product does not matter.

b) Use a computer algebra system to exponentiate a general matrix in the Lie algebra of this group (which is the entire matrix group plus the zero matrix). We can already foresee the result from the factoring of the exponential to a scaling by the exponential factor of the rotation matrix

$$e^{\theta^0 \underline{E}_0 + \theta^1 \underline{E}_1} = e^{\theta^0 \underline{E}_0} e^{\theta^1 \underline{E}_1}.$$

c) Note that if we compare the multiplication properties of 1 and i in the complex plane with \underline{E}_0 and \underline{E}_1 , we find the same relations. In fact the map

$$a\underline{E}_0 + b\underline{E}_1 \mapsto a + ib$$

is a faithful real representation of the complex number field which incorporates the rotation and conformal rescaling properties of the real plane into the algebra of the complex numbers. The special subgroup of unit determinant matrices corresponds to the exponentials of the tracefree Lie subalgebra, i.e., to the exponentials of \underline{E}_1 . Notice that complex conjugation corresponds to the matrix transpose.

e) Clearly one can take the identity matrix and any other matrix and form a 2-dimensional abelian Lie subalgebra whose exponential will give an Abelian subgroup of the general linear group. Picking the second matrix to be antisymmetric, or symmetric, or to have only 1 nonvanishing entry leads to rotations, boosts and null rotations of the plane. Let

$$\underline{E}_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Show that this group combines multiplication in one group parameter with addition in the other. ■

Chapter 2

Symmetry properties of tensors

We have already seen that symmetric $\binom{0}{2}$ -tensors are associated with the important notion of inner products from which the ideas of lengths and angles spring forth. We will see next that antisymmetric tensors of any rank, characterized by a change in sign of their components under a permutation of the index values as is the case for an antisymmetric matrix, are connected with the notion of area and volume as well as orientation (directional information) of planes and higher dimensional subspaces of \mathbb{R}^n . Antisymmetric $\binom{0}{2}$ -tensors turn out to be fundamental to the symmetry groups of inner products as well, and so have tremendous importance in the differential geometry of curvature.

2.1 Measure motivation and determinants

The determinant of a matrix $\underline{A} = \langle \underline{u}_{(1)} | \cdots | \underline{u}_{(n)} \rangle$ whose columns are the column matrices corresponding to vectors $u_{(i)} \in \mathbb{R}^n$ may be thought of as an antisymmetric multilinear real-valued function of n vector arguments, i.e., a $\binom{0}{n}$ -tensor on \mathbb{R}^n

$$\det \underline{A} = \det(\langle \underline{u}_{(1)} | \cdots | \underline{u}_{(n)} \rangle) \equiv \det(u_{(1)}, \cdots, u_{(n)}).$$

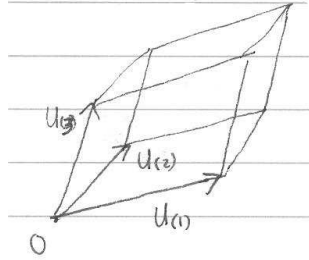


Figure 2.1: 3 vectors in \mathbb{R}^3 form a parallelepiped whose volume is the absolute value of the determinant of the matrix in which they are either rows or columns.

In multivariable calculus we learn that the triple scalar product of three linearly independent vectors in \mathbb{R}^3 is evaluated as the determinant of the matrix whose rows are those three vectors (the same as if they are the columns since the transpose does not change the determinant)

$$(u_{(1)} \times u_{(2)}) \cdot u_{(3)} = \det \langle \underline{u}_{(1)}^T, \underline{u}_{(2)}^T, \underline{u}_{(3)}^T \rangle = \det \langle \underline{u}_{(1)} | \underline{u}_{(2)} | \underline{u}_{(3)} \rangle$$

and that its value up to sign is the volume of the parallelepiped they determine. The sign itself is positive when the three vectors are ordered so that the third vector is on the side of the plane of the first two vectors determined by the right hand rule ($u_{(1)} \times u_{(2)}$ points on this side of the plane), since then its component along the cross-product of the first two will be positive. Such an ordered set of vectors is referred to as a right-handed set, like the standard basis of \mathbb{R}^3 . When the sign is negative, they are then a left handed set.

The length of the cross-product of two vectors alone is interpreted as the area of the parallelogram that they determine. For two vectors in the x - y plane within \mathbb{R}^3

$$a \times b = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & 0 \\ b_1 & b_2 & 0 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \mathbf{k}.$$

The magnitude of the cross product is just the absolute value of the 2x2 determinant formed by these two vectors, i.e., the area of the parallelogram they form, while the sign is positive if the right hand rule moving from the first to the second by a counterclockwise rotation by less than 180 degrees aligns the thumb with the positive z axis. Such an ordered pair of vectors is called right handed, like the standard basis is, and otherwise left handed.

This generalizes to \mathbb{R}^n , namely

$$\text{Vol}(u_{(1)}, \dots, u_{(n)}) = |\det(u_{(1)}, \dots, u_{(n)})|$$

has the interpretation as the volume of the n -parallelepiped formed with these vectors as the edges from the corner at the origin, while the sign of $\det(u_{(1)}, \dots, u_{(n)})$ indicates whether or not the vectors have the same “orientation” as the standard basis. The standard basis is said to have the positive orientation, while a basis with the opposite sign of the determinant is said to be negatively oriented. We’ll return to the orientation later after concentrating first on the area/volume/measure properties of the determinant.

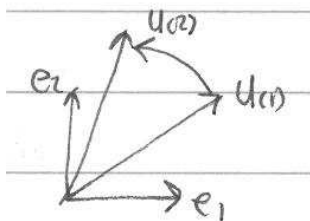


Figure 2.2: An orientation of the plane is a choice of clockwise (left handed, negative orientation) or counterclockwise (right handed, positive orientation) motion from the first to the second vector in an ordered set by an angle less than π . Here $\{u_{(1)}, u_{(2)}\}$ are positively oriented, like the standard basis.

Our notion of measure ($n = 1$: length, $n = 2$: area, $n = 3$: volume, etc.) given a way to measure lengths in 1 dimension generalizes to higher dimensional rectangular objects as the product of the lengths of the mutually orthogonal edges: area = length times width, 3-volume = area times height, 4-volume = 3-volume times height in the 4th direction, etc. For nonrectangular objects like parallelograms and parallelepipeds, only the height of the last edge or face of the object from the base matters as one adds the last dimension.

The connection of this basic notion of measure/volume with determinants comes from the way in which the following three elementary column operation properties characterize volume, modulo signs, which is the content of the fourth property

1. Invariance under adding multiples of other columns:

$$\det(u_{(1)}, \dots, \underbrace{u_{(i)} + au_{(j)}}_{\text{ith argument}}, \dots, u_{(j)}, \dots, u_{(n)}) = \det(u_{(1)}, \dots, u_{(i)}, \dots, u_{(j)}, \dots, u_{(n)})$$

[Iteration of this property leads to the property that adding any linear combination of the other vectors to a given vector in the determinant does not change its value.]

2. Scalar multiple factor from columns:

$$\det(u_{(1)}, \dots, au_{(i)}, \dots, u_{(n)}) = a \det(u_{(1)}, \dots, u_{(i)}, \dots, u_{(n)}).$$

3. Unit value on identity matrix:

$\det(e_1, \dots, e_n) = 1$, true for any orthonormal basis with the same orientation as the standard basis.

4. Changes sign when any two columns are swapped (interchanged in position):

$$\det(u_{(1)}, \dots, u_{(i)}, \dots, u_{(j)}, \dots, u_{(n)}) = -\det(u_{(1)}, \dots, u_{(j)}, \dots, u_{(i)}, \dots, u_{(n)}).$$

Properties (1) and (2) are independent of the Euclidean inner product and remain valid in defining volume with respect to any inner product. Property (3) basically fixes the scale of the volume function in terms of the choice of inner product, in this case, the Euclidean one, by assigning its values to be 1 or -1 on the set of all orthonormal bases of \mathbb{R}^n . The final property helps us keep track of some minimal information about the ordering of the vectors in the determinant.

Recall that we define the volume of a rectangular solid with perpendicular edges to be the product of the lengths of its orthogonal edges from any corner. This is then extended to a n -parallelepiped by that noticing one can always chop it up and re-assemble into a rectangular solid with the same volume. In the plane, for example, this property of volume is exactly equivalent to property (1). In figure 2.3 one has to add a multiple of the base vector to the second vector to make it orthogonal to the base yet have the same height, and in so doing, move the triangular piece from one side of the parallelogram to the other to form a rectangle with the same area.

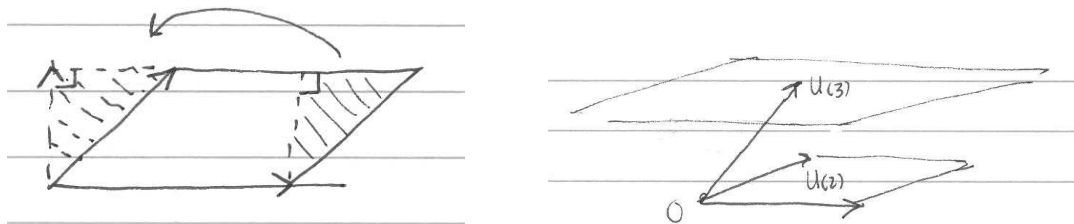


Figure 2.3: A parallelogram has the same area as the rectangle with the same height and base. Similarly a parallelepiped in \mathbb{R}^3 has the same volume as long as the height and base area remain the same. In each case we are free to add any linear combination of the previous vectors to the tip of the last vector without changing the height and thus the measure of the figure.

Correspondingly in \mathbb{R}^3 , as shown again in Fig. 2.3 we can move $u_{(3)}$ around anywhere in the plane through its tip parallel to the plane of $u_{(1)}$ and $u_{(2)}$ without changing the “height” relative to that plane and therefore not changing the volume of parallelepiped. In particular we can always move $u_{(3)}$ so that it is perpendicular to the plane of $u_{(1)}$ and $u_{(2)}$. Then by adding multiples of $u_{(1)}$ to $u_{(2)}$ we can make $u_{(2)}$ perpendicular to $u_{(1)}$ resulting in a rectangular solid. Such an iterative process can be used in \mathbb{R}^n to reduce any n -tuple of vectors to an orthogonal n -tuple with the same volume (closely related to the so called Gram-Schmidt process of orthogonalizing a set of vectors). Property (2) allows us to pull out the factors of the lengths

of the orthogonal edges, leaving the scale of the volume to be set by condition (3), that an orthonormal set of vectors has unit volume.

Property (1) is a direct result of the antisymmetry of the determinant, since antisymmetrization in a pair of identical or proportional objects always gives zero. It determines an equivalence relation on the ordered n -tuples of vectors which corresponds to having the same volume. Any orthogonal representative of such an equivalence class (i.e., any set of orthogonal vectors in it) then sets the value of the volume through properties (2) and (3) together. We will see that antisymmetrization of vectors is somehow equivalent to establishing a volume (or more generally measure) equivalence relation, while an inner product merely sets the scale.

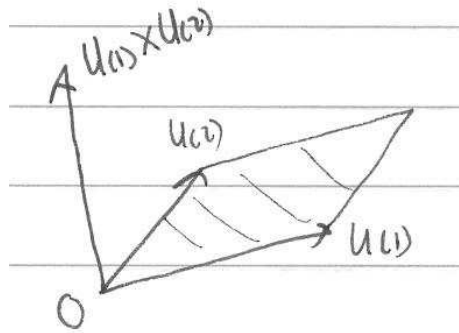


Figure 2.4: The cross product of two vectors in \mathbb{R}^3 .

We are also interested in the measure of p -dimensional objects in \mathbb{R}^n , like parallelograms in \mathbb{R}^3 . These turn out to be connected to subdeterminants. For example, an ordered pair of vectors $(u_{(1)}, u_{(2)})$ in \mathbb{R}^3 determines a parallelogram with a certain orientation in space. Consider the 3×2 matrix $\langle \underline{u}_{(1)} | \underline{u}_{(2)} \rangle$. It has three 2×2 subdeterminants obtained by eliminating the first, second, and the third rows respectively, and then alternating the sign to define the corresponding “minors” of the determinant. These define the components of the cross product of $u_{(1)}$ and $u_{(2)}$ whose magnitude gives the desired area information, and whose direction specifies the orientation of the plane of $u_{(1)}$ and $u_{(2)}$ within space as well as the relative orientation of the two vectors within their plane

$$u_{(1)} \times u_{(2)} = \left(\begin{vmatrix} u_{(1)}^2 & u_{(2)}^2 \\ u_{(1)}^3 & u_{(2)}^3 \end{vmatrix}, - \begin{vmatrix} u_{(1)}^1 & u_{(2)}^1 \\ u_{(1)}^3 & u_{(2)}^3 \end{vmatrix}, \begin{vmatrix} u_{(1)}^1 & u_{(2)}^1 \\ u_{(1)}^2 & u_{(2)}^2 \end{vmatrix} \right)$$

The sign in the second component will be explained later.

You also know that

$$\det(u_{(1)}, u_{(2)}, u_{(3)}) = (u_{(1)} \times u_{(2)}) \cdot u_{(3)},$$

i.e., the vector $u_{(1)} \times u_{(2)}$ is basically the partial evaluation of the determinant tensor leaving one vector argument free, i.e., a covector, which is then identified with a vector by the Euclidean inner product

$$\det(u_{(1)}, u_{(2)}, u) = [u_{(1)} \times u_{(2)}]^b(u).$$

To summarize, the properties of the determinant function and of “antisymmetrization” in general characterize both p -measure in \mathbb{R}^n up to a setting of scale which is accomplished via an inner product as well as the orientation of a set of ordered vectors within any subspace. We need to develop a notation that can more easily handle this kind of information.

2.2 Tensor symmetry properties

We have already used a symmetry condition for the class of inner products we have been considering, namely the order of the inner product $G(Y, X) = G(X, Y)$ of two vectors X, Y does not matter, and we have mentioned antisymmetric inner products which we will not pursue here. The determinant function, which is a multilinear function of the columns of a matrix, changes sign under swapping any two columns, so it is an antisymmetric tensor which is important for describing measure and orientation of subspaces of \mathbb{R}^n , the details of which we will learn below. Symmetry properties of tensors turn out to be extremely important for many reasons, so we need to develop a notation to handle them.

Symmetry properties involve the behavior of a tensor under the interchange of two or more arguments. Of course to even consider the value of a tensor after the permutation of some of its arguments, the arguments must be of the same type, i.e., covectors have to go in covector arguments and vectors in vector arguments and no other combinations are allowed.

The simplest case to consider are tensors with only 2 arguments of the same type. For vector arguments we have $\binom{0}{2}$ -tensors. For such a tensor T introduce the following terminology:

$$\begin{aligned} T(Y, X) &= T(X, Y), & T \text{ is symmetric in } X \text{ and } Y, \\ T(Y, X) &= -T(X, Y), & T \text{ is antisymmetric or "alternating" in } X \text{ and } Y. \end{aligned}$$

Letting $(X, Y) = (e_i, e_j)$ and using the definition of components, we get a corresponding condition on the components

$$\begin{aligned} T_{ji} &= T_{ij}, & T \text{ is symmetric in the index pair } (i, j), \\ T_{ji} &= -T_{ij}, & T \text{ is antisymmetric (alternating) in the index pair } (i, j). \end{aligned}$$

For an antisymmetric tensor, the last condition immediately implies that no index can be repeated without the corresponding component being zero

$$T_{ji} = -T_{ij} \rightarrow T_{ii} = 0.$$

Any $\binom{0}{2}$ -tensor can be decomposed into symmetric and antisymmetric parts by defining

$$\begin{aligned} [\text{SYM}(T)](X, Y) &= \frac{1}{2}[T(X, Y) + T(Y, X)], & (\text{"the symmetric part of } T\text{"}), \\ [\text{ALT}(T)](X, Y) &= \frac{1}{2}[T(X, Y) - T(Y, X)], & (\text{"the antisymmetric part of } T\text{"}), \\ T &= \text{SYM}(T) + \text{ALT}(T). \end{aligned}$$

The last equality holds since evaluating it on the pair (X, Y) immediately leads to an identity. [Check.]

Again letting $(X, Y) = (e_i, e_j)$ leads to corresponding component formulas

$$\begin{aligned} [\text{SYM}(T)]_{ij} &= \frac{1}{2}(T_{ij} + T_{ji}) \equiv T_{(ij)}, & (n(n+1)/2 \text{ independent components}), \\ [\text{ALT}(T)]_{ij} &= \frac{1}{2}(T_{ij} - T_{ji}) \equiv T_{[ij]}, & (n(n-1)/2 \text{ independent components}), \\ T_{ij} &= T_{(ij)} + T_{[ij]}, & (n^2 = n(n+1)/2 + n(n-1)/2 \text{ independent components}). \end{aligned}$$

Round brackets around a pair of indices denote the symmetrization operation, while square brackets denote antisymmetrization. This is a very convenient shorthand. All of this can be repeated for $\binom{2}{0}$ -tensors and just reflects what we already know about the symmetric and antisymmetric parts of matrices.

Exercise 2.2.1.

counting independent components

The component matrix (T_{ij}) of a $\binom{0}{2}$ -tensor or (T^{ij}) of a $\binom{2}{0}$ -tensor, with the left index i denoting the rows and the right index j the columns, has diagonal entries ($i = j$), offdiagonal entries ($i \neq j$), and upper ($i < j$) and lower ($i > j$) offdiagonal entries, and upper ($i \leq j$) and lower ($i \geq j$) triangular entries. Derive the number of independent components of a symmetric tensor, namely the number of diagonal plus upper offdiagonal entries of the corresponding matrix, i.e., the number of upper triangular entries, and an antisymmetric tensor, i.e., the number of upper offdiagonal entries of the corresponding matrix. Recall the identity $\sum_{j=1}^n j = n(n+1)/2$. ■

Exercise 2.2.2.

trace inner products and symmetry

a) Consider a tensor A whose matrix of components is

$$\underline{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

and evaluate the matrices of its symmetric and antisymmetric parts $\text{SYM}(\underline{A}) = \frac{1}{2}(\underline{A} + \underline{A}^T)$ and $\text{ALT}(\underline{A}) = \frac{1}{2}(\underline{A} - \underline{A}^T)$ and show that their sum equals the original matrix.

b) Since the trace of a matrix product obeys the cyclic property $\text{Tr } \underline{A} \underline{B} = \text{Tr } \underline{B} \underline{A} = \mathcal{G}(\underline{B}, \underline{A})$

$$\text{Tr } \underline{A} \underline{B} = A^i_j B^j_i = B^j_i A^i_j = \text{Tr } \underline{B} \underline{A},$$

this defines a symmetric bilinear function \mathcal{G} on pairs of square matrices and hence defines an inner product on the space of square matrices of a given dimension. The same statement applies to $\text{Tr } \underline{A}^T \underline{B} = \text{Tr } \underline{B}^T \underline{A} = G(\underline{B}, \underline{A})$, which defines a different inner product G . Evaluate each of these on the pair $\text{SYM}(\underline{A})$, $\text{ALT}(\underline{A})$ to show that these two matrices are orthogonal with respect to either of these two inner products. Thus the decomposition of the original matrix into its symmetric and antisymmetric parts is orthogonal with respect to both inner products.

c) What are the self-inner products of these two matrices under each such inner product? Notice their signs. Finally for the both inner products notice that the sum of the self-inner products of the two matrices equals the self-inner product of the original matrix, a Pythagorean theorem but with a sign change in the first case. This can be explored in general for arbitrary matrices, as the extended Exercise 1.6.9 of the previous chapter does, but it is not essential for us at this first pass through differential geometry.



In order to consider symmetries for more than a pair of indices we need to discuss the so called “symmetric group” S_n of permutations of the integers from 1 to n which maps the ordered integers from 1 to n to a reordering (rearrangement, permutation) of those integers

$$(1, 2, \dots, n) \mapsto (\sigma(1), \sigma(2), \dots, \sigma(n)).$$

This may also be written in a more suggestive form

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}$$

which indicates the integer $\sigma(i)$ which replaces i in a two row matrix, where the ordering of the columns clearly doesn’t matter. For example

$$\sigma \sim \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \sim \begin{pmatrix} 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix} \sim \begin{pmatrix} 3 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix}$$

all mean

$$(1, 2, 3) \mapsto (\sigma(1), \sigma(2), \sigma(3)) = (2, 3, 1).$$

The composition of two permutations

$$(1, 2, \dots, n) \mapsto (\pi \circ \sigma)(1, 2, 3, \dots, n) = \pi(\sigma(1), \dots, \sigma(n)) = (\pi(\sigma(1)), \dots, \pi(\sigma(n)))$$

is easily performed using the matrix algorithm: line up the upper rows of π below with lower row of σ on top, then erase these two rows to get the two rows of the “product” of two permutations

$$\begin{array}{lcl} \sigma & \sim & \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, & \pi & \sim & \begin{pmatrix} 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix} \sim \begin{pmatrix} 3 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix}, \\ \pi & \sim & \begin{pmatrix} 2 & 3 & 1 \\ 1 & 3 & 2 \end{pmatrix}, & & & \\ \pi \circ \sigma & \sim & \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} & & & \end{array}.$$

Fact.

Every permutation can be represented as a certain number N of transpositions (only two integers interchanged, all others fixed), the evenness or oddness of which is an invariant. One can therefore define the sign of a permutation to be

$$(-1)^N = \begin{cases} 1, & \text{even } N, \\ -1, & \text{odd } N, \end{cases}$$

and the permutations are correspondingly referred to as even or odd.

There are $n!$ permutations in S_n , half even, half odd (except for $n = 1$). For example the signs of the following permutations are

$$\begin{aligned} n = 1 : & \begin{pmatrix} 1 \\ 1 \end{pmatrix}_+, & n = 2 : & \begin{pmatrix} 12 \\ 12 \end{pmatrix}_+, \begin{pmatrix} 12 \\ 21 \end{pmatrix}_-, \\ n = 3 : & \begin{pmatrix} 123 \\ 123 \end{pmatrix}_+, \begin{pmatrix} 123 \\ 231 \end{pmatrix}_+, \begin{pmatrix} 123 \\ 312 \end{pmatrix}_+, \begin{pmatrix} 123 \\ 132 \end{pmatrix}_-, \begin{pmatrix} 123 \\ 213 \end{pmatrix}_-, \begin{pmatrix} 123 \\ 321 \end{pmatrix}_-. \end{aligned}$$

This permutation group itself has a rich structure which also has extremely important physical applications but we have enough information for our present purposes.

Exercise 2.2.3. counting transpositions

A cute extra piece of information about the sign of a permutation is that if you connect each integer in the top row by a straight line to its position in the bottom row (avoid multiple simultaneous intersections), the number of intersections of all these lines has the same parity (permutation sign) as the minimum number of transpositions necessary to obtain that permutation. Convince yourself this is true by imagining these lines as flexible strings, and one by one, begin untangling all the strings until they are all untangled, i.e., no longer crossing. ■

Suppose we have a $\binom{0}{3}$ -tensor T . For $\binom{0}{2}$ -tensors we defined the symmetric and antisymmetric parts by summing over all permutations of their arguments/indices, including the sign for the antisymmetric part, and dividing by the total number (2) of such permutations. For a $\binom{0}{3}$ -tensor the analogous definitions are

$$\begin{aligned} \left[\begin{pmatrix} \text{SYM} \\ \text{ALT} \end{pmatrix} (T) \right] (X, Y, Z) &= \frac{1}{3!} [T(X, Y, Z) + T(Y, Z, X) + T(Z, X, Y) \\ &\quad \pm T(X, Z, Y) \pm T(Y, X, Z) \pm T(Z, Y, X)], \end{aligned}$$

where the upper sign (lower sign) applies for the symmetric (antisymmetric) part. Clearly under any permutation of arguments, $\text{SYM}(T)$ is unchanged while $\text{ALT}(T)$ changes by the sign of the permutation. Letting $(X, Y, Z) = (e_i, e_j, e_k)$ leads to the component form

$$\left[\begin{pmatrix} \text{SYM} \\ \text{ALT} \end{pmatrix} (T) \right]_{ijk} = \frac{1}{3!} (T_{ijk} + T_{jki} + T_{kij} \pm T_{ikj} \pm T_{jik} \pm T_{kji}) \equiv \begin{pmatrix} T_{(ijk)} \\ T_{[ijk]} \end{pmatrix}.$$

It is important to note this crucial convention that putting rounded parentheses around a group of indices indicates the symmetrization over those indices, while square brackets indicates antisymmetrization, and a pair of vertical bars are placed around an index if it is not to be included in the group of indices being symmetrized or antisymmetrized. Thus $T_{[i|j|k]} = \frac{1}{2}(T_{ijk} - T_{kji})$ antisymmetrizes only the pair of indices (i, k) .

However, now in addition to the completely symmetric and completely antisymmetric parts of T , there are also mixed symmetries which were not possible with only two arguments

$$T = \text{SYM}(T) + \text{ALT}(T) + \cdots, \quad T_{ijk} = T_{(ijk)} + T_{[ijk]} + \cdots.$$

This question has to do with the representations of the symmetric group and belongs in a course on group theory.

A third rank tensor T is called symmetric if it equals its symmetric part, and antisymmetric if it equals its antisymmetric part

$$\begin{aligned} T = \text{SYM}(T) &\leftrightarrow T_{ijk} = T_{(ijk)}, & (\text{symmetric}) \\ T = \text{ALT}(T) &\leftrightarrow T_{ijk} = T_{[ijk]}. & (\text{antisymmetric}) \end{aligned}$$

We can extend this to totally covariant or totally contravariant tensors of any number of indices, or to a particular group of such indices on a tensor and say that the tensor is symmetric or antisymmetric in those particular indices or its corresponding arguments.

In general for a $\binom{0}{p}$ -tensor we define

$$\begin{aligned} [\text{SYM}(T)](X_{(1)}, \dots, X_{(p)}) &= \frac{1}{p!} \sum_{\sigma \in S_p} T(X_{\sigma(1)} \cdots X_{\sigma(p)}), \\ [\text{ALT}(T)](X_{(1)}, \dots, X_{(p)}) &= \frac{1}{p!} \sum_{\sigma \in S_p} (\text{sgn } \sigma) T(X_{\sigma(1)} \cdots X_{\sigma(p)}), \end{aligned}$$

where we use the parenthesis surrounded index in $\{X_{(i)}\}_{i=1, \dots, n}$ to list a set of n vectors to distinguish from the symbol X_i for the components of a single covector. Then letting $(X_{(1)}, \dots, X_{(p)}) = (e_{i_1}, \dots, e_{i_p})$ gives the component version

$$\begin{aligned} [\text{SYM}(T)]_{i_1 \dots i_p} &\equiv T_{(i_1 \dots i_p)} = \frac{1}{p!} \sum_{\sigma \in S_p} T_{i_{\sigma(1)} \dots i_{\sigma(p)}}, \\ [\text{ALT}(T)]_{i_1 \dots i_p} &\equiv T_{[i_1 \dots i_p]} = \frac{1}{p!} \sum_{\sigma \in S_p} (\text{sgn } \sigma) T_{i_{\sigma(1)} \dots i_{\sigma(p)}}. \end{aligned}$$

What we need now is a more efficient way of summing in these formulas using our index summation conventions. Remember, the sigma summation notation is something we suppressed once. We have to bury it again.

2.3 Epsilons and deltas

WARNING! FASTEN YOUR SEAT BELTS! or “epsilons and deltas, oh no!”

No fear, these are not the epsilons and deltas of advanced calculus.

The formal definition of a determinant of an $n \times n$ matrix $\underline{A} = (A^i_j)$ is a sum of products of elements, one taken from each column or row, preceded by a sign factor equal to the sign of the permutation

$$\begin{aligned} \det \underline{A} &= \sum_{\sigma \in S_p} \text{sgn}(\sigma) A^{\sigma(1)}_1 A^{\sigma(2)}_2 \dots A^{\sigma(n)}_n \quad (\text{one from each column}) \\ &= p! A^{[1}_1 A^2_2 \dots A^n]_n \\ &= \sum_{\sigma \in S_p} \text{sgn}(\sigma) A^1_{\sigma(1)} A^2_{\sigma(2)} \dots A^n_{\sigma(n)} \quad (\text{one from each row}) \\ &= p! A^1_{[1} A^2_2 \dots A^n_{n]} . \end{aligned}$$

However, the \sum -notation is bad news—we introduced the summation convention to eliminate the summation notation for contracted indices, and this is an even more complicated summation. By making some convenient definitions, we can also eliminate this explicit \sum -notation for determinants and for antisymmetrization, and which allow us to write down many associated identities that would be difficult to state in that former notation. The bracket notation is not sufficient by itself as a shorthand, since it is only an abbreviation for the corresponding sum over permutations. We can do better.

Generalized Kronecker Deltas

For $1 \leq p \leq n$ define

$$\delta^{i_1 \dots i_p}_{j_1 \dots j_p} \equiv p! \delta^{i_1}_{[j_1} \delta^{i_2}_{j_2} \dots \delta^{i_p}_{j_p]} \equiv p! \delta^{[i_1}_{j_1} \delta^{i_2}_{j_2} \dots \delta^{i_p]}_{j_p} .$$

or equivalently and more simply

$$\delta^{i_1 \dots i_n}_{j_1 \dots j_n} = \begin{cases} \text{sgn} \begin{pmatrix} i_1 & \dots & i_n \\ j_1 & \dots & j_n \end{pmatrix} & \text{if no repeated indices at either level,} \\ 0 & \text{otherwise.} \end{cases}$$

Once antisymmetrized on either the upper or lower indices, the result is automatically antisymmetric at both levels. [Check for $p = 2$ below!] Explicitly for $p = 1, 2, 3$ the antisymmetrized delta formulas are

$$\begin{aligned} p = 1 : & \quad \delta^{i_1}_{j_1} , \\ p = 2 : & \quad \delta^{i_1 i_2}_{j_1 j_2} = \delta^{i_1}_{j_1} \delta^{i_2}_{j_2} - \delta^{i_1}_{j_2} \delta^{i_2}_{j_1} \\ & \quad = \delta^{i_1}_{j_1} \delta^{i_2}_{j_2} - \delta^{i_2}_{j_1} \delta^{i_1}_{j_2} , \\ p = 3 : & \quad \delta^{i_1 i_2 i_3}_{j_1 j_2 j_3} = \delta^{i_1}_{j_1} \delta^{i_2}_{j_2} \delta^{i_3}_{j_3} + \delta^{i_1}_{j_2} \delta^{i_2}_{j_3} \delta^{i_3}_{j_1} + \delta^{i_1}_{j_3} \delta^{i_2}_{j_1} \delta^{i_3}_{j_2} \\ & \quad - \delta^{i_1}_{j_1} \delta^{i_2}_{j_3} \delta^{i_3}_{j_2} - \delta^{i_1}_{j_2} \delta^{i_2}_{j_1} \delta^{i_3}_{j_3} - \delta^{i_1}_{j_3} \delta^{i_2}_{j_2} \delta^{i_3}_{j_1} , \end{aligned}$$

In fact each generalized Kronecker delta is a determinant of a matrix whose entries are Kronecker deltas

$$\delta_{j_1 \dots j_p}^{i_1 \dots i_p} = \begin{vmatrix} \delta_{j_1}^{i_1} & \dots & \delta_{j_p}^{i_1} \\ \vdots & & \vdots \\ \delta_{j_1}^{i_p} & \dots & \delta_{j_p}^{i_p} \end{vmatrix} = \sum_{\sigma \in S_p} (\text{sgn } \sigma) \delta_{\sigma(j_1)}^{i_1} \delta_{\sigma(j_2)}^{i_2} \dots \delta_{\sigma(j_p)}^{i_p}.$$

Then

$$\frac{1}{p!} \delta_{i_1 \dots i_p}^{j_1 \dots j_p} T_{j_1 \dots j_p} = \delta^{j_1}_{[i_1} \delta^{j_2}_{i_2} \dots \delta^{j_p}_{i_p]} T_{j_1 \dots j_p} = T_{[i_1 \dots i_p]} = [\text{ALT}(T)]_{i_1 \dots i_p}$$

since each Kronecker delta contraction replaces a j -index by an i -index. Alternatively going backwards, we can think of shifting the antisymmetrization from the tensor indices to the indices of the tensor product of p Kronecker deltas

$$\begin{aligned} T_{i_1 \dots i_p} &= \delta^{j_1}_{i_1} \delta^{j_2}_{i_2} \dots \delta^{j_p}_{i_p} T_{j_1 \dots j_p}, & (\text{identity}) \\ T_{[i_1 \dots i_p]} &= \delta^{j_1}_{[i_1} \delta^{j_2}_{i_2} \dots \delta^{j_p}_{i_p]} T_{j_1 \dots j_p} & (\text{antisymmetrize over free indices}) \\ &\equiv \frac{1}{p!} \delta_{i_1 \dots i_p}^{j_1 \dots j_p} T_{j_1 \dots j_p}. & (\text{definition of generalized Kronecker delta}) \end{aligned}$$

Note that if a tensor is already antisymmetric, antisymmetrization does not change it (it is equal to its antisymmetric part), or

$$\text{ALT}(\text{ALT}(T)) = \text{ALT}(T)$$

equivalent to

$$\frac{1}{p!} \delta_{i_1 \dots i_p}^{j_1 \dots j_p} \left(\frac{1}{p!} \delta_{j_1 \dots j_p}^{k_1 \dots k_p} T_{k_1 \dots k_p} \right) = \frac{1}{p!} \delta_{i_1 \dots i_p}^{k_1 \dots k_p} T_{k_1 \dots k_p}$$

or

$$\delta_{i_1 \dots i_p}^{j_1 \dots j_p} \delta_{j_1 \dots j_p}^{k_1 \dots k_p} = p! \delta_{i_1 \dots i_p}^{k_1 \dots k_p}.$$

If a tensor is already antisymmetric, its component values change by the sign of a permutation applied to its indices

$$T_{\sigma(i_1) \dots \sigma(i_p)} = \text{sgn}(\sigma) T_{i_1 \dots i_p},$$

as long as the index values are all distinct, but if any two indices have the same value, then the component must be zero since interchanging them must change the sign of the component, but interchanging the indices does not change the component

$$T_{122} = -T_{122} \rightarrow T_{122} = 0.$$

Antisymmetrizing an antisymmetric tensor leads to a sum of $p!$ identical terms (changing the sign of the component is compensated by the change in sign of the Kronecker delta) which collapse to the original value once the factorial factor is divided out

$$T_{[i_1 \dots i_p]} = \frac{1}{p!} \delta_{i_1 \dots i_p}^{j_1 \dots j_p} T_{j_1 \dots j_p} = \frac{1}{p!} (p! T_{i_1 \dots i_p}) = T_{i_1 \dots i_p}.$$

What exactly are these generalized Kronecker deltas?

Each term in the expansion of a generalized Kronecker delta is a product of Kronecker deltas, which represent the components of certain tensor products of the Kronecker delta tensor (or unit tensor) with itself and so are themselves tensors, as is the entire sum of such terms. Thus $\delta_{j_1 \dots j_p}^{i_1 \dots i_p}$ are the components of a $\binom{p}{p}$ -tensor $\delta^{(p)}$ on our vector space V which has the same numerical values of its components in any basis

$$\delta^{(p)} = \delta_{j_1 \dots j_p}^{i_1 \dots i_p} e_{i_1} \otimes \dots \otimes e_{i_p} \otimes \omega^{j_1} \otimes \dots \otimes \omega^{j_p}.$$

Its value on p -vector arguments and p -covector arguments is just the determinant of the matrix of all possible evaluations of a covector on a vector

$$\delta^{(p)}(f^{(1)}, \dots, f^{(p)}, u_{(1)}, \dots, u_{(p)}) = |f^{(i)}(u_{(j)})|,$$

where the vertical bar delimiters are the standard notation for the determinant of the array of numbers between them.

It is useful to look explicitly at the $p = 2$ case to make this discussion more concrete. The components satisfy

$$\delta_{mn}^{ij} = \begin{vmatrix} \delta_m^i & \delta_n^i \\ \delta_m^j & \delta_n^j \end{vmatrix} = \delta_m^i \delta_n^j - \delta_n^i \delta_m^j$$

and the tensor is therefore

$$\delta^{(2)} = \delta_{mn}^{ij} e_i \otimes e_j \otimes \omega^m \otimes \omega^n = (\delta_m^i \delta_n^j - \delta_n^i \delta_m^j) e_i \otimes e_j \otimes \omega^m \otimes \omega^n = e_i \otimes e_j \otimes \omega^i \otimes \omega^j - e_i \otimes e_j \otimes \omega^j \otimes \omega^i,$$

which implies

$$\delta^{(2)}(f, g, X, Y) = f_i g_j \delta_{mn}^{ij} X^m Y^n = f_i X^i g_j Y^j - f_i Y^i g_j X^j = f(X)g(Y) - f(Y)g(X).$$

Example 2.3.1. For \mathbb{R}^3 with the dot product, this is just the scalar quadruple product of the corresponding vector fields

$$\delta^{(2)}(f, g, X, Y) = f^\sharp \cdot X g^\sharp \cdot Y - f^\sharp \cdot Y g^\sharp \cdot X = (f^\sharp \times g^\sharp) \cdot (X \times Y) = Q(f^\sharp, g^\sharp, X, Y).$$

$Q = \delta^{(2)\flat}$ is in fact the index lowered form of $\delta^{(2)}$. We will prove this identity for Q easily in a subsequent exercise. This tensor is useful for the following reason

$$\begin{aligned} Q(X, Y, X, Y) &= (X \cdot X)(Y \cdot Y) - (X \cdot Y)^2 = |X|^2 |Y|^2 - (|X||Y| \cos \theta)^2 \\ &= |X|^2 |Y|^2 (1 - \cos^2 \theta) = |X|^2 |Y|^2 \sin^2 \theta = (|X||Y| \sin \theta)^2 \\ &= |X \times Y|^2 = (X \times Y) \cdot (X \times Y) = \text{Area}(X, Y)^2. \end{aligned}$$

This is the square of the area of the parallelogram formed by the two vectors X and Y as sides, which is the interpretation of the magnitude of the cross product of two vectors. \square

Exercise 2.3.1.**quadruple scalar product**

On \mathbb{R}^n with any inner product $G(X, Y) = G_{ij}X^iY^j \equiv X \bullet Y$ define the corresponding scalar quadruple product by

$$\begin{aligned} Q(U, V, X, Y) &= (U \bullet X)(V \bullet Y) - (U \bullet Y)(V \bullet X) \\ &= \delta^{(2)b}(U, V, X, Y) = \delta_{ij,kl}U^iV^jX^kY^l \\ &= \delta^{(2)}(U^b, V^b, X, Y) = \delta^{ij}_{kl}U_iV_jX^kY^l, \end{aligned}$$

where

$$\delta_{ij,kl} = G_{im}G_{jn}\delta^{mn}_{kl}.$$

is the totally covariant form of this mixed tensor with respect to the metric, obtained by lowering the first two contravariant indices with the metric.

a) Using the definition $\delta^{ij}_{kl} = \delta^i_k\delta^j_l - \delta^i_l\delta^j_k$, show that the final line in the above equivalent definitions of the scalar quadruple product is equivalent to the previous lines.

b) Convince yourself that this tensor has the same symmetries as previously found for Q defined on \mathbb{R}^3 for the ordinary dot product, namely

$$\begin{aligned} \delta_{ij,kl} &= -\delta_{ji,kl} = -\delta_{ij,lk} && \text{(antisymmetry in each pair)} \\ &= \delta_{kl,ij}, && \text{(symmetry in pair interchange)} \\ 0 &= 3\delta_{i[j,kl]} = \delta_{ij,kl} + \delta_{ik,lj} + \delta_{il,jk}. && \text{(cyclic symmetry)} \end{aligned}$$

■

Exercise 2.3.2.**higher dimension contractions of the $p = 2$ generalized Kronecker delta**

This tensor plays an important role in curvature in any dimension, where its contractions have direct application. Derive the following formulas for \mathbb{R}^n

$$\delta^{ij}_{kj} = (n-1)\delta^i_k, \quad \delta^{ij}_{ij} = n(n-1).$$

■

Exercise 2.3.3.**Jacobian matrix**

a) On \mathbb{R}^2 for the new basis $\underline{B} = \langle b_1 | b_2 \rangle = \langle \langle 1, 1 | \langle -2, 1 \rangle \rangle$ with dual basis $\underline{B}^{-1} = \underline{A} = \langle \Omega^1, \Omega^2 \rangle = \frac{1}{3} \langle \langle 1 | 2 \rangle, \langle -1 | 1 \rangle \rangle$, evaluate $\delta^{(2)}(\omega^1, \omega^2, E_1, E_2) = \omega^1(E_1)\omega^2(E_2) - \omega^1(E_2)\omega^2(E_1) = \det \underline{B}$.

b) Letting $\{y^1, y^2\} = \{\Omega^1, \Omega^2\}$ denote the new coordinate functions, then show by partial differentiation of $x^i = B^i_j y^j$ that this same result represents the so called Jacobian determinant

$$\det \left(\frac{\partial x^i}{\partial y^j} \right) = \det \underline{B}.$$

The absolute value of this represents the area of the parallelogram formed the new basis vectors which is the unit parallelogram associated with the new coordinate grid. This is the “amplification factor” of grid box areas going from the old to new coordinates. ■

Finally define the alternating or permutation or Levi-Civita symbols by

$$\epsilon_{i_1 \dots i_n} = \delta_{i_1 \dots i_n}^{1 \dots n}, \quad \epsilon^{i_1 \dots i_n} = \delta_{1 \dots n}^{i_1 \dots i_n}.$$

These do not define the components of a single tensor but a different tensor for each choice of basis. For example δ^1_i are the components of $\delta^1_i \omega^i = \omega^1$, which is just the first dual basis vector in a dual basis. Similarly $\epsilon_{i_1 \dots i_n}$ are the components of $\delta_{i_1 \dots i_n}^{1 \dots n} \omega^{i_1} \otimes \dots \otimes \omega^{i_n} = p! \omega^1 \otimes \dots \otimes \omega^n \equiv \omega^{1 \dots n}$. We will return to these antisymmetric tensors later.

Meanwhile these alternating symbols are useful in representing determinants without explicit summation notation. Notice that from the definition of the determinant of a matrix

$$\begin{aligned} \det \underline{A} &= \sum_{\sigma \in S_p} (\text{sgn } \sigma) \underbrace{A^{\sigma(1)}_1 \dots A^{\sigma(n)}_n}_{\substack{\text{rearrange factors so} \\ \text{top order is } 1, 2, \dots, n}} \\ &= \sum_{\sigma \in S_p} (\text{sgn } \sigma) A^1_{\sigma^{-1}(1)} \dots A^n_{\sigma^{-1}(n)} \\ &= \sum_{\sigma \in S_p} (\text{sgn } \sigma) A^1_{\sigma(1)} \dots A^n_{\sigma(n)}, \end{aligned}$$

where the inverse appears when you re-order the lower row of the permutation above since

$$\sigma \sim \begin{pmatrix} 1 & \dots & n \\ \sigma(1) & \dots & \sigma(n) \end{pmatrix} \sim \begin{pmatrix} \sigma^{-1}(1) & \dots & \sigma^{-1}(n) \\ 1 & \dots & n \end{pmatrix}$$

and the sum over σ^{-1} for all $\sigma \in S_p$ is a sum over every permutation without the inverse since every permutation can be represented as the inverse of another permutation (group property) and since $\text{sgn } \sigma^{-1} = \text{sgn } \sigma$, the inverse can simply be dropped above. Thus we have just proven the equivalence of permuting either the rows or the columns in evaluating the determinant from this definition. But the alternating symbols provide the sign and allow one to sum over all permutations using our summation convention, permuting either the row indices or column indices, so the above formulas can be rewritten as

$$\det \underline{A} = \epsilon_{i_1 \dots i_n} A^{i_1}_1 \dots A^{i_n}_n = \epsilon^{i_1 \dots i_n} A^1_{i_1} \dots A^n_{i_n}.$$

Also since a permutation of the columns of \underline{A} changes $\det \underline{A}$ by its sign

$$(\det \underline{A}) \epsilon_{j_1 \dots j_n} = \epsilon_{i_1 \dots i_n} A^{i_1}_{j_1} \dots A^{i_n}_{j_n},$$

with a similar result for the rows

$$(\det \underline{A}) \epsilon^{j_1 \dots j_n} = \epsilon^{i_1 \dots i_n} A^{j_1}_{i_1} \dots A^{j_n}_{i_n},$$

or replacing \underline{A} by \underline{A}^{-1} in the first and not in the second (recall $\det \underline{A}^{-1} = (\det \underline{A})^{-1}$) and dividing both sides by the determinant leads to

$$\begin{aligned}\epsilon_{j_1 \dots j_n} &= (\det \underline{A}^{-1})^{-1} \underbrace{\epsilon_{i_1 \dots i_n} A^{-1 i_1}_{j_1} \dots A^{-1 i_n}_{j_n}}_{\text{(weight } W = -1)} \\ \epsilon^{j_1 \dots j_n} &= (\det \underline{A}^{-1}) \underbrace{\epsilon^{i_1 \dots i_n} A^{j_1}_{i_1} \dots A^{j_n}_{i_n}}_{\text{(weight } W = 1)}\end{aligned}$$

We can interpret these identities in the following way. The underbraced quantity is the correct transformation law for the tensor with components $\epsilon_{i_1 \dots i_n}$ or $\epsilon^{i_1 \dots i_n}$ respectively in the starting basis, under the change of basis $e_{i'} = A^{-1 j}_{i'} e_j$, but the transformed component values are no longer 1, -1, 0 in the new basis. The additional “weight W ” scaling factor of the (inverse matrix) determinant $(\det \underline{A}^{-1})^W$ restores these numerical values by changing to a new tensor in the new basis whose components have the same numerical values as the old tensor in the old basis.

So in fact these alternating symbols define the components of a 1-parameter family of proportional $\binom{0}{p}$ -tensors and $\binom{p}{0}$ -tensors respectively which together are referred to as a “tensor density of weight $W = -1$ and $W = 1$ ” respectively ($\det \underline{A} \neq 0$ for a change of basis, but it can assume all nonzero values). In fact these are not so unfamiliar. For any basis $\{e_i\}$ of our vector space V we can identify the components of vectors with column matrices

$$u = u^j e_j \longrightarrow \underline{u} = \begin{pmatrix} u^1 \\ \vdots \\ u^n \end{pmatrix}$$

and the value of the tensor $\epsilon_{i_1 \dots i_n} \omega^{i_1} \otimes \dots \otimes \omega^{i_n}$ on n vector arguments is just the determinant of the matrix whose columns are these column matrices

$$\epsilon_{i_1 \dots i_n} \omega^{i_1} \otimes \dots \otimes \omega^{i_n}(u_{(1)}, \dots, u_{(n)}) = \epsilon_{i_1 \dots i_n} u_{(1)}^{i_1} \dots u_{(n)}^{i_n} = \det \langle \underline{u}_{(1)} | \dots | \underline{u}_{(n)} \rangle.$$

But under a change of basis

$$u^{i'} = A^i_{j'} u^j, \quad \underline{u}' = \underline{A} \underline{u}$$

one finds

$$\begin{aligned}\det(\underline{u}'_{(1)} \dots \underline{u}'_{(n)}) &= \det \langle \underline{A} \underline{u}_{(1)} | \dots | \underline{A} \underline{u}_{(n)} \rangle \\ &= \det[\underline{A} \langle \underline{u}_{(1)} | \dots | \underline{u}_{(n)} \rangle] && \text{(definition of matrix product)} \\ &= (\det \underline{A}) \det \langle \underline{u}_{(1)} | \dots | \underline{u}_{(n)} \rangle, && \text{(product rule for matrix determinant)}\end{aligned}$$

where we have used the fact that the matrix product $\underline{A} \langle \underline{u}_{(1)} | \dots | \underline{u}_{(n)} \rangle$ is equivalent to multiplying each column by \underline{A} . Thus the determinant of the new column matrices differs from that of the old ones by the determinant of the transformation itself, explaining why one gets a different (but proportional) tensor for different choices of basis. One can look at the determinant in this context as the volume amplification factor (contraction if its absolute value is a proper fraction)

describing how the volume of the basic parallelepiped formed by the basis vectors changes under the change of basis, independent of what its value actually is (and which depends on having an inner product to set the scale).

Now I have to admit I have a sick love for lots of indices, but I wouldn't drag you through this index jungle if it weren't true that the algebra of antisymmetric tensors did not play a fundamental role in differential geometry.¹ Questions of measure for p -dimensional surfaces necessary for generalizing line integrals and surface integrals of vector fields and volume integrals of functions all involve this algebra in a way that will turn out to be very beautiful. Trust me.

What happened to symmetric tensors? Except for inner products they turn out not to be as important, so we don't need to develop machinery for them, which anyway involves the symmetric group in a much more nontrivial way. But we're not finished. First an easy formula

$$\delta_{j_1 \dots j_n}^{i_1 \dots i_n} = \epsilon^{i_1 \dots i_n} \epsilon_{j_1 \dots j_n} ,$$

which is true since the sign of the permutation of the upper indices relative to the lower indices of the left hand side is just the product of the signs of the permutation from the upper indices to $(1 \dots n)$ and then from $(1 \dots n)$ to the lower indices.

Next we state a hard formula (hard only because we have to do a counting game with the permutations, just accept it for now) and an easy consequence of it (the second formula follows from the previous easy formula) for index pair contractions of the generalized Kronecker deltas and hence for contractions of the product of the two Levi-Civita symbols

$$\begin{aligned} \delta_{j_1 \dots j_p}^{i_1 \dots i_p} &= \frac{1}{(n-p)!} \delta_{j_1 \dots j_p k_{p+1} \dots k_n}^{i_1 \dots i_p k_{p+1} \dots k_n} \\ &= \frac{1}{(n-p)!} \epsilon^{i_1 \dots i_p k_{p+1} \dots k_n} \epsilon_{j_1 \dots j_p k_{p+1} \dots k_n} \end{aligned}$$

that finishes the foundation. Next we build the house.

Remark.

By iteration of this ugly formula you can get

$$\delta_{j_1 \dots j_p}^{i_1 \dots i_p} = \frac{(n-q)!}{(n-p)!} \delta_{j_1 \dots j_p k_{p+1} \dots k_q}^{i_1 \dots i_p k_{p+1} \dots k_q} \quad 1 \leq p < q \leq n .$$

I had to sneak that in. I am not sure this is ever needed, but you can find it in books. We can extend this to $p = 0$ if by the Kronecker delta with no indices we mean the number 1.

Note that for $q = 2$ and $p = 1$ this reduces to the first result of Exercise 2.3.2

$$\delta_{j_1}^{i_1} = \frac{(n-2)!}{(n-1)!} \delta_{j_1 k}^{i_1 k} = \frac{1}{(n-1)} \delta_{j_1 k}^{i_1 k} ,$$

while its trace immediately leads to the second result of that exercise.

¹Maybe I would. . .



It helps to look at the case $n = 3$ to have a more concrete idea of what all this means, where there are 6 permutations and the signs are

$$\epsilon_{123} = \epsilon_{213} = \epsilon_{312} = 1, \quad \epsilon_{132} = \epsilon_{231} = \epsilon_{321} = -1.$$

Then the following identities hold for the definition of the generalized Kronecker deltas

$$\begin{aligned} \delta_{k\ell}^{ij} &= \begin{vmatrix} \delta_k^i & \delta_\ell^i \\ \delta_k^j & \delta_\ell^j \end{vmatrix} &= \delta_k^i \delta_\ell^j - \delta_\ell^i \delta_k^j = 2\delta_{[k}^i \delta_{\ell]}^j = 2\delta^{[i}_k \delta^{j]}_\ell, \\ \delta_{mn\ell}^{ijk} &= \begin{vmatrix} \delta_m^i & \delta_n^i & \delta_\ell^i \\ \delta_m^j & \delta_n^j & \delta_\ell^j \\ \delta_m^k & \delta_n^k & \delta_\ell^k \end{vmatrix} &= \delta_m^i \delta_n^j \delta_\ell^k + \delta_n^i \delta_\ell^j \delta_m^k + \delta_\ell^i \delta_m^j \delta_n^k \\ &&\quad - \delta_m^i \delta_\ell^j \delta_n^k - \delta_n^i \delta_m^j \delta_\ell^k - \delta_\ell^i \delta_n^j \delta_m^k, \end{aligned}$$

and for various their contractions

$$\begin{aligned} \delta_{mnk}^{ijk} &= \delta_m^i \delta_n^j \delta_k^k + \delta_n^i \delta_k^j \delta_m^k + \delta_k^i \delta_m^j \delta_n^k \\ &\quad - \delta_m^i \delta_k^j \delta_n^k - \delta_n^i \delta_m^j \delta_k^k - \delta_k^i \delta_n^j \delta_m^k \\ &= (3 - 1 - 1)\delta_m^i \delta_n^j - (3 - 1 - 1)\delta_n^i \delta_m^j = \delta_m^i \delta_n^j - \delta_n^i \delta_m^j = \delta^{ij}_{nm}, \\ \delta_{mjk}^{ijk} &= \delta^{ij}_{mj} = \delta_m^i \delta_j^j - \delta_j^i \delta_m^j = (3 - 1)\delta_m^i = 2\delta_m^i, \\ \delta_{ijk}^{ijk} &= 2\delta_i^i = 2 \cdot 3 = 6. \end{aligned}$$

Whew! Comparing with our previous contraction formula, we see that the final coefficient is as it should be just $(n - p)!$, where p is the number of uncontracted index pairs, namely $(3 - 2)! = 1! = 1$, $(3 - 1)! = 2! = 2$, and $(3 - 0)! = 3! = 6$ respectively.

Remark.

This Kronecker delta business is just a shorthand for giving compact expressions in tensor notation for 3×3 determinants and subdeterminants (and cofactors and minors, minors of minors, etc.). For example, we have the determinant

$$\begin{aligned} \delta_{mnp}^{123} X^m Y^n Z^p &= X^1 Y^2 Z^3 + X^2 Y^3 Z^1 + X^3 Y^1 Z^2 \\ &\quad - X^1 Y^3 Z^2 - X^2 Y^1 Z^3 - X^3 Y^2 Z^1 \\ &= \begin{vmatrix} X^1 & Y^1 & Z^1 \\ X^2 & Y^2 & Z^2 \\ X^3 & Y^3 & Z^3 \end{vmatrix} \end{aligned}$$

and one of the “minors” of the previous determinant, namely the determinant of the 2×2 matrix obtained by deleting the last row and column of the previous matrix

$$\delta_{mn}^{12} X^m Y^n = X^1 Y^2 - X^2 Y^1 = \begin{vmatrix} X^1 & Y^1 \\ X^2 & Y^2 \end{vmatrix}.$$

Minors and cofactors, etc., used to be important in evaluating determinants by hand, but not any more since technology is at our fingertips. However, the mathematics they represent still remains important in integration theory where these generalized Kronecker deltas reign. ▲

For the case in which $\underline{A} = \langle \underline{u}_{(1)} | \cdots | \underline{u}_{(n)} \rangle$ we get on \mathbb{R}^3

$$\begin{aligned} \det(\langle \underline{u}_{(1)} | \cdots | \underline{u}_{(n)} \rangle) &= \epsilon_{i_1 \cdots i_n} A^{i_1}_{(1)} \cdots A^{i_n}_{(n)} \\ &= \delta^{1 \cdots n}_{i_1 \cdots i_n} u^{i_1}_{(1)} \cdots u^{i_n}_{(n)} \\ &= n! u^{[1}_{(1)} \cdots u^{n]}_{(n)}. \end{aligned}$$

This is the single independent component of the antisymmetrized tensor product of the n vectors

$$\begin{aligned} n! [\text{ALT}(u_{(1)} \otimes \cdots \otimes u_{(n)})]^{i_1 \cdots i_n} &= \delta^{i_1 \cdots i_n}_{j_1 \cdots j_n} u^{j_1}_{(1)} \cdots u^{j_n}_{(n)} \\ &= \epsilon^{i_1 \cdots i_n}_{j_1 \cdots j_n} u^{j_1}_{(1)} \cdots u^{j_n}_{(n)} \\ &= \epsilon^{i_1 \cdots i_n} \det(\langle \underline{u}_{(1)} | \cdots | \underline{u}_{(n)} \rangle). \end{aligned}$$

The generalized Kronecker delta with p upper and p lower indices arises in a very simple way as the antisymmetrizer operator modulo the factorial factor

$$\begin{aligned} u^{i_1}_{(1)} \cdots u^{i_p}_{(p)} &= \delta^{i_1 \cdots i_p}_{j_1 \cdots j_p} u^{j_1}_{(1)} \cdots u^{j_p}_{(p)} \\ p! u^{[i_1}_{(1)} \cdots u^{i_p]}_{(p)} &= p! \delta^{[i_1 \cdots i_p]}_{j_1 \cdots j_p} u^{j_1}_{(1)} \cdots u^{j_p}_{(p)} \\ &= \delta^{i_1 \cdots i_p}_{j_1 \cdots j_p} u^{j_1}_{(1)} \cdots u^{j_p}_{(p)}. \end{aligned}$$

The antisymmetrized tensor product of p vectors in an n -dimensional vector space contains both information about the p -measure of the p -parallelepiped they form as well as its orientation within the space just like the cross product does in \mathbb{R}^3 (almost). An inner product merely sets the scale of the p -measure.

Example 2.3.2. On \mathbf{R}^3 with the usual dot product and using only positively oriented orthonormal frames for components, the triple scalar product is $X \cdot (Y \times Z) = \epsilon_{ijk} X^i Y^j Z^k$. It evidently does not need the dot product for its evaluation since it is just the determinant of the matrix whose rows or columns are the components of the three vector factors. The dot product cancels out from the combination of the cross product and dot product, since the cross product itself involves the index raising on the first index

$$[Y \times Z]^i = \delta^{im} \epsilon_{mjk} Y^j Z^k$$

so

$$\begin{aligned} X \cdot (Y \times Z) &= X^n \delta_{ni} [Y \times Z]^i = X^n \delta_{ni} \delta^{im} \epsilon_{mjk} Y^j Z^k = X^n \delta^n_m \epsilon_{mjk} Y^j Z^k \\ &= X^n \epsilon_{njk} Y^j Z^k. \end{aligned}$$

From the obvious properties of the determinant encoded in the Levi-Civita permutation symbol, one gets the usual identities for the cyclic and anticyclic permutations of the vector factors

$$\begin{aligned} X \cdot (Y \times Z) &= Y \cdot (Z \times X) = Z \cdot (X \times Y) \\ &= -X \cdot (Z \times Y) = -Y \cdot (X \times Z) = -Z \cdot (Y \times X). \end{aligned}$$

Classical vector analysis has a set of identities involving other combinations of at least two vector operations which can be described in this way. \square

Exercise 2.3.4.

quadruple scalar product again

a) The $n = 3$ summation formula $\epsilon^{ijk}\epsilon_{mnk} = \delta_{mn}^{ij} = \delta_m^i\delta_n^j - \delta_n^i\delta_m^j$ in classical vector analysis underlies the quadruple scalar product identity

$$Q(X, Y, Z, W) = (X \times Y) \cdot (Z \times W) = (X \cdot Z)(Y \cdot W) - (X \cdot W)(Y \cdot Z).$$

Show that the component form for these two expressions agree expressing the cross product and dot product in terms of the Levi-Civita permutation symbol and the Kronecker delta and then using the summation identity stated at the beginning of this problem, ignoring upper and lower index positions but maintaining our repeated index summation convention, possible only if we are working in an orthonormal basis. [Explicitly just make sure all contracted indices are an up/down pair and use the all up, all down epsilons to make all the other contractions agree in index position, like: $(X \times Y) \cdot (Z \times W) = \epsilon_{ijk}X^jY^k\epsilon^{imn}Z_mW_n = Q_{ij}^{mn}Z_mW_nX^jY^k$. What does this tell you the mixed $\binom{2}{2}$ -components of Q are? Thus the quadruple scalar product is really the 2 index pair Kronecker delta tensor!]

b) Now write the component form of the above identity with the Kronecker deltas in the right places to respect our index positioning conventions corresponding to lowering indices on $\delta^{(2)}$.

c) The double cross product, also called the triple vector product (it has two crosses, three vectors), satisfies a well known identity

$$X \times (Y \times Z) = (X \cdot Z)Y - (X \cdot Y)Z.$$

Use the same component technique and the same summation identity to establish that the component formulas of the left and right hand sides of this equation agree.

d) Show that

$$W \cdot (X \times (Y \times Z)) = Q(X, Y, Z, W).$$

Thus the double cross product is really just the tensor Q with its last argument left unevaluated. We already saw in Example 2.3.1 that the $\binom{0}{4}$ -tensor Q_{ijmn} is really the $\binom{2}{2}$ generalized Kronecker delta tensor δ_{mn}^{ij} with its first two indices lowered. In short these generalized Kronecker deltas are lying just below the surface of traditional vector analysis. \blacksquare

It would be a shame to move on from this point without deriving an important identity for the derivative of the determinant of a matrix with respect to one of its entries. Let's recap the story of the determinant and epsilons and deltas from the beginning. We introduced the Levi-Civita epsilon to extend our summation convention to the sum over permutations with signs

$$\det \underline{A} = \epsilon_{i_1 \dots i_n} A^{i_1}_1 \cdots A^{i_n}_n$$

and since a permutation of the columns of \underline{A} changes $\det \underline{A}$ by its sign we get

$$(\det \underline{A}) \epsilon_{j_1 \dots j_n} = \epsilon_{i_1 \dots i_n} A^{i_1}_{j_1} \dots A^{i_n}_{j_n},$$

so if we now contract this with another Levi-Civita epsilon we get

$$(\det \underline{A}) \epsilon_{j_1 \dots j_n} \epsilon^{j_1 \dots j_n} = \epsilon^{i_1 \dots i_n} \epsilon_{j_1 \dots j_n} A^{j_1}_{i_1} \dots A^{j_n}_{i_n} = \delta^{i_1 \dots i_n}_{j_1 \dots j_n} A^{j_1}_{i_1} \dots A^{j_n}_{i_n},$$

since the delta is defined as the product of the upper and lower epsilons. But $\epsilon_{j_1 \dots j_n} \epsilon^{j_1 \dots j_n} = n!$ since we are summing 1 over $n!$ permutations so finally

$$\begin{aligned} \det \underline{A} &= \frac{1}{n!} \delta^{i_1 \dots i_n}_{j_1 \dots j_n} A^{j_1}_{i_1} \dots A^{j_n}_{i_n} = \frac{1}{n} \left(\frac{1}{(n-1)!} \delta^{i_1 \dots i_n}_{j_1 \dots j_n} A^{j_1}_{i_1} \dots A^{j_{n-1}}_{i_{n-1}} \right) A^{j_n}_{i_n} \\ &\equiv \frac{1}{n} \Delta(\underline{A})^{i_n}_{j_n} A^{j_n}_{i_n}, \end{aligned}$$

which enables us to define the cofactor $\Delta(\underline{A})^{j_i}_i$ associated with the entry A^i_j as the determinant of the matrix obtained by eliminating its row and column. In fact the index positioning means that the matrix $(\Delta(\underline{A})^{j_i}_i)$ is already the transpose of the usual matrix of cofactors, so all we need to do is divide it by the determinant to get the inverse matrix according to the well known formula which we can verify

$$A^{-1i}_j = \frac{1}{\det(\underline{A})} \Delta(\underline{A})^i_j.$$

Using this formula for the moment we get from the previous relation

$$\det \underline{A} = \frac{1}{n} A^{-1i_n}_{j_n} A^{j_n}_{i_n} = \frac{1}{n} \delta^{i_n}_{i_n} = 1,$$

which is consistent with the previous definition of the inverse matrix but not a proof that the formula is justified. Using this formula it follows (using the various definitions) that

$$\begin{aligned} A^{-1i_n}_{j_n} A^{j_n}_k &= \frac{1}{\det(\underline{A})} \Delta(\underline{A})^{i_n}_{j_n} A^{j_n}_k \\ &= \frac{1}{\det(\underline{A})} \frac{1}{(n-1)!} \delta^{i_1 \dots i_n}_{j_1 \dots j_n} A^{j_1}_{i_1} \dots A^{j_{n-1}}_{i_{n-1}} A^{j_n}_k \\ &= \frac{1}{\det(\underline{A})} \frac{1}{(n-1)!} \epsilon^{i_1 \dots i_n} \epsilon_{j_1 \dots j_n} A^{j_1}_{i_1} \dots A^{j_{n-1}}_{i_{n-1}} A^{j_n}_k \\ &= \frac{1}{\det(\underline{A})} \frac{1}{(n-1)!} \epsilon^{i_1 \dots i_n} (\epsilon_{i_1 \dots i_{n-1} k} \det(\underline{A})) \\ &= \frac{1}{(n-1)!} \delta^{i_1 \dots i_n}_{i_1 \dots i_{n-1} k} = \delta^{i_n}_k, \end{aligned}$$

justifying the classic formula.

Now consider the differential using the product rule

$$\begin{aligned}
 d(\det(\underline{A})) &= d\left(\frac{1}{n!}\delta_{j_1\cdots j_n}^{i_1\cdots i_n}A^{j_1}_{i_1}\cdots A^{j_n}_{i_n}\right) \\
 &= \frac{1}{n!}\left(\delta_{j_1\cdots j_n}^{i_1\cdots i_n}dA^{j_1}_{i_1}\cdots A^{j_n}_{i_n} + \cdots + \delta_{j_1\cdots j_n}^{i_1\cdots i_n}A^{j_1}_{i_1}\cdots dA^{j_n}_{i_n}\right) \\
 &= \frac{1}{(n-1)!}\delta_{j_1\cdots j_n}^{i_1\cdots i_n}A^{j_1}_{i_1}\cdots A^{j_{n-1}}_{i_{n-1}}dA^{j_n}_{i_n} \\
 &= \Delta(\underline{A})^{i_n}_{j_n}dA^{j_n}_{i_n} \\
 &= \det(\underline{A})A^{-1i_n}_{j_n}dA^{j_n}_{i_n},
 \end{aligned}$$

so that in terms of the logarithmic determinant we have finally

$$d\ln|\det(\underline{A})| = A^{-1i}_j dA^j_i = \text{Tr } \underline{A}^{-1}d\underline{A} = \text{Tr } d\underline{A} \underline{A}^{-1},$$

where the last equality follows from the cyclic symmetry property of the trace of a matrix product. This means that as long as the trace of the matrix differential $\underline{A}^{-1}d\underline{A}$ is zero, the determinant of the matrix \underline{A} remains constant, which is a useful relation for matrix groups whose determinant is 1, called unimodular matrix groups. The other famous application of this formula applies to a symmetric matrix $\underline{g} = (g_{ij}) = (g_{ji})$ with inverse matrix $\underline{g}^{-1} = (g^{ij})$ where this becomes

$$d\ln|\det(\underline{g})| = g^{ij}dg_{ij}.$$

It is a crucial formula needed for deriving the Einstein equations from the Hilbert Lagrangian. Hilbert Lagrangian? Google it.

Exercise 2.3.5.

differential of the determinant

Justify each of the lines in the above derivation.

■

Exercise 2.3.6.

inverse matrix differential

a) Derive one further identity involving the inverse matrix by evaluating the differential of the identity

$$\underline{A}^{-1}\underline{A} = \underline{I}$$

using the product rule and then right multiplying by \underline{A}^{-1} to then solve for

$$d\underline{A}^{-1} = -\underline{A}^{-1}d\underline{A}\underline{A}^{-1}$$

b) If we apply this to a symmetric matrix \underline{g} of components of an inner product, show that this becomes

$$dg^{ij} = -g^{im}dg_{mn}g^{nj} = -g^{im}g^{in}dg_{mn}.$$

This is a second crucial formula for deriving the Einstein equations.

**Exercise 2.3.7.****relative differential rotations and boosts**

In exercise 1.4.1 we evaluated the differentials of the rotation and boost matrices in the plane. Here we evaluate the differential rotation and boost relative to the image point.

a) Consider the rotation matrix

$$\underline{R}(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

Evaluate $\underline{R}^{-1}d\underline{R}$ and show that its trace is zero, as it should be since the determinant of this matrix is identically 1.

b) Consider the hyperbolic rotation matrix

$$\underline{B}(\alpha) = \begin{pmatrix} \cosh \alpha & \sinh \alpha \\ \sinh \alpha & \cosh \alpha \end{pmatrix}.$$

Evaluate $\underline{B}^{-1}d\underline{B}$ and show that its trace is zero, as it should be since the determinant of this matrix is identically 1.

**Exercise 2.3.8.****antisymmetry of the electromagnetic field tensor**

Recall Exercise 1.6.6 introducing the electromagnetic field tensor matrix $\underline{F} = (F^i_j)$. Evaluate the index lowered matrix $F_{ij} = \eta_{ik}F^k_j$ and verify that it is an antisymmetric matrix. This is the condition that \underline{F} lie in the Lie algebra of the Lorentz group. The electric part of this matrix generates a boost while the magnetic part generates a rotation.



2.4 Antisymmetric tensors

We need names for the vector spaces of tensors of given types over an n -dimensional vector space V with basis e_i and dual basis ω^i . Let $T^{(p,q)}(V)$ be the vector space of $\binom{p}{q}$ -tensors over V . Each tensor S in this space is of the form

$$S = S_{j_1 \dots j_q}^{i_1 \dots i_p} \underbrace{e_{i_1} \otimes \dots \otimes e_{i_p} \otimes \omega^{j_1} \otimes \dots \otimes \omega^{j_q}}_{\text{basis tensors}}, \quad \text{where} \quad S_{j_1 \dots j_q}^{i_1 \dots i_p} = S(\omega^{i_1}, \dots, \omega^{i_p}, e_{j_1}, \dots, e_{j_q})$$

are its components with respect to the basis of V . The underbraced factor is a basis of $T^{(p,q)}(V)$, labeled by $p+q$ indices, each taking n values, so the dimension of this space of tensors is n^{p+q} , and the components of S with respect to this basis are what we refer to as components with respect to e_i .

The original vector space and its dual space are just $V = T^{(1,0)}(V)$ and $V^* = T^{(0,1)}(V)$ in this notation. For $0 \leq g \leq n$, let $\Lambda^{(p)}(V) = \text{ALT } T^{(p,0)}(V)$ and $\Lambda^{(p)}(V)^* = \text{ALT } T^{(0,p)}(V)$ be the linear subspaces of antisymmetric $\binom{p}{0}$ -tensors (called p -vectors) and antisymmetric $\binom{0}{p}$ -tensors (called p -covectors or p -forms) respectively. These tensors are of the form

$$\begin{aligned} T &= T^{i_1 \dots i_p} e_{i_1} \otimes \dots \otimes e_{i_p}, & T^{i_1 \dots i_p} &= T^{[i_1 \dots i_p]}, \\ S &= S_{i_1 \dots i_p} \omega^{i_1} \otimes \dots \otimes \omega^{i_p}, & S_{i_1 \dots i_p} &= S_{[i_1 \dots i_p]}. \end{aligned}$$

Antisymmetric tensors cannot have nonzero components with any repeated indices, since interchanging any pair of indices must change the sign of the result, but an interchange of an identical pair does change the component so it can only be zero. For example

$$S_{ijk} = -S_{jik} \longrightarrow S_{112} = -S_{112} \longrightarrow S_{112} = 0.$$

Thus an antisymmetric tensor can have at most n indices without being identically zero. The no-repeat condition tells us the dimension of the space of antisymmetric tensors of a given “rank” p , or equivalently the number of “independent components” of such a tensor. The number of p -tuples of distinct integers chosen from the set of integers $(1, \dots, n)$ is by definition the number of combinations of n things taken r at a time

$$\dim \Lambda^{(p)}(V) = \dim \Lambda^{(p)}(V)^* = \binom{n}{p} = \frac{n!}{p!(n-p)!}.$$

If we define $\Lambda^{(0)}(V) = \Lambda^{(0)}(V)^* = \mathbb{R}$, i.e. the $\binom{0}{0}$ -tensors or scalars are identified with antisymmetric tensors with no indices (1 index tensors are antisymmetric by default), then we have $(n+1)$ such spaces for the contravariant and covariant cases which pair off by dimension since

$$\binom{n}{p} = \binom{n}{n-p}.$$

So from the symmetries of these binomial coefficients, the cases $p = 0$ and $p = n$ are both 1-dimensional, $p = 1$ and $p = n - 1$ are both n -dimensional, $p = 2$ and $p = n - 2$ are both $n(n-1)/2$ -dimensional, etc.

Example 2.4.1. Consider the case $n = 3$. A scalar S has a single independent “component.” A vector $S^i e_i$ has 3 independent components (S^1, S^2, S^3). A 2-vector $S^{ij} e_i \otimes e_j$ has 3 independent components (S^{23}, S^{31}, S^{12}). A 3-vector $S^{ijk} e_i \otimes e_j \otimes e_k$ has single independent component S^{123} . \square

Any time we get our hands on a vector space, we try to find a convenient basis. We can do the same here. Consider the p -vector case, where we make the definition

$$\delta_{j_1 \dots j_p}^{i_1 \dots i_p} e_{i_1} \otimes \dots \otimes e_{i_p} \equiv e_{i_1 \dots i_p} \equiv p! e_{[i_1} \otimes \dots \otimes e_{i_p]} .$$

Using the antisymmetry condition

$$S^{i_1 \dots i_p} = S^{[i_1 \dots i_p]} = \frac{1}{p!} \delta_{j_1 \dots j_p}^{i_1 \dots i_p} S^{j_1 \dots j_p} ,$$

then substituting it into the expression for the tensor yields

$$\begin{aligned} S &= S^{i_1 \dots i_p} e_{i_1} \otimes \dots \otimes e_{i_p} = \frac{1}{p!} S^{j_1 \dots j_p} \delta_{j_1 \dots j_p}^{i_1 \dots i_p} e_{i_1} \otimes \dots \otimes e_{i_p} \\ &= \frac{1}{p!} S^{j_1 \dots j_p} e_{j_1 \dots j_p} = \sum_{i_1 < \dots < i_p} S^{j_1 \dots j_p} e_{j_1 \dots j_p} , \end{aligned}$$

since each distinct permutation in the sum is repeated $p!$ times with the same value, so it is enough to sum only over ordered p -tuplets of indices, without the $p!$ factor. For example

$$\frac{1}{6} S^{i_1 i_2 i_3} e_{i_1 i_2 i_3} = \frac{1}{6} (S^{123} e_{123} + S^{231} e_{231} + S^{312} e_{312} - S^{132} e_{132} - S^{213} e_{213} - S^{321} e_{321}) = S^{123} e_{123} ,$$

since both factors in each term only change sign with each permutation, leading to no change in their product.

The set $\{e_{i_1 \dots i_p}\}_{i_1 < \dots < i_p}$ is a basis for p -vectors since any p -vector can be expressed as a linear combination of them and they are linearly independent.

Exercise 2.4.1.

linear independence of basis p -vectors

We have only shown that any p -vector is a linear combination of these basis p -vectors. How do we show linear independence, i.e., that

$$\sum_{i_1 < \dots < j_p} S^{j_1 \dots j_p} e_{j_1 \dots j_p} = 0 \longrightarrow S^{j_1 \dots j_p} = 0$$

holds for all possible index values? Hint: evaluate this equation on $(\omega^{i_1}, \dots, \omega^{i_p})$. \blacksquare

Example 2.4.2. Consider the case $n = 3$, $p = 2$.

$$\begin{aligned}
 S &= S^{ij} e_i \otimes e_j \\
 &= S^{12} e_1 \otimes e_2 + S^{21} e_2 \otimes e_1 + S^{13} e_1 \otimes e_3 + S^{31} e_3 \otimes e_1 + S^{23} e_2 \otimes e_3 + S^{32} e_3 \otimes e_2 \\
 &= S^{12} \underbrace{(e_1 \otimes e_2 - e_2 \otimes e_1)}_{e_{12}} + S^{13} \underbrace{(e_1 \otimes e_3 - e_3 \otimes e_1)}_{e_{13}} + S^{23} \underbrace{(e_2 \otimes e_3 - e_3 \otimes e_2)}_{e_{23}} \\
 &= S^{12} e_{12} + S^{13} e_{13} + S^{12} e_{12} = S^{23} e_{23} + S^{31} e_{31} + S^{12} e_{12}
 \end{aligned}$$

For this case it turns out that the ordered basis $\{e_{23}, e_{31}, e_{12}\}$ is more useful because of its cyclic properties, as we will see later. □

Well, rather than write $\sum_{i_1 < \dots < i_p}$ (\sum -notation is bad, remember) we just sum over all orderings and divide by $p!$ when we represent a p -vector abstractly, or we introduce more notation

$$S = \frac{1}{p!} S^{i_1 \dots i_p} e_{i_1 \dots i_p} = \sum_{i_1 < \dots < i_p} S^{i_1 \dots i_p} e_{i_1 \dots i_p} \equiv S^{i_1 \dots i_p} e_{|i_1 \dots i_p|} \equiv S^{|i_1 \dots i_p|} e_{i_1 \dots i_p}.$$

Vertical bars enclosing a p -tuple of antisymmetric indices mean sum only over ordered p -tuple values and it clearly does not matter which set of indices is enclosed.

Exercise 2.4.2.

p -vectors in \mathbb{R}^4

For the case $n = 4$, write out explicitly the following sums

$$S = S^{ij} e_{|ij|} \quad (6 \text{ terms}), \quad T = T^{ijk} e_{|ijk|} \quad (4 \text{ terms}).$$

It is helpful to organize the terms in an order that groups them first by whether or not 4 is present as an index, and next by which member of the triplet $(1, 2, 3)$ is missing. ■

2.5 Symmetric tensors and multivariable Taylor series

While symmetric tensors are not quite as important as antisymmetric tensors, they still merit some attention! Multivariable Taylor series involve symmetric tensor coefficients and turn out to be quite useful for many reasons, one particularly physically interesting one of which is the theory of multipole moments (monopole, dipole, quadrupole, etc.) which characterize a distribution of mass or charge in some finite region around the origin of coordinates with respect to their long distance gravitational or electromagnetic effects. In studying curved surfaces, a multivariable Taylor series about a point of interest helps us analyze the local curvature, so it is a useful tool even for pure geometry.

Consider an infinitely differentiable function f of the Cartesian coordinates x^i on \mathbb{R}^n . We can represent it by a power series at the origin. It is almost as easy to establish a formula for a multivariable Taylor series as for the single variable case

$$f(x) = \sum_{p=0}^{\infty} T_{i_1 \dots i_p} x^{i_1} \dots x^{i_p} = T + T_i x^i + \frac{1}{2!} T_{ij} x^i x^j + \frac{1}{3!} T_{ijk} x^i x^j x^k + \dots$$

where the Taylor coefficients are the components of symmetric tensors

$$T_{i_1 \dots i_p} = \partial_{i_1} \dots \partial_{i_p} f(0) = \frac{\partial^p f}{\partial x^{i_1} \dots \partial x^{i_p}}(0) = T_{(i_1 \dots i_p)}$$

because the order of the partial derivatives does not matter. Clearly under linear transformations of the coordinates x^i , these must transform as the components of $\binom{0}{p}$ tensors so that their contraction with the p factors of the coordinate position vector is a scalar so that the Taylor expansion produces the same values of the function at a given position.

Establishing this formula for the Taylor coefficients is a simple calculation using the basic partial derivative formula $\partial_j x^i = \partial x^i / \partial x^j = \delta^i_j$ and the symmetry of the coefficients. The first few derivatives of the Taylor expansion are

$$\begin{aligned} \partial_m f(x) &= T_i \delta^i_m + \frac{1}{2!} T_{ij} (\delta^i_m x^j + x^i \delta^j_m) + \frac{1}{3!} T_{ijk} (\delta^i_m x^j x^k + x^i \delta^j_m x^k + x^i x^j \delta^k_m) + \dots \\ &= T_m + \frac{1}{2!} (T_{mj} x^j + T_{im} x^i) + \frac{1}{3!} (T_{mjk} x^j x^k + T_{imk} x^i x^k + T_{ijm} x^i x^j) + \dots \\ &= T_m + \frac{1}{1!} T_{mj} x^j + \frac{1}{2!} T_{mjk} x^j x^k + \dots \\ \partial_n \partial_m f(x) &= \frac{1}{1!} T_{mj} \delta^j_n + \frac{1}{2!} T_{mjk} (\delta^j_n x^k + x^j \delta^k_n) + \dots \\ &= T_{mn} + \frac{1}{2!} T_{mjk} (T_{mnk} x^k + T_{mjn} x^j) + \dots \\ &= T_{mn} + T_{mnk} x^k + \dots \end{aligned}$$

Evaluating these at $x = 0$ leads to

$$T = f(0), \quad T_m = \partial_m f(0), \quad T_{mn} = \partial_m \partial_n f(0), \dots, \quad T_{i_1 \dots i_p} = \partial_{i_1} \dots \partial_{i_p} f(0), \dots$$

Exercise 2.5.1.**multivariable Taylor series example**

Using a computer algebra system, evaluate the multivariable Taylor series approximation to $f(x, y) = \sin(x + 2y + x^2 + 4y^2)$ up through the third order terms. Compare the plots of the function and its approximation up to various orders on the rectangle $x = -2..2, y = -1..1$. Try deriving the coefficients up to the quadratic terms by hand. ■

Exercise 2.5.2.**Quadratic function graph approximation to sphere, ellipsoid at a pole**

a) Displacing a sphere of radius $a > 0$ from the origin a units up the z -axis makes it pass through the origin where its tangent plane is horizontal: $x^2 + y^2 + (z - a)^2 = a^2$. Solving this equation for the value of z on the lower hemisphere yields the function $f(x, y) = a - \sqrt{a^2 - x^2 - y^2}$, for which the value of the function and its first derivatives are zero at the origin, so that its Taylor series there starts at the quadratic terms. The quadratic coefficients define the symmetric matrix $\underline{T} = (T_{ij})$ as above (factor of 1/2 removed). Find this matrix. What are its eigenvalues? What is the value of the trace and determinant of this matrix? Use your computer algebra system Help to find the multivariable Taylor approximation command and check your hand results.

b) Repeat for the underside of the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{(z - c)^2}{c^2} = 1$$

at the origin, assuming $a > 0, b > 0, c > 0$.

c) Now we rotate the horizontal axes by 45 degrees in an explicit ellipsoid

$$\frac{((x - y)/\sqrt{2})^2}{4} + \frac{((x + y)/\sqrt{2})^2}{9} + (z - 1)^2 = 1$$

and solve for the underside value of z to define the function f

$$z = 1 - \frac{1}{12}\sqrt{144 - 26x^2 + 20xy - 26y^2} = f(x, y).$$

Evaluate the Taylor approximation to this function at the origin (use technology!) and identify the quadratic coefficient matrix (T_{ij}) , and find its eigenvalues and eigenvectors and use them to “find” the orthogonal transformation

$$x' = (x - y)/\sqrt{2}, \quad y' = (x + y)/\sqrt{2}$$

which diagonalizes it to read off its diagonal values. Recall that this matrix is the matrix of second derivatives at the origin. The diagonalized values of the second derivative matrix at a point where the first derivatives are zero are called the principal curvatures of the surface in this context and their product (the determinant of the matrix) is called the Gaussian curvature.

When not at a critical point of the function whose graph is the surface, the tangent plane is tilted rather than horizontal and a normalization factor must be taken into account as in the formula for the curvature of a plane curve at a point where the tangent line is not horizontal: $\kappa(x) = |f''(x)|/(1 + f'(x)^2)^{1/2}$. We'll get to this in Part 2. ■

Symmetric tensors and multipole moments in physics

This section can be safely ignored if you are not a physics student. It just shows that totally symmetric tensors do play some useful role in a limited number of applications but not as universal as antisymmetric tensors which are intimately tied to notions of linear independence and measure.

A distribution of mass or charge in a finite region of space creates a force field around it which is an integral of all the inverse square forces from each differential element of the source. This conservative force field outside the source can be obtained from the gradient of a potential function. By convention a sign change for the gradient is included so that the force points in the direction of decreasing potential. The actual distribution of the mass or charge can be replaced by an equivalent point source with an infinite tower of multipole moments which lead to the same force field as the actual distribution, in a way similar to the way in which the value of a function and all of its derivatives at a point can be used to reconstruct the whole function away from the point as an infinite series. For concreteness, consider the gravitational case.

The inverse square force per unit mass on a point particle of mass m at position \vec{a} by a point particle of mass M at position \vec{r} points from \vec{a} back towards \vec{r} along the unit vector \hat{n}

$$\frac{\vec{F}(\vec{a})}{m} = GM\hat{n}\frac{1}{|\vec{r} - \vec{a}|^2} = GM\frac{\vec{r} - \vec{a}}{|\vec{r} - \vec{a}|^3} = -\vec{\nabla}\Phi(\vec{a}),$$

where the potential function is $\Phi(\vec{a}) = -GM/|\vec{r} - \vec{a}|$. This is easily extended to a distribution of mass with a density function (mass per unit volume) $\rho(\vec{r})$ by applying these formulas to each differential of mass $dM = \rho(\vec{r})dV$ and integrating them over the whole distribution. Clearly it is easier to integrate up the scalar potential than the vector force field

$$\Phi(\vec{a})/G = - \int \frac{\rho(\vec{r})dV}{|\vec{r} - \vec{a}|}.$$

Next we expand the inverse factor of the relative distance in a Taylor series in \vec{r} about the origin and integrate the infinite series term by term. The Taylor expansion

$$\frac{1}{R} \equiv \frac{1}{|\vec{r} - \vec{a}|} = \sum_{p=0}^{\infty} \frac{1}{p!} \left(\partial_{i_1} \cdots \partial_{i_p} \frac{1}{R} \right) \Big|_{\vec{r}=0} x^{i_1} \cdots x^{i_p}$$

requires repeated differentiation involving the relation

$$\partial_i R = \partial_i [\delta_{mn}(x^m - a^m)(x^n - a^n)]^{1/2} = [\delta_{mn}(x^m - a^m)(x^n - a^n)]^{-1/2} \delta_{in}(x^n - a^n) = R^{-1}(x^i - a^i)$$

so the first two derivatives are

$$\partial_i R^{-1} = -R^{-2} \partial_i R = -R^{-3} (x^i - a^i)$$

and

$$\partial_j \partial_i R^{-1} = -\partial_j \left(\frac{x^i - a^i}{R^3} \right) = \dots = \frac{3(x^i - a^i)(x^j - a^j) - \delta^{ij} R^2}{R^5},$$

which has the property

$$\nabla^2 R^{-1} = \delta^{ij} \partial_i \partial_j R^{-1} = R^{-5} \delta_{ij} (3(x^i - a^i)(x^j - a^j) - \delta^{ij} R^2) = R^{-5} (3R^2 - 3R^2) = 0.$$

Evaluating these at the origin leads to $R|_{\vec{r}=0} = |\vec{a}|^{-1}$ and

$$\partial_i R^{-1}|_{\vec{r}=0} = |\vec{a}|^{-3} a^i, \quad \partial_j \partial_i R^{-1}|_{\vec{r}=0} = \frac{3a^i a^j - \delta^{ij} |\vec{a}|^2}{|\vec{a}|^5}, \quad \text{where } \delta^{ij} \partial_j \partial_i R^{-1}|_{\vec{r}=0} = 0.$$

Thus

$$\begin{aligned} -\Phi(\vec{a})/G &= \int \rho(\vec{r}) \sum_{p=0}^{\infty} \frac{1}{p!} \left(\partial_{i_1} \dots \partial_{i_p} \frac{1}{R} \right) \Big|_{\vec{r}=0} x^{i_1} \dots x^{i_p} dV. \\ &= \sum_{p=0}^{\infty} \frac{1}{p!} \left(\partial_{i_1} \dots \partial_{i_p} \frac{1}{R} \right) \Big|_{\vec{r}=0} \underbrace{\int \rho(\vec{r}) x^{i_1} \dots x^{i_p} dV}_{\equiv M^{i_1 \dots i_p}} \\ &= \frac{M}{|\vec{a}|} + \frac{M_i a^i}{|\vec{a}|^3} + \frac{M_{ij} (3a^i a^j - \delta^{ij} |\vec{a}|^2)}{2|\vec{a}|^5} + \dots \end{aligned}$$

defines an infinite family of symmetric tensors $M^{i_1 \dots i_p} = M^{(i_1 \dots i_p)}$ called the multipole moments of the mass distribution. The first multipole moment $M = \int \rho dV$ called the monopole is just the total mass. The second multipole moment $M^i = \int \rho x^i dV$ called the dipole defines the center of mass through $x_{CM}^i = M^i/M$, since the dipole vanishes with respect to the new origin at that point: $\int \rho (x^i - x_{CM}^i) dV = 0$. The third multipole moment $M^{ij} = \int \rho x^i x^j dV$ is called the quadrupole and is contracted with the tracefree coefficient $\partial_i \partial_j R^{-1}$ in the Taylor series, so only its tracefree part contributes

$$\begin{aligned} M_{ij} N^{ij} &= (M_{ij} - \frac{1}{3} \delta_{ij} \delta^{mn} M_{mn} + \frac{1}{3} \delta_{ij} \delta^{mn} M_{mn}) N^{ij} \\ &= (M_{ij} - \frac{1}{3} \delta_{ij} \delta^{mn} M_{mn}) N^{ij} + \frac{1}{3} \delta_{ij} \delta^{mn} M_{mn} N^{ij} \\ &= (M_{ij} - \frac{1}{3} \delta_{ij} \delta^{mn} M_{mn}) N^{ij} \end{aligned}$$

if N^{ij} is tracefree: $\delta_{ij} N^{ij} = 0$. This tracefree part of the second multipole is what is referred to as the quadrupole tensor

$$Q^{ij} = M_{ij} - \frac{1}{3} \delta_{ij} \delta^{mn} M_{mn}, \quad \delta_{ij} Q^{ij} = 0,$$

and in fact the same is true of higher moments, for which only the tracefree part contributes to the potential expansion since their coefficients are tracefree, a consequence of the fact that the function R^{-1} satisfies the Laplacian equation $\delta^{ij}\partial_i\partial_j R^{-1} = 0$.

Correspondingly the monopole term in the expansion of the potential is exactly the potential due to a point particle of mass M at the origin, while the dipole term is the additional part of the field which would result if one divided the mass into two equal parts $M/2$ displaced a small distance equidistant from the origin separated by the difference position vector M^i/M . The quadrupole term would instead result from dividing the mass into four equal parts equidistant from the origin such that the dipole moment is zero, etc. In this way the gravitational field of the entire actual continuous distribution of mass outside that distribution is represented by an equivalent point particle at the origin with complex limiting structure residing in the infinite set of multipole moments. It turns out that most significant contribution to gravitational radiation from a time-dependent mass distribution is proportional to the second time derivative of the quadrupole moment.

Our interest here is not in the details but only to see the context in which a large family of (tracefree) symmetric tensors play an important physical role in physical interactions. The same discussion applies to the electromagnetic field, where elementary particles are characterized by electric or magnetic monopole charge and electric or magnetic dipole moments, etc. In this case one can have a neutral particle (zero charge) with a dipole field since one has positive and negative charge and so can balance them, as in a water molecule, where the dipole moment of the electric field gives water many of its wonderful properties that sustain life.

Remark.

In the preceding derivation we essentially showed that the set of all functions

$$r^{-(2n+1)} \left(x^{i_1} \dots x^{i_n} - \frac{r^2}{3} \delta^{(i_1 i_2} x^{i_3} \dots x^{i_n)} \right) = r^{-(n+1)} \mathcal{Y}^{i_1 \dots i_n}$$

are solutions of Laplace's equation, which by definition are called harmonic functions. If express the position vector in spherical coordinates $\langle x^1, x^2, x^3 \rangle = \langle r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta \rangle$, then x^i/r is only a function of the angles (θ, ϕ) , i.e., is a function on the unit sphere, so the functions $\mathcal{Y}^{i_1 \dots i_n}$ are only functions on the unit sphere as well. Once we are familiar with spherical coordinates, we will return to this example to see the connection of these so called Cartesian harmonics with the spherical harmonics.

▲

Moments of inertia?

What are those crazy moment of inertia functions from multivariable calculus that were never explained? You just had to practice integration using their formulas perhaps, or that section was just completely ignored. Well, this is easy to explain and returns us to one of our familiar tensor examples.

In elementary physics you learn that the kinetic energy function is just one half the mass times the square of the speed $K = \frac{1}{2}mv^2 = \frac{1}{2}m\vec{v} \cdot \vec{v}$. This energy function is crucial in

understanding and solving the equations of motion for a body. Suppose our point mass body is rotating about a fixed axis with vector angular velocity $\vec{\Omega}$. If its position vector is $\vec{r} = \langle x^1, x^2, x^3 \rangle$ then you also learn that the velocity of the body is just the cross product $\vec{v} = \vec{\Omega} \times \vec{r}$, so the kinetic energy function is just the quadruple scalar product

$$\begin{aligned} K &= \frac{1}{2}m(\vec{\Omega} \times \vec{v}) \cdot (\vec{\Omega} \times \vec{v}) = \frac{1}{2}m(\vec{\Omega} \cdot \vec{\Omega} \vec{r} \cdot \vec{r} - \vec{\Omega} \cdot \vec{r} \vec{\Omega} \cdot \vec{r}) \\ &= \frac{1}{2}m(\delta^{ij}\delta_{mn}x^m x^n - x^i x^j)\Omega_i \Omega_j. \end{aligned}$$

Now suppose instead of a single point particle of mass m , we have localized rigidly rotating distribution of mass with density function ρ so that $dm = \rho dV$ represents a differential of mass. We must then integrate over the mass distribution to get the total kinetic energy

$$T = \frac{1}{2} \left(\int \rho(\delta^{ij}\delta_{mn}x^m x^n - x^i x^j) dV \right) \Omega_i \Omega_j \equiv \frac{1}{2} I^{ij} \Omega_i \Omega_j \equiv \frac{1}{2} I_{i'j'} \tilde{\Omega}^i \tilde{\Omega}^j.$$

Thus the kinetic energy is a quadratic form in the angular velocity whose coefficient matrix defines the components of the symmetric moment of inertia tensor. Reverting back to the more familiar notation (x, y, z) and $r = |\vec{r}| = (x^2 + y^2 + z^2)^{1/2}$, we have

$$\begin{aligned} I_{33} &= \int \rho(r^2 - z^2) dV = \int \rho(x^2 + y^2) dV, \\ I_{11} &= \int \rho(r^2 - x^2) dV = \int \rho(y^2 + z^2) dV, \\ I_{22} &= \int \rho(r^2 - y^2) dV = \int \rho(x^2 + z^2) dV, \\ I_{12} &= \int \rho(-xy) dV, \quad I_{13} = \int \rho(-xz) dV, \quad I_{23} = \int \rho(-yz) dV. \end{aligned}$$

For a homogeneous solid the density is constant: $\rho = \rho_0 = M/V$, where M is the total mass of the body and V is the total volume.

The diagonal components of this tensor are the integral against the square of the distance from the axis corresponding to the repeated index: I_{33} integrates the density against the square of the distance from the z -axis, etc. The offdiagonal components contain information necessary for axes not aligned with the coordinate axes in the same way that the offdiagonal components of the matrix of second partial derivatives are necessary to calculate the second partial derivative in a direction not aligned with the coordinate axes. However, for any surface of revolution about the z -axis, the off-diagonal components are zero by the reflection symmetry through the origin in the x - y plane, leaving only the diagonal components nonzero. When the moment of inertia tensor is diagonalized (which is always possible through a rotation since it is a symmetric matrix), the axes are referred to as the principal axes of the body. Thus for a surface of revolution about the vertical axis, the usual Cartesian axes are principal axes for the body.

Exercise 2.5.3.**moments of inertia of hemisphere**

Consider the upper hemisphere $x^2 + y^2 + z^2 = a^2, z \geq 0$ with a constant (homogenous) mass density $\rho_0 = M/V$ and volume $V = \frac{2}{3}\pi a^3$. Because of its rotational symmetry about the z axis, the x and y directions are equivalent, so $I_{11} = I_{22}$ and $I_{31} = I_{32}$, leaving only I_{33} and I_{12} remaining for a total of only 4 independent components of the 6 component matrix to be evaluated.

a) Use a computer algebra system to evaluate first the easier integral I_{33} in terms of M and a . (It would be smart to use cylindrical coordinates to iterate the integral.)

b) Now evaluate the harder one I_{11} .

c) While we are playing around, find the center of mass of the hemisphere by calculating the ratio of the dipole (along the z axis by symmetry) and the monopole (total mass) moments.

d) If you are only interested in a quickie, find the single independent component of the moment of inertial tensor for a whole sphere of mass M , a result found in any introductory physics textbook or at Wikipedia: $I = \frac{2}{5}Ma^2 = I_{11} = I_{22} = I_{33}$, where all the off-diagonal components are zero by symmetry. (Careful, now the mass M is twice the previous one, or the density is half the previous one.) Can you think why this result should be the same as the result for a hemisphere alone?

e) If we rotate the hemisphere by tilting the vertical axis, it would be difficult to evaluate these integrals to get the corresponding components of the moment of inertia tensor, but we don't need to since we can simple transform the components by the rotation to get its new more complicated matrix of components. Let's not do any calculation here and call it a day. (If you really insist, suppose we rotate by 45 degrees from the positive z -axis towards the positive x -axis. Calculate the new components of the tensor using an appropriate rotation matrix.) ■

Exercise 2.5.4.**moment of inertia for snow cone**

a) Use a computer algebra system to evaluate the moments of inertia tensor for a homogeneous snow cone of total mass M and volume V with vertex at the origin whose axis of symmetry is the z -axis, with base radius a and height h and lateral side length $R = \sqrt{a^2 + h^2}$, topped off by part of the sphere of radius R at the origin, namely the solid region inside the sphere of radius R above the plane $z = 0$ and inside the cone $z = (h/a)\sqrt{x^2 + y^2}$. The conical lateral surface can be described in cylindrical coordinates $(\rho, \phi, z) = (\sqrt{x^2 + y^2}, \arctan(y, x), z)$, where $\arctan(y, x)$ is a piecewise function

$$\arctan(y, x) = \begin{cases} \arctan(y/x) & x > 0 \text{ first, fourth quadrants} \\ \arctan(y/x) + \pi & x < 0, y > 0 \text{ second quadrant} \\ \arctan(y/x) - \pi & x < 0, y < 0 \text{ third quadrant} \end{cases}$$

by $z = \rho, 0 \leq \phi \leq 2\pi$ with $0 \leq \rho \leq a, 0 \leq \phi \leq 2\pi$, or in spherical coordinates $(r, \theta, \phi) = (\sqrt{x^2 + y^2 + z^2}, \arccos z/r, \phi)$ simply by $\theta = \arctan(h/a), 0 \leq r \leq R, 0 \leq \phi \leq 2\pi$. The cap can

be described in cylindrical coordinates by $z = \sqrt{R^2 - \rho^2}$, $0 \leq \phi \leq 2\pi$, $0 \leq \rho \leq a$ and in spherical coordinates by $r = R$, $0 \leq \theta \leq \arctan(h/a)$, $0 \leq \phi \leq 2\pi$. We will study these coordinates in detail in chapter 4, but based on your knowledge of multivariable calculus, evaluate the nonzero components $I_{11} = I_{22}$ and I_{33} . This can serve as a rotating top for a later problem.

b) Show that the result for the flat topped cone alone (a right circular cone) is

$$I_{11} = I_{22} = \frac{3}{5}M \left(\frac{a^2}{4} + h^2 \right), \quad I_{33} = \frac{3}{10}Ma^2$$

where M is the total mass.

c) Google “spinning top” for images of possible shapes for this old fashioned toy, which is described by the dynamics of a rigid body with one point fixed, as we will study once we have the appropriate tools.



Chapter 3

Time out

3.1 Whoa! Review of what we've done so far

Whoa! (This is a western movie cowboy expression for “stop,” usually directed at horses by their riders. Don't take my word for it, google the term.)

Okay, before going on, let's see what we've done so far to reassure ourselves that we have a general idea what we have done. Just let $V = \mathbb{R}^n$. We have been expanding on the following structure

- $(\mathbb{R}^n, “.”, \det)$
where
 - the vector space \mathbb{R}^n has the standard basis $\{e_i\}$,
 - “.” is the standard Euclidean dot product with components: $e_i \cdot e_j = \delta_{ij}$,
 - \det is the multilinear determinant function, whose absolute value gives the volume of n -parallelepipeds, and whose vanishing or nonvanishing tests the linear independence of a set of vectors, and whose sign tests the ordering of a set of vectors.
- We started down the road to tensors with the key notions
 - the dual space $(\mathbb{R}^n)^*$,
 - the dual basis $\{\omega^i\}$: just the “Cartesian coordinate functions” $\{x^i\}$,
 - duality $\omega^i(e_j) = \delta^i_j$: just the definition of the components of the standard basis vectors, equivalent to $x^i(0, \dots, 0, 1, 0, \dots, 0) = 0$ or 1 , the Cartesian components of the unit vectors along the axes: the i th component of e_i is 1 , the rest 0 ,
 - the dual of the dual space identification with the identity map

$$u(f) \equiv f(u) = f_i u^i, \quad e_j(\omega^i) \equiv \omega^i(e_j) = \delta^i_j.$$

So we know how to evaluate vectors on forms and vice versa, and the index pair contraction (summation) just symbolizes evaluation of a real-valued linear function of a vector or a covector when expressed in terms of its components

$$\begin{aligned} u &= u^i e_i, & u^i &= \omega^i(u), \\ f &= f_i \omega^i, & f_i &= e_i(f) = f(e_i). \end{aligned}$$

- We then generalize to multilinear real-valued functions accepting p covector arguments and q vector arguments

$$T^{(p,q)}(\mathbb{R}^n) \ni T = T_{j_1 \dots j_q}^{i_1 \dots i_p} e_{i_1} \otimes \dots \otimes e_{i_p} \otimes \omega^{j_1} \otimes \dots \otimes \omega^{j_q},$$

where the tensor product just holds the vectors and covectors apart in a certain order until they acquire arguments to be evaluated on, which results in a real number. The value of T is

$$\begin{aligned} T(f, g, \dots, u, v, \dots) &= T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p} e_{i_1}(f) e_{i_2}(g) \dots \omega^{j_1}(u) \omega^{j_2}(v) \dots \\ &\equiv T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p} f_{i_1} g_{i_2} \dots u^{j_1} v^{j_2} \dots \end{aligned}$$

The value of T on basis vectors and covectors defines its components

$$T(\omega^{i_1}, \omega^{i_2}, \dots, e_{j_1}, e_{j_2}, \dots) = T_{n_1 n_2 \dots n_q}^{m_1 m_2 \dots m_p} [\omega^{i_1}]_{m_1} [\omega^{i_2}]_{m_2} \dots [e_{j_1}]^{n_1} [e_{j_2}]^{n_2} \dots = T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p},$$

where

$$[\omega^{i_1}]_{m_1} = \delta^{i_1}_{m_1}, \quad [e_{j_1}]^{n_1} = \delta^{n_1}_{j_1}$$

define the components of the basis and dual basis vectors with respect to the same basis.

We agree to keep the covariant arguments first and the vector arguments last unless it is convenient to retain mixed positioning, as can occur with index raising and lowering. So for example

$$(S \otimes T)^{ij}_{kl} e_i \otimes e_j \otimes \omega^k \otimes \omega^l = (S^i_k e_i \otimes \omega^k) \otimes (T^j_l e_j \otimes \omega^l) = S^i_k T^j_l e_i \otimes e_j \otimes \omega^k \otimes \omega^l.$$

- The Kronecker delta δ^i_j is the component matrix of the identity tensor $I = \delta^j_i e_j \otimes \omega^i = e_i \otimes \omega^i$, which accepts a covector and vector argument and evaluates one against the other (in either order, since they are defined to be the same). This is just the natural evaluation of a covector on a vector to produce the value of a linear function, called contraction of the one up, one down index pair. This contraction operation can be extended to any pair of upper and lower indices on the same tensor or on tensor products of tensors. For example, the following contraction of the tensor product of two $\binom{1}{1}$ -tensors corresponds to matrix multiplication of their component matrices

$$\begin{aligned} S^i_k T^j_l e_i \otimes e_j \otimes \omega^k \otimes \omega^l &\mapsto S^i_k T^j_l e_i \otimes \omega^k(e_j) \omega^l \\ &= S^i_k T^j_l \delta^k_j e_i \otimes \omega^l = S^i_k T^k_l e_i \otimes \omega^l. \end{aligned}$$

These tensors describe linear transformations of \mathbb{R}^n into itself, with the identity tensor describing the identity transformation.

- The dot product

$$“\cdot” = G = G_{ij} \omega^i \otimes \omega^j = \delta_{ij} \omega^i \otimes \omega^j = \sum_{i=1}^n \omega^i \otimes \omega^i$$

is a symmetric tensor whose components in the standard basis numerically equal the Kronecker delta

$$G_{ij} = G(e_i, e_j) = e_i \cdot e_j = \delta_{ij}$$

so the self-inner product is a sum of squares

$$u \cdot u = G(u, u) = \sum_{i=1}^n [\omega^i(u)]^2 = \sum_{i=1}^n (u^i)^2$$

while the value on two vectors is the usual sum of products of like components

$$u \cdot v = G(u, v) = G_{ij} \omega^i(u) \otimes \omega^j(v) = G_{ij} u^i v^j = \delta_{ij} u^i v^j = \sum_{i=1}^n u^i v^i.$$

We interpret the dot product as a $\binom{0}{2}$ -tensor whose matrix of components in the standard basis equals the unit matrix, i.e., the standard basis is orthonormal. \mathbb{R}^n equipped with this natural inner product makes it into Euclidean space with its geometry of lengths and angles. We also considered more general inner products and in particular the Lorentz inner product on \mathbb{R}^n with diagonal metric matrix (η_{ij}) of all unit entries except the first which is -1 , extending the discussion to Minkowski spacetimes of any dimension. In our enthusiasm we also examined in Exercises two trace inner products on each space of $n \times n$ matrices, one of which is the Euclidean standard dot product on the equivalent \mathbb{R}^{n^2} space, and the other of which makes the symmetric subspace orthogonal to the antisymmetric subspace with positive self-inner products for the former space and negative self-inner products for the latter space. The latter inner product is actually important in the context of linear transformations.

- Determinants and antisymmetric tensors.

By linearity we can expand the determinant as a multilinear function of n vectors in \mathbb{R}^n

$$\begin{aligned}
 \det(u, v, \dots, w) &= \underbrace{\det(e_{i_1}, e_{i_2}, \dots, e_{i_n})}_{\equiv \epsilon_{i_1 i_2 \dots i_n}} u^{i_1} v^{i_2} \dots w^{i_n} && \text{(convenient definitions)} \\
 &\equiv \epsilon_{i_1 i_2 \dots i_n} \equiv \delta_{i_1 i_2 \dots i_n}^{12 \dots n} \\
 &= \epsilon_{i_1 i_2 \dots i_n} u^{i_1} v^{i_2} \dots w^{i_n} = \sum_{\sigma \in S_n} (\text{sgn } \sigma) u^{\sigma(1)} v^{\sigma(2)} \dots w^{\sigma(n)} && \text{(permutation definition of det)} \\
 &= \delta_{i_1 \dots i_n}^{1 \dots n} u^{i_1} v^{i_2} \dots w^{i_n} = n! u^{[1} v^2 \dots w^n] && \text{(definition of antisymmetrization)}
 \end{aligned}$$

so the determinant is the tensor

$$\det = \epsilon_{i_1 \dots i_n} \omega^{i_1} \otimes \dots \otimes \omega^{i_n} = n! \omega^{[1} \otimes \dots \otimes \omega^n].$$

This led to the introduction of the Levi-Civita permutation sign symbol $\epsilon_{i_1 \dots i_n}$ and the generalized Kronecker delta symbols.

Any antisymmetric $\binom{0}{n}$ -tensor T on \mathbb{R}^n (which just means it changes sign under the interchange of any two arguments) is completely determined by a single nonzero component

$$T_{i_1 \dots i_n} = t \epsilon_{i_1 \dots i_n}, \quad t = T_{12 \dots n}.$$

These are called n -forms, generalizing the single index 1-forms which are covectors. Positively (negatively) oriented n -forms are those for which $t > 0$ ($t < 0$). Correspondingly a basis $\{e_{i'}\}$ is called positively (negatively) oriented if $\det(e_{1'}, e_{2'}, \dots, e_{n'}) > 0$ (< 0). In \mathbb{R}^3 we refer to right-handed (+ orientation) and left-handed (− orientation) bases.

In fact

$$\underbrace{u^{i_1} v^{i_2} \dots w^{i_n}}_{n \text{ factors}} \equiv T^{i_1 i_2 \dots i_n}$$

is a $\binom{n}{0}$ -tensor with antisymmetric part $[\text{ALT}(T)]^{i_1 i_2 \dots i_n} = u^{[i_1} v^{i_2} \dots w^{i_n]}$ which has a single independent component

$$[\text{ALT}(T)]^{1 \dots n} = u^{[1} v^2 \dots w^n] = \frac{1}{n!} \det(u, v, \dots, w).$$

The antisymmetric tensor product turns out to be very useful. We'll get to it next. But notice that to get back to our useful determinant function we have to multiply the antisymmetric part by $n!$.

- Matrix notation.

We did all this stuff first in matrix notation. Let's go back to it to remind ourselves. The basic index suppression mechanism is the introduction of row and column matrices and matrix multiplication of adjacent column (left) and row (right) matrices

$$\begin{aligned} u \in \mathbb{R}^n &\mapsto \underline{u} = (u^i), & (\text{vector} = \text{column matrix}) \\ f \in [\mathbb{R}^n]^* &\mapsto \underline{f}^T = (f_i), & (\text{covector} = \text{row matrix}) \\ f(u) &= f_i u^i = \underline{f}^T \underline{u}, & (\text{matrix product gives evaluation}) \\ u \otimes f &\mapsto \underline{u} \underline{f}^T = (u^i f_j) & (\text{tensor product}) \end{aligned}$$

where to keep indices correctly positioned without indices, we need the transpose operation

$$\begin{aligned} u_i &\equiv \delta_{ij} u^j = \text{components of } u^\flat \in [\mathbb{R}^n]^* \mapsto \underline{u}^T, & (\text{column to row matrix}) \\ u^i &= \delta^{ij} u_j \quad \underline{u}^T \mapsto (\underline{u}^T)^T = \underline{u}. & (\text{row to column matrix}) \end{aligned}$$

The transpose corresponds to raising and lowering indices in this correspondence. The dot product is then

$$G(u, v) \equiv u \cdot v = \delta_{ij} u^i v^j = u^i \delta_{ij} v^j = \underline{u}^T \underline{I} \underline{v} = \underline{u}^T \underline{v}$$

and the multilinear determinant function is

$$\det(u, v, \dots, w) = \det(\underbrace{(\underline{u} | \underline{v} | \dots | \underline{w})}_{\text{matrix}}). \quad (\text{in the original matrix determinant sense})$$

- Change of basis.

Let $A = B^{-1}$ be an active (invertible) linear transformation of \mathbb{R}^n , under which all the points of the space u move to new positions $\bar{u} = A(u)$, namely $\bar{u}^i = A^i_j u^j$. If instead we apply its inverse B to the basis vectors to define new basis vectors with primed indices: $e_{i'} = A^{-1j}_i e_j = B^j_i e_j$, we obtain a passive coordinate transformation in which the points $u = u^i e_i = u^{i'} e_{i'}$ remain fixed but their components with respect to the basis change since the basis changes: $u^{i'} = A^i_j u^j$.

Thus the components of u with respect to $\{e_{i'}\}$ equal the components of \bar{u} with respect to $\{e_i\}$ as should be clear from Fig. 3.1: $u^{i'} = \bar{u}^i$. In words, the components of a given vector with respect to the new basis are the same as the components of the new actively transformed vector with respect to the old basis. We can work through the basis and dual basis point of view for any vectors space V to see how the components behave this way.

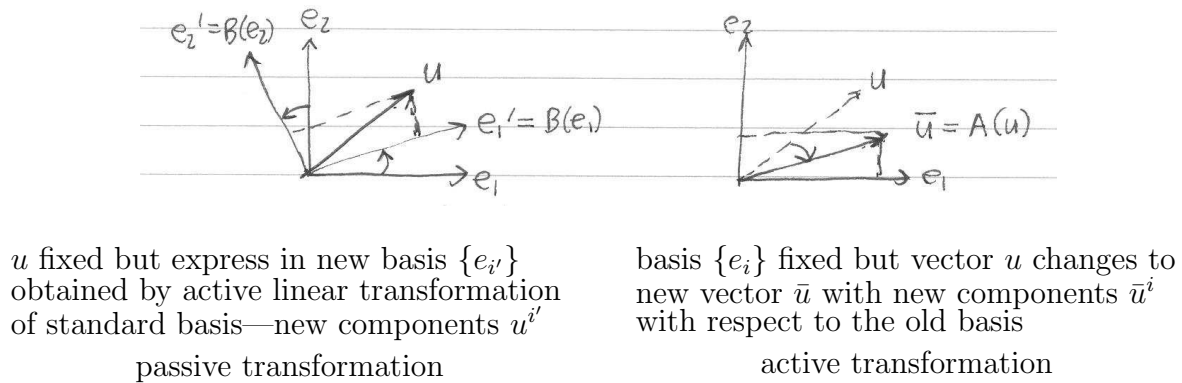


Figure 3.1: The idea of a passive linear transformation versus an active linear transformation.

Consider the transformation and its inverse for both V and V^* interpreted as a change of the basis.

$$\begin{aligned} e_{i'} &= B^j_i e_j = A^{-1j}_i e_j, & e_i &= B^{-1j}_i e_{j'} = A^j_i e_{j'}, \\ \omega^{i'} &= B^{-1i}_j \omega^j = A^i_j \omega^j, & \omega^i &= B^i_j \omega^{j'} = A^{-1i}_j \omega^{j'}. \end{aligned}$$

Notice the row-column symmetry in these relations. The columns of \underline{B} are the old components of the new basis vectors, while the rows of its inverse \underline{A} are the old components of the new dual basis. Similarly the columns of \underline{A} are the new components of the old basis vectors, while the rows of \underline{B} are the new components of the old dual basis.

Then evaluating the transformation of dual basis relation on u

$$\omega^{i'}(u) = u^{i'}, \quad \omega^j(u) = u^j$$

gives the transformation of its components under the change of basis

$$u^{i'} = A^i_j u^j$$

or in matrix form

$$\underline{u}' = \underline{A} \underline{u} \quad \text{or} \quad \underline{u} = \underline{A}^{-1} \underline{u}'.$$

But by definition $A^i_j u^j$ are the components of the active linear transformation of u by A , i.e., $u^{i'} = \bar{u}^i$.

We can use these relations to re-express the dot product on \mathbb{R}^n

$$\begin{aligned} u \cdot v &= \underline{u}^T \underline{I} \underline{v} = (\underline{A}^{-1} \underline{u}')^T \underline{I} (\underline{A}^{-1} \underline{v}') = (\underline{u}')^T (\underline{A}^{-1})^T \underline{I} \underline{A}^{-1} \underline{v}' = (\underline{u}')^T \underline{G}' \underline{v}', \\ G_{i'j'} &= A^{-1m}_i \delta_{mn} A^{-1n}_j = A^{-1m}_i A^{-1n}_j \delta_{mn}, \end{aligned}$$

which is the “tensor transformation law” for a $\binom{0}{2}$ -tensor. Alternatively, one has

$$G_{i'j'} = e_{i'} \cdot e_{j'} = A^{-1m}_i A^{-1n}_j e_m \cdot e_n = A^{-1m}_i A^{-1n}_j \delta_{mn}.$$

The orthogonal matrices leave the components of the dot product unchanged, taking the orthonormal standard basis of \mathbb{R}^n into new orthonormal bases. For other inner products this defines the associated orthogonal matrix group, like the Lorentz group for the Minkowski spacetimes. We saw that the Lie algebra of these matrix groups satisfied a simply condition, that the index-lowered matrices were just antisymmetric.

- The determinant as a tensor rather than a function on matrices.

From its ordinary interpretation in terms of components of vectors with respect to the standard basis of \mathbb{R}^n , the determinant defines a $\binom{0}{n}$ -tensor

$$\begin{aligned}\det(u, v, \dots, w) &= \det(\langle \underline{u} | \underline{v} | \dots | \underline{w} \rangle) \\ &= \det(\langle \underline{A}^{-1} \underline{u}' | \underline{A}^{-1} \underline{v}' | \dots | \underline{A}^{-1} \underline{w}' \rangle) \\ &= \det(\underline{A}^{-1} \langle \underline{u}' | \underline{v}' | \dots | \underline{w}' \rangle) \\ &= \det(\underline{A}^{-1}) \det(\langle \underline{u}' | \underline{v}' | \dots | \underline{w}' \rangle) \\ &= [\det(\underline{A}^{-1}) \epsilon_{i_1 \dots i_n}] u^{i_1'} \dots w^{i_n'}\end{aligned}$$

so as a tensor it has components $\epsilon_{i_1 \dots i_n}$ with respect to the standard basis but not in general

$$\begin{aligned}\det &= (\det \underline{A}^{-1}) \epsilon_{i_1 \dots i_n} \omega^{i_1'} \otimes \dots \otimes \omega^{i_n'} &&= \epsilon_{i_1 \dots i_n} \omega^{i_1} \otimes \dots \otimes \omega^{i_n} \\ &= (\det \underline{A}^{-1}) \underbrace{n! \omega^{[1'} \otimes \dots \otimes \omega^{n']}}_{\text{takes det of new component matrix}} &&= n! \omega^{[1} \otimes \dots \otimes \omega^{n]}.\end{aligned}$$

Thus the factor $\det \underline{A}^{-1} = \det \underline{B}$ corrects the value of the determinant of the matrix of new components to give the value of the determinant tensor on \mathbb{R}^n , which is independent of basis. Another way of looking at this is that the determinant of the new matrix gives the volume of the n -parallelepiped associated with the n vectors relative to the n -parallelepiped of the new basis vectors, but they already have volume amplified by the factor $\det \underline{A}^{-1} = \det \underline{B}$ relative to the standard basis vectors, so the product of the correction factor and the determinant of the matrix of new components gives the absolute volume of the former parallelepiped with respect to the Euclidean geometry of \mathbb{R}^n (modulo sign changes that come from the sign of $\det \underline{A}^{-1}$).

Example 3.1.1. Consider the vectors $u_1 = \langle -1, 2 \rangle = b_1 + b_2$, $u_2 = \langle -3, 0 \rangle = -b_1 + b_2$ which form the parallelogram shown in Fig. 3.2, where $b_1 = \langle 1, 1 \rangle$ and $b_2 = \langle -2, 1 \rangle$. The factoring law for determinants leads to the following relationship between the determinants of the matrices of old and new components of the pair of vectors

$$\begin{aligned}\det \langle \underline{u}_1 | \underline{u}_2 \rangle &= \underbrace{\begin{vmatrix} -1 & -3 \\ 2 & 0 \end{vmatrix}}_6 = |\langle \underline{B} \underline{u}'_1 | \underline{B} \underline{u}'_2 \rangle| = |\underline{B} \langle \underline{u}'_1 | \underline{u}'_2 \rangle| \\ &= \det(\underline{B}) |\langle \underline{u}'_1 | \underline{u}'_2 \rangle| = \underbrace{\begin{vmatrix} 1 & -2 \\ 1 & 1 \end{vmatrix}}_3 \underbrace{\begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix}}_2.\end{aligned}$$

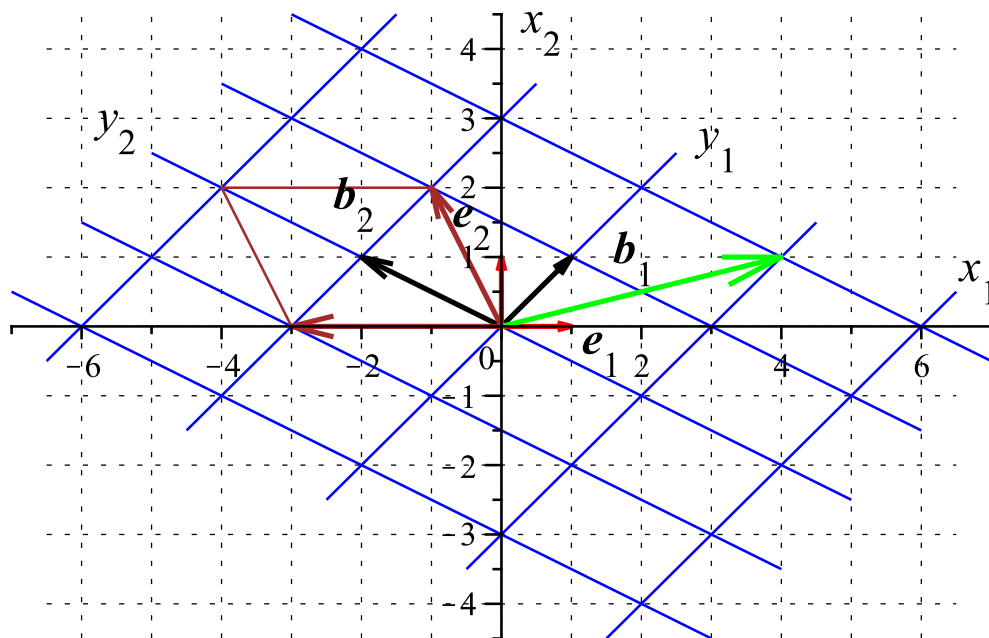


Figure 3.2: The new coordinate grid of the basis $b_1 = \langle 1, 1 \rangle$ and $b_2 = \langle -2, 1 \rangle$. Modulo signs the true area of the parallelogram shown (just the determinant of the matrix of its old components) is its area with respect to the new grid (just the determinant of the matrix of its new components), times the determinant of the matrix of new basis vectors, which is the area of the unit parallelogram of the new grid with respect to the old grid which defines area in terms of its orthonormal coordinates. This is the geometric interpretation of the determinant product rule: the determinant of a product of two matrices is the product of the determinants. The components with respect to either the original or new grid of the vectors $u_1 = \langle -1, 2 \rangle = b_1 + b_2$, $u_2 = \langle -3, 0 \rangle = -b_1 + b_2$, $u_3 = \langle 4, 1 \rangle = 2b_1 - b_2$ are easily read off from these grids.

The factor $\det \underline{B} = (\det \underline{A})^{-1} = 3$ is the area amplification factor between the grids.

Similarly to calculate the dot product of the two vectors $u_1 \cdot u_2 = (\underline{u}_1)^T \underline{u}_2$ we must take into account the dot products of the new basis vectors

$$\begin{aligned} 3 &= \langle -1, 2 \rangle^T \langle -3, 0 \rangle = (\underline{u}_1)^T \underline{u}_2 = (\underline{B} \underline{u}'_1)^T (\underline{B} \underline{u}'_2) = (\underline{u}'_1)^T (\underline{B}^T \underline{B}) \underline{u}'_2 = (\underline{u}'_1)^T \underline{G}' \underline{u}'_2 \\ &= \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = 3. \end{aligned}$$

Thus one has the new matrix of inner products

$$\underline{G}' = \underline{B}^T \underline{I} \underline{B} = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}$$

of the nonorthonormal basis, which is the matrix form of the transformation law for the $\binom{0}{2}$ -tensor G whose matrix $\underline{G} = \underline{I}$ is the unit matrix with respect to the standard basis. The connection between this and the area amplification factor is the relation

$$\det \underline{G}' = \det(\underline{B}^T \underline{B}) = (\det \underline{B})^2 \rightarrow |\det \underline{G}'|^{1/2} = |\det \underline{B}| = |\det \underline{A}^{-1}| |\det \underline{G}|,$$

namely the square root of the absolute value of the metric determinant (9 in this case) is the area amplification factor (3 in this case). In fact this last relationship can be interpreted as the transformation law of a weight 1 scalar density, with an extra wrinkle since it is the absolute value of determinant of the basis changing matrix rather than the determinant itself, so we invent a new name: oriented weight 1 scalar density. [This just means that it has an extra sign factor in the transformation law equal to the sign of the determinant of the transformation matrix, in order to make the new component come out positive when that determinant is negative.]

The new matrix \underline{G}' of components of the dot product is necessary to lower indices in the new coordinates, namely $(\underline{u}') \rightarrow \underline{G}'(\underline{u}')$. For example lowering the index on the vectors u_1 and u_2 leads to

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -3 \\ 6 \end{pmatrix}.$$

In other words the covector index-lowered from the vector $u_2 = \langle -3, 0 \rangle$ is $-3x^1 = -3y^1 + 6y^2$, while the covector index-lowered from $u_1 = \langle -1, 2 \rangle$ is $-x^1 + 2x^2 = y^1 + 4y^2$. Obtaining these from the transformation law instead of by index lowering from the vector components we have the same component results of course

$$v_{i'} = v_j B^j_{i'} : \quad \begin{pmatrix} -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 \end{pmatrix}, \quad \begin{pmatrix} -3 & 0 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} -3 & 6 \end{pmatrix}.$$

□

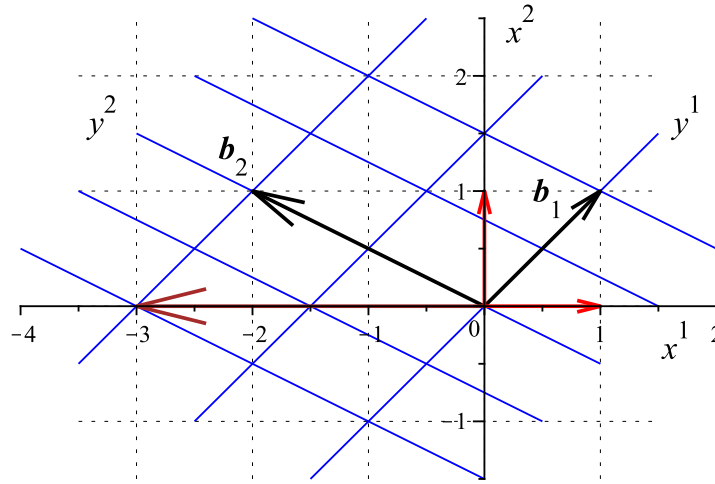


Figure 3.3: Grid for use in drawing in the representative line for the covector corresponding to the vector $u_1 = \langle -3, 0 \rangle = -2b_1 + b_2$ and building up the same covector representative line $3x^1 = 1 = 3(-y^1 + 2y^2)$ in the new grid as described in the text.

Exercise 3.1.1. covector addition

Using the grid in Fig. 3.3, first draw in the representative covector lines $-3x^1 = 1, 0$. Next draw in the covector lines $-y^1 = 1, 0$ and $2y^2 = 1, 0$. Next using the covector cross-diagonal parallelogram addition, draw in the covector lines $-y^1 + 2y^2 = 1, 0$ and finally scale it up by a factor of 3 using the geometric interpretation of scalar multiplication for covectors to obtain the lines $3(-y^1 + 2y^2) = 1, 0$ which should agree with the starting covector line pair.

■

• Linear transformations.

Suppose we have any linear transformation L of \mathbb{R}^n into itself. In the various notations

$$u \mapsto L(u), \quad u^i \mapsto L^i_j u^j, \quad \underline{u} \mapsto \underline{L} \underline{u},$$

where $L^i_j = \omega^i(L(e_j))$ or equivalently $L(e_j) = L^i_j e_i$. However, a vector-valued linear function of vectors can be identified with a $\binom{1}{1}$ -tensor in a natural way simply by adding an extra covector argument to the linear function L : define \mathbb{L} by

$$\mathbb{L}(f, u) = f_i L^i_j u^j = \underline{f} \underline{L} \underline{u}$$

so that $L = \mathbb{L}(, u)$ is the partial evaluation of \mathbb{L} (thinking of the vector $L(u)$ as waiting for a covector argument). Then

$$\mathbb{L} = L^i_j e_i \otimes \omega^j, \quad L^i_j = \mathbb{L}(\omega^i, e_j)$$

so multiplying both \underline{u} and $\underline{L(u)}$ by \underline{A} to obtain their new components

$$\underline{A}(\underline{u} \mapsto \underline{L u}) \longrightarrow \underbrace{\underline{A} u}_{u'} \mapsto \underbrace{\underline{A} \underline{L u}}_{\underline{A}^{-1} u'}$$

so

$$u' \mapsto (\underline{A} \underline{L} \underline{A}^{-1}) \underline{u}', \quad L^{i'}_{j'} = A^i_m L^m_n A^{-1n}_j = A^i_m A^{-1n}_j L^m_n$$

so we get the “tensor-transformation” law for a $\binom{1}{1}$ -tensor.

Both inner products (which are $\binom{0}{2}$ -tensors) and linear transformations (which are $\binom{1}{1}$ -tensors) are represented by matrices, but their different mathematical structure is reflected in the different matrix transformation laws. Our index notation makes these differences explicit. To go beyond objects with two indices, we need index notation to handle them intelligently.

Remark.

What is the difference between δ_{ij} and δ^i_j ?

It depends on the interpretation. The values for each index pair (i, j) are identical *BUT* we interpret δ^i_j as the components

$$\delta^i_j = \text{EVAL}(\omega^i, e_j) = \text{IDENTITY}(\omega^i, e_j)$$

of a tensor $e_i \otimes \omega^i$ which does not depend on the choice of basis, i.e., has the same components no matter what basis we choose, *while* $\delta_{ij} = G(e_i, e_j)$ are the components in a special basis of a given tensor G (independent of the choice of basis) but which change under a general change of basis—alternatively the component values δ_{ij} in every choice of basis do not define a single tensor but a family of different tensors. In each frame this tensor is $\delta_{ij} \omega^i \otimes \omega^j = \omega^1 \otimes \omega^1 + \dots \omega^n \otimes \omega^n$, which of course represents a different tensor in different bases unless the two are related by an orthogonal matrix change of basis. The same interpretation applies to $\epsilon_{i_1 \dots i_n}$. It defines a different tensor $\epsilon_{i_1 \dots i_n} \omega^{i_1} \otimes \dots \otimes \omega^{i_n}$ in each basis.

Are δ^1_i the components of a covector?

Again it depends. Since $\omega^1 = \delta^1_i \omega^i$, these are the components of the first covector in our chosen basis, so if we change the basis, the covector ω^1 will no longer (in general) have such simple components in terms of the new basis, but still it defines a unique tensor, namely ω^1 . On the other hand the numerical values δ^1_i define different covectors in different bases. One really needs to qualify our opening question so that one of these two interpretations is clear. Then we can answer the question.

▲

- The dot product, duality, and index shifting.

These operations can be extended in a natural way to each space $T^{(p,q)}(\mathbb{R}^n)$ which is itself a Euclidean vector space isomorphic to $(\mathbb{R}^{n^{p+q}}, \cdot)$. Such tensors have $p+q$ indices, and

n choices for each index value so there are n^{p+q} independent components. Listing them in a certain order establishes an isomorphism with $\mathbb{R}^{n^{p+q}}$, which has its own dot product. This dot product and index shifting correspond exactly to the ones we have established for $\binom{p}{q}$ -tensors.

For example, let $U = T^{(2,0)}(\mathbb{R}^n)$, with basis $\{e_i \otimes e_j\} \equiv \{E_{ij}\}$, and $T = T^{ij}e_i \otimes e_j = T^{ij}E_{ij}$. Instead of using an index, say A, B, C, \dots which runs from 1 to n^2 , we can use n^2 index pairs (i, j) to label the distinct basis vectors in U and also the components of vectors in U .

I claim the dual basis can be identified with $W^{ij} \equiv \omega^i \otimes \omega^j$ and the dual space with $U^* = T^{(0,2)}(\mathbb{R}^n)$

$$W^{ij}(E_{mn}) = [\omega^i \otimes \omega^j](e_m \otimes e_n) \equiv \omega^i(e_m)\omega^j(e_n) = \delta^i_m \delta^j_n \equiv I^{ij}_{mn}.$$

I^{ij}_{mn} are the components of the Kronecker delta on U in this notation

$$I^{ij}_{mn} = \begin{cases} 1, & \text{if } (i, j) = (m, n), \\ 0, & \text{otherwise.} \end{cases}$$

So $\mathcal{F} = \mathcal{F}_{ij}W^{ij}$ is a “covector” with the evaluation given by

$$\mathcal{F}(T) = \mathcal{F}_{ij}T^{ij}.$$

Let us define

$$E_{ij} \cdot E_{mn} = \delta_{im}\delta_{jn} \equiv \delta_{ij,mn} \equiv \mathcal{G}_{ij,mn}.$$

Then this corresponds to an inner product tensor

$$\begin{aligned} \mathcal{G} &= \mathcal{G}_{ij,mn}W^{ij} \otimes W^{mn} \\ \mathcal{G}(T, u) &= \mathcal{G}_{ij,mn}T^{ij}u^{mn} = \delta_{im}\delta_{jn}T^{ij}u^{mn} = T^{ij}u_{ij} \end{aligned}$$

which is how we defined the inner product previously.

Note that the “ \otimes ” in $\mathcal{G} = \mathcal{G}_{ij,mn}W^{ij} \otimes W^{mn}$ is the tensor product for U , not \mathbb{R}^n , since it is holding the “covectors” (with respect to U) W^{ij} and W^{mn} apart until they accept “vector” (with respect to U) arguments, but this distinction doesn’t matter. The $\binom{0}{4}$ -tensor

$$\mathcal{G} = \mathcal{G}_{ij,mn}\omega^i \otimes \omega^j \otimes \omega^m \otimes \omega^n$$

can be contracted against two $\binom{2}{0}$ -tensors to yield a real number which is exactly $\mathcal{G}(T, u) = \mathcal{G}_{ij,mn}T^{ij}u^{mn}$. Our notation identifies these different interpretations. We just need to allow for “contraction” of any number of indices of a tensor with those of another.

For example, what “contractions” are allowed between T^{ij}_{klm} and S^{pq}_r ? First define

$$[T \otimes S]^{ijpq}_{klmn} = T^{ij}_{klm}S^{pq}_r.$$

We can then contract any subset of contravariant indices with any subset of covariant indices of the same number, to yield tensors of various ranks less than $5 + 3$

$$T^{ik}{}_{j\ell m} S^{\ell m}{}_k \sim \binom{1}{1}\text{-tensor},$$

for example. Furthermore index-shifting on U corresponds to index-shifting of pairs of indices with the inner product G on \mathbb{R}^n

$$T = T^{ij} E_{ij} \longrightarrow T^\flat = T_{ij} W^{ij}$$

where

$$T_{ij} = \mathcal{G}_{ij, mn} T^{mn} = \delta_{im} \delta_{jn} T^{mn}$$

as was already defined above. Generalizations of this inner product on the space of symmetric 2-index tensors turns out to be extremely important in understanding the dynamics of general relativity. Exercise 3.1.2 explores this.

We can repeat this discussion for all the tensor spaces and their “dual” tensor spaces

$$U = T^{(p,q)}(\mathbb{R}^n) \in T$$

and

$$U^* = T^{(q,p)}(\mathbb{R}^n) \in S$$

with natural evaluation of one on the other defined by

$$T(S) = T^{i_1 \dots i_p}{}_{j_1 \dots j_q} S^{j_1 \dots j_q}{}_{i_1 \dots i_p}, \text{ etc.}$$

Looks like I snuck in a few new thoughts on you in this review of our progress so far. Anyway, our extended Exercise 1.6.9 with the vector space $V = gl(n, \mathbb{R}^n)$ of $n \times n$ matrices at the end of Chapter 1 develops matrix operations relevant to both linear transformations which are $\binom{1}{1}$ -tensors and to $\binom{0}{2}$ -tensors and $\binom{2}{0}$ -tensors interpreted as linear maps between \mathbb{R}^n and its dual space. In that discussion, our starting vector space V has the natural basis $\underline{e}^i{}_j$ such that $\underline{A} = A^i{}_j \underline{e}^j{}_i$. Here we have component indices associated with V which are “1 up, 1 down” index pairs taken from the integers from 1 to n instead of single indices taken from the integers from 1 to n^2 . The correspondence

$$\underline{A} = A^i{}_j \underline{e}^j{}_i \longrightarrow A = A^i{}_j e_i \otimes \omega^j$$

maps the standard matrix basis $\{\underline{e}^j{}_i\}$ onto the basis

$$\{E^j{}_i\} = \{e_i \otimes \omega^j\} \text{ of } T^{(1,1)}(\mathbb{R}^n),$$

which is a natural identification. The moral of the story is to stay flexible with notation to allow it to fit the circumstances.

Looking ahead for applications to differential geometry, once we have identified any starting basis $\{e_i\}$ of each tangent space to a curved space M , i.e., a field of bases, one defined at each point of M , we can then carry over all the tensor algebra we have developed for \mathbb{R}^n in the standard basis to each such tangent space. This will lead to tensor fields over the curved space itself whose interpretation at each tangent space derives from our preceding discussion of tensors over a single vector space.

Exercise 3.1.2.**deWitt inner product for symmetric tensors**

Inner products on the space of 6-dimensional space of symmetric 2-index tensors over \mathbb{R}^3 turns out to be of physical interest for time-dependent gravitational fields and quantum gravity. Suppose we have a positive-definite inner product $G = G_{ij}\omega^i \otimes \omega^j$ on \mathbb{R}^3 that is used to raise and lower indices. Define the de Witt inner product of $J = J_{ij}\omega^i \otimes \omega^j$ and $K = J_{ij}\omega^i \otimes \omega^j$ by

$$G_{\text{dW}}(J, K) = \text{Tr}(\underline{J} \underline{K}) - \text{Tr}(\underline{J}) \text{Tr}(\underline{K}) = G_{\text{dS}}^{ijkl} J_{ij} K_{kl}.$$

a) Express the components G_{dS}^{ijkl} in terms of G^{ij} .

b) Show that if we decompose the $\binom{0}{2}$ -tensors into their pure trace and tracefree parts with respect to the inner product G by

$$K_{ij}^{\text{tr}} = \frac{1}{3} K^k{}_k G_{ij}, \quad K_{ij}^{\text{trfree}} = K_{ij} - \frac{1}{3} K^k{}_k G_{ij}$$

which corresponds exactly to multiples of the identity matrix and tracefree matrices when the indices are raised to the mixed position $K^i{}_j$. Show that self-inner products are negative for the pure trace tensors but positive for the remaining tracefree tensors, thus defining a Lorentz inner product on the space of symmetric tensors.

c) Show that for $G_{ij} = \delta_{ij}$ so that index raising and lowering does not change the component values, the tensors with the following component matrices form an orthogonal basis of the diagonal subspace space with the same self-inner products apart from sign

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \sqrt{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \sqrt{3} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \sqrt{3} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \sqrt{3} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Use a computer algebra system to do the trace product evaluations. What is the common factor we have to divide these matrices by to normalize them and arrive at an orthonormal basis?

d) Show that the contravariant inner product for symmetric $\binom{2}{0}$ -tensors corresponds to

$$G_{\text{dW}}^{-1}(J, K) = \text{Tr}(\underline{J} \underline{K}) - \frac{1}{2} \text{Tr}(\underline{J}) \text{Tr}(\underline{K}) = G_{\text{dW}}^{-1ijkl} J^{ij} K^{kl},$$

namely

$$G_{\text{dW}}^{-1ijkl} G_{\text{dW}mnkl} = \delta^i_{(k} \delta^j_{l)}.$$

e) Show that index lowering from $\binom{0}{2}$ -tensors to $\binom{2}{0}$ -tensors with this inner product has the following action

$$J^{ij} = G_{\text{dW}}^{ijkl} K_{kl} = K^{ij} - K^k{}_k G^{ij}.$$

This is “index lowering” in the sense that $\binom{0}{2}$ -tensors are a vector space whose contravariant up index is instead represented by a covariant symmetric index pair, which would become a single contravariant index if we listed the basis tensors in the usual single index format

■

Chapter 4

Antisymmetric tensors, subspaces and measure

Antisymmetric tensors play a fundamental role not only in how we measure area, volume, and higher dimensional analogs of these quantities but in symmetry properties of the geometry of lengths and angles. Determinant theory quantifies the former, giving a geometric interpretation to arbitrary rank antisymmetric tensors, while providing a tool to characterize subspaces of vector spaces associated with these measures. In a completely different context differential rotations and pseudo-rotations which characterize the symmetry properties of flat Euclidean or non-Euclidean geometries are intimately associated with second rank antisymmetric tensors. In this chapter we delve into both topics in some detail.

4.1 Determinants gone wild

The theory of determinants is a classic area of mathematics which even I did not appreciate until writing this book. In modern applied mathematics with determinants so easily evaluated with technology, it is enough to understand how to evaluate determinants through row reduction operations, which then explain why the determinant being zero or nonzero is important to solving a linear system with a given square coefficient matrix. However, in pre-technology days cofactor expansions of the determinant were used for evaluation purposes. Minors of a determinant are introduced in which one deletes a row and column to get a determinant of one order less with a certain alternating sign which is associated with it according to its position in the matrix, the product of the minor and this sign being called the associated cofactor. As we will see, all of this structure in modern language will define the wedge product algebra on the spaces of p -vectors for all $0 \leq p \leq n$ that is essential for integration over regions of an n -dimensional space where these antisymmetric tensors are called differential forms. We have already encountered these in line integrals in multivariable calculus where we learned how to integrate differentials of functions and linear combinations of the differentials of the Cartesian coordinates over curves. A serious multivariable calculus course would also introduce surface integrals. We will get to these issues in Part 2.

It helps to start with a familiar concrete example in \mathbb{R}^3 to motivate the discussion for any dimension.

Example 4.1.1. Consider the following string of equalities which define the minors and cofactors a 3×3 matrix

$$\begin{aligned} \det(\underline{A}) &= \begin{vmatrix} A^1_1 & A^1_2 & A^1_3 \\ A^2_1 & A^2_2 & A^2_3 \\ A^3_1 & A^3_2 & A^3_3 \end{vmatrix} = \det(\langle \underline{A}_1 | \underline{A}_2 | \underline{A}_3 \rangle) \\ &= \det(\underline{A}_1, \underline{A}_2, \underline{A}_3) \quad (\text{as a linear function on column matrices}) \\ &= A^1_1(A^2_2A^3_3 - A^3_2A^2_3) + A^2_1(A^3_3A^1_3 - A^1_2A^3_3) + A^3_1(A^1_2A^2_3 - A^2_2A^1_3) \\ &= A^1_1M^1_1 - A^2_1M^1_2 + A^3_1M^1_3 = A^1_1C^1_1 + A^2_1C^1_2 + A^3_1C^1_3, \end{aligned}$$

where the minor M^j_i of entry A^i_j is the 2×2 determinant obtained by crossing out the j th column and i th row of the full determinant array $|\underline{A}|$ (note the transposition of index position from matrix to minor) and the cofactor $C^j_i = (-1)^{i+j}M^j_i$ of that entry A^i_j differs only by a sign which alternates between 1 and -1 as one goes along any row or column. The last line called the expansion of the determinant along the first column is the linear combination of the entries of the first column with the corresponding vector of cofactors. One can easily write the original 6 terms of the determinant in 2 more such column expansions along the remaining columns and as 3 corresponding such row expansions, all equal to the same determinant. Continuing in this fashion, each of the minors can be evaluated in terms of similar column or row expansions in terms of 1×1 determinants (now just the entries themselves which are left after removing a row and column from the 2×2 submatrices).

Now look at the last equality of the above displayed equation. It says that the determinant function, unevaluated on the first column matrix argument so that it becomes a covector:

$\det(\quad, \underline{A}_2, \underline{A}_3)$, is represented by the component covector $\langle C^1_1 | C^1_2 | C^1_3 \rangle$ since these are the corresponding coefficients of the entries of the first column when one evaluates that covector on it. Now consider the matrix with the first column removed and delete one row at a time from top to bottom to obtain three 2×2 determinants and alternate the sign to obtain the cofactors of the original matrix along the first column

$$\begin{pmatrix} A^1_2 & A^1_3 \\ A^2_2 & A^2_3 \\ A^3_2 & A^3_3 \end{pmatrix} \rightarrow \begin{vmatrix} A^2_2 & A^2_3 \\ A^3_2 & A^3_3 \end{vmatrix}, - \begin{vmatrix} A^1_2 & A^1_3 \\ A^3_2 & A^3_3 \end{vmatrix}, \begin{vmatrix} A^1_2 & A^1_3 \\ A^2_2 & A^2_3 \end{vmatrix} = C^1_1, C^1_2, C^1_3$$

However, if we start with a 3×2 matrix, there is no original 3×3 matrix to identify these determinants as cofactors. Instead they represent the cofactors of any matrix we obtain by inserting another column at the beginning. So taking cofactors of a matrix of two three-entry columns is equivalent to dealing with the determinant tensor with its first argument left unevaluated. If all these 2×2 determinants vanish, it means no matter what column we insert at the beginning, the set of columns is linearly dependent because the resulting 3×3 determinant must be zero, so at least one such 2×2 determinant must be nonzero to establish linear independence of those two column matrices. If we remove another column, we are down to a single column, where at least one entry must be nonzero for it to be linearly independent (nonzero).

Note that to find the equation of the plane spanned by the final two columns we can simply evaluate the unevaluated argument on a variable vector $\langle x^1, x^2, x^3 \rangle$

$$0 = \begin{vmatrix} x^1 & A^1_2 & A^1_3 \\ x^2 & A^2_2 & A^2_3 \\ x^3 & A^3_2 & A^3_3 \end{vmatrix} = C^1_1 x^1 + C^1_2 x^2 + C^1_3 x^3.$$

Recall that this equals the triple scalar product of the 3 columns, interpreted as the signed volume of the parallelepiped formed by the 3 column vectors, so the covector $\langle C^1_1 | C^1_2 | C^1_3 \rangle = \underline{A}_2 \times \underline{A}_3$ has a magnitude with respect to the dot product which is equal to the area spanned by the two column vectors spanning the subspace. This follows since the signed volume reduces to the signed area of the parallelogram formed by these two vectors if the third column is a unit vector.

We can go one further step to examine the case of two unevaluated arguments of the determinant

$$0 = \begin{vmatrix} y^1 & x^1 & A^1_3 \\ y^2 & x^2 & A^2_3 \\ y^3 & x^3 & A^3_3 \end{vmatrix} = y^1 \begin{vmatrix} x^2 & A^2_3 \\ x^3 & A^3_3 \end{vmatrix} - y^1 \begin{vmatrix} x^1 & A^1_3 \\ x^3 & A^3_3 \end{vmatrix} + y^1 \begin{vmatrix} x^1 & A^1_3 \\ x^2 & A^2_3 \end{vmatrix}.$$

Since this must be zero for any y^i , the three coefficients must all be zero if this is to be zero for a given x^i

$$0 = \begin{vmatrix} x^2 & A^2_3 \\ x^3 & A^3_3 \end{vmatrix} = \begin{vmatrix} x^1 & A^1_3 \\ x^3 & A^3_3 \end{vmatrix} = \begin{vmatrix} x^1 & A^1_3 \\ x^2 & A^2_3 \end{vmatrix}.$$

Introducing the abbreviated notation for the last remaining vector argument $\langle a^1, a^2, a^3 \rangle = \langle A^1_3, A^2_3, A^3_3 \rangle$, these 3 equations can be written in matrix form

$$\begin{pmatrix} 0 & a^3 & -a^2 \\ -a^3 & 0 & a^1 \\ a^2 & -a^1 & 0 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which is equivalent to the familiar condition $\underline{a} \times \underline{x} = \underline{0}$, which forces \underline{x} to be proportional to \underline{a} . The Levi-Civita notation enables us to see this immediately. The determinant unevaluated on its first argument defines a covector, which must vanish for any vector x , namely

$$\epsilon_{ijk} x^j A^k_3 = 0.$$

The relation between the vector A^k_3 and the antisymmetric tensor $A_{ij} = \epsilon_{ijk} A^k_3$ is exactly what we explored in Exercise 1.2.4. It is this antisymmetric tensor which determines the plane—the corresponding vector just helps us interpret it in terms of a normal vector using the dot product to raise and lower indices to achieve this interpretation. The antisymmetric tensor itself can be realized at the antisymmetrized tensor product of any two vectors in that plane. Antisymmetrization as well as going from a vector to an antisymmetric 2-tensor and back, are the key tools we need to develop in general to describe all the p -planes through the origin of \mathbb{R}^n .

If we add one more dimension to consider \mathbb{R}^4 , then starting from 4×4 matrices of 4 column vectors, if we unevaluate the determinant on one vector argument, we determine the hyperplane spanned by the remaining 3 vectors. If we unevaluate on two vector arguments, we determine an ordinary plane spanned by the remaining two vector arguments, and finally if we unevaluate the determinant on 3 vector arguments, we determine a line spanned by the single vector argument remaining. At each step at least one of all the possible subdeterminants must be nonzero for the set of remaining columns to be linearly independent.

In practice we are primarily interested in at most 4-dimensions for the elementary applications we have in mind so the following generic dimension discussion just allows us to understand the general structure of this antisymmetric algebra, which consists of two parts: antisymmetrized tensor products and a duality of complementary indices that comes out of the all the possible unevaluated determinant tensors, which in 3 dimensions is simple

$$S = \epsilon_{ijk} a^i b^j c^k, \quad S_i = \epsilon_{ijk} b^j c^k, \quad S_{ij} = \epsilon_{ijk} c^k,$$

and one level more complicated in 4 dimensions

$$S = \epsilon_{ijkl} a^i b^j c^k d^l, \quad S_i = \epsilon_{ijk} b^j c^k d^l, \quad S_{ij} = \epsilon_{ijk} c^k d^l, \quad S_{ijk} = \epsilon_{ijkl} d^l.$$

□

Now go to the corresponding discussion for \mathbb{R}^n .

Start with an $n \times n$ matrix whose columns represent n vectors in \mathbb{R}^n . The determinant of this matrix is nonzero if the n vectors are linearly independent and zero otherwise. Remove one column from the matrix to get an $n \times (n-1)$ matrix representing $(n-1)$ vectors. Are these

linearly independent? The only way the determinant function can enter into the discussion is by deleting one row of this matrix (which can be done in n different ways), which corresponds to projecting the set of $(n - 1)$ vectors down to \mathbb{R}^{n-1} , where one can test for their linear independence by taking the determinant. At least one such subdeterminant must be nonzero for the original $(n - 1)$ vectors in \mathbb{R}^n to be linearly independent; otherwise it is straightforward to show that those original vectors are linearly dependent. Arranging these n minor determinants in a column, each entry multiplied by its appropriate alternating sign to make it into the associated cofactor, the original $n \times n$ determinant is obtained as the dot product of the omitted column with the column of cofactors, called the expansion of the determinant along the omitted column. The transpose of the column vector of cofactors is the row vector which simply represents the determinant as a linear function of the last column vector (having fixed the first $(n - 1)$ columns), which is a covector.

Suppose one starts with $n - 1$ vectors in \mathbb{R}^n arranged as the columns of a matrix. Form the row vector of its n sequential $(n - 1) \times (n - 1)$ subdeterminants (obtained by deleting each row in succession) multiplied by an alternating sign to become the cofactors of an $n \times n$ matrix with an additional final column (okay, there is also an overall sign $(-1)^n$ since it is the last column). This row vector acts as the set of coefficients of a linear function of an n th vector which produces the determinant of the $n \times n$ matrix in which that last column vector is augmented to the original $n \times (n - 1)$ matrix as an extra column, whose interpretation is the signed volume of the n -parallelepiped formed by the n vectors. If the last vector is a linear combination of the remaining vectors, the volume is zero, so requiring this linear function (if nonzero, which means that the $n - 1$ vectors are linearly independent) to be zero defines the linear condition that defines the hyperplane through the origin (subspace) determined by the span of those $n - 1$ linearly independent vectors. In this way the row of cofactors determines a linear function of a single vector which determines hyperplanes through the origin. Interpreting the value of the linear function on the vector as the dot product of the coefficient vector with the input vector, the zero value of this dot product gives the coefficient vector the geometric interpretation as a normal to that hyperplane.

Continue iterating this process. Delete another column from the $n \times n$ matrix to obtain an $n \times (n - 2)$ matrix. For these $n - 2$ columns to be linearly independent and determine an $(n - 2)$ -plane through the origin, at least one of the $n(n - 1)$ distinct subdeterminants obtained by deleting two rows from the $n \times (n - 2)$ matrix must be nonzero. One can then interpret these determinants with an appropriate sign as the components of an antisymmetric bilinear function of two additional vectors which complete the set to n vectors to give the value of the determinant of the full set of vectors. If the partial evaluation of this 2-covector on a single vector is a zero covector, that additional vector must lie in the $(n - 2)$ plane of the remaining vectors. In this way the determinant function, a multilinear function of n vectors, if only partially evaluated on $n - 1$ vectors, determines the subspace spanned by those $n - 1$ vectors, and if only partially evaluated on $n - 2$ vectors, determines the subspace spanned by those $n - 2$ vectors, and so on. This method of determining subspaces of \mathbb{R}^n directly in terms of a span of a set of vectors $\{\underline{N}_a\}$ is an alternative to the complementary approach of specifying what vectors (called normals \underline{N}_a , $a = 1..r$) its elements should be orthogonal to in the usual dot product, which requires determining the null space of a matrix with rank r

linearly independent rows (namely just the solution space of $\langle \underline{N}_1^T, \dots, \underline{N}_r^T \rangle \underline{x} = 0$ which defines the subspace as the subspace orthogonal to the span of the subspace spanned by the set of normals). The determinant approach instead directly uses the vectors which span the subspace to find the condition that any other vector lie in that subspace.

In this way the determinant function through partial evaluation determines a series of multilinear functions of p vectors for $1 \leq p \leq n$ which determine all the p -dimensional subspaces. These multiforms, that is, antisymmetric $\binom{0}{p}$ -tensors, also determine the p -measure of p -parallelepipeds in those subspaces in a way that we still must flush out. The multiforms arise from the antisymmetrization of the $(n - p)$ remaining columns in the $n \times n$ determinant. Thus antisymmetrizing a tensor product of p vectors or covectors amounts to this family of successive determinants of submatrices of components with appropriate signs tossed in. This iterative evaluation process for determinants will define the wedge product which captures the orientation and measure information about subspaces and bases for those subspaces.

Example 4.1.2. The cross product $\vec{a} \times \vec{b}$ of two vectors $\vec{a} = \langle a^1, a^2, a^3 \rangle$ and $\vec{b} = \langle b^1, b^2, b^3 \rangle$ in R^3 has two uses. Its direction determines uniquely the orientation of the plane through the origin containing both vectors, while its magnitude $|\vec{a} \times \vec{b}|$ equals the area of the parallelogram formed with the two vectors as edges from the common vertex at the origin. Form the 3×2 matrix $\langle \vec{a} | \vec{b} \rangle$ whose columns are the standard basis components, and augment this matrix with an additional column whose entries are the symbols for the standard basis vectors to form a 3×3 matrix which has a determinant. The components of their cross product vector consists of the three subdeterminants of this matrix $\langle \vec{a} | \vec{b} \rangle$ (called minors of the 3×3 matrix) multiplied by an alternating sign (to form the corresponding cofactors of the 3×3 matrix), which can be expressed as the abovementioned 3×3 determinant (although traditionally one uses rows rather than columns in order to only have to deal with row reduction techniques, but the determinant is invariant under interchanging rows and columns)

$$\vec{a} \times \vec{b} = \begin{vmatrix} a^1 & b^1 & \vec{e}_1 \\ a^2 & b^2 & \vec{e}_2 \\ a^3 & b^3 & \vec{e}_3 \end{vmatrix} = \begin{vmatrix} a^2 & b^2 \\ a^3 & b^3 \end{vmatrix} \vec{e}_1 - \begin{vmatrix} a^1 & b^1 \\ a^3 & b^3 \end{vmatrix} \vec{e}_2 + \begin{vmatrix} a^1 & b^1 \\ a^2 & b^2 \end{vmatrix} \vec{e}_3. \quad (4.1)$$

These cofactors are obtained by deleting each row in turn from the 3×2 matrix to form square 2×2 matrices which have a determinant. Clearly at least one of these determinants must be nonzero or it will imply that all three projections of the pair of vectors onto the three coordinate planes are proportional, implying that the vectors themselves are proportional and hence determine a degenerate parallelogram of zero area. Thus the nonzero value of the cross product guarantees the linear independence of the two vectors.

The magnitude of this cross product is the area of the parallelogram formed by the two vectors.

$$|\vec{a} \times \vec{b}|^2 = |\vec{a}|^2 |\vec{b}|^2 \sin^2 \theta = |\vec{a}|^2 |\vec{b}|^2 (1 - \cos^2 \theta) = (\vec{a} \cdot \vec{a})(\vec{b} \cdot \vec{b}) - (\vec{a} \cdot \vec{b})^2 = \begin{vmatrix} \vec{a} \cdot \vec{a} & \vec{a} \cdot \vec{b} \\ \vec{b} \cdot \vec{a} & \vec{b} \cdot \vec{b} \end{vmatrix}. \quad (4.2)$$

In the vector space R^3 , the subspaces are lines and planes through the origin and the whole space, of dimensions 1, 2 and 3. The measure of the rectangular objects of these dimensions

are length, area and volume, which in R^n we have to generalize to p -measure for $p = 1, 2, \dots, n$. Each p -dimensional subspace can be characterized by a linear independent set of p vectors which span the subspace, which make $n \times p$ matrices $\langle \vec{a}_{(1)} | \dots | \vec{a}_{(p)} \rangle$. This matrix determines the orientation of the p -plane through the origin containing these vectors, and the magnitude of the p -parallelloiped formed by the vectors. To capture these two pieces of information we need to generalize the cross product to the wedge product, and introduce a magnitude for that wedge product to give the p -measure of that parallelloiped. Looking ahead to differential geometry, once we can evaluate p -measures of p -parallelloipeds in the tangent space, we can integrate up differential p -measures: arclength, surface area and volume of curves, surfaces and solid regions in R^3 , for example.

One can write an equation for the plane of these two vectors as

$$(\vec{a} \times \vec{b}) \cdot \vec{x} = 0. \quad (4.3)$$

This also generalizes with the wedge product to give an equation for the p -subspaces in R^n . Here in this example the triple scalar product of 3 vectors has the interpretation as the signed volume of the corresponding parallelloiped, so if the third vector lies in the plane of the first 2, then one gets zero for this volume, hence this condition forces the variable \vec{x} to lie in that plane. Thus determinants are key to both orientation and measure for subspaces in this example, and we just need to play a little bit with its generalization. \square

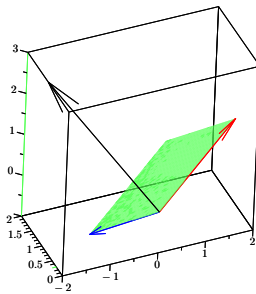


Figure 4.1: The parallelogram formed by two vectors \vec{a} and \vec{b} together with their cross product vector $\vec{a} \times \vec{b}$.

4.2 The wedge product

We are now ready to introduce the wedge product “ \wedge ”. This is the “obvious” antisymmetrized tensor product, but it only becomes obvious *after* you understand it. For each integer value $0 \leq p \leq n$, we have p -vectors (antisymmetric $\binom{p}{0}$ -tensors) and p -covectors (antisymmetric $\binom{0}{p}$ -tensors) also called p -forms, where $\binom{0}{0}$ -tensors are just scalars, i.e., real numbers. We will also consider scalars as both 0-forms and 0-vectors for completeness. If we consider the tensor product of two antisymmetric tensors of the same index level (both covariant or both contravariant), the resulting tensor will not be antisymmetric, but we can take its antisymmetric part. This is essentially the wedge product but we must take into account overcounting issues which leads to an at first mysterious factorial factor in the definition but which later becomes clear.

We will be continually using the fact that when a group of indices is contracted with a group of antisymmetric indices only the antisymmetric part contributes to the sum. For example, if $T^{ij} = T^{[ij]}$ is antisymmetric then

$$\begin{aligned}
 S_{ij}T^{ij} &= S_{ij}T^{[ij]} && \text{(antisymmetry of } T) \\
 &= S_{ij} \left(\frac{1}{2} \delta_{mn}^{ij} T^{mn} \right) && \text{(definition of antisymmetric part)} \\
 &= \left(\frac{1}{2} \delta_{mn}^{ij} S_{ij} \right) T^{mn} && \text{(reapplication to } S) \\
 &= S_{[ij]} T^{ij}, && \text{(antisymmetrization of } S)
 \end{aligned}$$

where the parentheses are not needed but show how the antisymmetrizer is transferred to the other set of indices. Thus in any contraction with an antisymmetric group of indices, only the antisymmetric part of the contracting factor contributes.

Consider covariant antisymmetric tensors, i.e., $\text{ALT}(T) = T$ or in component form: $T_{[i_1 \dots i_p]} = T_{i_1 \dots i_p}$. A component can be nonzero only when all indices are distinct, and any two components with the same set of index values differ by the sign of the permutation which takes one to the other. By making the convenient definition

$$\delta_{i_1 \dots i_p}^{j_1 \dots j_p} \omega^{i_1} \otimes \dots \otimes \omega^{i_p} = p! \omega^{[j_1} \otimes \dots \otimes \omega^{j_p]} \equiv \omega^{j_1 \dots j_p} = \omega^{[j_1 \dots j_p]},$$

then any such tensor can be expressed as a linear combination of the p -forms $\omega^{j_1 \dots j_p}$

$$\begin{aligned}
 T &= T_{i_1 \dots i_p} \omega^{i_1} \otimes \dots \otimes \omega^{i_p} = T_{[i_1 \dots i_p]} \omega^{i_1} \otimes \dots \otimes \omega^{i_p} && \text{(antisymmetric tensor)} \\
 &= \frac{1}{p!} T_{j_1 \dots j_p} \delta_{i_1 \dots i_p}^{j_1 \dots j_p} \omega^{i_1} \otimes \dots \otimes \omega^{i_p} && \text{(definition of antisymmetric part)} \\
 &= T_{i_1 \dots i_p} \omega^{[i_1} \otimes \dots \otimes \omega^{i_p]} && \text{(reapplied to omega indices)} \\
 &= \frac{1}{p!} T_{i_1 \dots i_p} \omega^{i_1 \dots i_p} && \text{(definition of antisymmetric basis)} \\
 &= \sum_{i_1 < \dots < i_p} T_{i_1 \dots i_p} \omega^{i_1 \dots i_p} \equiv T_{|i_1 \dots i_p|} \omega^{i_1 \dots i_p}. && \text{(sum over independent components)}
 \end{aligned}$$

Notice how antisymmetrizing on the lower indices is equivalent to antisymmetrizing on the upper indices. The tensors $\{\omega^{i_1 \dots i_p}\}_{i_1 < \dots < i_p}$ are a basis for the space of p -forms, but since ordered sums are inconvenient (more notation), we sum over all indices and divide by $p!$ to compensate for including $p!$ terms in the sum which repeat each other. Alternatively we can introduce the convention that surrounding a set of antisymmetric indices with vertical bars in a summation indicates restriction of the sum to the index ordered summation. The choice of ordered indices is not essential, it is enough to pick one ordering of each set of possible indices, as we do in 3-dimensions where the cyclic order 23, 31, 12 is more useful than 23, 13, 12 because of the cyclic properties and the signs associated with them, as we will learn below.

Now $\omega^{i_1 \dots i_p}$ is itself on antisymmetrized tensor product of covectors, multiplied by a counting factor. Why is the counting factor (namely $p!$) included? Well, if $V = R^n$ and $p = n$, then we saw above that

$$\det = \epsilon_{i_1 \dots i_n} \omega^{i_1} \otimes \dots \otimes \omega^{i_n} = \epsilon_{i_1 \dots i_n} \omega^{[i_1} \otimes \dots \otimes \omega^{i_n]} = \frac{1}{n!} \epsilon_{i_1 \dots i_n} \omega^{i_1 \dots i_n} = \omega^{1 \dots n},$$

i.e., the single independent n -form $\omega^{1 \dots n}$ of this family $\{\omega^{i_1 \dots i_n}\}$ is exactly the determinant function, which is more interesting than the determinant function divided by $n!$, which is instead equal to the antisymmetrized tensor product with no counting factor modifying it.

It turns out to be useful to introduce an antisymmetrized tensor product, modified by some counting factor coefficient, of any number of factors which are themselves antisymmetric tensors of the same index level (all covariant or all contravariant so that we can take the antisymmetric part).

For example, $S_{ijk} = T_{ij}f_k$ are the components of $S = T \otimes f$ which are clearly antisymmetric in (i, j) if T is antisymmetric but not in all three indices. However, $\text{ALT}(S) = \text{ALT}(T \otimes f)$ with components $S_{[ijk]} = T_{[ij}f_{k]}$ is antisymmetric. We would like to introduce a new product “ \wedge ” called the wedge product (since the symbol visually resembles a wedge) so that we can write $T \wedge f = (\text{some factor}) \text{ALT}(T \otimes f)$, where the factor is chosen conveniently.

Suppose we make the definition

$$(I) \quad \omega^{i_1} \wedge \dots \wedge \omega^{i_p} \equiv p! \omega^{[i_1} \otimes \dots \otimes \omega^{i_p]} = \delta_{j_1 \dots j_p}^{i_1 \dots i_p} \omega^{j_1} \otimes \dots \otimes \omega^{j_p} = \omega^{i_1 \dots i_p}$$

and extend this by linearity to the wedge product of p covectors

$$\begin{aligned} f^{(1)} \wedge \dots \wedge f^{(p)} &= (f_{i_1}^{(1)} \omega^{i_1}) \wedge \dots \wedge (f_{i_p}^{(p)} \omega^{i_p}) \\ &= f_{i_1}^{(1)} \dots f_{i_p}^{(p)} \omega^{i_1} \wedge \dots \wedge \omega^{i_p} = f_{i_1}^{(1)} \dots f_{i_p}^{(p)} \omega^{i_1 \dots i_p} = f_{i_1}^{(1)} \dots f_{i_p}^{(p)} \omega^{[i_1 \dots i_p]} \\ &= f_{i_1}^{(1)} \dots f_{i_p}^{(p)} \frac{1}{p!} \delta_{j_1 \dots j_p}^{i_1 \dots i_p} \omega^{j_1 \dots j_p} = f_{[j_1}^{(1)} \dots f_{j_p]}^{(p)} \omega^{j_1 \dots j_p} \\ &= \frac{1}{p!} \left\{ p! f_{[j_1}^{(1)} \dots f_{j_p]}^{(p)} \right\} \omega^{j_1 \dots j_p} \end{aligned}$$

where we have used the antisymmetry property $\omega^{i_1 \dots i_p} = \omega^{[i_1 \dots i_p]}$ and the last line enables us to identify the components of the resulting antisymmetric tensor

$$[f^{(1)} \wedge \dots \wedge f^{(p)}]_{i_1 \dots i_p} = p! f_{[i_1}^{(1)} \dots f_{i_p]}^{(p)} = \delta_{i_1 \dots i_p}^{j_1 \dots j_p} f_{j_1}^{(1)} \dots f_{j_p}^{(p)}.$$

With this definition, then for the case $p = n$, the single independent component

$$\begin{aligned} [f^{(1)} \cdots f^{(n)}]_{1\dots n} &= n! f^{(1)}_{[1} \cdots f^{(n)}_{n]} \\ &= \det \begin{pmatrix} \underline{f^{(1)}} \\ \vdots \\ \underline{f^{(n)}} \end{pmatrix} \end{aligned}$$

is just the determinant of the matrix whose rows are the components of the covectors in this set. For the case $p < n$, these are the possible subdeterminants of size $p \times p$ taken from the $p \times n$ matrix $f^{(i)}_j$, $1 \leq i \leq p$, $1 \leq j \leq n$ whose rows are the components of the p 1-forms, i.e., the determinants of the matrices consisting of all possible subsets of p columns taken from the original matrix.

Similarly for p -vectors we can define a basis by $e_{i_1 \dots i_p} \equiv e_{i_1} \wedge \cdots \wedge e_{i_p}$ and find that

$$[u_{(1)} \wedge \cdots \wedge u_{(n)}]^{1\dots n} = \det(\underline{u}_{(1)} \dots \underline{u}_{(n)}),$$

which is the determinant of the matrix $(u^j_{(i)})$ whose columns are the components of the vectors. For $V = \mathbb{R}^n$ with the usual Euclidean geometry, this is just the volume of the parallelepiped they form. So in each case the factorial factor eliminates an ugly counting factor to give something more interesting, namely the determinant. For the case $p < n$, these are the possible subdeterminants of size $p \times p$ taken from the $n \times p$ matrix $e^j_{(i)}$, $1 \leq i \leq p$, $1 \leq j \leq n$ whose columns are the components of the p vectors, i.e., the determinants of the matrices consisting of all possible subsets of p rows taken from the original matrix.

However, we still don't know to take the wedge product of higher rank antisymmetric tensors. Our notation implicitly tells us how to do this since as long as we assume the wedge product is associative, then

$$\omega^{i_1 \dots i_p j_1 \dots j_q} = \omega^{i_1} \wedge \cdots \wedge \omega^{i_p} \wedge \omega^{j_1} \wedge \cdots \wedge \omega^{j_q} = \omega^{i_1 \dots i_p} \wedge \omega^{j_1 \dots j_q}$$

suggests how to wedge two basis tensors together in a way consistent with the notation. This can then be extended by linearity to any two antisymmetric tensors

$$\begin{aligned} T \wedge S &= \left(\frac{1}{p!} T_{i_1 \dots i_p} \omega^{i_1 \dots i_p} \right) \wedge \left(\frac{1}{q!} S_{j_1 \dots j_q} \omega^{j_1 \dots j_q} \right) \\ &= \frac{1}{p!q!} T_{i_1 \dots i_p} S_{j_1 \dots j_q} \omega^{i_1 \dots i_p} \wedge \omega^{j_1 \dots j_q} \\ &= \frac{1}{p!q!} T_{[i_1 \dots i_p} S_{j_1 \dots j_q]} \omega^{i_1 \dots i_p j_1 \dots j_q} \\ &= \frac{1}{(p+q)!} \left[\frac{(p+q)!}{p!q!} T_{[i_1 \dots i_p} S_{j_1 \dots j_q]} \right] \omega^{i_1 \dots i_p j_1 \dots j_q}, \end{aligned}$$

where in the third line, only the antisymmetric part of the tensor product contributes to the sum since $\omega^{i_1 \dots i_p j_1 \dots j_q} = \omega^{[i_1 \dots i_p j_1 \dots j_q]}$ is antisymmetric. Identifying the components of the $(p+q)$ -form from the identity

$$T \wedge S = \frac{1}{(p+q)!} [T \wedge S]_{i_1 \dots i_{p+q}} \omega^{i_1 \dots i_{p+q}} = [T \wedge S]_{i_1 \dots i_{p+q}} \omega^{[i_1 \dots i_{p+q}]},$$

this leads to the definition

$$(II) \quad [T \wedge S]_{i_1 \dots i_p j_1 \dots j_q} = \frac{(p+q)!}{p!q!} T_{[i_1 \dots i_p} S_{j_1 \dots j_q]} \quad \text{or} \quad T \wedge S = \frac{(p+q)!}{p!q!} \text{ALT}(T \otimes S).$$

In exactly the same way we could have partitioned the indices into 3 (or more) subsets and found

$$T \wedge S \wedge R = \frac{(p+q+r)!}{p!q!r!} \text{ALT}(T \otimes S \otimes R)$$

(T is a p -form, S is a q -form, R is an r -form) and so on (the pattern is clear). The extreme case of this is the wedge product of n 1-forms

$$f^{(1)} \wedge \dots \wedge f^{(n)} = n! \text{ALT}(f^{(1)} \otimes \dots \otimes f^{(n)}).$$

Example 4.2.1. only for the brave:

generalized Kronecker delta formula manipulation

Our notation assumes the wedge product is associative since no parentheses are necessary to evaluate $T \wedge S \wedge R$. Is this consistent? Do we have

$$(III) \quad T \wedge S \wedge R = (T \wedge S) \wedge R = T \wedge (S \wedge R)?$$

Yes, we've defined it to be true, but let's check as an exercise. First we factor out the common numerical factors, letting T , S and R be p , q , and r -forms respectively. Using the definition (II) for each wedge product

$$\begin{aligned} (T \wedge S) \wedge R &= \frac{((p+q)+r)!}{(p+q)!r!} \text{ALT}((T \wedge S) \otimes R) = \frac{(p+q+r)!}{p!q!r!} \text{ALT}(\text{ALT}(T \otimes S) \otimes R), \\ T \wedge (S \wedge R) &= \frac{(p+(q+r))!}{p!(q+r)!} \text{ALT}(T \otimes (S \wedge R)) = \frac{(p+q+r)!}{p!q!r!} \text{ALT}(T \otimes \text{ALT}(S \otimes R)). \end{aligned}$$

Thus the second equality of (III), is equivalent to

$$(IV) \quad \text{ALT}(\text{ALT}(T \otimes S) \otimes R) = \text{ALT}(T \otimes \text{ALT}(S \otimes R)).$$

If we had defined the wedge product by (II) as is usually done, then we would need to verify (IV) in order to show that it is an associative operation, i.e., to prove (III). Let's just check that (IV) is indeed true.

$$\begin{aligned} &\{\text{ALT}(\text{ALT}(T \otimes S) \otimes R)\}_{i_1 \dots i_p j_1 \dots j_q k_1 \dots k_r} \\ &= \frac{1}{(p+q+r)!} \delta_{i_1 \dots i_p j_1 \dots j_q k_1 \dots k_r}^{m_1 \dots m_p n_1 \dots n_q l_1 \dots l_r} [\text{ALT}(T \otimes S)]_{m_1 \dots m_p n_1 \dots n_q} R_{l_1 \dots l_r} \\ &= \frac{1}{(p+q+r)!} \delta_{i_1 \dots i_p j_1 \dots j_q k_1 \dots k_r}^{m_1 \dots m_p n_1 \dots n_q l_1 \dots l_r} \frac{1}{(q+r)!} \delta_{m_1 \dots m_p n_1 \dots n_q}^{a_1 \dots a_p b_1 \dots b_q} T_{a_1 \dots a_p} S_{b_1 \dots b_q} R_{l_1 \dots l_r} \\ &= \frac{1}{(p+q+r)!} \delta_{i_1 \dots i_p j_1 \dots j_q k_1 \dots k_r}^{[a_1 \dots a_p b_1 \dots b_q] l_1 \dots l_r} T_{a_1 \dots a_p} S_{b_1 \dots b_q} R_{l_1 \dots l_r} \\ &= \frac{1}{(p+q+r)!} \delta_{i_1 \dots i_p j_1 \dots j_q k_1 \dots k_r}^{a_1 \dots a_p b_1 \dots b_q l_1 \dots l_r} T_{a_1 \dots a_p} S_{b_1 \dots b_q} R_{l_1 \dots l_r}, \end{aligned}$$

where the last equality follows since the Kronecker delta is already antisymmetric, so antisymmetrizing has no effect. Similarly the right hand side of (IV) reduces to the same expression, showing their equivalence. \square

Exercise 4.2.1.

successive antisymmetrization and the wedge product

Only if you feel up to the task, repeat for the right hand side of (IV) to obtain the same expression. \blacksquare

These factorials are really a nuisance right? Right. They come from summing over all indices rather than ordered indices. If we agree only to sum over ordered indices, they disappear! Recall our double vertical bar notation

$$T = \frac{1}{p!} T_{i_1 \dots i_p} \omega^{i_1 \dots i_p} = \sum_{i_1 < \dots < i_p} T_{i_1 \dots i_p} \omega^{i_1 \dots i_p} \equiv T_{i_1 \dots i_p} \omega^{|i_1 \dots i_p|} \equiv T_{|i_1 \dots i_p|} \omega^{i_1 \dots i_p}$$

This just tells us to only sum over those p -tuplets (i_1, \dots, i_p) whose values are ordered. Using this notation

$$\begin{aligned} [T \wedge S]_{i_1 \dots i_p j_1 \dots j_q} &= \frac{(p+q)!}{p!q!} T_{[i_1 \dots i_p} S_{j_1 \dots j_q]} = \frac{1}{p!q!} \delta_{i_1 \dots i_p j_1 \dots j_q}^{m_1 \dots m_p n_1 \dots n_q} T_{m_1 \dots m_p} S_{n_1 \dots n_q} \\ &= \delta_{i_1 \dots i_p j_1 \dots j_q}^{m_1 \dots m_p n_1 \dots n_q} T_{|m_1 \dots m_p|} S_{|n_1 \dots n_q|} \end{aligned}$$

and similarly

$$[T \wedge S \wedge R]_{i_1 \dots i_p j_1 \dots j_q k_1 \dots k_r} = \delta_{i_1 \dots i_p j_1 \dots j_q k_1 \dots k_r}^{m_1 \dots m_p n_1 \dots n_q l_1 \dots l_r} T_{|m_1 \dots m_p|} S_{|n_1 \dots n_q|} R_{|l_1 \dots l_r|}.$$

There are no factorials if you don't overcount redundant terms in sums over antisymmetric indices!

Example 4.2.2. The index formulas for the wedge product look scary but in practice one does not really need them when actually taking the wedge product of concrete p -forms (as opposed to doing a derivation of general properties)! Consider the following calculation not with components but with the invariant p -forms. Let $E = E_1 \omega^1 + E_2 \omega^2 + E_3 \omega^3$ be a 1-form and $B = B_1 \omega^2 \wedge \omega^3 + B_2 \omega^3 \wedge \omega^1 + B_3 \omega^1 \wedge \omega^2$ be a 2-form on \mathbb{R}^3 . Then all we have to do is expand out the product using associativity of the wedge product and its antisymmetry to set terms with repeated indices to zero

$$\begin{aligned} E \wedge B &= (E_1 \omega^1 + E_2 \omega^2 + E_3 \omega^3) \wedge (B_1 \omega^2 \wedge \omega^3 + B_2 \omega^3 \wedge \omega^1 + B_3 \omega^1 \wedge \omega^2) \\ &= E_1 B_1 \omega^1 \wedge \omega^2 \wedge \omega^3 + E_2 B_2 \omega^2 \wedge \omega^3 \wedge \omega^1 + E_3 B_3 \omega^3 \wedge \omega^1 \wedge \omega^2 \\ &= (E_1 B_1 + E_2 B_2 + E_3 B_3) \omega^1 \wedge \omega^2 \wedge \omega^3. \end{aligned}$$

Simple, no? Notice what looks like a dot product. This is no accident and will be explored below.

□

Example 4.2.3. Let's repeat the above problem using index notation

$$[E \wedge B]_{ijk} = \delta_{ijk}^{mnl} E_m B_{nl} = \delta_{ijk}^{123} E_1 B_{23} + \delta_{ijk}^{213} E_2 B_{13} + \delta_{ijk}^{312} E_3 B_{12}$$

and

$$\begin{aligned} [E \wedge B]_{123} &= E_1 B_{23} - E_2 B_{13} + E_3 B_{12} & (\text{"ordered sum" has alternating sign}) \\ &= E_1 B_{23} + E_2 B_{31} + E_3 B_{12}. & (\text{"cyclic sum" has positive signs}) \end{aligned}$$

Notice that in the previous example we made the identification $(B_1, B_2, B_3) = (B_{23}, B_{31}, B_{12})$. This too will be explored below. We need the complications of the component formulas for the wedge product in order to see how it fits into other mathematical contexts, but for computation purposes, doing the wedge product of the representation of the p -form or p -vector factors in terms of the basis-wedged forms or vectors as in the previous example is easy and more efficient. □

Exercise 4.2.2.

wedges in \mathbb{R}^3

Suppose $u = \langle 1, 2, 3 \rangle = u^i e_i$ and $v = \langle -1, 1, 2 \rangle = v^i e_i$ on \mathbb{R}^3 .

a) What are the three independent components $(u \wedge v)^{23}, (u \wedge v)^{31}, (u \wedge v)^{12}$?

How are these related to the cross product $u \times v$?

b) If $w = \langle 1, 1, 1 \rangle$, what is the single independent component of $u \wedge v \wedge w$?

How is this related to their triple scalar product?

c) Suppose $B = B^1 e_{23} + B^2 e_{31} + B^3 e_{12}$ and $E = E^1 e_1 + E^2 e_2 + E^3 e_3$.

What is $[B \wedge E]^{123}$?

■

Exercise 4.2.3.

wedges in \mathbb{R}^4

On \mathbb{R}^4 :

a) Evaluate $F \wedge F$ and $F \wedge H$ for

$$\begin{aligned} F &= E^1 e_{14} + E^2 e_{24} + E^3 e_{34} + B^1 e_{23} + B^2 e_{31} + B^3 e_{12}, \\ H &= -B^1 e_{14} - B^2 e_{24} - B^3 e_{34} + E^1 e_{23} + E^2 e_{31} + E^3 e_{12}. \end{aligned}$$

b) Simplify $(B^1 e_{234} + B^2 e_{314} + B^3 e_{124}) \wedge (E^1 e_1 + E^2 e_2 + E^3 e_3)$.

■

Exercise 4.2.4.**wedges in \mathbb{R}^5**

On \mathbb{R}^5 simplify the following, expressing each as a linear combination of ordered basis tensors

- (a) $e_3 \wedge e_5 \wedge e_{24}$,
- (b) $e_2 \wedge e_3 \wedge e_{62}$,
- (c) $e_1 \wedge (e_{14} + e_{64})$,
- (d) $(e + 3e_4 - e_6) \wedge (2e_{23} + e_{36}) \wedge e_{45}$,
- (e) $(e_{12} + e_{13}) \wedge (e_{34} + e_{25}) \wedge (e_{56} + e_{46})$.



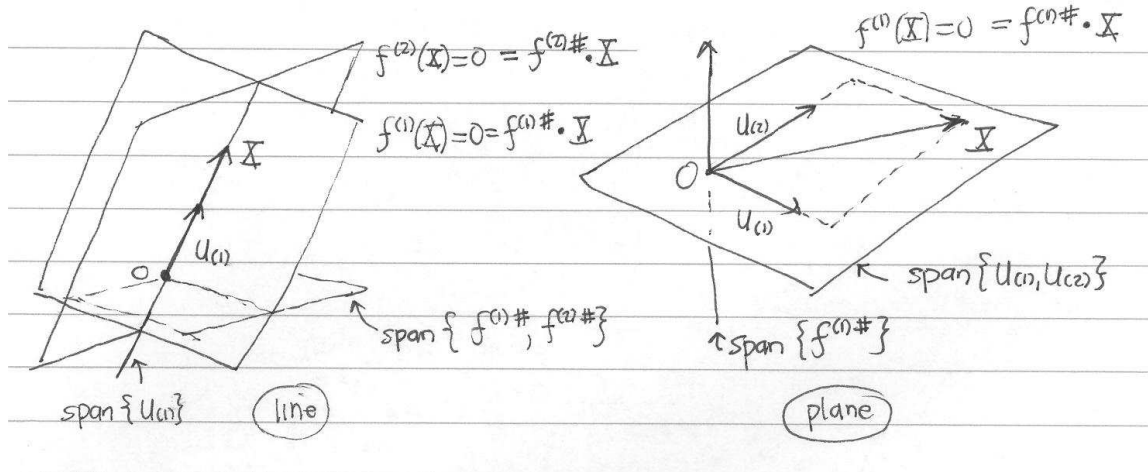


Figure 4.2: A line through the origin of \mathbb{R}^3 can be viewed either as the span of a nonzero vector or the intersection of two distinct planes (zero value surfaces of two independent covectors). A plane through the origin may be viewed as the span of two independent vectors or the set of vectors perpendicular to a given vector (zero value surface of one covector).

4.3 Subspace orientation and a new star duality

If $\{u_{(1)}, \dots, u_{(p)}\}$ is a collection of p vectors in V , then $\text{span}\{u_{(1)}, \dots, u_{(p)}\}$ is the set of all possible linear combinations of these vectors—yielding a vector or linear subspace of V , whose dimension is p if the set is linearly independent. We can think of such a subspace as a “ p -plane” through the origin. We would like to describe the orientation or “direction” of the p -plane.

In \mathbb{R}^3 there are three ways to do this: two involve only linearity, while the third uses the Euclidean inner product. The nontrivial subspaces are lines ($p = 1$) and planes ($p = 2$) through the origin.

A subspace can be specified *explicitly* by giving a basis, which may be used to parametrize it, i.e., represent a point in the subspace as an arbitrary linear combination of the basis vectors, or *implicitly* as the simultaneous solution of a system of linear equations, i.e., the intersection of the zero value level surfaces of a set of linearly independent covectors. These relationships are complementary—a p -subspace is determined directly by p linearly independent vectors through their span or by $n - p$ covectors indirectly through their zero value sets. This is the substance of a new star duality (named for its superscript symbol) between p -subspaces of vectors and $(n - p)$ -subspaces of covectors, in contrast to our existing duality between a vector space and its dual space of the same dimension.

Consider the case of a plane through the origin in \mathbb{R}^3 (a 2-dimensional subspace) determined by a set of two linearly independent vectors $\{u_{(1)}, u_{(2)}\}$. Any two linearly independent combinations of this basis are just as good as the original two—either set specifies the same plane. The 2-vector $u_{(1)} \wedge u_{(2)}$ at most changes by a scalar multiple under such a change of basis. Letting $a, b = 1, 2$, one uses the antisymmetry properties to simplify the transformed

wedge product

$$\begin{aligned} u_{(1)'} &= A^b_a u_{(b)} \longrightarrow u_{(1)'} \wedge u_{(2)'} = A^a_1 A^b_2 u_{(i)} \wedge u_{(j)} = A^{[a}_1 A^{b]}_2 u_{(i)} \wedge u_{(j)} \\ &= 2! A^{[1}_1 A^{2]}_2 u_{(1)} \wedge u_{(2)} = \det \underline{A} u_{(1)} \wedge u_{(2)}. \end{aligned}$$

In fact it just changes by the determinant of the matrix of the change of basis, which is nonzero if the new vectors are linearly independent as assumed. The condition that a vector X belong to the plane is equivalent to

$$X = c_1 u_{(1)} + c_2 u_{(2)} \longrightarrow u_{(1)} \wedge u_{(2)} \wedge X = u_{(1)} \wedge u_{(2)} \wedge (c_1 u_{(1)} + c_2 u_{(2)}) = 0$$

since any wedge product with repeated factors vanishes. In index notation this condition is

$$u_{(1)}^{[i} u_{(2)}^{j]} X^k = 0 \leftrightarrow \delta^{ijk}_{mnq} u_{(1)}^m u_{(2)}^n X^q = 0,$$

or since we are in \mathbb{R}^3 where there is only a single independent component 123 of three antisymmetric indices, and $\delta^{123}_{mnq} = \epsilon_{mnq}$, we can write instead

$$\epsilon_{ijk} u_{(1)}^i u_{(2)}^j X^k = 0,$$

which leads us to introduce the covector $f^{(1)}$ for which this plane is the zero value surface $f^{(1)}_k X^k = 0$, namely

$$f^{(1)}_k \equiv \epsilon_{ijk} u_{(1)}^i u_{(2)}^j \equiv [^{(*)}(u_{(1)} \wedge u_{(2)})]_k = \frac{1}{2} \epsilon_{kij} [u_{(1)} \wedge u_{(2)}]^{ij},$$

where for convenience we have permuted the indices of the Levi-Civita symbol by two transpositions which does not change its value $\epsilon_{ijk} = \epsilon_{kij}$. This component formula defines a covector which specifies the same plane implicitly and a natural star duality operation $(^*)$ from 2-vectors to 1-forms. One can go backwards from the covector $f^{(1)} = (^*)(u_{(1)} \wedge u_{(2)})$ to the 2-vector $u_{(1)} \wedge u_{(2)} = (^*)f^{(1)}$, which extends the operation $(^*)$ in the opposite direction

$$[^{(*)}f^{(1)}]^{mn} \equiv f^{(1)}_k \epsilon^{kmn} = \epsilon_{ijk} u_{(1)}^i u_{(2)}^j \epsilon^{kmn} = \epsilon^{mnk} \epsilon_{ijk} u_{(1)}^i u_{(2)}^j = \delta^{mn}_{ij} u_{(1)}^i u_{(2)}^j = [u_{(1)} \wedge u_{(2)}]^{mn},$$

where we have used the identity

$$\epsilon^{mnk} \epsilon_{ijk} = \delta^{mn}_{ij} = \delta^{mn}_{ij}$$

as well as permuted the indices of the Levi-Civita symbol again, with no sign change, showing that in this instance $(^*)(^*) = Id$.

The map $(^*)$ from a 2-vector to a covector or from a covector to a 2-vector is called (at least by me) the “natural dual” operation since it does not rely on the existence of an inner product at all, and in fact it extends the natural duality of covectors and vectors which exists in the relationship between a basis of the vector space and its dual basis. However, since its definition involves the Levi-Civita symbol which “transforms” as a tensor density, this operation depends on the choice of basis leading from a given 2-vector to a covector differing by the determinant

of the basis changing matrix in a new basis, but this does not affect its zero value surface. For example, vectors $x = x^3 e_3$ in the 1-dimensional subspace spanned by e_3 in \mathbb{R}^3 satisfy $\omega^1(x) = 0 = \omega^2(x)$ according to the duality relations. If we change the other two basis vectors in the basis, the new basis vectors $\omega^{1'}, \omega^{2'}$ must still satisfy $\omega^{1'}(x) = 0 = \omega^{2'}(x)$ by duality. The two proportional 2-forms $\omega^1 \wedge \omega^2 \propto \omega^{1'} \wedge \omega^{2'}$ both determine the same 1-dimensional subspace of vectors which give zero upon evaluation of those 2-forms on x in either argument. The first 2-form is the dual $(^*)e_3$ of e_3 with respect to the first basis and the second 2-form is the dual of e_3 with respect to the new basis.

For the case of a line through the origin of \mathbb{R}^3 (a 1-dimensional subspace), by the same reasoning two linearly independent covectors $\{f^{(1)}, f^{(2)}\}$ are required to specify it implicitly as the intersection of their zero value planes: $f^{(1)}(X) = 0 = f^{(2)}(X)$, and only their wedge product $f^{(1)} \wedge f^{(2)}$ is needed to specify the orientation of the line, since any two linearly independent covectors which specify the line will have a wedge product differing only by a nonzero multiple of $f^{(1)} \wedge f^{(2)}$. The natural dual of this 2-covector defines a vector

$$u_{(1)}^i = \frac{1}{2} \epsilon^{imn} [f^{(1)} \wedge f^{(2)}]_{mn} = \epsilon^{imn} f_m^{(1)} f_n^{(2)} = [^{(*)}(f^{(1)} \wedge f^{(2)})]^i$$

which lies along the line since

$$[f^{(1)} \wedge f^{(2)}]_{mn} = \epsilon_{mni} u_{(1)}^i = [^{(*)}u_{(1)}]_{mn}$$

but $[f^{(1)} \wedge f^{(2)}]_{mn} X^n = 0$ for X along the line (since the contraction of X with either factor vanishes), hence

$$0 = [f^{(1)} \wedge f^{(2)}]_{mn} X^n \epsilon^{mij} = \epsilon_{mnl} u_{(1)}^\ell X^n \epsilon^{mij} = \delta_{nl}^{ij} u_{(1)}^\ell X^n = -[u_{(1)} \wedge X]^{ij}.$$

The wedge product of 2 vectors being zero means that they are linearly dependent, i.e., they lie along the same direction.

Thus in \mathbb{R}^3 , a p -plane through the origin for $p = 1, 2$ (the nontrivial linear subspaces of \mathbb{R}^3) is specified by the wedge product of a basis of p -vectors or by the wedge product of $(n - p)$ -linearly independent covectors which implicitly give the p -plane. The natural dual map “ $(^*)$ ” relates the p -vector and $(n - p)$ -covector to each other.

By “raising” the indices on the covectors with the Euclidean metric one makes the change

$$0 = f^{(i)}(X) = f^{(i)\sharp} \cdot X$$

converting the zero evaluation of the covector on the vector to a vanishing dot product with the vector corresponding to the covector instead. Thus X is orthogonal to each of the vectors obtained from the covectors in this way and to the entire $(n - p)$ -plane they determine—which in turn is specified by the $(n - p)$ -covector which is the wedge product of these vectors. This is the third way of specifying the orientation of a p -plane—by giving its orthogonal $(n - p)$ -plane.

This also leads to a “metric dual” operation which is the natural dual followed by shifting all the indices to the opposite level, i.e., the same level as before the natural dual changed the index level. Thus from the vector specifying a line, we get the 2-vector specifying the orthogonal plane, while from the 2-vector specifying a plane we get a vector orthogonal to the plane.

The same statement applies to the various covectors and p -covectors. For the case of a line in \mathbb{R}^3 , from the 2-covector one gets a vector along the line by the natural dual and a covector by lowering its index. For a plane in \mathbb{R}^3 , from the covector specifying the plane one gets the 2-vector by the natural dual and finally a 2-covector by lowering its indices.

The natural dual takes p -vectors into $(n-p)$ -covectors and vice versa, while the metric dual takes p -vectors into $(n-p)$ -vectors and p -covectors into $(n-p)$ -covectors. Note that because ϵ_{ijk} are not the components of a tensor, the natural dual depends on the choice of basis and changes by a scalar factor under a change of basis. This is okay since the overall scale of any of these tensors is irrelevant to the orientation of the subspaces. However, using the metric we can convert this to a duality operation “ $*$ ” where the overall scale is fixed so that the magnitude of the p -vector determines its p -measure. Instead of using ϵ_{ijk} for the duality operation in \mathbb{R}^3 , i.e.,

$$\frac{1}{3!}\epsilon_{ijk}\omega^{ijk} = \omega^{123},$$

namely the basis 3-covector which of course changes with a change of basis, one can use the “unit” 3-vector which reduces to ω^{123} for any (oriented) orthonormal basis—in particular, for the standard basis of \mathbb{R}^3 . In other words we fix $\eta = \omega^{123}$ in the standard basis of \mathbb{R}^3 and then one can express the fixed 3-covector η in any other basis by the tensor transformation law which will involve the determinant of the transformation for a 3-covector

$$\eta_{i'j'k'} = A^{-1m}{}_i A^{-1n}{}_j A^{-1\ell}{}_k \underbrace{\eta_{mnl}}_{= \epsilon_{mnl} \text{ in the standard basis}} = \epsilon_{ijk}(\det \underline{A}^{-1}),$$

using the definition of the determinant. As long as $\det \underline{A} = 1$, one will have

$$\eta_{i'j'k'} = \epsilon_{ijk} \quad \text{or} \quad \eta = \omega^{1'2'3'} = \omega^{1'} \wedge \omega^{2'} \wedge \omega^{3'},$$

otherwise there will be a correction factor.

Suppose one takes an orthonormal basis of \mathbb{R}^3 (oriented as well) adapted to a subspace, i.e., $\{e_1, e_2, e_3\}$ where $\{e_1\}$ is basis for a 1-dimensional subspace, or $\{e_1, e_2\}$ is a basis for a 2-dimensional subspace. For the latter case $e_1 \wedge e_2$ specifies the plane and the metric dual $*(e_1 \wedge e_2) = e_3$ (we’ll see this more easily below) will give a unit normal to the plane. For the line, $*e_1 = e_2 \wedge e_3$ will give the 2-vector which specifies the orthogonal 2-plane. In each case the “magnitude” of these tensors, divided by the usual overcounting factorial factor, will give the p -measure of the p -parallelepiped formed by the basis vectors. The same will extend by linearity to any adapted basis.

The preceding discussion is just motivation for giving the formulas for the general case. For p -vectors and p -covectors, $0 \leq p \leq n$, define the natural dual by

$$\begin{aligned} [^{(*)}T]_{i_{p+1}\dots i_n} &= \frac{1}{p!} T^{i_1\dots i_p} \epsilon_{i_1\dots i_p i_{p+1}\dots i_n} \quad (\text{contraction of } T \text{ with first } p \text{ indices of } \omega^{1\dots n} \text{ divided by } p!) \\ &= T^{[i_1\dots i_p]} \epsilon_{i_1\dots i_p i_{p+1}\dots i_n}, \\ [^{(*)}S]^{i_{p+1}\dots i_n} &= \frac{1}{p!} S_{i_1\dots i_p} \epsilon^{i_1\dots i_p i_{p+1}\dots i_n} \quad (\text{contraction of } S \text{ with first } p \text{ indices of } e_{1\dots n} \text{ divided by } p!) \\ &= S_{[i_1\dots i_p]} \epsilon^{i_1\dots i_p i_{p+1}\dots i_n}, \end{aligned}$$

where recalling the identities

$$\begin{aligned}\omega^{1\dots n} &= \omega^1 \wedge \dots \wedge \omega^n = \epsilon_{i_1\dots i_n} \omega^{i_1} \otimes \dots \otimes \omega^{i_n}, \\ e_{1\dots n} &= e_1 \wedge \dots \wedge e_n = \epsilon^{i_1\dots i_n} e_{i_1} \otimes \dots \otimes e_{i_n}\end{aligned}$$

explain the comments about contraction, namely the components of the n -form $\omega^1 \wedge \dots \wedge \omega^n$ as a tensor are $\epsilon_{i_1\dots i_n}$ and similarly for the n -vector. The factorial factor avoids overcounting repeated terms in the sum.

Exercise 4.3.1.

double natural dual sign

a) Use the identity

$$\delta_{j_1\dots j_p j_{p+1}\dots j_n}^{i_1\dots i_p j_{p+1}\dots j_n} = \epsilon^{i_1\dots i_p j_{p+1}\dots j_n} \epsilon_{j_1\dots j_p j_{p+1}\dots j_n} = (n-p)! \delta_{j_1\dots j_p}^{i_1\dots i_p}$$

and the permutation result

$$\epsilon^{j_{p+1}\dots j_n i_1\dots i_p} = (-1)^{p(n-p)} \epsilon^{i_1\dots i_p j_{p+1}\dots j_n}$$

(where the sign comes from the $(n-p)$ transpositions needed to move one j index across the group of indices $i_1 \dots i_p$, but there are p indices to move across this group for a total of $p(n-p)$ transpositions) to show that

$$(*)^{(*)}T = (-1)^{p(n-p)}T$$

for a p -vector T .

b) What is $(-1)^{p(3-p)}$ for all values of p : $0 \leq p \leq 3$? What is $(-1)^{p(4-p)}$ for all values of p : $0 \leq p \leq 4$? ■

Note that if $T = \frac{1}{p!} T^{i_1\dots i_p} e_{i_1\dots i_p}$ and

$$(*)T = \frac{1}{(n-p)!} {}^*T_{i_{p+1}\dots i_n} \omega^{i_{p+1}\dots i_n} = \frac{1}{p!} \frac{1}{(n-p)!} T^{i_1\dots i_p} \epsilon_{i_1\dots i_p i_{p+1}\dots i_n} \omega^{i_{p+1}\dots i_n},$$

then by the linearity of the natural dual one has

$$(*)T = \frac{1}{p!} T^{i_1\dots i_p} (**)e_{i_1\dots i_p},$$

so equating the two expressions for $(*)T$ one gets

$$(**)e_{i_1\dots i_p} = \frac{1}{(n-p)!} \epsilon_{i_1\dots i_p i_{p+1}\dots i_n} \omega^{i_{p+1}\dots i_n} = \epsilon_{i_1\dots i_p i_{p+1}\dots i_n} \omega^{|i_{p+1}\dots i_n|}$$

Similarly one finds

$$(*)\omega^{i_1\dots i_p} = \epsilon^{i_1\dots i_p i_{p+1}\dots i_n} e_{|i_{p+1}\dots i_n|}.$$

Exercise 4.3.2.**natural dual index approach**

Verify the formula for $(*)e_{i_1 \dots i_p}$ using the component relations

$$[e_{i_1 \dots i_p}]^{j_1 \dots j_p} = \delta_{i_1 \dots i_p}^{j_1 \dots j_p}, \quad [\omega^{1 \dots n}]_{j_1 \dots j_n} = \delta_{j_1 \dots j_n}^{1 \dots n} = \epsilon_{j_1 \dots j_n}$$

and the component formula for $(*)T$. ■

Like the wedge product in practice, the natural dual in practice is simple. Table 4.1 lists of all the natural duals of the bases of the various p -vector/covector spaces for any 3-dimensional vector space once a basis and its dual basis are given. Note how the cyclic permutations of (1,2,3) dominate every formula.

	p -vectors to $(n-p)$ -covectors	p -covectors to $(n-p)$ -vectors
	$(*) : \Lambda^{(p)}(V) \longrightarrow \Lambda^{(n-p)}(V^*)$	$(*) : \Lambda^{(p)}(V^*) \longrightarrow \Lambda^{(n-p)}(V)$
$p = 0$	$(*)1 = \epsilon_{123}\omega^{123} = \omega^{123}$	$(*)1 = \epsilon^{123}e_{123} = e_{123}$
$p = 1$	$(*)e_1 = \epsilon_{123}\omega^{23} = \omega^{23}$ $(*)e_2 = \epsilon_{231}\omega^{31} = \omega^{31}$ $(*)e_3 = \epsilon_{312}\omega^{12} = \omega^{12}$	$(*)\omega^1 = \epsilon^{123}e_{23} = e_{23}$ $(*)\omega^2 = \epsilon^{231}e_{31} = e_{31}$ $(*)\omega^3 = \epsilon^{312}e_{12} = e_{12}$
$p = 2$	$(*)e_{23} = \epsilon_{231}\omega^1 = \omega^1$ $(*)e_{31} = \epsilon_{312}\omega^2 = \omega^2$ $(*)e_{12} = \epsilon_{123}\omega^3 = \omega^3$	$(*)\omega^{23} = \epsilon^{231}e_1 = e_1$ $(*)\omega^{31} = \epsilon^{312}e_2 = e_2$ $(*)\omega^{12} = \epsilon^{123}e_3 = e_3$
$p = 3$	$(*)e_{123} = \epsilon_{123} = 1$	$(*)\omega^{123} = \epsilon^{123} = 1$

Table 4.1: The table of natural duals on the standard basis and dual basis of $V = \mathbb{R}^3$ or for any 3-dimensional vector space V in a given basis $\{e_i\}$.

Exercise 4.3.3.**natural duals**

- (i) If $n = 3$ and $B = B_{23}\omega^{23} + B_{31}\omega^{31} + B_{12}\omega^{12}$, what is $(*)B$? If $E = E^i e_i$, what is $E \wedge (*)B$?
(ii) If $n = 4$, what is $(*)[\omega^{12} + \omega^{34}]$?

What is

$$(2e_{12} + 3e_{13} - e_{23}) \wedge (*)[\omega^{12} + \omega^{34}]?$$

What is $(*)[e_{123} - e_{412} + 2e_{431}]$?

- (iii) Repeat (i) for $n = 4$.

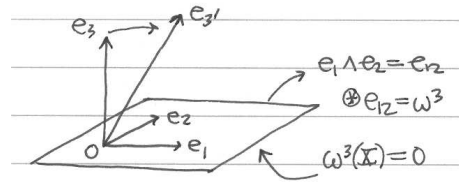


Figure 4.3: If the basis vector e_3 changes, then ω^3 cannot change its orientation (direction) since the plane of e_1 and e_2 does not change, so only its magnitude can change (to maintain the relation $\omega^3(e_3) = 1$).

■

A *decomposable* p -vector or p -covector is one which can be represented as the wedge product of p vectors or covectors. An *adapted basis* of a vector space V , adapted to a p -dimensional subspace W , is a basis of V such that the first p basis vectors are a basis of W . Each adapted basis determines a direct sum of V into W and a complementary subspace which is the span of the last $(n - p)$ basis vectors. Although this changes with a change of adapted basis, the last $(n - p)$ dual basis covectors still determine the given subspace W .

In the \mathbb{R}^3 example in this diagram the plane of e_1 and e_2 is determined by $\omega^3(X) = 0$. If e_3 is changed to $e_{3'}$, keeping e_1 and e_2 fixed, then ω^3 can at most change to $\omega^{3'} = c\omega^3$ to preserve $\omega^{3'}(e_{3'}) = 1$, but its orientation (direction) must stay the same to maintain $\omega^{3'}(e_1) = \omega^{3'}(e_2) = 0$.

In general if $\{e_1, \dots, e_p\}$ determine a p -dimensional subspace, then $(*)\{e_1 \wedge \dots \wedge e_p\} = \omega^{p+1 \dots n}$ can at most change by a determinant factor since $\{\omega^{p+1'}, \dots, \omega^{n'}\}$ must be linearly independent linear combinations of $\{\omega^{p+1}, \dots, \omega^n\}$ alone so that the duality relations giving zero along $\{e_1', \dots, e_p'\}$ are preserved (these must only be linear combinations of $\{e_1, \dots, e_p\}$ to be an adapted basis).

Exercise 4.3.4.

dual of decomposable p -vector

Using the idea of an adapted basis, explain why the natural dual of a decomposable p -vector must itself be decomposable. ■

Rescaled inner product for antisymmetric tensors

In our motivating example, the cross product of two vectors in R^3 determines the orientation of the plane of the two vectors, and its magnitude gives the area of the parallelogram they form. Now we have p -vectors which determine the orientation of the p -planes that they determine. It remains only to introduce their length in a natural way, which will then give the p -measure of the p -parallelopiped they form.

If we have an inner product on V , we have shown how to get an inner product on any of the tensor spaces over V . If T and S are both $\binom{p}{q}$ -tensors, then their inner product is the scalar

$$G_{im} \dots G^{jn} \dots T_{j \dots}^{i \dots} S_{n \dots}^{m \dots} = T_{m \dots j \dots} S^{m \dots j \dots} = T^{i \dots n \dots} S_{i \dots n \dots}.$$

For antisymmetric tensors this overcounts the number of independent component terms in these sums, so it is natural to divide by the number of repetitions in the sum. For p -vectors and p -covectors, define

p -vectors:

$$\langle T, S \rangle = \frac{1}{p!} G_{i_1 j_1 \dots} G_{i_p j_p \dots} T^{i_1 \dots i_p} S^{j_1 \dots j_p} = T_{j_1 \dots j_p} S^{[j_1 \dots j_p]},$$

p -covectors:

$$\langle T, S \rangle = \frac{1}{p!} G^{i_1 j_1 \dots} G^{i_p j_p \dots} T_{i_1 \dots i_p} S_{j_1 \dots j_p} = T^{[j_1 \dots j_p]} S_{j_1 \dots j_p}.$$

For the Euclidean metric, the self inner product of a p -vector is just the sum of the squares of its ordered-indexed components.

Exercise 4.3.5.

self-inner products of p -vectors

What is the self-inner product on \mathbb{R}^3 (Euclidean metric) of the following p -vectors?

- (i) $E^1 e_1 + E^2 e_2 + E^3 e_3$.
- (ii) $B^1 e_{23} + B^2 e_{31} + B^3 e_{12}$.
- (iii) $3e_{123}$.
- (iv) $F = \frac{1}{2} F^{ij} e_{ij}$.

■

Exercise 4.3.6.

quadruple scalar product and area

In Exercise 2.3.4 we saw that the quadruple scalar product definition was easily shown to correspond to the index-shifted component formula and resulting identity

$$Q(U, V, X, Y) = \delta_{mn}^{ij} U_i V_j X^m Y^n = (U \cdot X)(V \cdot Y) - (U \cdot Y)(V \cdot X),$$

where we shift indices as convenient in the standard basis of \mathbb{R}^3 .

a) Show that this the scalar quadruple product is just the full evaluation of the index-shifted wedge product of two vectors, index-lowered of course to be able evaluate them on the final pair of vectors

$$Q(U, V, X, Y) = (U \wedge V)^b(X, Y) = Q(X, Y, U, V).$$

As a $\binom{0}{4}$ -tensor, its pair exchange symmetry together with its antisymmetry in the first and second index pairs makes it a symmetric bilinear form on the space of 2-vectors, often called bi-vectors because of their importance in geometry in capturing not only the orientation information about 2-planes but also the area associated with the parallelogram formed by 2 vectors.

b) Show that

$$Q(X, Y, X, Y) = \frac{1}{2} \|X \wedge Y\|^2 = \langle X \wedge Y, X \wedge Y \rangle.$$

The factor of two avoids overcounting because of the antisymmetry repetition in the sum.

c) Convince yourself that

$$|\langle X \wedge Y, X \wedge Y \rangle| = \text{Area}(X, Y)^2$$

not only holds for the usual \mathbb{R}^3 dot product case but for any inner product \bullet with component matrix (G_{ij})

$$Q(X, Y, X, Y) = (X \bullet X)(Y \bullet Y) - (X \bullet Y)^2.$$

For example the area in 2-dimensional Minkowski spacetime formed by the two parallelograms formed by the vector pairs $\langle 1, 0 \rangle, \langle 1, 2 \rangle$ and $\langle 0, 1 \rangle, \langle 1, 2 \rangle$ which have area 2 and 1 respectively in the corresponding Euclidean geometry have the same areas for the Minkowski geometry using this formula.

■

The unit n -form on an oriented vector space with inner product

We first used the Levi-Civita symbols to give a compact expression for the determinant of any matrix \underline{A}

$$\det \underline{A} = \epsilon_{i_1 \dots i_n} A^{i_1}_1 \dots A^{i_n}_n \quad \text{or} \quad \det \underline{A} \epsilon_{j_1 \dots j_n} = \epsilon_{i_1 \dots i_n} A^{i_1}_1 A^{i_2}_{j_1} \dots A^{i_n}_{j_n}.$$

The index level (up or down) of the alternating symbol is just a convenience here to use our summation convention. If we apply these to the matrix $\underline{G} = (G_{ij})$ of components of an inner product, we can write

$$\begin{aligned} \det \underline{G} \epsilon_{j_1 \dots j_n} &= \epsilon^{i_1 \dots i_n} G_{i_1 j_1} \dots G_{i_n j_n} \\ \det \underline{G}^{-1} \epsilon^{j_1 \dots j_n} &= \epsilon_{i_1 \dots i_n} G^{i_1 j_1} \dots G^{i_n j_n} \end{aligned}$$

Obviously we get into trouble if we try to extend our index-shifting convention to the pair of Levi-Civita symbols themselves, since the results of index-raising or lowering either one differs from the other by the determinant factors on the left hand side in these relations.

However, suppose we have an oriented vector space and define an indexed object

$$\eta_{i_1 \dots i_n} = \pm |\det \underline{G}|^{\frac{1}{2}} \epsilon_{i_1 \dots i_n} \quad \begin{cases} + \text{ sign for an oriented basis} \\ - \text{ sign for oppositely oriented basis} \end{cases}$$

and define all other index positions for the object to be obtained from this fully covariant form of the object by the usual rules for index raising. In particular

$$\begin{aligned} \eta^{i_1 \dots i_n} &= G^{i_1 \dots j_1} \dots G^{i_n j_n} \eta_{j_1 \dots j_n} = G^{i_1 \dots j_1} \dots G^{i_n j_n} \epsilon_{i_1 \dots i_n} (\pm |\det \underline{G}|^{\frac{1}{2}}) \\ &= (\det \underline{G}^{-1}) \epsilon^{i_1 \dots i_n} (\pm |\det \underline{G}|^{\frac{1}{2}}) = \pm (\text{sgn } \det \underline{G}) |\det \underline{G}|^{-1/2} \epsilon^{i_1 \dots i_n}. \end{aligned} \quad (4.4)$$

I claim that

$$\eta = \eta_{i_1 \dots i_n} \omega^{i_1 \dots i_n} = \eta_{1 \dots n} \omega^{1 \dots n}$$

is a uniquely defined tensor, independent of which particular basis we use to define it.

First recall

$$G_{i'j'} = A^{-1m}_i A^{-1n}_j G_{mn} \longleftrightarrow \underline{G}' = (\underline{A}^{-1})^T \underline{G} \underline{A}^{-1}$$

so

$$\underline{G}' = (\det \underline{A}^{-1})^2 \det \underline{G}.$$

(Why?) Furthermore

$$|\det \underline{G}'|^{1/2} = (\text{sgn } \det \underline{A}) \det \underline{A}^{-1} |\det \underline{G}|^{1/2}.$$

According our discussion of scalar densities, $\det \underline{G}$ acts like a scalar density of weight $W = 2$, while $|\det \underline{G}|^{1/2}$ acts like a scalar density of weight $W = 1$ except that it changes sign under a change of orientation ($\det \underline{A} < 0$), so it is called an “oriented” scalar density.

Next using the previous relation, evaluate

$$\begin{aligned}
 A^{-1j_1}_{i_1} \cdots A^{-1j_n}_{i_n} \eta_{j_1 \cdots j_n} &= (\pm) |\det \underline{G}|^{1/2} \underbrace{\epsilon_{j_1 \cdots j_n} A^{-1j_1}_{i_1} \cdots A^{-1j_n}_{i_n}}_{\epsilon_{i_1 \cdots i_n} \det \underline{A}^{-1}} \\
 &= \underbrace{(\operatorname{sgn} \det \underline{A})(\pm)}_{\text{appropriate sign for new basis}} |\det \underline{G}'|^{1/2} \epsilon_{i_1 \cdots i_n} \equiv \eta_{i'_1 \cdots i'_n}.
 \end{aligned}$$

If $\det A < 0$ this switches the orientation sign as it should, so in fact the transformation law for a $\binom{0}{n}$ -tensor holds, i.e., the above component definition defines the same tensor for every choice of basis

$$\eta = \frac{1}{n!} \eta_{i_1 \cdots i_n} \omega^{i_1 \cdots i_n} = \eta_{1 \cdots n} \omega^{1 \cdots n}.$$

This is called the unit n -form for the oriented inner product vector space. It does two things:

1. It carries the orientation information, with $c \eta$ positively oriented if $c > 0$ and negatively oriented if $c < 0$.
2. It measures n -volume by setting the scale as explained above.

Note that independent of the orientation of the frame, the product satisfies

$$\begin{aligned}
 \eta^{i_1 \cdots i_n} \eta_{j_1 \cdots j_n} &= \operatorname{sgn}(\det \underline{G}) \epsilon^{i_1 \cdots i_n} \epsilon_{j_1 \cdots j_n} = \underbrace{\operatorname{sgn}(\det \underline{G})}_{\equiv (-1)^M} \delta^{i_1 \cdots i_n}_{j_1 \cdots j_n} \\
 &\equiv (-1)^M
 \end{aligned}$$

An *orthonormal basis* $\{e_i\}$ with respect to a given inner product G is one for which each basis vector is a unit vector (with sign $\pm : G_{ij} = G(e_i, e_j) = \pm 1$) orthogonal to the rest ($G_{ij} = 0, i \neq j$). The difference $s = P - M$ (Plus/Minus) in the number of positive and negative signs is called the signature and is fixed for a given inner product (accept as a fact for now; these are just the signs of the eigenvalues). A “positive-definite” inner product has all positive signs, i.e., signature $s = n$, while a “negative-definite” inner product has all negative signs, i.e., signature $s = -n$. An “indefinite” inner product has a signature s in between these two extreme values. A “Lorentz” inner product has only one negative sign or only one positive sign (the choice depends on prejudice, motivated by convenience of competing demands) and so the absolute value of the signature is $|s| = (n - 1) - 1 = n - 2$. Since $n = P + M$, one gets the relation $M = (n - s)/2$.

Remark.

Useful observation

For an orthonormal basis, $|\det \underline{G}|^{1/2} = 1$, so $\eta = \omega^{1 \cdots n}$ if the basis is positively-oriented (i.e., has the same orientation as the chosen one) and $\eta = -\omega^{1 \cdots n}$ otherwise. [On \mathbb{R}^n with the standard inner product and orientation, then $\eta = \omega^{1 \cdots n}$.] Thus η is the n -covector which assigns unit volume to a unit hypercube—the parallelepiped formed by an orthonormal basis.

In an oriented orthonormal frame the sign of the metric determinant is $\det \underline{G} = (-1)^M = \det \underline{G}^{-1}$ so starting from $\eta_{i_1 \cdots i_n} = \epsilon_{i_1 \cdots i_n}$ and raising all its indices leads to $\eta^{i_1 \cdots i_n} = (-1)^M \epsilon^{i_1 \cdots i_n}$. However, in any frame the following relation holds

$$\eta^{i_1 \cdots i_n} \eta_{j_1 \cdots j_n} = (-1)^M \epsilon^{i_1 \cdots i_n} \epsilon_{j_1 \cdots j_n} = (-1)^M \delta^{i_1 \cdots i_n}_{j_1 \cdots j_n}.$$

So what?

Well, now we can define a metric duality operation that has tensor character by using $\eta_{i_1 \dots i_n}$ instead of $\epsilon_{i_1 \dots i_n}$. We will obtain a unique tensor by taking the metric dual, independent of the choice of basis. This will automatically tell us both about p -measure's orientation that generalizes our "counterclockwise" orientation in a plane and its connection to the right handed normal in \mathbb{R}^3 (inner and outer orientations).

▲

The metric dual

We are now in a position to define the metric duality operation associated with an inner product (metric) on our vector space, taking p -vectors to $(n - p)$ -vectors and p -covectors to $(n - p)$ -covectors

$$\begin{aligned} * : \Lambda^{(p)}(V) &\longrightarrow \Lambda^{(n-p)}(V), \\ * : \Lambda^{(p)}(V^*) &\longrightarrow \Lambda^{(n-p)}(V^*). \end{aligned}$$

We modify the natural dual using $\eta_{i_1 \dots i_n}$ in place of $\epsilon_{i_1 \dots i_n}$ and then shift all the indices back to their original level using our inner product (metric). In both cases we contract the unit n -form η or unit n -vector η^\sharp on the left with the components of the p -form or p -vector whose dual is being taken

$$\begin{aligned} \text{for } p\text{-covectors:} \quad [*T]_{i_{p+1} \dots i_n} &= \frac{1}{p!} T_{i_1 \dots i_p} \eta^{i_1 \dots i_p}{}_{i_{p+1} \dots i_n} = T_{i_1 \dots i_p} \eta^{i_1 \dots i_p}{}_{i_{p+1} \dots i_n} \\ \text{for } p\text{-vectors:} \quad [*T]^{i_{p+1} \dots i_n} &= \frac{1}{p!} T^{i_1 \dots i_p} \eta_{i_1 \dots i_p}{}^{i_{p+1} \dots i_n} = T^{i_1 \dots i_p} \eta_{i_1 \dots i_p}{}^{i_{p+1} \dots i_n}, \end{aligned}$$

which is equivalent to the following relations on the basis p -vectors and p -covectors

$$\begin{aligned} *e_{i_1 \dots i_p} &= \eta_{i_1 \dots i_p}{}^{i_{p+1} \dots i_n} e_{i_{p+1} \dots i_n} \\ *\omega^{i_1 \dots i_p} &= \eta^{i_1 \dots i_p}{}_{i_{p+1} \dots i_n} \omega^{i_{p+1} \dots i_n}. \end{aligned}$$

These latter relations simply follow from linearity of this operation, for example,

$$\begin{aligned} *T &= *(T^{i_1 \dots i_p} e_{i_1 \dots i_p}) = T^{i_1 \dots i_p} *e_{i_1 \dots i_p} && \text{(linearity of dual)} \\ &= [*T]^{i_{p+1} \dots i_n} e_{i_{p+1} \dots i_n} && \text{(components of dual)} \\ &= T^{i_1 \dots i_p} \eta_{i_1 \dots i_p}{}^{i_{p+1} \dots i_n} e_{i_{p+1} \dots i_n} && \text{(component definition of dual)}, \end{aligned}$$

so comparing the first and last lines, we can equate the coefficients of $T^{i_1 \dots i_p}$ to get the desired relation.

To extend the dual map to 0-forms and 0-vectors when we are working with p -forms or p -vectors (but not both at the same time since then this becomes ambiguous), we can set $*1 = \eta$ or $*1 = \eta^\sharp$ respectively.

Exercise 4.3.7.**dual of the unit n -form**

Show that $*\eta = (-1)^M$ and that $*(\eta^\sharp) = (-1)^M$.

■

Exercise 4.3.8.**duals in \mathbb{R}^3 : self wedge with dual in \mathbb{R}^3**

On \mathbb{R}^3 with the standard basis, inner-product and orientation, we can make the following Table 4.2 of the duals of all the basis tensors.

	p -vectors to $(n-p)$ -vectors	p -covectors to $(n-p)$ -covectors
	$* : \Lambda^{(p)}(V) \longrightarrow \Lambda^{(n-p)}(V)$	$* : \Lambda^{(p)}(V^*) \longrightarrow \Lambda^{(n-p)}(V^*)$
$p = 0$	$*1 = e_{123}$	$*1 = \omega^{123}$
$p = 1$	$*e_1 = e_{23}$ $*e_2 = e_{31}$ $*e_3 = e_{12}$	$*\omega^1 = \omega^{23}$ $*\omega^2 = \omega^{31}$ $*\omega^3 = \omega^{12}$
$p = 2$	$*e_{23} = e_1$ $*e_{31} = e_2$ $*e_{12} = e_3$	$*\omega^{23} = \omega^1$ $*\omega^{31} = \omega^2$ $*\omega^{12} = \omega^3$
$p = 3$	$*e_{123} = 1$	$*\omega^{123} = 1$

Table 4.2: The duals of the basis multivectors and multicovectors for \mathbb{R}^3 with the usual dot product for which $\eta_{123} = \eta^{123} = 1$.

Note that each of the basis tensors T in the Table 4.2 satisfies $T \wedge *T = e_{123}$. This is no accident. Check.

■

Exercise 4.3.9.**duals for M^3**

How do the signs in Table 4.2 change for 3-dimensional Minkowski spacetime M^3 with $-G_{00} = G_{11} = G_{22} = 1$ and $\eta_{012} = -\eta^{012} = 1$. How does $T \wedge *T$ depend on p ?

■

Exercise 4.3.10.**cross product on \mathbb{R}^3 and M^3**

We can define a cross product for any 3-dimensional vector space with an inner product G

$$X \times Y = *(X \wedge Y) \leftrightarrow [X \times Y]^i = G^{im} \eta_{mjk} X^j Y^k \equiv \eta^i{}_{jk} X^j Y^k.$$

This is just the usual cross-product for \mathbb{R}^3 with its dot product in an orthonormal basis like the standard basis where it is simply $[X \times Y]^i = \epsilon_{ijk} X^j Y^k$.

This cross product is also useful on 3-dimensional Minkowski spacetime M^3 with the inner product matrix $\underline{G} = \text{diag}(-1, 1, 1) = \underline{G}^{-1}$, where the span of any two spacelike vectors defines a local rest space for an observer moving orthogonally to them in spacetime. Their cross product necessarily defines a timelike vector which defines the local time line for this local rest space, namely a normal to the 2-plane they form. The unit normal reversed in sign defines the corresponding future-pointing spacetime-velocity.

To be more concrete since $\eta_{012} = 1$ sets the unit-volume 3-form, raising the indices in an orthonormal basis leads to $\eta^{012} = -1 = \eta^{023}$, so $[X \times Y]^i = \eta^i{}_{jk} X^j Y^k$ has a timelike component $[X \times Y]^0 = \eta^0{}_{12}(X^1 Y^2 - X^2 Y^1)$. Thus for spacelike vectors X, Y which in the Euclidean case have an upward cross-product in that order, the Minkowski cross-product will point downward because $\eta^0{}_{12} = -1$, so to re-establish the right hand rule giving an upward future-pointing normal to the plane of these two vectors in this order, one has to reverse the overall vector: $-X \times Y$ satisfies the right hand rule.

Show this by considering orientation by considering $-\langle 0, 1, 0 \rangle \times \langle 0, 0, 1 \rangle$ and seeing that its first component is positive, i.e., points in the future time direction. The index raising of the natural dual simply changes the sign of the first component of the Euclidean cross product, so changing the overall sign of all the components restores the right hand rule. This is not necessary for the cross product of a timelike and a spacelike vector since the result is spacelike. Check this with the standard basis vectors, say $\langle 1, 0, 0 \rangle$ and $\langle 0, 1, 0 \rangle$. ■

Example 4.3.1. On \mathbb{R}^4 with the standard basis and standard orientation, consider instead the special relativity inner product $\underline{G} = \text{diag}(1, 1, 1, -1)$, i.e., e_4 has a negative sign. Then

$$\eta_{1234} = 1 = \eta^1{}_{234} = \eta^{12}{}_{34} = -\eta_{123}{}^4 = -\eta_{12}{}^{34}$$

etc. since \underline{G} is diagonal, so

$$\eta_{123}{}^4 = -\eta_{1234} G^{44} = -\eta_{1234}$$

etc. □

Exercise 4.3.11. double dual sign

Starting from Exercise 4.2.1 for the double natural dual and the definition of $\eta^{i_1 \dots i_n}$, show that in an oriented frame, the double dual of a p -form T satisfies

$$**T = (-1)^{M+p(n-p)} T,$$

where $(-1)^M = \text{sgn det } \underline{G}$.

From this one can introduce the inverse dual map $*^{-1}$ by $*^{-1}*T = T = **^{-1}T$ and conclude by comparing the second of these equalities with the double dual identity that

$$*^{-1}T = (-1)^{M+p(n-p)}*T.$$

■

The variable part of this sign can be rewritten as $(-1)^{p(n-p)} = (-1)^{pn}(-1)^{p^2}$ but $(-1)^{p^2}$ has the same sign as $(-1)^p$ (since if p is even it is positive, but if p is odd, so is p^2 , and the sign is negative) so $(-1)^{p(n-p)} = (-1)^{pn}(-1)^p = (-1)^{(n-1)p}$. Thus if n is odd like $n = 3$, this sign is positive, independent of p . If n is even like $n = 4$, then the sign is $(-1)^p$, which alternates as p is increased from 0 to 4.

Exercise 4.3.12.

inverse of dual

If S is a 1-form and T is a p -form, simplify the expression $*^{-1}(S \wedge *T)$, i.e., what is the appropriate sign factor needed to replace $*^{-1}$ by $*$ here? What is the degree q of the resulting q -form? This sign formula will return when we discuss the codifferential in Part 2.

■

Exercise 4.3.13.

double dual sign for $n = 4$

The metric dual is a map from p -vectors (-covectors) to $(n-p)$ -vectors (-covectors) so when n is even, the dual maps $(n/2)$ -vectors (-covectors) to $(n/2)$ -vectors (-covectors) and is thus a linear transformation of this space into itself which preserves inner products. For \mathbb{R}^2 then the dual of vectors are again (orthogonal) vectors, while for \mathbb{R}^4 , the dual of 2-vectors are again (orthogonal) 2-vectors. The double dual on these objects satisfies $** = (-1)^{M+n/2}$. One can then consider the eigenvalue problem $*X = \lambda X$ for such objects, with $\lambda^2 = (-1)^{M+n/2}$.

a) For $n = 4$ and $M = 0$, the case of the usual Euclidean inner product on \mathbb{R}^4 , then for 2-vectors, $\lambda^2 = 1$, so $\lambda = \pm 1$. Find a basis of the space of 2-vectors which are eigenvectors of the duality operation. The two eigenspaces of opposite signed eigenvalues are called self-dual and anti-self dual 2-vectors.

b) For $n = 4$ and $M = 1$, the case of the Lorentian inner product on \mathbb{R}^4 , then for 2-vectors, $\lambda^2 = -1$, so $\lambda = \pm i$. Find a basis of the space of complex 2-vectors which are eigenvectors of the duality operation. The two eigenspaces of opposite signed eigenvalues are called self-dual and anti-self dual 2-vectors, although here the overall sign correlation depends on other conventions.

■

Inner product, duality and wedge product relations

Suppose V is an n -dimensional vector space with basis $\{e_i\}_{i=1,\dots,n}$ and W is an m -dimensional vector space with basis $\{E_\alpha\}_{\alpha=1,\dots,m}$. Let $\{\omega^i\}$ and $\{W^\alpha\}$ be the respective dual basis. Let $A : V \rightarrow W$ be a linear map. Then by linearity

$$A(v) = A(v^i e_i) = v^i A(e_i),$$

i.e., the map is completely determined by its values on the basis vectors. For each i , $A(e_i) \in W$ can be expressed in terms of its components with respect to $\{E_\alpha\}$

$$A(e_i) = A^\alpha_i E_\alpha, \quad A^\alpha_i = W^\alpha(A(e_i)).$$

Thus

$$w = A(v) = v^i A^\alpha_i E_\alpha = [A^\alpha_i v^i] E_\alpha$$

becomes

$$w^\alpha = A^\alpha_i v^i \quad \text{in components, equivalent to}$$

$$A(e_i) = A^\alpha_i E_\alpha \quad \text{on the basis vectors.}$$

The matrix $\underline{A} = (A^\alpha_i)$ is called the matrix representation of A with respect to the bases $\{e_i\}$ and $\{E_\alpha\}$ of V and W respectively. If either basis changes, the matrix will change in an “obvious” way (obvious when you see it). Consider the following changes of basis on both spaces

$$\begin{aligned} e_{i'} &= B^{-1j}_i e_j, & \omega^{i'} &= B^i_j \omega^j, \\ E_{\alpha'} &= C^{-1\beta}_\alpha E_\beta, & W^{\alpha'} &= C^\alpha_\beta W^\beta. \end{aligned}$$

Then

$$\begin{aligned} A^{\alpha'}_{i'} &= W^{\alpha'}(A(e_{i'})) = C^\alpha_\beta W^\beta(A(e_j B^{-1j}_i)) \\ &= C^\alpha_\beta W^\beta(A(e_j)) B^{-1j}_i = C^\alpha_\beta A^\beta_j B^{-1j}_i \end{aligned}$$

or

$$\underline{A}' = \underline{C} \underline{A} \underline{B}^{-1}.$$

When $V = W$ are $e_i = E_i$, this reduces to the more familiar result

$$A' = \underline{C} \underline{A} \underline{C}^{-1}.$$

For a given vector space V each space $T^{(p,q)}(V)$ of tensors of a given “index type,” or subspaces with certain symmetries like $\Lambda^{(p)}(V)$ and $\Lambda^{(p)}(V^*)$, is a vector space in its own right. However, instead of labeling the basis vectors in these spaces by a subscript label taking values between 1 and the dimension of the space, we use collections of indices associated with the underlying space V . The linear operations we have introduced all correspond to various linear maps between these spaces which can be expressed either in “component” form or as a relation between the new and old basis vectors, which defines the “matrix” of the linear transformation—but matrix in this generalized sense of one index corresponding to a collection of indices.

The “index-shifting” maps associated with an inner product or “metric” G are a perfect example. Considering the “lowering” map on V

$$\flat : V \longrightarrow V^* \quad [X^\flat]_i \equiv X_i = G_{ij}X^j \quad (\text{component relation})$$

or evaluating the index lowering in two different ways

$$\begin{aligned} X^\flat &= X_i \omega^i = G_{ij} X^j \omega^i \\ &= (X^j e_j)^\flat = X^j e_j^\flat \quad (\text{by linearity}) \end{aligned}$$

one finds that

$$e_j^\flat = G_{ij} \omega^i \quad (\text{basis relation}).$$

Similarly

$$\begin{aligned} \sharp : V^* \longrightarrow V \quad [f^\sharp]^i &= f^i = G^{ij} f_j \quad (\text{component relation}), \\ \omega^{j\sharp} &= G^{ij} e_j \quad (\text{basis relation}). \end{aligned}$$

Exercise 4.3.14.

index shifting

Verify this as above. ■

The index shifting maps can be extended to any collection of indices for any space of tensors of a given type. The \flat and \sharp notation will always indicate shifting all the indices down and up respectively. In particular for p -vectors and p -covectors one can translate the component relations

$$\begin{aligned} T_{i_1 \dots i_p} &= G_{i_1 j_1} \dots G_{i_p j_p} T^{i_1 \dots i_p}, \\ T^{i_1 \dots i_p} &= G^{i_1 j_1} \dots G^{i_p j_p} T_{i_1 \dots i_p} \end{aligned}$$

into the basis relations

$$\begin{aligned} e_{i_1 \dots i_p}^\flat &= G_{i_1 j_1} \dots G_{i_p j_p} \omega^{i_1 \dots i_p}, \\ \omega^{i_1 \dots i_p \sharp} &= G^{i_1 j_1} \dots G^{i_p j_p} e_{i_1 \dots i_p} \quad (\text{Exercise: verify these}) \end{aligned}$$

for the maps

$$\flat : \Lambda^{(p)}(V) \longrightarrow \Lambda^{(p)}(V^*), \quad \sharp : \Lambda^{(p)}(V^*) \longrightarrow \Lambda^{(p)}(V).$$

In fact it is natural to interpret $\Lambda^{(p)}(V^*)$ as the dual space to $\Lambda^{(p)}(V)$ since the natural contraction

$$T_{i_1 \dots i_p} S^{i_1 \dots i_p} = \frac{1}{p!} T_{i_1 \dots i_p} S^{i_1 \dots i_p}$$

is linear both in the p -vector S and the p -covector T , so fixing either factor produces a real-valued linear function of the other.

$\{\omega^{i_1 \dots i_p}\}$ is the basis dual to $\{e_{i_1 \dots i_p}\}$, and their duality relation is that $\omega^{i_1 \dots i_p}$ contracted on $e_{i_1 \dots i_p}$ equals $\delta^{i_1 \dots i_p}_{j_1 \dots j_p}$ (which is just the Kronecker delta for these two spaces).

The inner product on $\Lambda^{(p)}(V)$

$$\langle T, S \rangle = \frac{1}{p!} T^{i_1 \dots i_p} G_{i_1 j_1} \dots G_{i_p j_p} S^{j_1 \dots j_p} \equiv \frac{1}{p!} T_{j_1 \dots j_p} S^{j_1 \dots j_p}$$

induces the inner product on $\Lambda^{(p)}(V^*)$

$$\langle T, S \rangle = \frac{1}{p!} T_{i_1 \dots i_p} G^{i_1 j_1} \dots G^{i_p j_p} S_{j_1 \dots j_p} \equiv \frac{1}{p!} T^{j_1 \dots j_p} S_{j_1 \dots j_p}$$

for which the above relations are the component and basis relations for the two index shifting maps between these two spaces (for each p). The “matrix” of this inner product is

$$\langle e_{i_1 \dots i_p}, e_{j_1 \dots j_p} \rangle = \delta^{k_1 \dots k_p}_{i_1 \dots i_p} G_{k_1 j_1} \dots G_{k_p j_p}.$$

Both the natural dual and the metric dual are linear maps among these spaces which are completely determined by their values on the basis tensors

$$\begin{aligned} (*) e_{i_1 \dots i_p} &= e_{i_1 \dots i_p i_{p+1} \dots i_n} \omega^{i_{p+1} \dots i_n}, \\ {}^* e_{i_1 \dots i_p} &= \eta_{i_1 \dots i_p i_{p+1} \dots i_n} \omega^{i_{p+1} \dots i_n} \# = \eta_{i_1 \dots i_p} {}^{i_{p+1} \dots i_n} e_{i_{p+1} \dots i_n} \\ &= \eta_{i_1 \dots i_p i_{p+1} \dots i_n} G^{i_{p+1} j_{p+1}} \dots G^{i_n j_n} e_{j_{p+1} \dots j_n}, \\ {}^* \omega^{i_1 \dots i_p} &= \eta^{i_1 \dots i_p}_{i_{p+1} \dots i_n} \omega^{i_{p+1} \dots i_n} \\ &= \eta^{i_1 \dots i_p i_{p+1} \dots i_n} G_{i_{p+1} j_{p+1}} \dots G_{i_n j_n} \omega^{j_{p+1} \dots j_n}. \end{aligned}$$

In an oriented orthonormal frame, the natural and metric duals are very closely related, with only the diagonal components of the metric nonzero and equal to $G_{ii} = e_i \cdot e_i = \pm 1 = G^{ii}$ and $\eta_{1 \dots n} = 1 = (-1)^M \eta^{1 \dots n}$, where $(-1)^M = G_{11} \dots G_{nn}$. The metric dual then simplifies to

$$\begin{aligned} {}^* e_{i_1 \dots i_p} &= \eta_{i_1 \dots i_p} {}^{i_{p+1} \dots i_n} e_{i_{p+1} \dots i_n} = \eta_{i_1 \dots i_p i_{p+1} \dots i_n} G^{i_{p+1} j_{p+1}} \dots G^{i_n j_n} e_{j_{p+1} \dots j_n}, \\ {}^* \omega^{i_1 \dots i_p} &= \eta^{i_1 \dots i_p}_{i_{p+1} \dots i_n} \omega^{i_{p+1} \dots i_n} = \eta^{i_1 \dots i_p i_{p+1} \dots i_n} G_{i_{p+1} j_{p+1}} \dots G_{i_n j_n} \omega^{j_{p+1} \dots j_n}, \end{aligned}$$

which only has sign changes relative to the Euclidean inner product for other signature inner products.

Exercise 4.3.15.

2-vector duals in \mathbb{R}^4

For a 4-dimensional vector space, 2-vectors are mapped into 2-vectors by the metric dual associated with any inner product. This is a linear transformation of the space of 2-vectors into itself, so if we specify a basis, we can find the matrix of this linear transformation.

On \mathbb{R}^4 with the standard orthonormal basis with respect to the Euclidean dot product, introduce the following ordered basis of the space of 2-vectors

$$\{E_1, E_2, E_3, E_4, E_5, E_6\} = \{e_{23}, e_{31}, e_{12}, e_{14}, e_{24}, e_{34}\}.$$

a) Show that the matrix of the metric dual on this vector space defined by ${}^*E_\alpha = D^\beta{}_\alpha E_\beta$, $\alpha, \beta = 1, 2, 3, 4, 5, 6$ is

$$\underline{D} = \begin{pmatrix} \underline{0}_{3,3} & \underline{I}_3 \\ \underline{I}_3 & \underline{0}_{3,3} \end{pmatrix},$$

where $\underline{0}_{3,3}$ is the 3×3 zero matrix and \underline{I}_3 is the 3×3 identity matrix.

b) Show that this basis is orthonormal: $E_\alpha \cdot E_\beta = \delta_{\alpha\beta}$.

c) Suppose we change to the Lorentzian inner product where $e_i \cdot e_j = \eta_{ij}$, with $-\eta_{00} = \eta_{11} = \eta_{22} = \eta_{33} = 1$. Show that

$$\underline{D} = \begin{pmatrix} \underline{0}_{3,3} & -\underline{I}_3 \\ \underline{I}_3 & \underline{0}_{3,3} \end{pmatrix},$$

and that the basis E_α is orthonormal, with self-dot products

$$E_1 \cdot E_1 = E_2 \cdot E_2 = E_3 \cdot E_3 = 1, \quad E_4 \cdot E_4 = E_5 \cdot E_5 = E_6 \cdot E_6 = -1.$$

■

The metric dual turns out to be very closely related to the inner product for p -vectors and p -covectors. The following calculation establishes that simple relationship. If T and S are both p -covectors, then *S is an $(n-p)$ -covector, so $T \wedge {}^*S$ is an n -covector, with only one independent component, which is in turn its dual, a real number, thus defining an inner product. Using the contraction identity for a pair of alternating symbols in the second line, one has

$$\begin{aligned} T \wedge {}^*S &= \left(\frac{1}{p!} T_{i_1 \dots i_p} \omega^{i_1 \dots i_p} \right) \wedge \left(\frac{1}{(n-p)! p!} S^{j_1 \dots j_p} \eta_{j_1 \dots j_p i_{p+1} \dots i_n} \omega^{i_{p+1} \dots i_n} \right) \\ &= \left(\frac{1}{p!} \right)^2 T_{i_1 \dots i_p} S^{j_1 \dots j_p} \underbrace{\eta_{j_1 \dots j_p i_{p+1} \dots i_n}}_{\eta_{1 \dots n} \epsilon_{j_1 \dots j_p i_{p+1} \dots i_n}} \underbrace{\frac{1}{(n-p)!} \omega^{i_{p+1} \dots i_n}}_{\epsilon^{i_1 \dots i_n} \omega^{1 \dots n}} \\ &= \left(\frac{1}{p!} \right)^2 T_{i_1 \dots i_p} S^{j_1 \dots j_p} \underbrace{\frac{1}{(n-p)!} \delta_{j_1 \dots j_p i_{p+1} \dots i_n}^{i_1 \dots i_n}}_{\delta_{j_1 \dots j_p}^{i_1 \dots i_p}} \underbrace{\eta_{i_1 \dots i_n} \omega^{1 \dots n}}_{\eta} \\ &= \underbrace{\left(\frac{1}{p!} \right) T_{i_1 \dots i_p} S^{i_1 \dots i_p}}_{\langle T, S \rangle} = \langle T, S \rangle \eta \end{aligned}$$

Thus we get the natural inner product we already have defined, but now defined without indices in terms of the wedge product and duality operation

$$T \wedge {}^*S = \langle T, S \rangle \eta,$$

The same formula holds for p -vectors with η^\sharp in place of η .

In terms of the basis tensors, in an oriented orthonormal frame, one has

$$e_{i_1 \dots i_p} \wedge {}^*e_{i_1 \dots i_p} = \langle e_{i_1 \dots i_p}, e_{i_1 \dots i_p} \rangle \eta^\sharp,$$

where the left factor is the product of the signs of the basis vectors

$$\langle e_{i_1 \dots i_p}, e_{i_1 \dots i_p} \rangle = (e_{i_1} \cdot e_{i_1}) \cdots (e_{i_p} \cdot e_{i_p})$$

and

$$\eta^\sharp = G^{11} \cdots G^{nn} e_{1 \dots n} = (-1)^M e_{1 \dots n}$$

is the product of all possible such signs times $e_{1 \dots n}$, while $*e_{i_1 \dots i_p}$ is the $(n-p)$ -covector needed to wedge into $e_{i_1 \dots i_p}$ from the right to get the product of signs associated with “complementary indices” times $e_{1 \dots n}$, namely the product of signs $(e_{i_{p+1}} \cdot e_{i_{p+1}}) \cdots (e_{i_n} \cdot e_{i_n})$.

Exercise 4.3.16.

inner product of two duals for a general inner product; EDIT THIS

- (i) Why “complementary indices”?
- (ii) Use the component definitions to show that $\langle *T, *S \rangle = (-1)^M \langle T, S \rangle$ for two p -vectors T and S .
- (iii) On \mathbb{R}^3 define the cross product by $*(X \wedge Y) = X \times Y$, for any two vectors X and Y . What are the components $[X \times Y]^i$? (Verify that they are what you expect.)

■

Exercise 4.3.17.

M^4 duals with indices 0,1,2,3

For $n = 4$ with the Lorentz inner product (metric) $-\eta_{00} = \eta_{11} = \eta_{22} = \eta_{33} = 1$, and $\eta_{0123} = 1 = -\eta^{0123}$, one has the following table 4.3 of all the metric duals of the bases of the various p -vector/covector spaces, where again the cyclic permutations of (1,2,3) dominate many formulas which are naturally categorized by having 0 or 1 index values which are 0.

a) Verify the relation $**T = (-1)^{1+p(4-p)}T = (-1)^{p-1}T$ for all the multivector or multiform basis tensors.

b) For a 1-form

$$U = U_0 \omega^0 + U_a \omega^a, \quad (a = 1, 2, 3)$$

evaluate $*U$.

c) For a 2-form

$$F^\flat = E_a \omega^{a0} + B_1 \omega^{23} + B_2 \omega^{31} + B_3 \omega^{12},$$

evaluate $*F^\flat$, $F^\flat \wedge *F^\flat$, and $F^\flat \wedge F^\flat$. Evaluate the matrix of mixed components $\underline{F} = (F^\alpha_\beta)$ and $*\underline{F} = (*F^\alpha_\beta)$ and compare with Exercise 4.6.6.

d) For a 3-form

$$J = J^0 \omega^{123} - J^1 \omega^{023} + J^2 \omega^{031} - J^3 \omega^{012},$$

evaluate $*J$.

e) Evaluate $U \wedge J$.

■

	p -vectors to $(n-p)$ -covectors	p -covectors to $(n-p)$ -vectors
	$* : \Lambda^{(p)}(V) \longrightarrow \Lambda^{(n-p)}(V^*)$	$* : \Lambda^{(p)}(V) \longrightarrow \Lambda^{(n-p)}(V^*)$
$p = 0$	$*1 = \eta^{0123}e_{0123} = -e_{0123}$	$*1 = \eta_{0123}\omega^{0123} = \omega^{0123}$
$p = 1$	$*e_1 = \eta_1^{023}e_{023} = e_{023}$ $*e_2 = \eta_2^{031}e_{031} = e_{031}$ $*e_3 = \eta_3^{012}e_{012} = e_{012}$	$*\omega^1 = \eta^1_{023}\omega^{023} = -\omega^{023}$ $*\omega^2 = \eta^2_{031}\omega^{031} = -\omega^{031}$ $*\omega^3 = \eta^3_{012}\omega^{012} = -\omega^{012}$
	$*e_0 = \eta_0^{123}e_{123} = e_{123}$	$*\omega^0 = \eta^0_{123}\omega^{123} = -\omega^{123}$
$p = 2$	$*e_{23} = \eta_{23}^{01}e_{01} = -e_{01}$ $*e_{31} = \eta_{31}^{02}e_{02} = -e_{02}$ $*e_{12} = \eta_{12}^{03}e_{03} = -e_{03}$	$*\omega^{23} = \eta^{23}_{01}\omega^{01} = \omega^{01}$ $*\omega^{31} = \eta^{31}_{02}\omega^{02} = \omega^{02}$ $*\omega^{12} = \eta^{12}_{03}\omega^{03} = \omega^{03}$
	$*e_{01} = \eta_{01}^{23}e_{23} = e_{23}$ $*e_{02} = \eta_{02}^{31}e_{31} = e_{31}$ $*e_{03} = \eta_{03}^{12}e_{12} = e_{12}$	$*\omega^{01} = \eta^{01}_{23}\omega^{23} = -\omega^{23}$ $*\omega^{02} = \eta^{02}_{31}\omega^{31} = -\omega^{31}$ $*\omega^{03} = \eta^{03}_{12}\omega^{12} = -\omega^{12}$
$p = 3$	$*e_{123} = \eta_{123}^0e_0 = -e_0$	$*\omega^{123} = \eta^{123}_0\omega^0 = \omega^0$
	$*e_{023} = \eta_{023}^1e_1 = -e_1$ $*e_{031} = \eta_{031}^2e_2 = -e_2$ $*e_{012} = \eta_{012}^3e_3 = -e_3$	$*\omega^{023} = \eta^{023}_1\omega^1 = \omega^1$ $*\omega^{031} = \eta^{031}_2\omega^2 = \omega^2$ $*\omega^{012} = \eta^{012}_3\omega^3 = \omega^3$
$p = 4$	$*e_{0123} = \eta_{0123} = 1$	$*\omega^{0123} = \eta^{0123} = -1$

Table 4.3: The table of metric duals for the Lorentz inner product on M^4 with $-G_{00} = 1 = G_{11} = G_{22} = G_{33}$ and $\eta_{0123} = -\eta^{0123} = 1$. Notice that $** = (-1)^{p-1}$.

Exercise 4.3.18.

M^4 duals with indices 1,2,3,4

For $n = 4$ with the Lorentz metric $\eta_{11} = \eta_{22} = \eta_{33} = 1 = -\eta_{44}$, and $\eta_{1234} = 1 = -\eta^{1234}$, one has the following table 4.4 of all the natural duals of the bases of the various p -vector/covector spaces, where again the cyclic permutations of (1,2,3) dominate many formulas which are naturally categorized by having 0 or 1 index values which are 4.

a) Verify the relation $**T = (-1)^{1+p(4-p)}T = (-1)^{p-1}T$ for all the multivector or multiform basis tensors.

b) For a 1-form

$$U = U_a\omega^a + U_4\omega^4, \quad (a = 1, 2, 3)$$

evaluate $*U$.

c) For a 2-form

$$F^\flat = E_a\omega^{a4} + B_1\omega^{23} + B_2\omega^{31} + B_3\omega^{12},$$

	p -vectors to $(n-p)$ -covectors	p -covectors to $(n-p)$ -vectors
	$* : \Lambda^{(p)}(V) \longrightarrow \Lambda^{(n-p)}(V^*)$	$* : \Lambda^{(p)}(V) \longrightarrow \Lambda^{(n-p)}(V^*)$
$p = 0$	$*1 = \eta^{1234}e_{1234} = -e_{1234}$	$*1 = \eta_{1234}\omega^{1234} = \omega^{1234}$
$p = 1$	$*e_1 = \eta_1^{234}e_{234} = -e_{234}$ $*e_2 = \eta_2^{314}e_{314} = -e_{314}$ $*e_3 = \eta_3^{124}e_{124} = -e_{124}$	$*\omega^1 = \eta^1_{234}\omega^{234} = \omega^{234}$ $*\omega^2 = \eta^2_{314}\omega^{314} = \omega^{314}$ $*\omega^3 = \eta^3_{124}\omega^{124} = \omega^{124}$
	$*e_4 = \eta_4^{123}e_{123} = -e_{123}$	$*\omega^4 = \eta^4_{123}\omega^{123} = \omega^{123}$
$p = 2$	$*e_{23} = \eta_{23}^{14}e_{14} = -e_{14}$ $*e_{31} = \eta_{31}^{24}e_{24} = -e_{24}$ $*e_{12} = \eta_{12}^{34}e_{34} = -e_{34}$	$*\omega^{23} = \eta^{23}_{14}\omega^{14} = \omega^{14}$ $*\omega^{31} = \eta^{31}_{24}\omega^{24} = \omega^{24}$ $*\omega^{12} = \eta^{12}_{34}\omega^{34} = \omega^{34}$
	$*e_{14} = \eta_{14}^{23}e_{23} = e_{23}$ $*e_{24} = \eta_{24}^{31}e_{31} = e_{31}$ $*e_{34} = \eta_{34}^{12}e_{12} = e_{12}$	$*\omega^{14} = \eta^{14}_{23}\omega^{23} = -\omega^{23}$ $*\omega^{24} = \eta^{24}_{31}\omega^{31} = -\omega^{31}$ $*\omega^{34} = \eta^{34}_{12}\omega^{12} = -\omega^{12}$
$p = 3$	$*e_{123} = \eta_{123}^4e_4 = -e_4$	$*\omega^{123} = \eta^{123}_4\omega^4 = \omega^4$
	$*e_{234} = \eta_{234}^1e_1 = -e_1$ $*e_{314} = \eta_{314}^2e_2 = -e_2$ $*e_{124} = \eta_{124}^3e_3 = -e_3$	$*\omega^{234} = \eta^{234}_1\omega^1 = \omega^1$ $*\omega^{314} = \eta^{314}_2\omega^2 = \omega^2$ $*\omega^{124} = \eta^{124}_3\omega^3 = \omega^3$
$p = 4$	$*e_{1234} = \eta_{1234} = 1$	$*\omega^{1234} = \eta^{1234} = -1$

Table 4.4: The table of metric duals for the Lorentz inner product on \mathbb{R}^4 with $1 = G_{11} = G_{22} = G_{33} = -G_{44}$ and $\eta_{1234} = -\eta^{1234}$. Notice that $** = (-1)^{p-1}$.

evaluate $*F^\flat$, $F^\flat \wedge *F^\flat$, and $F^\flat \wedge F^\flat$. Evaluate the matrix of mixed components $\underline{F} = (F^\alpha_\beta)$ and $*\underline{F} = (*F^\alpha_\beta)$ and compare with Exercise 1.6.6.

d) For a 3-form

$$J = J_1\omega^{234} + J_2\omega^{314} + J_3\omega^{124} + J_4\omega^{123},$$

evaluate $*J$.

e) Evaluate $U \wedge J$.

■

Exercise 4.3.19.

Euclidean \mathbb{R}^4 duals

How do the signs in table 4.3 change if instead of the Lorentian inner product on \mathbb{R}^4 , we use the usual Euclidean inner product? This changes $\eta^{1234} = -1$ to $\eta^{1234} = 1$ and $G_{44} = -1$ to

$\delta_{44} = 1$. Check the signs in Table 4.5. ■

	p -vectors to $(n-p)$ -covectors	p -covectors to $(n-p)$ -vectors
	$* : \Lambda^{(p)}(V) \longrightarrow \Lambda^{(n-p)}(V^*)$	$* : \Lambda^{(p)}(V) \longrightarrow \Lambda^{(n-p)}(V^*)$
$p = 0$	$*1 = \eta^{1234}e_{1234} = e_{1234}$	$*1 = \eta_{1234}\omega^{1234} = \omega^{1234}$
$p = 1$	$*e_1 = \eta_1^{234}e_{234} = e_{234}$ $*e_2 = \eta_2^{314}e_{314} = e_{314}$ $*e_3 = \eta_3^{124}e_{124} = e_{124}$	$*\omega^1 = \eta^1_{234}\omega^{234} = \omega^{234}$ $*\omega^2 = \eta^2_{314}\omega^{314} = \omega^{314}$ $*\omega^3 = \eta^3_{124}\omega^{124} = \omega^{124}$
	$*e_4 = \eta_4^{123}e_{123} = -e_{123}$	$*\omega^4 = \eta^4_{123}\omega^{123} = -\omega^{123}$
$p = 2$	$*e_{23} = \eta_{23}^{14}e_{14} = e_{14}$ $*e_{31} = \eta_{31}^{24}e_{24} = e_{24}$ $*e_{12} = \eta_{12}^{34}e_{34} = e_{34}$	$*\omega^{23} = \eta^{23}_{14}\omega^{14} = \omega^{14}$ $*\omega^{31} = \eta^{31}_{24}\omega^{24} = \omega^{24}$ $*\omega^{12} = \eta^{12}_{34}\omega^{34} = \omega^{34}$
	$*e_{14} = \eta_{14}^{23}e_{23} = e_{23}$ $*e_{24} = \eta_{24}^{31}e_{31} = e_{31}$ $*e_{34} = \eta_{34}^{12}e_{12} = e_{12}$	$*\omega^{14} = \eta^{14}_{23}\omega^{23} = \omega^{23}$ $*\omega^{24} = \eta^{24}_{31}\omega^{31} = \omega^{31}$ $*\omega^{34} = \eta^{34}_{12}\omega^{12} = \omega^{12}$
$p = 3$	$*e_{123} = \eta_{123}^4e_4 = e_4$	$*\omega^{123} = \eta^{123}_4\omega^4 = \omega^4$
	$*e_{234} = \eta_{234}^1e_1 = -e_1$ $*e_{314} = \eta_{314}^2e_2 = -e_2$ $*e_{124} = \eta_{124}^3e_3 = -e_3$	$*\omega^{234} = \eta^{234}_1\omega^1 = -\omega^1$ $*\omega^{314} = \eta^{314}_2\omega^2 = -\omega^2$ $*\omega^{124} = \eta^{124}_3\omega^3 = -\omega^3$
$p = 4$	$*e_{1234} = \eta_{1234} = 1$	$*\omega^{1234} = \eta^{1234} = 1$

Table 4.5: The table of metric duals for the Euclidean inner product on \mathbb{R}^4 , where $\eta_{1234} = \eta^{1234}$ in the standard orthonormal basis. Notice that now $** = (-1)^p$.

Exercise 4.3.20.

complex plane and real wedge products

The field \mathbb{C} of complex numbers is a 2-dimensional real vector space isomorphic to \mathbb{R}^2 through the isomorphism $z = x + iy \leftrightarrow (x, y)$ which associates the basis $\{1, i\}$ with the standard basis $\{e_1 = (1, 0), e_2 = (0, 1)\}$. Let $C(u) = C((u^1, u^2)) = u^1 + iu^2$ be the explicit isomorphism, and denote the complex conjugate by $\bar{z} = x - iy$.

a) Show that complex multiplication encodes both the dot product and determinant functions nicely by showing that

$$\overline{C(u)}C(v) = u \cdot v + i^*(u \wedge v),$$

where $u \cdot v = u^1 v^1 + u^2 v^2$ and

$$*(u \wedge v) = \det(u, v) = \begin{vmatrix} u^1 & v^1 \\ u^2 & v^2 \end{vmatrix} = u^1 v^2 - u^2 v^1.$$

Thus the imaginary part of the product of the complex conjugate of one complex number with another complex number is just the signed area of the parallelogram formed by the corresponding vectors in \mathbb{R}^2 , while the real part is just their dot product.

b) For a unit vector \hat{u} , $C(\hat{u})$ is a unit complex number: $|C(u)| = (\overline{C(u)}C(u))^{1/2} = 1$. Show that by introducing polar coordinates of points in \mathbb{R}^2 : $u^1 = r \cos \theta$, $u^2 = r \sin \theta$, the corresponding complex number is put into polar form with its magnitude given by r and its “argument” given by θ : $C(u) = r e^{i\theta}$. Thus the unit circle in \mathbb{R}^2 corresponds to the unit circle of unit complex numbers in the complex plane.

c) For two unit vectors \hat{u}, \hat{v} , introduce a signed angle $\theta \in [-\pi, \pi]$ between them measured in the counterclockwise direction from u to v . Then $\hat{u} \cdot \hat{v} = \cos \theta$ while $*(u \wedge v) = \sin \theta$, so this becomes

$$\overline{C(\hat{u})}C(\hat{v}) = \cos \theta + i \sin \theta = e^{i\theta}$$

so by linearity

$$\overline{C(u)}C(v) = |u||v|(\cos \theta + i \sin \theta) = |u||v|e^{i\theta}.$$

Cute. To justify this verify that if $u = |u|(\cos \theta_1, \sin \theta_1)$ and $v = |v|(\cos \theta_2, \sin \theta_2)$, then

$$*(u \wedge v) = \det(u, v) = \begin{vmatrix} u^1 & v^1 \\ u^2 & v^2 \end{vmatrix} = |u||v| \sin(\theta_2 - \theta_1) \equiv |u||v| \sin \theta.$$

■

Remark.

This stuff is too good to let slide. The idea generalizes in various ways. Quaternions are the simplest next step in which one considers a product on \mathbb{R}^4

$$\begin{aligned} u &= u^i e_i + u^4 e_4 \equiv \vec{u} + u^4 e_4, \quad v = v^i e_i + v^4 e_4 \equiv \vec{v} + v^4 e_4, \\ uv &= {}^{*(3)}(\vec{u} \wedge \vec{v}) + u^4 \vec{v} + v^4 \vec{u} + (u^4 v^4 - \vec{u} \cdot \vec{v}) e_4 \end{aligned}$$

where ${}^{*(3)}\vec{u} \wedge \vec{v} = \vec{u} \times \vec{v}$ involves the 3-dimensional duality operation on 3-vectors to get the cross product and $i = 1, 2, 3$. In analogy with the complex numbers quaternions are usually written

$$Q(u) = u^1 i + u^2 j + u^3 k + u^4 1$$

under which one has the isomorphism which sends the standard basis of \mathbb{R}^4 to $\{i, j, k, 1\}$, putting the “real part” last so we can take advantage of the $(1, 2, 3)$ indices correlating with i, j, k . Introduce the conjugate of u and $Q(u)$ by $\bar{u} = -u^i e_i + u^4 e_4$,

$$\overline{Q(u)} = Q(\bar{u}) = -u^1 i - u^2 j - u^3 k + u^4 1.$$

The magnitude of a quaternion is just the magnitude of the corresponding vector with the usual Euclidean dot product on \mathbb{R}^4

$$|Q(u)| = (\overline{Q(u)}Q(u))^{1/2} = (\bar{u}u)^{1/2} = (u \cdot u)^{1/2} = |u|,$$

so unit quaternions correspond to the unit sphere S_3 in \mathbb{R}^4 .

Another possible next step is “geometric algebra” in which you don’t take the dual of $u \wedge v$ but work with a linear combination of multivectors of different rank. You can easily find information about either quaternions or geometric algebra on the web. Both approaches turn out to be extremely efficient in describing the mathematics of the rotation group, and so are not just mathematical games playing. Remember, computer graphics relies heavily on being able to rotate objects around in space freely, so even computer animated films and the computer gaming industry need this stuff.

For example, one can calculate that the unit quaternion

$$q(\theta, \hat{n}) = \cos \theta/2 + \sin \theta/2 (n^1 i + n^2 j + n^3 k)$$

has the effect

$$q(x^1 i + x^2 j + x^3 k)\bar{q} = (\underline{R}x)^1 i + (\underline{R}x)^2 j + (\underline{R}x)^3 k,$$

where \underline{R} is an active rotation by the angle θ about the direction \hat{n} . Thus rotations become part of the quaternion arithmetic. ▲

Determining subspaces

All this symbol manipulation seems like a big waste of time, right? Let’s pull it together with the basic problem of how to represent the subspaces of $V = \mathbb{R}^n$, namely the p -planes through the origin. The same discussion will apply to any vector space once we establish a basis.

A p -plane is clearly determined by a basis for the subspace consisting of n linearly independent vectors which span the subspace: $\{u_{(1)}, \dots, u_{(p)}\}$. Let $\underline{U} = \langle \underline{u}_{(1)} | \dots | \underline{u}_{(p)} \rangle$ be the $p \times n$ matrix of component column vectors. This matrix has rank p since its columns are linearly independent so it row reduces to the first p columns of the identity matrix. Any vector lying in this p -plane can be expressed uniquely as a linear combination of the basis vectors: $x = c^A u_{(A)}$, $A = 1, \dots, p$.

The wedge product of the basis vectors $u_{(1)} \wedge \dots \wedge u_{(p)}$ is a p -vector. Then the $(p+1)$ -vector obtained by the wedge product

$$\begin{aligned} u_{(1)} \wedge \dots \wedge u_{(p)} \wedge x &= u_{(1)} \wedge \dots \wedge u_{(p)} \wedge (c^A u_{(A)}) && \text{(expand in basis)} \\ &= c^A (u_{(1)} \wedge \dots \wedge u_{(p)} \wedge u_{(A)}) && \text{(linearity)} \\ &= 0 && \text{(repeated factors in each term)} \end{aligned}$$

is zero since it consists of a sum of p terms, each of which contains two repeated factors in the wedge product, which separately vanish. This condition exactly captures all the vectors which lie in the subspace as the solution of a set of linear conditions when expressed in terms

of components, namely, a p -plane is the result of imposing $n - p$ independent linear conditions on n variables. Since the right hand side of this relation will remain zero under the subsequent operations we do, the sign and factorial factors are irrelevant and can be ignored, which will be indicated by using \sim to indicate proportionality.

In components

$$[u_{(1)} \wedge \cdots \wedge u_{(p)} \wedge x]^{i_1 \cdots i_p j} = u^{i_1}_{(1)} \cdots u^{i_p}_{(p)} x^j = 0.$$

Taking the natural dual of this $(p + 1)$ -vector leads to an $(n - p + 1)$ -form

$$[^{(*)}(u_{(1)} \wedge \cdots \wedge u_{(p)} \wedge x)]^{j_1 \cdots j_{p-1}}_{i_1 \cdots i_p j} \sim \epsilon_{j_1 \cdots j_{n-p+1} i_1 \cdots i_p j} u^{i_1}_{(1)} \cdots u^{i_p}_{(p)} x^j = 0,$$

which is the natural dual of the p -vector alone evaluated on x in its last argument

$$^{(*)}(u_{(1)} \wedge \cdots \wedge u_{(p)} \wedge x) = ^{(*)}(u_{(1)} \wedge \cdots \wedge u_{(p)}) \lrcorner x$$

where the symbol $S \lrcorner x$ will indicate evaluating a multiform S on x in its last argument, or in components

$$[^{(*)}(u_{(1)} \wedge \cdots \wedge u_{(p)})]_{j_1 \cdots j_{n-p+1} j} x^j = 0.$$

The solution of these linear conditions on the n -variables x^j yields the p -plane, i.e., the natural dual $(n - p)$ -form determines the p -plane implicitly, and one can think of $x = c^A u_{(A)}$ as a parametrized solution of this $(n - p) \times n$ system of linear equations.

The metric dual raises the indices on the $(n - p)$ -form $^{(*)}(u_{(1)} \wedge \cdots \wedge u_{(p)})$ to make an $(n - p)$ -vector $[^{(*)}(u_{(1)} \wedge \cdots \wedge u_{(p)})]^\sharp$ (and corrects it with the metric determinant factor), which is such that its inner product with the vector x is zero (provided we agree that $S^{i_1 \cdots i_{p-1} i_p} G_{i_p j} x^j$ represents such an inner product of a p -vector and a vector). If one takes a basis $\{u_{(A)}\}$, $A = p + 1, \dots, n$ of the orthogonal $(n - p)$ -plane, this translates into stating that x is individually orthogonal to each of these complementary basis vectors, changing the nature of the conditions from linear conditions to geometrical conditions. For a hyperplane where $p = n - 1$, the natural dual is a 1-form, and lowering its index produces a normal vector field. The p -measure of the p -parallelepiped formed by the original p vectors is exactly the magnitude of the corresponding dual p -vector, which in turn equals the length of the corresponding dual p -vector.

Take the familiar case of an ordinary plane in \mathbb{R}^3 using the standard basis, where

$$\begin{aligned} u_{(1)} \wedge u_{(2)} &= u^{[i}_{(1)} \wedge u^{j]}_{(2)} e_i \wedge e_j \\ &= \begin{vmatrix} u^2_{(1)} & u^2_{(2)} \\ u^3_{(1)} & u^3_{(2)} \end{vmatrix} e_2 \wedge e_3 + \begin{vmatrix} u^3_{(1)} & u^3_{(2)} \\ u^1_{(1)} & u^1_{(2)} \end{vmatrix} e_3 \wedge e_1 + \begin{vmatrix} u^1_{(1)} & u^1_{(2)} \\ u^2_{(1)} & u^2_{(2)} \end{vmatrix} e_1 \wedge e_2, \\ ^{(*)}(u_{(1)} \wedge u_{(2)}) &= \begin{vmatrix} u^2_{(1)} & u^2_{(2)} \\ u^3_{(1)} & u^3_{(2)} \end{vmatrix} \omega^1 + \begin{vmatrix} u^3_{(1)} & u^3_{(2)} \\ u^1_{(1)} & u^1_{(2)} \end{vmatrix} \omega^2 + \begin{vmatrix} u^1_{(1)} & u^1_{(2)} \\ u^2_{(1)} & u^2_{(2)} \end{vmatrix} \omega^3, \\ *(u_{(1)} \wedge u_{(2)}) &= \begin{vmatrix} u^2_{(1)} & u^2_{(2)} \\ u^3_{(1)} & u^3_{(2)} \end{vmatrix} e_1 + \begin{vmatrix} u^3_{(1)} & u^3_{(2)} \\ u^1_{(1)} & u^1_{(2)} \end{vmatrix} e_2 + \begin{vmatrix} u^1_{(1)} & u^1_{(2)} \\ u^2_{(1)} & u^2_{(2)} \end{vmatrix} e_3 \equiv u_{(1)} \times u_{(2)}. \end{aligned}$$

The components of the natural dual

$$n_i = [^{(*)}(u_{(1)} \wedge u_{(2)})]_i = \epsilon_{ijk} u^j_{(1)} u^k_{(2)}$$

are the coefficients of the variables in the single linear equation determining the plane: $n_i x^i = 0$. The dual is instead a normal vector $n^i = [*(u_{(1)} \wedge u_{(2)})]^i$ which is orthogonal to all vectors x in the 2-plane: $\delta_{ij} n^i x^j = 0$. This normal vector is just the ordinary cross product of the two vectors determining the plane, and its length is exactly the area of the parallelogram that they form. The mathematics of multivectors and the metric dual generalize this familiar fact from ordinary vector geometry to subspaces of all dimensions in all of the \mathbb{R}^n spaces. Instead of a normal 1-vector to a hyperplane, one has a normal $(n - p)$ -vector to a p -plane carrying p -measure information.

In fact this continues to work for any nondegenerate inner product on \mathbb{R}^3 since the inner product of a vector X with a normal vector $N = *(X \wedge Y)$ is proportional to the natural evaluation of the corresponding natural dual 1-form on the vector:

$$N \bullet Z = (*(X \wedge Y))(Z) = |\det(\underline{G})|^{1/2} \epsilon_{ijk} X^i Y^j Z^k.$$

In particular we can use this for 3-dimensional Minkowski spacetime to determine the local time direction for a spatial 2-plane. See Appendix D.

Exercise 4.3.21.

2-planes in \mathbb{R}^4 and wedge products

In \mathbb{R}^4 with the usual Euclidean inner product for which the standard basis vectors $\{e_i\}$ are orthonormal, consider the two vectors

$$X_{(1)} = \langle 1, 3, 5, 7 \rangle, \quad X_{(2)} = \langle 2, 4, 6, 8 \rangle,$$

and let

$$\underline{X} = \langle \underline{X}_{(1)} | \underline{X}_{(2)} \rangle = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{pmatrix}$$

be the 2×4 matrix for which they are the columns. They determine a 2-plane through the origin and a parallelogram in that plane with area A .

a) Evaluate the 2-vector $P = X_{(1)} \wedge X_{(2)} = X^i_{(1)} \wedge X^j_{(2)} e_i \wedge e_j$ whose 6 independent components can be labeled by the ordered index pairs $(m, n) = (2, 3), (1, 3), (1, 2), (1, 4), (2, 4), (3, 4)$, and are just the determinants of the 2×2 submatrices of \underline{X} formed by the ordered rows (m, n) . The length squared of the 2-vector P is $|X_{(1)} \wedge X_{(2)}|^2 = [X_{(1)} \wedge X_{(2)}]_{ij} [X_{(1)} \wedge X_{(2)}]^{ij}$, which is just the sum of the squares of these 6 ordered components. Evaluate this length.

b) Evaluate the dual 2-vector $Q = *(X_{(1)} \wedge X_{(2)})$ and their wedge product $P \wedge Q$, thus confirming the relation $P \wedge *P = |P|^2 e_1 \wedge e_2 \wedge e_3 \wedge e_4$.

c) What are the antisymmetric matrices of components \underline{P} of $X_{(1)} \wedge X_{(2)}$ and \underline{Q} of its dual Q ? Evaluate their product $\underline{P}\underline{Q}$. (It should be zero since they are constructed from orthogonal vectors!) What is the rank of these two matrices, i.e., how many nonzero rows do they have when you row reduce them to reduced row echelon form? The rank is the number of independent linear equations they impose on the variable x when you solve the equations $\underline{A}x = 0$ for each of these matrices as the coefficient matrix \underline{A} . The latter solution space for \underline{Q} is exactly the

original 2-plane and the solution space for \underline{P} is its orthogonal 2-plane, explaining these results. Use a computer algebra system to solve these two homogeneous systems of linear equations (the command `LinearSolve(A, <0, 0, 0, 0>)` in the `LinearAlgebra` package of Maple) and read off the coefficients of the arbitrary parameters in the solution as an ordered basis of each subspace

$$\underline{A}\underline{x} = 0 \rightarrow \underline{x} = t^A E_{(A)} \rightarrow \underline{E} = \langle E_{(1)} | E_{(2)} \rangle.$$

For the second such linear system $\underline{Q}\underline{x} = 0$, start with the basis matrix \underline{E} so obtained, and use a computer algebra system (the command `LinearSolve(E, X)` in the `LinearAlgebra` package of Maple) to find the change of basis matrix \underline{B} which satisfies $\underline{E}\underline{B} = \underline{X}$ and which expresses the old basis of the original 2-plane as linear combinations of these new basis vectors. Use this matrix \underline{B} to express the vector

$$Z = X_{(1)} + X_{(2)} = \underline{X} \langle 1, 1 \rangle = \underline{E}\underline{B} \langle 1, 1 \rangle = \underline{E} \langle y^1, y^2 \rangle = y^1 E_{(1)} + y^2 E_{(2)}$$

in terms of this new basis.

d) Use the Gram-Schmidt procedure to replace the second vector in the set by its orthogonal projection with respect to the first vector

$$(X_{(1)}, X_{(2)}) \rightarrow (Y_{(1)}, Y_{(2)}) = (X_{(1)}, Y_{(2)}), \quad Y_{(2)} = X_{(2)} - \left(\frac{X_{(2)} \cdot X_{(1)}}{X_{(1)} \cdot X_{(1)}} \right) X_{(1)}.$$

Since these two new vectors are orthogonal, the product of their lengths is the area of the rectangle they form, but this must be the same as the area of the parallelogram formed by the original pair since the Gram-Schmidt operation adds a multiple of first vector to the second, leaving the orthogonal component of the second unchanged, whose magnitude is the “height” of the parallelogram whose base has length equal to the magnitude of the first vector. See figure 2.3. Evaluate this area.

d) Now compare the length $|X_{(1)} \wedge X_{(2)}|$ calculated in the first two parts of this problem. They should be equal. The length of a decomposable p -vector formed from the wedge product of p vectors equals the measure of the p -parallelepiped formed by those vectors.

e) A 2-plane in \mathbb{R}^4 has two independent normal vectors which in turn span the orthogonal 2-plane. One can find a pair of such normals using the Gram-Schmidt procedure applied to a set of 4 vectors which complete the $X_{(1)}, X_{(2)}$ to a basis of the whole space. Adding the standard basis vectors e_3, e_4 to this set yields a square matrix with nonzero determinant, so we can apply Gram-Schmidt to the ordered set $\{X_{(1)}, X_{(2)}, e_3, e_4\}$ to obtain a new orthogonal basis $Y_{(1)}, Y_{(2)}, Y_{(3)}, Y_{(4)}$ (the first two vectors are $X_{(1)}$ and the vector $Y_{(2)}$ you already calculated in part d) above!). Do this with a computer algebra system finding the appropriate GramSchmidt command from the Help. Then for ease of computation, rescale the last two vectors of the set to obtain integer component vectors (multiply by the least common denominator of each set of components), and evaluate their wedge product M . The 2-vector M should determine the same orthogonal 2-plane as $Q = *(X_{(1)} \wedge X_{(2)})$ in the sense that these two are proportional. Are they?

f) Normalize the orthogonal basis $\{Y_{(i)}\}$ adapted to this orthogonal decomposition of \mathbb{R}^4 to obtain an orthonormal basis $\{Y_{(\hat{i})}\}$. Then $Y_{(\hat{1})} \wedge Y_{(\hat{2})}$ is a unit 2-vector determining the plane and $Y_{(\hat{3})} \wedge Y_{(\hat{4})}$ a unit normal 2-vector. Therefore up to sign, they are duals of each other.



4.4 Wedge and star duality on \mathbb{R}^n in practice

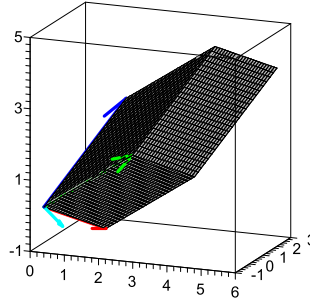


Figure 4.4: The parallelopiped formed by 3 vectors u, v, w plotted with the unit normal $\widehat{u \times v}$.

Now that we have had a chance to play with these concepts a bit, we should conclude by making them more intuitive and interpret them a bit more clearly. Let's only worry about the case of \mathbb{R}^n with its Euclidean dot product in its standard orthonormal coordinates $x^i = \omega^i$ associated with the standard basis e_i , where index raising and lowering with the Kronecker delta $G_{ij} = \delta_{ij}$ removes the need to distinguish upper and lower indices, and we can temporarily allow a repeated pair of indices at the same level to be summed over, letting $u_i = u^i$, so that $u^i u_i = u^i u^i$, etc. Also the components of the unit volume n -form $\eta_{i_1 \dots i_n} = \epsilon_{i_1 \dots i_n}$ coincide with the Levi-Civita alternating symbol, where again we can ignore index positioning, so the duality operation is

$$*(e_{i_1} \wedge \dots \wedge e_{i_p}) = \epsilon_{i_1 \dots i_p i_{p+1} \dots i_n} e_{i_{p+1}} \wedge \dots \wedge e_{i_n}.$$

In fact the unit volume n -form $\eta = \det$ is the determinant tensor itself.

To be concrete, consider an explicit example of 2 or 3 vectors in \mathbb{R}^3

$$\begin{aligned} u &= \langle 1, 2, -1 \rangle, \quad v = \langle 2, 1, 3 \rangle, \quad w = \langle 3, -2, 1 \rangle, \\ u \cdot v &= 1, \quad u \times v = \langle 7, -5, -3 \rangle, \quad (u \times v) \cdot w = \det \langle \underline{u} | \underline{v} | \underline{w} \rangle = 20. \end{aligned}$$

The determinant of the matrix with these three vectors as its columns in this order is the signed volume of the parallelopiped they form, in this case positive since the third vector is on the same side of the plane of the first two as their ordered cross product.

The 3-vector $u \wedge v \wedge w = 20e_1 \wedge e_2 \wedge e_3$ is just 20 times the unit volume 3-vector $\eta^\sharp = e_1 \wedge e_2 \wedge e_3$, and the single independent component is just the determinant

$$20 = (\omega^1 \wedge \omega^2 \wedge \omega^3)(u, v, w) = *(u \wedge v \wedge w).$$

This is constructed from the antisymmetrization of the tensor product of the three vectors

$$u \otimes v \otimes w = u^i e_i \otimes v^j e_j \otimes w^k e_k = u^i v^j w^k e_i \otimes e_j \otimes e_k.$$

The 27 basis $\binom{0}{3}$ -tensors $\omega^i \otimes \omega^j \otimes \omega^k$ pick out the products of the components of the 3 vectors, which are the components of the tensor product $u \otimes v \otimes w$

$$(\omega^i \otimes \omega^j \otimes \omega^k)(u, v, w) = u^i v^j w^k$$

The wedge product is instead just the sum over all possible permutations of the indices with a sign factor

$$\omega^i \wedge \omega^j \wedge \omega^k = \delta_{mnp}^{ijk} \omega^m \otimes \omega^n \otimes \omega^p,$$

introducing the $\binom{3}{3}$ -indexed generalized Kronecker delta which is just the collection of components of the wedge product of the basis covectors in this case. However, there is only one independent 3-form $\omega^1 \wedge \omega^2 \wedge \omega^3$, so

$$(\omega^1 \wedge \omega^2 \wedge \omega^3)(u, v, w) = \delta_{ijk}^{123} (\omega^i \otimes \omega^j \otimes \omega^k)(u, v, w) = \delta_{ijk}^{123} u^i v^j w^k = \det \langle \underline{u} | \underline{v} | \underline{w} \rangle,$$

which is a sum of 6 terms giving the determinant. This single number is the dual of the 3-vector $u \wedge v \wedge w$: $*(u \wedge v \wedge w) = 20$.

If we delete any one of the columns of the 3×3 matrix $\langle \underline{u} | \underline{v} | \underline{w} \rangle$, say \underline{w} , we can get at most 2×2 submatrices by deleting one of the rows of the resulting 3×2 matrix of the first two vectors

$$\underline{A} = \langle \underline{u} | \underline{v} \rangle = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ -1 & 3 \end{pmatrix}.$$

In an obvious notation

$$\det(\text{DelRow}(\underline{A}, 1)) = \begin{vmatrix} 2 & 1 \\ -1 & 3 \end{vmatrix} = 7,$$

$$\det(\text{DelRow}(\underline{A}, 2)) = \begin{vmatrix} 1 & 2 \\ -1 & 3 \end{vmatrix} = 5,$$

$$\det(\text{DelRow}(\underline{A}, 3)) = \begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = -3.$$

Ignoring the first components projects the two vectors orthogonally onto the x^2 - x^3 plane, where the determinant gives the signed area of the parallelogram formed by the two projected vectors in the plane, and the sign indicates whether one moves counterclockwise (+) or clockwise (−) from the projection of u to the projection of v . Remarkably the sum of the squares of the areas of these 3 projections of the original unprojected parallelogram is the square of the area of that original parallelogram.

The 3 independent basis 2-forms (with components which are the $\binom{2}{2}$ -Kronecker deltas) evaluate exactly to these three subdeterminants. The tensor product $\omega^2 \otimes \omega^3$ simply picks out the product components: $(\omega^2 \otimes \omega^3)(u, v) = u^2 v^3$. The wedge product $\omega^2 \wedge \omega^3 = \omega^2 \otimes \omega^3 - \omega^3 \otimes \omega^2$ instead evaluates the first of these subdeterminants, deleting the first components of both vectors

$$(\omega^2 \wedge \omega^3)(u, v) = (\omega^2 \otimes \omega^3 - \omega^3 \otimes \omega^2)(u, v) = u^2 v^3 - u^3 v^2 = \begin{vmatrix} u^2 & v^2 \\ u^3 & v^3 \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ -1 & 3 \end{vmatrix}.$$

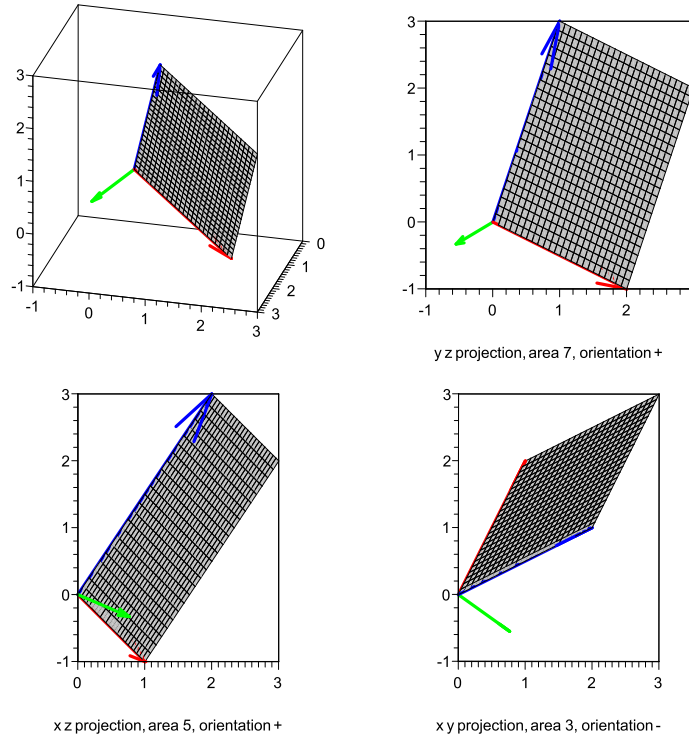


Figure 4.5: The parallelogram formed by 2 vectors u, v plotted with the unit normal $\widehat{u \times v}$ projects onto parallelograms in each of the three coordinate planes. The projected vectors rotate from u to v in the counterclockwise direction when the orientation is positive (normal out of the page) and in the clockwise direction (normal into the page) when the orientation is negative.

In general the 3 independent such forms, either the ordered set $\omega^2 \wedge \omega^3, \omega^1 \wedge \omega^3, \omega^1 \wedge \omega^2$, or the cyclic set $\omega^2 \wedge \omega^3, \omega^3 \wedge \omega^1, \omega^1 \wedge \omega^2$, evaluate to these subdeterminants or an alternating sign times them

$$\begin{aligned} \omega^i \wedge \omega^j &= \delta_{mn}^{ij} \omega^m \otimes \omega^n, \\ (\omega^i \wedge \omega^j)(u, v) &= \delta_{mn}^{ij} (\omega^m \otimes \omega^n)(u, v) = \delta_{mn}^{ij} u^m v^n = (\delta_m^i \delta_n^j - \delta_n^i \delta_m^j) u^m v^n = u^u v^j - u^n v^m. \end{aligned}$$

These are exactly the coefficients of the 2-vector $u \wedge v$ when expressed in terms of the basis 2-forms

$$u \wedge v = 7e_2 \wedge e_3 + 5e_1 \wedge e_3 - 3e_1 \wedge e_2 = 7e_2 \wedge e_3 - 5e_3 \wedge e_1 - 3e_1 \wedge e_2.$$

The dual swaps a pair of indices (a, b) from the cyclic triplet (a, b, c) (namely a positive permutation of $(1, 2, 3)$) for the third index, thus economically dealing only with the three independent components of the antisymmetric tensor $u \wedge v$ which has 3 zero components and 3 pairs of

components differing only by sign in each pair: $*e_a \wedge e_b = \epsilon_{abc}e_c$ or

$$\begin{aligned} *(u \wedge v) &= *(7e_2 \wedge e_3 - 5e_3 \wedge e_1 - 3e_1 \wedge e_2) = 7*(e_2 \wedge e_3) - 5*(e_3 \wedge e_1) - 3*(e_1 \wedge e_2) \\ &= 7e_1 - 5e_2 - 3e_3 = u \times v, \end{aligned}$$

which is the familiar cross product vector. It is well known that its length represents the area of the parallelogram formed by u and v

$$Area(u, v) = |u \times v| = |u \wedge v|$$

whose square is the sum of the squares of the three 2×2 subdeterminants representing the signed projected areas, which is the square of the length of the 2-vector as well.

If we delete two of the three vectors in the original 3×3 matrix, say v and w , we are left with a 3×1 matrix of one column vector u . The largest square submatrices contained in a single column are the three 1×1 submatrices whose determinant is just the single matrix entry. These are picked out by the three basis 1-forms ω^i . These represent the signed lengths of the orthogonal projections onto the three respective axes. Thus the wedge algebra is a way of capturing the volume, area and length of the original parallelepiped and of the orthogonal projections of its faces onto the coordinate planes, and the orthogonal projections of its edges onto the coordinate axes, all evaluated from the set of all possible subdeterminants of the original square matrix of the three vectors of our set. Each face is represented by a 2-vector whose dual is a normal vector to the face, and whose length is the area of the face.

The duality operation is also very natural. The lines and planes through the origin are the subspaces of \mathbb{R}^3 . Think of the orthonormal basis e_1, e_2, e_3 of \mathbb{R}^3 as a simple set of vectors to discuss this concept. One way to specify the x -axis (all multiples of e_1) is to say that it is the 1-dimensional subspace consisting of all vectors orthogonal to the y - z coordinate plane, which is its complementary orthogonal 2-dimensional subspace, in turn described by the conditions: $y = 0, z = 0$. e_2 and e_3 are a basis of that plane, but to specify a plane, no particular choice of basis matters, so we can simply take their wedge product to represent that subspace, since any other basis will lead to the same 2-vector $e_2 \wedge e_3$ modulo a scalar, a scalar which is just the determinant of the 2×2 matrix of components of the new basis of that plane. The dual of e_1 is the 2-vector $e_2 \wedge e_3$ (itself a unit tensor since it represents the area of a unit square whose edges are the two orthogonal unit vectors) representing the orthogonal plane such that the wedge product of the two gives the unit volume 3-vector on \mathbb{R}^3 : $\eta^\sharp = e_1 \wedge e_2 \wedge e_3$. In other words it is the missing factor needed to complete e_1 to $e_1 \wedge e_2 \wedge e_3$.

Conversely, how can we specify a plane for which we have a basis? Usually we specify a plane by giving a normal vector to the plane. For the y - z plane which has e_2, e_3 as a basis and therefore $e_2 \wedge e_3$ as a representative 2-vector, the dual is the missing factor e_1 whose wedge with it produces the unit 3-vector. There is only one orthogonal direction, so a unit normal is all we need to specify the orthogonal direction and the dual of the unit 2-vector $e_2 \wedge e_3$ produces one of the two possible unit normals e_1 , chosen so that the wedge on the right produces the unit volume 3-vector $\eta^\sharp = e_1 \wedge e_2 \wedge e_3$ and not the one differing by a negative sign. For lines and planes not aligned with the coordinate axes and planes, the duality operation still does the same thing geometrically. Modulo a factor, the dual of any p -vector is a $(3 - p)$ vector representing

the orthogonal directions, and the factor itself is chosen so that their wedge product produces the square of the length of the p -vector as the multiple of η^\sharp , a factor which represents the square of the length of the vector ($p = 1$) or the area of the parallelogram formed by the two vector factors ($p = 2$).

In \mathbb{R}^n , a p -plane through the origin can be specified by giving a basis of $(n - p)$ normal vectors for the orthogonal $(n - p)$ -plane, or simply the wedge product of those normals, and requiring any vector that belong to the p -plane be orthogonal to that set of vectors, or to their wedged $(n - p)$ -vector. If we start with an orthonormal basis of the p -plane, then their wedge is a unit p -vector specifying that p -plane, and its dual is a unit $(n - p)$ -vector which specifies the orthogonal $(n - p)$ -plane, and can be represented as the wedge product of an orthonormal basis of the $(n - p)$ -plane.

Even in the simplest case of \mathbb{R}^2 , we are all pretty familiar with the way in which we get a line through the origin at 90 degrees to an existing line through the origin: they have negative reciprocal slopes. In fact as illustrated in Figure 4.6, if we take the original vector specifying such a line through the origin and swap the components with an additional minus sign introduced in one of them, we get a negative reciprocal slope. This is just the duality operation! The dual of a vector is a $(2 - 1)$ -vector, just another vector, which is orthogonal to the first vector and has the same length. The sign is chosen so that the vector and its dual form an oriented basis of the plane, namely one moves from the first to the second counterclockwise through an angle less than π .

The spaces of p -vectors and $(n - p)$ -vectors in \mathbb{R}^n both have dimension

$$\binom{n}{p} = \frac{n!}{p!(n-p)!} = \binom{n}{n-p},$$

and the star duality operation allows us to represent the latter in terms of the former. The dimension is 1 for $p = 0$ (scalars) and $p = n$, with an n -vector representable as the dual of a scalar. This dimension is n for 1-vectors and $(n - 1)$ -vectors, and the latter are representable as the dual of a vector. The dimension is $n(n - 1)/2$ for 2-vectors (whose components are antisymmetric matrices) and $(n - 2)$ -vectors, and the latter are representable as the dual of a 2-vector. This is all we need for $n < 5$. In the plane $n = 2$, where $n - 1 = 1$, we only need vectors, and duality takes vectors to vectors. For $n = 3$ we only need vectors, since the dual of a 2-vector is a vector (explaining the cross product). For $n = 4$, we only need vectors and 2-vectors, since the dual of a 3-vector is a vector, while 2-vectors are essentially just antisymmetric matrices, still familiar. Thus the whole business of wedging and duality is not really such a big deal in practice.

Suppose we consider an $(n - p) \times n$ matrix

$$\underline{A} = \langle \underline{u}_{(1)} | \dots | \underline{u}_{(n-p)} \rangle = \langle \underline{a}^{(1)T}, \dots, \underline{a}^{(n-p)T} \rangle$$

of rank $r = n - p$ consisting of n columns $\underline{u}_{(i)}$ and $n - p$ linearly independent rows $\underline{a}^{(B)T}$, $B = 1 \dots n - p$. Then these rows are a basis of the $(n - p)$ -dimensional row space of the matrix consisting of the span of its set of rows, which is the subspace of \mathbb{R}^n orthogonal to the p -dimensional null space of the matrix with respect to the dot product. The null space,

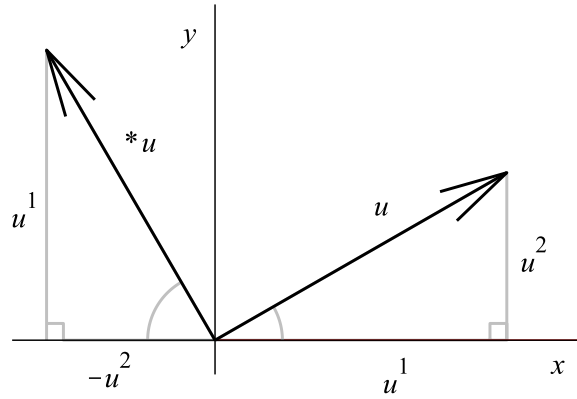


Figure 4.6: The duality operation in the Euclidean plane is simply a rotation by 90 degrees in the counterclockwise direction: $*\langle u^1, u^2 \rangle = *(u^1 e_1 + u^2 e_2) = u^1 *e_1 + u^2 *e_2 = u^1 e_2 - u^2 e_1 = \langle -u^2, u^1 \rangle$, yielding a normal vector to the linear subspace spanned by the original vector. The natural dual $(*)\langle u^1, u^2 \rangle = (*)(u^1 e_1 + u^2 e_2) = u^1 (*)e_1 + u^2 (*)e_2 = u^1 \omega^2 - u^2 \omega^1$ on the other hand, when used to find those vectors on which it evaluates to zero, describes the original subspace by its linear equation: $0 = -u^2 \omega^1 + u^1 \omega^2 = -u^2 x + u^1 y$, or $y = (u^2/u^1)x$ if $u^1 \neq 0$.

which is a p -plane in \mathbb{R}^n through the origin consisting of all those vectors sent to 0 by matrix multiplication by \underline{A} , has a basis of p vectors $v_{(1)}, \dots, v_{(p)}$ provided by the reduced row echelon algorithm. The wedge product p -vector $v_{(1)} \wedge \dots \wedge v_{(p)}$ completely determines this p -plane. On the other hand the row space is an $(n-p)$ -plane through the origin of \mathbb{R}^n which is determined by the wedge product $(n-p)$ -vector $a^{(1)} \wedge \dots \wedge a^{(n-p)}$. The dual of each is proportional to the other. This same geometry extends to any nondegenerate inner product on \mathbb{R}^n .

Exercise 4.4.1.

transforming wedge products and star duals in the plane

This exercise gives us some hands on contact with actual numbers instead of just juggling formulas. Given the standard basis of \mathbb{R}^2 and a given vector X , one can evaluate some of the various maps: index lowering maps the vector onto a covector X^\flat . The natural dual maps the vector onto another covector $(*)X$. The metric dual maps the vector onto another vector $*X$. Given a new basis of the plane determined by the change of basis matrix \underline{A} , one can calculate the new components of the vector and of the dot product tensor G and use the latter to lower the vector's index to obtain the new components of the corresponding covector. However, one can also transform the components of the index lowered covector X^\flat to get the same result and they should agree. In other words, we can lower the index on the vector either before or after

transforming its components from the old basis to the new basis. For each of the other maps one can repeat the exercise, performing the map either before or after transforming components.

For example, as illustrated in the following diagram we can either transform the components and then lower the index (right, then down) or lower the index and then transform the components (down, then right) and we should get the same result in either order, leading to a “commutative diagram”

$$\begin{array}{ccc} X^i & \xrightarrow{\underline{A}} & X^{i'} \\ \underline{I} \downarrow \flat & & \flat \downarrow \underline{G}' \\ X_i & \xrightarrow{\underline{A}^{-1}} & X_{i'}. \end{array}$$

A similar diagram describes the component independence of the natural dual $(*)$, which is also a linear map from \mathbb{R}^2 to $(\mathbb{R}^2)^*$. The metric dual map $*$ on the other hand is a linear map from \mathbb{R}^2 to itself

$$\begin{array}{ccc} X^i & \xrightarrow{\underline{A}} & X^{i'} \\ * \downarrow & & * \downarrow \\ [*X]^i & \xrightarrow{\underline{A}} & [*X]^{i'}. \end{array}$$

Indeed the same diagram would describe the component independence of any linear map L from the vector space into itself. The present problem will calculate these two different ways of getting to the opposite corner of the diagram for concrete values of the components and the various maps.

In doing the change of basis, one has two choices: work with the set of components \underline{X} or \underline{I} and use component/matrix methods to transform them, like $\underline{X}' = \underline{A} \underline{X}$ or $\underline{G}' = (\underline{A}^{-1})^T \underline{I} \underline{A}^{-1}$, or work with the index-free object $X = X^i e_i = X^{i'} e_{i'}$ or $G = \delta_{ij} \omega^i \otimes \omega^j = G_{i'j'} \omega^{i'} \otimes \omega^{j'}$ and simply substitute the old basis vectors and covectors expressed in terms of the new ones and expand to get the new components. Similarly the unit area 2-form $\eta = \omega^1 \wedge \omega^2$ can be so re-expressed to evaluate its new single independent component and compared with the component definition in a general basis.

Consider the new left handed basis on the plane: $e_{1'} = \langle 1, 2 \rangle, e_{2'} = \langle 3, 1 \rangle$. It would not hurt to make a diagram showing the new coordinate grid and the various vectors, lines and 1-forms discussed below, and start by writing out the four transformations

$$\begin{aligned} e_{i'} &= e_j A^{-1j}{}_i, & e_i &= e_{j'} A^j{}_{i'}, \\ \omega^{i'} &= A^i{}_{j'} \omega^{j'}, & \omega^i &= A^{-1i}{}_{j'} \omega^{j'}, \end{aligned}$$

and the matrices \underline{A} and \underline{A}^{-1} .

a) Evaluate the matrix of inner products $\underline{G}' = (e_{i'} \cdot e_{j'})$ and its determinant $\det \underline{G}'$. Express the inner product tensor $G = \delta_{ij} \omega^i \otimes \omega^j = G_{i'j'} \omega^{i'} \otimes \omega^{j'}$ in terms of the new dual covectors using your result for the new components, i.e., just replace the four numbers $G_{i'j'}$ in this expression for G by the values you found.

b) Evaluate the unit volume form $\eta = \omega^1 \wedge \omega^2 = \eta_{1'2'} \omega^{1'} \wedge \omega^{2'}$ by wedging together ω^1 and ω^2 expressed in terms of the new dual basis, and read off the component $\eta_{1'2'}$ as the coefficient

of $\omega^{1'} \wedge \omega^{2'}$. Then compare with the formula $\eta_{1'2'} = \pm(\det \underline{G}')^{1/2} \epsilon_{12}$, where the plus or minus is appropriate if the transformation matrix from the oriented standard basis has positive or negative determinant.

c) Express the vector $X = -2e_1 + e_2$ and its corresponding 1-form $X^b = -2\omega^1 + \omega^2$ in terms of the new basis, and show that the new components are related to each other by index lowering with the matrix \underline{G}' .

d) Evaluate the natural dual of $X = -2e_1 + e_2$ to obtain the equation of the subspace spanned by this vector as a linear homogeneous equation on its coordinates x^1, x^2 , namely $^{(*)}X(\langle x^1, x^2 \rangle) = 0$. Evaluate the metric dual of X to obtain a normal vector to this line and show that its dot product with X is indeed zero.

e) Repeat d) but working in terms of the components X with respect to the new basis.

f) Consider the tensor $L = -2e_1 \otimes \omega^1 + e_1 \otimes \omega^2 - e_2 \otimes \omega^1 + e_2 \otimes \omega^2$ representing a linear transformation of the plane, and the vector obtained from its evaluation in its second argument on $X = -2e_1 + e_2$: $Y = L(\ , X)$. Express this tensor in terms of the new basis and use its new matrix of components to multiply the component matrix of X expressed in terms of the new basis and get the new coordinates of the vector Y . Check that these are the same as you would obtain by directly transforming the components of Y .

■

4.5 Matrix generators of the generalized orthogonal matrix groups

We have seen that the Lie algebra of the generalized orthogonal matrix group which leaves invariant the canonical matrix of an inner product G in an orthonormal basis of \mathbb{R}^n under a change of basis consists of the set of antisymmetric $n \times n$ matrices with their first index raised with the metric. The commutation relations of these Lie algebra matrices completely determine the local structure of the matrix groups themselves. Now that we are armed with the necessary delta machinery, we can easily write down an explicit formula for a standard basis and their commutation relations.

In Section 1.6 we discussed the orthogonal matrix group $O(P, M)$ and its unit determinant subgroups, the special orthogonal matrix group $SO(P, M)$, with $P + M = n$. These groups which map among themselves the orthonormal bases of an inner product with a diagonal orthonormal component matrix consisting of P positive entries 1 and M negative entries -1 . We can assume $P \geq M$ without loss of generality since the overall sign has no influence on the corresponding matrix group, and we can order an orthonormal basis so that the negative signs are all first. The component matrix of the inner product then takes the form

$$\underline{G} = \text{diag}(\underbrace{-1, \dots, -1}_P, \underbrace{1, \dots, 1}_M).$$

The corresponding generalized orthogonal Lie algebra matrices $\underline{B} \in so(P, M)$ are off-diagonal and satisfy

$$\text{sgn}(G_{ii})K^i_j = -\text{sgn}(G_{jj})K^j_i.$$

The dimensions of these Lie algebras are all $n(n-1)/2$, the sum of the first $n-1$ natural numbers, which in turn corresponds to the number of entries above the main diagonal of an $n \times n$ matrix.

$$K_{ij} = -K_{ji} \quad \text{or more explicitly} \quad G_{ik}K^k_j = -G_{jk}K^k_i.$$

If we examine this condition in an orthonormal basis in which \underline{G} is diagonal and $G_{ii} = \pm 1$, then this forces \underline{K} to be an off-diagonal matrix since $i = j$ implies $K_{ii} = 0$ so $K^i_i = 0$ (no sum on i). If $i \neq j$ and G_{ii} and G_{jj} have the same sign, K is antisymmetric in the index pair (i, j) and the matrix $-\underline{e}^j_i + \underline{e}^i_j$ generates an ordinary “active” rotation in the x^i - x^j plane from the x^i axis towards the x^j axis, while if G_{ii} and G_{jj} have the opposite sign, \underline{K} is symmetric in the index pair (i, j) and the matrix $\underline{e}^j_i + \underline{e}^i_j$ generates a hyperbolic rotation or “active boost” in the x^i - x^j plane which squeezes these two positive coordinate axes into the first quadrant.

This is almost the whole story except that in some cases (like relativity) the whole is greater than the sum of its parts. This tale has to do with the indefinite trace inner product on the space of off-diagonal square matrices, where antisymmetric matrices have negative sign while symmetric matrices have positive sign, and certain combinations of the two have zero drlg-inner product, making them more interesting. Since rotations push points in the plane of the rotation around circles which are special cases of ellipses (negative sign), while boosts in a plane push points along hyperbolas (positive sign), the only conic remaining is the parabola,

which manifests itself in this context as what are called null rotations (remember, null is a word signaling the presence of zero in the concept). These “orbits” (the set of points related to each other under a transformation) require 3 dimensions to realize.

The $n = 2$ case Lie algebras $so(2)$ and $so(1, 1)$ are 1-dimensional corresponding to the generators of ordinary rotations or hyperbolic rotations of the plane respectively as discussed in Appendix A. For $n = 3$ there are only two inequivalent cases $so(3)$ and $so(2, 1)$, the Lie algebras of the rotation group and the Lorentz group of 3-dimensional Minkowski spacetime and the latter allows us to see an example of a null rotation. There is no need to be intimidated by the rest of this section. The only fact we need to take away is that the (pseudo-)orthogonal matrix groups which describe the freedom in choosing a (pseudo-)orthonormal basis of the vector space are generated by the $\binom{1}{1}$ tensors whose component matrices are antisymmetric when index-lowered or raised to the fully covariant or contravariant position, generated in the sense that the matrix exponential of these matrices in the matrix Lie algebra yields elements of the matrix group itself. The rest of this section are interesting details about related issues that you can read for amusement or just blast through so you can consider yourself exposed to the material.

Exercise 1.2.4 explored the commutator relations of the rotation group Lie algebra $so(3, \mathbb{R})$, where we used the duality operation to replace the antisymmetric pair of indices $\underline{L}_{ij} = -\underline{e}^j_i + \underline{e}^i_j$ associated with the plane of the rotation with a single vector index by defining

$$\underline{L}_k = \frac{1}{2} \delta_{kl} \epsilon^{lij} \underline{L}_{ij} \leftrightarrow [\underline{L}_k]^i_j = \epsilon_{ikj} = -\epsilon_{kij}.$$

Explicitly

$$\begin{aligned} \underline{\omega} &= \begin{pmatrix} 0 & -\omega^3 & \omega^2 \\ \omega^3 & 0 & -\omega^1 \\ -\omega^2 & \omega^1 & 0 \end{pmatrix} = \omega^1 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} + \omega^2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} + \omega^3 \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &\equiv \omega^1 \underline{L}_1 + \omega^2 \underline{L}_2 + \omega^3 \underline{L}_3. \end{aligned}$$

A rotation in the x^1 - x^2 plane is interpreted as about the remaining x^3 axis. We showed there that

$$[\underline{L}_2, \underline{L}_3] = \underline{L}_1, \quad [\underline{L}_3, \underline{L}_1] = \underline{L}_2, \quad [\underline{L}_1, \underline{L}_2] = \underline{L}_3,$$

or

$$[\underline{L}_{31}, \underline{L}_{12}] = \underline{L}_{23}, \quad [\underline{L}_{12}, \underline{L}_{23}] = \underline{L}_{31}, \quad [\underline{L}_{23}, \underline{L}_{31}] = \underline{L}_{12},$$

The higher dimensional orthogonal groups basically replicate these relations, namely, for any triplet of distinct natural numbers (i, j, k) one has

$$[\underline{L}_{ki}, \underline{L}_{ij}] = \underline{L}_{jk}$$

if the two planes of the rotations include a common direction, and zero otherwise.

In the other (Lorentz) case of $so(2, 1)$ with $\underline{G} = \text{diag}(-1, 1, 1)$ and coordinates (x^0, x^1, x^2) corresponding to 3-dimensional Minkowski spacetime, we have instead two boost generators

mixing the time coordinate x^0 with the two spatial coordinates $\{x^A\} = \{x^1, x^2\}$ and one rotation generator in the plane of those two spatial coordinates

$$\begin{aligned} \underline{\theta} &= \begin{pmatrix} 0 & \theta^1 & \theta^2 \\ \theta^1 & 0 & -\theta^3 \\ \theta^2 & \theta^3 & 0 \end{pmatrix} = \theta^1 \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \theta^2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} + \theta^3 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \\ &\equiv \theta^1 \underline{B}_1 + \theta^2 \underline{B}_2 + \theta^3 \underline{L}_3. \end{aligned}$$

Exercise 4.5.1.

exponentiating boost matrices

In Exercise 1.7.9 the matrix representing a rotation in space in terms of the axis of rotation and angle about it was evaluated through summing the matrix exponential of a Lie algebra matrix through an iteration formula. One can do the same for a general boost in 3-dimensional spacetime. Let $A, B, C = 1, 2$ and let n^A be the components of a unit spacelike 2-vector: $\delta_{AB} n^A n^B = 1$.

a) Show that $(n^C \underline{B}_C)^3 = n^C \underline{B}_C$, a simple sign change compared to the rotation case in Exercise 1.7.9, which allows one to collapse all powers in the exponential series $e^{\alpha n^C \underline{B}_C}$ to at most the quadratic power.

b) Following the same approach as the rotation case, show that

$$\begin{aligned} \underline{L} &= e^{\alpha n^C \underline{B}_C} = \underline{I} + \sinh \theta n^C \underline{B}_C + (\cosh \theta - 1)(n^C \underline{B}_C)^2 \\ &= \underbrace{\underline{I} - (n^C \underline{B}_C)^2}_{1\text{-identity}} + \sinh \theta n^C \underline{B}_C + \cosh \theta \underbrace{(n^C \underline{B}_C)^2}_{2\text{-identity}}. \end{aligned}$$

This corresponds to the identity transformation along the direction perpendicular to the plane of the boost, and a simple hyperbolic rotation in the plane of the boost.

c) Define the speed $v = \tanh \alpha$, the velocity components $v^A = \tanh \alpha n^A$ and the gamma factor $\gamma = \cosh \alpha$, so that $\gamma v^A = \sinh \alpha n^A$. Rewrite the previous result in terms of these new boost velocity parameters. v^A represents the components of the spatial velocity of the boost transformation. Show that

$$L^0_0 = \gamma, \quad L^A_0 = \gamma v^A, \quad L^A_B = \delta^A_B + (\gamma - 1)v^{-2}v^A v_B.$$

Adding a third spatial direction x^3 and letting $A, B = 1, 2, 3$ extends these same formulas to 4-dimensional Minkowski spacetime. Compare these formulas with what you can find on Lorentz transformations on the web. ■

Exercise 4.5.2.

null rotations in 3 dimensions

a) Evaluate the three commutators $[\underline{B}_2, \underline{L}_3]$, $[\underline{L}_3, \underline{B}_1]$ and $[\underline{B}_1, \underline{B}_2]$ and re-express them in terms of $\underline{L}_{01} = \underline{B}_1$, $\underline{L}_{02} = \underline{B}_2$ and $\underline{L}_{12} = -\underline{L}_3$.

Notice that unlike rotations, boosts do not form a subgroup since boosts in two orthogonal directions in different orders do not lead to a third boost but to a rotation, as signaled by the commutator $[B_1, B_2]$.

b) Confirm the identity:

$$\text{Tr}(\underline{\theta}^2) = -(\theta^0)^2 + (\theta^1)^2 + (\theta^2)^2.$$

We get a null result for $\theta^0 = \theta^2$, say both equal 1, while $\theta^1 = 0$:

$$\underline{N} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix}$$

This generates a simultaneous boost in the x^2 direction and a rotation in the x^2 - x^3 plane. Show that $\underline{N}^3 = 0$ so that the exponential series for this matrix truncates to a quadratic expression in the coefficient. Evaluate this explicitly

$$e^{t\underline{N}} = \underline{I} + t\underline{N} + \frac{t^2}{2}\underline{N}^2.$$

c) Let $\vec{r}(t) = e^{t\underline{N}} \langle x^0, x^1, x^2 \rangle$ be the orbit of the point $\langle x^0, x^1, x^2 \rangle$ in 3-dimensional Minkowski spacetime \mathbb{M}^3 under this 1-parameter group of null rotations. Show that $\vec{r}'(t) \times \vec{r}''(t)$ is a constant (covariant) vector, which proves that this is a plane curve since its binormal is constant (but its first index needs to be raised, i.e., changed in sign to get the actual bi-normal in the Lorentzian geometry) and in fact this bi-normal is independent of the initial point undergoing this null rotation except for the special points $x^0 = -x^1$, where this degenerates to a straight null line. Plot the curves for $\langle x^0, x^1, x^2 \rangle = \langle 1, 0, 0 \rangle$ and $\langle x^0, x^1, x^2 \rangle = \langle 0, 1, 0 \rangle$ for $t = -1..1$. These look like parabolas. Can you think of a way to prove that they are? ■

To solve this problem once and for all we can define the $so(P, M)$ matrix Lie algebra basis in an orthonormal frame (diagonal components ± 1 as above) by

$$\begin{aligned} \underline{L}_{ij} &= -\delta_{ij}^m G_{kn} \underline{e}^n_m = -(\delta_i^m \delta_j^k - \delta_j^m \delta_i^k) G_{kn} \underline{e}^n_m \\ &= -(\delta_i^m G_{jn} - \delta_j^m G_{in}) \underline{e}^n_m \\ &= -(\delta_i^m \delta_{jn} \text{sgn}(G_{jj}) - \delta_j^m \delta_{in} \text{sgn}(G_{ii})) \underline{e}^n_m. \end{aligned}$$

Exercise 4.5.3.

antisymmetric 3×3 matrices and the negative dual vector a) For \mathbb{R}^3 with the usual dot product show that the definition $\underline{L}_{ij} = -\delta_{ij}^m \delta_{kn} \underline{e}^n_m$ means that the matrix $(\underline{L}_3)^1_2 = (\underline{L}_{12})^1_2 = -1$, which extends cyclicly to the other two pairs (2,3) and (3,1). Thus defining the dual indexed matrices $\underline{L}_k = \delta_{kl} \epsilon^{lij} \underline{L}_{ij}$, we obtain the three matrices of Exercise 1.2.4 which generate active rotations in the counterclockwise direction in their planes as determined by the right hand rule.

b) If we expand any antisymmetric matrix $\underline{\omega} = (\omega^i_j) = \omega^a \underline{L}_a$ in terms of this basis of antisymmetric matrices, show that the relation between the matrix entries and the components of the coefficient vector (ω^a) is the sign-reversed dual

$$\omega^k = -\frac{1}{2} \epsilon^{kij} \delta_{im} \omega^m_j.$$

■

Exercise 4.5.4.

commutators of the Lorentz group Lie algebra

For \mathbb{M}^4 with the Lorentz dot product show that the definition $\underline{L}_{ij} = -\delta^{mk}_{ij} G_{kn} \underline{e}^n_m$ means that the matrix $(\underline{L}_{10})^1_2 = 1$, which extends cyclicly to the other two pairs $(2, 0)$ and $(3, 0)$. These matrices \underline{B}_i generate the 3 active boosts along the positive spatial axes. These together with the 3 rotation matrices are a standard basis of the 6-dimensional Lie algebra $so(3, 1)$ of the Lorentz group $SO(3, 1)$.

Evaluate their commutators $[\underline{B}_i, \underline{B}_j]$, $[\underline{L}_i, \underline{B}_j]$, $[\underline{L}_i, \underline{L}_j]$.

■

Exercise 4.5.5.

commutators of the (pseudo-)orthogonal group Lie algebras

a) For a warmup, show that for the Euclidean dot product case where $G_{ij} = \delta_{ij}$ and $\underline{L}_{ij} = \delta^{mk}_{ij} \delta_{kn} \underline{e}^n_m$ or $\underline{L}^i_j = \delta^{mk}_{pj} \delta_{kn} \delta^{pi} \underline{e}^n_m$ that

$$[\underline{L}^{ij}, \underline{L}_{kl}] = \delta^{[i}_{[k} \underline{L}^m_{n]}.$$

which is the simplest formula one can write that has the proper antisymmetries in the index pairs.

b) Show that by shifting the indices back we get

$$[\underline{L}_{ij}, \underline{L}_{kl}] = C^{mn}_{ij,kl} \underline{L}_{mn}, \quad C^{mn}_{ij,kl} = 4\delta^m_{[i} \delta_{j][k} \delta^n_{l]}.$$

c) Show that the the same formulas hold with the substitution of $\delta_{ij} \rightarrow G_{ij}$

$$[\underline{L}^{ij}, \underline{L}_{kl}] = \delta^{[i}_{[k} \underline{L}^m_{n]}.$$

which can also be written

$$[\underline{L}_{ij}, \underline{L}_{kl}] = C^{mn}_{ij,kl} \underline{L}_{mn}, \quad C^{mn}_{ij,kl} = 4\delta^m_{[i} G_{j][k} \delta^n_{l]}.$$

■

Exercise 4.5.6.
rotations in \mathbb{R}^4

Consider the rotations of \mathbb{R}^4 , with the 6 matrices defined above which generate the Lie algebra $so(4, \mathbb{R}^4)$ of the special orthogonal group $SO(4, \mathbb{R})$. Define

$$\begin{aligned}\underline{E}_i &= \frac{1}{2}(\underline{L}_{4i} - \underline{L}_{jk}), \quad (i, j, k) \text{ cyclic permutation of } (1, 2, 3) \\ \tilde{\underline{E}}_i &= \frac{1}{2}(\underline{L}_{4i} + \underline{L}_{jk}), \quad (i, j, k) \text{ cyclic permutation of } (1, 2, 3).\end{aligned}$$

These generate simultaneous rotations in two orthogonal planes, in the same and opposite senses. Use a computer algebra system to show that their commutators are the following

$$[\underline{E}_a, \underline{E}_b] = -C^a_{bc}\underline{E}_a, \quad [\tilde{\underline{E}}_a, \tilde{\underline{E}}_b] = C^a_{bc}\tilde{\underline{E}}_a, \quad [\tilde{\underline{E}}_a, \underline{E}_b] = 0,$$

where $C^a_{bc} = \epsilon_{abc}$. Thus this 6-dimensional Lie algebra decomposes into 2 commuting subalgebras, each of which has cyclicly related commutators like those of the 3-dimensional rotation group generated by $\underline{L}_{23}, \underline{L}_{31}, \underline{L}_{12}$ (apart from an overall sign which can be changed reversing the sign of all the basis vectors) which generate rotations that leave the last coordinate x^4 -axis invariant. Note that by reversing the sign of the tilde matrices, we reverse the sign of all the commutator coefficients. ■

Exercise 4.5.7.
differentials of rotation matrices

Recall Exercise 1.7.10 where the differentials

$$\underline{R}^{-1}d\underline{R} = \omega^a \underline{L}_a, \quad d\underline{R} \underline{R}^{-1} = \tilde{\omega}^a \underline{L}_a$$

were evaluated for a general rotation matrix in $SO(3, \mathbb{R})$, where $\underline{R}^{-1} = \underline{R}^T$.

a) Starting from the identity

$$\epsilon_{ijk} R^{-1i}_a R^{-1j}_b R^{-1k}_c = \det(\underline{R}) \epsilon_{abc} = \epsilon_{abc}$$

which holds for the rotation matrices in $SO(3, \mathbb{R})$, multiply this by R^c_m to get

$$R^{-1i}_a \epsilon_{ijm} R^{-1j}_b = \epsilon_{abc} R^c_m.$$

Next using the definition of the basis of the Lie algebra of the rotation group

$$(\underline{L}_a)^b_c = \delta^{bd} \epsilon_{dac} = \epsilon_{bac} = -\epsilon_{abc},$$

replace $\epsilon_{ijm} = \epsilon_{mij}$ by $-(\underline{L}_m)_{ij}$ and similarly $\epsilon_{abc} = \epsilon_{cab}$ by $-(\underline{L}_c)_{ab}$ to obtain (canceling the common minus sign)

$$R^{-1i}_a (\underline{L}_m)_{ij} R^{-1j}_b = (\underline{L}_c)_{ab} R^c_m.$$

Returning to matrix notation this is

$$\underline{R}^{-1T} \underline{L}_m \underline{R}^{-1} = \underline{L}_c R^c_m$$

or since $\underline{R}^{-1} = \underline{R}^T$

$$\underline{R} \underline{L}_m \underline{R}^{-1} = \underline{L}_c R^c_m.$$

This just says that the so called adjoint action of the matrix Lie group on its Lie algebra $\underline{K} \rightarrow \text{Ad}(\underline{R})\underline{K} = \underline{R} \underline{K} \underline{R}^{-1}$ leads to a rotation of the basis of the Lie algebra. This is in fact the identity derived in general in Exercise 1.7.8 where we introduced the adjoint action $\text{AD}(\underline{A}) \underline{B} = \underline{A} \underline{B} \underline{A}^{-1}$. If we let $\underline{R} = e^{\theta^a \underline{L}_a}$ and $\lambda \underline{X} = \theta^a \underline{L}_a$, then that identity becomes

$$\text{AD}(e^{\lambda \underline{X}}) \underline{Y} = e^{\lambda \text{ad}(\underline{X})} \underline{Y} \leftrightarrow \text{AD}(e^{\theta^c \underline{L}_c}) Y^b \underline{L}_b = \underline{L}_a (e^{\theta^c \text{ad}(\underline{L}_c)})^a_b Y^b$$

but in fact for the rotation group we have $(\underline{L}_a)^b_c = \epsilon_{bac} = (\underline{k}_a)^b_c$, i.e., the linear adjoint group of the rotation group is itself.

b) Use the above identity for $\text{AD}(\underline{R})$ applied to the differential relationships above, applying $\text{Ad}(\underline{R})$ to the first one to obtain the second one and comparing coefficients of the basis matrices, show that

$$\tilde{\omega}^a = R^a_b \omega^b.$$

c) Then show that

$$\delta_{ab} \tilde{\omega}^a \tilde{\omega}^b = \delta_{ab} \omega^a \omega^b \equiv 4ds^2.$$

d) Show that left multiplication by a constant rotation leaves the relation $\underline{R}^{-1} d\underline{R} = \omega^a \underline{L}_a$ invariant, so the differentials ω^a are invariant under such a “left translation” of the group. They are said to be “left invariant.” Repeat for the other differentials $\tilde{\omega}^a$ which are called right invariant. Thus the previous part c) defines a so called “bi-invariant” (both left and right invariant) metric on the rotation group, also invariant under the inversion map which sends each group element to its inverse, interchanging left and right on the group. We will see next that this is essentially just the metric on the unit 3-sphere S^3 within \mathbb{R}^4 and hence multiplying it by a^2 yields the metric on the 3-sphere of radius a .

■

Hermitian and unitary matrices

Complex matrix groups are also very useful in the real world, especially in quantum physics where complex scalar wave functions Ψ define probability distributions through the square of their absolute value $|\Psi|^2 = \Psi^* \Psi \geq 0$. On the complex vector space \mathbb{C}^n of n -tuples of complex numbers with standard basis \vec{e}_i , identical to the real basis of \mathbb{R}^n , there are two natural inner products which make this standard basis orthonormal. One is the usual bilinear dot product for which

$$\langle \underline{z}, \underline{z} \rangle = \langle z^1, \dots, z^n \rangle \cdot \langle z^1, \dots, z^n \rangle = \underline{z}^T \underline{z} = \delta_{ab} z^a z^b \in \mathbb{C}$$

but the following one (referred to as sesquilinear instead of bilinear, since it is linear in one argument, but “antilinear” in the other) enables the result of self-inner products to be interpreted as probability densities in quantum mechanics (overbars represent complex conjugates of complex numbers)

$$\langle \underline{z}, \underline{z} \rangle_s = \langle \overline{z^1}, \dots, \overline{z^n} \rangle \cdot \langle z^1, \dots, z^n \rangle = \underline{\bar{z}}^T \underline{z} = \delta_{ab} \bar{z}^a z^b \geq 0.$$

The latter inner product requires a complex conjugate acting on the matrix \underline{A} of a linear transformation when transposed from the right factor to the left factor

$$\langle \underline{w}, \underline{z} \rangle_s = \underline{\bar{w}}^T \underline{A} \underline{z} = (\underline{\bar{A}}^T \underline{\bar{w}})^T \underline{z} = \langle \underline{\bar{A}}^T \underline{\bar{w}}, \underline{z} \rangle_s$$

so it is useful to define the Hermitian conjugate of a matrix as the complex conjugate (overbar) of the transpose

$$\underline{A}^\dagger \equiv \underline{\bar{A}}^T \quad (\text{Hermitian conjugate}).$$

This is also called the adjoint matrix.

The adjoint operation on complex matrices plays the same role for the sesquilinear dot product that the transpose operation plays for the bilinear dot product. Hermitian matrices are those matrices which equal their adjoint and are sometimes referred to as self-adjoint

$$\underline{H}^\dagger = \underline{H} \quad (\text{Hermitian condition}).$$

These generalize the symmetric matrices for the ordinary dot product. Anti-Hermitian matrices

$$\underline{K}^\dagger = -\underline{K} \quad (\text{anti-Hermitian condition})$$

then play the role of the antisymmetric matrices for the ordinary dot product.

Note that if $\underline{K} = i\underline{\mathcal{K}}$ is anti-Hermitian, i.e., $\underline{\mathcal{K}} = -i\underline{K}$, then multiplying it by i makes it Hermitian

$$\underline{\mathcal{K}}^\dagger = (-i\underline{K})^\dagger = i(-\underline{K}) = \underline{\mathcal{K}}.$$

If a matrix is real and symmetric, the transpose and complex conjugate operations do nothing to it so it is Hermitian. If a matrix is purely imaginary and antisymmetric, they both change the sign so together they do nothing and the matrix is Hermitian. Thus Hermitian matrices consist of all real linear combinations of real symmetric matrices (dimension $n(n+1)/2$ and purely imaginary antisymmetric matrices (dimension $n(n-1)/2$, a real direct sum vector space of dimension n^2 . The tracefree subspace has dimension $n^2 - 1$. Thus the Lie algebra $su(n)$ has dimension $n^2 - 1$: $\dim(su(2)) = 3, \dim(su(3)) = 8$.

Remark.

Exercise 1.2.2 introduced the Paoli matrices plus the identity matrix as a basis of the subspace $h(2)$ of 2×2 complex matrices. $\underline{I}, \underline{\sigma}_1, \underline{\sigma}_3$ span the 3-dimensional subspace of real symmetric matrices and $\underline{\sigma}_2$ spans the 1-dimensional subspace of purely imaginary antisymmetric matrices, thus providing a basis for the 2×2 Hermitian matrices. Thus $h(2)$ is an appropriate symbol for this real vector space of complex matrices.

In two dimensions, there is a single real tracefree diagonal (symmetric) matrix, one real off-diagonal symmetric matrix, and one real antisymmetric matrix, which lead to the three matrices introduced by Pauli as though by magic and explains why one of them is multiplied by i . Finally after 40 years I understand how natural they are from this point of view of matrix symmetry properties, rather than just blindly using them for their important connection with rotations explored next.

▲

Why are Hermitian matrices useful? Well, we already know from our experience in the plane that a rotation generator (antisymmetric matrix) has purely imaginary eigenvalues, while a boost generator (symmetric matrix) has real eigenvalues, as do the real diagonal matrices, so multiplying the rotation generator by i makes it have real eigenvalues too, so Hermitian 2 matrices have real eigenvalues. Suppose $\underline{K} \underline{x} = \lambda \underline{K} \underline{x}$ for a Hermitian matrix $\underline{K} = \underline{K}^\dagger$. Then

$$\underline{x}^\dagger \underline{K} \underline{x} = \lambda \underline{x}^\dagger \underline{K} \underline{x} = \lambda |\underline{x}|^2$$

but the Hermitian conjugate of the left hand side (a 1×1 matrix, so just the complex conjugate) is itself since like the transpose $(\underline{A} \underline{B})^\dagger = \underline{B}^\dagger \underline{A}^\dagger$, etc. (for more factors)

$$\overline{(\underline{x}^\dagger \underline{K} \underline{x})} (\underline{x}^\dagger \underline{K} \underline{x})^\dagger = \underline{x}^\dagger \underline{K}^\dagger \underline{x} = \underline{x}^\dagger \underline{K} \underline{x}.$$

So it equals its complex conjugate, and is real, but the right hand side must also be real, so λ is real.

Physically observable quantities in physics must be real, so matrices which represent some physical quantity in quantum mechanics must have real eigenvalues, which turn out to be the observables in quantum mechanics. Angular and linear momentum are examples, where the quantum operators are related to the generators of rotations and translations by a factor of i . They describe how wave functions behave under dragging along by these symmetry transformations of space.

Matrix transformations of the basis which preserve the bilinear inner product lead to the complex orthogonal condition $\underline{Q}^T \underline{Q} = \underline{I}$ or equivalently $\underline{Q}^T = \underline{Q}^{-1}$ as in the real case, defining the complex orthogonal and special orthogonal groups $O(n, \mathbb{C})$ and $SO(n, \mathbb{C})$, the latter of which has complex antisymmetric matrices as its matrix Lie algebra just like in the real case.

In the second case of the sesquilinear inner product, this leads instead to the unitary condition which preserves the inner product of vectors like the orthogonal matrices in the real case

$$\underline{U}^\dagger \underline{U} = \underline{I} \leftrightarrow \underline{U}^\dagger = \underline{U}^{-1},$$

This condition guarantees that probability densities are invariant under the linear transformation of the complex vector on which it acts, called spinors. The group of $n \times n$ unitary matrices is denoted by $U(n)$. The special unitary subgroup $SU(n)$ satisfies the additional condition $\det \underline{U} = 1$, which leads to the tracefree condition on its Lie algebra

$$\underline{U} = e^{i\underline{K}} : \quad \det \underline{U} = 1 \rightarrow \text{Tr } \underline{K} = 0.$$

2-component spinor wave function fields are used to represent particles with half-integral spin (“fermions”) because of how they transform under the special unitary group $SU(2)$ either directly through its defining action on \mathbb{C}^2 through matrix multiplication or through its representation on the space \mathbb{C}^4 of 4-component Dirac spinors through the Dirac matrix algebra. $SU(2)$ is locally isomorphic to the rotation group $S(3, \mathbb{R})$, as will be explored in the next exercise, and spin is a concept intimately connected to rotations: a spinning body is a rotating body.

Exercise 4.5.8.

unitary groups

a) Show by differentiation of a curve in $\underline{U}(\lambda) = e^{i\lambda \underline{K}}$ through the identity $\underline{U}(0) = \underline{I}$ at $\lambda = 0$ leads to the anti-Hermitian condition on its Lie algebra matrices $i\underline{K}$

$$(i\underline{K})^\dagger = -i\underline{K} \quad (\text{anti-Hermitian condition}),$$

which explains the factor of i , without which the matrices are Hermitian

$$\underline{K}^\dagger = \underline{K} \quad (\text{Hermitian condition}).$$

Show that this second condition follows from the first. Physicists like Hermitian matrices since they are associated with “observable quantities” in quantum mechanics. ■

Exercise 4.5.9.

the special unitary group $SU(2)$ and $SO(3, \mathbb{R})$

FIX ??

In Exercise 1.2.2, we introduced the three tracefree Pauli matrices

$$\underline{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \underline{\sigma}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \underline{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

a) Use a computer algebra system to show that together with the identity matrix these generate an algebra, namely all products of these matrices can be expressed as linear combinations of themselves so that the vector space they generate is closed under multiplication?? Namely establish the product formula by first considering the self-products and then all distinct factor products

$$\underline{\sigma}_a \underline{\sigma}_b = \delta_{ab} \underline{I} + i \sum_{c=1}^3 \epsilon_{abc} \underline{\sigma}_c.$$

This contains symmetric and antisymmetric parts

$$\underline{\sigma}_{(a} \underline{\sigma}_{b)} = \delta_{ab} \underline{I}, \quad \underline{\sigma}_{[a} \underline{\sigma}_{b]} = i \sum_{c=1}^3 \epsilon_{abc} \underline{\sigma}_c.$$

Show that the latter is equivalent to the following commutation relations for the rescaled matrices

$$[\underline{E}_a, \underline{E}_b] = \sum_{c=1}^3 \epsilon_{abc} \underline{E}_c \quad \underline{E}_a = \frac{i}{2} \sigma_a.$$

b) Use the symmetric part of their matrix products to evaluate the matrix exponential

$$e^{i\theta n^a \underline{\sigma}_a} / 2, \quad \underline{n} \cdot \underline{n} = \delta_{ab} n^a n^b = 1.$$

by summing the even and odd powers of θ separately as in Exercise 1.7.9, recognizing the power series for the cosine and sine as coefficients. Show that the result is

$$e^{i\theta n^c \underline{\sigma}_c} = \cos \theta \underline{I} + i \sin \theta n^c \underline{\sigma}_c.$$

This generalizes Euler's formula for the exponential of 1×1 anti-Hermitian matrices, i.e., purely imaginary numbers: $e^{i\theta} = \cos \theta + i \sin \theta$.

This result takes the form

$$\underline{U} = e^{i\theta n^c \underline{\sigma}_c} = a^4 \underline{I} + i a^c \underline{\sigma}_c = \begin{pmatrix} a^4 + i a^3 & i a^1 + a^2 \\ -a^1 + a^2 & a^4 - i a^3 \end{pmatrix}, \quad .$$

Show that $\underline{U}^\dagger = \underline{U}$ is unitary, as must be the case since multiplying the Paoli matrices by i makes them anti-Hermitian. Since the Paoli matrices are tracefree their exponentials have unit determinant and lie in the special unitary group $SU(2)$ for which they form a basis of the Lie algebra $su(2)$ once multiplied by i .

c) Show that the special unitary condition of unit determinant takes the form

$$1 = \det(\underline{U}) = \dots = (a^1)^2 + (a^2)^2 + (a^3)^2 + (a^4)^2,$$

which is the equation for the unit sphere S^3 in Euclidean \mathbb{R}^4 in Cartesian coordinates a^α ($\alpha = 1, 2, 3, 4$). Thus the group manifold of $SU(2)$ can be identified with the 3-sphere.

d) The summation sign due to the wrong index positioning for summation is easily fixed by introducing the components of the structure constant tensor of the Lie algebra in this basis

$$[\underline{E}_a, \underline{E}_b] = C_{ab}^c \underline{E}_c, \quad C_{ab}^c = \epsilon_{cab}.$$

Thus these matrices have the same commutation relations as the rotation matrices $\{\underline{L}_a\}$. This in turn means that the rotation group is the linear adjoint group of the Lie algebra $su(2)$

$$\text{AD}(e^{\frac{i}{2}\theta^c \underline{\sigma}_c}) \underline{\sigma}_a = \underline{\sigma}_b (e^{\theta^c \underline{k}_c})^b{}_a.$$

Re-express the above exponential to get

$$e^{\theta n^c \underline{E}_c} = \cos\left(\frac{\theta}{2}\right) \underline{I} + \sin\left(\frac{\theta}{2}\right) n^c \underline{E}_c.$$

Note that the factor of 2 requires $\theta = 4\pi$ to return to the identity

$$e^{4\pi n^c \underline{E}_c} = \underline{I}.$$

e) Note that the adjoint action of the unitary group

$$a^a \underline{\sigma}_a \rightarrow \underline{U} \underline{\sigma}_a \underline{U}^{-1}$$

leaves the above trace inner product of these Pauli matrices invariant and hence must result in a rotation of the basis

$$\underline{U} \underline{\sigma}_a \underline{U}^{-1} = \underline{\sigma}_b R^b_a.$$

This defines a 2-to-1 map $(\underline{U}, -\underline{U}) \rightarrow \underline{R}$ from the 3-dimensional matrix group $SU(2)$ and hence from the unit 3-sphere S^3 to the 3-dimensional matrix group $SO(3, \mathbb{R})$ which clearly satisfies

$$\underline{U}_1 \underline{U}_2 \rightarrow \underline{R}_1 \underline{R}_2$$

Verify this composition law. This is said to define a Lie group homomorphism which is locally 1 to 1 near the identity matrices, and which corresponds to identifying antipodal points on the unit sphere to correspond to a single rotation matrix. The result stated in part b) makes this linear adjoint group relationship more explicit in terms of an exponential representation of the matrices in both groups. This relationship turns out to be extremely important in describing the spin of most of the elementary particles in nature (which have half integral spin). The elements of \mathbb{C}^2 on which $SU(2)$ acts by matrix multiplication are called spinors and are instrumental in describing spin states in nonrelativistic quantum mechanics. The above result of part d) shows that while spinors undergo a rotation by an angle $\theta/2$ about the axis \vec{n} , the corresponding adjoint matrix rotates the Pauli basis by a rotation by angle θ about that axis. However, spinors must be rotated by 4π to return to their original state, while ordinary 3-vectors return to their original state after a single revolution by 2π . ■

Exercise 4.5.10.

$SL(2, \mathbb{R})$ and the Lorentz group in 3 dimensions

Define the three tracefree matrices

$$\underline{\rho}_1 = \underline{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \underline{\rho}_2 = i\underline{\sigma}_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \underline{\rho}_3 = \underline{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

which are a basis of the vector space of tracefree 2×2 matrices, the Lie algebra $sl(2, \mathbb{R})$ of the special linear group $SL(2, \mathbb{R})$ in 2 dimensions. Let

$$\underline{S} = a^4 \underline{I} + a^a \underline{\rho}_a = \begin{pmatrix} a^4 + a^3 & a^1 + a^2 \\ a^1 - a^2 & a^4 - a^3 \end{pmatrix}.$$

a) Show that

$$1 = \det(\underline{S}) = \dots = -(a^1)^2 + (a^2)^2 - (a^3)^2 + (a^4)^2$$

which is a unit hyperboloid in \mathbb{R}^4 in Cartesian coordinates a^α ($\alpha = 1, 2, 3, 4$).

b) Evaluate the commutation relations of the basis $\{\frac{1}{2}\underline{\rho}_a\}$ of the Lie algebra

$$\left[\frac{1}{2}\rho_a, \frac{1}{2}\rho_b\right] = C^c{}_{ab}\frac{1}{2}\rho_c, \quad C^1{}_{23} = -C^2{}_{31} = C^3{}_{12} = 1.$$

c) Show that

$$(\text{Tr } \underline{\rho}_a \underline{\rho}_b) = 2 \text{diag}(1, -1, 1),$$

and even further

$$\left(\frac{1}{2}(\underline{\rho}_a \underline{\rho}_b + \underline{\rho}_b \underline{\rho}_a)\right) = \text{diag}(1, -1, 1)$$

so that

$$(\theta^a \underline{\rho}_a)(\theta^a \underline{\rho}_a) = (\theta^1)^2 - (\theta^2)^2 + (\theta^3)^2.$$

Use this to sum the exponential series by separating out the even and odd series to recognize the power series for the hyperbolic cosine and sine to obtain (setting $\theta^a = \alpha n^a$ with unit vector \vec{n} satisfying $(n^1)^2 - (n^2)^2 + (n^3)^2 = 1$)

$$\underline{S}(\alpha \vec{n}) = e^{\frac{1}{2}\alpha n^a \underline{\rho}_a} = \cosh\left(\frac{\alpha}{2}\right) \underline{I} + \sinh\left(\frac{\alpha}{2}\right) n^a \underline{\rho}_a = a^4 \underline{I} + a^a \underline{\rho}_a.$$

d) Note that the adjoint action of the special linear group leaves the above trace inner product of these Lie algebra basis matrices invariant and hence must result in a Lorentz transformation of the basis

$$\underline{S} \underline{\rho}_a \underline{S}^{-1} = \underline{\rho}_b L^b{}_a.$$

This defines a 2-to-1 map $(\underline{S}, -\underline{S}) \rightarrow \underline{L}$ from the 3-dimensional matrix group $SL(2, \mathbb{R})$ to the 3-dimensional matrix group $SO(2, 1)$ exactly as in the previous exercise.

In fact the complex $SL(2, \mathbb{C})$ matrix group is a 6-dimensional Lie group which is related to the full 6-dimensional Lorentz group in 4 spacetime dimensions in exactly this same way. ■

Exercise 4.5.11. quaternions?

The real linear combinations of the Pauli matrices together with the unit matrix not only have a Lie algebra structure (the unit matrix commutes with any other matrix, and together they are a basis of the Lie algebra $u(2)$ of the unitary group $U(2)$, not a rock band) as explored in Exercise 4.5.9, but actually generate a field called the quaternions. The real numbers are a field, and the complex numbers are a field, equivalent to an algebra structure on a real 2-dimensional vector space, and the quaternions are a field, equivalent to an algebra structure on a real 4-dimensional vector space. In an algebra, the elements can be both added and multiplied resulting again in an element of the same space, obeying all the usual rules that we know for real numbers except for the commutative rule for multiplication, which is an extra condition. Indeed the quaternions are the first noncommutative field in the hierarchy of increasing real

dimension. The basis of the quaternion algebra is usually denoted by $\{1, i, j, k\}$ but the matrices $\{\underline{E}_\alpha\} = \{\underline{I}, i\underline{\sigma}_a\}$, where $\alpha = 0, 1, 2, 3$ are a faithful representation of the algebra they generate (in the sense that products of quaternions correspond exactly to products of the corresponding matrices).

a) Show that

$$\underline{E}_0 \underline{E}_0 = \underline{E}_0, \quad \underline{E}_0 \underline{E}_a = \underline{E}_a = \underline{E}_a \underline{E}_0, \quad \underline{E}_a \underline{E}_b = -\delta_{ab} \underline{E}_0 + \epsilon_{abc} \underline{E}_c.$$

b) Using the notation $a^\alpha \underline{E}_\alpha = a^0 \underline{E}_0 + [\vec{a}]^c \underline{E}_c$ for a general quaternion (with real coefficients a^α), show that the multiplication of two quaternions incorporates both the dot and cross products that are so crucial to the structure of the rotation group on \mathbb{R}^3 .

$$(a^\alpha \underline{E}_\alpha) (b^\beta \underline{E}_\beta) = (a^0 b^0 - \vec{a} \cdot \vec{b}) \underline{E}_0 + a^0 b^c \underline{E}_c + b^0 a^c \underline{E}_c + [\vec{a} \times \vec{b}]^c \underline{E}_c.$$

c) \underline{E}_0 is a unit quaternion playing the role of the number 1 in both the real and complex numbers. As in the complex numbers one can introduce the conjugate quaternion by

$$\underline{A} = a^\alpha \underline{E}_\alpha \rightarrow \overline{\underline{A}} = a^0 \underline{E}_0 - a^a \underline{E}_a$$

and magnitude by

$$|\underline{A}|^2 = \overline{\underline{A}} \underline{A} = (a^0)^2 + \delta_{ab} a^a a^b.$$

Like the unit complex numbers which reside on the unit circle S^1 and can be represented as the exponential of a purely imaginary number, the unit quaternions reside on the unit sphere S^3 ($(a^0)^2 + \delta_{ab} a^a a^b = 1$) and can be represented as the exponential of a “purely spatial quaternion” $a^a \underline{E}_a$, namely

$$e^{a^a \underline{E}_a}.$$

Show that quaternion “conjugation” of the spatial quaternions by a fixed spatial quaternion is equivalent to a rotation of the spatial quaternions through what we have learned in the adjoint discussion of $SU(2)$ in Exercise 4.7.11. The quaternions incorporate the mathematics of both $SU(2)$ and $SO(3, \mathbb{R})$ and are fascinating in their own right, but enough said.

The slight modification of the Pauli matrices over the complex numbers associated with $SL(2, \mathbb{R})$ as described in Exercise 4.5.10 leads to a hyperbolic analog of the trigonometric geometry of the rotation group associated with the 3-dimensional Lorentz group of the Minkowski plane, and to the “Gödel quaternions” (also called “split quaternions”) which played a role in the Gödel solution of the Einstein equations involving rotating matter which surprised even Einstein in 1949.

■

Exercise 4.5.12.

squared angular momentum L^2

Given a basis \underline{E}_a of a matrix Lie algebra \mathfrak{g} , with corresponding adjoint matrices $\underline{k}_a = (C^b_{ac})$, then $\gamma_{ab} = \text{Tr}(\underline{k}_a \underline{k}_b) = C^m_{an} C^n_{bm}$ transforms as a symmetric covariant tensor under changes of basis. When this matrix of components has nonzero determinant, this or any multiple of it

represents a nondegenerate inner product on the Lie algebra called the Killing form. This is true for the standard basis of $so(3, \mathbb{R})$.

a) Since the adjoint representation is the identity for $so(3, \mathbb{R})$, namely $\underline{k}_a = \underline{L}_a$, one has the identity

$$\underline{R} \underline{k}_m \underline{R}^{-1} = \underline{k}_c R^c{}_m.$$

By writing this out in component form, show that this implies that the structure constant tensor itself is invariant under the adjoint action

$$C^a{}_{bc} = R^a{}_p C^p{}_{mn} R^{-1m}{}_b R^{-1n}{}_c$$

Thus the Killing form $\gamma_{ab} = -2\delta_{ab}$ is also invariant, which is true in general.

b) Since the adjoint representation is the identity representation for $SO(3, \mathbb{R})$: $(\underline{k}_a)^b{}_c = \epsilon_{bac}$, one has the following evaluation in which we raise and lower indices as we please

$$\gamma_{ab} = \epsilon_{man} \epsilon^{nbm} = \epsilon_{man} \epsilon^{mnb} = \delta^{nb}{}_{an} = -2\delta^a{}_b = -2\delta_{ab}.$$

Thus the matrix $\underline{L}^2 = \delta^{ab} \underline{L}_a \underline{L}_b$ is rotationally invariant. Evaluate this matrix and show that it is equal to $-s(s+1)\underline{I}$, where $s = 1$. Since this is a multiple of the identity matrix, this means it commutes with the individual basis matrices

$$[\underline{L}^2, \underline{L}_a] = 0.$$

This squared magnitude angular momentum operator L^2 turns out to be extremely important in classifying electronic states in the atom.

c) Evaluate γ for the structure constant tensor found in Exercise 4.5.10 for $SL(2, \mathbb{R})$, to find that it is proportional to the Lorentz inner product, which must be invariant under the adjoint action, which explains why the Lorentz group shows its face here. ■

Exercise 4.5.13.

unitary groups again

a) Suppose we consider a general “seuilinear” inner product on \mathbb{C}^n introduced in Exercise 4.5.8 and use angle bracket pairing to denote its evaluation

$$\langle X, Y \rangle = G_{ij} \overline{X^i} Y^j = \overline{\underline{X}}^T \underline{G} \underline{Y}$$

and require that the self-inner product of a vector be real

$$\overline{\langle X, X \rangle} = \langle X, X \rangle.$$

Expressing this in components, show that this requires that the matrix satisfy $\underline{G} = \overline{\underline{G}}^T \equiv \underline{G}^\dagger$, i.e., it should be a Hermitian matrix. This generalizes the symmetry condition for a real inner product, and extends the real pseudo-orthogonal groups which preserve them to the complex case.

b) Suppose we have a linear transformation A of \mathbb{C}^n into itself

$$\begin{aligned}\langle X, AY \rangle &= G_{ij} \overline{X^i} A^j_k Y^k = G_{km} A_i^m \overline{X^i} Y^k \\ &= \langle \text{Ad}_G(A) X, Y \rangle\end{aligned}$$

Show that the adjoint matrix with respect to this inner product is the ordinary Hermitian conjugate with its indices then raised and lowered by the inner product matrix in a particular way (since the inner product is not symmetric but Hermitian, the ordering of its component indices matters)

$$\underline{\text{Ad}}_G(A) = \underline{G}^{-1} \underline{A}^\dagger \underline{G} \quad \leftrightarrow \quad (\text{Ad}_G(A))^i_j = G^{ik} \overline{A^m_k} G_{mj}.$$

When the inner product component matrix is the identity matrix $\underline{G} = \underline{I}$ in an orthonormal basis, this reduces to the Hermitian conjugate. In the case of real matrices, this condition corresponds to the ordinary transpose of the index-lowered matrix, which is the operation used to describe the antisymmetry of the matrices in the Lie algebra of the pseudo-orthogonal groups which preserve symmetric inner products. Correspondingly for Hermitian inner products in an orthonormal basis, anti-Hermitian matrices generate the unitary matrix groups which are symmetries of that inner product.

■

Part II

CALCULUS

Chapter 5

From multivariable calculus to the
foundation of differential geometry

5.1 The tangent space in multivariable calculus

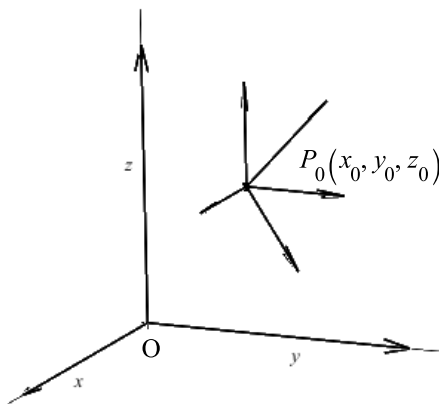


Figure 5.1: The tangent space at $P_0 \in \mathbb{R}^3$ consists of all difference vectors $\overrightarrow{OP} - \overrightarrow{OP}_0$, i.e., with initial point at P_0 .

The space \mathbb{R}^3 consisting of all triplets (a, b, c) of real numbers has many different mathematical structures. Most simply, it can be thought of as a space of points with no additive structure, with the values of the three Cartesian coordinates $\{x, y, z\}$ at a point serving to locate that point relative to the standard orthogonal axes on the space. Alternatively one can think of \mathbb{R}^3 as a space of vectors, i.e., as a real vector space with vector addition and scalar multiplication. In this case the points P of \mathbb{R}^3 are reinterpreted as directed line segments \overrightarrow{OP} or “arrows” with initial point at the origin O and terminal point at the point P . The notation $\vec{\mathbf{r}} = \langle x, y, z \rangle$ for the position vector of the point $P(x, y, z)$ emphasizes this vector interpretation of the point $(x, y, z) \in \mathbb{R}^3$.

In this case the Cartesian coordinates of a point are reinterpreted as the components of the corresponding vector with respect to the standard basis $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$ of \mathbb{R}^3 , often designated respectively by $\hat{i}, \hat{j}, \hat{k}$, where the “overhat” is a reminder that these are “unit vectors.” The Cartesian coordinates $\{x, y, z\}$ are real-valued linear functions on \mathbb{R}^3 which pick out the associated component of a vector with respect to the standard basis: $x((a, b, c)) = a$, $y((a, b, c)) = b$, $z((a, b, c)) = c$. In other words they are just the basis dual to the standard basis of \mathbb{R}^3 when thought of as a vector space: $\{x, y, z\} = \{\omega^1, \omega^2, \omega^3\}$, in the context of which the notation $\omega^1(\langle u^1, u^2, u^3 \rangle) = x(u^1, u^2, u^3) = u^1$ is more suggestive.

The closest one comes to the terminology “tangent space” in a first pass at multivariable calculus is the tangent plane to the graph of a function of two variables or to the level surface of a function of three variables. The idea of the tangent space to a point P_0 is not formally

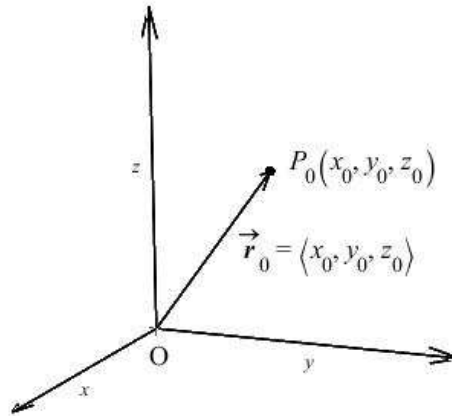


Figure 5.2: The dual interpretation of elements (x_0, y_0, z_0) of \mathbb{R}^3 as points $P_0(x_0, y_0, z_0)$ and vectors $\overrightarrow{OP_0} = \langle x_0, y_0, z_0 \rangle \equiv \vec{r}_0$.

introduced but it is nonetheless used and understood. It is just the space of all difference vectors $\vec{r} - \vec{r}_0 = \langle x - x_0, y - y_0, z - z_0 \rangle$ for all points $P(x, y, z)$ of \mathbb{R}^3 . These difference vectors are pictured as arrows with initial point at P_0 and terminal point at P , i.e., as the directed line segments $\overrightarrow{P_0P}$. They are called tangent vectors at P_0 .

A basis for this tangent space is the standard basis $\{e_1, e_2, e_3\}$ of \mathbb{R}^3 thought of as directed line segments with their initial points at P_0 . To recall this interpretation, the symbols $\{e_1|_{P_0}, e_2|_{P_0}, e_3|_{P_0}\}$ can be used. Each such tangent space is a real vector space isomorphic to \mathbb{R}^3 itself and usually no distinction is made between them in multivariable calculus. However, tangent vectors are discussed in relation to curves and surfaces in \mathbb{R}^3 . The tangent vector to a parametrized curve is always thought of as attached to the point on the curve at which it is defined, while a normal vector determining the orientation of the tangent plane to a surface at a point is always thought of as attached to that point. Each of these are examples of tangent vectors.

The Cartesian coordinate differentials $\{dx, dy, dz\}$ at the point P_0 are sometimes introduced as new Cartesian coordinates translated from the origin to P_0 , but notationally the point P_0 is suppressed

$$\langle dx, dy, dz \rangle|_{P_0} = \langle x - x_0, y - y_0, z - z_0 \rangle = \vec{r} - \vec{r}_0 \equiv d\vec{r}|_{P_0}.$$

The value of these new coordinates at a point $P_1(x_1, y_1, z_1)$

$$dx|_{P_0}(P_1) = x_1 - x_0, \quad dy|_{P_0}(P_1) = y_1 - y_0, \quad dz|_{P_0}(P_1) = z_1 - z_0,$$

are just the components of the difference vector $\overrightarrow{P_0P_1}$ with respect to the basis $\{e_1|_{P_0}, e_2|_{P_0}, e_3|_{P_0}\}$ of the tangent space at P_0 . In other words the Cartesian coordinate differentials $\{dx|_{P_0}, dy|_{P_0}, dz|_{P_0}\}$

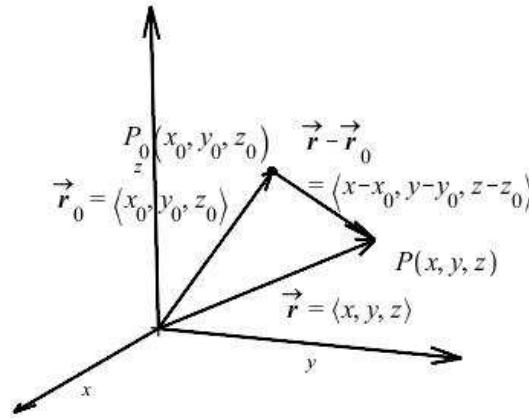


Figure 5.3: Tangent vectors at P_0 thought of as difference vectors with a fixed initial point at P_0 .

form the dual basis to this basis of the tangent space at P_0 . However, to interpret the differentials as the dual basis, we must agree to evaluate them on the difference vectors relative to P_0 rather than on their terminal points. If

$$\vec{X} = X^1 e_1|_{P_0} + X^2 e_2|_{P_0} + X^3 e_3|_{P_0}$$

is a tangent vector at P_0 , then

$$dx|_{P_0}(\vec{X}) = X^1, \quad dy|_{P_0}(\vec{X}) = X^2, \quad dz|_{P_0}(\vec{X}) = X^3$$

are its components evaluated using the dual basis. The differential of an arbitrary (differentiable) function f at point P_0 is defined in terms of the partial derivatives of f at P_0

$$df|_{P_0} = f_x(x_0, y_0, z_0)dx|_{P_0} + f_y(x_0, y_0, z_0)dy|_{P_0} + f_z(x_0, y_0, z_0)dz|_{P_0},$$

where we use interchangeably the subscript and partial notations for partial derivatives

$$f_x = \partial_x f = \frac{\partial f}{\partial x}.$$

The differential $df|_{P_0}$ is a real valued linear function on the tangent space at P_0 , i.e., a covector, also called a “1-form.” When evaluated on a tangent vector \vec{X} as above, it produces the result

$$df|_{P_0}(\vec{X}) = f_x(x_0, y_0, z_0)X^1 + f_y(x_0, y_0, z_0)X^2 + f_z(x_0, y_0, z_0)X^3 = \vec{X} \cdot \vec{\nabla} f(x_0, y_0, z_0),$$

where

$$\vec{\nabla} f(x_0, y_0, z_0) = f_x(x_0, y_0, z_0)e_1|_{P_0} + f_y(x_0, y_0, z_0)e_2|_{P_0} + f_z(x_0, y_0, z_0)e_3|_{P_0}$$

is the gradient of f at P_0 . The differential $df|_{P_0}$ represents the linear approximation to the function $f - f(x_0, y_0, z_0)$ at P_0 , as a function of the difference vectors relative to P_0 .

Remark.

Notice that the differential is defined as a covector on the tangent space, taking a linear combination of the components of the tangent vector \vec{X} in that space, but the dot product in our multivariable calculus notation represents that linear evaluation instead as a dot product “ $\vec{\nabla} f \cdot \vec{X}$ ” with a vector $\vec{\nabla} f$ whose components are the same as the covector differential df . This new vector is the gradient vector, and it is related to the differential by index raising with respect to the dot product. If we use indexed notation $(x^1, x^2, x^3) = (x, y, z)$, then $df = f_i dx^i$ has components $f_i = \partial_i f \equiv f_{,i}$ using the comma subscript to denote partial differentiation, where $f_i = df(e_i)$, while the gradient vector is $\vec{\nabla} f = f^i e_i$ with $f^i = \delta^{ij} f_j$, all objects referred to a particular tangent space. ▲

Two important uses of the tangent space in multivariable calculus occur in the discussion of tangent vectors to parametrized curves and in directional derivatives of functions, and they come together in the chain rule. Given a parametrized curve

$$\vec{\mathbf{r}} = \vec{\mathbf{r}}(t) = \langle x(t), y(t), z(t) \rangle$$

which passes through a point $P_0(x_0, y_0, z_0)$ at $t = t_0$, one produces a tangent vector at P_0 by differentiating it to produce the tangent vector to the curve at P_0

$$\vec{\mathbf{r}}'(t_0) = \langle x'(t_0), y'(t_0), z'(t_0) \rangle = x'(t_0)e_1|_{P_0} + y'(t_0)e_2|_{P_0} + z'(t_0)e_3|_{P_0},$$

where the last equality reminds us notationally of the connection of the tangent vector to the point P_0 . This distinction is never made but it is an integral part of the intuitive picture one has of the tangent vector.

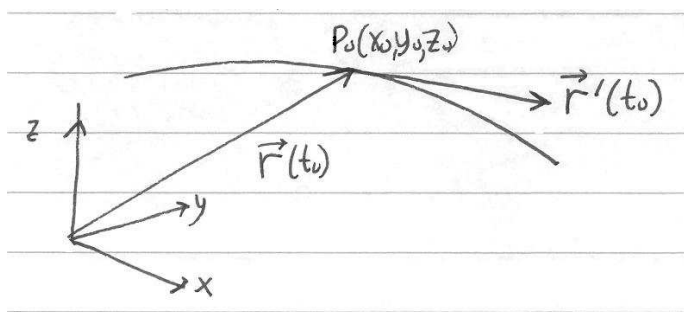


Figure 5.4: A parametrized curve $\vec{\mathbf{r}}(t)$ through P_0 with its tangent vector $\vec{\mathbf{r}}'(t)$ there.

One can think of the tangent space at P_0 as the space of tangent vectors at P_0 to all possible parametrized curves passing through P_0 . This is in fact a useful idea which generalizes to more complicated settings.

The chain rule evaluates the derivative of a function f on \mathbb{R}^3 along the parametrized curve as a function of the parameter

$$\begin{aligned}\frac{d}{dt}f(\vec{\mathbf{r}}(t))|_{t=t_0} &= f_x(x_0, y_0, z_0)x'(t_0) + f_y(x_0, y_0, z_0)y'(t_0) + f_z(x_0, y_0, z_0)z'(t_0) \\ &= \vec{\mathbf{r}}'(t_0) \cdot \vec{\nabla} f(x_0, y_0, z_0).\end{aligned}\quad (5.1)$$

However, this linear combination needs no inner product as we explained above—it is just the value of the differential of f at P_0 on the tangent vector $\vec{\mathbf{r}}'(t_0)$, namely

$$\frac{d}{dt}f(\vec{\mathbf{r}}(t))|_{t=t_0} = df|_{P_0}(\vec{\mathbf{r}}'(t_0)).$$

One can also differentiate a function along a given direction at P_0 without having an explicit parametrized curve. For this one introduces the directional derivative which generalizes the partial derivatives to an arbitrary direction specified by a unit vector $\hat{u} = \langle u^1, u^2, u^3 \rangle$, $\hat{u} \cdot \hat{u} = 1$. (We immediately adopt the correct index positioning u^i rather than the usual multivariable calculus notation u_i .) Taking the arclength parametrized straight line in the direction \hat{u} at P_0

$$\vec{\mathbf{r}} = \vec{\mathbf{r}}(s) = \langle x(s), y(s), z(s) \rangle : \quad x = x_0 + su^1, \quad y = y_0 + su^2, \quad z = z_0 + su^3,$$

for which the tangent vector is the given unit vector

$$\vec{\mathbf{r}}'(s) = \langle u^1, u^2, u^3 \rangle = \hat{u},$$

one defines the directional derivative $D_{\hat{u}}f$ of a function f at P_0 in the direction \hat{u} by an application of the chain rule

$$\begin{aligned}D_{\hat{u}}f(x_0, y_0, z_0) &\equiv \frac{d}{dt}f(\vec{\mathbf{r}}(s))|_{s=0} = \hat{u} \cdot \vec{\nabla} f(x_0, y_0, z_0) \\ &= u^1 f_x(x_0, y_0, z_0) + u^2 f_y(x_0, y_0, z_0) + u^3 f_z(x_0, y_0, z_0) \\ &= \left(u^1 \frac{\partial}{\partial x} \Big|_{P_0} + u^2 \frac{\partial}{\partial y} \Big|_{P_0} + u^3 \frac{\partial}{\partial z} \Big|_{P_0} \right) f = df|_{P_0}(\hat{u}).\end{aligned}$$

Along the coordinate directions this reduces to the ordinary partial derivatives, using the $\hat{i}, \hat{j}, \hat{k}$ notation for the basis unit vectors (the standard basis of \mathbb{R}^3)

$$D_{\hat{i}}f = f_x, D_{\hat{j}}f = f_y, D_{\hat{k}}f = f_z.$$

Note that the directional derivative $D_{\hat{u}}f(x_0, y_0, z_0)$ may be interpreted either as the result of allowing the first order differential operator

$$u^1 \frac{\partial}{\partial x} \Big|_{P_0} + u^2 \frac{\partial}{\partial y} \Big|_{P_0} + u^3 \frac{\partial}{\partial z} \Big|_{P_0} = \hat{u} \cdot \vec{\nabla} \Big|_{P_0}$$

to act on function f or by evaluating the differential of the function f at P_0 on the unit tangent vector \hat{u} . However, the usual multivariable calculus condition that \hat{u} be a unit vector requires

the use of the dot product, so if we want to generalize the directional derivative to a setting where no inner product is required, this restriction must be dropped.

So introduce the derivative of f at P_0 along any tangent vector \vec{X} there by

$$\begin{aligned}\nabla_{\vec{X}} f &= X^1 f_x(x_0, y_0, z_0) + X^2 f_y(x_0, y_0, z_0) + X^3 f_z(x_0, y_0, z_0) \\ &= \left(X^1 \frac{\partial}{\partial x} \Big|_{P_0} + X^2 \frac{\partial}{\partial y} \Big|_{P_0} + X^3 \frac{\partial}{\partial z} \Big|_{P_0} \right) f = df|_{P_0}(\vec{X}).\end{aligned}\quad (5.2)$$

In this way the chain rule links the derivative of a function along a parametrized curve to the derivative along its tangent vector

$$\frac{d}{dt} f(\vec{\mathbf{r}}(t))|_{t=t_0} = (\nabla_{\vec{\mathbf{r}}'(t_0)} f)(x_0, y_0, z_0) = \left(x'(t_0) \frac{\partial}{\partial x} \Big|_{P_0} + y'(t_0) \frac{\partial}{\partial y} \Big|_{P_0} + z'(t_0) \frac{\partial}{\partial z} \Big|_{P_0} \right) f.$$

This in turn may be interpreted as the result of a uniquely associated first order differential operator at P_0 acting on the function, namely the final expression in the large parentheses.

Still we haven't gone far enough with the tangent space idea. The notation of a tangent vector as a difference vector requires an underlying vector space in order to realize it as a difference of vectors. If we want to generalize this tangent vector idea to a setting without vector space structure (like the points on the surface of a sphere), the difference vector interpretation must be abandoned. The solution to this problem of disconnecting tangent vectors from their interpretation as difference vectors lies with tangent vectors to parametrized curves and derivatives of functions along them.

One can always differentiate functions along parametrized curves and the chain rule shows that this is equivalent to the derivative of those functions along the corresponding tangent vectors, regardless of how we try to interpret those tangent vectors. In fact with every vector at a point P_0 there is a uniquely associated first order linear differential operator which accomplishes the derivatives of functions along that tangent vector. It is just the linear combination of the partial derivatives at P_0 whose coefficients are the corresponding components of the tangent vector.

Why not simply define this differential operator to be the tangent vector? Indeed this is exactly what we will do.

This definition makes the coordinate partial derivative operators a basis of the tangent space at each point. The components of a tangent vector with respect to this basis are exactly what we've been calling the components all along. So this definition can be looked at as a bookkeeping trick. It turns out to be extremely useful. Thus our previous expansion of a tangent vector at P_0

$$\vec{X} = X^1 e_1|_{P_0} + X^2 e_2|_{P_0} + X^3 e_3|_{P_0}$$

can still be used if we re-interpret the symbols $e_i|_{P_0}$ to mean the corresponding partial derivatives at P_0 : $\partial/\partial x^i|_{P_0}$. The index notation $\{x^1, x^2, x^3\} = \{x, y, z\}$ for the three coordinate variables has to be introduced so that indexed equations using the summation convention can make formula writing simpler.

Finally index positioning must be respected. We can remind ourselves of the differential operator interpretation for a tangent vector \vec{X} by dropping the arrow notation and just let X denote the above tangent vector

$$\vec{X} = X^i e_i|_{P_0} \mapsto X = X^i \frac{\partial}{\partial x^i} \Big|_{P_0}.$$

Also, since we have changed our definition of tangent vectors, and since differentials were defined to be dual to tangent vectors, i.e., real valued linear functions on tangent vectors, their definition must be changed if we insist on maintaining duality. The differential of a function f at P_0 will be defined by an equation already used above but now acting on the new interpretation of a tangent vector as a first order differential operator

$$df|_{P_0}(X) = Xf \equiv X^i \frac{\partial f}{\partial x^i} \Big|_{P_0}.$$

The right hand side is a real valued linear function of the tangent vector X and so defines a covector or 1-form at P_0 . The coordinate differential $dx|_{P_0}$ is no longer a new Cartesian coordinate $x - x_0$, but the real valued linear map obtained by letting a tangent vector X act on the function x : $dx|_{P_0}(X) = Xx = (X^1\partial_x + X^2\partial_y + X^3\partial_z)x = X^1$ and so on, or in terms of the indexed coordinates $dx^i|_{P_0}(X) = Xx^i = X^i\partial_i x = X^i$. Thus the coordinate differentials merely pick out the components of tangent vectors in the coordinate derivative basis. This identification of the tangent space and its dual enables us to extend the concept to spaces which are locally like \mathbb{R}^n , called manifolds.

Example 5.1.1. Let $P_0(1, 2, 3)$ and $P_1(2, 0, 7)$ be two points in space with position vectors $\vec{r}_0 = \overrightarrow{OP_0} = \langle 1, 2, 3 \rangle$ and $\vec{r}_1 = \overrightarrow{OP_1} = \langle 2, 0, 7 \rangle$. Let's talk about the tangent space at the first point. The difference vector

$$\vec{X} = \overrightarrow{P_0P_1} = \vec{r}_1 - \vec{r}_0 = \langle 1, -2, 4 \rangle = e_1|_{P_0} - 2e_2|_{P_0} + 4e_3|_{P_0}$$

belongs to this tangent space. Its components are $dx^1|_{P_0}(X) = 1$, $dx^2|_{P_0}(X) = -2$, $dx^3|_{P_0}(X) = 4$.

It is also the tangent vector at P_0 to the parametrized curve which is the straight line connecting the two points parametrized in the most economical way:

$$\vec{r}(t) = \vec{r}_0 + t(\vec{r}_1 - \vec{r}_0) = \langle 1 + t, 2 - 2t, 3 + 4t \rangle.$$

Its constant tangent vector is $\vec{r}'(t) = \langle 1, -2, 4 \rangle$ in the usual calculus notation but now that we are distinguishing tangent vectors by their initial points we must write

$$\vec{r}'(0) = e_1|_{P_0} - 2e_2|_{P_0} + 4e_3|_{P_0}.$$

Next we move up the the differential operator interpretation of a tangent vector: $X = \partial/\partial x^1|_{P_0} - 2\partial/\partial x^2|_{P_0} + 4\partial/\partial x^3|_{P_0}$, dropping the arrow to distinguish it from the former mathematical quantity. It acts on a function

$$f = 4x^2 - 2xy + \frac{5}{2}y^2 + z^2 = 4(x^1)^2 - 2x^1x^2 + \frac{5}{2}(x^2)^2 + (x^3)^2$$

by

$$\begin{aligned} Xf &= \left(1 \frac{\partial}{\partial x^1} \Big|_{P_0} - 2 \frac{\partial}{\partial x^2} \Big|_{P_0} + 4 \frac{\partial}{\partial x^3} \Big|_{P_0} \right) f = 1 \frac{\partial f}{\partial x^1} \Big|_{P_0} - 2 \frac{\partial f}{\partial x^2} \Big|_{P_0} + 4 \frac{\partial f}{\partial x^3} \Big|_{P_0} \\ &= 1((8x - 2y) - 2(-2x + 5y) + 4(2z))|_{(1,2,3)} = 1(8(1) - 2(2)) - 2(-2(1) + 5(2)) + 4(6) \\ &= 1(4) - 2(8) + 4(6) = 12, \end{aligned}$$

or equivalently

$$\begin{aligned} df &= 1(8x - 2y) dx + (-2x + 5y) dy + (2z) dz, \\ df|_{P_0} &= (4 dx + 8 dy + 6 dz)|_{P_0}, \\ df|_{P_0}(X) &= 4(1) + 8(-2) + 6(4) = 12. \end{aligned}$$

Our mental image of the three basis tangent vectors as tangent vectors to the three coordinate curves pictured as arrows with initial points at $(1, 2, 3)$ remains, but the actual mathematical object will be the directional derivative along those arrow tangent vectors waiting for a function to act on. For a fixed function f and variable tangent vector X in this fixed tangent space, the quantity Xf is a linear function and hence a covector: $df|_{P_0}(X) = (4dx^1 + 8dx^2 + 6dz)|_{P_0}(X) = 4X^1 + 8X^2 + 6X^3$, which is the interpretation of the differential of a function evaluated at the point in question. \square

Exercise 5.1.1.

Some problems from 3-d calculus

Note the correspondence $\vec{X} = X^i e_i \longleftrightarrow X = X^i \partial/\partial x^i|_{P_0}$ for the tangent space at the point P_0 .

- Suppose $x = t$, $y = t^2 + 1$, $z = 2 - t$, or $\vec{\mathbf{r}}(t) = \langle t, t^2 + 1, 2 - t \rangle$.
Evaluate $\vec{\mathbf{r}}'(t)$. What is the tangent vector at $t = 1$?
Express it as a first order linear differential operator, call it $\mathbf{r}'(1)$ with no over arrow (and $\mathbf{r}'(t)$ in general).
- Consider the function $f(x, y, z) = x^2 + y^2 - 3z^2$.
What is $df(x, y, z)$? $df(1, 2, 1)$?
What is $\vec{\mathbf{r}}'(1)f$? [the action of the derivative operator on the function f].
- Find expressions for x, y, z as functions of t for some other parametrized curve which has the same tangent at $t = 0$ as the previous curve (such that $t = 0 : x = 0, y = 1, z = 2$ as with the previous curve). (This is easy! Try the tangent line as the other curve!)
- If $X = 2 \frac{\partial}{\partial x} \Big|_{(1,2,1)} - \frac{\partial}{\partial y} \Big|_{(1,2,1)} + 4 \frac{\partial}{\partial z} \Big|_{(1,2,1)}$, what is $df(1, 2, 1)(X)$?
If $\Theta = dx|_{(1,2,1)} + 2dy|_{(1,2,1)} - dz|_{(1,2,1)}$, what is $\Theta(\mathbf{r}'(1))$? $\Theta(X)$?

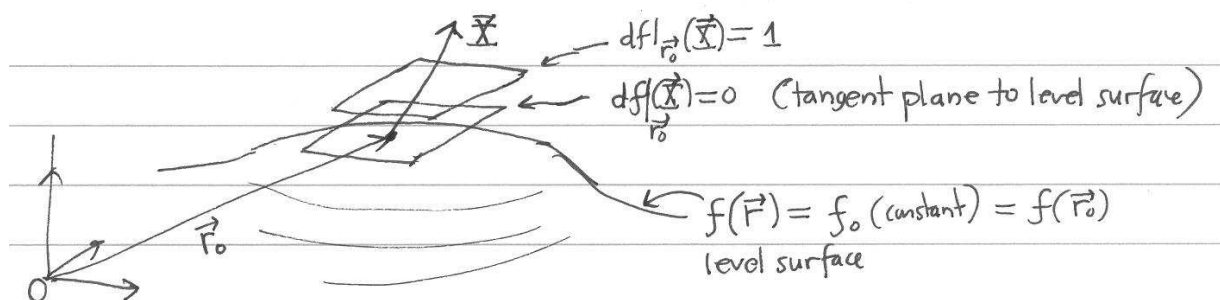


Figure 5.5: Visualizing the differential of a function at a point on a level surface.

The differential $df|_{\vec{r}_0}$ of f at \vec{r}_0 can be represented by the pair of planes in \mathbb{R}^3 shown in the diagram. The value $df|_{\vec{r}_0}(\vec{X})$ using the calculus meanings of differential and tangent vector (as a difference vector) is the number of integer spaced planes of this family containing these two planes which are pierced by \vec{X} , interpolating between the integers. The value on tangent vectors belonging to the tangent plane is zero since it does not pierce any of the planes in this family. In the old fashioned language

$$df|_{\vec{r}_0}(\vec{X}) = \vec{X} \cdot \vec{\nabla} f|_{\vec{r}_0} = 0$$

means that \vec{X} is perpendicular to the gradient of f which itself is orthogonal to the tangent plane to the level surface, making \vec{X} belong to this tangent plane.

The new meaning of the differential and tangent vector

$$df|_{P_0}(X) = Xf, \quad X \in T\mathbb{R}^3_{P_0}$$

tells us if $Xf = 0$, then X belongs to the tangent plane to the level surface of f at the point P_0 . Suppose we have instead a vector field X such that $Xf = 0$. This means $X|_P$ is tangent to the level surface of f through P at every point P .

Remark. vector fields

Suppose we take a vector field in the multivariable calculus sense $\vec{X}(x, y, z)$, for example the position vector defines a vector at each point of space $\vec{r} = \langle x, y, z \rangle$, whose value at any point defines a tangent vector there which we can picture as having its initial point located at that point, thus determining a field of tangent vectors, all pointing radially away from the origin. Similarly $X = x\partial_x + y\partial_y + z\partial_z$ is a vector field in the new sense, whose value at any point defines a tangent vector there in the new sense of a derivative operator whose coefficients are exactly the component functions of the corresponding vector field in the multivariable calculus sense.



Exercise 5.1.2.
tangent to level surfaces

Show that

$$X = x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}$$

is tangent to the level surfaces of the functions

$$f(x, y, z) = \frac{y}{x} \text{ and } g(x, y, z) = \frac{z}{\sqrt{x^2 + y^2}}.$$

Compute df and dg and $df(X)$, $dg(X)$ as well as Xf and Xg .

Note: Recall that $f = \tan(\phi)$ and $\arccos(g) = \theta$ relate these two functions to the polar coordinate functions: the azimuthal angle ϕ around the z -axis and the polar angle θ down from the z -axis, although these Greek letters are interchanged in calculus textbooks. This explains why the vector field whose components are the radial position vector components is tangent to their level surfaces. It lies in the cones of constant θ and in the vertical half planes of constant ϕ . Don't worry, we will study these coordinates soon enough if you don't recall their details now.



Exercise 5.1.3.
elliptical level curves

Consider the function $f(x, y) = 4x^2 - 2xy + \frac{5}{2}y^2$ from Exercise 1.6.12 and the two vector fields

$$X = (8x - 2y) \frac{\partial}{\partial x} + (-2x + 5y) \frac{\partial}{\partial y}, \quad Y = -(-2x + 5y) \frac{\partial}{\partial x} + (8x - 2y) \frac{\partial}{\partial y},$$

where the vector field Y at each point is related to X by a local rotation of the tangent space by a 90 degree angle counterclockwise: $\langle Y^1, Y^2 \rangle = \langle -X^2, X^1 \rangle = \langle \langle 0 | -1 \rangle, \langle 1 | 0 \rangle \rangle \langle X^1, X^2 \rangle$.

a) Compute df , then $df(X)$ and Xf , and then $df(Y)$ and Yf . What is $X \cdot X = |X|^2$, i.e., the length squared of this vector field? What is $X \cdot Y$?

b) Consider also the parametrized curve: $x(t) = \cos(6t) + \frac{1}{3}\sin(6t)$, $y(t) = \frac{4}{3}\sin(6t)$. What is the tangent vector $Z(t)$ to this curve? Evaluate $df(Z(t)) = Z(t)f$.



5.2 More motivation for the re-interpretation of the tangent space

The standard Cartesian coordinates $\{x^i\}$ on \mathbb{R}^n are those functions which pick out the individual components u^i of vectors $\vec{u} = \langle u^1, \dots, u^n \rangle = u^i e_i$ —these are just the dual basis covectors which are dual to the standard basis e_i

$$x^i \equiv \omega^i, \quad x^i(\langle u^1, \dots, u^n \rangle) = u^i.$$

However, since we are going to emphasize a different mathematical structure on \mathbb{R}^n , we will use a different notation. If $P = (u^1, \dots, u^n)$ is a point in \mathbb{R}^n , de-emphasizing its vector nature using a capital letter as we conventionally do for points, then

$$x^i|_P \equiv x^i(P) = u^i$$

will indicate the value of x^i at P .

We can now re-interpret a change of basis on \mathbb{R}^n as a change of Cartesian coordinates. A change of basis from the (old) standard basis e_i to a (new) basis $e_{i'}$ involving a (constant) matrix $\underline{A} = (A^i_j)$ is given by

$$\omega^{i'} = A^i_j \omega^j, \quad e_{i'} = A^{-1j}_i e_j.$$

The second equation shows that the columns (labeled by the lower right index) of the inverse matrix $\underline{B} = \underline{A}^{-1}$ are the old components of the new basis vectors $e_{i'}$. Rewritten in terms of Cartesian coordinates this becomes

$$x^{i'} = A^i_j x^j \quad \text{or} \quad x^i = A^{-1i}_j x^{j'},$$

which when evaluated at a particular point (i.e., on a particular vector) become

$$u^{i'} = A^i_j u^j \quad \text{or} \quad u^i = A^{-1i}_j u^{j'}.$$

By the definition of partial differentiation for a given coordinate system

$$\frac{\partial x^j}{\partial x^i} = \delta^j_i, \quad \frac{\partial x^{j'}}{\partial x^{i'}} = \delta^{j'}_{i'}$$

so

$$\frac{\partial x^{i'}}{\partial x^j} = \frac{\partial}{\partial x^j} (A^{i'}_k x^k) = A^{i'}_k \frac{\partial x^k}{\partial x^j} = A^{i'}_k \delta^k_j = A^{i'}_j$$

and

$$\frac{\partial x^i}{\partial x^{j'}} = \dots (\text{same calculation}) \dots = A^{-1i}_{j'}.$$

Thus the components of vectors transform by

$$u^{i'} = \frac{\partial x^{i'}}{\partial x^j} u^j \quad \text{or} \quad u^i = \frac{\partial x^i}{\partial x^{j'}} u^{j'},$$

which is said to be the transformation law for a contravariant vector. The matrix

$$\underline{J} = \left(\frac{\partial x^{i'}}{\partial x^j} \right)$$

is called the Jacobian matrix of the coordinate transformation and its inverse

$$\underline{J}^{-1} = \left(\frac{\partial x^i}{\partial x^{j'}} \right)$$

is just the Jacobian matrix of the inverse transformation. Although these are just constant matrices for such a simple transformation of Cartesian coordinates, they will depend on position in a transformation to more general “curvilinear” coordinate systems which are not adapted to the vector space structure of \mathbb{R}^3 like linear Cartesian coordinates are.

Suppose now that (u^i) are n real-valued functions on \mathbb{R}^n and we introduce the partial derivative operator $u = u^i \partial / \partial x^i$ on real-valued differentiable functions on \mathbb{R}^n , acting in the obvious way to produce new functions

$$f \mapsto uf = \left(u^i \frac{\partial}{\partial x^i} \right) f = u^i \frac{\partial f}{\partial x^i}.$$

By the chain rule and using the same vector transformation law for the “contravariant vector field” u^i , one finds

$$uf = u^i \frac{\partial f}{\partial x^i} = u^i \frac{\partial x^{j'}}{\partial x^i} \frac{\partial f}{\partial x^{j'}} = u^{j'} \frac{\partial f}{\partial x^{j'}} = \left(u^{j'} \frac{\partial}{\partial x^{j'}} \right) f.$$

Thus the vector transformation law for the coefficient functions in the linear differential operator guarantees that this differential operator has the same form in both coordinate systems, i.e., is “independent of coordinates,” producing the same result when acting on a function.

Conversely if one has a set of components which transform in this way, then using the chain rule in the form

$$\frac{\partial}{\partial x^{i'}} = \frac{\partial x^j}{\partial x^{i'}} \frac{\partial}{\partial x^j} = A^{-1j}_i \frac{\partial}{\partial x^j},$$

the combination

$$u^{i'} \frac{\partial}{\partial x^{i'}} = (A^i_k u^k) \left(A^j_i \frac{\partial}{\partial x^j} \right) = (A^{-1j}_i A^i_k) u^k \frac{\partial}{\partial x^j} = \delta^j_k u^k \frac{\partial}{\partial x^j} = u^j \frac{\partial}{\partial x^j}$$

is invariant. The same chain rule applied to a function

$$\frac{\partial f}{\partial x^{i'}} = \frac{\partial x^j}{\partial x^{i'}} \frac{\partial f}{\partial x^j} = A^{-1j}_i \frac{\partial f}{\partial x^j}$$

is said to define the transformation law of a covariant vector (field). In fact these are just the coefficients of the differential df expressed in terms of the coordinate differentials

$$df = \frac{\partial f}{\partial x^i} dx^i = \frac{\partial f}{\partial x^{i'}} dx^{i'}$$

which of course doesn't depend on which coordinates are used to express it. Thus expressing the coordinate independent operations uf and df in particular Cartesian coordinate systems leads to the usual coordinate transformation of the components of vectors and covectors point by point on \mathbb{R}^n using the Jacobian matrix and its inverse. By referring to the differential operator $u = u^i \partial/\partial x^i$ as the vector field instead of the collection of components (u^i) , the vector field enjoys the same invariant status as the differential of a function or even of an ordinary vector

$$v = \langle v^1, \dots, v^n \rangle = v^i e_i \in \mathbb{R}^n,$$

which is a quantity v independent of the choice of basis, whose components merely change with a change of basis.

On the other hand, although everybody knows the rules for evaluating differentials, the meaning of the differentials of the coordinates themselves is often lost on students or poorly presented in textbooks. We all remember that we plug in increments in the coordinates for them when we use the differential approximation, but the mathematical interpretation of the differentials themselves we quickly forget. It should therefore cause no great objection if we redefine what they mean mathematically, although the rules for taking differentials will remain the same.

Linearity of differentiation means

$$(a u + b v)f = \left(a u^i \frac{\partial}{\partial x^i} + b v^i \frac{\partial}{\partial x^i} \right) f = a u^i \frac{\partial f}{\partial x^i} + b v^i \frac{\partial f}{\partial x^i} = a u f + b v f,$$

so

$$(a u + b v)|_P f = a u|_P f + b v|_P f, \quad (\text{where } a, b \text{ are constants}).$$

This linearity condition means that associating to such a derivative operator at a given point of \mathbb{R}^n the value of its derivative of a particular function there

$$u|_P \mapsto u|_P f = u^i|_P \left. \frac{\partial f}{\partial x^i} \right|_P \in \mathbb{R}$$

is a real-valued linear function of the derivative operator. The space of all such derivative operators at a given point P is clearly an n -dimensional vector space isomorphic to \mathbb{R}^n

$$u^i|_P e_i \in \mathbb{R}^n \longleftrightarrow u^i|_P \left. \frac{\partial}{\partial x^i} \right|_P$$

so as a real-valued linear function, $u|_P \mapsto u|_P f$ defines a covector on that vector space. $\{\partial/\partial x^i|_P\}$ is then a basis of the space of these operators at P , and with respect to this basis, the components of the covector are $\partial f/\partial x^i|_P$ since

$$u|_P f = \left. \frac{\partial f}{\partial x^i} \right|_P u^i|_P.$$

By making the simple definition

$$df|_P(u) = u|_P f$$

for the covector, thereby defining the differential of the function f at P , we get a meaning for the differentials of the coordinates themselves

$$dx^i|_P(u) = u^j \frac{\partial}{\partial x^j} x^i = u^j \delta^i_j = u^i$$

as the covectors which pick out the components of these linear operators with respect to the basis $\{\partial/\partial x^i|_P\}$, i.e., the dual basis

$$dx^i|_P \left(\frac{\partial}{\partial x^j} \Big|_P \right) = \frac{\partial}{\partial x^j} \Big|_P x^i = \frac{\partial x^i}{\partial x^j} = \delta^i_j.$$

Of course this is exactly what we did in the previous section for \mathbb{R}^3 , presented in a slightly different way here for \mathbb{R}^n .

Let's use the notation $T\mathbb{R}_P^n$ for the tangent space to \mathbb{R}^n at the point P (T for tangent!) identified with the space of linear differential (derivative!) operators there. Then by this new definition of differential, the coordinate differentials form a basis for the dual space $(T\mathbb{R}_P^n)^*$, called the cotangent space at P . Thus at each point P of \mathbb{R}^n , we have the tangent space $V = T\mathbb{R}_P^n$ with basis $\{\partial/\partial x^i|_P\}$ and its dual basis $\{dx^i|_P\}$ of the dual cotangent space $V^* = (T\mathbb{R}_P^n)^*$, and we are free to consider all the spaces of $\binom{p}{q}$ -tensors over each such V and independent changes of basis at different points. Objects defined at each point of a space are called “fields.” Picking out (smoothly) a tangent vector at each point P leads to the already familiar concept of a vector field (familiar at least for $n = 2$ and $n = 3$).

The differential of a (smooth = differentiable) function leads to a covector field or “1-form” on \mathbb{R}^n . Fields of p -covectors are often called p -forms, or just “differential forms” without being specific about the number of indices. The differential is a special 1-form since its components come from the partial derivatives of a function. Given n functions Θ_i on \mathbb{R}^n , then $\Theta = \Theta_i dx^i$ defines a general 1-form field (or just a 1-form, with “field” understood by context).

The constant vector fields $e_i = \partial_i$ resulting from the standard basis vectors of \mathbb{R}^n comprise a field of bases, which is often called a frame, in the sense that it provides a local reference frame with which to measure the geometry near each point. The standard dual basis of 1-forms $\omega^i = dx^i$ is then referred to as the dual frame. One can smoothly pick any set of n vector fields which are linearly independent at each point of the space to introduce more general frames and the corresponding dual frames. The coordinate components of these frames and dual frames are simply inverse matrices. Note that the original Cartesian coordinates x^i we interpreted as the dual basis to the standard basis of \mathbb{R}^n as a vector space—now their differentials give us the dual basis to the standard basis of each tangent space of \mathbb{R}^n .

Similarly the Euclidean dot product tensor on \mathbb{R}^n itself enables us to define a corresponding constant tensor field, the Euclidean metric tensor field

$$G = \delta_{ij} dx^i \otimes dx^j, \quad \delta_{ij} = G \left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right)$$

which tells us how to take the lengths of vector fields by defining the basis $\{\partial/\partial x^i|_P\}$ at each point P to be orthonormal. This just reproduces the usual inner product (the dot product on \mathbb{R}^n)

$$G(u, v) = \delta_{ij} u^i v^j$$

for two vector fields u, v when expressed in terms of components. Inner product tensor fields are referred to as metrics since they provide a way to measure the local geometry at each point of space, the word “metric” being associated with standards of measurement. Conventionally a metric tensor field is denoted by a lower case g , a notation we will soon adopt, but for now we retain the uppercase letter from its origin in a symmetrix matrix \underline{G} .

A p -covector field or simply p -form has the expression

$$S = \frac{1}{p!} S_{i_1 \dots i_P} dx^{i_1} \wedge \dots \wedge dx^{i_P} \equiv \frac{1}{p!} S_{i_1 \dots i_P} dx^{i_1 \dots i_P},$$

while a p -vector field is of the form

$$T = \frac{1}{p!} T^{i_1 \dots i_P} \frac{\partial}{\partial x^{i_1}} \wedge \dots \wedge \frac{\partial}{\partial x^{i_P}},$$

where the components are now functions on \mathbb{R}^n . In particular the unit n -form associated with the dot product induced metric tensor field G is just

$$\eta = \eta_{1\dots n} dx^{1\dots n}.$$

with $\eta_{1\dots n} = \epsilon_{1\dots n} = 1$ in an oriented basis, and represents the determinant function in each tangent space.

Exercise 5.2.1.

polar coordinate calculations

Consider the functions and vector fields

$$r = (x^2 + y^2)^{1/2}, \quad \theta = \arctan(y/x), \quad X = x\partial/\partial x + y\partial/\partial y, \quad Y = x\partial/\partial y - y\partial/\partial x.$$

- Evaluate dr , $d\theta$ and $r dr \wedge d\theta$.
- Evaluate $X \wedge Y$, $dr(X)$, $dr(Y)$, $d\theta(X)$, $d\theta(Y)$ and $(r dr \wedge d\theta)(X, Y)$.
- r and θ are just polar coordinates in the plane. Evaluate $dr \otimes dr + r^2 d\phi \otimes d\phi$, simplifying it until you recognize the metric tensor field $G = dx \otimes dx + dy \otimes dy$.
- Now do the opposite easier calculation, evaluating and simplifying $G = dx \otimes dx + dy \otimes dy$ in terms of the polar coordinates using the inverse coordinate transformation: $(x, y) = (r \cos \theta, r \sin \theta)$. Then evaluate and simplify $dx \wedge dy$ in terms of the polar coordinates.

■

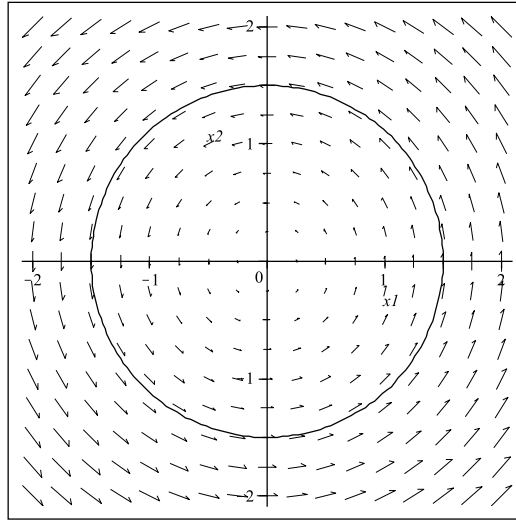


Figure 5.6: Visualizing the flow of the vector field $x \partial/\partial y - y \partial/\partial x$ in the plane. In this representation, the length of the vector field is scaled down so that its largest value fits within a grid box.

5.3 Flow lines of vector fields

The reinterpretation of a tangent vector as a directional derivative at a point makes sense for another good reason, the flow lines associated with a vector field. In a typical first course in differential equations, students learn to solve a linear homogeneous system of ordinary differential equations. For example, consider this system in the plane

scalar form: $\frac{dx^1}{dt} = -x^2, \quad x^1(0) = x_0^1$ $\frac{dx^2}{dt} = x^1, \quad x^2(0) = x_0^2$	vector form $\frac{d\underline{x}}{dt} = \underline{A} \underline{x}$ $\underline{x}(0) = \underline{x}_0$	index form: $\frac{dx^i}{dt} = A^i_j x^j$ $x^i(0) = x_0^i$
--	---	---

where

$$\underline{A} = (A^i_j) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \underline{x} = \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}.$$

By introducing the vector field $\xi = A^i_j x^j \partial/\partial x^i = x^1 \partial/\partial x^2 - x^2 \partial/\partial x^1$, i.e., with components $(\xi^1, \xi^2) = (-x^2, x^1)$, this system of differential equations states that the tangent to a solution curve parametrized by the variable t equals the value of the vector field at each point along it

$$\frac{dx^i(t)}{dt} = \xi^i(\underline{x}(t)).$$

The solution curves are called the integral curves of the vector field, or its “flow lines.” There is a unique such flow line through each point of the space, as long as the vector field is well

behaved everywhere. Indeed with modern mathematics technology, one can easily visualize both the vector field and its flow lines in the plane. Fig. 5.6 illustrates this for this particular rotational vector field, whose flow lines are circles about the origin.

This system is solved by finding an eigenbasis $\{\underline{b}_1, \underline{b}_2\}$ of the coefficient matrix: $\underline{A}\underline{b}_i = \lambda_i \underline{b}_i$, in terms of which the transformed matrix of coefficients is diagonal

$$\underline{A}_B = \underline{B}^{-1} \underline{A} \underline{B} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \quad \underline{B} = \langle \underline{b}_1 | \underline{b}_2 \rangle.$$

Then by introducing a linear change of variables: $\underline{x} = \underline{B}\underline{y}$, $\underline{y} = \underline{B}^{-1}\underline{x}$, the differential equations take the new form

$$\frac{d\underline{y}}{dt} = \underline{B}^{-1} \frac{d\underline{x}}{dt} = \underline{B}^{-1} \underline{A} \underline{x} = \underline{B}^{-1} \underline{A} \underline{B} \underline{y} = \underline{A}_B \underline{y}.$$

This decouples the differential equations for the new variables: $dy^i/dt = \lambda_i y^i$, which have exponential solutions $y^i = c^i e^{\lambda_i t}$ whose arbitrary constant coefficients are the initial values of the new variables at $t = 0$: $y^i = y_0^i e^{\lambda_i t}$, $y^i(0) = y_0^i = c^i$. Backsubstituting these expressions for the new variables into the matrix product which yields the old variables gives explicitly the general solution.

By going a small step farther than time usually allows in an undergraduate class on differential equations, writing the solution for the new variables in vector form with the exponential factors factored out into a diagonal matrix factor

$$\underline{y} = \begin{pmatrix} e^{\lambda_1 t} c^1 \\ e^{\lambda_2 t} c^2 \end{pmatrix} = \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix} \begin{pmatrix} c^1 \\ c^2 \end{pmatrix} = \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix} \underline{c},$$

one can backsubstitute using matrix notation to get the matrix form of the general solution

$$\underline{x} = \underline{B} \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix} \underline{c}.$$

Solving the initial condition transforms the coordinates of the initial position in the plane: $\underline{x}(0) = \underline{B}\underline{y}(0)$ leads to $\underline{B}\underline{c} = \underline{x}_0$ with solution $\underline{c} = \underline{B}^{-1}\underline{x}_0 = \underline{y}_0$, representing a change of coordinates of the initial point. Thus the matrix form of the solution of the initial value problem solution is

$$\underline{x} = \underline{B} \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix} \underline{B}^{-1} \underline{x}_0.$$

By defining $\underline{A}^0 = \underline{I}$ to be the identity matrix and introducing the matrix exponential

$$e^{\underline{A}} = \sum_{k=0}^{\infty} \underline{A}^k / k! = \underline{I} + \underline{A} + \frac{1}{2} \underline{A}^2 + \dots$$

one sees that for a diagonal matrix this consists of the diagonal entries which are the exponentials of the diagonal entries

$$\begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix} = e^{t \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}} = e^{t \underline{A}_B}.$$

Also, it is obvious from the property $\underline{B} \underline{A}^n \underline{B}^{-1} = (\underline{B} \underline{A} \underline{B}^{-1})(\underline{B} \underline{A} \underline{B}^{-1}) \cdots (\underline{B} \underline{A} \underline{B}^{-1}) = (\underline{B} \underline{A} \underline{B}^{-1})^n$ which applied term by term to the infinite series for the exponential leads to

$$\underline{B} e^{\underline{A}} \underline{B}^{-1} = e^{\underline{B} \underline{A} \underline{B}^{-1}}.$$

Putting this all together, we find

$$\begin{aligned} \underline{x} &= \underline{B} \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix} \underline{B}^{-1} \underline{x}_0 \\ &= e^{t \underline{B} \underline{A} \underline{B}^{-1}} \underline{x}_0 = e^{t \underline{A}} \underline{x}_0. \end{aligned}$$

Of course this final result is just the matrix version of the well known scalar exponential initial value problem $dx/dt = kx$, $x(0) = x_0$, which has solution $x = x_0 e^{kt}$, and the matrix result follows from the obvious “chain rule” derivative property

$$\frac{d}{dt} e^{t \underline{A}} = \underline{A} e^{t \underline{A}} = e^{t \underline{A}} \underline{A}$$

which follows from differentiating the infinite series term by term, so

$$\frac{d}{dt} (e^{t \underline{A}} \underline{x}_0) = \underline{A} (e^{t \underline{A}} \underline{x}_0)$$

shows that this is indeed a solution of the matrix differential equation directly.

Exercise 5.3.1.

matrix exponential chain rule

Verify this “chain rule” derivative property of the matrix exponential using its series representation. ■

For this particular matrix that we started with, the eigenvalues and eigenvectors are complex, but the final result is real. Using the property

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^2 = -\underline{I}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^3 = -\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

one can evaluate the matrix exponential by separating it into its even and odd terms

$$\begin{aligned} e^{t \underline{A}} &= \sum_{k=0}^{\infty} \frac{(t \underline{A})^k}{k!} = \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} \underline{A}^{2k} + \sum_{k=0}^{\infty} \frac{t^{2k+1}}{(2k+1)!} \underline{A}^{2k+1} \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k}}{(2k)!} \underline{I} + \sum_{k=0}^{\infty} (-1)^{2k+1} \frac{t^{2k+1}}{(2k+1)!} \underline{A} = \cos t \underline{I} + \sin t \underline{A}. \end{aligned}$$

Suppose we introduce the function $f^i(\underline{x}_0, t)$ for the solution of this initial value problem for any matrix \underline{A} . Then in this case

$$\begin{pmatrix} f^1(\underline{x}, t) \\ f^2(\underline{x}, t) \end{pmatrix} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} x^1 \cos t - x^2 \sin t \\ x^1 \sin t + x^2 \cos t \end{pmatrix}$$

which corresponds to a counterclockwise rotation of the plane by an angle t . Notice that the radius of the circle through a point (x_0^1, x_0^2) is just $r_0 = ((x_0^1)^2 + (x_0^2)^2)^{1/2}$, which is the magnitude of the vector field $x^1 \partial / \partial x^2 - x^2 \partial / \partial x^1$ on this circle, the vector field for which the circle is a flow line.

In general the map $x^i \rightarrow f^i(\underline{x}, t)$ is a point transformation of the plane into itself in which each point “flows” along a solution curve a parameter interval t from initial point with coordinates x^i to final point with coordinates $f^i(\underline{x}, t)$. These transformations form a 1-parameter group of transformations of the plane into itself, referred to as the flow of the vector field. In this case the flow is the group of rotations of the plane about the origin, each point moving around a circle centered at the origin, with the origin itself a fixed point.

One may directly solve the flow line differential equations formally by expressing the function f as a power series in t , using the chain rule to differentiate functions along the flow line

$$\frac{dx^i(t)}{dt} = \xi^i(x(t)) = (\xi x^i)|_{\underline{x}(t)}, \quad \frac{d^2 x^i(t)}{dt^2} = \frac{dx^k(t)}{dt} \frac{\partial}{\partial x^k} (\xi x^i)|_{\underline{x}(t)} = (\xi^2 x^i)|_{\underline{x}(t)}, \dots$$

so that

$$(x^i)^{(k)}(0) = \left. \frac{d^k x^i(t)}{dt^k} \right|_{t=0} = \xi^k x^i$$

and hence the exponential series representation of the solution is

$$x^i(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} (x^i)^{(k)}(0) = \sum_{k=0}^{\infty} \frac{t^k}{k!} (\xi^k x^i)|_{t=0} = \left(\sum_{k=0}^{\infty} \frac{t^k}{k!} \xi^k \right) x^i = e^{t\xi} x^i = f^i(\underline{x}, t).$$

Thus the exponential of the first order differential operator corresponding to the multivariable calculus vector field, when acting on the coordinate functions, produces the coordinates of a new point a unit parameter interval along the flow line through an initial point with the given coordinates. An immediate consequence of this is the formula for the derivative of a scalar along the flow lines

$$\left. \frac{d}{dt} \right|_{t=0} F(x(t)) = \frac{\partial F}{\partial x^i}(x) \left. \frac{dx^i}{dt}(t) \right|_{t=0} = \frac{\partial F}{\partial x^i}(x) \frac{dx^i}{dt}(0) = \xi^i(x) \frac{\partial F}{\partial x^i}(x) = (\xi F)(x).$$

The coordinate vector fields $\partial / \partial x^i$ themselves generate translations along the coordinate lines since the infinite series terminates at the first power term

$$e^{t\partial / \partial x^j} x^i = x^i + t\delta_j^i.$$

The corresponding system of differential equations can be trivially integrated to produce this result directly.

Exercise 5.3.2.**hyperbolic rotations via matrix exponential**

a) Repeat the above analysis for the matrix and associated linear vector field

$$\underline{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \xi = x^1 \partial / \partial x^2 + x^2 \partial / \partial x^1,$$

but first find the eigenvalues $\lambda = \lambda_1, \lambda_2$ and corresponding matrix of eigenvectors $\underline{B} = \langle b_1 | b_2 \rangle$ and evaluate directly the matrix product $\underline{B} e^{t \underline{A} \underline{B}} \underline{B}^{-1}$. Then evaluate the exponential $e^{t \underline{A}}$ using the corresponding properties of the powers of \underline{A} to get the same result. In each case rewrite your expressions in terms of the hyperbolic cosine and sine

$$\cosh t = \frac{e^t + e^{-t}}{2}, \quad \sinh t = \frac{e^t - e^{-t}}{2}.$$

One could repeat this same calculation for the previous rotation case in terms of complex exponentials, in terms of which one would need the identities

$$\cos t = \frac{e^{it} + e^{-it}}{2}, \quad \sin t = \frac{e^{it} - e^{-it}}{2i}.$$

b) Use technology to view the direction field of this DE system on the window $x^1 \in [-3, 3]$, $x^2 \in [-3, 3]$ and include solution curves for the eight initial data points

$$\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

In this case the flow lines are hyperbolas centered at the origin, except for the degenerate cases along the eigenvector directions $x^2 = \pm x^1$. This hyperbolic analog of the circular geometry of trigonometry is just the mathematics of special relativity in one space and one time dimension, explored in Appendices A.1 and A.2.

■

Exercise 5.3.3.**space rotations via the matrix exponential**

a) Use a computer algebra system to find the integral curves of the linear vector fields introduced at the beginning of this Section 5.2, i.e., to solve the first order system of differential equations associated with the vector fields generating the rotations of the plane about the origin, imposing the generic initial conditions $x^1(0) = x_0^1, x^2(0) = x_0^2$.

b) Since these are linear vector fields, they generate linear transformations, so one can use the matrix exponential to find the general solution of the same initial value problem. Do this and compare with your previous result. Both of these calculations can be done by hand as illustrated in this section, but if the goal is to obtain the results for other purposes, there is no need to waste the time.

c) Since we like math, maybe we want to waste some time for fun, provided we can learn something new from the experience. To evaluate the matrix exponential by hand we need to be able to express \underline{A}^2 and \underline{A}^3 in terms of \underline{I} and \underline{A} and in fact once we do the first task, the second follows for free by multiplying through the previous relation and substituting for \underline{A}^2 . Can you solve the equation $\underline{A}^2 = a\underline{I} + b\underline{A}$ for the unknown coefficients a, b , i.e., $\underline{A}^2 - a\underline{I} - b\underline{A} = 0$? Compare your resulting quadratic equation satisfied by the matrix \underline{A} with the characteristic equation satisfied by its eigenvalues

$$\begin{aligned} 0 &= (\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2 \\ &= \lambda^2 - (\text{Tr } \underline{A})\lambda + \det(\underline{A}). \end{aligned}$$

They are the same! This result, that the matrix satisfies its characteristic equation using the identity matrix in its constant term (called the Cayley-Hamilton theorem), is therefore crucial in reducing the infinite series of the matrix exponential function to a more manageable linear combination of \underline{I} and \underline{A} whose coefficients are scalar infinite series, as in our explicit example in this section. ■

Remark.

It is easy to establish this result by first showing that the diagonalized matrix $\underline{A}_B = \underline{B}^{-1}\underline{A}\underline{B}$ satisfies the equation $\text{CharPoly}(\underline{A}_B) = 0$ (easy since each diagonal entry satisfies it separately as an eigenvalue) and then “conjugating” this matrix equation by \underline{B}

$$\underline{B} \text{CharPoly}(\underline{A}_B) \underline{B}^{-1} = 0,$$

and using the fact that $\underline{A} = \underline{B}^{-1}\underline{A}_B\underline{B}$, $\underline{A}^2 = \underline{B}^{-1}\underline{A}_B^2\underline{B}$, etc. to convert it to the characteristic equation for \underline{A} . ▲

Exercise 5.3.4.

Cayley-Hamilton theorem for $n = 3$

For 3×3 matrices expanding the characteristic equation in terms of its eigenvalue roots one finds

$$\begin{aligned} 0 &= (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3) = \lambda^3 - (\lambda_1 + \lambda_2 + \lambda_3)\lambda^2 + (\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1)\lambda - (\lambda_1\lambda_2\lambda_3) \\ &= \lambda^3 - (\text{Tr } \underline{A})\lambda^2 + \frac{1}{2}(\text{Tr}^2 \underline{A} - \text{Tr } \underline{A}^2)\lambda - \det(\underline{A}). \end{aligned}$$

a) Convince yourself that the first power coefficient in this third degree polynomial is correctly expressed in terms of the original matrix.

b) Recall Exercise 4.5.2, where a tracefree matrix \underline{N} with zero self-inner product under the trace inner product $\text{Tr } \underline{N}^2 = 0$ and zero determinant was seen to generate null rotations

of 3-dimensional Minkowski space. This formula shows why it must be a so called nilpotent matrix—one for which a positive integer power equals the zero matrix—in this case $\underline{N}^3 = \underline{0}$. What extra condition is required for 4×4 matrices to have $\underline{N}^4 = \underline{0}$? It is a fun challenge to express the new first power coefficient Q in the characteristic equation in terms of the trace of the first three powers of the matrix

$$\lambda^4 - (\text{Tr } \underline{A})\lambda^3 + \frac{1}{2}(\text{Tr}^2 \underline{A} - \text{Tr } \underline{A}^2)\lambda^2 + Q\lambda + \det(\underline{A}).$$

■

Remark.

A direct sum decomposition of a vector space into two subspaces $V = V_1 \oplus V_2$ was introduced near the end of Section 1.7. Recall that this just means that every vector in V can be decomposed uniquely into a sum of two vectors, one in each subspace: $v = v_1 + v_2, v_1 \in V_1, v_2 \in V_2$. If there is an inner product on V , there is an orthogonal direct sum if the vectors in each subspace are mutually orthogonal $G(v_1, v_2) = 0$. Any such a decomposition is accompanied by two linear projection maps: $P_1(v_1 + v_2) = v_1, P_2(v_1 + v_2) = v_2$, which then satisfy

$$P_1^2 = P_1, P_2^2 = P_2, P_1 P_2 = P_2 P_1 = 0.$$

This direct sum structure can be extended to any number of summand subspaces in an obvious way. When an inner product is available, one can take an orthonormal basis and express vectors in terms of this basis, and any partition of the terms in that linear combination leads to a corresponding orthogonal direct sum of the vector space. For example, the vector space of $n \times n$ matrices $gl(n, \mathbb{R}^n)$ decomposes into an orthogonal direct sum of the subspace of multiples of the identity matrix, the subspace of antisymmetric matrices, and the subspace of symmetric tracefree matrices as explored in Exercise 1.6.9, with respect to either of the two natural trace inner products on that space. For example on \mathbb{R}^3 , any 3×3 matrix can be represented in terms of the following orthonormal basis with respect to either trace inner product

$$\underline{A} = \underbrace{A^0 \underline{I}/\sqrt{3}}_{\in \text{Trace}(3)} + \underbrace{C^6 \text{diag}(1, 1, -2)/\sqrt{6} + C^7 \text{diag}(1, -1, 0)/\sqrt{2} + |\epsilon^{ijk}| C^i e^j_k / \sqrt{2}}_{\in \text{SymTraceFree}(3)} + \underbrace{\epsilon^{ijk} B^i e^j_k / \sqrt{2}}_{\in \text{ASym}(3)}.$$

Thus one has an orthogonal direct sum into 3 subspaces of dimensions 1, 5, and 3 respectively: the pure trace matrices $\text{Trace}(3)$ which are multiples of the identity, the symmetric tracefree matrices $\text{SymTracefree}(3)$, and the antisymmetric (tracefree) matrices $\text{ASym}(3)$.

The next exercise deals with such an orthogonal direct sum. ▲

Exercise 5.3.5.

space rotations as solutions of a system of differential equations

Consider the system of constant coefficient differential equations

$$\frac{dx^i(t)}{dt} = \xi^i(\underline{x}(t)), \quad \xi^i = \Omega^i_j x^j, \quad \Omega^i_j = \delta^{im} \epsilon_{mjk} \omega^k, \quad \langle \omega^1, \omega^2, \omega^3 \rangle = \langle 1, 1, 1 \rangle,$$

or explicitly

$$\frac{dx^1(t)}{dt} = -x^2(t) + x^3(t), \quad \frac{dx^2(t)}{dt} = x^1(t) - x^3(t), \quad \frac{dx^3(t)}{dt} = -x^1(t) + x^2(t),$$

with initial conditions

$$x^1(0) = x_0^1, \quad x^2(0) = x_0^2, \quad x^3(0) = x_0^3.$$

Note that the coefficient matrix $\underline{\Omega}$ is antisymmetric and so determines a 2-vector $\Omega = \Omega^{ij} e_i \wedge e_j$ which spans a plane whose normal is the dual $\omega = *\Omega = \omega^i e_i$ of this 2-vector, both of which have magnitude $|\Omega| = |\omega| = (\delta_{ij} \omega^i \omega^j)^{1/2} = \sqrt{3}$, and hence $n = \hat{\omega} = \omega/|\omega| = \langle 1, 1, 1 \rangle / \sqrt{3}$ is a unit normal to this plane. Let $\hat{\Omega} = \Omega/|\Omega|$. Then by definition $\underline{\Omega} \underline{x}$ corresponds to $\omega \times x$ and $\hat{\Omega} \underline{x}$ corresponds to $n \times x$. [By the magnitude of the antisymmetric matrix, we mean the one from the self-inner product without overcounting: $\Omega^{[ij]} \Omega_{ij} = \frac{1}{2} \text{Tr } \underline{\Omega}^T \underline{\Omega}$.]

a) Use technology to verify that $\det \underline{\Omega} = 0$, which means that 0 is an eigenvalue of the matrix corresponding to a direction along which the matrix product yields zero, and hence equilibrium points of the system of differential equations (constant value solutions for the three variables of the system), that is, fixed points of the flow of the corresponding vector field ξ . Verify that $*\Omega$ is an eigenvector with this eigenvalue.

b) Use technology to solve the initial value problem. The result looks pretty messy, no? How do we make sense of it? Keep reading.

c) Use technology to find the eigenvalues and eigenvectors of the matrix $\underline{\Omega}$: $\lambda_1 = |\omega|i, \lambda_2 = -|\omega|i, \lambda_3 = 0$, $\underline{B} = \langle b_1 | b_2 | b_3 \rangle$. Verify that $\text{Re } b_1 \times \text{Im } b_1 \propto \omega$, so the real plane determined by the complex conjugate pair of eigenvectors through their real and imaginary parts coincides with the plane determined by the 2-vector Ω or its normal ω . Define a new real basis of this plane by taking the real and imaginary parts of the complex conjugate of the eigenvector b_1 corresponding to the positive imaginary eigenvalue, and divide each one by its length, and complete it to a basis of the whole space by adding the normalized third eigenvector

$$r_1 = \text{Re } b_1 / |\text{Re } b_1|, \quad r_2 = -\text{Im } b_1 / |\text{Im } b_1|, \quad r_3 = b_3 / |b_3|,$$

Define the matrix $\underline{R} = \langle r_1 | r_2 | r_3 \rangle$. Verify that this is an orthogonal matrix: $\underline{R}^T \underline{R} = \underline{I}$ and that $\det \underline{R} = 1$, the reason for using the complex conjugate. This latter condition guarantees that this basis is right-handed: $r_1 \times r_2 = r_3$ or equivalently $(r_1 \times r_2) \cdot r_3 = 1$. Check it. Use technology to evaluate the new matrix of $\underline{\Omega}_R = \underline{R}^{-1} \underline{\Omega} \underline{R}$ with respect to this basis.

d) Use technology to obtain the characteristic polynomial of $\underline{\Omega}$ whose roots are the eigenvalues and verify that it has the form: $\lambda(\lambda^2 + |\omega|^2) = 0$. This implies that the matrix itself satisfies $\underline{\Omega}^3 = -|\omega|^2 \underline{\Omega}$ whose iteration ($\underline{\Omega}^4 = -|\omega|^2 \underline{\Omega}^2$, etc.) allows us to express every power of $\underline{\Omega}$ from the third on up in terms of $\underline{\Omega}$ and $\underline{\Omega}^2$ so that the matrix exponential power series can be expressed in terms of a linear combination of these two matrices alone (with coefficients

which are scalar power series) together with the identity matrix (the first term in the series). Show that the higher order terms can be grouped as follows

$$\begin{aligned}
 e^{t\hat{\Omega}} &= \underline{I} + \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} t^{2k+1} \hat{\Omega}^{2k+1} + \sum_{k=1}^{\infty} \frac{1}{(2k)!} t^{2k} \hat{\Omega}^{2k} \\
 &= \underline{I} + \left(t - \frac{t^3}{3!} |\omega|^2 + \dots \right) \hat{\Omega} + \left(\frac{t^2}{2!} - \frac{t^4}{4!} |\omega|^2 + \dots \right) \hat{\Omega}^2 + \dots \\
 &= \underline{I} + \sin(|\omega|t) \frac{\hat{\Omega}}{|\omega|} + (1 - \cos(|\omega|t)) \frac{\hat{\Omega}^2}{|\omega|^2} \\
 &= \left(\underline{I} + \hat{\Omega}^2 \right) + \sin(|\omega|t) \hat{\Omega} + \cos(|\omega|t) \left(-\hat{\Omega}^2 \right).
 \end{aligned}$$

This represents a rotation by an angle $\theta = |\omega|t$ in the plane of the 2-vector $\hat{\Omega}$ with angular velocity $d\theta/dt = |\omega|$ (namely $\sqrt{3}$ in our explicit example) if t is interpreted as the classical time. This is said to be a rotation “about the axis” n , which is the unit normal to this plane, fixed under all these rotations. The angle of the rotation is increasing in the counterclockwise direction in this plane as oriented by the right hand rule using n : with the thumb along n , the fingers curl in the direction of the time-dependent rotation.

e) The matrix $\underline{P}_{\perp} = -\hat{\Omega}^2$ is the projection onto this two plane and acts as the identity on vectors which lie in this plane, while the linear transformation $\underline{x} \rightarrow \hat{\Omega} \underline{x} = \underline{n} \times \underline{x}$ acts as a 90 degree counterclockwise rotation in this plane. Check these statements by multiplying r_1 , r_2 and r_3 by the matrix $\hat{\Omega}$. The matrix $\underline{P}_{\parallel} = \underline{I} + \hat{\Omega}^2$ projects along the normal direction, where it acts as the identity. Check this by evaluating this matrix explicitly and using it to multiply r_1 , r_2 and r_3 , verifying these statements. The pair of matrices $\underline{P}_{\perp}, \underline{P}_{\parallel}$ together determine an orthogonal decomposition of $\mathbb{R}^3 = (\mathbb{R}^3)_{\parallel} \oplus (\mathbb{R}^3)_{\perp}$ into an orthogonal direct sum of two subspaces, which are the eigenspaces of dimension 2 and 1 of the matrix $\hat{\Omega}$.

Show that

$$\underline{P}_{\perp}^2 = \underline{P}_{\perp}, \quad \underline{P}_{\parallel}^2 = \underline{P}_{\parallel}, \quad \underline{P}_{\perp} \underline{P}_{\parallel} = \underline{0} = \underline{P}_{\parallel} \underline{P}_{\perp}$$

and

$$\underline{I} = \underline{P}_{\perp} + \underline{P}_{\parallel}.$$

Then show that

$$\underline{P}_{\perp} = r_1 r_1^T + r_2 r_2^T, \quad \underline{P}_{\parallel} = r_3 r_3^T$$

The fact that the sum of these two matrices is the identity merely reflects the component expression of the identity tensor in terms of an orthonormal basis $Id = r_1 \otimes r_1^b + r_2 \otimes r_2^b + r_3 \otimes r_3^b$ as the sum of 1-dimensional mutually orthogonal projections along the orthonormal basis vectors.

f) The rotation of the standard basis e_i by the orthogonal matrix \underline{R} adapts the coordinates of \mathbb{R}^3 to this rotation by sending the standard basis to the basis \underline{r}_i . If we introduce new coordinates $\underline{x} = \underline{R}^{-1} \underline{y}$, then the transformed matrix of the differential equation system

$$\underline{R} \hat{\Omega} \underline{R}^{-1} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Use technology to verify this. In terms of these new coordinates the family of rotations takes the simple form

$$\begin{pmatrix} y^1(t) \\ y^2(t) \\ y^3(t) \end{pmatrix} = \begin{pmatrix} \cos(|\omega|t) & -\sin(|\omega|t) & 0 \\ \sin(|\omega|t) & \cos(|\omega|t) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_0^1 \\ y_0^2 \\ y_0^3 \end{pmatrix}.$$

Thus the dual ω of a 2-vector Ω represented as a linear transformation $\Omega x = \omega \times x$ through the cross product with the dual vector is something that is used in the first physics course when describing a body rotating about an axis with a given angular velocity. The magnitude of this 2-vector Ω is the scalar angular velocity $|\omega|$, while its dual is the vector angular velocity ω , and the 2-plane it determines is the plane of the rotation, whose normal is the direction unit vector $n = \hat{\omega}$ associated with the angular velocity vector. ■

Remark.

Given any orthonormal basis $e_{i'} = e_j B^j_{i'}$ of \mathbb{R}^n , where e_i is the standard basis, the identity tensor partially evaluated on a vector decomposes it into a sum of vector components along each 1-dimensional vector subspace spanned by the individual basis vectors. This in turn can be re-expressed in terms of the dot product

$$\text{Id}(\cdot, X) = \delta^j_{i'} e_{j'} \otimes \omega^{i'}(\cdot, X) = \omega^{j'}(X) e_{j'} = G^{i'j'}(e_{i'} \cdot X) e_{j'} = \sum_{j=1}^n \frac{(e_{j'} \cdot X) e_{j'}}{G_{j'j'}},$$

where of course $G_{j'j'} = \delta_{jj}$. However, for any signature metric G with $G_{ii} = \pm 1$, this latter formula continues to hold, so that the components along basis vectors with $G_{ii} = -1$ have an extra minus sign to reverse the sign of the inner product in its term in the projection. Thus in Minkowski spacetimes, the vector projection of a vector X along a timelike unit vector u satisfying $G(u, u) = -1$ is $-G(X, u)u$, or in components $X_{||}^i = -X^j u_j u^i \equiv T(u)^i_j X^j$. Then $P(u) = Id - T(u)$ projects to the hyperplane orthogonal to u . ▲

Exercise 5.3.6.

Local rest space decomposition in \mathbb{M}^4

On Minkowski spacetime \mathbb{M}^4 any unit timelike vector u (namely $G(u, u) = u_\alpha u^\alpha = -1$) decomposes the vector space into the 1-dimensional subspace of its own multiples and the 3-dimensional orthogonal subspace $LR S_u$ of spacelike vectors called the local rest space associated with the observer whose 4-velocity is u .

a) If one completes $u = E_0$ to an orthonormal frame E_α , $\alpha = 0, 1, 2, 3$ then any vector can be expressed as $X^\alpha e_\alpha = Y^0 E_0 + Y^i E_i$, $i = 1, 2, 3$, where e_α is the standard basis. Then $P_{||}(X) = Y^0 E_0$, where $Y^0 = -u_\alpha X^\alpha$, and $P_\perp(X) = Y^i E_i = X - P_{||}(X)$ defines the two projection maps for this decomposition. Show that the tensors which represent these projections

are $P_{\parallel}^{\alpha}{}_{\beta} = -u^{\alpha}u_{\beta}$ and $P_{\perp}(X)^{\alpha}{}_{\beta} = \delta^{\alpha}{}_{\beta} + u^{\alpha}u_{\beta}$ and that they satisfy the projection relations (multiplication of linear maps means contraction of adjacent indices as in matrix multiplication)

$$P_{\parallel}^2 = P_{\parallel}, \quad P_{\perp}^2 = P_{\perp}, \quad P_{\parallel}P_{\perp} = 0.$$

b) Suppose $u = u^0 e_0 + u^i e_i$ is a timelike future pointing ($u^0 > 0$) vector in the standard orthonormal basis of \mathbb{M}^4 , namely the 4-velocity of some observer in relative motion with the observer associated with these inertial coordinates (having 4-velocity e_0). The 3 vectors e_i span the local rest space $LR S_{e_0}$. The unit condition is $-(u^0)^2 + \delta_{ij}u^i u^j = -1$ is identically satisfied by setting $u^0 = \cosh \beta = \gamma$ and $\sinh \beta = \sqrt{\delta_{ij}u^i u^j}$, with $\beta \geq 0$. Then define the 3-velocity $v^i = u^i/u^0$ and speed $|v| = \sqrt{\delta_{ij}v^i v^j} = \tanh \beta$ and let $\hat{v}^i = v^i/|v|$ be its unit vector so that the 4-velocity has the representation

$$u = \cosh \beta e_0 + \sinh \beta \hat{v}^i e_i = \cosh \beta (e_0 + \tanh \beta \hat{v}^i e_i) = \gamma(e_0 + v^i e_i)$$

Show that the gamma factor is related to the spatial speed by $\gamma = (1 - |v|^2)^{-1/2}$.

c) While this exercise can be thought of as living on the vector space \mathbb{M}^4 involving its points thought of as displacement vectors from the origin, once we start considering world lines in that space it is more appropriate to consider it in the context of each tangent space along the world line, as in the examples in Appendix C. In this case u would be the future-pointing timelike unit tangent to the world line called its 4-velocity, and the local rest space would be a subspace of the local tangent space, which of course can also be considered a hyperplane in \mathbb{R}^4 itself.

For a test particle of rest mass m , the 4-momentum is defined as

$$p = mu = m\gamma(u + v^i e_i) \equiv Eu + p^i e_i,$$

thus defining the energy $E = m\gamma$ as the timelike component and the spatial momentum $\vec{p} = p^i e_i = m\gamma \vec{v}$ as the orthogonal component with respect to the inertial observer attached to the inertial coordinates (just the ordinary Cartesian coordinates of the standard basis if \mathbb{R}^4 interpreted as \mathbb{M}^4 with the Lorentz inner product. The relativistic equations of motion for a charged particle with charge q in an electromagnetic field F and 4-velocity $u^{\alpha} = dx^{\alpha}/d\tau$ along such a world line parametrized by its proper time τ are then

$$\frac{dp^{\alpha}}{d\tau} = qF^{\alpha}{}_{\beta}u^{\beta}.$$

Since u is a unit vector, it can only change by an orthogonal transformation (Lorentz transformation, which includes rotations), and thus its rate of change along its world line has to be generated by an antisymmetric matrix, which is the role played by the electromagnetic field.

Show that using the chain rule relation $dt/d\tau = u^0 = \gamma$ and the decomposition of the electromagnetic field given in Exercise 1.6.6, the components of this equation can be written (don't confuse the energy scalar E with the electric field $\vec{E} = E^i e_i$)

$$\begin{aligned} \frac{dE}{dt} &= E_i v^i, \\ \frac{dp^i}{dt} &= q(\vec{E} + \vec{v} \times \vec{B})^i. \end{aligned}$$

**Exercise 5.3.7.****logarithmic spiral group**

The 2-parameter family of logarithmic spiral curves satisfies the equation $r = r_0 e^{k\theta}$ in the plane in polar coordinates (r, θ) , so called since the natural log of the radial coordinate $\ln r = \ln r_0 + k\theta$ is a linear function of the polar angle. For a given k value, this is an integral curve (flow line) of a vector field which has a simple expression in polar coordinates. First choosing to parametrize it by $t = \theta - \theta_0$, we get

$$\vec{r}(t) = \langle r_0 e^{kt}, t + \theta_0 \rangle, \quad \vec{r}(0) = \langle r_0, \theta_0 \rangle, \quad \vec{r}'(t) = \langle kr_0 e^{kt}, 1 \rangle = \langle kr, 1 \rangle \equiv \langle \xi^r, \xi^\theta \rangle.$$

Define the associated vector field $\xi = kr \partial_r + \partial_\theta$. By definition the flow lines of this vector field starting at (r_0, θ_0) at $t = 0$ are the members of the logarithmic spiral family for that fixed k value. The origin is a fixed point of this vector field where it vanishes, so the flow of the vector field only acts on the plane excluding the origin.

a) We will study polar coordinates in detail soon, but use coordinate transformation relations the chain rule

$$\begin{aligned} x &= r \cos \theta, \quad y = r \sin \theta, \\ \frac{\partial}{\partial r} &= \frac{\partial x}{\partial r} \frac{\partial}{\partial x} + \frac{\partial y}{\partial r} \frac{\partial}{\partial y}, \quad \frac{\partial}{\partial \theta} = \frac{\partial x}{\partial \theta} \frac{\partial}{\partial x} + \frac{\partial y}{\partial \theta} \frac{\partial}{\partial y}, \end{aligned}$$

to show that in Cartesian coordinates one has

$$\xi = (kx - y)\partial_x + (x + ky)\partial_y.$$

b) This is a linear vector field. What is its matrix \underline{K} ? Use a computer algebra system to evaluate $e^{t\underline{K}}$. If you consider the characteristic equation satisfied by the matrix itself, one can repeat the steps taken with the rotations of the plane to sum the exponential power series for this matrix, but it is a bit tedious, no, complicated by the parameter k . In fact one can separate \underline{K} into a multiple of the identity matrix which generates a radial scaling of the points and a simple rotation generator which commutes with the identity matrix multiples, so that one can factor the exponential into the product of the two separate exponentials, which is easy. This transformation transforms figures into the plane into similar figures of the same shape but different scale. These are called conformal transformations of the Euclidean plane.



5.4 Frames and dual frames and Lie brackets

As already mentioned above, a smooth choice of basis for the tangent spaces to \mathbb{R}^n is called a frame, and consists of n vector fields whose values are n linearly independent tangent vectors of each point of \mathbb{R}^n . The corresponding choice of dual basis is called the dual frame. $\{\partial/\partial x^i\}$ is such a frame, usually called a coordinate frame since the individual frame vector fields are just partial derivatives with respect to the coordinates, and $\{dx^i\}$ is its dual frame.

All the linear algebra we developed for a single vector space we can apply to each tangent space to \mathbb{R}^n independently, although we must assume that what we do at different tangent spaces is a continuous or even differentiable function of position.

For example, we can change the frame, i.e., perform a change of basis on each tangent space in a continuous or differentiable fashion

$$E_i = A^{-1j}_i \frac{\partial}{\partial x^j}, \quad W^i = A^i_j dx^j, \quad (\text{i.e., } E^j_i = A^{-1j}_i, \quad W^i_j = A^i_j),$$

where now \underline{A} is a matrix-valued function on \mathbb{R}^n , with everywhere nonzero determinant of course so that the frame vectors are linearly independent. The components of tensor fields will change according to the same formulas as before except that now both the components of the tensors and the matrix of the transformation are functions on \mathbb{R}^n . The special case of constant \underline{A} describes the change to a new frame which is the coordinate frame associated with the new Cartesian coordinates $x^{i'} = A^i_j x^j$ so that $E_i = \partial/\partial x^{i'}$. A more general special case corresponds to the change to a frame associated with a non-Cartesian coordinate system, but the most general case cannot be associated with any coordinate system as we will see.

Suppose $\{x^{i'}\}$ are n functions on \mathbb{R}^n such that the matrix $A^i_j = \partial x^{i'}/\partial x^j$ of partial derivatives has nonzero determinant at each point of \mathbb{R}^n . Then the chain rule says that $E_i = \partial/\partial x^{i'}$ are partial derivatives with respect to the new coordinates.

Thus we have Cartesian coordinate frames, non-Cartesian coordinate frames, and “non-coordinate” frames, namely frames for which no system of coordinates can be found so that the frame vector fields can be represented as coordinate derivatives. There is a simple way to tell whether a frame is noncoordinate or not. We all know that partial derivatives commute, i.e., as long as a function f is well behaved, the order of the partial derivatives does not matter

$$\frac{\partial}{\partial x^i} \frac{\partial}{\partial x^j} f = \frac{\partial}{\partial x^j} \frac{\partial}{\partial x^i} f$$

or

$$\left(\frac{\partial}{\partial x^i} \frac{\partial}{\partial x^j} - \frac{\partial}{\partial x^j} \frac{\partial}{\partial x^i} \right) f = 0 \text{ for all such } f$$

or

$$\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right] \equiv \frac{\partial}{\partial x^i} \frac{\partial}{\partial x^j} - \frac{\partial}{\partial x^j} \frac{\partial}{\partial x^i} = 0$$

when acting on such well behaved functions.

For any operators A and B , their commutator is defined by

$$[A, B] = AB - BA$$

and when it vanishes, their order doesn't matter: $AB = BA$, and they are said to "commute." Any commutator has the obvious properties

$$[A, A] = AA - AA = 0, \quad [B, A] = BA - AB = -(AB - BA) = -[A, B].$$

Note also that this commutator $[A, B]$ is bilinear in its two inputs or arguments A and B , namely for any two constants c_1, c_2 one has a distributive law in each argument, for example

$$\begin{aligned} [c_1A + c_2B, C] &= (c_1A + c_2B)C - C(c_1A + c_2B) = (c_1AC + c_2BC) - (c_1CA + c_2CB) \\ &= c_1(AC - CA) + c_2(BC - CB) = c_1[A, C] + c_2[B, C]. \end{aligned}$$

Another very useful commutator identity is the Jacobi identity which results from expanding a cyclic combination of three double commutators into twelve terms which all cancel in pairs

$$\begin{aligned} & [[A, B], C] + [[B, C], A] + [[C, A], B] \\ &= \begin{aligned} & (AB - BA)C - C(AB - BA) & (ABC - BAC - CAB + CBA) \\ & + (BC - CB)A - A(BC - CB) & + (BCA - CBA - ABC + ACB) \\ & + (CA - AC)B - B(CA - AC) & + (CAB - ACB - BCA + BAC) \end{aligned} = 0. \end{aligned}$$

Define the commutator of any two vector fields u and v by the same formula $[u, v] = uv - vu$. This is a differential operator on functions. What is it? We can express it in components, when acting on a well-behaved (i.e., differentiable) function

$$\begin{aligned} [u, v]f &= (uv - vu)f = uvf - vuf = u^i \frac{\partial}{\partial x^i} \left(v^j \frac{\partial f}{\partial x^j} \right) - v^i \frac{\partial}{\partial x^i} \left(u^j \frac{\partial f}{\partial x^j} \right) \\ &= u^i \frac{\partial v^j}{\partial x^i} \frac{\partial f}{\partial x^j} + u^i v^j \frac{\partial^2 f}{\partial x^i \partial x^j} - v^i \frac{\partial u^j}{\partial x^i} \frac{\partial f}{\partial x^j} - v^i u^j \frac{\partial^2 f}{\partial x^i \partial x^j} \\ &= \underbrace{\left(u^i \frac{\partial v^j}{\partial x^i} - v^i \frac{\partial u^j}{\partial x^i} \right) \frac{\partial f}{\partial x^j}}_{\text{new vector field acting on } f} + \underbrace{u^i v^j \left(\frac{\partial^2 f}{\partial x^i \partial x^j} - \frac{\partial^2 f}{\partial x^j \partial x^i} \right)}_{0 \text{ for well-behaved } f}. \end{aligned}$$

The final formula defines a new vector field $[u, v]$ acting on f whose components are

$$[u, v]^j = u^i \frac{\partial v^j}{\partial x^i} - v^i \frac{\partial u^j}{\partial x^i} = uv^j - vu^j$$

called the Lie bracket of u and v .

Thus if $E_{i'} = A^i_{i'} \partial / \partial x^i$ can be represented as coordinate derivatives for some coordinate system: $E_{i'} = \partial / \partial x^{i'}$, then

$$[E_{i'}, E_{j'}] = \left[\frac{\partial}{\partial x^{i'}}, \frac{\partial}{\partial x^{j'}} \right] = 0$$

since partial derivatives commute. A necessary condition for this therefore is the vanishing of the Lie brackets of all pairs of distinct frame vector fields.

When the frame vectors do not commute, then one can express them in the same frame, leading to the definition of the component functions of those frame vectors

$$[E_{i'}, E_{j'}] = C^{k'}_{i'j'} E_{k'}, \quad C^{k'}_{j'i'} = -C^{k'}_{i'j'}.$$

These “structure functions of the frame” determine certain geometrical properties of the frame. They are the components of a $\binom{1}{2}$ -tensor but one which is frame-dependent, like the Levi-Civita symbols. All coordinate frames have zero structure functions.

Exercise 5.4.1.

Lie bracket evaluation

Compute the nonzero Lie brackets among the following sets of vector fields on \mathbb{R}^2

- a) $X_1 = \partial_1, X_2 = \partial_2, X_3 = x^1 \partial_2 - x^2 \partial_1,$
- b) $X_1 = \partial_1, X_2 = \partial_2, X_3 = x^1 \partial_2 + x^2 \partial_1.$

Note that the flow lines of X_1, X_2 are just the Cartesian coordinate lines, while the final vector field in each set corresponds respectively to rotations about the origin or hyperbolic rotations about the origin, as discussed in Section 5.3. Notice that in this case these Lie brackets are constant linear combinations of the same set of vector fields, i.e., as a set of 3 vector fields, this is a 3-dimensional vector space which is closed under the Lie bracket. Such sets of vector fields are said to define a Lie algebra of vector fields. ■

Exercise 5.4.2.

Lie bracket evaluation

- a) Compute the nonzero Lie brackets among the following vector fields on \mathbb{R}^3

$$u = x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}, \quad v = y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y}, \quad w = (x^2 + y^2) \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} \right) + \frac{\partial}{\partial z},$$

namely, $[u, v]$, $[u, w]$ and $[v, w]$.

- b) On \mathbb{R}^2 do the same for $u = (x^2 + y^2)^{-1/2}(x \partial/\partial x + y \partial/\partial y)$ and $v = -y \partial/\partial x + x \partial/\partial y$. If we introduce a new frame by $E_1 = u, E_2 = v$ then

$$\underline{A}^{-1} = \begin{pmatrix} \frac{x}{\sqrt{x^2 + y^2}} & -y \\ \frac{y}{\sqrt{x^2 + y^2}} & x \end{pmatrix}.$$

What is $\det \underline{A}^{-1}$? Can this vanish? What does this mean? What does your result for $[u, v]$ tell you?

Note that the ordinary vector functions $\vec{u} = (x^2 + y^2)^{-1/2}(x, y) = \vec{r}/\|\vec{r}\| = \hat{r}$, and $\vec{v} = (-y, x)$, which satisfy $\vec{u} \cdot \vec{v} = 0$, are the unit outward radial vector field and the counterclockwise pointing vector field tangent to the circles about the origin illustrated in Fig. 5.6.

■

Example 5.4.1. Consider the product rule

$$\frac{\partial}{\partial x^i} \left(f \frac{\partial}{\partial x^j} \right) = \frac{\partial f}{\partial x^i} \frac{\partial}{\partial x^j} + f \frac{\partial^2}{\partial x^i \partial x^j}$$

and use it to expand the Lie bracket

$$\begin{aligned} \left[xy \frac{\partial}{\partial x}, \sin(x+y) \frac{\partial}{\partial y} \right] &= xy \frac{\partial}{\partial x} \left(\sin(x+y) \frac{\partial}{\partial y} \right) - \sin(x+y) \frac{\partial}{\partial y} \left(xy \frac{\partial}{\partial x} \right) \\ &= xy \cos(x+y) \frac{\partial}{\partial y} + xy \sin(x+y) \frac{\partial^2}{\partial x \partial y} - \sin(x+y) x \frac{\partial}{\partial x} - \sin(x+y) xy \frac{\partial^2}{\partial x \partial y} \\ &= xy \cos(x+y) \frac{\partial}{\partial y} - x \sin(x+y) \frac{\partial}{\partial x}. \end{aligned}$$

So the commutator of the vector fields on \mathbb{R}^2 with the Cartesian coordinate components $(xy, 0)$ and $(0, \sin(x+y))$ has components $(-x \sin(x+y), xy \cos(x+y))$. □

Exercise 5.4.3.

linear vector field Lie brackets

If $\underline{A} = (A^i_j)$ and $\underline{B} = (B^i_j)$ are constant matrices and $\vec{b} = (b^i)$ and $\vec{c} = (c^i)$ are constant vectors in \mathbb{R}^n , define the four vector fields

$$X = A^i_j x^j \frac{\partial}{\partial x^i}, \quad Y = B^m_n x^n \frac{\partial}{\partial x^m}, \quad Z = b^l \frac{\partial}{\partial x^l}, \quad W = c^k \frac{\partial}{\partial x^k}.$$

Evaluate $[X, Y]$ and $[X, Z]$ (using $\partial x^i / \partial x^j = \delta^j_i$) and $[Z, X]$. The result $[X, Y] = -[\underline{A}, \underline{B}]^i_j x^j \partial / \partial x^i$ shows that the commutators of the matrices are directly reflected in the commutators of the corresponding vector fields.

These results allow us evaluate the Lie brackets of any vector fields whose components are linear functions of the coordinates in the nonhomogeneous sense: $A^i_j x^j + b^i$. ■

Exercise 5.4.4.

rotation generator Lie brackets

Define three vector fields on \mathbb{R}^3 of the type discussed in the previous problem by $L_i = \epsilon_{ijk} x^j \partial / \partial x^k$ associated with the basis $[S_k]^i_j = \epsilon_{ikj}$ of antisymmetric matrices

$$\begin{aligned} \underline{S}_1 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \underline{S}_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \underline{S}_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ L_1 &= x^2 \frac{\partial}{\partial x^3} - x^3 \frac{\partial}{\partial x^2}, \quad L_2 = x^3 \frac{\partial}{\partial x^1} - x^1 \frac{\partial}{\partial x^3}, \quad L_3 = x^1 \frac{\partial}{\partial x^2} - x^2 \frac{\partial}{\partial x^1}. \end{aligned}$$

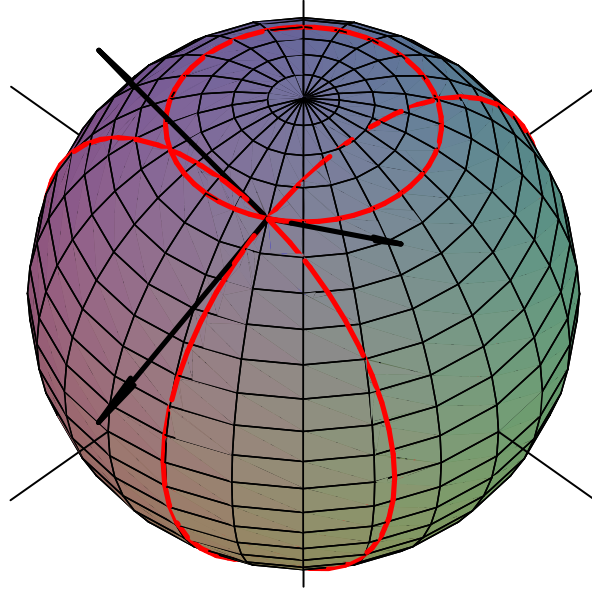


Figure 5.7: The three vector fields L_1, L_2, L_3 are shown at a single point of a sphere of radius r centered at the origin. Their flow lines are circles about the three coordinate axes in \mathbb{R}^3 contained in each such sphere and their magnitudes equal the radii of the corresponding circles.

a) Evaluate directly the commutators $[S_1, S_2]$ and $[L_1, L_2]$ and compare with the previous problem result. Then instead use the epsilon formulas to derive the formulas

$$[S_i, S_j] = \epsilon_{ijk} S_k, \quad [L_i, L_j] = -\epsilon_{ijk} L_k,$$

which are explicitly

$$\begin{aligned} [\underline{S}_2, \underline{S}_3] &= \underline{S}_1, & [\underline{S}_3, \underline{S}_1] &= \underline{S}_2, & [\underline{S}_1, \underline{S}_2] &= \underline{S}_3, \\ [L_2, L_3] &= -L_1, & [L_3, L_1] &= -L_2, & [L_1, L_2] &= -L_3. \end{aligned}$$

b) Show that $\omega^i L_i = [\vec{\omega} \times \vec{x}]^i \partial / \partial x^i$, where the cross product is defined in terms of the usual vectors in \mathbb{R}^3 . Thus the differential equations

$$\frac{d\vec{x}}{dt} = \vec{\omega} \times \vec{x}$$

describing a constant rotation with constant angular velocity $\vec{\omega}$ corresponds to the flow lines of the vector field $\omega^i L_i$. Each vector field L_i generates rotations in the x^j - x^k plane (where i, j, k are distinct) exactly as the 2-dimensional example of the previous section, leaving fixed the x^i -axis.

c) Introduce three constant vector fields $p_i = \partial / \partial x^i$. These are just the usual unit vectors along the three coordinate axes thought of as vector fields. Evaluate $[L_i, p_j]$.

■

Exercise 5.4.5.**Laplacian**

2nd order linear differential operators are also useful. Define the Laplacian

$$\nabla^2 = \vec{\nabla} \cdot \vec{\nabla} = \delta^{ij} \frac{\partial^2}{\partial x^i \partial x^j} = \sum_{i=1}^n \frac{\partial^2}{\partial (x^i)^2}.$$

For $n = 2$ this is

$$\nabla^2 = \frac{\partial^2}{\partial (x^1)^2} + \frac{\partial^2}{\partial (x^2)^2}.$$

Evaluate the commutator $\left[\nabla^2, x^1 \frac{\partial}{\partial x^2} - x^2 \frac{\partial}{\partial x^1} \right]$.

Hint:

$$\frac{\partial^2}{\partial x^2} \left(x \frac{\partial}{\partial y} \right) = \frac{\partial}{\partial x} \left(\frac{\partial}{\partial x} \left(x \frac{\partial}{\partial y} \right) \right) = \frac{\partial}{\partial x} \left(\frac{\partial}{\partial y} + x \frac{\partial^2}{\partial x \partial y} \right)$$

etc. ■

Exercise 5.4.6.**total angular momentum operator and the Laplacian**

a) In traditional notation on \mathbb{R}^3 , the position vector $\vec{r} = \langle x^1, x^2, x^3 \rangle$ has length $r = |\vec{r}| = (\delta_{ij} x^i x^j)^{1/2}$ and the direction unit vector $\hat{r} = \vec{r}/r$. These determine two vector fields whose associated derivatives are indicated by $\vec{r} \cdot \vec{\nabla}$ and $\hat{r} \cdot \vec{\nabla} = D_{\hat{r}}$, the latter being the true directional derivative along the radial direction from the origin, interpreted as the derivative with respect to arclength along the radial line from the origin. In our new notation, these correspond to the vector fields $x^i \partial / \partial x^i$ and $(x^i/r) \partial / \partial x^i \equiv D_r$. Show that $D_r r = 1$ and that $D_r (x^i/r) = 0$.

b) Similarly the three vector fields $\vec{L}_1 = \langle 0, -x^3, x^2 \rangle$, $\vec{L}_2 = \langle x^3, 0, -x^1 \rangle$, $\vec{L}_3 = \langle -x^2, x^1, 0 \rangle$, clearly satisfy $\vec{L}_i \cdot \vec{r} = 0$ by inspection, which is due to the fact that each is tangent to a circle about one of the three axes and hence is tangent to the sphere of radius r at each point which contains these circles, and the radial direction is orthogonal to the tangent plane to the sphere. Show that the corresponding vector field operators $L_i = \epsilon_{ijk} x^j \partial / \partial x^k$ satisfy $L_i r = 0$, which must be the case since these vector fields are tangent to the level surfaces (spheres) of the function r , and that

$$\left[L_i, x^k \frac{\partial}{\partial x^k} \right] = 0, \quad [L_i, D_r] = 0.$$

c) Consider the second order operator in Cartesian coordinates on \mathbb{R}^3

$$\begin{aligned} L^2 &= \delta^{ij} L_i L_j = \delta^{ij} \epsilon_{imn} \epsilon_{jpk} x^m \frac{\partial}{\partial x^n} \left(x^p \frac{\partial}{\partial x^k} \right) = \delta^{pq}_{mn} x^m \left(\delta^p_n \frac{\partial}{\partial x^q} + x^p \frac{\partial^2}{\partial x^n \partial x^q} \right) \\ &= (\delta^p_m \delta^q_n - \delta^q_m \delta^p_n) \left(\delta^p_n \frac{\partial}{\partial x^q} + x^p \frac{\partial^2}{\partial x^n \partial x^q} \right) = \dots \\ &= 2x^i \frac{\partial}{\partial x^i} + (\delta^{ij} x_k x^k - x^i x^j) \frac{\partial^2}{\partial x^i \partial x^j}, \end{aligned}$$

so that upon solving this for the Laplacian $\nabla^2 = \delta^{ij} \partial^2 / \partial x^i \partial x^j$, one finds

$$\nabla^2 = \frac{L^2}{r^2} + \frac{x^i x^j}{r^2} \frac{\partial^2}{\partial x^i \partial x^j} + \frac{2x^k}{r^2} \frac{\partial}{\partial x^k}.$$

Fill in the dots above and then show that the final two terms in this formula can be rewritten in terms of the derivative $D_r \equiv r^{-1} x^i \partial / \partial x^i$ as

$$\frac{D_r(r^2 D_r)}{r^2} = \frac{x^i x^j}{r^2} \frac{\partial^2}{\partial x^i \partial x^j} + \frac{2x^k}{r^2} \frac{\partial}{\partial x^k},$$

so that one obtains the formula

$$\nabla^2 = \frac{L^2}{r^2} + \frac{D_r(r^2 D_r)}{r^2}.$$

d) Use the following product rule for commutators

$$[A, BC] = ABC - BCA = ABC - BAC + BAC - BCA = [A, B]C + B[A, C]$$

and the commutation relations for the L_i from a previous exercise to show that the following commutator vanishes

$$\begin{aligned} [L_3, L^2] &= [L_3, L_1^2 + L_2^2 + L_3^2] = [L_3, L_1^2 + L_2^2] \quad (\text{note: } [A, A^n] = AA^n - A^nA = 0) \\ &= [L_3, L_1]L_1 + L_1[L_3, L_1] \\ &\quad + [L_3, L_2]L_2 + L_2[L_3, L_2] = \dots = 0. \end{aligned}$$

Clearly the latter result holds for all three L_i : $[L_i, L^2] = 0$ because of the symmetry with which they enter the formula.

e) Since L_i also commutes with D_r and since $L_i r = 0$, it follows from the final formula of part c) that the Laplacian ∇^2 also commutes with L_i : $[L_i, \nabla^2] = 0$. Convince yourself that this is true. Thus ∇^2 , L^2 and L_3 form a set of commuting operators. It turns out that once one adds a radial potential term to the Laplacian, these are associated with the three quantum numbers (n, l, m) that characterize the electronic wave function states of atoms, from which the periodic table and all of chemistry follows. ■

Remark.

Why do we care if linear operators like matrices acting by matrix multiplication or derivative operators acting on functions commute? The eigenvector/eigenfunction technique is very powerful. Suppose two matrices have simultaneous eigenvalues:

$$\underline{A}\underline{x} = \lambda_A \underline{x}, \quad \underline{B}\underline{x} = \lambda_B \underline{x},$$

then

$$[\underline{A}\underline{B}]\underline{x} = (\underline{A}\underline{B} - \underline{B}\underline{A})\underline{x} = (\lambda_B \lambda_A - \lambda_B \lambda_A)\underline{x} = \underline{0}.$$

This only shows that the commutator must have \underline{x} as an eigenvector with eigenvalue 0, but if both matrices are diagonalizable, i.e., share an eigenbasis, then the commutator must have 0 eigenvalue for all the eigenbasis vectors, i.e., it is identically zero. Such matrices are “simultaneously diagonalizable.” When instead we deal with linear derivative operators acting on functions, we call the eigenvectors eigenfunctions. We already showed in Exercise 1.7.12 that the space of 3×3 matrices could be decomposed into eigenspaces of $\underline{L}^2 = \delta^{ab} \underline{L}_a \underline{L}_b$. Since it and \underline{L}_3 commute they are simultaneously diagonalizable and we can choose eigenbases of each such subspace which are eigenmatrices of \underline{L}_3 as well. \blacktriangle

Exercise 5.4.7.
spherical basis?

Consider the Hermitian matrix (just the second Pauli matrix in the upper 2×2 block of the matrix, see Exercise 1.7.12)

$$\underline{\mathcal{L}}_3 = i\underline{L}_3 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Find its eigenvalues and show that the following combinations of the standard basis vectors

$$\underline{e}_\pm = \underline{e}_1 \pm i\underline{e}_2, \underline{e}_0 = \underline{e}_3$$

of \mathbb{R}^3 form an eigenbasis, called the spherical basis. What are their respective eigenvalues? What is the result of a rotation of this basis by an angle θ

$$e^{\theta \underline{L}_3} = e^{-i\theta \underline{\mathcal{L}}_3}?$$

This spherical basis is important for creating the vector spherical harmonics by combining them properly with the scalar spherical harmonics. \blacksquare

5.5 Non-Cartesian coordinates on \mathbb{R}^n (polar coordinates in \mathbb{R}^2)

The dual basis covectors $\omega^i \equiv x^i$ to the standard basis $\{e_i\}$ of \mathbb{R}^n are the standard Cartesian coordinates on \mathbb{R}^n . Any change of basis of this vector space

$$\begin{aligned} e_{i'} &= A^{-1j}{}_i e_j, & e_i &= A^j{}_i e_{j'}, \\ \omega^{i'} &= A^i{}_j \omega^j, & \omega^i &= A^{-1i}{}_j \omega^j \end{aligned}$$

leads to a new set of Cartesian coordinates $x^{i'} = A^i{}_j x^j$, where $\underline{A} = (A^i{}_j)$ is a constant matrix.

The Cartesian coordinates also induce a basis $\{\partial/\partial x^i|_P\}$ of the tangent space at each point of \mathbb{R}^n , with dual basis $\{dx^i|_P\}$ of the corresponding cotangent space. The set of vector fields $\{\partial/\partial x^i\}$ is a frame on \mathbb{R}^n in terms of which any tensor field may be expressed

$$T = T^{i_1 \dots i_p}_{j_1 \dots j_q} \frac{\partial}{\partial x^{i_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{i_p}} \otimes \omega^{j_1} \otimes \dots \otimes \omega^{j_q},$$

with the component functions defined by

$$T^{i_1 \dots i_p}_{j_1 \dots j_q} = T \left(dx^{i_1}, \dots, dx^{i_p}, \frac{\partial}{\partial x^{j_1}}, \dots, \frac{\partial}{\partial x^{j_q}} \right),$$

which are functions on \mathbb{R}^n . The “constant” tensor fields on \mathbb{R}^n whose Cartesian coordinate component functions are just constants are in a 1–1 correspondence with the tensors on the vector space \mathbb{R}^n . Their components are clearly constants in any Cartesian frame.

For example, $X = \partial/\partial x + \partial/\partial y + \partial/\partial z$ is a constant vector field on \mathbb{R}^3 , while $G = dx \otimes dx + dy \otimes dy + dz \otimes dz = \delta_{ij} dx^i \otimes dx^j$ is a constant metric field, the Euclidean metric tensor field on \mathbb{R}^3 . The self-inner product of X with itself

$$\begin{aligned} G(X, X) &= dx \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \right) dx \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \right) \\ &\quad + dy \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \right) dy \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \right) \\ &\quad + dz \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \right) dz \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \right) \\ &= 1(1) + 1(1) + 1(1) = 3 = \langle 1, 1, 1 \rangle \cdot \langle 1, 1, 1 \rangle \end{aligned}$$

is just the self-inner product of the corresponding vector $\langle 1, 1, 1 \rangle \in \mathbb{R}^3$. The covector field $\theta = 4dx$ is a constant 1-form on \mathbb{R}^3 .

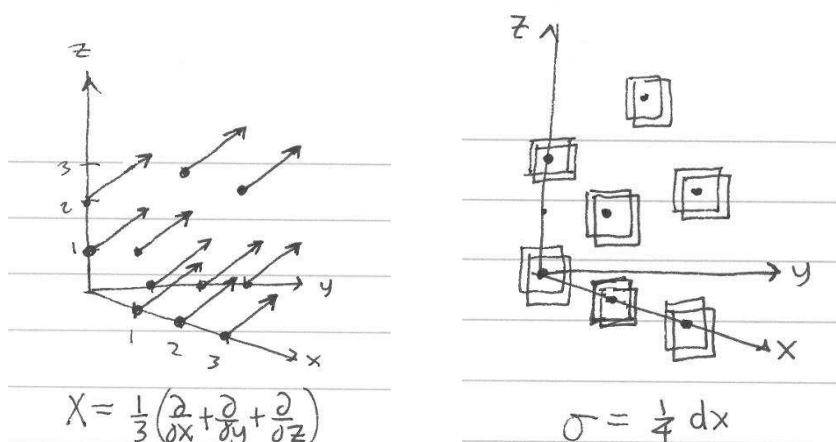


Figure 5.8: Picturing vector fields and 1-forms as a field of representative arrows or plane pairs on a grid. [correction: change figure to $4dx$]

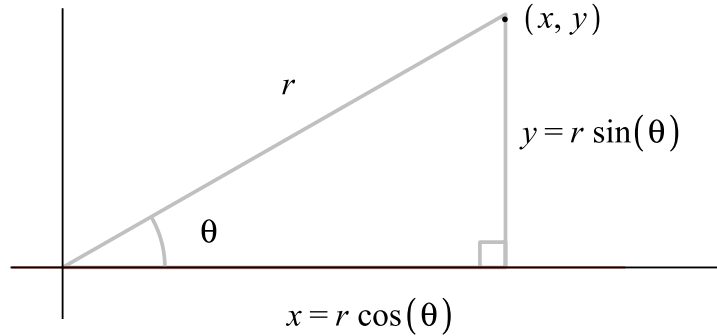
Figure 5.9: Polar coordinates on \mathbb{R}^2 .

Figure 5.8 shows how we can picture the 1-form as a field of pieces of the pair of planes which represent its value at each point just like we picture a vector field as a field of arrows with initial point at the point where they represent a value. Similar pictures hold for nonconstant vector and covector fields.

Non-Cartesian coordinates on \mathbb{R}^n often prove useful, especially when a problem under consideration has a symmetry associated with special families of surfaces like concentric circles, spheres or cylinders. Polar coordinates $\{r, \theta\}$ on \mathbb{R}^2 are the most familiar example, followed by cylindrical coordinates $\{\rho, \phi, z\}$ on \mathbb{R}^3 and spherical coordinates $\{r, \theta, \phi\}$ on \mathbb{R}^3 .

Consider polar coordinates on \mathbb{R}^2 . The usual picture is illustrated by Fig. 5.9, using the more familiar coordinate symbols x, y , with the coordinate transformation and its inverse given by

$$\begin{array}{ll} a) \quad \begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases} & b) \quad \begin{cases} r = \sqrt{x^2 + y^2} \geq 0 \\ \tan \theta = y/x \end{cases} \\ \text{(parametrization map)} & \text{(coordinate map)} \end{array}$$

If we agree to choose $\theta \in (-\pi, \pi]$, we get a unique polar angle for every point except the origin, a function on the plane which we can designate by Θ

$$\theta = \Theta \equiv \begin{cases} \tan^{-1}(y/x) & x > 0, \\ \tan^{-1}(y/x) + \pi & x < 0, y > 0, \\ \tan^{-1}(y/x) - \pi & x < 0, y < 0, \\ \pi/2 & x = 0, y > 0, \\ -\pi/2 & x = 0, y < 0, \\ \pi & x < 0, y = 0. \end{cases}$$

Thus a unique pair of values of the polar coordinates characterize every point in the plane except the origin where $r = 0$ but θ is undetermined and no choice of value for θ there will make the function continuous at the origin. This is called a “coordinate singularity”.

WARNING: We use the symbols x and y or r and θ for different things! We interpret x and y as functions on the plane, but we also use (x, y) to represent a particular point in the plane, i.e., the pair of values of the Cartesian coordinate functions at that point. This sloppy habit means we have to be careful so that in any given situation we understand which meaning is intended. Otherwise in order to be clear we can use notation which distinguishes them. For example $x = r \cos \theta$ or $r = (x^2 + y^2)^{1/2}$ are each relationships among three functions on the plane which happen to express one “as a function of the others”, or we can think of them as relationships among 3 “variables.” To make explicit the functional relationship, we must name explicitly the function, which we will do below, for example $r = \Phi^1(x, y), \theta = \Phi^2(x, y)$.

What is really going on with the above picture and relationships between the Cartesian and polar coordinates? Well, first of all we have two distinct copies of \mathbb{R}^2 , a “physical space” of points which has a lot of mathematical structure, and a “coordinate space” on which operations involving the polar coordinates occur. The relationship between the two sets of coordinates define two maps between these spaces going in opposite directions. The “coordinate map” $(r, \theta) = (\Phi^1(x, y), \Phi^2(x, y))$ associates with each point (x, y) in the “physical space” a point (r, θ) in the coordinate space which is the pair of values of the polar coordinates there. The “parametrization map” $(x, y) = (\Psi^1(r, \theta), \Psi^2(r, \theta))$ associates with each point (r, θ) in the coordinate space, the point in the (x, y) “physical space” that it represents.

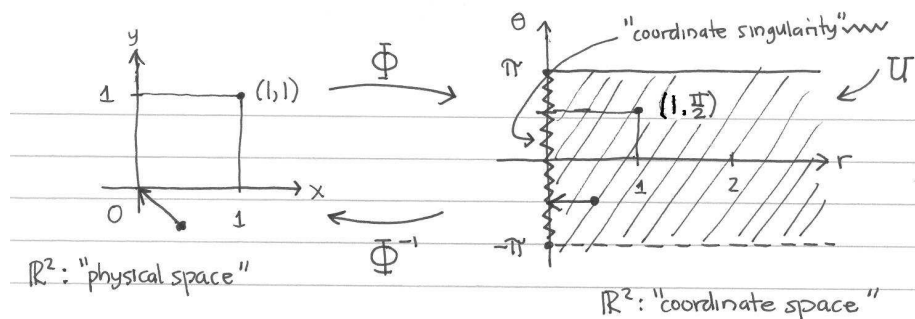


Figure 5.10: Polar coordinates on \mathbb{R}^2 : the physical space and the coordinate space for the plane. The origin in the physical space corresponds to an entire line segment in the coordinate space, where the 1-1 nature of the relationship breaks down, called a “coordinate singularity.” [oops, the figure should show the polar coordinate point $(\sqrt{2}, \pi/4)$ corresponding to the Cartesian coordinate point $(1, 1)$,]

For the polar coordinates as shown in Fig. 5.10, the coordinate map Φ maps physical space to its coordinate representation

$$\begin{aligned}\Phi : U = \mathbb{R}^2 - \{\vec{0}\} &\longrightarrow \mathcal{U} \subset \mathbb{R}^2, \\ \Phi(u^1, u^2) &= (\sqrt{(u^1)^2 + (u^2)^2}, \Theta(u^1, u^2)),\end{aligned}$$

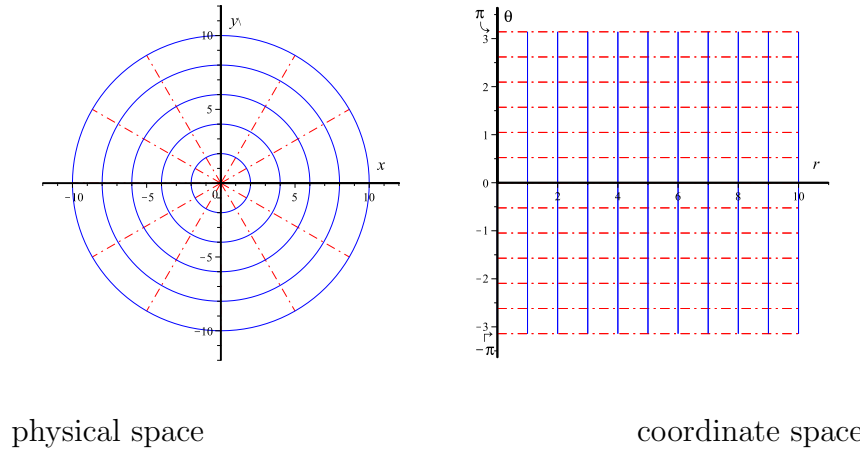


Figure 5.11: The polar coordinate grid on physical space: $(x, y) = \Psi(r, \theta) = (r \cos \theta, r \sin \theta)$. Hold θ fixed, vary r : r coordinate lines (half rays from origin). Hold r fixed, vary θ : θ coordinate lines (circle centered at origin). The map Ψ maps from right to left, while Φ maps from left to right.

which takes the open set U excluding the origin from \mathbb{R}^2 onto the subset $\mathcal{U} = (0, \infty) \times (-\pi, \pi]$ of another copy of \mathbb{R}^2 . (An open set is simply a set of points that does not include its boundary points. A closed set includes its boundary points.)

Its inverse, the parametrization map Ψ , maps the coordinate representation of a physical point onto that point

$$\begin{aligned}\Psi : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2, \\ \Psi(u^1, u^2) &= (u^1 \cos u^2, u^1 \sin u^2),\end{aligned}$$

which maps all of \mathbb{R}^2 onto all of \mathbb{R}^2 an infinite number of times unless we restrict it to the subset \mathcal{U} . In both cases we denote a point in \mathbb{R}^2 by the neutral symbols (u^1, u^2) .

Notice that $\Phi \circ \Psi$ maps all of the coordinate space onto the subset \mathcal{U} which is the image of the coordinate map Φ . For points in \mathcal{U} , this is the identity map. For points outside of \mathcal{U} this associates our specific choice of polar coordinates with any other possible choices, like fixing $\theta \in [0, 2\pi)$, or allowing negative r . The vertical segment between $-\pi$ and π on the θ axis (not in \mathcal{U}) corresponds to the origin in physical space in the sense that approaching it from any nearby point of \mathcal{U} corresponds to approaching the origin in physical space in a certain direction. The parametrization map collapses this whole line segment to a single point in the physical space, so the map is no longer 1-1 as it must be to faithfully represent distinct points with distinct coordinates.

The map $\Psi \circ \Phi$ maps all of physical space except for the origin onto itself, where it is the identity map. Restricting Ψ to the set \mathcal{U} makes it the inverse of the map Φ . The parametrization map Ψ represents the plane as a 2-parameter family of parametrized curves (“coordinate lines”) which make up the polar coordinate grid, illustrated in Fig. 5.11

We can easily compute the tangents to these parametrized curves as ordinary vector func-

tions and as tangent vectors defined along the curves

$$\begin{aligned}\vec{\mathcal{E}}_1(r, \theta) &= \left(\frac{\partial x}{\partial r}, \frac{\partial y}{\partial r} \right) = (\cos \theta, \sin \theta), \\ \vec{\mathcal{E}}_2(r, \theta) &= \left(\frac{\partial x}{\partial \theta}, \frac{\partial y}{\partial \theta} \right) = (-r \sin \theta, r \cos \theta), \\ \mathcal{E}_1(r, \theta) &= \cos \theta \frac{\partial}{\partial x} \Big|_{(r \cos \theta, r \sin \theta)} + \sin \theta \frac{\partial}{\partial y} \Big|_{(r \cos \theta, r \sin \theta)}, \\ \mathcal{E}_2(r, \theta) &= -r \sin \theta \frac{\partial}{\partial x} \Big|_{(r \cos \theta, r \sin \theta)} + r \cos \theta \frac{\partial}{\partial y} \Big|_{(r \cos \theta, r \sin \theta)}.\end{aligned}$$

Suppose $f(x, y) = x^2 - y^2$ is a function on \mathbb{R}^2 . Then the tangent vector $\mathcal{E}_1(r, \theta)$ acts on it to produce the number

$$\begin{aligned}\mathcal{E}_1(r, \theta)f &= \cos \theta \frac{\partial}{\partial x} \Big|_{(r \cos \theta, r \sin \theta)} (x^2 - y^2) + \sin \theta \frac{\partial}{\partial y} \Big|_{(r \cos \theta, r \sin \theta)} (x^2 - y^2) \\ &= \cos \theta (2x) \Big|_{(r \cos \theta, r \sin \theta)} + \sin \theta (-2y) \Big|_{(r \cos \theta, r \sin \theta)} \\ &= 2r(\cos^2 \theta - \sin^2 \theta) = 2r \cos 2\theta\end{aligned}$$

for given values of r and θ . Note that this is the same as first evaluating f in terms of the new coordinates and just taking the r partial derivative

$$\begin{aligned}f(r \cos \theta, r \sin \theta) &= x^2 - y^2 = (r \cos \theta)^2 - (r \sin \theta)^2 = r^2(\cos^2 \theta - \sin^2 \theta) = r^2 \cos 2\theta, \\ \frac{\partial}{\partial r} f(r \cos \theta, r \sin \theta) &= \frac{\partial}{\partial r} (r^2 \cos 2\theta) = 2r \cos 2\theta.\end{aligned}$$

The ordinary dot products of the vector functions $\vec{\mathcal{E}}_1$ and $\vec{\mathcal{E}}_2$ (multivariable calculus notation) or equivalently the inner products of the vector fields \mathcal{E}_1 and \mathcal{E}_2 (as differential operators) using the Euclidean metric tensor G are

$$\begin{aligned}\vec{\mathcal{E}}_1(r, \theta) \cdot \vec{\mathcal{E}}_1(r, \theta) &= \cos^2 \theta + \sin^2 \theta = 1 = G|_{(r \cos \theta, r \sin \theta)}(\mathcal{E}_1(r, \theta), \mathcal{E}_1(r, \theta)), \\ \vec{\mathcal{E}}_2(r, \theta) \cdot \vec{\mathcal{E}}_2(r, \theta) &= r^2 \sin^2 \theta + r^2 \cos^2 \theta = r^2 = G|_{(r \cos \theta, r \sin \theta)}(\mathcal{E}_2(r, \theta), \mathcal{E}_2(r, \theta)), \\ \vec{\mathcal{E}}_1(r, \theta) \cdot \vec{\mathcal{E}}_2(r, \theta) &= -r \cos \theta \sin \theta + r \cos \theta \sin \theta = 0 = G|_{(r \cos \theta, r \sin \theta)}(\mathcal{E}_1(r, \theta), \mathcal{E}_2(r, \theta)),\end{aligned}$$

so $\mathcal{E}_1(r, \theta)$, $\mathcal{E}_2(r, \theta)$ are mutually orthogonal tangent vectors of lengths 1 and r respectively.

Now going back to the sloppy notation $x = r \cos \theta$, $y = r \sin \theta$ suppressing functional arguments, then

$$\cos \theta = \frac{x}{r} = \frac{x}{\sqrt{x^2 + y^2}}, \quad \sin \theta = \frac{y}{r} = \frac{y}{\sqrt{x^2 + y^2}}.$$

These relations enable us to re-express these two tangent vector fields entirely in terms of the Cartesian coordinates, so define the vector fields on physical space by

$$E_1 = (x^2 + y^2)^{-1/2} \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} \right), \quad E_2 = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}.$$

Their values at $(r \cos \theta, r \sin \theta)$ are just $\mathcal{E}_1(r, \theta)$ and $\mathcal{E}_2(r, \theta)$

$$E_i|_{(r \cos \theta, r \sin \theta)} = \mathcal{E}_i(r, \theta), \quad i = 1, 2.$$

In other words E_1 at a given point equals the tangent vector to the curve through the point corresponding to translations in the coordinate r , while E_2 is the same for θ . Their action on a function is equivalent to partial differentiation once it is re-expressed in term of the new coordinates

$$E_1|_{(r \cos \theta, r \sin \theta)} f = \mathcal{E}_1(r, \theta) f = \frac{\partial}{\partial r} f(r \cos \theta, r \sin \theta) = \frac{\partial}{\partial r} [f \circ \Psi](r, \theta),$$

$$E_2|_{(r \cos \theta, r \sin \theta)} f = \mathcal{E}_2(r, \theta) f = \frac{\partial}{\partial \theta} f(r \cos \theta, r \sin \theta) = \frac{\partial}{\partial \theta} [f \circ \Psi](r, \theta).$$

The function $f \circ \Psi$ on the coordinate space is just the function one gets by expressing f in terms of the coordinate functions r and $\theta = \Theta$ (namely $r^2 \cos 2\theta$ in our explicit example above) and we write

$$E_1 = \frac{\partial}{\partial r}, \quad E_2 = \frac{\partial}{\partial \theta}.$$

In other words $\{E_i\}$ is the coordinate frame associated with the polar coordinates. The change in frame (the matrix columns are the Cartesian coordinate components of the new frame vectors)

$$E_i = A^{-1j}_i \frac{\partial}{\partial x^j}, \quad \underline{A}^{-1} = \begin{pmatrix} \frac{x}{(x^2 + y^2)^{1/2}} & -y \\ \frac{y}{(x^2 + y^2)^{1/2}} & x \end{pmatrix}, \quad \det \underline{A}^{-1} = (x^2 + y^2)^{1/2}$$

is invertible everywhere except at the origin where $E_2 = 0$ and E_1 has no unique limiting value, with inverse

$$\underline{A} = (x^2 + y^2)^{-1/2} \begin{pmatrix} x & y \\ -\frac{y}{(x^2 + y^2)^{1/2}} & \frac{x}{(x^2 + y^2)^{1/2}} \end{pmatrix},$$

whose rows are the Cartesian coordinate components of the corresponding dual basis.

If we return to the indexed notation $(x^1, x^2) = (x, y)$ and let $(x^{1'}, x^{2'}) = (r, \theta)$, then from the identification

$$\frac{\partial}{\partial x^{i'}} = \frac{\partial x^j}{\partial x^{i'}} \frac{\partial}{\partial x^j} = A^{-1j}_i \frac{\partial}{\partial x^j}, \quad dx^{i'} = \frac{\partial x^{i'}}{\partial x^j} dx^j = A^i_j dx^j,$$

correlates the change of frame matrix with the two so called Jacobian matrices of partial derivatives of one set of coordinates with respect to the other

$$\frac{\partial x^i}{\partial x^{j'}} = A^{-1i}_j, \quad \frac{\partial x^{i'}}{\partial x^j} = A^i_j.$$

In terms of the explicit coordinate variables, one can express the new partial derivative operators in terms of the old ones (and vice versa) so

$$\frac{\partial}{\partial r} = \frac{\partial x}{\partial r} \frac{\partial}{\partial x} + \frac{\partial y}{\partial r} \frac{\partial}{\partial y}, \quad \frac{\partial}{\partial \theta} = \frac{\partial x}{\partial \theta} \frac{\partial}{\partial x} + \frac{\partial y}{\partial \theta} \frac{\partial}{\partial y}.$$

Using polar coordinates basically means re-expressing everything in terms of them, i.e., moving over to the coordinate space, and doing calculus operations there. The polar coordinate frame vectors

$$E_1 = (x^2 + y^2)^{-1/2} \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} \right) = \frac{\partial}{\partial r} \quad , \quad E_2 = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} = \frac{\partial}{\partial \theta}$$

do not form a frame at the origin. E_1 is not defined and has no limit there, while E_2 vanishes. If we remove the factor $(x^2 + y^2)^{-1/2}$ from E_1 , it is defined but also equal to zero at the origin. This means we cannot use them to express tangent vectors at the origin. Everywhere else they are fine. We therefore need the idea of a *local frame* and a *local coordinate patch* to handle frames and coordinate systems which have problems at certain points of space, using the word “local” to distinguish them from the global Cartesian frames which are valid everywhere in space.

A local frame (defined on an open set $U \subset \mathbb{R}^n$) will be a set of n vector fields $e_i = e^j_i(x) \partial / \partial x^j$ which form a basis of the tangent space at each point of U , i.e., $\det(e^j_i(x))$ never vanishes in this set. If $U = \mathbb{R}^n$, it will be called a global frame or simply a frame. A local coordinate patch will be an open set $U \subset \mathbb{R}^n$ and a set of n coordinate functions such that the associated coordinate vector fields form a local frame on U . If $\{x^i\}$ are Cartesian coordinates and

$$\frac{\partial}{\partial x^{i'}} = \frac{\partial x^j}{\partial x^{i'}} \frac{\partial}{\partial x^j} = A^{-1j}_{i'} \frac{\partial}{\partial x^j} \quad ,$$

this requires that the Jacobian matrix determinant be nonzero everywhere on the set U

$$\det \underline{A}^{-1} = \det \left(\frac{\partial x^j}{\partial x^{i'}} \right) \neq 0 \quad \text{on } U \quad .$$

For polar coordinates $U = \mathbb{R}^2 - \{\vec{0}\}$ is an open set on the physical space. Note that the subset \mathcal{U} of the coordinate space (r - θ plane), which was the range of the coordinate functions from all of physical space (including the origin which is mapped to the line segment $r = 0$), contained the boundary points $\theta = \pi$ and was not an open set.

In order to deal with tensor fields or tangent tensors at a given point of \mathbb{R}^n , it must be an interior point of the open set U of the local frame or of the local coordinate patch we wish to use or we must play special games to circumvent the difficulties associated with the bad boundary points.

Example 5.5.1. Let U be the interior of a circle of radius $\epsilon > 0$ about the origin in the plane. The Cartesian coordinates $\{x, y\}$ are local coordinates on U , for every value of $\epsilon > 0$. In order to use polar coordinates, which fail at the origin, we must use some other local coordinate patch like one of this family which contains the origin in order to handle that problem point. The polar coordinates themselves are local coordinates on the plane minus the origin. This local coordinate patch has to be supplemented by some other patch like one of these local Cartesian coordinate patches in order to figure out what's going on at the origin. The two patches together then form a “coordinate covering” of the plane, with each point “covered” by at least one local coordinate patch.

In fact the situation is a bit more complicated than this, since one has a discontinuity in the angular coordinate at $\theta = \pi$ where a jump of 2π occurs, so we need to make sure functions of θ that we deal with are always periodic. However, we don't need to worry much about these technicalities at this introductory level, and in practice it is hardly ever necessary as well. \square

When a coordinate patch just missed being global on some set “of measure zero” (like a single point or a curve or a surface), the points where it fails are called coordinate singularities. The origin is a coordinate singularity for polar coordinates. The coordinate map Φ fails to be well-defined since the parametrization map is no longer 1-1, and although $r = 0$, one has many choices for the polar angle θ to be assigned there.

We can also redefine the infinite range radial coordinate to one which has a finite interval of values. Suppose we introduce a new radial coordinate χ by

$$r = \tan \chi, \quad \chi = \tan^{-1} r, \quad r \in [0, \infty), \quad \chi \in [0, \pi/2).$$

Although the coordinate lines are still the same, all of physical space except for the origin is mapped onto a rectangle in coordinate space with the new “edge” representing “infinity.” One must be careful with the boundary of a coordinate patch where the limiting boundary points in the coordinate space can be deceptive as far as what points they represent in the physical space. In this case one side of the boundary of this coordinate rectangle corresponds to the circle at infinity and the opposite side to a single point (the origin). The remaining two opposing sides must be identified since they correspond to the same line in physical space (the negative x -axis).

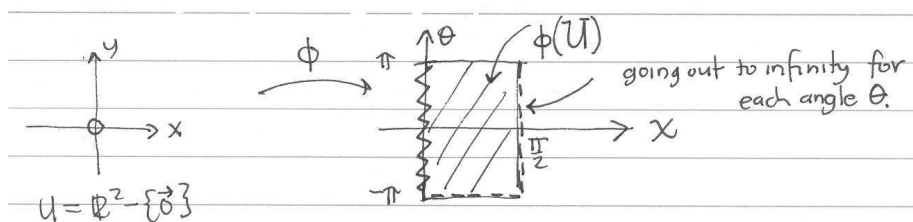


Figure 5.12: Mapping the radial polar coordinate $r = \tan \chi$ onto the closed interval $[0, \pi/2]$ using the parametrization map $(x, y) = (\tan \chi \cos \theta, \tan \chi \sin \theta)$.

Exercise 5.5.1.

polar coordinates and circles not centered at the origin

a) It is easy to re-express the Cartesian equations of horizontal and vertical lines in the plane $y = y_0$ and $x = x_0$ in terms of polar coordinates, as well as the lines through the origin (trivially angular coordinate half lines). Do this and solve them for r as a function of θ . Similarly it is easy to re-express the Cartesian equations of circles with centers on the x and y axes passing through the origin $(x - a)^2 + y^2 = a^2$, $x^2 + (y - a)^2 = a^2$, as well as those with center at the origin (trivially the radial coordinate circles). Expand these equations and transform them to polar

coordinates and solve them for r as a function of θ . What is the range of θ over one of these circles?

b) Suppose we move those circles out farther from the origin along those axes, say $(x-a)^2 + y^2 = b^2$, $b > a$. Now there are two values of r for each value of θ along these circles. What is the range of θ for the case $b > a > 0$? Express the two values of r as functions of θ for that range.

If you use a computer algebra system, you want the equation expressed in terms of the sine as is natural in a hand calculation, and not in terms of the cosine, which happens when you simplify it in Maple.



Exercise 5.5.2.

polar coordinates and multipetal curves

a) Consider the flower petal curve $r = a \cos(n\theta)$. What is the range of θ about 0 over one lobe of this curve (between successive values $r = 0$).

b) For $n = 2$, re-express this equation in terms of Cartesian coordinates as a polynomial condition using the double angle formula for the sine. [Hint: first multiply both sides of the equation by r^2 , then square both sides.]

c) Repeat for $r = 2 \sin(2\theta)$.



5.6 Cylindrical and spherical coordinates on \mathbb{R}^3

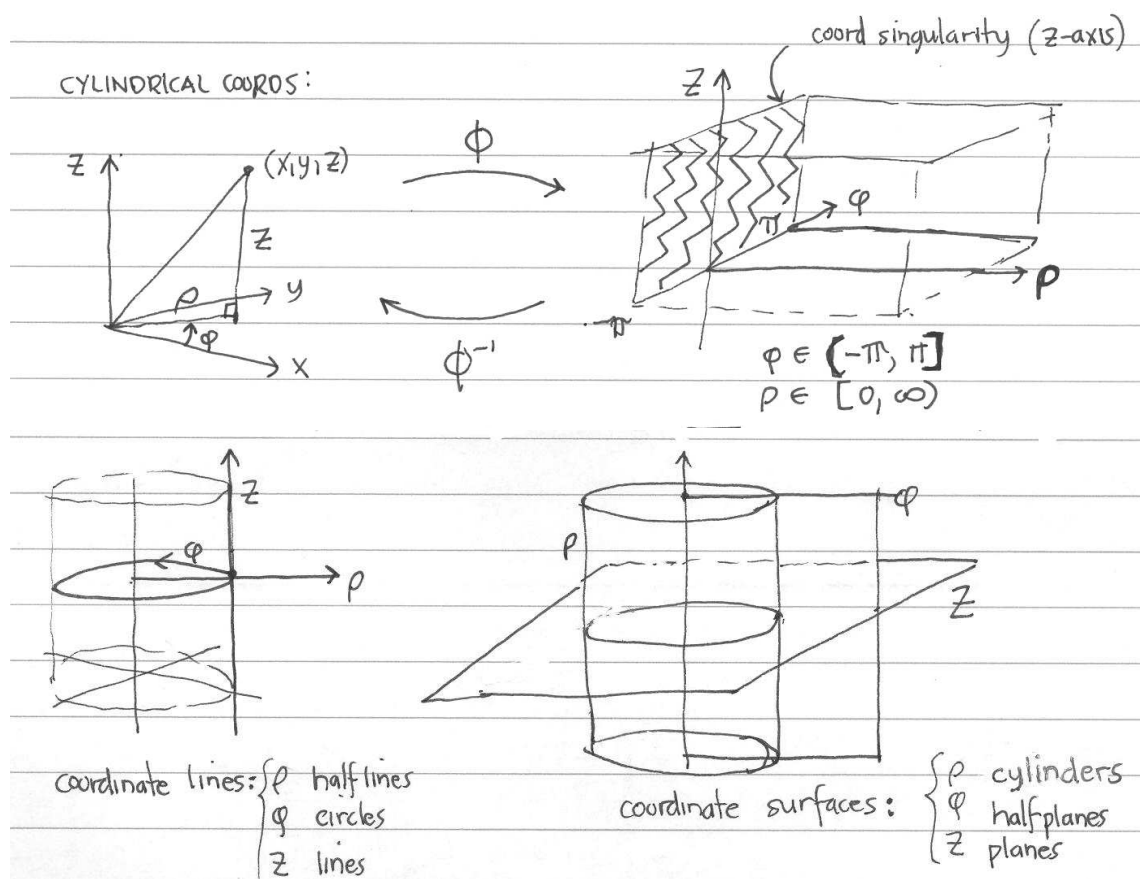


Figure 5.13: Cylindrical coordinates on \mathbb{R}^3 and their coordinate lines and surfaces.

2-dimensional space is a bad example for some issues since $1 = n - 1$, so lines and planes as well as curves and surfaces coincide. 3-dimensional space gives us a better picture of what occurs in higher dimensions.

Cylindrical and spherical coordinates in 3-dimensions generalize polar coordinates in 2-dimensions and are covered in every course in multivariable calculus. They provide us with a convenient springboard to jump into the more general topic of curvilinear coordinates in curved spaces. Coming from training as a physicist, I will use the physics convention for naming the angular coordinates in this context. The “azimuthal angle” measured around the vertical z -axis will be called ϕ instead of θ as in polar coordinates and to which it reduces in the horizontal x - y plane, while θ will be used for the polar angle measured down from the upward vertical axis (the “North pole” on any sphere centered at the origin!). In multivariable calculus it is easier to just add one more angle with a new name than switch the names on students. Physicists actually use these coordinates extensively and have a long tradition of certain conventions which are useful to respect. Similarly r (the length of the position vector \vec{r}) is always used for the radial

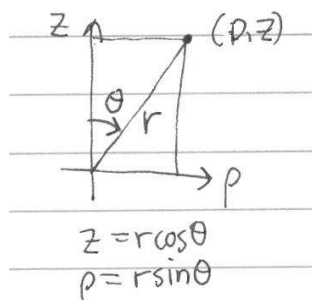


Figure 5.14: One passes from cylindrical coordinates (ρ, ϕ, z) to spherical coordinates (r, θ, ϕ) by introducing polar coordinates in the ρ - z half plane.

distance from the origin in \mathbb{R}^3 , so the cylindrical coordinate giving the radial distance from the z -axis will be designated by a new name ρ .

For each coordinate system we need to specify a parametrization map Ψ representing the Cartesian coordinate functions in terms of the new coordinates and a coordinate map Φ expressing the new coordinate functions in terms of the Cartesian coordinates, well-defined on some open set $U \subset \mathbb{R}^3$ covering “almost all of space.”

Cylindrical coordinates

$$\Psi : \begin{cases} x = \rho \cos \phi \\ y = \rho \sin \phi \\ z = z \end{cases} \quad \Phi : \begin{cases} \rho = \sqrt{x^2 + y^2} \\ \tan \phi = \frac{y}{x} \\ z = z \end{cases} \quad (\text{same solution as before})$$

The cylindrical coordinates (ρ, ϕ) are just the polar coordinates (r, θ) of the projection of a point vertically downward or upward to the x - y plane (see Fig. 5.13). The open set of \mathbb{R}^3 on which they are uniquely defined (assuming $-\pi < \phi \leq \pi$) is $U = \mathbb{R}^3 - \{(x, y, z) | y = 0, x \leq 0\}$, which excludes the z -axis and the vertical half plane through negative x -axis where the angular coordinate is respectively not defined and discontinuous (jumping in value by 2π). This just means that when we consider these bad points, we have to be careful about what we are doing.

Spherical coordinates

To go from cylindrical coordinates to spherical coordinates, one introduces polar coordinates $\{r, \theta\}$ in the ρ - z half plane, with the same open set U as before

$$\Psi : \begin{cases} x = r \sin \theta \cos \phi \\ y = r \sin \theta \sin \phi \\ z = r \cos \theta \end{cases} \quad \Phi : \begin{cases} r = \sqrt{x^2 + y^2 + z^2} \\ \theta = \cos^{-1} \left(\frac{z}{\sqrt{x^2 + y^2 + z^2}} \right) \\ \tan \phi = \frac{y}{x} \end{cases} \quad (\text{same solution as before})$$

The intermediate coordinate transformation from cylindrical to spherical coordinates which takes place in the ρ - z half plane is pictured in Fig. 5.14 and is very useful in visualizing the latter coordinates

$$\begin{cases} \rho = r \sin \theta \\ z = r \cos \theta \end{cases}, \quad \begin{cases} r = \sqrt{\rho^2 + z^2} \\ \theta = \cos^{-1} \left(\frac{z}{\sqrt{\rho^2 + z^2}} \right) \end{cases}.$$

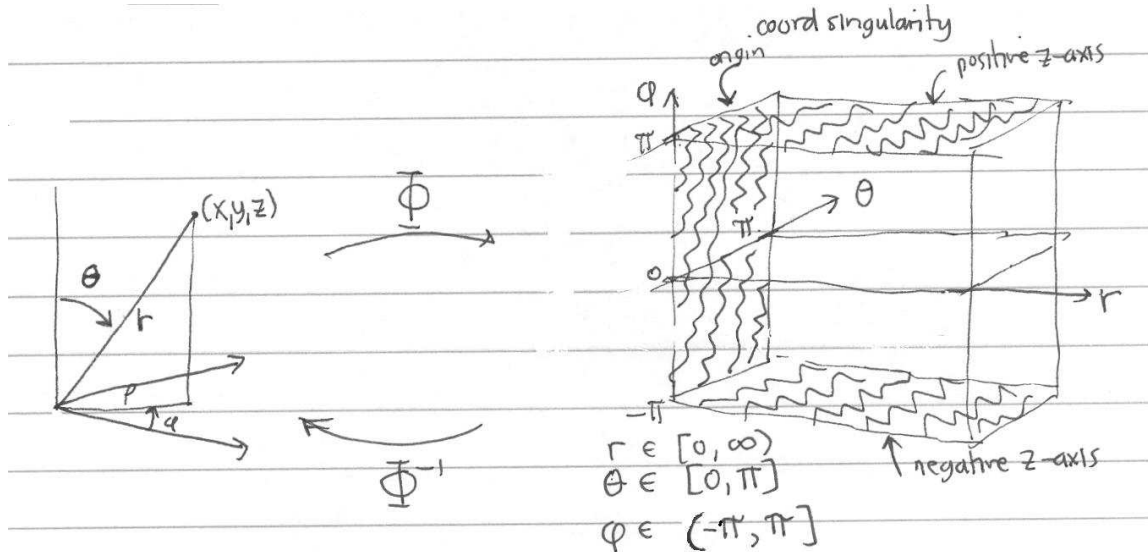


Figure 5.15: Spherical coordinates on \mathbb{R}^3 .

In both of these cases the coordinate map Φ is discontinuous on the half plane $\{(x, y, z) | x < 0, y = 0\}$ where ϕ has a jump of 2π and undefined on the z -axis where the angular coordinate ϕ is not defined. A “coordinate singularity” occurs at the z -axis for this reason, while at the origin θ is also undefined, making the singularity worse. The parametrization map maps a line segment (different ϕ values) onto each point on the z -axis except at the origin where a rectangle (all θ and ϕ values) is mapped onto a single point.

I haven’t been consistent about the open set U of a local coordinate patch. For polar coordinates I included the negative x -axis where the jump in the angular coordinate occurs, but not the corresponding half plane for cylindrical and spherical coordinates. If we require ϕ to be continuous then we must exclude points of discontinuity. The coordinate frame and dual frame are perfectly fine there, however, because of periodicity, so the local frame $\{E_1, E_2, E_3\}$ expressed in terms of Cartesian coordinates is valid everywhere except on the z -axis, i.e., its U includes the discontinuous points of ϕ .

Now that we have the two new coordinate systems defined, the first thing to do for each is compute the new coordinate frame vector fields and dual 1-forms. The next step is to evaluate their inner products and re-express the Euclidean metric, which shows them to be orthogonal coordinate systems, namely those for which the coordinate lines have orthogonal tangent vectors. The orthogonal coordinate frame vector fields can be normalized to orthonormal frame

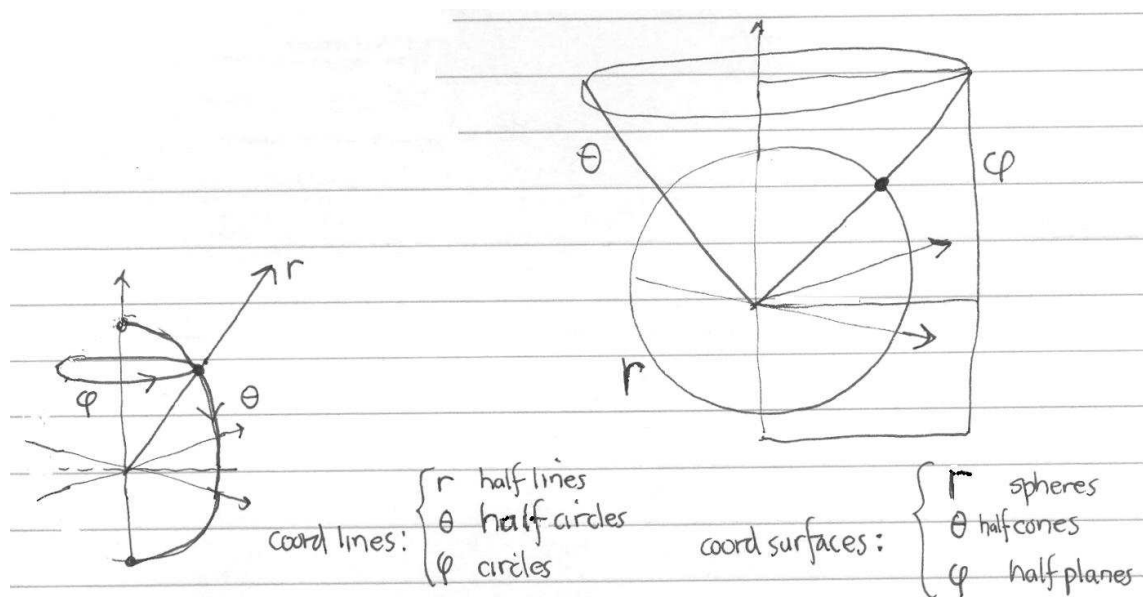


Figure 5.16: Coordinate lines and surfaces for spherical coordinates on \mathbb{R}^3 .

vector fields simply by dividing by their lengths, and these frames must be explored since they are natural to use in calculations. Finally we can consider derivative operators expressed in the new coordinates and in their closely associated orthonormal frames.

Exercise 5.6.1.

mathematical wedding band surface boundaries

Consider the region of space bounded by the surfaces

$$\begin{aligned} \text{sphere:} \quad & x^2 + y^2 + z^2 = 4^2, \\ \text{cylinder:} \quad & x^2 + y^2 = 3^2. \end{aligned}$$

This is roughly the shape of simple wedding band ring, though exaggerated to have simple numbers, and it is invariant under rotations around the z -axis, i.e., is independent of the azimuthal angle ϕ . Draw a rough sketch of the cross-section of this region in a vertical plane of constant ϕ and evaluate the boundary values of the remaining cylindrical and spherical coordinates on the two intersection rings of these two surfaces. Note that this is equivalent to a plane problem in polar coordinates for a vertical line and a circle, treated in Exercise 5.5.1.

a) In cylindrical coordinates (ρ, ϕ, z) describe this region first by explicit relations of the form $\rho_1 \leq \rho \leq \rho_2$, $z_1(\rho) \leq z \leq z_2(\rho)$ and second by relations of the form $z_1 \leq z \leq z_2$, $\rho_1(z) \leq \rho \leq \rho_2(z)$.

b) In spherical coordinates (r, θ, ϕ) , describe this region by explicit relations of the form $\theta_1 \leq \theta \leq \theta_2$, $r_1(\theta) \leq r \leq r_2(\theta)$.

■

5.7 Cylindrical coordinate frames

The cylindrical coordinate differentials, using multiple notations, are

$$\begin{aligned} W^1 \equiv \omega^\rho &= d\rho = d(x^2 + y^2)^{1/2} = \frac{d(x^2 + y^2)}{2(x^2 + y^2)^{1/2}} = \frac{x dx + y dy}{(x^2 + y^2)^{1/2}}, \\ W^2 \equiv \omega^\phi &= d\phi = d\left(\tan^{-1} \frac{y}{x} + \text{const}\right) = \frac{d(y/x)}{1 + (y/x)^2} = \frac{-y dx + x dy}{(x^2 + y^2)}, \\ W^3 \equiv \omega^z &= dz. \end{aligned}$$

Comparison with

$$\begin{aligned} (x^{1'}, x^{2'}, x^{3'}) &\equiv (\rho, \phi, z) \quad (x^1, x^2, x^3) \equiv (x, y, z) \\ dx^{i'} &= A^i_j dx^j = \frac{\partial x^{i'}}{\partial x^i} dx^i, \quad \frac{\partial}{\partial x^{i'}} = A^{-1j}_i \frac{\partial}{\partial x^j} \end{aligned}$$

shows that the rows of the “Jacobian matrix” (A^i_j) of partial derivatives are the old components of the new basis 1-forms, so one can read off its entries expressed in terms of the old coordinates, which are then easily re-expressed in terms of the new coordinates

$$(A^i_j) = \left(\frac{\partial x^{i'}}{\partial x^j} \right) = \begin{pmatrix} \frac{x}{(x^2 + y^2)^{1/2}} & \frac{y}{(x^2 + y^2)^{1/2}} & 0 \\ \frac{-y}{x^2 + y^2} & \frac{x}{x^2 + y^2} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\rho^{-1} \sin \phi & \rho^{-1} \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and hence using the convenient formula for the inverse of a 2×2 matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = (ad - bc)^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

one obtains the inverse Jacobian matrix expressed in terms of the old coordinates, which is then easily re-expressed in terms of the new coordinates

$$(A^{-1i}_j) = \left(\frac{\partial x^i}{\partial x^{j'}} \right) = \begin{pmatrix} \frac{x}{(x^2 + y^2)^{1/2}} & -y & 0 \\ \frac{y}{(x^2 + y^2)^{1/2}} & x & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos \phi & -\rho \sin \phi & 0 \\ \sin \phi & \rho \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The columns of this matrix are the old components of the new coordinate vector fields

$$\begin{aligned} E_1 \equiv e_\rho &= \frac{\partial}{\partial \rho} = \frac{\partial}{\partial x^{1'}} = A^{-1i}_1 \frac{\partial}{\partial x^i} = (x^2 + y^2)^{-1/2} \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} \right), \\ E_2 \equiv e_\phi &= \frac{\partial}{\partial \phi} = \frac{\partial}{\partial x^{2'}} = A^{-1i}_2 \frac{\partial}{\partial x^i} = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}, \\ E_3 \equiv e_z &= \frac{\partial}{\partial z} = \frac{\partial}{\partial x^{3'}} = A^{-1i}_3 \frac{\partial}{\partial x^i} = \frac{\partial}{\partial z}. \end{aligned}$$

On the other hand, by the chain rule, the tangents to the parametrized coordinate lines are

$$\begin{aligned}
 \left. \frac{\partial}{\partial \rho} \right|_{(\rho \cos \phi, \rho \sin \phi, z)} &= \left[\frac{\partial x}{\partial \rho} \frac{\partial}{\partial x} + \frac{\partial y}{\partial \rho} \frac{\partial}{\partial y} + \frac{\partial z}{\partial \rho} \frac{\partial}{\partial z} \right] \Big|_{(\rho \cos \phi, \rho \sin \phi, z)} \\
 &= \left[\cos \phi \frac{\partial}{\partial x} + \sin \phi \frac{\partial}{\partial y} \right] \Big|_{(\rho \cos \phi, \rho \sin \phi, z)}, \\
 \left. \frac{\partial}{\partial \phi} \right|_{(\rho \cos \phi, \rho \sin \phi, z)} &= \left[\frac{\partial x}{\partial \phi} \frac{\partial}{\partial x} + \frac{\partial y}{\partial \phi} \frac{\partial}{\partial y} + \frac{\partial z}{\partial \phi} \frac{\partial}{\partial z} \right] \Big|_{(\rho \cos \phi, \rho \sin \phi, z)} \\
 &= - \left[\rho \sin \phi \frac{\partial}{\partial x} + \rho \cos \phi \frac{\partial}{\partial y} \right] \Big|_{(\rho \cos \phi, \rho \sin \phi, z)}, \\
 \left. \frac{\partial}{\partial z} \right|_{(\rho \cos \phi, \rho \sin \phi, z)} &= \left. \frac{\partial}{\partial z} \right|_{(\rho \cos \phi, \rho \sin \phi, z)}.
 \end{aligned}$$

This gives a direct way to evaluate the matrix \underline{A}^{-1} in terms of the new coordinates as partial derivatives of the parametrization map; the components of these vector fields are its columns. Similarly one could evaluate directly \underline{A} in terms of the old coordinates by partial differentiation of the coordinate map.

Note that the coordinate vector fields $\{E_1, E_2, E_3\}$ fail to be linearly independent on the z -axis where E_2 vanishes, while at the origin E_1 vanishes. This leads to W^1 not having a well-defined limit at the z -axis and causes W^2 to have components which become infinite there.

Now we need a change in notation for the Euclidean metric. We have been using g for functions and G for symmetric inner product tensors. By convention one uses g for symmetric inner product tensors. The Euclidean metric tensor field is

$$g = \delta_{ij} dx^i \otimes dx^j, \quad g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \delta_{ij}.$$

We can re-express it in terms of the new frame

$$g = g_{i'j'} dx^{i'} \otimes dx^{j'}, \quad g_{i'j'} = g\left(\frac{\partial}{\partial x^{i'}}, \frac{\partial}{\partial x^{j'}}\right) = A^{-1m}{}_i A^{-1n}{}_j \delta_{mn} = [(\underline{A}^{-1})^T \underline{I} \underline{A}^{-1}]_{ij}.$$

One can directly take the inner products of $\{E_i\}$ or use the matrix transformation law to obtain the new components as functions of the Cartesian coordinates, or one can just evaluate the differentials of x, y, z which will lead to expressions in terms of the new components. The matrix calculation (exercise) yields

$$(g_{i'j'}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & x^2 + y^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} g_{\rho\rho} & g_{\rho\phi} & g_{\rho z} \\ g_{\phi\rho} & g_{\phi\phi} & g_{\phi z} \\ g_{z\rho} & g_{z\phi} & g_{zz} \end{pmatrix},$$

while

$$\begin{aligned}
 dx &= d(\rho \cos \phi) = \cos \phi d\rho - \rho \sin \phi d\phi, \\
 dy &= d(\rho \sin \phi) = \sin \phi d\rho + \rho \cos \phi d\phi, \\
 dz &= d(z) = dz.
 \end{aligned}$$

Comparing with

$$dx^i = \frac{\partial x^i}{\partial x^{j'}} dx^{j'} = A^{-1i}{}_{j'} dx^{j'}$$

shows that

$$\underline{A}^{-1}(\underline{x}') = \begin{pmatrix} \cos \phi & -\rho \sin \phi & 0 \\ \sin \phi & \rho \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then backsubstituting the coordinate differentials in the metric and simplifying leads to

$$\begin{aligned} g &= dx \otimes dx + dy \otimes dy + dz \otimes dz \\ &= (\cos \phi d\rho - \rho \sin \phi d\phi) \otimes (\cos \phi d\rho - \rho \sin \phi d\phi) \\ &\quad + (\sin \phi d\rho + \rho \cos \phi d\phi) \otimes (\sin \phi d\rho + \rho \cos \phi d\phi) + dz \otimes dz \\ &= (\cos^2 \phi + \sin^2 \phi) d\rho \otimes d\rho + \rho(\cos \phi \sin \phi - \sin \phi \cos \phi)(d\rho \otimes d\phi + d\phi \otimes d\rho) \\ &\quad + \rho^2(\cos^2 \phi + \sin^2 \phi) d\phi \otimes d\phi + dz \otimes dz \\ &= \underbrace{1}_{g_{\rho\rho}} d\rho \otimes d\rho + \underbrace{\rho^2}_{g_{\phi\phi}} d\phi \otimes d\phi + \underbrace{1}_{g_{zz}} dz \otimes dz. \end{aligned}$$

Thus the coordinate frame is orthogonal (since mutual inner products vanish, i.e., the metric component matrix is diagonal): $e_\rho = \partial/\partial\rho$ and $e_z = \partial/\partial z$ are in fact unit vector fields, while $e_\phi \partial/\partial\phi$ has length ρ , making $\rho^{-1}\partial/\partial\phi$ a unit vector. Thus

$$\{e_{\hat{\rho}}, e_{\hat{\phi}}, e_{\hat{z}}\} = \left\{ \frac{\partial}{\partial\rho}, \frac{1}{\rho} \frac{\partial}{\partial\phi}, \frac{\partial}{\partial z} \right\}$$

is an orthonormal frame naturally associated with cylindrical coordinates, with dual frame

$$\{\omega^{\hat{\rho}}, \omega^{\hat{\phi}}, \omega^{\hat{z}}\} = \{d\rho, \rho d\phi, dz\}.$$

Note that the cross product of the first such frame vector with the second equals the third (which can be evaluated component-wise using the corresponding Cartesian component triplet vectors, or geometrically from Fig. 5.17 representing the orthogonal coordinate lines), making this a right handed frame like the original Cartesian coordinate frame: $e_{\hat{\rho}} \times e_{\hat{\phi}} = e_{\hat{z}}$. This is reflected in the fact that $\det \underline{A} > 0$.

The two matrices \underline{A} and \underline{A}^{-1} expressed in terms of both the old and new coordinates (four matrices in all) may be used to transform any tensor field from one coordinate system to the other. For example, to transform the components of a vector field from old to new coordinates, one must a) re-express the old coordinates as functions of the new coordinates in the component functions, and b) change from old to new frame components, which requires the Jacobian matrix

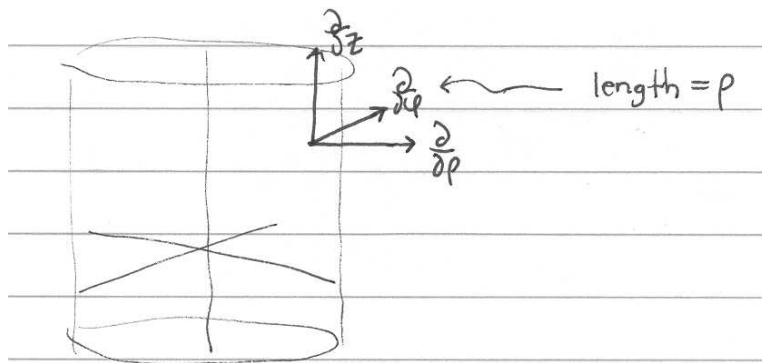


Figure 5.17: The coordinate frame vectors for cylindrical coordinates. Only ∂_ϕ is not a unit vector, having length equal to ρ . The cross product of $\partial/\partial\rho$ with $\partial/\partial\phi$ is along $\partial/\partial z$, making the ordering (r, ϕ, z) a right handed coordinate system.

to be expressed in terms of the new coordinates

$$\begin{aligned}
 X^{i'}(x') &= \underbrace{\frac{\partial x^{i'}}{\partial x^j}(x(x'))}_{b)} \underbrace{X^j(x(x'))}_{a)} \\
 X_{i'}(x') &= \underbrace{X_j(x(x'))}_{a)} \underbrace{\frac{\partial x^j}{\partial x^{i'}}(x')}_{b)}.
 \end{aligned}$$

In traditional “old fashioned” tensor analysis, tensors are defined by the “transformation law” under changes of coordinates which follows from these two basic relations, suppressing coordinate dependence of the component functions and the Jacobian matrices

$$T^{i' \dots}_{j' \dots} = \frac{\partial x^{i'}}{\partial x^m} \dots \frac{\partial x^n}{\partial x^{j'}} \dots T^{m \dots}_{n \dots}.$$

Example 5.7.1. transforming a vector field and 1-form

The vector field $X = y \partial/\partial x + x \partial/\partial y$ has components

$$(X^i) = \begin{pmatrix} y \\ x \\ 0 \end{pmatrix} = \begin{pmatrix} \rho \sin \phi \\ \rho \cos \phi \\ 0 \end{pmatrix}$$

so its new components are

$$\begin{aligned}(X^{i'}) &= \begin{pmatrix} X^\rho \\ X^\phi \\ X^z \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi / \rho & \cos \phi / \rho & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \rho \sin \phi \\ \rho \cos \phi \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 2\rho \sin \phi \cos \phi \\ \cos^2 \phi - \sin^2 \phi \\ 0 \end{pmatrix} = \begin{pmatrix} \rho \sin 2\phi \\ \cos 2\phi \\ 0 \end{pmatrix}\end{aligned}$$

so

$$X = \rho \sin 2\phi \frac{\partial}{\partial \rho} + \cos 2\phi \frac{\partial}{\partial \phi}.$$

Similarly $X^\flat = y dx + x dy$ can be transformed

$$\begin{aligned}(X_i') &= (X_j A^{-1j}_i) = (\rho \sin \phi \quad \rho \cos \phi \quad 0) \begin{pmatrix} \cos \phi & -\rho \sin \phi & 0 \\ \sin \phi & \rho \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= (2\rho \sin \phi \cos \phi \quad \rho^2(\cos^2 \phi - \sin^2 \phi) \quad 0) = (\rho \sin 2\phi \quad \rho^2 \cos 2\phi \quad 0)\end{aligned}$$

so

$$X^\flat = \rho \sin 2\phi d\rho + \rho^2 \cos 2\phi d\phi.$$

The same result could have been obtained using the re-expressed metric to lower the indices of the vector field X . Since the frame is orthogonal, index lowering reduces to multiplication of each vector component by the corresponding diagonal metric component. Similarly index raising simply divides each 1-form component by that diagonal metric component.

Notice that since these had no z components, nor depended on z , this was really just a polar coordinate problem in the plane.

□

Exercise 5.7.1.

transforming a vector field and 1-form

Repeat this for $X = x\partial_x + y\partial_y$.

■

We can also just re-express the old coordinate frame vector fields in terms of the new ones using

$$\frac{\partial}{\partial x^i} = A^j_i \frac{\partial}{\partial x^{j'}},$$

namely

$$\frac{\partial}{\partial x} = \cos \phi \frac{\partial}{\partial \rho} - \rho^{-1} \sin \phi \frac{\partial}{\partial \phi}, \quad \frac{\partial}{\partial y} = \sin \phi \frac{\partial}{\partial \rho} + \rho^{-1} \cos \phi \frac{\partial}{\partial \phi}, \quad \frac{\partial}{\partial z} = \frac{\partial}{\partial z}.$$

To transform all the way to the orthonormal components associated with cylindrical coordinates, one must divide the rows of \underline{A} (which are the Cartesian coordinate components of the new dual 1-forms) and the columns of \underline{A}^{-1} (which are the Cartesian components of the new frame vector fields) by their lengths as vectors in \mathbb{R}^3 , leading to

$$\begin{pmatrix} \omega^x \\ \omega^y \\ \omega^z \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \omega^{\hat{\rho}} \\ \omega^{\hat{\phi}} \\ \omega^{\hat{z}} \end{pmatrix} = \underline{A} \begin{pmatrix} \omega^{\hat{\rho}} \\ \omega^{\hat{\phi}} \\ \omega^{\hat{z}} \end{pmatrix}$$

$$(e_{\hat{\rho}} \ e_{\hat{\phi}} \ e_{\hat{z}}) = (e_x \ e_y \ e_z) \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} = (e_x \ e_y \ e_z) \underline{A}^{-1}$$

Notice that the matrix which transforms from the orthonormal Cartesian coordinate frame to the orthonormal cylindrical coordinate frame is a rotation matrix representing a rotation by the angle ϕ in the horizontal plane of each tangent space taking e_x, e_y to $e_{\hat{\rho}}, e_{\hat{\phi}}$.

Exercise 5.7.2.

Laplacian in cylindrical coordinates

Evaluate the 2nd order linear differential operator

$$\nabla^2 = \delta^{ij} \frac{\partial^2}{\partial x^i \partial x^j} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

by substituting the above expressions for the Cartesian coordinate vector fields. This operator called the Laplacian plays a very important role in many physically interesting partial differential equations.

Show that the result can be re-expressed in the form

$$\nabla^2 = \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2} = \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}.$$

This is really just the Laplacian expressed in polar coordinate in the plane, with the extra z term, and for a brute force evaluation like this, a computer algebra system is better than a hand calculation. Later we will understand this in a much better way. ■

Summary of what we did for cylindrical coordinates

$$\begin{aligned} (1) \quad dx^{i'} &= \frac{\partial x^{i'}}{\partial x^j}(x) dx^j = A^i_{j'}(x) dx^j \\ (2) \quad \frac{\partial}{\partial x^{i'}} &= \frac{\partial x^j}{\partial x^{i'}}(x'(x)) \frac{\partial}{\partial x^j} = A^{-1j}_{i'}(x'(x)) \frac{\partial}{\partial x^j} \\ (3) \quad dx^i &= \frac{\partial x^i}{\partial x^{j'}}(x') dx^{j'} = A^{-1i}_{j'}(x') dx^{j'} \\ (4) \quad \frac{\partial}{\partial x^i} &= \frac{\partial x^{j'}}{\partial x^i}(x(x')) \frac{\partial}{\partial x^{j'}} = A^j_{i'}(x(x')) \frac{\partial}{\partial x^{j'}} \end{aligned}$$

In words, we defined the new coordinate frame entirely in terms of the old Cartesian coordinates by taking the differential (1) of the coordinate map $\Phi(x) = (x^1(x), x^2(x), x^3(x))$ to yield the new coordinate dual frame in terms of the Cartesian coordinates and the Jacobian matrix $A^i_j(x)$ which can be inverted to give $A^{-1i}_j(x'(x))$ and Cartesian coordinate expressions (2) for the new coordinate vector fields.

Then substitution of parametrization map $\Psi(x') = (x^1(x'), x^2(x'), x^3(x'))$ into these two matrices re-expresses them in terms of the new coordinates which may then be used to represent the old coordinate frame (4) and dual frame (3) in terms of the new coordinates.

(5) Alternatively one can take the differential (3) of the parametrization map to directly yield $A^{-1i}_j(x')$ which can be inverted to get $A^i_j(x(x'))$ expressing the old coordinate frame and dual frame in terms of the new coordinates. Using the coordinate map to re-express the Jacobian matrix, one could then represent the new coordinate frame and dual frame in terms of the old coordinates.

(6) Then we re-expressed the Euclidean metric

$$\begin{aligned} g &= \delta_{ij} dx^i \otimes dx^j \\ &= \delta_{ij} \frac{\partial x^i}{\partial x^{m'}} \frac{\partial x^j}{\partial x^{n'}} dx^{m'} \otimes dx^{n'} \\ &= g_{m'n'} dx^{m'} \otimes dx^{n'} \end{aligned}$$

either by substituting the differential (3) of the parametrization map into the metric g or by using the equivalent matrix transformation of its components

$$\underline{g'} = \underline{A}^{-1}(x')^T \underline{I} \underline{A}^{-1}(x') \equiv (g_{m'n'})$$

(7) Then we evaluated $\nabla^2 = \delta^{ij} \partial^2 / \partial x^i \partial x^j = \dots$

Exercise 5.7.3.

paracylindrical coordinates

Many other useful orthogonal coordinate systems exist in the plane and in ordinary space, built on the geometry of other interesting curves, many of which are conics. For example, generalized cylindrical coordinates exist which keep the Cartesian coordinate z like ordinary cylindrical coordinates, and replace the polar coordinates in the x - y plane by another system of coordinates. A simple example of these are paracylindrical coordinates, where “parabolic coordinates” in the plane are defined by

$$x = \frac{1}{2}(u^2 - v^2), \quad y = uv.$$

The coordinate lines for both coordinates are families of parabolas with a common symmetry axis on which all the vertices lie, namely the y axis. Since X is really a vector field in the x - y plane, one can convert it to paracylindrical coordinates in the same way it would be converted to parabolic coordinates in that plane alone. Do this for both X and X^b .

a) Evaluate the 2-dimensional matrices \underline{A} and \underline{A}^{-1} expressed in terms of both old and new coordinates for the corresponding coordinates in the plane.

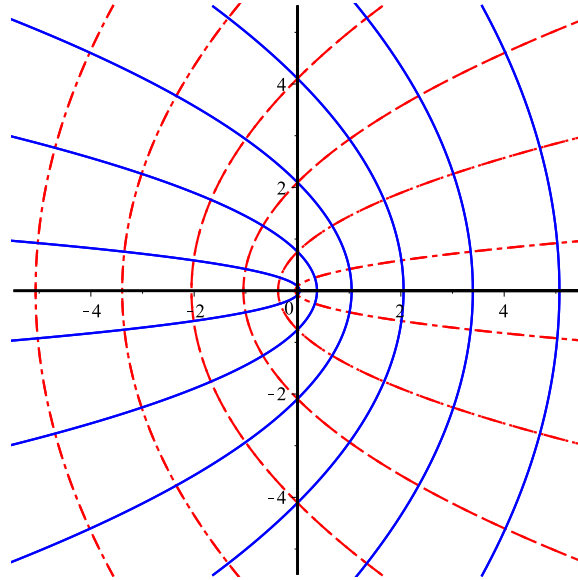


Figure 5.18: One passes from Cartesian coordinates in the x - y plane to parabolic coordinates (u, v) by introducing coordinates built from two families of parabolas with a common symmetry axis and the same foci. The u coordinate lines open to the right, while the v coordinate lines open to the left.

b) Evaluate the Euclidean metric tensor field in the new coordinates on the plane, showing that this is an orthogonal coordinate system.

c) Evaluate the the 2-dimensional matrices $\underline{\mathcal{A}}$ and $\underline{\mathcal{A}}^{-1}$ of components of the normalized orthogonal coordinate frame vectors and dual frame 1-forms expressed in terms of both old and new coordinates for the corresponding coordinates in the plane.

■

5.8 Spherical coordinate frames

To keep you on your toes (it is not good to be married to a single notation, versatility is key in this subject), we switch from primed index to barred variable notation for new coordinates: $x^{i'} \rightarrow \bar{x}^i$.

We now repeat step (5) for spherical coordinates, namely we differentiate the parametrization map to get the Jacobian matrix $\underline{A}^{-1}(x(\bar{x}))$

$$\begin{aligned}\underline{\mathbf{r}} &= \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r \sin \theta \cos \phi \\ r \sin \theta \sin \phi \\ r \cos \theta \end{pmatrix} \\ \underline{A}^{-1}(x(\bar{x})) &= \left(\frac{\partial x^i}{\partial \bar{x}^j}(\bar{x}) \right) = \begin{pmatrix} \partial x / \partial r & \partial x / \partial \theta & \partial x / \partial \phi \\ \partial y / \partial r & \partial y / \partial \theta & \partial y / \partial \phi \\ \partial z / \partial r & \partial z / \partial \theta & \partial z / \partial \phi \end{pmatrix} \\ &= \begin{pmatrix} \sin \theta \cos \phi & r \cos \theta \cos \phi & -r \sin \theta \sin \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix}\end{aligned}$$

The first column is just $\hat{\mathbf{r}} = r^{-1}\underline{\mathbf{r}}$, corresponding to the vector $(x^2 + y^2 + z^2)^{-1/2}\langle x, y, z \rangle$. The first two entries of the second column correspond to $\cot \theta = z/\rho = z/(x^2 + y^2)^{1/2}$ times $\langle x, y, 0 \rangle$ (see Fig. 5.14 for the trig ratio), while the last entry is just $-\rho$ (obvious from the same figure). The last column obviously corresponds to the vector $\langle -y, x, 0 \rangle$. Thus re-expressing the Jacobian matrix in terms of the old coordinates gives

$$\underline{A}^{-1}(x) = \left(\frac{\partial x^i}{\partial \bar{x}^j}(\bar{x}(x)) \right) = \begin{pmatrix} \frac{x}{(x^2 + y^2 + z^2)^{1/2}} & \frac{xz}{(x^2 + y^2)^{1/2}} & -y \\ \frac{y}{(x^2 + y^2 + z^2)^{1/2}} & \frac{yz}{(x^2 + y^2)^{1/2}} & x \\ \frac{z}{(x^2 + y^2 + z^2)^{1/2}} & \frac{-(x^2 + y^2)}{(x^2 + y^2)^{1/2}} & 0 \end{pmatrix}.$$

On the other hand repeating step (1) of differentiating the coordinate map

$$\begin{pmatrix} r \\ \theta \\ \phi \end{pmatrix} = \begin{pmatrix} (x^2 + y^2 + z^2)^{1/2} \\ \cos^{-1} \left(\frac{z}{(x^2 + y^2 + z^2)^{1/2}} \right) \\ \tan^{-1} \frac{y}{x} + C \end{pmatrix}$$

will lead to the inverse Jacobian. In carrying out this step, one needs the derivative formulas

$$\frac{d}{du} \cos^{-1} u = -\frac{1}{\sqrt{1-u^2}}, \quad \frac{d}{du} \tan^{-1} u = \frac{1}{1+u^2},$$

together with the quotient rule and some simplification, finally obtaining the result

$$\begin{aligned} \underline{A}(x) &= \left(\frac{\partial \bar{x}^i}{\partial x^j}(x) \right) = \begin{pmatrix} \partial r / \partial x & \partial r / \partial y & \partial r / \partial z \\ \partial \theta / \partial x & \partial \theta / \partial y & \partial \theta / \partial z \\ \partial \phi / \partial x & \partial \phi / \partial y & \partial \phi / \partial z \end{pmatrix} \\ &= \begin{pmatrix} \frac{x}{(x^2 + y^2 + z^2)^{1/2}} & \frac{y}{(x^2 + y^2 + z^2)^{1/2}} & \frac{z}{(x^2 + y^2 + z^2)^{1/2}} \\ \frac{xz}{(x^2 + y^2 + z^2)(x^2 + y^2)^{1/2}} & \frac{yz}{(x^2 + y^2 + z^2)(x^2 + y^2)^{1/2}} & \frac{-(x^2 + y^2)}{(x^2 + y^2 + z^2)(x^2 + y^2)^{1/2}} \\ -\frac{y}{x^2 + y^2} & \frac{x}{x^2 + y^2} & 0 \end{pmatrix}. \end{aligned}$$

Finally to re-express this in terms of the spherical coordinates one notices that the first row corresponds to the vector $\langle x, y, z \rangle / r$, the first two entries of the second row are just $z(x, y) / (r^2 \rho) = (r \cos \theta)(x, y) / (r^3 \sin \theta)$ while the last entry is $-\rho / r^2 = -r \sin \theta / r^2$, and finally the third row corresponds to $(-y, x) / (r^2 \sin^2 \theta)$. Finishing the details leads to the result

$$\underline{A}(\bar{x}) = \left(\frac{\partial \bar{x}^i}{\partial x^j}(x(\bar{x})) \right) = \begin{pmatrix} \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\ \frac{1}{r} \cos \theta \cos \phi & \frac{1}{r} \cos \theta \sin \phi & -\frac{1}{r} \sin \theta \\ -\frac{\sin \phi}{r \sin \theta} & \frac{\cos \phi}{r \sin \theta} & 0 \end{pmatrix}.$$

Exercise 5.8.1.

Jacobian matrices for spherical coordinates

Check all the details in the previous calculations of the two Jacobian matrices. ■

The matrix

$$\underline{A}^{-1}(\bar{x}) = \begin{pmatrix} \frac{\partial x^i}{\partial r} & \frac{\partial x^i}{\partial \theta} & \frac{\partial x^i}{\partial \phi} \end{pmatrix}$$

has as its columns the Cartesian coordinate components of the tangents to the new coordinate lines parametrized by those coordinates. The first column are the old fashioned components of the tangent vector of the curve which results from holding θ and ϕ fixed and varying r , for example. Since the new coordinate system is orthogonal, these three tangent vectors are orthogonal as one can verify by taking dot products of the corresponding vectors in \mathbb{R}^3 . In fact

$$\underline{g} = [\underline{A}^{-1}(\bar{x})]^T \underline{A}^{-1}(\bar{x})$$

is the matrix of all possible inner products of these vectors, namely the coordinate components of the Euclidean metric. By orthogonality this matrix will be diagonal. The diagonal elements will be the self-dot products of the three tangent vectors, i.e., the lengths of the three column matrices thought of as vectors in \mathbb{R}^3 .

If we divide each column by its lengths, the new columns will be orthonormal. A square matrix whose columns are mutually orthogonal unit vectors in \mathbb{R}^n are called orthogonal matrices.

Exercise 5.8.2.**spherical coordinate frame rotation**

Show that dividing the rows of \underline{A} (whose entries are the Cartesian components of the differential of the corresponding new coordinate) by their lengths as vectors in \mathbb{R}^3 yields the matrix

$$\underline{A} = \begin{pmatrix} \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\ \cos \theta \cos \phi & \cos \theta \sin \phi & -\sin \theta \\ -\sin \phi & \cos \phi & 0 \end{pmatrix}$$

and show that it is orthogonal by verifying that $\underline{A}^T \underline{A} = \underline{I}$. This matrix takes the orthonormal Cartesian coordinate frame to the orthonormal frame associated with spherical coordinates and represents a rotation.

Show that this rotation $\underline{A} = \underline{A}_3 \underline{A}_2 \underline{A}_1$ is the product of two simple rotations and a swap: first a rotation in the x - y plane of the tangent space by the azimuthal angle ϕ from x towards y , which takes e_x, e_y to $e_{\hat{\rho}}, e_{\hat{\phi}}$, then followed by a rotation in the ρ - z plane of the tangent space by the polar angle θ from z towards ρ which rotates the vertical direction to the radial direction taking $e_z, e_{\hat{\rho}}$ to $e_{\hat{r}}, e_{\hat{\theta}}$

$$\underbrace{\begin{pmatrix} d\rho \\ \rho d\phi \\ dz \end{pmatrix}}_{\begin{pmatrix} \omega^{\hat{\rho}} \\ \omega^{\hat{\phi}} \\ \omega^z \end{pmatrix}} = \underbrace{\begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\underline{A}_1} \underbrace{\begin{pmatrix} dx \\ dy \\ dz \end{pmatrix}}_{\begin{pmatrix} \omega^x \\ \omega^y \\ \omega^z \end{pmatrix}}, \quad \underbrace{\begin{pmatrix} dr \\ r \sin \theta d\phi \\ r d\theta \end{pmatrix}}_{\begin{pmatrix} \omega^{\hat{r}} \\ \omega^{\hat{\phi}} \\ \omega^{\hat{\theta}} \end{pmatrix}} = \underbrace{\begin{pmatrix} \sin \theta & 0 & \cos \theta \\ 0 & 1 & 0 \\ \cos \theta & 0 & -\sin \theta \end{pmatrix}}_{\underline{A}_2} \begin{pmatrix} d\rho \\ \rho d\rho \\ dz \end{pmatrix},$$

and finally an exchange of the two angular directions $e_{\hat{\phi}}, e_{\hat{\theta}}$ to make the final frame ordering $\{e_{\hat{r}}, e_{\hat{\theta}}, e_{\hat{\phi}}\}$ right handed (since $\det \underline{A}_2 = -1 = \underline{A}_3$ while $\det \underline{A}_1 = 1$, one has $\det \underline{A} = 1$)

$$\begin{pmatrix} dr \\ r d\theta \\ r \sin \theta d\phi \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}}_{\underline{A}_3} \begin{pmatrix} dr \\ r \sin \theta d\phi \\ r d\theta \end{pmatrix}.$$

■

The rows of $\underline{A}^{-1}(\bar{x})$ are the new components of the differentials of the old coordinates

$$dx^i = \frac{\partial x^i}{\partial \bar{x}^j} d\bar{x}^j : \quad \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} = \underline{A}^{-1}(\bar{x}) \begin{pmatrix} dr \\ d\theta \\ d\phi \end{pmatrix} = \begin{pmatrix} \sin \theta \cos \phi dr + r \cos \theta \cos \phi d\phi - r \sin \theta \sin \phi d\phi \\ \sin \theta \sin \phi dr + r \cos \theta \sin \phi d\phi + r \sin \theta \cos \phi d\phi \\ \cos \theta dr - r \sin \theta d\theta \end{pmatrix}.$$

The columns of $\underline{A}(\bar{x})$ are the new components of the old coordinate frame vectors

$$\frac{\partial}{\partial x^i} = \frac{\partial \bar{x}^j}{\partial x^i} \frac{\partial}{\partial \bar{x}^j} : \quad \begin{pmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial r} & \frac{\partial}{\partial \theta} & \frac{\partial}{\partial \phi} \end{pmatrix} \underline{A}(\bar{x}).$$

These, together with the parametrization map $x^i = x^i(\bar{x})$ are needed to transform the components of tensor fields.

For a vector field the matrix form of the transformation from old to new components is accomplished by left multiplication of the column matrix of components by the matrix \underline{A}^{-1}

$$X = X^i \frac{\partial}{\partial x^i} = \bar{X}^i \frac{\partial}{\partial \bar{x}^i} : \quad \bar{X}^i(\bar{x}) = \frac{\partial \bar{x}^i}{\partial x^j}(\bar{x}) X^j(x(\bar{x})) \longleftrightarrow \begin{pmatrix} X^r \\ X^\theta \\ X^\phi \end{pmatrix} = \underline{A}(\bar{x}) \begin{pmatrix} X^1 \\ X^2 \\ X^3 \end{pmatrix}.$$

For a covector field the matrix form of the transformation from old to new components is instead accomplished by right multiplication of the row matrix of components by the matrix \underline{A} , re-expressing everything in terms of the new coordinates

$$X^\flat = X_i dx^i = \bar{X}_i d\bar{x}^i : \\ \bar{X}_i(\bar{x}) = \frac{\partial x^j}{\partial \bar{x}^i}(\bar{x}) X_j(x(\bar{x})) \longleftrightarrow (X_r \ X_\theta \ X_\phi) = (X_1 \ X_2 \ X_3) \underline{A}^{-1}(\bar{x}).$$

Example 5.8.1. transformation and index shifting

We used matrix methods to express the vector field and 1-form of Exercise 5.7.1 in spherical coordinates, first re-expressing the matrix of Cartesian components in the new coordinates and then left multiplying by the appropriate Jacobian matrix.

$$X = y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} \longleftrightarrow \begin{pmatrix} X^1 \\ X^2 \\ X^3 \end{pmatrix} = \begin{pmatrix} y \\ x \\ 0 \end{pmatrix} = \begin{pmatrix} r \sin \theta \sin \phi \\ r \sin \theta \cos \phi \\ 0 \end{pmatrix} \\ \begin{pmatrix} X^r \\ X^\theta \\ X^\phi \end{pmatrix} = \begin{pmatrix} \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\ \frac{1}{r} \cos \theta \cos \phi & \frac{1}{r} \cos \theta \sin \phi & -\frac{1}{r} \sin \theta \\ -\frac{\sin \phi}{r \sin \theta} & \frac{\cos \phi}{r \sin \theta} & 0 \end{pmatrix} \begin{pmatrix} r \sin \theta \sin \phi \\ r \sin \theta \cos \phi \\ 0 \end{pmatrix} \\ = \begin{pmatrix} r \sin^2 \theta \cos \phi \sin \phi + r \sin^2 \theta \sin \phi \cos \phi \\ \sin \theta \cos \theta \sin \phi \cos \phi + \sin \theta \cos \theta \sin \phi \cos \phi \\ -\sin^2 \phi + \cos^2 \phi \end{pmatrix} = \begin{pmatrix} r \sin^2 \theta \sin 2\phi \\ \sin \theta \cos \theta \sin 2\phi \\ \cos 2\phi \end{pmatrix}$$

so

$$X = \sin \theta \sin 2\phi \left(r \sin \theta \frac{\partial}{\partial r} + \cos \theta \frac{\partial}{\partial \theta} \right) + \cos 2\phi \frac{\partial}{\partial \phi},$$

so that lowering the index leads to

$$X^\flat = \sin \theta \sin 2\phi (r \sin \theta dr + r^2 \cos \theta d\theta) + r^2 \sin^2 \theta \cos 2\phi d\phi.$$

Now transforming the corresponding 1-form

$$\begin{aligned} X^\flat = y dx + x dy &\longleftrightarrow (X_1 \ X_2 \ X_3) = (y \ x \ 0) = (r \sin \theta \sin \phi \ r \sin \theta \cos \phi \ 0), \\ (X^r \ X^\theta \ X^\phi) &= (r \sin \theta \sin \phi \ r \sin \theta \cos \phi \ 0) \begin{pmatrix} \sin \theta \cos \phi & r \cos \theta \cos \phi & -r \sin \theta \sin \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix} \\ &= \begin{pmatrix} r \sin^2 \theta \sin \phi \cos \phi + r \sin^2 \theta \sin \phi \cos \phi & r^2 \sin \theta \cos \phi \sin \phi \cos \phi + r^2 \sin \theta \cos \phi \sin \phi \cos \phi \\ & -r^2 \sin^2 \theta \sin^2 \phi + r^2 \sin^2 \theta \cos^2 \phi \end{pmatrix} \\ &= (r \sin^2 \theta \sin 2\phi \ r^2 \sin \theta \cos \phi \sin 2\phi \ r^2 \sin^2 \theta \cos 2\phi), \end{aligned}$$

where the 3×1 row matrix above is too long for its three entries to fit on one row, so we obtain the previous expression derived above, here factored

$$X^\flat = r \sin \theta [\sin 2\phi (\sin \theta dr + r \cos \theta d\theta) + r \cos 2\phi \sin \theta d\phi].$$

□

Exercise 5.8.3.

differential, gradient in cylindrical, spherical coordinates

EDIT THIS.

Consider the function

$$f = xy = \rho^2 \sin \phi \cos \phi = \frac{1}{2} \rho^2 \sin 2\phi = \frac{1}{2} r^2 \sin^2 \theta \sin 2\phi.$$

Then

$$\begin{aligned} df &= y dx + x dy = X^\flat \\ \vec{\nabla} f &= [df]^\sharp = y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} = X \end{aligned}$$

yields our friend X from previous exercises where we saw that

$$\begin{aligned} X &= \rho \sin 2\phi \frac{\partial}{\partial \rho} + \cos 2\phi \frac{\partial}{\partial \phi} = \sin \theta \sin 2\phi \left(r \sin \theta \frac{\partial}{\partial r} + \cos \theta \frac{\partial}{\partial \theta} \right) + \cos 2\phi \frac{\partial}{\partial \phi} \\ X^\flat &= \rho \sin 2\phi d\rho + \rho^2 \sin 2\phi d\phi = \sin \theta \sin 2\phi (r \sin \theta dr + r^2 \cos \theta d\theta) + r^2 \sin^2 \theta \cos 2\phi d\phi \end{aligned}$$

$$\left[\frac{\partial}{\partial r} \right]_i = g_{ij} \left[\frac{\partial}{\partial r} \right]^j = g_{ir} \longrightarrow \left[\frac{\partial}{\partial r} \right]^\flat = g_{ir} d\bar{x}^i = g_{rr} dr = dr$$

and similarly

$$\begin{aligned}\left[\frac{\partial}{\partial\phi}\right]^b &= g_{\phi i} d\bar{x}^i = g_{\phi\phi} d\phi = r^2 \sin^2 \theta d\phi \\ \left[\frac{\partial}{\partial\theta}\right]^b &= g_{\theta i} d\bar{x}^i = g_{\theta\theta} d\theta = r^2 d\theta\end{aligned}$$

In general

$$e_i{}^b = g_{kj} e^j{}_i \omega^k = g_{ik} \omega^k$$

so that

$$X^b = (X^i e_i)^b = X^i e_i{}^b = X^i g_{ik} \omega^k = X_k \omega^k.$$

Similarly

$$[\omega^i]^\sharp = g^{ij} e_j$$

holds for an orthogonal frame, index shifting the frame vectors and dual frame covectors yields the corresponding basis covector or vector multiplied by the diagonal metric component or its reciprocal.

Compute df and $\text{grad } f = \vec{\nabla} f$ in cylindrical and then spherical coordinates and verify that you get our previous results quoted above. ■

Similarly to transform the metric one can evaluate the differentials and expand their products

$$\begin{aligned}g &= \delta_{ij} dx^i \otimes dx^j \\ &= (\sin \theta \cos \phi dr + \dots) \otimes (\sin \theta \cos \phi dr + \dots) + \dots, \text{ etc.},\end{aligned}$$

but matrix methods are more efficient

$$\begin{aligned}(\bar{g}_{ij}(\bar{x})) &= \left(\frac{\partial x^m}{\partial \bar{x}^i}(\bar{x}) \delta_{mn} \frac{\partial x^n}{\partial \bar{x}^j}(\bar{x}) \right) = \underline{A}^{-1}(\bar{x})^T \underline{A}^{-1}(\bar{x}) \\ &= \begin{pmatrix} \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\ r \cos \theta \cos \phi & r \cos \theta \sin \phi & -r \sin \theta \\ -r \sin \theta \sin \phi & r \sin \theta \cos \phi & 0 \end{pmatrix} \begin{pmatrix} \sin \theta \cos \phi & r \cos \theta \cos \phi & -r \sin \theta \sin \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix} \\ &= \dots \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2 \theta \end{pmatrix}\end{aligned}$$

so

$$g = \underbrace{1}_{g_{rr}} dr \otimes dr + \underbrace{r^2}_{g_{\theta\theta}} d\theta \otimes d\theta + \underbrace{r^2}_{g_{\phi\phi}} \sin^2 \theta d\phi \otimes d\phi.$$

To understand why this simple result must come out from the matrix multiplication, notice that the first column of the right matrix are just the Cartesian components of the unit radial

vector, while the second column is a Cartesian component vector with length r (pointing along the ϕ coordinate circles), and the final one with length $r \sin \theta$ (pointing along the θ coordinate circles). Multiplying by the transpose yields the square of these factors along the diagonal, and the obvious orthogonality of these three Cartesian vectors leads to the off-diagonal entries of the matrix product all vanishing.

The square root of the determinant of the diagonal metric matrix is then

$$|\det \underline{g}|^{1/2} = (g_{rr}g_{\theta\theta}g_{\phi\phi})^{1/2} = r^2 \sin \theta$$

and hence the unit volume element is

$$\eta = |\det \underline{g}|^{1/2} dr \wedge d\theta \wedge d\phi = r^2 \sin \theta dr \wedge d\theta \wedge d\phi$$

while the contravariant metric is

$$g^{-1} = \frac{\partial}{\partial r} \otimes \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial}{\partial \theta} \otimes \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial}{\partial \phi} \otimes \frac{\partial}{\partial \phi}.$$

In the above example, we can now easily check that X^\flat is related to X by index lowering

$$X_r = g_{rr}X^r = X^r, \quad X_\theta = g_{\theta\theta}X^\theta = r^2(\sin \theta \cos \theta \sin 2\phi), \quad X_\phi = g_{\phi\phi}X^\phi = (r^2 \sin^2 \theta)(\cos 2\phi).$$

The matrices $\underline{A}(x)$ and $\underline{A}^{-1}(x)$ are necessary for transforming in the opposite direction, from spherical to Cartesian components. The rows of $\underline{A}(x)$ are the old components of the differentials of the new coordinates

$$\begin{pmatrix} dr \\ d\theta \\ d\phi \end{pmatrix} = \underline{A}(x) \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} = \cdots = \begin{pmatrix} xdx + ydy + zdz \\ z(xdx + ydy) - (x^2 + y^2)dz \\ -ydx + xdy \end{pmatrix} \equiv \begin{pmatrix} \omega^r \\ \omega^\theta \\ \omega^\phi \end{pmatrix}$$

The columns of $\underline{A}^{-1}(x)$ are old components of the new coordinate frame vector fields (we use the simpler partial derivative notation $\partial/\partial r = \partial_r$, etc.)

$$\begin{aligned} (\partial_r \quad \partial_\theta \quad \partial_\phi) &= (\partial_x \quad \partial_y \quad \partial_z) \underline{A}^{-1}(x) = \cdots \\ &= \left(\frac{(x\partial_x + y\partial_y + z\partial_z)}{(x^2 + y^2 + z^2)^{1/2}} \quad \frac{z[x\partial_x + y\partial_y] - [x^2 + y^2]\partial_z}{(x^2 + y^2)^{1/2}} \quad -y\partial_x + x\partial_y \right) \equiv (e_r \quad e_\theta \quad e_\phi) \end{aligned}$$

Exercise 5.8.4.

spherical coordinates back to Cartesian coordinates

Re-express the following vector field in terms of Cartesian coordinates

$$Y = r \frac{\partial}{\partial r} + \frac{\partial}{\partial \phi}.$$

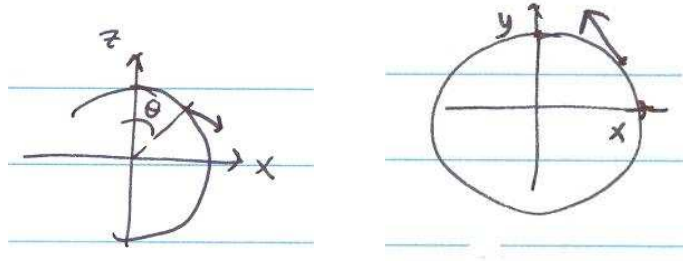


Figure 5.19: Left: the representation of ∂_θ for the case $y = 0, z > 0$. Right: the representation of ∂_ϕ for the case $y = 0, x > 0$.

If we take the expressions $\{e_r, e_\theta, e_\phi\}$ for the spherical coordinate frame vector fields in Cartesian coordinates and their dual frame covector fields $\{\omega^r, \omega^\theta, \omega^\phi\}$, we can use them as an orthogonal frame on \mathbb{R}^3 , forgetting about their representation in terms of spherical coordinates.

Clearly e_r is undefined at the origin and e_θ along the z -axis, while e_ϕ is zero on the z -axis, so the frame is only valid off the z -axis. We can normalize this orthogonal frame to an orthonormal frame by dividing by the lengths of the frame vector fields

$$e_{\hat{x}^i} \equiv (\bar{g}_{ii})^{-1/2} e_{\bar{x}^i}, \quad \omega^{\hat{x}^i} \equiv (\bar{g}_{ii})^{1/2} \omega^{\bar{x}^i},$$

namely

$$\{e_{\hat{r}}, e_{\hat{\theta}}, e_{\hat{\phi}}\} = \left\{ e_r, \frac{1}{(x^2 + y^2 + z^2)^{1/2}} e_\theta, \frac{1}{(x^2 + y^2)^{1/2}} e_\phi \right\} = \left\{ \frac{\partial}{\partial r}, \frac{1}{r} \frac{\partial}{\partial \theta}, \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi} \right\}.$$

Components in an orthonormal frame are called “physical components” since given the orientation of the frame at a point we can visualize a vector in terms of its orthonormal components. For example, for our vector field $X = y \partial/\partial x + x \partial/\partial y$ we have

$$X^{\hat{r}} = X_{\hat{r}} = r \sin^2 \theta \sin 2\phi$$

$$X^{\hat{\theta}} = X_{\hat{\theta}} = r \sin \theta \cos \theta \sin 2\phi$$

$$X^{\hat{\phi}} = X_{\hat{\phi}} = r \sin \theta \cos 2\phi$$

and

$$X = r \sin \theta [\sin 2\phi (\sin \theta e_{\hat{r}} + \cos \theta e_{\hat{\theta}}) + \cos 2\phi e_{\hat{\phi}}].$$

Note that

$$X \cdot X = \delta_{ij} X^i X^j = y^2 + x^2 = r^2 \sin^2 \theta,$$

so $\|X\| = r \sin \theta$, which is the factor outside the square brackets. We can also visualize the contents of the square brackets in terms of two successive rotations of $e_{\hat{\theta}}$, yielding a unit vector, as illustrated in Fig. 5.20.

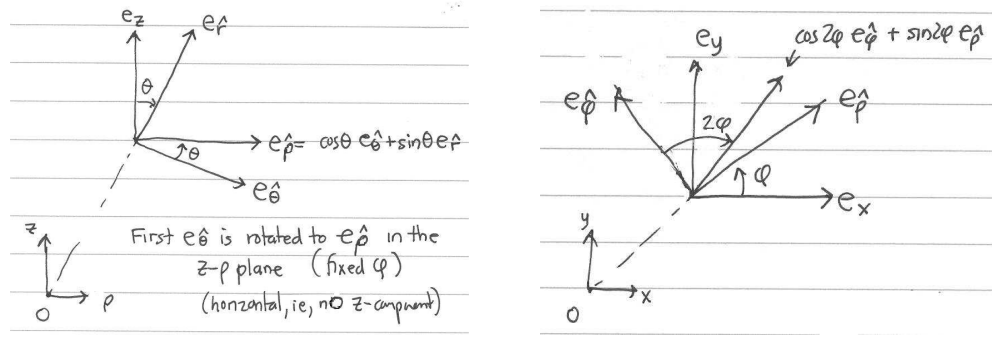


Figure 5.20: Left: First $e_{\hat{\theta}}$ is rotated to $e_{\hat{\rho}}$ in the ρ - z plane of fixed ϕ (horizontal, i.e., no z component). Right: Then $e_{\hat{\rho}}$ is rotated in the horizontal plane by an angle 2ϕ in the opposite direction. Scaling the result by $r \sin \theta$ gives the vector field X .

As an example of a tedious coordinate derivative calculation, consider evaluating the Laplacian in spherical coordinates by brute force as follows

$$\begin{aligned}
 \nabla^2 &= \left(\frac{\partial}{\partial x} \right)^2 + \left(\frac{\partial}{\partial y} \right)^2 + \left(\frac{\partial}{\partial z} \right)^2 \\
 &= \left(\sin \theta \cos \phi \frac{\partial}{\partial r} + \frac{\cos \theta \cos \phi}{r} \frac{\partial}{\partial \theta} - \frac{\sin \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \left(\sin \theta \cos \phi \frac{\partial}{\partial r} + \frac{\cos \theta \cos \phi}{r} \frac{\partial}{\partial \theta} - \frac{\sin \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \\
 &\quad + \left(\sin \theta \sin \phi \frac{\partial}{\partial r} + \frac{\cos \theta \sin \phi}{r} \frac{\partial}{\partial \theta} - \frac{\cos \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \left(\sin \theta \sin \phi \frac{\partial}{\partial r} + \frac{\cos \theta \sin \phi}{r} \frac{\partial}{\partial \theta} - \frac{\cos \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \\
 &\quad + \left(\cos \theta \frac{\partial}{\partial r} - \frac{\sin \theta}{r} \frac{\partial}{\partial \theta} \right) \left(\cos \theta \frac{\partial}{\partial r} - \frac{\sin \theta}{r} \frac{\partial}{\partial \theta} \right) = \dots \\
 &= \underbrace{(\sin^2 \theta \cos^2 \phi + \sin^2 \theta \sin^2 \phi + \cos^2 \theta)}_1 \frac{\partial^2}{\partial r^2} + r^{-2} \underbrace{(\cos^2 \theta \cos^2 \phi + \cos^2 \theta \sin^2 \phi + \sin^2 \theta)}_1 \frac{\partial^2}{\partial \theta^2} \\
 &\quad + (r^2 \sin^2 \theta)^{-1} \underbrace{(\sin^2 \phi + \cos^2 \phi)}_1 \frac{\partial^2}{\partial \phi^2} \\
 &\quad + 0 \frac{\partial^2}{\partial r \partial \theta} + 0 \frac{\partial^2}{\partial \theta \partial \phi} + 0 \frac{\partial^2}{\partial r \partial \phi} \quad [\text{due to orthogonality of coefficient vectors}] \\
 &\quad - r^{-2} \underbrace{[\sin \theta \cos \phi (\cos \theta \cos \phi) + \sin \theta \sin \phi (\cos \theta \sin \phi) + \cos \theta (-\sin \theta)]}_0 \frac{\partial}{\partial \theta} + \dots \\
 &\quad + r^{-1} \left[\cos \theta \cos \phi (\cos \theta \cos \phi) + \cos \theta \sin \phi (\cos \theta \sin \phi) - \sin \theta (-\sin \theta) \right. \\
 &\quad \quad \left. - \frac{\sin \phi}{\sin \theta} (-\sin \theta \sin \phi) + \frac{\cos \phi}{\sin \theta} (\sin \theta \cos \phi) \right] \frac{\partial}{\partial r} + \dots
 \end{aligned}$$

2

However, later we will see that there is a better way to do this, although it doesn't hurt to check some of the coefficients by this brute force method as an exercise. The result can be expressed as

$$\begin{aligned}\nabla^2 &= \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{\cot \theta}{r^2} \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \\ &= \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2}.\end{aligned}$$

We will return to this in Chapter 7.

Exercise 5.8.5.

spherical coordinate Laplacian

Check the steps and fill in the dots in the previous calculation.



5.9 Lie brackets and noncoordinate frames

While the Lie brackets of the orthogonal cylindrical and spherical coordinate frame vectors vanish like those of the Cartesian coordinate frame vectors, it is convenient to work with the associated orthonormal frame vectors whose components have a direct physical interpretation like Cartesian vector components. These noncoordinate frame vectors have nonvanishing Lie brackets, which turn out to be useful in evaluating curvature related quantities. The orthonormalized vectors differ from coordinate frame vectors only by a scalar factor, so it is useful to have a formula for the Lie brackets of two vector fields rescaled by function factors. This can then be easily extended by linearity to provide a formula for the frame components of the Lie bracket of any two vector fields.

Lie bracket product rule

When the vector fields in a Lie bracket are multiplied by scalar functions, the derivative product rule generates two extra terms besides the original Lie bracket multiplied by those functions. In the following calculation the parentheses are essential to remove the ambiguity of each expression

$$\begin{aligned}
 [fX, hY] &= fX(hY) - hY(fX) \\
 &= f(Xh)Y + fhXY - h(Yf)X - hfYX \\
 &= fh(XY - YX) + f(Xh)Y - h(Yf)X \\
 &= fh[X, Y] + f(Xh)Y - h(Yf)X.
 \end{aligned}$$

In the first line for example, $fX(hY)$ means that one first applies the operation hY to a function and then fX to the result. In the second line $f(Xh)Y = (Xh)fY$, the factor (Xh) is a function which simply multiplies fY ; the parentheses in this case serve to limit the action of the derivative to h alone. The result of this calculation is the sum of the linear term $fh[X, Y]$ in which neither scalar factor is differentiated plus the two terms which result when an outer derivative differentiates an inner scalar factor.

Example 5.9.1. We verify that

$$\left[\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta} \right] = [e_r, e_\theta] = 0$$

entirely in Cartesian coordinates. We use the simpler partial derivative notation $\partial_x = \partial/\partial x$, etc. This calculation is mildly tedious but certainly doable. We start by using the previous result extended (using the bilinearity of the Lie bracket) to sums of products of functions and vector fields, with the three respective terms simplified in three successive lines in each equality below. The first term simplifies considerably with the cancellation of six terms in the numerator.

$$\begin{aligned}
[e_r, e_\theta] &= \left[\frac{x\partial_x + y\partial_y + z\partial_z}{(x^2 + y^2 + z^2)^{1/2}}, \frac{z(x\partial_x + y\partial_y) - (x^2 + y^2)\partial_z}{(x^2 + y^2)^{1/2}} \right] \\
&= (x^2 + y^2 + z^2)^{-1/2} (x^2 + y^2)^{-1/2} [x\partial_x + y\partial_y + z\partial_z, z(x\partial_x + y\partial_y) - (x^2 + y^2)\partial_z] \\
&\quad + (x^2 + y^2 + z^2)^{-1/2} [(x\partial_x + y\partial_y + z\partial_z)(x^2 + y^2)^{-1/2}] [z(x\partial_x + y\partial_y) - (x^2 + y^2)\partial_z] \\
&\quad - (x^2 + y^2)^{-1/2} \{ [z(x\partial_x + y\partial_y) - (x^2 + y^2)\partial_z](x^2 + y^2 + z^2)^{-1/2} \} [x\partial_x + y\partial_y + z\partial_z] \\
&= \frac{x(z\partial_x - 2x\partial_z) + y(z\partial_y - 2y\partial_z) + z(x\partial_x + y\partial_y) - zx\partial_x - zy\partial_y + (x^2 + y^2)\partial_z}{(x^2 + y^2 + z^2)^{1/2}(x^2 + y^2)^{1/2}} \\
&\quad - \frac{1}{2} \frac{(2x^2 + 2y^2)[z(x\partial_x + y\partial_y) - (x^2 + y^2)\partial_z]}{(x^2 + y^2 + z^2)^{3/2}(x^2 + y^2)^{1/2}} \\
&\quad - \frac{1}{(x^2 + y^2)^{1/2}} \left(-\frac{1}{2(x^2 + y^2 + z^2)^{3/2}} \right) [zx(2x) + zy(2y) - (x^2 + y^2)(2z)] \\
&= \frac{z(x\partial_x + y\partial_y - (x^2 + y^2)\partial_z)}{(x^2 + y^2 + z^2)^{1/2}(x^2 + y^2)^{1/2}} \\
&\quad - \frac{z(x\partial_x + y\partial_y - (x^2 + y^2)\partial_z)}{(x^2 + y^2 + z^2)^{1/2}(x^2 + y^2)^{1/2}} \\
&\quad + 0 \\
&= 0.
\end{aligned}$$

□

Exercise 5.9.1.**spherical coordinate commutator using Cartesian coordinates**Try the easier calculation $[e_r, e_\phi] = 0$.

■

The Lie bracket product rule is easily extended to apply to linear combinations of vector fields with scalar function coefficients. This is necessary to evaluate the coordinate components of the Lie bracket. The passage from the first line to the second line below written explicitly

in the summation notation uses the bilinearity of the Lie bracket in each argument

$$\begin{aligned}
 \left[\underbrace{X^i}_{\substack{f \\ X}} \frac{\partial}{\partial x^i}, \underbrace{Y^j}_{\substack{h \\ Y}} \frac{\partial}{\partial x^j} \right] &= \left[\sum_{i=1}^n X^i \frac{\partial}{\partial x^i}, \sum_{j=1}^n Y^j \frac{\partial}{\partial x^j} \right] \\
 &= \sum_{i=1}^n \sum_{j=1}^n \left[X^i \frac{\partial}{\partial x^i}, Y^j \frac{\partial}{\partial x^j} \right] \\
 &= \sum_{i=1}^n \sum_{j=1}^n \left(X^i Y^j \left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right] + X^i \frac{\partial Y^j}{\partial x^i} \frac{\partial}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j} \frac{\partial}{\partial x^i} \right) \\
 &= X^i \frac{\partial Y^j}{\partial x^i} \frac{\partial}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j} \frac{\partial}{\partial x^i} \\
 &= (XY^j) \frac{\partial}{\partial x^j} - (YX^i) \frac{\partial}{\partial x^i} \\
 &= (XY^i - YX^i) \frac{\partial}{\partial x^i}.
 \end{aligned}$$

Thus the coordinate formula

$$[X, Y]^i = (XY^i - YX^i) = (Y^i_{,j} X^j - X^i_{,j} Y^j)$$

is a consequence of this basic rule for how Lie brackets behave when you stick function factors in them, where the comma partial derivative notation $f_{,i} = \partial f / \partial x^i$ will prove useful in the next chapter. This result can be extended to the component formula for the Lie bracket with respect to a general noncoordinate frame.

Given any frame $\{e_i\}$, define a set of “structure functions” for the frame in the following way. Each of the commutators $[e_i, e_j]$ is itself a vector field which may be expressed as a linear combination of $\{e_i\}$ so

$$[e_i, e_j] = C^k_{ij} e_k.$$

Clearly $C^k_{ij} = 0$ for $i = j$ and $C^k_{ij} = -C^k_{ji}$ for $i \neq j$, i.e., this object is antisymmetric in its lower indices. If we wish we can define a tensor having these components

$$C(e) = C^i_{jk} e_i \otimes \omega^j \otimes \omega^k = \frac{1}{2} C^i_{jk} e_i \otimes \omega^k,$$

but it changes if we change the frame, i.e., does not define the same tensor in every frame like δ^i_j does.

Exercise 5.9.2.

structure functions of cylindrical and spherical coordinates

It is enough to list $\{C^k_{ij}\}_{i < j}$ to specify all the independent structure functions. Do this for the normalized cylindrical and spherical coordinate frames, i.e.,

$$\{e_{\hat{r}}, e_{\hat{\phi}}, e_{\hat{z}}\}, \quad \{e_{\hat{r}}, e_{\hat{\theta}}, e_{\hat{\phi}}\},$$

■

Now the same calculation done above for the coordinate frame, but not setting to zero the self commutators, gives the component formula for the Lie bracket in a general frame where again we make explicit the summation notation to emphasize the role of linearity in this calculation

$$\begin{aligned}
 [X^i e_i, Y^j e_j] &= \left[\sum_{i=1}^n X^i e_i, \sum_{j=1}^n Y^j e_j \right] \\
 &= \sum_{i=1}^n \sum_{j=1}^n [X^i e_i, Y^j e_j] \\
 &= \sum_{i=1}^n \sum_{j=1}^n (X^i Y^j [e_i, e_j] + (X^i e_i Y^j) e_j - (Y^j e_j X^i) e_i) \\
 &= (XY^j) e_j - (YX^i) e_i + C^k_{ij} X^i Y^j e_k \\
 &= (XY^i) e_i - (YX^i) e_i + C^i_{jk} X^j Y^k e_i \\
 &= [XY^i - YX^i + C^i_{jk} X^j Y^k] e_i,
 \end{aligned}$$

so using the frame derivative notation $e_i f = f_{,i}$, we have the result

$$\begin{aligned}
 [X, Y]^i &= XY^i - YX^i + C^i_{jk} X^j Y^k \\
 &= Y^i_{,j} X^j - X^i_{,j} Y^j + C^i_{jk} X^j Y^k.
 \end{aligned}$$

This formula is just what is needed to evaluate Lie brackets of vector fields in the orthonormal frame associated with cylindrical and spherical coordinates.

Exercise 5.9.3.

Lie brackets in cylindrical coordinates

Consider the vector field X from Exercise 5.7.1 and Y obtained by multiplying ∂_ρ by $(x^2 + y^2)^{1/2}$

$$\begin{aligned}
 X &= y\partial_x + x\partial_y = \rho \sin 2\phi \partial_\rho + \cos 2\phi \partial_\phi = \rho \sin 2\phi e_{\hat{\rho}} + \rho \cos 2\phi e_{\hat{\phi}}, \\
 Y &= x\partial_x + y\partial_y = \rho\partial_\rho = \rho e_{\hat{\rho}},
 \end{aligned}$$

which identifies their nonzero orthonormal components $X^{\hat{\rho}}$, $X^{\hat{\phi}}$ and $Y^{\hat{\rho}}$.

Compute the components $[X, Y]^{\hat{\rho}}$, $[X, Y]^{\hat{\phi}}$, $[X, Y]^{\hat{z}}$ from the above formula using the fact that

$$C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} = -\frac{1}{\rho} = -C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}}$$

are the only nonzero structure functions.

Compare with the result $[X, Y]$ easily done entirely in terms of Cartesian coordinates. Repeat for

$$Y = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} = \frac{\partial}{\partial \phi} = \rho e_{\hat{\phi}}$$

and the same X . ■

Exercise 5.9.4.

Lie brackets in spherical coordinate orthonormal frame

Compute all of the Lie brackets of $\{e_{\hat{r}}, e_{\hat{\theta}}, e_{\hat{\phi}}\}$ working in spherical coordinates (this is easy). ■

Exercise 5.9.5.

Lie brackets in cylindrical coordinate orthonormal frame

Normalizing the cylindrical coordinate frame leads to the orthonormal frame

$$e_{\hat{\rho}} = \frac{\partial}{\partial \rho}, \quad e_{\hat{\phi}} = \frac{1}{\rho} \frac{\partial}{\partial \phi}, \quad e_{\hat{z}} = \frac{\partial}{\partial z}.$$

Evaluate the nonzero Lie brackets of these vector fields using their cylindrical coordinate expressions (almost trivial). ■

Exercise 5.9.6.

duality practice

The duality map for the basis p -vectors and p -covectors in an orthonormal basis $\{e_i\}$ of \mathbb{R}^3 were given in Table 4.2. These are valid pointwise for the orthonormal frames $\{e_x, e_y, e_z\}$, $\{e_{\hat{\rho}}, e_{\hat{\phi}}, e_{\hat{z}}\}$, $\{e_{\hat{r}}, e_{\hat{\theta}}, e_{\hat{\phi}}\}$.

Evaluate the following duals:

x, y, z :

$*1$ (3-form)

$*(x_1 dx + x_2 dy + x_3 dz)$

$*(x_{23} dy \wedge dz + x_{31} dz \wedge dx + x_{12} dx \wedge dy)$

$*(x_{123} dx \wedge dy \wedge dz)$

$\hat{\rho}, \hat{\phi}, \hat{z}$:

$*1$ (3-form)

$*(x_{\hat{\rho}} \omega^{\hat{\rho}} + x_{\hat{\phi}} \omega^{\hat{\phi}} + x_{\hat{z}} \omega^{\hat{z}})$

$*(x_{\hat{\phi}\hat{z}} \omega^{\hat{\phi}\hat{z}} + x_{\hat{z}\hat{\rho}} \omega^{\hat{z}\hat{\rho}} + x_{\hat{\rho}\hat{\phi}} \omega^{\hat{\rho}\hat{\phi}})$

$*(x_{\hat{r}\hat{\phi}\hat{z}} \omega^{\hat{r}\hat{\phi}\hat{z}})$

$\hat{r}, \hat{\theta}, \hat{\phi} :$

$$\begin{aligned} & *1 \quad (3\text{-form}) \\ & *(x_{\hat{r}}\omega^{\hat{r}} + x_{\hat{\theta}}\omega^{\hat{\theta}} + x_{\hat{\phi}}\omega^{\hat{\phi}}) \\ & *(x_{\hat{\theta}\hat{\phi}}\omega^{\hat{\theta}\hat{\phi}} + x_{\hat{\phi}\hat{r}}\omega^{\hat{\phi}\hat{r}} + x_{\hat{r}\hat{\theta}}\omega^{\hat{r}\hat{\theta}}) \\ & *(x_{\hat{r}\hat{\theta}\hat{\phi}}\omega^{\hat{r}\hat{\theta}\hat{\phi}}) \end{aligned}$$

■

Exercise 5.9.7.

dual of 2-form in \mathbb{R}^3

If

$$X = \frac{1}{2} \bar{X}_{jk} d\bar{x}^j \wedge d\bar{x}^k = X_{\theta\phi} d\theta \wedge d\phi + X_{\phi r} d\phi \wedge dr + X_{r\phi} dr \wedge d\theta$$

use the formula $(*\bar{X})^i = \frac{1}{2} \eta^{ijk} X_{jk}$ to evaluate the vector field

$$*X^\sharp = (*X^\sharp) \frac{\partial}{\partial r} + (*X^\sharp) \frac{\partial}{\partial \theta} + (*X^\sharp) \frac{\partial}{\partial \phi} .$$

What are its physical components?

■

Exercise 5.9.8.

Lie brackets and the derivatives of the frame transformation matrix

Suppose we consider how the Lie brackets of a general frame are related to their component matrix with respect to a coordinate frame, like a Cartesian coordinate frame

$$e_i = B^j{}_i \partial_j, \quad \partial_i = B^{-1j}{}_i e_j .$$

Show that

$$C^k{}_{ij} = B^{-1k}{}_m (B^m{}_{j,i} - B^m{}_{i,j}) = 2B^{-1k}{}_m B^m{}_{[j,i]} .$$

We will see later that this antisymmetrized derivative has a geometrical meaning too.

■

Exercise 5.9.9.

rotation generator Lie brackets in spherical coordinates

The columns of the matrix $\underline{A} = \underline{B}^{-1}$ are the new components of the Cartesian coordinate frame vector fields. Expressing it in term of the new coordinates in the spherical coordinate

case gives the representation of the Cartesian frame vector fields in those coordinates given in Section 5.8

$$\underline{A}(\bar{x}) \begin{pmatrix} \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\ \frac{1}{r} \cos \theta \cos \phi & \frac{1}{r} \cos \theta \sin \phi & -\frac{1}{r} \sin \theta \\ -\frac{\sin \phi}{r \sin \theta} & \frac{\cos \phi}{r \sin \theta} & 0 \end{pmatrix}.$$

a) Use this and $\langle x^1, x^2, x^3 \rangle = \langle r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta \rangle$ to re-express in spherical coordinates the vector fields of Exercise 5.4.4 which generate the rotations, leading to the following results

$$\begin{aligned} L_1 &= x^2 \frac{\partial}{\partial x^3} - x^3 \frac{\partial}{\partial x^2}, & &= -\sin \phi \frac{\partial}{\partial \theta} - \cos \phi \cot \theta \frac{\partial \phi}{\partial}, \\ L_2 &= x^3 \frac{\partial}{\partial x^1} - x^1 \frac{\partial}{\partial x^3}, & &= -\cos \phi \frac{\partial}{\partial \theta} + \sin \phi \cot \theta \frac{\partial \phi}{\partial}, \\ L_3 &= x^1 \frac{\partial}{\partial x^2} - x^2 \frac{\partial}{\partial x^1} & &= \frac{\partial}{\partial \phi}. \end{aligned}$$

Do this by using a computer algebra system to matrix multiply \underline{A} by the component vectors of these vector fields $\langle 0, -x^3, x^2 \rangle$, etc., expressed in spherical coordinates.

b) Now evaluate their Lie brackets in spherical coordinates to re-obtain the results of that Exercise.

c) Introduce the ladder operators $L_{\pm} = L_1 \pm iL_2$ and show that they can be written

$$L_{\pm} = e^{\pm i\phi} \left(\pm \frac{\partial}{\partial \theta} + i \cot \theta \frac{\partial \phi}{\partial} \right).$$

d) Evaluate the Lie brackets $[L_{\pm}, L_3]$.

■

Chapter 6

Covariant derivatives

Flat space has globally constant frames, namely the coordinate frames associated with any Cartesian coordinate system. These frames can be used to characterize constant vector fields: those which have constant components in any such globally constant frame of vector fields. However, if the frame we use is not constant but depends on position, then a constant vector field must have components which vary with position to compensate for the position-dependence of the vectors in terms of which it is expressed as a linear combination. Thus when working with non-Cartesian coordinate systems we need a way to determine when a vector field is constant in spite of having nonconstant components. This easily leads us to define a covariant derivative with correction terms contributed to the derivative of the vector field by the derivatives of the frame vectors in terms of which it is expressed. In Cartesian coordinate systems, this covariant derivative just reduces to the ordinary partial derivative. The correction terms in a general frame are easily evaluated in terms of the derivatives of those frame vector fields with respect to the globally constant Cartesian frames, without which we would not so easily be able to recognize when objects are “constant.”

Once we understand how to differentiate vector fields covariantly, we can then easily extend the rule to covector fields and then to their tensor products to get the covariant derivative of any tensor field. This covariant derivative is independent of the coordinate system or frame used to express it, hence the adjective “covariant.”

However, the globally constant frames also have constant lengths and inner products, so the metric which describes this geometry will have to be covariant constant. It will turn out that the correction terms for the covariant derivative can alternatively be obtained through the derivatives of the components of the metric tensor instead of the components of the frame vectors. This latter approach will work even in a space that is not flat. If we specialize to an orthonormal frame, it will turn out that we can express the correction terms entirely in terms of the Lie brackets of the frame vector fields, the primary reason for our interest in those Lie brackets.

6.1 Covariant derivatives on \mathbb{R}^n with Euclidean metric

It's clear what we mean by a “constant tensor field” on \mathbb{R}^n — namely one whose Cartesian coordinate components are constant in the standard Cartesian coordinate system $\{x^i\}$ on \mathbb{R}^n (or in fact for any Cartesian coordinates). There is a 1–1 correspondence between tensors on the vector space \mathbb{R}^n and such constant tensor fields on \mathbb{R}^n .

$$S = \underbrace{S^{i_1 \dots i_p}_{j_1 \dots j_q}}_{\text{constants}} e_{i_1} \otimes \dots \otimes \omega^{j_1} \otimes \dots \in T^{(p,q)}(\mathbb{R}^n),$$

$$\tilde{S} = S^{i_1 \dots i_p}_{j_1 \dots j_q} \frac{\partial}{\partial x^{i_1}} \otimes \dots \otimes dx^{j_1} \otimes \dots = \binom{p}{q}\text{-tensor field on } \mathbb{R}^n.$$

Such a constant tensor field is also characterized by the vanishing of the Cartesian coordinate derivatives of its Cartesian coordinate components

$$S^{i_1 \dots i_p}_{j_1 \dots j_q, k} \equiv \frac{\partial}{\partial x^k} S^{i_1 \dots i_p}_{j_1 \dots j_q} = 0.$$

This is an important concept for the geometry of \mathbb{R}^n , since if we take a tangent vector at the origin and translate it all over space without changing its length or direction, we obtain a constant vector field. In other words constant tensor fields tell us something about how to move the tangent and cotangent spaces around in space without changing length or orientation information. The usual dot product on each tangent space tells us about the relative geometry of lengths and orientations of vectors in the same tangent space, but it is the ability to compare tangent vectors at different points of space which defines the global geometry.

The Cartesian coordinate frame and dual frame consist of such constant tensor fields, so constant linear combinations of them and the corresponding constant basis tensor fields formed by their tensor products are consequently constant. However, if we use a general coordinate system $\{\bar{x}^i\}$, the new basis frame vector fields *will not be* constant, and hence a constant tensor field must have nonconstant components to compensate for the changing length and orientation of the frame vectors. How can we test for constancy of a tensor field in a general coordinate system where its components hide this property? The position-dependent expressions for $\{\partial/\partial x, \partial/\partial y, \partial/\partial z, dx, dy, dz\}$ in cylindrical or spherical coordinates given in the previous chapter are a good example to consider in this context.

Before we go on, a warning is necessary. The following calculations involving first and second partial derivatives of the coordinate transformation are exactly what gave old fashioned tensor analysis a bad name. Fortunately, we will re-examine the result in a more modern way that allows us to leave those kinds of transformation law nightmares in the past.

So suppose we differentiate the transformation law for a vector field

$$\bar{X}^i = \frac{\partial \bar{x}^i}{\partial x^k} X^k = A^i_j X^k$$

with respect to the new coordinates using the chain rule to re-express those new partial derivatives in terms of the old ones

$$\bar{e}_j = \frac{\partial}{\partial \bar{x}^j} = \frac{\partial x^i}{\partial \bar{x}^j} \frac{\partial}{\partial x^i} = A^{-1i}_j \frac{\partial}{\partial x^i} = B^i_j \frac{\partial}{\partial x^i}$$

in order to see how the component derivatives transform. We find

$$\bar{X}^i_{\cdot j} \equiv \frac{\partial \bar{X}^i}{\partial \bar{x}^j} = \left(\frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial}{\partial x^n} \right) \left(\frac{\partial \bar{x}^i}{\partial x^m} X^m \right) = \frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial \bar{x}^i}{\partial x^m} X^m_{\cdot n} + \frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial^2 \bar{x}^i}{\partial x^n \partial x^m} X^m.$$

The first linear homogenous term corresponds to the transformation law as though $X^m_{\cdot n}$ defined the components of some $\binom{1}{1}$ -tensor field in every coordinate system. The second term is an extra nonhomogeneous term

$$\frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial^2 \bar{x}^i}{\partial x^n \partial x^m} X^m$$

breaking the tensor transformation law. If X is a constant vector field, i.e., $X^m_{\cdot n} = 0$ in the Cartesian coordinates, then

$$\bar{X}^i_{\cdot j} = \frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial^2 \bar{x}^i}{\partial x^n \partial x^m} X^m.$$

Only if $\bar{x}^i = A^i_j x^j$ define new Cartesian coordinates (in general not orthonormal coordinates) so that the Jacobian matrix is constant

$$\frac{\partial \bar{x}^i}{\partial x^m} = A^i_m, \quad A^i_{m,n} = \frac{\partial^2 \bar{x}^i}{\partial x^n \partial x^m} = 0$$

will the new coordinate components of the vector field be constants: $\bar{X}^i_{\cdot j} = 0$. Otherwise the new partial derivatives of the new components will not be identically zero, i.e., the vector will have nonconstant components in the new coordinates.

We can turn this around to have a quantity which must be zero in the new coordinate system if a vector field is to be constant. We just need to re-express the extra term entirely in terms of the new coordinates, a term which must compensate the new coordinate component derivatives

$$\frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial \bar{x}^i}{\partial x^m} X^m_{\cdot n} = \bar{X}^i_{\cdot j} - \frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial^2 \bar{x}^i}{\partial x^n \partial x^m} X^m = \bar{X}^i_{\cdot j} - \frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial^2 \bar{x}^i}{\partial x^n \partial x^m} \frac{\partial x^m}{\partial \bar{x}^l} \bar{X}^l.$$

Re-expressing the second derivative requires first doing a preliminary calculation, differentiating the relation which states that the two Jacobian matrices are inverse matrices, and then solving for the second term in the product rule result, and recombining the factor $(\partial x^n / \partial \bar{x}^j) \partial / \partial x^n = \partial / \partial \bar{x}^j$. These steps follow

$$\begin{aligned} \frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial}{\partial x^n} \left[\frac{\partial \bar{x}^i}{\partial x^m} \frac{\partial x^m}{\partial \bar{x}^\ell} = \delta^i_\ell \right] &\longrightarrow \frac{\partial x^n}{\partial \bar{x}^j} \left[\frac{\partial^2 \bar{x}^i}{\partial x^n \partial x^m} \frac{\partial x^m}{\partial \bar{x}^\ell} + \frac{\partial \bar{x}^i}{\partial x^m} \frac{\partial^2 x^m}{\partial x^n \partial \bar{x}^\ell} \right] = 0, \\ &\longrightarrow - \frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial^2 \bar{x}^i}{\partial x^n \partial x^m} \frac{\partial x^m}{\partial \bar{x}^\ell} = \frac{\partial \bar{x}^i}{\partial x^m} \frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial}{\partial x^n} \left(\frac{\partial x^m}{\partial \bar{x}^\ell} \right) = \frac{\partial \bar{x}^i}{\partial x^m} \frac{\partial^2 x^m}{\partial \bar{x}^j \partial \bar{x}^\ell} \\ &= A^i_m A^{-1m}_{\ell, \bar{j}} = B^{-1i}_m B^m_{\ell, \bar{j}} = B^{-1i}_m dB^m_\ell(\bar{e}_j) \equiv \bar{\Gamma}^i_{j\ell} = \bar{\Gamma}^i_{\ell j}, \end{aligned}$$

where $f_{\cdot \bar{j}} = \partial f / \partial \bar{x}^j$ is a convenient shorthand. This defines a three index object

$$\bar{\Gamma}^i_{j\ell} = \bar{\Gamma}^i_{\ell j} = B^{-1i}_m dB^m_\ell(\bar{e}_j) = \bar{\omega}^i_m(\bar{e}^m_{\ell, \bar{j}})$$

which is symmetric in its lower indices due to the commutivity of partial derivatives, representing the new components of the derivatives of the Cartesian coordinate components $B^m_\ell = \bar{e}^m_\ell$ of the new coordinate frame vectors. The original transformation can be re-expressed in the form

$$\frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial \bar{x}^i}{\partial x^m} X^m_{,n} = \bar{X}^i_{,j} + \bar{\Gamma}^i_{j\ell} \bar{X}^\ell \equiv [\bar{\nabla} \bar{X}]^i_j \equiv \bar{X}^i_{;j} \equiv \bar{\nabla}_j \bar{X}^i.$$

These represent the new components of a $\binom{1}{1}$ -tensor field ∇X whose Cartesian coordinate components are the ordinary derivatives $X^m_{,n}$.

In other words we can define a tensor field in any coordinate system by transforming the tensor field

$$\nabla X = X^m_{,n} \frac{\partial}{\partial x^m} \otimes dx^n$$

from Cartesian coordinates to any other coordinates. The result will vanish any time the Cartesian components vanish, i.e., it will be zero in any coordinate system. Constant vector fields have $\nabla X = 0$ in any coordinate system. The additional term in the new components of ∇X is a correction term to compensate for the nonconstant frame vector fields. It will be interpreted below directly in terms of the derivatives of these nonconstant frame vector fields. In fact the columns of the matrix $(A^{-1m}_i) = (B^m_i) = (\bar{e}^m_i)$ are the old components of new frame vectors

$$\bar{e}_i = \frac{\partial}{\partial \bar{x}^i} = \frac{\partial x^m}{\partial \bar{x}^i} \frac{\partial}{\partial x^m} = B^m_i \frac{\partial}{\partial x^m} = \bar{e}^m_i \frac{\partial}{\partial x^m}$$

and this new object

$$\bar{\Gamma}^i_{j\ell} = B^{-1i}_m B^m_{\ell,\bar{j}} = \bar{\omega}^i_m (\bar{e}^m_{\ell,\bar{j}}).$$

consists of their new coordinate derivatives, re-expressed in terms of the new frame vectors. We need some notation to make this precise.

6.2 Notation for covariant derivatives

Given any $\binom{p}{q}$ -tensor field on \mathbb{R}^n

$$S = S^{i_1 \dots i_p}_{j_1 \dots j_q} e_{i_1} \otimes \dots \otimes \omega^{j_q}$$

expressed in any frame $\{e_i\}$ with dual frame $\{\omega^i\}$, define the following two tensor fields, if $X = X^i e_i$ is any vector field:

$$\begin{aligned} \nabla S &= [\nabla S]^{i_1 \dots i_p}_{j_1 \dots j_q k} e_{i_1} \otimes \dots \otimes \omega^{j_q} \otimes \omega^k : \quad \binom{p}{q+1}\text{-tensor field, covariant derivative of } S \\ \nabla_X S &= [\nabla_X S]^{i_1 \dots i_p}_{j_1 \dots j_q} e_{i_1} \otimes \dots \otimes \omega^{j_q} : \quad \binom{p}{q}\text{-tensor field, covariant derivative of } S \text{ along } X \end{aligned}$$

where

$$\begin{aligned} [\nabla S]^{i_1 \dots i_p}_{j_1 \dots j_q k} &\equiv S^{i_1 \dots i_p}_{j_1 \dots j_q; k} \equiv [\nabla_{e_k} S]^{i_1 \dots i_p}_{j_1 \dots j_q}, \\ [\nabla_X S]^{i_1 \dots i_p}_{j_1 \dots j_q} &\equiv S^{i_1 \dots i_p}_{j_1 \dots j_q; k} X^k \end{aligned}$$

are alternate more convenient notations for these fields. The operation ∇_X generalizes the directional derivative from scalar fields to tensor fields. Both these tensor derivatives are defined so that their components reduce to ordinary derivatives in a Cartesian coordinate system

$$S^{i_1 \dots i_p}_{j_1 \dots j_q; k} = S^{i_1 \dots i_p}_{j_1 \dots j_q, k}.$$

∇S is a new tensor field with one more vector argument which accepts the tangent vector along which the covariant derivative of the tensor is evaluated, generalizing the differential of a function $df(X) = Xf$ which adds the vector argument to accept the tangent vector along which the derivative of the function is evaluated.

A tensor field which vanishes in a single frame, by definition vanishes in every frame, i.e., zero components in one frame define a zero tensor field which must have zero components in every frame. Constant tensor fields have vanishing covariant derivative and will be called “covariant constant.” The components of the covariant derivative in any non-Cartesian coordinate frame may be calculated in two ways:

1. by transformation from Cartesian coordinates,
2. by being clever (see below).

First notice that the covariant derivative obeys obvious product rules which are inherited from partial derivatives in Cartesian coordinates. In Cartesian coordinates one has

$$(S^{i \dots}_{j \dots} T^{k \dots}_{\ell \dots})_{,m} = S^{i \dots}_{j \dots, m} T^{k \dots}_{\ell \dots} + S^{i \dots}_{j \dots} T^{k \dots}_{\ell \dots, m}$$

so in any frame

$$(S^{i \dots}_{j \dots} T^{k \dots}_{\ell \dots})_{;m} = S^{i \dots}_{j \dots; m} T^{k \dots}_{\ell \dots} + S^{i \dots}_{j \dots} T^{k \dots}_{\ell \dots; m}$$

since a direct transformation of the first line yields the second line in the new frame. Thus, dropping indices

$$\nabla(S \otimes T) = \nabla S \otimes T + S \otimes \nabla T,$$

or

$$\nabla_X(S \otimes T) = (\nabla_X S) \otimes T + S \otimes (\nabla_X T).$$

The same product rule holds for any number of contractions of up/down index pairs in this equation. For example

$$(S^i_j T^j)_{;k} = S^i_{j;k} T^j + S^i_j T^j_{;k},$$

$$(S^i_j T^j)_{;i} = S^i_{j;i} T^j + S^i_j T^j_{;i}.$$

For a $\binom{0}{0}$ -tensor, i.e., a function f we have $f_{;k} = f_{,k}$ in Cartesian coordinates, so transforming it to a new coordinate system $\{\bar{x}^i\}$ we get

$$f_{;\bar{k}} = \frac{\partial x^\ell}{\partial \bar{x}^k} f_{,\ell} = \frac{\partial x^\ell}{\partial \bar{x}^k} \frac{\partial f}{\partial x^\ell} = \frac{\partial f}{\partial \bar{x}^k} = f_{,\bar{k}},$$

i.e., the covariant derivative of a function equals the 1-form whose components are the corresponding partial derivatives of the function in any coordinate system, i.e., the differential of the function

$$\begin{aligned} \nabla f &= \bar{f}_{;\bar{k}} d\bar{x}^k = f_{,\bar{k}} d\bar{x}^k = df, \\ \nabla_X f &= \bar{f}_{;\bar{k}} \bar{X}^k = df(X) = Xf, \end{aligned}$$

while the covariant derivative of a function along X is just the ordinary derivative of f along X .

In a given frame $\{e_i\}$, consider the covariant derivative of e_j along e_i , itself a new vector field which can therefore be expressed in terms of its components in this frame

$$\nabla_{e_i} e_j = \Gamma^k_{ij} e_k,$$

thus defining a three index symbol called the “components of the covariant derivative” ∇ or more commonly the “components of the connection” (terminology which will be explained in the next chapter)

$$\Gamma^k_{ij} = \omega^k(\nabla_{e_i} e_j)$$

which is the k th component of the vector field $\nabla_{e_i} e_j$. Notice that the first covariant index is associated with the direction of differentiation, while the other mixed pair of indices are associated with a linear transformation of the tangent space.

Because of the duality relation $\omega^j(e_k) = \delta^j_k$ whose right hand side is a set of constant functions whose derivative is zero, applying the product rule for the covariant derivative to the left hand side yields a formula for the covariant derivatives of the dual frame covector fields

$$0 = \nabla_{e_i}(\delta^j_k) = \nabla_{e_i}[\omega^j(e_k)] = (\nabla_{e_i} \omega^j)(e_k) + \omega^j(\nabla_{e_i} e_k) = (\nabla_{e_i} \omega^j)(e_k) + \Gamma^j_{ik},$$

The first term is the k th component of the covector $\nabla_{e_i}\omega^j$ so solving for it one finds

$$\nabla_{e_i}\omega^j = -\Gamma_{ik}^j\omega^k.$$

Now use the product rule for an arbitrary tensor field

$$S = S_{j_1\cdots j_q}^{i_1\cdots i_p} e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes \omega^{j_1} \otimes \cdots \otimes \omega^{j_q}$$

yielding in general three kinds of terms: the ordinary frame derivatives of the scalar component functions, the derivatives of the frame vector fields if contravariant indices are present, and the derivatives of the dual 1-forms if covariant indices are present

$$\begin{aligned} \nabla_{e_k} S &= (\nabla_{e_k} S_{j_1\cdots j_q}^{i_1\cdots i_p}) e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes \omega^{j_1} \otimes \cdots \otimes \omega^{j_q} \\ &\quad + S_{j_1\cdots j_q}^{i_1\cdots i_p} (\nabla_{e_k} e_{i_1}) \otimes \cdots \otimes e_{i_p} \otimes \omega^{j_1} \otimes \cdots \otimes \omega^{j_q} + \cdots \\ &\quad + S_{j_1\cdots j_q}^{i_1\cdots i_p} e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes \nabla_{e_k} \omega^{j_1} \otimes \cdots \otimes \omega^{j_q} + \cdots \\ &= [e_k S_{j_1\cdots j_q}^{i_1\cdots i_p} + \Gamma_{km}^{i_1} S_{j_1\cdots j_q}^{m\cdots i_p} + \cdots - \Gamma_{kj_1}^m S_{m\cdots j_q}^{i_1\cdots i_p} - \cdots] e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes \omega^{j_1} \otimes \cdots \otimes \omega^{j_q} \\ &\equiv S_{j_1\cdots j_q; k}^{i_1\cdots i_p} e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes \omega^{j_1} \otimes \cdots \otimes \omega^{j_q} \end{aligned}$$

or if we extend the comma index notation to denote the frame derivatives $f_{,k} = e_k f$, we have the component formula

$$S_{j_1\cdots j_q; k}^{i_1\cdots i_p} = S_{j_1\cdots j_q, k}^{i_1\cdots i_p} + \Gamma_{km}^{i_1} S_{j_1\cdots j_q}^{m\cdots i_p} + \cdots - \Gamma_{kj_1}^m S_{m\cdots j_q}^{i_1\cdots i_p} - \cdots.$$

It consists of three kinds of terms: the first ordinary frame derivative of the component functions, a positive correction term for each contravariant index, and a negative correction term for each covariant index. The correction terms are all zero in a Cartesian coordinate frame where the frame vector fields and 1-forms are covariant constant and so the covariant derivative reduces to the ordinary derivative of their components. In a general frame, they simply represent the contribution to the rate of change of the tensor due to the changing frame vectors and dual frame covectors through the product rule.

Thus once we calculate the components of the covariant derivatives of the frame vector fields, i.e., Γ_{ij}^k , we can evaluate the covariant derivative of any tensor field. These were evaluated above for general coordinate frames in terms of the transformation from Cartesian coordinates to the new coordinates $\{x^i\}$

$$\frac{\partial}{\partial \bar{x}^l} = A^{-1m}{}_l \frac{\partial}{\partial x^m} \rightarrow \bar{\Gamma}_{j\ell}^i = A^i{}_m A^{-1m}{}_{l\bar{j}} = [\underline{A} d\underline{A}^{-1}]^i{}_{\ell}(\bar{e}_j),$$

which led to frightening relations involving first and second partial derivatives of the two sets of coordinates in various directions and combinations.

If we instead consider the transformation to a general frame

$$e_l = A^{-1m}{}_l \frac{\partial}{\partial x^m} = B^m{}_l \frac{\partial}{\partial x^m},$$

by covariant differentiation of this relation using the covariant constancy of the Cartesian coordinate frame vectors

$$\begin{aligned}\nabla_{e_j} e_l &= (e_j B^m_l) \frac{\partial}{\partial x^m} = B^{-1i}{}_m (e_j B^m_l) e_i, \\ \Gamma^i_{j\ell} &= B^{-1i}{}_m (e_j B^m_l) = B^{i-1}{}_m dB^m_l(e_j) = [\underline{B}^{-1} d\underline{B}]^i{}_\ell(e_j),\end{aligned}$$

where the frame derivatives of the matrix \underline{B} whose columns are the old components of the new frame vector fields have been re-expressed as the values of its differential on the frame vector fields: for a function $f_j = df(e_j)$, for the matrix $e_j B^j_l = dB^j_l(e_j)$. This is exactly the previous relation when the frame is actually a coordinate frame: $e_i = \partial/\partial \bar{x}^i$.

Matrix methods when applicable are much more efficient than working with individual components. The n^3 components Γ^i_{jk} of the covariant derivative operator, or “connection” as it is called for reasons to become clear later, may be packaged in a more user friendly format by introducing the matrix of connection 1-forms for a given frame $\{e_i\}$

$$\omega^i{}_k \equiv \Gamma^i_{jk} \omega^j, \quad \underline{\omega} \equiv (\omega^i{}_k) = \underline{B}^{-1} d\underline{B}(e_j) \omega^j = \underline{B}^{-1} d\underline{B}.$$

Each entry in this matrix $\underline{\omega}$ is a covector field or 1-form (we will use the two terms interchangeably). This formula makes it easy to calculate $\underline{\omega}$. Once the matrix is evaluated, the (i, k) entry gives the 1-form $\omega^i{}_k = \Gamma^i_{jk} \omega^j$ and the j -th component of this 1-form is the connection component Γ^i_{jk} .

Both cylindrical and spherical coordinates are examples of orthogonal coordinates, namely those for which the coordinate frame vectors are orthogonal, and hence when divided by their lengths become an orthonormal frame

$$\begin{aligned}\bar{e}_i &= B^j{}_i \frac{\partial}{\partial x^j}, \quad g_{ii} = g(e_i, e_i) \\ \rightarrow \bar{e}_{\hat{i}} &= (g_{ii})^{-1/2} e_i = (g_{ii})^{-1/2} B^j{}_i \frac{\partial}{\partial x^j} \equiv \mathcal{B}^j{}_i \frac{\partial}{\partial x^j}.\end{aligned}$$

The columns of \underline{B} are normalized in this way to produce the orthogonal matrix \mathcal{B} . The dual 1-forms are also orthogonal and can be normalized by dividing by their lengths

$$\omega^i = d\bar{x}^i = A^i{}_j dx^j, \quad \bar{g}^{ii} = g(\omega^i, \omega^i) = (\bar{g}_{ii})^{-1},$$

which corresponds to normalizing the rows of \underline{A} to obtain the orthogonal matrix $\underline{\mathcal{A}} = \underline{\mathcal{B}}^{-1}$

$$\omega^{\hat{i}} = (\bar{g}^{ii})^{-1/2} \omega^i = (\bar{g}_{ii})^{1/2} A^i{}_j dx^j \equiv \mathcal{A}^i{}_j dx^j.$$

Replacing $\underline{A} = \underline{B}^{-1}$ by $\underline{\mathcal{A}} = \underline{\mathcal{B}}^{-1}$ in the formula for the connection 1-forms gives the formula valid for the associated orthonormal frame

$$\underline{\hat{\omega}} = \underline{\mathcal{B}}^{-1} d\underline{\mathcal{B}} = (\omega^{\hat{i}}{}_{\hat{k}}) = (\Gamma^{\hat{i}}{}_{\hat{j}\hat{k}} \omega^{\hat{j}}).$$

Thus from the j th orthonormal component of the connection 1-form in the ik entry of this matrix, one reads off the connection component $\Gamma_{\hat{j}\hat{k}}^{\hat{i}}$. This is the matrix of the linear transformation of the orthonormal frame vectors corresponding to the derivative along the unit vector $e_{\hat{j}}$, so it must belong to the Lie algebra of the orthogonal group associated with the signature of the metric, i.e.,

$$\nabla e_{\hat{j}} e_{\hat{l}} = \Gamma_{\hat{j}\hat{l}}^{\hat{i}} e_{\hat{i}} \rightarrow \Gamma_{\hat{i}\hat{j}\hat{l}} = g_{\hat{i}\hat{i}} \Gamma_{\hat{j}\hat{l}}^{\hat{i}} = -g_{\hat{l}\hat{l}} \Gamma_{\hat{j}\hat{i}}^{\hat{l}} = \Gamma_{\hat{l}\hat{j}\hat{i}} \rightarrow (g \underline{\hat{\omega}})^T = -g \underline{\hat{\omega}}$$

or

$$\omega_{\hat{i}\hat{j}} = -\omega_{\hat{j}\hat{i}}.$$

With the index lowering convention this just means that the totally covariant components of the connection are antisymmetric in their outer indices associated with the linear transformation of the tangent space, or simply that the index-lowered connection 1-form matrix is antisymmetric. For the usual dot product in \mathbb{R}^n , the connection 1-form matrix $\underline{\hat{\omega}}$ itself is antisymmetric. This gives a direct interpretation to this matrix-valued 1-form. Its value on the tangent vector X , namely $\underline{\hat{\omega}}(X)$, describes the rate of rotation of the orthonormal frame as one moves in the direction of the tangent vector X . For \mathbb{R}^3 the dual of this antisymmetric matrix defines an angular velocity vector 1-form using the usual correspondence between antisymmetric matrices and vectors in an orthonormal frame—its value on a tangent vector X defines the axis of the rotation in the tangent space, and the rate of rotation around that axis, for motion along a curve whose tangent vector is X . For more general inner products on \mathbb{R}^n with some negative signs associated with the self-inner products of orthonormal frame vectors, the connection 1-form matrix describes pseudo-rotations of the frame, namely Lorentz transformations (which include ordinary rotations and hyperbolic rotations called boosts) in a Minkowski spacetime.

Note that the tensor $\Gamma_{ij}^k e_k \otimes \omega^i \otimes \omega^j$ is the invariant object whose component functions we are using, but it is a different tensor in different frames. We will see how it “transforms” in the next section.

Exercise 6.2.1.

cylindrical and spherical frame connection components

Let's agree to drop the bar for the non-Cartesian coordinates $\{x^i\} = \{\rho, \phi, z\}$ or $\{r, \theta, \phi\}$ here.

a) Use the formula

$$\underline{B}^{-1} d\underline{B} = (\Gamma_{jk}^i dx^j) = \begin{pmatrix} \omega_{\rho}^{\rho} & \omega_{\phi}^{\rho} & \omega_z^{\rho} \\ \omega_r^{\phi} & \omega_{\phi}^{\phi} & \omega_{\phi}^{\phi} \\ \omega_r^z & \omega_{\theta}^z & \omega_z^z \end{pmatrix}$$

to calculate by hand the nonzero components $\Gamma_{jk}^i = (\omega_k^i)_j$ of the connection in cylindrical coordinates using the expressions given in the previous chapter

$$\underline{B} = \begin{pmatrix} \cos \phi & -\rho \sin \phi & 0 \\ \sin \phi & \rho \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \underline{B}^{-1} = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\rho^{-1} \sin \phi & \rho^{-1} \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The results should be

$$\Gamma^\rho_{\phi\phi} = -\rho, \quad \Gamma^\phi_{\phi\rho} = \rho^{-1} = \Gamma^\phi_{\rho\phi} = \rho^{-1}.$$

b) Do the same in the associated orthonormal frame and dual frame

$$\{e_{\hat{\rho}}, e_{\hat{\phi}}, e_{\hat{z}}\} = \left\{ \frac{\partial}{\partial \rho}, \frac{1}{\rho} \frac{\partial}{\partial \phi}, \frac{\partial}{\partial z} \right\}, \quad \{\omega^{\hat{\rho}}, \omega^{\hat{\phi}}, \omega^{\hat{z}}\} = \{d\rho, \rho d\phi, dz\}.$$

using the expressions

$$\underline{\mathcal{B}} = \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \underline{\mathcal{B}}^{-1} = \underline{\mathcal{B}}^T$$

given in the previous chapter. Show that the result is

$$\underline{\hat{\omega}} = \underline{\mathcal{B}}^{-1} d\underline{\mathcal{B}} = d\phi \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This has a simple interpretation. The azimuthal angle ϕ corresponds to a right hand rule rotation around the z -axis which rotates the orthonormal cylindrical frame vectors into position from the Cartesian frame at $\phi = 0$. The differential rotation by the additional angle $d\phi$ simply continues this same rotation of those frame vectors. This accounts for two of the three nonzero connection components in the cylindrical coordinate frame (which are scaled by normalizing factors compared to the orthonormal frame components). The extra nonzero component due to the fact that the azimuthal frame vector increases in length by its distance ρ from the origin, so one gets a $d \ln \rho$ contribution to the relative rate of change of this frame vector in the radial direction.

c) With a little more effort you can repeat this exercise by hand for spherical coordinates, but this is insane. Use a computer algebra system to do this almost effortlessly. The orthogonal matrices of this orthonormal frame are

$$\underline{\mathcal{B}} = \begin{pmatrix} \sin \theta \cos \phi & \cos \theta \cos \phi & -\sin \phi \\ \sin \theta \sin \phi & \cos \theta \sin \phi & \cos \phi \\ \cos \theta & -\sin \theta & 0 \end{pmatrix}, \quad \underline{\mathcal{B}}^{-1} = \underline{\mathcal{B}}^T.$$

Recall that the columns of $\underline{\mathcal{B}}$ are the old components of the new basis vectors $e_{\hat{r}}, e_{\hat{\theta}}, e_{\hat{\phi}}$, while the columns of $\underline{\mathcal{B}}^{-1}$ (rows of $\underline{\mathcal{B}}$) are the new components of the old basis vectors $e_{\hat{z}}, e_{\hat{\theta}}, e_{\hat{r}}$.

Show that

$$\underline{\mathcal{B}}^{-1} d\underline{\mathcal{B}} = \begin{pmatrix} 0 & -d\theta & -\sin \theta d\phi \\ d\theta & 0 & -\cos \theta d\phi \\ \sin \theta d\phi & \cos \theta d\phi & 0 \end{pmatrix} = \begin{pmatrix} 0 & \omega^{\hat{r}}_{\hat{\theta}} & \omega^{\hat{r}}_{\hat{\phi}} \\ \omega^{\hat{\theta}}_{\hat{r}} & 0 & \omega^{\hat{\theta}}_{\hat{\phi}} \\ \omega^{\hat{\phi}}_{\hat{r}} & \omega^{\hat{\phi}}_{\hat{\theta}} & 0 \end{pmatrix} \equiv \Omega^{\hat{i}} \underline{L}_{\hat{i}}.$$

The corresponding vector-valued 1-form (the metric dual on the pair of antisymmetric indices) is therefore

$$\vec{\underline{\Omega}} = \langle \Omega^{\hat{r}}, \Omega^{\hat{\theta}}, \Omega^{\hat{\phi}} \rangle = \langle \cos \theta d\phi, -\sin \theta d\phi, d\theta \rangle,$$

but these are components in the orthonormal spherical frame so as a vector field valued 1-form, this is

$$\Omega = (\cos \theta e_{\hat{r}} - \sin \theta e_{\hat{\theta}}) d\phi + e_{\hat{\phi}} d\theta = e_{\hat{z}} d\phi + e_{\hat{\phi}} d\theta.$$

The last relation follows from comparing the coefficients with the final column of the matrix $\underline{\mathcal{B}}^{-1}$ (final row of $\underline{\mathcal{B}}$), which are the new components of $e_{\hat{z}}$. The component relations between the antisymmetric matrix $\underline{\omega}$ and the vector $\vec{\Omega}$ in the spherical orthonormal frame just reflect the metric dual $*$ of the corresponding 2-form to define a 1-form or vector field depending on the index positioning.

This final formula makes it really easy to interpret the differential rotation which occurs as one moves along any given angular direction, and makes it obvious why this result had to be the way it is. The rotation of the entire r - θ coordinate half plane for fixed ϕ rotates the radial direction down from the vertical by the polar angle, so incrementing that angle by $d\theta$ simply rotates the $e_{\hat{r}}$ - $e_{\hat{\theta}}$ plane of the tangent space by an additional angle $d\theta$ about the $e_{\hat{\phi}}$ axis, rotating $e_{\hat{r}}$ down towards $e_{\hat{\theta}}$. Similarly the azimuthal angle rotates the plane of constant z by the azimuthal angle about the upwards vertical axis, so incrementing that angle by $d\phi$ simply rotates the tangent space about the $e_{\hat{z}}$ axis by that angle.

d) Use a computer algebra system to evaluate the Cartesian components (in the local Cartesian coordinate frame) of the matrix of the linear transformation represented by $\underline{\hat{\omega}}$ and verify that it can be written in the form

$$\underline{\mathcal{B}} \underline{\hat{\Omega}} \underline{\mathcal{B}}^{-1} = (-\sin \theta \underline{L}_1 + \cos \theta \underline{L}_2) d\theta + \underline{L}_3 d\phi = \hat{e}_{\hat{\phi}}^i \underline{L}_i d\theta + \underline{L}_3 d\phi.$$

This shows how moving in the polar angle direction generates a rotation about the axis $\hat{e}_{\hat{\phi}}$, from $e_{\hat{r}}$ to $e_{\hat{\theta}}$, while moving in the azimuthal direction generates a rotation about the vertical axis, from $e_{\hat{r}}$ to $e_{\hat{\phi}}$.

e) Now use a computer algebra system to obtain the connection 1-form matrix for the unnormalized spherical coordinate frame and read off the corresponding nonzero connection components

$$\begin{aligned} \Gamma^r_{\theta\theta} &= -r, \quad \Gamma^r_{\phi\theta} = -r \sin^2 \theta, \\ \Gamma^\theta_{\theta r} &= \frac{1}{r}, \quad \Gamma^\theta_{r\theta} = \frac{1}{r}, \quad \Gamma^\theta_{\phi\phi} = -\sin \theta \cos \theta, \\ \Gamma^\theta_{\phi r} &= \frac{1}{r}, \quad \Gamma^\theta_{\phi\theta} = \cot \theta, \quad \Gamma^\theta_{r\phi} = \frac{1}{r}, \quad \Gamma^\theta_{\theta\phi} = \cot \theta. \end{aligned}$$

Besides the reverse scaling by the lengths of the nonzero orthonormal components, the extra nonzero components of the connection along the diagonal of $\underline{\omega}$ correspond to the increase in the lengths of e_θ and e_ϕ with r (since $(g_{\theta\theta})^{1/2} = r$ and $(g_{\phi\phi})^{1/2} = r \sin \theta$ and the change in length of e_ϕ with θ , increasing and then decreasing. ■

Remark.

If in part a) of this exercise we ignore z and rename the remaining cylindrical coordinates $\{\rho, \phi\}$ to their original polar variable names $\{r, \theta\}$, we get the expressions for the nonzero components of the covariant derivative in polar coordinates in the plane \mathbb{R}^2 in the usual polar coordinates

$$\Gamma^r_{\phi\phi} = -r, \quad \Gamma^\phi_{\phi r} = r^{-1} = \Gamma^\phi_{r\phi} = r^{-1},$$

associated with the Euclidean metric

$$g = dr \otimes dr + r^2 d\theta \otimes d\theta.$$

One can also obtain the same result from the spherical coordinate expressions by ignoring ϕ , but here the correspondence with \mathbb{R}^2 is with the vertical r - θ plane of constant azimuthal spherical coordinate rather than the horizontal z -coordinate plane as in cylindrical coordinates. On the other hand one can set $\theta = \pi/2$ and get a direct correspondence with that horizontal plane as in cylindrical coordinates, renaming only ϕ to θ . \blacktriangle

Example 6.2.1. normalized orthogonal coordinate frame Lie brackets

For simplicity we drop the barred notation to refer to the new coordinates. The following calculation applies to any orthogonal coordinates x^i .

The Lie brackets of the normalized coordinate frame vector fields lead to logarithmic derivatives of the normalization factors

$$\begin{aligned} [e_{\hat{i}}, e_{\hat{j}}] &= [(g_{ii})^{-1/2} e_i, (g_{jj})^{-1/2} e_j] = (\ln(g_{jj})^{-1/2})_{,\hat{i}} e_{\hat{j}} - (\ln(g_{ii})^{-1/2})_{,\hat{j}} e_{\hat{i}} \\ &= (\ln(g_{jj})^{-1/2})_{,\hat{i}} \delta^k_j e_{\hat{k}} - (\ln(g_{ii})^{-1/2})_{,\hat{j}} \delta^k_i e_{\hat{k}} = C^{\hat{k}}_{\hat{i}\hat{j}} e_{\hat{k}}. \end{aligned}$$

Thus the structure functions of the orthonormal frame can be written

$$C^{\hat{k}}_{\hat{i}\hat{j}} e_{\hat{k}} = [(\ln(g_{jj})^{-1/2})_{,\hat{i}} \delta^k_j - (\ln(g_{ii})^{-1/2})_{,\hat{j}} \delta^k_i] e_{\hat{k}}.$$

We shall see below that one can then directly evaluate the components of the connection in the orthonormal frame in terms of these structure functions alone.

Alternatively, with only the knowledge of the orthogonal coordinate components of the connection, one can repeat this Lie bracket derivative calculation for the covariant derivatives of the normalized vectors in terms of the orthogonal coordinate frame vectors

$$\begin{aligned} \nabla_{e_{\hat{i}}} e_{\hat{j}} &= \nabla_{(g_{ii})^{-1/2} e_i} [(g_{jj})^{-1/2} e_j] = (g_{ii})^{-1/2} \nabla_{e_i} [(g_{jj})^{-1/2} e_j] \\ &= (g_{ii})^{-1/2} [(\nabla_{e_i} \ln(g_{jj})^{-1/2}) e_j + (g_{jj})^{-1/2} \Gamma^k_{ij} e_k] \\ &= (g_{ii})^{-1/2} [(\nabla_{e_i} \ln(g_{jj})^{-1/2}) \delta^k_j + (g_{jj})^{-1/2} \Gamma^k_{ij} (g_{kk})^{1/2}] e_{\hat{k}} \\ &= \Gamma^{\hat{k}}_{\hat{i}\hat{j}} e_{\hat{k}}. \end{aligned}$$

This can be re-expressed in terms of the connection 1-form matrix

$$\hat{\omega} = (\omega^{\hat{k}}_{\hat{j}}) = (\Gamma^{\hat{k}}_{\hat{i}\hat{j}} \omega^{\hat{i}}) = (d[\ln(g_{jj})^{-1/2}] \delta^k_j + (g_{kk})^{1/2} \Gamma^k_{ij} (g_{jj})^{-1/2} (g_{ii})^{-1/2} \omega^{\hat{i}}).$$

The second term is just a rescaling of the components as though Γ^k_{ij} were the components of a tensor (which it is, but a different tensor in different choices of frames!) while the first term which is a diagonal matrix, is present to subtract the diagonal part of the second term as a matrix because as we will see below, in an orthonormal frame, $\hat{\omega}$ must have zero diagonal entries, and when index lowered (by at most sign changes since this is an orthonormal frame), must be antisymmetric.

□

6.3 Covariant differentiation and the general linear group

Suppose \square is any linear (homogeneous) derivative operator which produces a $\binom{p}{q}$ -tensor field $\square T$ from a $\binom{p}{q}$ -tensor T and which obeys the product rule for tensor products and contractions thereof. Then \square is entirely determined by how it acts on functions and how it acts on the frame vector fields.

Consider $\square e_i$ for the frame vector fields. Each such field $\square e_i$ is again a vector field so it can be expressed as a linear combination of the frame vector fields

$$\square e_i = \square^j_i e_j, \quad \square^j_i = \omega^j(\square e_i).$$

The entries of the square matrix of functions (\square^j_i) are the “components of \square with respect to the frame,” where \square^j_i is the j th component of the vector field $\square e_i$. This matrix represents a linear transformation of the tangent space with respect to the frame.

However, by duality

$$\omega^i(e_j) = \delta^i_j$$

the contraction (or evaluation) of the dual frame covector fields with the frame vector fields are constant functions (0 or 1) whose derivative must be zero no matter what \square actually is, so by the product rule

$$\begin{aligned} 0 = \square \delta^i_j &= \square(\omega^i(e_j)) = (\square \omega^i)(e_j) + \underbrace{\omega^i(\square e_j)}_{\equiv \square^i_j} \rightarrow \omega^i(\square e_j) = -\square^i_j. \\ &\equiv \square^i_j \end{aligned}$$

The first term is the j th component of the covector field $\square \omega^i$, so this covector field is just

$$\square \omega^i = [(\square \omega^i)(e_j)] \omega^j = -\square^i_j \omega^j.$$

Thus the same matrix determines the components of $\square \omega^i$ but with a minus sign.

Now take any $\binom{p}{q}$ -tensor field

$$T = T^{i\dots}_{j\dots} \underbrace{e_i \otimes \dots}_{p \text{ factors}} \otimes \underbrace{\omega^j \otimes \dots}_{q \text{ factors}}.$$

This is a sum of products of functions (the components of T), and tensor products of frame vector fields and dual frame covector fields, so by the frame product rule

$$\begin{aligned} \square T &= (\square T^{i\dots}_{j\dots}) e_i \otimes \dots \otimes \omega^j \otimes \dots + T^{i\dots}_{j\dots} (\square e_i) \otimes \dots \otimes \omega^j \otimes \dots + \underbrace{\dots}_{p-1 \text{ terms}} \\ &\quad + T^{i\dots}_{j\dots} e_i \otimes \dots \otimes (\square \omega^j) \otimes \dots + \underbrace{\dots}_{q-1 \text{ terms}} \\ &= \square(T^{i\dots}_{j\dots}) e_i \otimes \dots \otimes \omega^j \otimes \dots + \dots + T^{i\dots}_{j\dots} (\square^k_i e_k) \otimes \dots \otimes \omega^j \otimes \dots + \dots \\ &\quad + T^{i\dots}_{j\dots} e_i \otimes \dots \otimes (-\square^j_k \omega^k) \otimes \dots + \dots \\ &= \square(T^{i\dots}_{j\dots}) e_i \otimes \dots \otimes \omega^j \otimes \dots + \square^i_k T^{k\dots}_{j\dots} e_i \otimes \dots \otimes \omega^j \otimes \dots + \dots \\ &\quad - \square^k_j T^{i\dots}_{k\dots} e_i \otimes \dots \otimes \omega^j \otimes \dots - \dots \\ &= [\square(T^{i\dots}_{j\dots}) + \square^i_k T^{k\dots}_{j\dots} + \dots - \square^k_j T^{i\dots}_{k\dots} - \dots] e_i \otimes \dots \otimes \omega^j \otimes \dots \end{aligned}$$

one obtains the component formula

$$[\square T]_{j\dots}^{i\dots} = \square(T_{j\dots}^{i\dots}) + \underbrace{\square^i_k T_{j\dots}^{k\dots} + \dots}_{p \text{ terms}} - \underbrace{\square^k_j T_{k\dots}^{i\dots} - \dots}_{q \text{ terms}},$$

consisting of three kinds of terms: the \square derivative applied to the component functions, and a linear transformation of each contravariant index by \square^i_j and of each covariant index by $-\square^i_j$, representing the contributions from the frame vector fields and dual frame 1-form fields. A compact notation for these corrective terms can be introduced by defining

$$[\sigma(\square)T]_{j\dots}^{i\dots} = \square^i_k T_{j\dots}^{k\dots} + \dots - \square^k_j T_{k\dots}^{i\dots} - \dots,$$

so that

$$[\square T]_{j\dots}^{i\dots} = \square[T_{j\dots}^{i\dots}] + [\sigma(\square)T]_{j\dots}^{i\dots}.$$

WARNING: In old fashioned tensor analysis, which only uses components and does not introduce frame vectors or dual frame covectors, one drops the clarifying grouping delimiters and

$$\square T_{j\dots}^{i\dots} \text{ means } [\square T]_{j\dots}^{i\dots} \text{ not } \square[T_{j\dots}^{i\dots}],$$

i.e., \square is understood to be applied to the tensor and not just to the scalar component functions. This notation no longer distinguishes between the derivative of the components and the components of the derivative, so one has to be careful. This is true for the covariant derivative $\square = \nabla$ where $T_{j\dots;k}^{i\dots}$ refers to the components $[\nabla T]_{j\dots;k}^{i\dots}$ of the covariant derivative ∇T of T , not to the covariant derivative of the component functions (which are always the ordinary derivatives $T_{j\dots,k}^{i\dots}$).

The product rule and transformation matrices

The form of the corrective terms in the component formula for a derivative operator which arise from the derivative of the frame and dual frame factors has a simple origin related to how a linear transformation A on an n -dimensional vector space V induces an active linear transformation $\rho^{(p,q)}(A)$ on each of the tensor spaces of $\binom{p}{q}$ -tensors above it, as already discussed at the end of section 1.7. In terms of components with respect to a given basis e_i , this is given by

$$x^i \rightarrow A^i_j x^j \text{ or } \underline{x} \rightarrow \underline{A} \underline{x} : \quad T_{j\dots}^{i\dots} \rightarrow [\rho^{(p,q)}(\underline{A})T]_{j\dots}^{i\dots} = A^i_k \dots A^{-1\ell}_j \dots T_{\ell\dots}^{k\dots}.$$

As the points of the vector space move around, so too do the points in each of the vector spaces of tensors of a given rank. However, this discussion requires some terminology, and we might as well assume we are talking about $V = \mathbb{R}^n$. None of this is necessary to accomplish our goals but it does help explain why certain formulas arise naturally.

As already introduced in chapters 1 and 4, the set of all real $n \times n$ matrices \underline{A} is an n^2 -dimensional vector space called $gl(n, R)$, which is easily identified with the space \mathbb{R}^{n^2} by listing the entries of the rows consecutively. The single condition $\det \underline{A} = 0$ on this space determines a hypersurface through the zero matrix. The remaining open set of matrices in this vector space whose determinant is instead nonzero is a group called the “general linear group” $GL(n, R)$,

where matrix multiplication is the group multiplication. This group in turn acts on the n -dimensional vector space \mathbb{R}^n as a group of linear “point transformations” of this space into itself (namely a 1-1 mapping of points of the space into each other) by left multiplication of vectors in \mathbb{R}^n interpreted as column matrices

$$\underline{x} \rightarrow \rho^{(1,0)}(\underline{A})(\underline{x}) = \underline{A} \underline{x}.$$

This map $\underline{A} \rightarrow \rho^{(1,0)}(\underline{A})$ from the matrix group into the group of linear transformations of \mathbb{R}^n into itself is called the identity representation of the matrix group, and under composition of transformations it satisfies

$$\rho^{(1,0)}(\underline{A} \underline{B}) = \rho^{(1,0)}(\underline{A}) \circ \rho^{(1,0)}(\underline{B}).$$

This just means that the result of two successive transformations on the space is the transformation corresponding to the matrix product.

In general a representation of a group is simply a map ρ from the group into the group of linear transformations of some vector space which respects the group multiplication law: $\rho(A) \circ \rho(B) = \rho(AB)$, i.e., the composition of two successive such linear transformations corresponds to the group product of the corresponding group elements. When a basis is chosen in the vector space, one has a matrix representing each such group element acting in this way, hence the name “representation.” Maps between groups which satisfy this composition condition

$$\rho(AB) = \rho(A) \circ \rho(B)$$

are called homomorphisms. The determinant function satisfies $\det AB = \det A \det B$ and is therefore is a homomorphism from $GL(n, R)$ into the multiplicative group of nonzero real numbers, for example.

Given the identity representation $\rho^{(1,0)}$ of the matrix group $GL(n, R)$, one automatically has an infinite tower of tensor representations $\rho^{(p,q)}$ above it on each of the vector spaces of $\binom{p}{q}$ -tensors. For example, covectors identified with row matrices \underline{x}^T satisfy

$$\rho^{(0,1)}(\underline{A})(\underline{x}^T) = \underline{x}^T \underline{A}^{-1},$$

and hence the composition requirement is satisfied

$$\rho^{(0,1)}(\underline{A} \underline{B})(\underline{x}^T) = \underline{x}^T (\underline{A} \underline{B})^{-1} = \underline{x}^T \underline{B}^{-1} \underline{A}^{-1} = \rho^{(0,1)}(\underline{A}) \circ \rho^{(0,1)}(\underline{B})(\underline{x}^T).$$

Suppose one has a one-parameter family of nonsingular matrices $\underline{A}(t)$ with $\underline{A}(0) = \underline{I}$. This is just a curve through the identity matrix of the group $GL(n, R)$, which we can think of as \mathbb{R}^{n^2} , and the derivative of $\rho(\underline{A}(t))$ at $t = 0$ is the tangent vector to this curve identified with a matrix in the same way tangent vectors to curves in \mathbb{R}^{n^2} are identified with a vector in the same space in the multivariable calculus approach. As illustrated suggestively in Fig. 6.1, then $\rho(\underline{A}(t))$ is a curve through the identity matrix of the representation group and one can calculate its tangent vector at $t = 0$ using the product rule, which leads to one term for each factor of \underline{A}

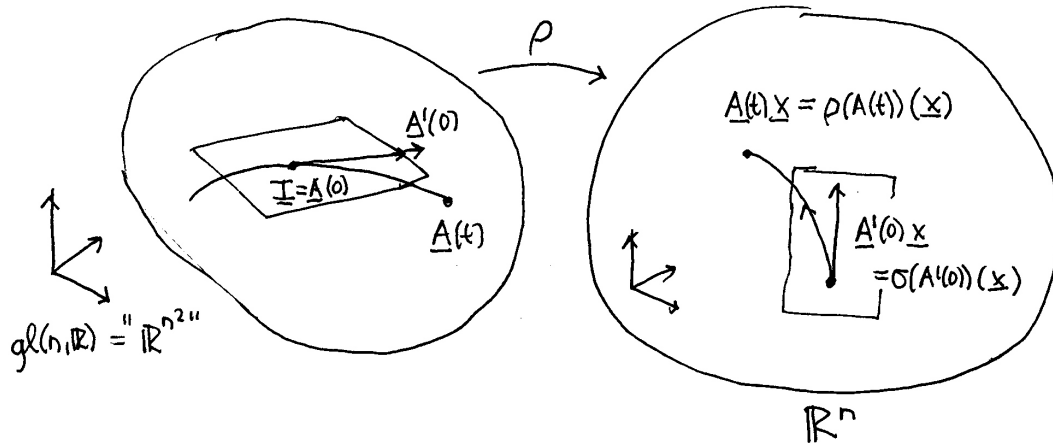


Figure 6.1: A suggestive diagram of a curve $A(t)$ of nonsingular matrices passing through the identity matrix I acting on each of the points x of \mathbb{R}^n by left multiplication (the map $\rho^{(1,0)}$). The derivative $A'(0)$ of the curve at the identity matrix in the group is mapped onto the derivative of the corresponding curve through each point x by the linear map σ , which defines a linear vector field $\sigma(A'(0))(x)$ on \mathbb{R}^n .

or \underline{A}^{-1} in the expression for $\rho(\underline{A})T$

$$\begin{aligned} \frac{d}{dt} [\rho^{(p,q)}(\underline{A}(t))] \Big|_{t=0} &= A^i_k{}'(0) T_{j\dots}^{k\dots} + \dots + A^{-1k}_j{}'(0) T_{k\dots}^{i\dots} + \dots \\ &= A^i_k{}'(0) T_{j\dots}^{k\dots} + \dots - A^k_j{}'(0) T_{k\dots}^{i\dots} - \dots \\ &= [\sigma^{(p,q)}(\underline{A}'(0))T]_{j\dots}^{i\dots}, \end{aligned}$$

where the derivative of the inverse matrix $\underline{A}^{-1}{}'(0) = -\underline{A}'(0)$ follows from differentiating the relation

$$A^i_k A^{-1k}_j = \delta^i_j \rightarrow 0 = A^i_k{}'(0) A^{-1k}_j(0) + A^i_k(0) A^{-1k}_j{}'(0) = A^i_j{}'(0) + A^{-1i}_j{}'(0).$$

The map $\underline{B} \rightarrow \sigma^{(p,q)}(\underline{B})$ defined above by

$$[\sigma^{(p,q)}(\underline{B})T]_{j\dots}^{i\dots} = B^i_k T_{j\dots}^{k\dots} + \dots - B^k_j T_{k\dots}^{i\dots} - \dots$$

on the space $gl(n, \mathbb{R})$ of all $n \times n$ real matrices, which is the derivative of the map $\rho^{(p,q)}$ at the identity matrix, is exactly how the extra corrective terms in the covariant derivative arise

$$T_{j\dots;k}^{i\dots} = T_{j\dots,k}^{i\dots} + [\sigma^{(p,q)}(\underline{\omega}_k)T]_{j\dots}^{i\dots},$$

where $\underline{\omega}_k = ([\omega^i_j]_k) = (\Gamma^i_{kj})$ is the k th component of the connection 1-form matrix. This latter matrix 1-form, once evaluated on a vector field to yield a matrix function, acts as a linear

transformation of each tangent space, lifted to a linear transformation on the tensor spaces above it by the map $\sigma^{(p,q)}$.

The space of all $n \times n$ real matrices $gl(n, R)$, within which $GL(n, R)$ is the open set for which the determinant function is nonzero, is referred to as the Lie algebra of the “Lie group” $GL(n, R)$. Recall that a Lie algebra is a vector space with a commutator, which in this case is just the matrix commutator $[\underline{A}, \underline{B}] = \underline{A}\underline{B} - \underline{B}\underline{A}$. Much of the advances of twentieth century physics was due to the mathematics of Lie groups and their Lie algebras.

This process of evaluating the tangent to a curve of matrices through the identity produces an element of the Lie algebra $gl(n, \mathbb{R})$. The corresponding process of differentiating the matrix Lie group representation map $\rho^{(p,q)}$ produces the matrix Lie algebra representation map $\sigma^{(p,q)}$, but now it is the Lie bracket matrix commutator which reflects the commutators of the matrix Lie algebra itself.

Exercise 6.3.1.

matrix Lie algebra representation map

Convince yourself that the map $\sigma^{(p,q)}$, say for $(p, q) = (0, 1)$, satisfies

$$[\sigma(\underline{A}), \sigma(\underline{B})] = \sigma([\underline{A}, \underline{B}]),$$

i.e., the commutator can be done either before or after the map. This makes it a “Lie algebra” homomorphism since it respects the commutator relations of the matrices. ■

Exercise 6.3.2.

rotation generator

Consider \mathbb{R}^2 and the curve of rotation matrices from section 5.3

$$\underline{S}_3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \underline{A}(t) = e^{t\underline{A}} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

Show that $\underline{S}_3 = \underline{A}'(0)$ and that $\underline{A}'(0)\underline{x} = \underline{L}_3$, where the components of the vector field \underline{L}_3 are $\underline{L}_3 = \langle -y, x \rangle$. This vector field $\underline{L}_3 = x\partial/\partial y - y\partial/\partial x$ is said to generate the rotations about the origin, as discussed in Section 5.2. Under a rotation by an angle t , points in the plane flow along the flow lines of this vector field by a parameter interval t . ■

Exercise 6.3.3.

pseudoorthogonal generators

We already saw in the context of the Lorentz inner product in Section 1.6 that if a matrix \underline{A} preserves an inner product G by transforming it into itself under a change of basis

$$G_{ij} = A^{-1m}{}_i G_{mn} A^{-1n}{}_j, \quad \underline{G} = \underline{A}^{-1T} \underline{G} \underline{A}^{-1},$$

then it satisfies a generalized orthogonality condition

$$(\underline{G} \underline{A})^T = \underline{G} \underline{A}^{-1}.$$

since $(\underline{G} \underline{A})^T = \underline{A}^T \underline{G}^T = \underline{A}^T \underline{G}$. The set of matrices \underline{A} form a group called the orthogonal group relative to that inner product. They preserve lengths and angles and in particular, if one starts with an orthonormal basis with respect to the inner product, it maps orthonormal bases into orthonormal bases. When \underline{G} is the identity matrix as for the usual dot product expressed in an orthonormal basis, this reduces to the orthogonal matrix condition that the transpose equal the inverse.

a) Suppose $\underline{A}(t)$ is a curve of such matrices through the identity $\underline{A}(0) = \underline{I}$ and take the derivative $\underline{B} = \underline{A}'(0)$ of this relation at $t = 0$, again using the result from above that $\underline{A}^{-1'}(0) = -\underline{A}'(0)$, to show that the tangent matrix \underline{B} must then satisfy a generalized antisymmetry condition, namely the index lowered matrix is antisymmetric

$$(\underline{G} \underline{B})^T = -\underline{G} \underline{B}, \quad G_{jk} B^k{}_i = -G_{ik} B^k{}_j, \quad B_{ji} = -B_{ij}.$$

b) Next show that the commutator of two matrices satisfying this condition again satisfies the same condition,

$$(\underline{G} [\underline{B}, \underline{C}])^T = -\underline{G} [\underline{B}, \underline{C}] \leftrightarrow G_{jm} (B^m{}_n C^n{}_i - C^m{}_n B^n{}_i) = -G_{im} (B^m{}_n C^n{}_j - C^m{}_n B^n{}_j)$$

establishing the result that the set of matrices which are antisymmetric with respect to the inner product (in exactly this sense) form a Lie algebra, which is just a vector space with a commutator product which produces a new vector in the space from any two vectors which belong to the space. This is the Lie algebra of the corresponding orthogonal matrix group. Hint: insert factors of the identity matrix $\underline{I} = \underline{G}^{-1} \underline{G}$ in between the two matrices and express all the factors of the two matrices in terms of their index-lowered form so that one can use the antisymmetry of the factor matrices to infer the antisymmetry of the commutator matrix. ■

Suppose e_j is a general frame on \mathbb{R}^n and we consider a transformation to another such frame $\bar{e}_i = A^{-1j} e_j$ with dual frame $\bar{\omega}^i = A^i{}_j \omega^j$, where \underline{A} is a position dependent matrix. Then it is straightforward to calculate step by step a transformation law for expressing how the new components of the covariant derivative are related to the old components. Starting with the

definition of the new components, we have

$$\begin{aligned}
\bar{\Gamma}^k_{ij} &= \bar{\omega}^k(\nabla_{\bar{e}_i}\bar{e}_j) && \text{(definition)} \\
&= A^k_p \omega^p(\nabla_{A^{-1m}_i e_m}(A^{-1n}_j e_n)) && \text{(substitution)} \\
&= A^k_p A^{-1m}_i \omega^p(\underbrace{\nabla_{e_m}(A^{-1n}_j e_n)}_{(\nabla_{e_m} A^{-1n}_j) e_n + A^{-1n}_j \underbrace{\nabla_{e_m} e_n}_{\Gamma^q_{mn} e_q}}) && \begin{aligned} &\text{(linearity of } \nabla: \nabla_{X^i e_i} Y = X^i \nabla_{e_i} Y) \\ &\text{(product rule)} \\ &\text{(definition)} \end{aligned} \\
&= A^k_p A^{-1m}_i [(\nabla_{e_m} A^{-1n}_j) \delta^p_n + A^{-1n}_j \Gamma^q_{mn} \delta^p_q] && \text{(evaluation: } \omega^p(e_q) = \delta^p_q) \\
&= A^k_p A^{-1m}_i [(\nabla_{e_m} A^{-1p}_j) + A^{-1n}_j \Gamma^p_{mn}] && \text{(contraction with delta)} \\
&= A^k_p A^{-1m}_i [(\nabla_{e_m} A^{-1p}_q) A^q_n + \Gamma^p_{mn}] A^{-1n}_j && \text{(clever, see afternote)} \\
&= A^k_p A^{-1m}_i [(dA^{-1p}_q A^q_n)(e_m) + \Gamma^p_{mn}] A^{-1n}_j, && \text{(definition: } \nabla_{e_m} f = df(e_m)) \\
&= A^k_p dA^{-1p}_j(e_m) A^{-1m}_i + A^k_p \Gamma^p_{mn} A^{-1m}_i A^{-1n}_j, && \text{(expand out)}
\end{aligned}$$

where in the second to last line we introduced the Kronecker delta in a clever way

$$\nabla_{e_m} A^{-1p}_j = (\nabla_{e_m} A^{-1p}_q) \delta^q_j = (\nabla_{e_m} A^{-1p}_q) A^q_n A^{-1n}_j.$$

The right term of the final line of the previous calculation corresponds exactly to the tensor transformation law for a $\binom{1}{2}$ -tensor field Γ^i_{jk} , but the first inhomogeneous additive term breaks this transformation law. This means that the components of the connection are the components of a different tensor in each different choice of frame.

Introducing the connection 1-form matrix

$$\begin{aligned}
\bar{\Gamma}^k_{ij} &= A^k_p [(dA^{-1p}_q A^q_n)(e_m) + [\omega(e_m)]^k_n] A^{-1m}_i A^{-1n}_j \\
&= A^k_p [(dA^{-1p}_q A^q_n)(\bar{e}_i) + [\omega(\bar{e}_i)]^k_n] A^{-1m}_i A^{-1n}_j,
\end{aligned}$$

this takes the matrix form $\bar{\omega} = \bar{\omega}(\bar{e}_i) \bar{\omega}^i$, namely the transformation law for the connection 1-form matrix

$$\begin{aligned}
\bar{\omega} &= \underline{A}(\underline{\omega} + d\underline{A}^{-1} \underline{A}) \underline{A}^{-1} = \underline{A} \underline{\omega} \underline{A}^{-1} + \underline{A} d\underline{A}^{-1} \\
&= \underline{B}^{-1} \underline{\omega} \underline{B} + \underline{B}^{-1} d\underline{B}.
\end{aligned}$$

where the last line merely restates the result in terms of the inverse matrix $\underline{B} = \underline{A}^{-1}$.

Exercise 6.3.4.

efficient use of connection 1-forms

By using the matrix of connection 1-forms, we can effortlessly reduce the previous somewhat involved calculation (reminiscent of the ugly coordinate calculation in section 6.1) to a few lines. Assume we have already evaluated the connection 1-form matrix $\omega_1 = \underline{B}_1^{-1} d\underline{B}_1$ for a frame $(e_1)_i = B_1^j{}_i \partial_j$ related to some Cartesian coordinate frame, and consider a new such frame $(e_2)_i = (\underline{B}_2)^k{}_i (e_1)_k = (\underline{B}_2)^k{}_i (\underline{B}_1)^j{}_k \partial_j = (\underline{B}_1 \underline{B}_2)^j{}_i \partial_j$.

Use the product rule and the product property of the inverse matrix to expand and simplify in three successive lines arriving at

$$\begin{aligned}
 \underline{\omega}_2 &= (\underline{B}_1 \underline{B}_2)^{-1} d(\underline{B}_1 \underline{B}_2) \\
 &= \dots \\
 &= \dots \\
 &= \underline{B}_2^{-1} \underline{\omega}_1 \underline{B}_2 + \underline{B}_2^{-1} d\underline{B}_2 \\
 &= \underline{A}_2 \underline{\omega}_1 \underline{A}_2^{-1} + \underline{A}_2 d\underline{A}_2^{-1}.
 \end{aligned}$$

This shows the power of combining matrix notation with frame quantities. The first linear term represents the tensor transformation law for the connection components (supressing one explicit covariant index transformation by the contraction with the dual frame 1-forms), while the second inhomogeneous term is essential for being able to make those components vanish in a covariant constant frame. ■

Exercise 6.3.5.

properties of the vector covariant derivative component formula

a) Use the frame component formula

$$\nabla_X Y = Y^i{}_{;j} X^j e_i = (Y^i{}_{,j} + \Gamma^i{}_{jk} Y^k) X^j$$

to show that

$$[\nabla_X Y]^i = dY^i(X) + \omega^i{}_j(X).$$

This relies on the linearity of the covariant derivative in the differentiating vector field X , a fact used at the beginning of the above long derivation for the transformation of the connection components.

b) The covariant derivative is not linear over scalar function coefficients in the field being differentiated but satisfies a product rule $\nabla_X(fY) = (\nabla_X f)Y + f\nabla_X Y$. Show this using the component formula.

c) Furthermore, show that $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$, just to remind ourselves of the sum rule for covariant differentiation. ■

The flat space \mathbb{R}^n is characterized by the fact that it has a class of globally constant frames for which the connection 1-form matrix is identically zero, but if one starts out in a general frame expressed in general coordinates on \mathbb{R}^n , not knowing how it is expressed in terms of some Cartesian coordinate system, it is not obvious that the geometry defined by this covariant derivative is really flat, and we will need some test of this. Characterizing flatness in this way leads to the idea of curvature, which will be in a later chapter.

6.4 Covariant constant tensor fields

Suppose X has is a constant vector field having constant components with respect to our original standard Cartesian coordinates on \mathbb{R}^n : $X^m_{,n} = 0$. We saw above that the new coordinate component derivatives are

$$\bar{X}^i_{,j} \equiv \frac{\partial \bar{X}^i}{\partial \bar{x}^j} = \frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial \bar{x}^i}{\partial x^m} X^m_{,n} + \frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial^2 \bar{x}^i}{\partial x^n \partial x^m} X^m = \frac{\partial x^n}{\partial \bar{x}^j} \frac{\partial^2 \bar{x}^i}{\partial x^n \partial x^m} X^m.$$

These vanish only if

$$\frac{\partial^2 \bar{x}^i}{\partial x^n \partial x^m} = 0.$$

The general solution of this condition viewed as a differential equation with unknowns \bar{x}^i is

$$\bar{x}^i = A^i_j x^j + b^i,$$

where A^i_j and b^i are constants. This corresponds to allowing new Cartesian coordinates adapted to any basis of the vector space \mathbb{R}^n (thus without any orthonormality restrictions) and with any choice of origin. The mathematical structure associated with this larger class of Cartesian coordinate systems for which no preferred origin exists is called an “affine structure”. An affine space is basically a vector space modulo a choice of origin. Difference vectors between points in the space make sense, but no absolute position vector does since that requires first an arbitrary choice of origin.

When we think of “physical 3-space” whether doing calculus or physics, it is really this affine space (since we have to arbitrarily choose an origin for our axes) together with the Euclidean inner product for difference vectors that we work with. This inner product picks out the orthonormal Cartesian coordinate systems as preferred, and are always assumed in undergraduate multivariable calculus.

The general Cartesian coordinate systems are tied to the global Euclidean geometry of flat space. Any constant vector field simply has constant components with respect to such a coordinate system, thus enabling the identification of all the tangent spaces with the same vector space which in turn can be identified with the whole space, hiding the distinction between vectors and tangent vectors that is typical of the first pass at undergraduate multivariable calculus. The covariant derivative enables us to maintain this connection among the tangent spaces at different points of space even in non-Cartesian coordinate systems.

A Cartesian coordinate frame $\{e_i\}$ is globally covariant constant in the sense that its covariant derivatives ∇e_i are everywhere zero. We will see below that the existence of such a globally constant frame is what characterizes the flatness of space.

Some covariant constant tensors: $\nabla T = 0$

The zero tensor of any rank is obviously covariant constant, but this is rather uninteresting. Nonzero covariant constant tensors may not even exist in curved geometry, as we will see below, so they are worth examining.

Some tensors are automatically covariant constant. The Kronecker delta tensor field $\delta = \delta^i_j e_i \otimes \omega^j = e_i \otimes \omega^i$ has constant components 1 or 0 in every frame $\{e_i\}$, not only in a Cartesian coordinate frame, so the component derivative term in the formula for its covariant derivative is zero and the remaining terms automatically cancel

$$\delta^i_{j;k} = \underbrace{\delta^i_{j,k}}_{\delta^i_{j,k}=0} + \Gamma^i_{k\ell} \delta^\ell_j - \Gamma^\ell_{kj} \delta^i_\ell = \Gamma^i_{kj} - \Gamma^i_{kj} = 0.$$

This confirms its covariant constancy, i.e., $\nabla \delta = 0$. This constancy allows the covariant derivative to commute with contraction, which is the fundamental operation associated with linearity that leads to the idea of a tensor in the first place.

The covariant derivative is defined to satisfy a product rule applying to tensor products of tensors

$$\nabla(T \otimes S) = (\nabla T) \otimes S + T \otimes (\nabla S)$$

as well as to be linear: the covariant derivative of a constant coefficient linear combination of tensor fields equals that linear combination of the covariant derivatives of those tensor fields. Thus any tensor constructed from tensor products of covariant constant tensor fields or from constant linear combinations of such constant tensor fields will be covariant constant. The generalized Kronecker deltas are constructed from ordinary Kronecker deltas in this way

$$\delta^{(p)} = \delta^{i_1 \dots i_p}_{j_1 \dots j_p} e_{i_1} \otimes \dots \otimes e_{i_p} \omega^{j_1} \otimes \dots \otimes \omega^{j_p}, \quad \delta^{i_1 \dots i_p}_{j_1 \dots j_p} = p! \delta^{i_1}_{[j_1} \delta^{i_p}_{j_p]}.$$

All of these are therefore constant tensor fields on \mathbb{R}^n .

Exercise 6.4.1.

covariant constancy of generalized Kronecker delta

Use the component formula for the covariant derivative to calculate the covariant derivative of $\delta^{(2)}$ for an arbitrary frame $\{e_i\}$ to convince yourself that this is true

$$\delta^{(2)} = \delta^{ij}_{mn} e_i \otimes e_j \otimes \omega^m \otimes \omega^n, \quad \delta^{ij}_{mn;k} = \dots = 0.$$

■

Exercise 6.4.2.

constant fields in cylindrical coordinates

The only nonzero connection components in cylindrical coordinates are $\bar{\Gamma}^\rho_{\phi\phi} = -\rho$, $\bar{\Gamma}^\phi_{\phi\rho} = \rho^{-1} = \bar{\Gamma}^\phi_{\rho\phi}$. Since the connection has no components along z , the covariant derivatives of $\partial/\partial z = \bar{e}_z$ and $dz = \omega^z$ are zero:

$$\nabla_{\bar{e}_i} \bar{e}_z = \bar{\Gamma}^j_{iz} \bar{e}_j = 0, \quad \nabla_{\bar{e}_i} d\omega^z = -\Gamma^z_{ik} \omega^k = 0,$$

i.e., $\partial/\partial z$ and dz are covariant constant as we already knew before from our definition of covariant differentiation as ordinary differentiation in Cartesian coordinates. This is not obvious for the remaining two Cartesian coordinate derivatives when expressed in cylindrical coordinates.

From the expressions given above, the nonzero cylindrical coordinate components are

$$\begin{aligned}\frac{\partial}{\partial x} &= \cos \phi \frac{\partial}{\partial \rho} - \frac{\sin \phi}{\rho} \frac{\partial}{\partial \phi} \leftrightarrow \left[\frac{\partial}{\partial x} \right]^\rho = \cos \phi, \quad \left[\frac{\partial}{\partial x} \right]^\phi = -\frac{\sin \phi}{\rho}, \\ \frac{\partial}{\partial y} &= \sin \phi \frac{\partial}{\partial \rho} + \frac{\cos \phi}{\rho} \frac{\partial}{\partial \phi} \leftrightarrow \left[\frac{\partial}{\partial y} \right]^\rho = \sin \phi, \quad \left[\frac{\partial}{\partial y} \right]^\phi = \frac{\cos \phi}{\rho}.\end{aligned}$$

Use the coordinate formula $X^i{}_{;j} = X^i{}_{,j} + \bar{\Gamma}^i{}_{jk}X^k = 0$ to show that the covariant derivatives of both $\partial/\partial x$ and $\partial/\partial y$ are zero. Repeat for the dual 1-forms, whose nonzero components are

$$\begin{aligned}dx &= \cos \phi d\rho - \rho \sin \phi d\phi \leftrightarrow [dx]_\rho = \cos \phi, \quad [dx]_\phi = -\rho \sin \phi, \\ dy &= \sin \phi d\rho + \rho \cos \phi d\phi \leftrightarrow [dy]_\rho = \sin \phi, \quad [dy]_\phi = \rho \cos \phi.\end{aligned}$$

■

6.5 The clever way of evaluating the components of the covariant derivative

Suppose $\{x^i\}$ are Cartesian coordinates on \mathbb{R}^n , and $g = g_{ij}dx^i \otimes dx^j$ is any constant metric on \mathbb{R}^n coming from an inner product on the vector space \mathbb{R}^n , i.e., the constant matrix $\underline{g} = (g_{ij})$ is symmetric and has nonzero determinant so that it can be inverted. Then its covariant derivative must vanish. Expressing this in a general coordinate system leads to

$$\begin{aligned} 0 = \nabla g &\longleftrightarrow 0 = \bar{g}_{ij;k} = \bar{g}_{ij,k} - \bar{g}_{\ell j} \bar{\Gamma}_{ki}^{\ell} - \bar{g}_{i\ell} \bar{\Gamma}_{kj}^{\ell} \\ &\equiv \bar{g}_{ij,k} - \bar{\Gamma}_{jki} - \bar{\Gamma}_{ikj}, \end{aligned}$$

where the last two terms have been rewritten using the index lowering notation, so solving for the metric derivative term leads to

$$\bar{g}_{ij,k} = \bar{\Gamma}_{jki} + \bar{\Gamma}_{ikj} = 2\bar{\Gamma}_{(j|k|i)},$$

where the vertical lines exclude the middle index k from the symmetrization over the outer indices. The left hand side must be symmetric in these indices ij so the right hand side must agree and it does.

This result expresses the ordinary derivatives of the metric components in terms of a certain symmetric part of the index-lowered form of the components of the covariant derivative. By definition

$$\bar{\Gamma}_{jk}^i = \frac{\partial \bar{x}^i}{\partial x^m} \frac{\partial^2 \bar{x}^m}{\partial \bar{x}^j \partial \bar{x}^k} = \bar{\Gamma}_{kj}^i$$

is symmetric in its lower index pair (since partial derivatives commute) which have $n(n+1)/2$ independent components for each of the n values of the upper index for a total of $n^2(n+1)/2$ independent components. But the collection of partial derivatives of \bar{g}_{ij} (namely $n(n+1)/2$ independent components times n independent derivatives) has the same number of independent components, so it is not surprising that they are equivalent, i.e., contain the same information in different packaging—in other words one can invert the above relationship to express the components of the covariant derivative in terms of the ordinary derivatives of the metric components.

This is a classic calculation of differential geometry. One forms the “anticyclic” linear combination (suggestively: $ijk - jki + kij$ compared to the cyclic combination $ijk + jki + kij$) of this same equation

$$\begin{aligned} \bar{g}_{ij,k} &= \bar{\Gamma}_{ikj} + \bar{\Gamma}_{jki}, \\ -\bar{g}_{jk,i} &= -\bar{\Gamma}_{jik} - \bar{\Gamma}_{kij}, \\ \bar{g}_{ki,j} &= \bar{\Gamma}_{kji} + \bar{\Gamma}_{ijk}, \end{aligned}$$

adding them and regrouping terms in pairs which have the same first index, using the symmetry in the last two indices to get the simple result

$$(*) \quad \bar{g}_{ij,k} - \bar{g}_{jk,i} + \bar{g}_{ki,j} = (\bar{\Gamma}_{ikj} + \bar{\Gamma}_{ijk}) + (\bar{\Gamma}_{jki} - \bar{\Gamma}_{jik}) - (\bar{\Gamma}_{kij} - \bar{\Gamma}_{kji}) = 2\bar{\Gamma}_{ijk}$$

so

$$\bar{\Gamma}_{ijk} = \frac{1}{2}(\bar{g}_{ij,k} - \bar{g}_{jk,i} + \bar{g}_{ki,j})$$

and then by raising the first index back to its original position

$$\bar{\Gamma}_{jk}^i = \bar{g}^{i\ell} \bar{\Gamma}_{\ell jk} = \frac{1}{2} \bar{g}^{i\ell} (\bar{g}_{\ell j,k} - \bar{g}_{jk,\ell} + \bar{g}_{k\ell,j}).$$

In other words as long as the metric is a covariant constant nondegenerate symmetric $\binom{0}{2}$ -tensor field on \mathbb{R}^n (corresponding to a symmetric nondegenerate inner product on the vector space \mathbb{R}^n), the components of the covariant derivative can be represented in terms of the components of the metric and their derivatives in a given coordinate system. This can be turned around. If a connection is to have the property that a given metric is covariant constant, **and** if it is a symmetric connection as automatically occurs for the flat space \mathbb{R}^n , then it must have exactly the coordinate components given in this formula. Such a connection is called a metric connection. Every metric determines a symmetric connection through this condition of covariant constancy and symmetry of the connection, defined in a coordinate system by this formula.

No matter what nonsingular constant symmetric matrix (g_{ij}) we start with, the connection components in a Cartesian coordinate system will be all zero and hence the class of covariant constant tensor fields will not change. The covariant derivative on \mathbb{R}^n is really only connected with the *symmetry* of \mathbb{R}^n under the inhomogeneous general linear group of translations and linear transformations about any point, not with the Euclidean inner product.

Key to Riemannian geometry and surface geometry in R^3

The fact that the components of the connection are completely determined by the metric means that if all we know about a space is its metric in some coordinate system, we can introduce the associated symmetric connection and covariant differentiation and study the geometry exactly as in R^3 in a non-Cartesian coordinate system. It is not necessary to have an underlying flat geometry that gives rise to the metric as in our development.

In fact the first example where we can use this idea which we encounter already in multivariable calculus is the case of parametrized surfaces in R^3 , where the three Cartesian coordinates are parametrized by two variables (u, v)

$$x = x(u, v), \quad y = y(u, v), \quad z = z(u, v),$$

or using indexed coordinates

$$x^1 = x^1(u^1, u^2), \quad x^2 = x^2(u^1, u^2), \quad x^3 = x^3(u^1, u^2).$$

If we substitute these expressions into the Euclidean metric and expand the differentials, we collapse the metric to a 2-dimensional metric describing only the geometry of the surface displacements

$$ds^2 = \delta_{ij} dx^i(u^1, u^2) dx^j(u^1, u^2) = \delta_{ij} \frac{\partial x^i}{\partial u^M} \frac{\partial x^j}{\partial u^N} du^M du^N = g_{MN} du^M du^N,$$

where $M, N = 1, 2$. This is the starting point for the study of the intrinsic geometry of the surface, which is one of the key topics of classical differential geometry. The expression ds^2 is called the line element associated with the metric but often is sloppily referred to as the metric, which of course is instead the second rank tensor

$$g = g_{MN} du^M \otimes du^N.$$

In particular we can fix our attention on any two of the three new coordinates in cylindrical or spherical coordinates, holding the remaining one fixed, to get the intrinsic metric on the coordinate surfaces, the most interesting of which describes the geometry of the spheres of radius r_0 in spherical coordinates, although the cones of constant θ_0 are also interesting. Appendix A.4 discusses parametrized surfaces in \mathbb{R}^3 with both the Euclidean and Minkowski geometry.

Exercise 6.5.1.

orthogonal coordinate connection components

For an orthogonal coordinate system the metric component matrix is diagonal, so index shifting amounts to scaling components by certain diagonal metric component factors, so the above formula is easy to evaluate in practice. Show that it reduces to the following formula

$$\bar{\Gamma}^i_{jk} = \frac{1}{2}[(\ln \bar{g}_{ii})_{,j} \delta^i_k + (\ln \bar{g}_{ii})_{,k} \delta^i_j - \bar{g}^{ii} \bar{g}_{jj,i} \delta_{jk}].$$

■

Exercise 6.5.2.

symmetry of connection components

Show that the general formula for $\bar{\Gamma}^i_{jk}$ is symmetric in its lower indices jk .

■

Exercise 6.5.3.

differential log metric determinant

Contract the following formula with g^{ij}

$$\bar{g}_{ij,k} = \bar{\Gamma}_{jki} + \bar{\Gamma}_{ikj} = 2\bar{\Gamma}_{(j|k|i)}$$

and use the previously derived formula $d \ln(\det \bar{g})^{1/2} = \frac{1}{2} \bar{g}^{ij} d\bar{g}_{ij}$ to derive the relation

$$(\ln(\det \bar{g})^{1/2})_{;k} = \bar{\Gamma}^i_{ki}.$$

This relation is needed below to show that the unit volume n -form is covariant constant, and to simplify the expression for the divergence of a vector field in general coordinates.

Note that although $\epsilon_{i_1 \dots i_n}$ has the same numerical values in each frame, it represents different tensors in different frames, so it is not covariant constant. However, the unit n -form η with

components $\bar{\eta}_{i_1 \dots i_n}$ is covariant constant, as are the metric \bar{g}_{ij} and its inverse \bar{g}^{ij} . Suppose e_i is an oriented coordinate frame so that the formula

$$\bar{\eta}_{i_1 \dots i_n} = (\det \bar{g})^{1/2} \epsilon_{i_1 \dots i_n}$$

holds. Now evaluate $\bar{\eta}_{i_1 \dots i_n; k}$ using the previous result to show that it vanishes. To simplify the resulting formula, it is enough to evaluate the component $\eta_{1 \dots n; k}$ since it remains antisymmetric in those indices and this is the only nonvanishing component. Notice that the n negative terms which follow the first derivative term in the formula end up combining into the trace $\bar{\Gamma}_{ki}^i$ of the outer indices of the components of the covariant derivative, which exactly cancels the first term. ■

Exercise 6.5.4.

trace of the connection components

Using the relation

$$d\bar{g}^{-1} = \bar{g}^{-1} d\bar{g} \bar{g}^{-1}, \quad \bar{g}^{ij}{}_{,k} = -\bar{g}^{im} \bar{g}_{mn,k} \bar{g}^{nj}$$

which follows from the differential of $\bar{g}^{-1} \bar{g} = \underline{I}$, show that

$$\bar{\Gamma}^{ik}{}_{,k} = \bar{\Gamma}^i{}_{jk} \bar{g}^{jk} = \bar{g}^{ik}{}_{,k}.$$

■

Exercise 6.5.5.

cylindrical coordinate connection components

Use the formula involving the metric partial derivatives to recalculate the nonzero components $\bar{\Gamma}^i{}_{jk}$ of the covariant derivative in cylindrical coordinates in terms of the nonzero components $\bar{g}_{\rho\rho} = 1 = \bar{g}_{zz}$, $\bar{g}_{\phi\phi} = \rho^2$ of the Euclidean metric. If you feel ambitious, repeat for spherical coordinates. ■

Exercise 6.5.6.

covariant constant tensor

For either polar coordinates in the plane or cylindrical coordinates in space, the constant symmetric tensor

$$\begin{aligned} T &= dx \otimes dx = (\cos \phi d\rho - \rho \sin \phi d\phi) \otimes (\cos \phi d\rho - \rho \sin \phi d\phi) \\ &= \underbrace{\cos^2 \phi d\rho \otimes d\rho}_{T_{\rho\rho}} + \underbrace{\rho^2 \sin^2 \phi d\phi \otimes d\phi}_{T_{\phi\phi}} - \underbrace{\rho \sin \phi \cos \phi (d\rho \otimes d\phi + d\phi \otimes d\rho)}_{T_{\rho\phi}=T_{\phi\rho}} \end{aligned}$$

is covariant constant: $\bar{T}_{ij;k} = 0$. Verify that all components of this covariant derivative are indeed zero in cylindrical coordinates using the result of the previous exercise. ■

6.6 Noncoordinate frames

Well, we are in good shape for computing the components of the covariant derivative in a coordinate frame entirely in terms of the metric in that coordinate system alone, but what is the corresponding formula for a more general frame with nonzero structure functions?

The expression for the Lie bracket in a coordinate frame

$$[X, Y]^k = XY^k - YX^k = Y^k_{,i} X^i - X^k_{,i} Y^i$$

is the “ordinary derivative” commutator of two vector derivatives. Suppose we introduce the corresponding covariant derivative commutator still in a coordinate frame, which is the “comma goes to semi-colon” version of the previous formula

$$\begin{aligned} [\nabla_X Y - \nabla_Y X] &= Y^k_{,i} X^i - X^k_{,i} Y^i \\ &= Y^k_{,i} X^i - X^k_{,i} Y^i + \Gamma^k_{ij} X^i Y^j - \underbrace{\Gamma^k_{ij} Y^i X^j}_{\Gamma^k_{ji} X^i Y^j} \\ &= [X, Y]^k + 2 \underbrace{\Gamma^k_{[ij]}}_{=0} X^i Y^j \end{aligned}$$

where a convenient relabeling of the indices permits a factoring of the vector field factors, leading to the antisymmetric part of the covariant components of the connection, which are zero in a coordinate frame, so we have the result

$$\nabla_X Y - \nabla_Y X = [X, Y].$$

Although obtained in a coordinate frame, this equation is valid for arbitrary vector fields X and Y since it is a frame independent formula involving tensor fields. Another way of obtaining this result is to notice that in a Cartesian coordinate frame, there is no difference between ordinary and covariant differentiation so the equation which holds there must be a tensor equation valid independent of what frame we choose to use.

Thus applying it to the frame vectors of a general frame

$$[e_i, e_j] = C^k_{ij} e_k,$$

whose Lie brackets define the structure functions C^k_{ij} for the frame, we obtain

$$\begin{aligned} \underbrace{\nabla_{e_i} e_j}_{\Gamma^k_{ij} e_k} - \underbrace{\nabla_{e_j} e_i}_{\Gamma^k_{ji} e_k} &= \underbrace{[e_i, e_j]}_{C^k_{ij} e_k} \\ [\Gamma^k_{ij} - \Gamma^k_{ji}] e_k &= C^k_{ij} e_k \rightarrow \Gamma^k_{ij} - \Gamma^k_{ji} = C^k_{ij} \\ \text{or} \\ \Gamma^k_{[ij]} &= \frac{1}{2} C^k_{ij}. \end{aligned}$$

In a noncoordinate frame, the antisymmetric part of the components of the covariant derivative in the lower index pair equals the corresponding structure function which is itself antisymmetric in those indices. The symmetry of the connection components in those indices in a coordinate frame then follows from the vanishing of the structure functions for a coordinate frame.

Let us reconsider the above derivation of the formula for the components of the connection in terms of the metric components in a coordinate frame. There we used the symmetry of $\bar{\Gamma}_{jk}^i$ in its lower indices in a coordinate frame to go on and invert the relationship between $\bar{\Gamma}_{jk}^i$ and the derivative of the metric. Let's drop the bar notation. The anticyclic combination of the condition resulting from the covariant constancy of the metric is the point where we can continue the derivation in a noncoordinate frame but now using the result for the antisymmetric part of the connection components

$$g_{ij,k} - g_{jk,i} + g_{ki,j} = \underbrace{(\Gamma_{ikj} + \Gamma_{ijk})}_{\Gamma_{ijk} - C_{ijk}} + \underbrace{(\Gamma_{jki} - \Gamma_{jik})}_{C_{jki}} - \underbrace{(\Gamma_{kij} - \Gamma_{kji})}_{C_{kij}},$$

where we extend index-shifting to the structure functions

$$C_{ijk} = g_{i\ell} C_{jk}^{\ell}$$

so that the antisymmetric part is

$$\Gamma_{ijk} - \Gamma_{ikj} = C_{ijk} \quad \text{or} \quad \Gamma_{ijk} = \Gamma_{ikj} + C_{ijk}.$$

Solving the above equation for Γ_{ijk} in terms of the derivatives of the metric components now yields

$$\begin{aligned} \Gamma_{ijk} &= \frac{1}{2} [g_{ij,k} - g_{jk,i} + g_{ki,j} + C_{ijk} - C_{jki} + C_{kij}] \\ &= \frac{1}{2} [g_{ij,k} - g_{jk,i} + g_{ki,j} + C_{ijk} + C_{jik} + C_{kij}], \\ \Gamma_{jk}^i &= \underbrace{\frac{1}{2} g^{i\ell} (g_{ij,k} - g_{jk,i} + g_{ki,j})}_{\equiv \{_{jk}^i\}} + \frac{1}{2} (C_{jk}^i + C_j^i{}_k + C_k^i{}_j). \end{aligned}$$

Exercise 6.6.1.

antisymmetric part of connection components

The first part of this formula $\{_{jk}^i\}$ is called a Christoffel symbol and we saw above that it is symmetric in its lower indices. Show that the antisymmetric part of the second contribution involving the structure functions reduces to $\frac{1}{2} C_{jk}^i$ as it should. In the early stages of differential geometry only coordinate frames were used, so there was no distinction between Christoffel symbols and connection components.

■

The formula for the components of the connection in a general frame has two extremes. In a coordinate frame $C^i{}_{jk} = 0$ so the components of the connection reduce to the Christoffel symbols alone $\Gamma^i{}_{jk} = \{^i{}_{jk}\}$. In an orthonormal frame $g_{ij} = \delta_{ij}$ so the component derivative terms are zero $g_{ij,k} = 0$ and $\{^i{}_{jk}\} = 0$ and only the contributions from the structure functions remain

$$\begin{aligned}\Gamma^i{}_{jk} &= \{^i{}_{jk}\}, & (\text{coordinate frame}) \\ \Gamma^i{}_{jk} &= \frac{1}{2}(C^i{}_{jk} - C_{jk}{}^i + C_{kj}{}^i). & (\text{orthonormal frame}) \\ & & (\text{or constant metric components})\end{aligned}$$

In any other kind of frame in which the metric components are not constants, both parts contribute. Such frames on the rotation group manifold $SO(3, \mathbb{R})$ are necessary to describe the motion of a rigid body, and will serve as a useful example of many aspects of metric geometry that we will explore later.

For an orthonormal frame $\{e_i\}$, the covariant constancy of the metric $g_{ij;k} = 0$ and constancy of its components $g_{ij,k} = 0$ imply

$$0 = g_{ij;k} = g_{ij,k} - g_{\ell j}\Gamma_{ki}^\ell - g_{i\ell}\Gamma_{kj}^\ell = 2\bar{\Gamma}_{(j|k|i)}.$$

In other words the symmetric part of the index-lowered connection 1-form matrix is zero, so that it is an antisymmetric matrix

$$\underline{\omega} = (\omega^i{}_j) = (\Gamma^i{}_{kj}\omega^j), \quad \omega_{ij} + \omega_{ji} = 0.$$

When the metric is positive-definite as in the Euclidean metric on \mathbb{R}^n , index raising does not change components so that the original connection 1-form matrix is antisymmetric.

Exercise 6.6.2.

cylindrical coordinate orthonormal frame connection components

As an alternative to example 6.6.2??, use the above formula and the values of the structure functions evaluated in an earlier exercise to evaluate the components of the covariant derivatives of the orthonormal frame vector fields $\{e_{\hat{r}}, e_{\hat{\phi}}, e_{\hat{z}}\}$ associated with cylindrical coordinates.

Then express your results in terms of the matrix of connection 1-forms, obtaining the result

$$\underline{\omega} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} d\phi.$$

■

Exercise 6.6.3.**constant metric component connection**

Some applications have orthogonal frames with constant inner products which are not equal to ± 1 like orthonormal frames, but simply constants of the appropriate sign. Suppose we have a new metric on a 3-dimensional space in such an orthogonal frame $\{e_a\}$ with a constant metric component matrix $\underline{g} = \text{diag}(g_{11}, g_{22}, g_{33})$ and having constant structure functions C^a_{bc} which vanish unless (a, b, c) is a permutation of $(1, 2, 3)$.

a) Show that

$$\Gamma_{(ca)b} = \frac{1}{2} (g_{cc} C^c_{ab} - g_{aa} C^a_{bc}) .$$

b) Show that for a vector field $X = X^a e_a$

$$\begin{aligned} [\nabla_X X^b]_a &= X^b \nabla_b X_a = \dots \\ &= X_{a,b} X^b - \frac{1}{2} (g_{cc} C^c_{ab} - g_{aa} C^a_{bc}) X^c X^b . \end{aligned}$$

c) Assume $C^c_{ab} = \pm \epsilon_{cab}$ and $g_{aa} \equiv I_a$. Show that this becomes

$$[\nabla_X X^b]_a = \underbrace{I_a dX^a(X) - (I_c - I_a) C^c_{ab} X^c X^b}_{(a,b,c) \text{ cyclic permutation of } (1,2,3), \text{ no sums}} .$$

This turns out to be the key to the equations of motion of a rigid body in terms of the diagonalized moment of inertia tensor, called Euler's equations. We'll get to this in chapter 8. Note that for the isotropic case when $I_1 = I_2 = I_3 = I$, this reduces to simply to $I dX^a(X)$. For a symmetric body two of the three are equal, say $I_1 = I_2$, so only the last formula simplifies: $[\nabla_X X^b]_a = I_3 dX^3(X)$. ■

Exercise 6.6.4.**the torsion tensor**

We showed above that the right hand side of the following expression defines an antisymmetric tensor since it is bilinear in X and Y and changes sign with their interchange

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y] = -T(Y, X)$$

whose components are defined by (twice) the antisymmetric part of the connection components on the lower indices in a coordinate frame, or that quantity minus the structure function components in the general case, namely

$$[T(X, Y)]^k = [\nabla_X Y - \nabla_Y X - [X, Y]]^k = (2\Gamma^k_{[ij]} - C^k_{ij}) X^i Y^j \equiv T^k_{ij} X^i Y^j ,$$

leading to the formula

$$2\Gamma^k_{[ij]} = C^k_{ij} + T^k_{ij}.$$

This “torsion tensor” is identically zero for the flat geometry of \mathbb{R}^n with any covariant constant metric. The structure component functions C^k_{ij} are the components of a $\binom{1}{2}$ -tensor, but which depends on the frame, in contrast with this torsion tensor which is independent of the frame or coordinate system used to evaluate it. It can be introduced as an additional independent field to generalize the geometry in a way that is compatible with a metric, which remains covariant constant. In that case using this last formula to replace the antisymmetric part of the connection components in our above general derivation adds one more anticyclic set of terms to the general formula for the components of the connection entering exactly like the structure functions. Show that you get

$$\Gamma^i_{jk} = \frac{1}{2}g^{i\ell}(g_{ij,k} - g_{jk,i} + g_{ki,j}) + \frac{1}{2}(C^i_{jk} + C_j^i{}_k + C_k^i{}_j) + \frac{1}{2}(T^i_{jk} + T_j^i{}_k + T_k^i{}_j).$$

Torsion theories of gravity attempted to relate this additional field to spin densities in matter, but were not too successful. ■

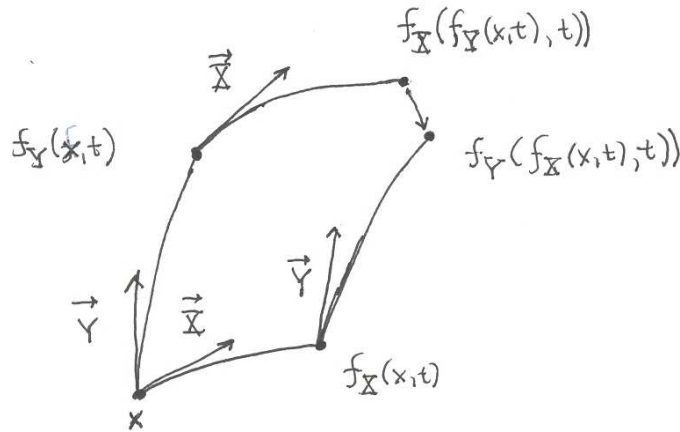


Figure 6.2: Starting from an initial point and flowing along the flow lines of two vector fields in succession in opposite orders in general leads to two different terminal points. The failure of the quadrilateral formed by these flow lines for the same value $t = \epsilon$ of the flow line parameter to close at the far corner is described by the tangent vector $\epsilon^2[X, Y]$ as $\epsilon \rightarrow 0$.

6.7 Geometric interpretation of the Lie bracket

We introduced the Lie bracket as a convenient computational quantity first for characterizing the coordinate frame vector fields and second for evaluating the components of the connection in a noncoordinate frame, but the Lie bracket is actually a fundamental geometrical operation associated with the mathematics of transformation groups like the group of rotations and translations of ordinary space. Its geometrical interpretation also tells us something about the formula for the curvature tensor in a noncoordinate frame.

In section 5.3 we introduced the flow lines of a vector field X and its one-parameter group of transformations that allow points of space to simultaneously flow along these flow lines by natural parameter intervals

$$x^i \rightarrow \bar{x}^i = f_X^i(x, t) = e^{tX} x^i = x^i + tX^i + \dots$$

Suppose we have two vector fields and consider flowing first along one by a parameter interval t and then along the other by the same interval. If instead we switch the order of the vector fields in this sequence, we end up in general at different final points, starting from the same initial point, as illustrated in Fig. 6.2. In the limit $t \rightarrow 0$ we can identify this difference with a difference vector in the tangent space at the starting point, and it is easily evaluated using the simple approximate formula which amounts to a power series approximation to the change in any function F along a vector field, equivalently the interpretation of the directional derivative along X via the chain rule

$$F(e^{tX}) = e^{tX} F(x) = (1 + tX + \frac{t^2}{2} X^2 + \dots) F(x) = F(x) + tX F(x) + \frac{t^2}{2} X^2 F(x) \dots$$

We can apply this to the transformation functions themselves, keeping only terms up to order two in the series expansion. The difference is given by

$$\begin{aligned}
& f_Y^i(f_X(x, t), t) - f_X^i(f_Y(x, t), t) \\
&= e^{tX} f_Y^i(x, t) - e^{tY} f_X^i(x, t) = (e^{tX} e^{tY} - e^{tY} e^{tX}) x^i \\
&= (1 + tX + \frac{t^2}{2} X^2 + \dots)(1 + tY + \frac{t^2}{2} Y^2 + \dots) x^i \\
&\quad - (1 + tY + \frac{t^2}{2} Y^2 + \dots)(1 + tX + \frac{t^2}{2} X^2 + \dots) x^i \\
&= (1 + tY + \frac{t^2}{2} Y^2 + \dots) x^i - (1 + tX + \frac{t^2}{2} X^2 + \dots) \\
&\quad + (tX + t^2 XY + \dots) x^i - (tY + t^2 YX + \dots) \\
&\quad + (\frac{t^2}{2} X^2 + \dots) x^i - \frac{t^2}{2} Y^2 + \dots \\
&\approx t^2 [X, Y] x^i = t^2 [X, Y]^i.
\end{aligned}$$

In other words if the two vector fields do not commute, the four curve segments do not close at second order in t . This does not occur for a pair of coordinate vector fields, where this forms a parallelogram through all orders in t since the translations along one coordinate line and then along another coordinate line commute.

Lie brackets and transformation groups

When a set of vector fields does not commute, the next more general special relationship they can have is to be a set which is closed under the Lie bracket. In other words the Lie bracket of any two elements of the set again belongs to the set. When the set of vector fields forms a finite-dimensional vector space \mathfrak{g} with a basis $\{E_{(a)}\}$, $a = 1, \dots, r$ such that any element of the space can be represented as $X = X^a E_{(a)}$, where X^a are constants, then closure means that the Lie bracket defines a $\binom{1}{2}$ -tensor (antisymmetric in its lower indices) on this vector space called the structure constant tensor, whose components are defined by

$$[E_{(a)}, E_{(b)}] = C^c{}_{ab} E_{(c)}, \quad C^c{}_{ab} = -C^c{}_{ba}.$$

These components $C^c{}_{ab}$ are constants called the structure constants of the Lie algebra, and they transform like a $\binom{1}{2}$ -tensor under a change of basis of this Lie algebra. The Lie brackets of any two elements of the Lie algebra are then evaluated by multilinearity, using the fact that the components with respect to this basis of the Lie algebra are constants

$$[X, Y] = [X^a E_{(a)}, Y^b E_{(b)}] = X^a Y^b [E_{(a)}, E_{(b)}] = C^c{}_{ab} X^a Y^b E_{(c)}.$$

Because of the cyclic Jacobi identity satisfied by any commutation operation, these structure constants satisfy a quadratic identity. By multilinearity

$$[[E_{(a)}, E_{(b)}], E_{(c)}] = C^d{}_{ab} [E_{(d)}, E_{(c)}] = C^d{}_{ab} C^e{}_{dc} E_{(e)}$$

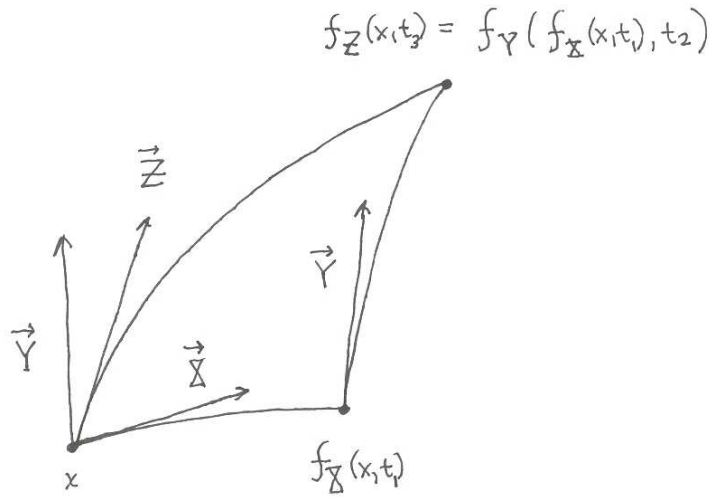


Figure 6.3: Starting from an initial point and flowing along the flow lines of two vector fields of a Lie algebra in succession should be equivalent to flowing along the flow lines of a third vector field of the Lie algebra.

so using this in the Jacobi identity one has

$$\begin{aligned} 0 &= [[E_{(a)}, E_{(b)}], E_{(c)}] + [[E_{(b)}, E_{(c)}], E_{(a)}] + [[E_{(c)}, E_{(a)}], E_{(b)}] \\ &= [C^d_{ab} C^e_{dc} + C^d_{bc} C^e_{da} + C^d_{ca} C^e_{db}] E_{(e)}, \end{aligned}$$

so

$$C^d_{ab} C^e_{dc} + C^d_{bc} C^e_{da} + C^d_{ca} C^e_{db} = -2C^d_{[ab} C^e_{c]d} = 0.$$

Exercise 6.7.1.

Jacobi identity components

Use the antisymmetry of the structure constant tensor in its lower indices, plus the definition of the antisymmetric part of a 3 index tensor to verify the previous shortened component form of the Jacobi identity. ■

Any vector field on \mathbb{R}^n determines a 1-parameter group of transformations defined by allowing the points of the space to flow along its flow lines as discussed in section 5.3 for the case of vector fields whose Cartesian coordinate components are linear functions of those coordinates. The vector field is said to “generate” this group of point transformations, each of which is a 1-1 map of the space into itself. In terms of the Cartesian coordinates, the coordinates of the new points are related to their old coordinates locally by an exponential relationship

$$x^i \rightarrow e^{tX} x^i.$$

This is true for each element of a Lie algebra \mathfrak{g} . If the set of these 1-parameter groups themselves form a group together, then following one such transformation by one for another vector field in the set should be equivalent to a single transformation associated with a third vector field in the set as illustrated in Fig. 6.3.

One of the most important results about transformation groups is that the set of all 1-parameter groups associated with the vector fields in an r -dimensional Lie algebra of vector fields form an r -dimensional group of transformations called a Lie group. The vector fields of the Lie algebra are said to generate this group of transformations. When the structure constant tensor vanishes, the vector fields commute and they generate an Abelian group of commuting transformations, and one can find local coordinates in which a basis of those commuting vector fields are coordinate derivatives and the transformations reduce to translations in those coordinates (i.e., adding constants to those coordinates).

For example consider the case of \mathbb{R}^3 where we have introduced three commuting vector fields $p_i = \partial/\partial x^i$, just the standard Cartesian coordinate derivatives, and three noncommuting vector fields $L_i = \epsilon_{ijk}x^j\partial/\partial x^k$, which satisfy the cyclic commutator relationships

$$[L_i, L_j] = -\epsilon_{ijk}L_k, \quad [L_i, p_j] = -\epsilon_{ijk}p_k, \quad [p_i, p_j] = 0.$$

Remark.

The translation vector fields are denoted by the suggestive notation \vec{p} , in terms of which the rotation vector fields are $\vec{r} \times \vec{p}$. This is the notation reserved for linear momentum and angular momentum respectively. There is good reason for this but we are not ready to appreciate the reasons why yet. ▲

Exercise 6.7.2.
commutators of rotations and translations

We already calculated the first subset of these commutation relations earlier, which correspond to the (nonzero independent) structure constant values $C^1_{23} = C^2_{31} = C^3_{12} = -1$, while the final subset are obvious since the p_j are partial derivative operators and commute. Verify the second subset of commutation relations for the generators of the translations and rotations. ■

These relationships tell us that $\{p_i\}$ are the basis of the 3-dimensional Abelian Lie algebra of the constant vector fields on \mathbb{R}^3 , while the three vector fields $\{L_i\}$ are the basis of the 3-dimensional Lie algebra associated with the rotations of the space into itself, and together they form a 6-dimensional Lie algebra in which both are Lie subalgebras. The flow of the first set of vector fields are the translations of space into itself

$$x^i \rightarrow e^{a^j p_j} x^i = x^i + a^i \equiv T^i(x, a).$$

The flow of the second set are the rotations of the space about the origin

$$x^i \rightarrow e^{\theta n^k L_k} x^i = [e^{\theta n^k S_k}]^i_j x^j \equiv R^i(x, \theta n).$$

since

$$L_k x^i = \epsilon_{kjm} x^j \frac{\partial}{\partial x^m} x^i = \epsilon_{kjm} x^j \frac{\partial x^i}{\partial x^m} = \epsilon_{kjm} x^j \delta_m^i = \epsilon_{kji} x^j = \epsilon_{ikj} x^j \equiv [S_k]^i_j x^j.$$

This transformation represents a rotation of space by an angle θ about the axis through the origin with direction equal to the unit vector n , with the direction of the rotation about this axis determined by the right hand rule. Together the rotations and translations form a 6-dimensional group called the Euclidean group or inhomogeneous orthogonal group $E(3) = IO(3, \mathbb{R})$ of \mathbb{R}^3 .

Exercise 6.7.3.

Lie brackets of linear transformation generating vector fields

Consider the action of a matrix group on \mathbb{R}^n by matrix multiplication. For any $n \times n$ matrix \underline{A} in the Lie algebra of the matrix group, its family of matrix exponentials act on \mathbb{R}^n with group parameter t which is additive for successive matrix multiplications, making this an Abelian group

$$\underline{x} \rightarrow e^{t\underline{A}} \underline{x}.$$

For each matrix in its Lie algebra we can associate a linear vector field

$$\xi(\underline{A}) = A^i_j x^j \partial_i.$$

a) Show that the map $-\xi$ is a Lie bracket isomorphism, namely that it maps the matrix Lie bracket onto the corresponding vector field Lie bracket

$$[-\xi(\underline{A}), -\xi(\underline{B})] = -\xi([\underline{A}, \underline{B}]) \quad \leftrightarrow \quad [\xi(\underline{A}), \xi(\underline{B})] = -\xi([\underline{A}, \underline{B}]).$$

b) Show that this explains the opposite signs of the Lie brackets for the rotation group vector fields $L_k = (\underline{L}_k)^i_j x^j \partial_i = \epsilon_{ikj} x^j \partial_i = \epsilon_{kji} x^j \partial_i$ compared to the matrix commutators of the corresponding matrices. We will understand the significance of this sign below. ■

Exercise 6.7.4.

polar coordinate vector fields

We already showed in Exercise 5.4.6 that the vector field $D_r = (x^i/r)\partial/\partial x^i$ (where $r = (\delta_{ij}x^i x^j)^{1/2}$) commutes with all three L_i vector fields generating rotations in \mathbb{R}^3 . Show that the analogous vector field in \mathbb{R}^2 commutes with L_3 . This means that locally one can find coordinates in which they reduce to coordinate derivatives, and locally the two vector fields generate an Abelian transformation group. However, the adapted coordinates are just polar coordinates in the plane (in terms of which $D_r = \partial/\partial r, L_3 = \partial/\partial \phi$), and while translations in the polar angle are fine, translations in the radial direction have a big problem with the origin where points inside a circle about the origin crash into each other under translations by a negative number, and under translations by a positive number, some points inside a circle near the origin have no point from which they are translated. Thus such transformations fail to be 1-1. Clearly, one needs some global considerations as well to avoid such problems.

Figure 6.4: Needs figure.

■

Exercise 6.7.5.

3-sphere vector fields

We investigated the 2-to-1 relationship between the rotations $SO(3, \mathbb{R})$ and the unit 3-sphere S^3 in \mathbb{R}^4 through the group $SU(2)$ in Exercise 4.5.9. The 3-sphere with its antipodal points identified is in a 1-1 correspondence with the 3-dimensional rotation group.

a) We investigated the 4-dimensional rotation group $SO(4, \mathbb{R})$ action on the 3-sphere in Exercise 4.5.6. There we defined its matrix generators by

$$(\underline{L}_{ij})^{mn} = -\delta_{ij}^{mn} \quad \text{or} \quad (\underline{L}_{ij})^m{}_n = -\delta_{ij}^{mk} \delta_{kn}$$

and introduced the following new basis of the Lie algebra

$$\underline{E}_a = \frac{1}{2}(\underline{L}_{4a} - \underline{L}_{bc}), \quad \tilde{\underline{E}}_a = \frac{1}{2}(\underline{L}_{4a} + \underline{L}_{bc}). \quad (a, b, c) \text{ cyclic permutation of } (1, 2, 3)$$

If you did not do so then, use a computer algebra system to show that the two vector subspaces spanned by $\{\underline{E}_a\}$ and $\{\tilde{\underline{E}}_a\}$ are mutually commuting matrix Lie subalgebras with the following commutation relations

$$[\underline{E}_a, \underline{E}_b] = -\epsilon_{abc} \underline{E}_c, \quad [\underline{E}_a, \tilde{\underline{E}}_b] = 0, \quad [\tilde{\underline{E}}_a, \tilde{\underline{E}}_b] = \epsilon_{abc} \tilde{\underline{E}}_c.$$

which have the same commutation relations under the correspondence $\underline{E}_a \leftrightarrow -\tilde{\underline{E}}_a$. Thus the 6-dimensional Lie algebra of $SO(4, \mathbb{R})$ is the direct sum of two mutually commuting copies of the 3-dimensional Lie algebra of $SO(3, \mathbb{R})$. These turn out to correspond to the right and left translations of the group $SU(2)$ into itself, which we will not pursue here.

b) Define the corresponding vector field generators by

$$L_{ij} = (\underline{L}_{ij})^m{}_n x^n \partial_m = x^i \partial_j - x^j \partial_i, \quad i, j, k, \dots = 1, 2, 3, 4$$

(check this evaluation!) and the related vector fields

$$e_a = \frac{1}{2}(L_{4a} - L_{bc}), \quad \tilde{e}_a = \frac{1}{2}(L_{4a} + L_{bc}), \quad (a, b, c) \text{ cyclic permutation of } (1, 2, 3)$$

which have the sign-reversed commutation relations compared to the corresponding matrices

$$[E_a, E_b] = \epsilon_{abc} E_c, \quad [E_a, \tilde{E}_b] = 0, \quad [\tilde{E}_a, \tilde{E}_b] = -\epsilon_{abc} \tilde{E}_c.$$

c) Show that the Cartesian components of these vector fields are

$$\begin{aligned}\vec{e}_1 &= \frac{1}{2}\langle x^4, x^3, -x^2, -x^1 \rangle, & \vec{\tilde{e}}_1 &= \frac{1}{2}\langle x^4, -x^3, x^2, -x^1 \rangle, \\ \vec{e}_2 &= \frac{1}{2}\langle -x^3, x^4, x^1, -x^2 \rangle, & \vec{\tilde{e}}_2 &= \frac{1}{2}\langle x^3, x^4, -x^1, -x^2 \rangle, \\ \vec{e}_3 &= \frac{1}{2}\langle x^2, -x^1, x^4, -x^3 \rangle, & \vec{\tilde{e}}_3 &= \frac{1}{2}\langle -x^2, x^1, x^4, -x^3 \rangle,\end{aligned}$$

satisfying

$$e_a \cdot e_b = \frac{1}{4} (\delta_{ij} x^i x^j) \delta_{ab} = \tilde{e}_a \cdot \tilde{e}_b.$$

Then show that these are all orthogonal to the radial position vector field

$$e_r = x^i \partial_i.$$

Thus $\{2e_a\}$ and $\{2\tilde{e}_a\}$ are both orthonormal 3-frames on the unit 3-sphere tangent to the 3-sphere which both reduce to $\{\partial_a\}$ at the North pole $(0, 0, 0, 1)$ which corresponds to the identity matrix of $SU(2)$.

d) Defining the radius $r = (\delta_{ij} x^i x^j)^{1/2}$, show that lowering the index on the above frame vectors with the Euclidean metric and dividing by the square of the length yields the dual 3-frame

$$\begin{aligned}\omega^1 &= \frac{2}{r^2} (x^4 dx^1 + x^3 dx^2 - x^2 dx^3 - x^1 dx^4), \\ \omega^2 &= \frac{2}{r^2} (-x^3 dx^1 + x^4 dx^2 + x^1 dx^3 - x^2 dx^4), \\ \omega^3 &= \frac{2}{r^2} (x^2 dx^1 - x^1 dx^2 + x^4 dx^3 - x^3 dx^4),\end{aligned}$$

with similar expressions for $\{\tilde{\omega}^a\}$. Thus the metric on the unit sphere is

$$\frac{1}{4} \delta_{ab} \omega^a \omega^b = \frac{1}{4} \delta_{ab} \tilde{\omega}^a \tilde{\omega}^b.$$

This is no accident. For this group the components of the metric in either frame are proportional to

$$C^c{}_{ad} C^d{}_{bc} = \text{Tr}(\underline{k}_a \underline{k}_b) = \epsilon^{cad} \epsilon_{dbc} = -2\delta_{ab},$$

where $C^c{}_{ab} = \epsilon_{cab}$. As the group manifold of $SU(2)$, the 3-sphere with its natural geometry as a hypersurface in Euclidean space has a metric which is invariant under left and right translations on the group, an interesting fact which we do not have time to explore here. [It turns out that under the correspondence of Exercise 4.5.9, the Lie algebra of “left invariant” vector fields $\{e_a\}$ generates the right translations under which the “right invariant” vector fields $\{\tilde{e}_a\}$ are invariant, while the latter generate the left translations under which the former are invariant, and since the left and right translations on any group commute since they do not interfere with each other, these Lie algebras must commute.]

Who would have ever guessed that the equation of a sphere in 4-dimensions could hide so much mathematical structure?

■

6.8 Isometry groups and Killing vector fields

The symmetry of the geometry we know in our everyday lives is one of its defining characteristics and hence is a fundamental aspect of the metric geometry which reflects its properties in the limiting domain of the differential geometry of curves, surfaces, and their generalizations to higher dimensional spaces and unfamiliar settings. The rotations and translations of space are symmetries of its Euclidean geometry which have important physical consequences for how nature works. Conservation laws for linear and angular momentum hold for systems with such symmetry, while for systems exhibiting invariance under time translation (i.e., are independent of time), the law of conservation of energy holds. These are the underlying principles which enable us to make sense of the world.

In the large, rotational and translational symmetry means that we can move things around with these symmetry operations without changing their size or shapes. The distance formula of Euclidean space enables us to quantify these statements. The metric tensor field of Euclidean space reflects that geometry in the small, in the tangent spaces at each point of space. Invariance of the metric tensor under these transformations reflects the symmetry which holds in the large. But how do vectors and tensors transform under transformations of space into itself?

In section 5.7 we described passive coordinate transformations $x^{i'} = x^{i'}(x)$ of the components of tensor fields in which the points of space do not move, but we change the coordinate grid underneath the points so that the component functions must be re-expressed in the new coordinates (step *a*) below) and then remapped by the appropriate linear transformation associated with the change in the coordinate frames used to express those components (step *b*) below)

$$\begin{aligned}
 X^{i'}(x') &= \underbrace{\frac{\partial x^{i'}}{\partial x^j}(x(x'))}_{a)} \underbrace{X^j(x(x'))}_{a)} \\
 X_{i'}(x') &= \underbrace{X_j(x(x'))}_{a)} \underbrace{\frac{\partial x^j}{\partial x^{i'}}(x')}_{b)} .
 \end{aligned}$$

For a general tensor field suppressing coordinate dependence of the component functions one has the transformation law for $\binom{p}{q}$ tensor field components

$$T^{i' \dots}_{j' \dots} = \frac{\partial x^{i'}}{\partial x^m} \dots \frac{\partial x^n}{\partial x^{j'}} \dots T^{m \dots}_{n \dots} = [\rho^{(p,q)}(\underline{\partial x' / \partial x}) T]^{i' \dots}_{j' \dots} .$$

In contrast we would like to see how to transform tensor fields under an active transformation of the points of a space $x^i \rightarrow \varphi^i(x) = x^i(\varphi(x)) = (x^i \circ \varphi)(x)$ where the points x move to $\varphi(x)$ but the coordinate system (i.e., the coordinate functions x^i) remains fixed.

To understand this geometry, we introduce the coordinate system dragged along by the transformation so that the new point has values of the new coordinates there which equal the values of the old coordinates at the old point

$$x^{i'}(\varphi(x)) = x^i(x) .$$

This is exactly what we see when we take a polar coordinate grid and rotate it around the origin by some fixed angle. The grid associated with the dragged along coordinates is the result of that rotation of the original grid. From the coordinate grid, all the tangent structure is dragged along, since the coordinate frame vectors are obtained from the tangents to the coordinate grid, so this automatically provides a way to drag along tangent vectors from the old point to the new point. The new coordinate line of $x^{i'}$ through the new point $\varphi(x)$ with old coordinates $\varphi^i(x) = x^i(\varphi(x))$ is $\varphi(x^j + \delta^j_{i'}t)$, so by definition its derivative there of a function is the derivative by the tangent vector at $t = 0$, namely

$$\begin{aligned} \frac{\partial f}{\partial x^{i'}}(\varphi(x)) &= \left. \frac{d}{dt} \right|_{t=0} f \circ \varphi(x^j + \delta^j_{i'}t) \\ &= \left. \frac{\partial f}{\partial x^j}(\varphi(x^j + \delta^j_{i'}t)) \frac{\partial \varphi^j}{\partial x^k}(x^j + \delta^j_{i'}t) \frac{d}{dt}(x^k + \delta^k_{i'}t) \right|_{t=0} \\ &= \frac{\partial f}{\partial x^j}(\varphi(x)) \frac{\partial \varphi^j}{\partial x^k}(x) \delta^k_{i'} = \frac{\partial \varphi^j}{\partial x^i}(x) \frac{\partial f}{\partial x^j}(\varphi(x)) \end{aligned}$$

or suppressing coordinate dependence

$$\frac{\partial f}{\partial x^{i'}} = \frac{\partial \varphi^j}{\partial x^i} \frac{\partial f}{\partial x^j},$$

where the left hand side is at the new point and the right hand side is at the old point, so that if $X^i \partial_i$ is a tangent vector at x , the corresponding tangent vector dragged along to $\varphi(x)$ is defined to be the new tangent vector which has the same components as at the old point (anchored into the new grid in exactly the same way as the original tangent vector is anchored into the old grid)

$$\begin{aligned} X^i(x) \frac{\partial f}{\partial x^{i'}}(\varphi(x)) &= \underbrace{\frac{\partial \varphi^j}{\partial x^i}(x) X^i(x)}_{\equiv d\varphi(X(\varphi(x)))^j} \frac{\partial f}{\partial x^j}(\varphi(x)). \\ &\equiv d\varphi(X(\varphi(x)))^j \end{aligned}$$

The expression defined by the underbrace is the value of the components of the dragged along tangent vector at the new point in the old coordinates.

$$d\varphi(X(\varphi(x)))^j = \frac{\partial \varphi^j}{\partial x^i}(x) X^i(x)$$

or since this is a vector field X defined on the whole space, we can drag along from $\varphi^{-1}(x)$ to x by this same transformation to get the value of the dragged along field φX at x in the original coordinates

$$[\varphi X(x)]^i = \frac{\partial \varphi^j}{\partial x^i}(\varphi^{-1}(x)) X^j(\varphi^{-1}(x)).$$

This transformation law has the same form as a coordinate transformation but has a completely different interpretation, and has two crucial parts as well, first the evaluation of the component functions and the Jacobian matrix at the old point which is sent to the new point x by the transformation φ and second the linear transformation of those components due to the change

of the coordinate grid under the dragging along from the old point to the new point, using the Jacobian matrix. To get the point dependence of the component functions for dragged along 1-forms, consider the identity and its derivative

$$\varphi^i(\varphi^{-1}(x)) = x^i \quad \rightarrow \quad \frac{\partial \varphi^i}{\partial x^k}(\varphi^{-1}(x)) \frac{\partial \varphi^{-1k}}{\partial x^i}(x) = \delta^j_i,$$

which tells us exactly what the inverse Jacobian matrix components are as functions. Suppressing this point dependence we can extend these two transformations to any $\binom{p}{q}$ tensor field in the obvious way

$$[\varphi T]^{i\dots}_{j\dots} = \frac{\partial \varphi^i}{\partial x^m} \dots \frac{\partial \varphi^{-1n}}{\partial x^j} \dots T^{m\dots}_{n\dots} \circ \varphi^{-1} = [\rho^{(p,q)}(\partial \varphi / \partial x) T]^{i\dots}_{j\dots} \circ \varphi^{-1}.$$

This is most interesting to evaluate for the 1-parameter family of transformations

$$\varphi^i(x) = e^{t\xi} x^i = x^i + t\xi^i + \dots$$

generated by a vector field ξ whose Jacobian matrix is

$$\frac{\partial \varphi^i}{\partial x^j} = \delta^i_j + t \frac{\partial \xi^i}{\partial x^j} + \dots$$

so that

$$-\left. \frac{d}{dt} \right|_{t=0} \frac{\partial \varphi^i}{\partial x^j} = -\frac{\partial \xi^i}{\partial x^j},$$

and since $t \rightarrow -t$ inverts the transformation, we get

$$-\left. \frac{d}{dt} \right|_{t=0} \frac{\partial \varphi^{-1i}}{\partial x^j} = \frac{\partial \xi^i}{\partial x^j}.$$

One can evaluate the derivative of the dragged along fields $e^{t\xi} T$ with respect to the parameter t at $t = 0$ to obtain a derivative operator on tensor fields, reversed in sign for good reason

$$[\mathcal{L}_\xi T]^{i\dots}_{j\dots} = -\left. \frac{d}{dt} \right|_{t=0} [e^{t\xi} T]^{i\dots}_{j\dots}$$

In section 5.3 it was shown that for a scalar

$$\left. \frac{d}{dt} \right|_{t=0} f(e^{t\xi} x) = \xi f,$$

so the Lie derivative of a scalar is just the ordinary derivative along ξ

$$\mathcal{L}_\xi f = -\left. \frac{d}{dt} \right|_{t=0} f(e^{-t\xi} x) = \xi f.$$

For a general tensor field one merely has to add the terms for the derivatives of the Jacobian matrix

$$[\mathcal{L}_\xi T]^{i\dots}_{j\dots} = \xi T^{i\dots}_{j\dots} + \xi^i{}_{,k} T^{k\dots}_{j\dots} + \dots - \xi^k{}_{,j} T^{i\dots}_{k\dots} + \dots = \xi T^{i\dots}_{j\dots} + [\sigma^{(p,q)}(\partial\xi)T]^{i\dots}_{j\dots}.$$

For a vector field this reduces to the Lie bracket

$$[\mathcal{L}_\xi X]^i = \xi X^i - \xi^i{}_{,j} X^j = [\xi, X]^i.$$

For a metric tensor field, the formula is

$$[\mathcal{L}_\xi g]_{ij} = \xi g_{ij} + \xi^k{}_{,i} g_{kj} + \xi^k{}_{,j} g_{ik}.$$

This is the key formula.

A tensor field T is invariant under a transformation φ when the dragged along field equals the original field $\varphi^*T = T$. When T is invariant under the 1-parameter family of transformations generated by the vector field ξ , clearly its t -derivative must be zero, so the Lie derivative is zero

$$\mathcal{L}_\xi T = 0.$$

For a metric g , this is called Killing's equation

$$\mathcal{L}_\xi g = 0,$$

named after the mathematician who first introduced it. When a metric is invariant under a transformation, the transformation is called an isometry, and a vector field which generates a 1-parameter family of isometries is called a Killing vector. A Lie algebra of r linearly independent Killing vector fields generates an r -parameter family of isometries. The generators of translations and rotations of \mathbb{R}^3 are all Killing vector fields.

Note that a coordinate frame vector itself is a Killing vector field when the metric components don't depend on that coordinate, since all the terms in the formula are then identically zero. Such coordinates are said to be adapted to the Killing vector field, like the azimuthal angle θ in polar coordinates in the plane or ϕ in cylindrical or spherical coordinates in space.

Exercise 6.8.1.

(pseudo-) orthogonal group generators are Killing vector fields

Show that for metric whose coordinate components g_{ij} are constant and a matrix linear transformation $\underline{x} \rightarrow e^{t\mathbf{A}}\underline{x}$ generated by a matrix \mathbf{A} with generating vector field $\xi = A^i{}_j x^j \partial_i$, the Killing vector condition reduces to

$$0 = g_{kj} A^k{}_i + g_{ik} A^k{}_j = A_{ji} + A_{ij} = 2A_{(ij)}.$$

Thus those matrices whose index-lowered form is antisymmetric lead to Killing vector fields. For a vector field ξ whose coordinate components are constants, the Lie derivative formula reduces to the scalar formula since $\xi^i{}_{,j} = 0$, so the vector fields ∂_i which generate translations

in these coordinates are automatically Killing vectors. This describes the Lie algebras of the continuous isometry groups of all the metrics on \mathbb{R}^n whose metrics have constant components in the Cartesian coordinates of the space, namely the (pseudo-) orthogonal groups (describing the isotropy of the space) together with the translations (describing the homogeneity of the space) forming the combined inhomogeneous (pseudo-) orthogonal groups $IO(P, M)$. ■

Exercise 6.8.2.

comma to semicolon rule for Lie derivative of a metric

Confirm the “comma to semicolon rule” for the Lie derivative of the metric in a coordinate frame where one can replace the partial derivative in the Lie derivative component formula by the covariant derivative to get the same result due to the cancellation of the extra connection components due to their symmetry $\Gamma^k_{ij} = \Gamma^k_{ji}$

$$[\mathcal{L}_\xi g]_{ij} = g_{ij,k} \xi^k + \xi^k_{;i} g_{kj} + \xi^k_{;j} g_{ik} = g_{ij;k} \xi^k + \xi^k_{;i} g_{kj} + \xi^k_{;j} g_{ik} = \xi_{i;j} + \xi_{j;i}.$$

This just says that the matrix of covariant components of the covariant derivative of the vector field must be antisymmetric if the metric itself is invariant, i.e., has zero Lie derivative, so Killing’s equation takes the equivalent form

$$\mathcal{L}_\chi g_{ij} = \xi_{i;j} + \xi_{j;i} = 0.$$

For vector fields whose component matrix is linear as in the previous Exercise, like the generators of rotations (or pseudo-rotations) on \mathbb{R}^n , this just reduces to the requirement that the index-lowered matrix of components is antisymmetric, as we know already know. However, Killing’s equation is very useful in nonflat geometries where the idea of linear component vector fields does not apply. ■

Exercise 6.8.3.

1-form Lie derivative

a) The coordinate frame components of the Lie derivative of a 1-form are

$$[\mathcal{L}_\xi \sigma]_i = \xi \sigma_i + \sigma_j \xi^j_{;i}.$$

Show that this agrees with the coordinate independent formula

$$(\mathcal{L}_\xi \sigma)(X) = \xi \sigma(X) - \sigma([\xi, X]).$$

b) Use this last formula to evaluate the Lie derivatives of the dual frame 1-forms $\{\omega^i\}$ to a frame $\{e_i\}$

$$\left(\mathcal{L}_\xi \omega^c \right) (e_b) = -\omega^c([\xi, e_b]) = -[\mathcal{L}_\xi e_b]^c \leftrightarrow \mathcal{L}_\xi \omega^c = -[\xi, e_b]^c \omega^b.$$

c) Use the same formula to evaluate the Lie derivatives of the dual frame 1-forms $\{\omega^i\}$ to a frame $\{e_i\}$ with respect to the same frame vectors

$$(\mathcal{L}_{e_a} \omega^c)(e_b) = -C_{ab}^c \leftrightarrow \mathcal{L}_{e_a} \omega^c = -C_{ab}^c \omega^b.$$

■

The set of nice (differentiable) functions (scalar fields) on \mathbb{R}^n is an infinite-dimensional vector space: linear combinations of functions with constant coefficients define new functions in the same space. Any group which acts as a Lie transformation group on \mathbb{R}^n generated by some Lie algebra of vector fields $\{\xi_a\}$ also acts on this infinite-dimensional vector space as well as the corresponding vector spaces of tensor fields of a given rank. These are all therefore infinite-dimensional representations of the Lie group on these vector spaces with a corresponding Lie algebra representation for its corresponding Lie algebra. In other words, the discussion of Section 6.3 for the action of $GL(n, \mathbb{R})$ on \mathbb{R}^n as a vector space and on the tower of tensor spaces above it can be extended to the corresponding discussion of its action on each of the infinite-dimensional vector spaces of tensor fields of all possible types (including scalar fields) which live over \mathbb{R} as a manifold.

The 1-parameter subgroups $x \rightarrow e^{t\theta^a \xi_a} x$ for fixed constants θ^a of this transformation group are represented on these infinite dimensional spaces by

$$T \rightarrow e^{-t \mathcal{L}_{\theta^a \xi_a}} T.$$

Recall that when we dragged functions along by an active point transformation we had to compose them with the inverse transformation in order to bring the value at the old point to the new point where we evaluate the dragged along field. This is true for dragging along any tensor field, which explains why we need that minus sign in front of the generating vector field Lie derivative.

For the action of a matrix group with matrix Lie algebra basis $\underline{\xi}_a$, we saw that the corresponding vector field generators ξ_a had an extra minus sign in their commutators, or reversing all their signs, we got the same commutation relations

$$[\underline{\xi}_a, \underline{\xi}_b] = C_{ab}^c \underline{\xi}_c \leftrightarrow [\xi_a, \xi_b] = -C_{ab}^c \xi_c \leftrightarrow [-\xi_a, -\xi_b] = C_{ab}^c (-\xi_c).$$

In fact this generalizes to the corresponding Lie derivative operators for any tensor field as it must since they are the generators of the representations on those spaces

$$[-\mathcal{L}_{\xi_a}, -\mathcal{L}_{\xi_b}] = C_{ab}^c (-\mathcal{L}_{\xi_c}),$$

which can be rewritten in the form

$$[\mathcal{L}_{\xi_a}, \mathcal{L}_{\xi_b}] = -\mathcal{L}_{C_{ab}^c \xi_c} = \mathcal{L}_{[\xi_a, \xi_b]}.$$

This just states that the Lie derivative operation on tensor fields is a direct representation of the vector field Lie algebra, both of which are reversed in sign compared to the original

matrix Lie algebra generators. For scalar fields the Lie derivative just reduces to the vector field derivative, so this Lie algebra operator relation just reduces to the corresponding group generating vector field Lie bracket relations, which have to have that minus sign compared to the original matrix generators used to define them in order to represent their action on the space of scalar fields. The Lie derivative extends this action to tensor fields whose indices also transform under dragging along.

For ordinary Euclidean space \mathbb{R}^3 the Lie derivative with respect to the generators of the translations and rotations act on scalar, vector and 2nd rank tensor fields. Consider only vector fields where the Lie derivative formula for $\xi = L_a = S_a^i x^j \partial_i$ and $\xi^i_{;j} = S_a^i{}_{;j}$ becomes

$$[-\mathcal{L}_{L_a} X]^i = \underbrace{-\xi X^i}_{-L_a X^i} + \underbrace{\xi^i_{;j} X^j}_{(S_a)^i{}_j X^j} \equiv [(-L_a + S_a)X]^i \equiv [J_a X]^i.$$

The opposite signs here make sense because the active component transformation is generated by \underline{S}_a (forward direction), while the dragging of the function values comes from reaching back by the inverse transformation (backwards direction). Indeed the commutation relations of the sign-reversed vector fields $-L_a$ agrees with that of the corresponding matrices \underline{S}_a . The term involving the derivative of the components is said to be the result of the orbital angular momentum operator, while the term involving the linear transformation of the components is said to be the result of the spin angular momentum operator and their “sum” is called the “total angular momentum” operator. Indeed in physics with applications to complex-valued wavefunctions, one defines

$$\mathcal{L} = -iL_a, \quad \underline{\mathcal{S}}_a = i\underline{S}_a, \quad \mathcal{J}_a = \mathcal{L}_a + \mathcal{S}_a$$

to find the classic angular momentum commutation relations that appear in every text on quantum mechanics

$$[\mathcal{L}_a, \mathcal{L}_b] = i\epsilon_{abc}\mathcal{L}_c, \quad [\underline{\mathcal{S}}_a, \underline{\mathcal{S}}_b] = i\epsilon_{abc}\underline{\mathcal{S}}_c, \quad [\mathcal{J}_a, \mathcal{J}_b] = i\epsilon_{abc}\mathcal{J}_c.$$

This makes the operators “Hermitian” as described for matrices alone in Exercise 4.5.8 and in the Remark following it. For scalar fields one must extend the inner product to an integral over space of the pointwise product $\bar{\Psi}_1 \Psi$ of two complex fields in order to speak about Hermitian operators.

Similarly we could decompose the Lie derivative of a tracefree symmetric $\binom{0}{2}$ -tensor field on which the spin 2 representation of the rotation group would transform the components. Spin 1/2 “spinor fields” are necessarily complex, but we could extend this discussion to them as well since we know how to “rotate” their components with $SU(2)$ which is rigidly linked to the rotation group. None of these issues is ever very well explained in either classical electrodynamics or in quantum mechanics because there simply isn’t time and typical students don’t have the mathematical tools to handle it. Unfortunately neither do we have the time here to do justice to these claims, but we are so close, you may follow this up on your own if it interests you.

Exercise 6.8.4.**Lie derivative and the Jacobi identity**

For vector fields the Lie derivative is the vector field Lie bracket and applying the above identity for the Lie derivative commutator to another element of the vector field Lie algebra is just the Jacobi identity

$$[-\mathcal{L}_{\xi_a}, -\mathcal{L}_{\xi_b}]\xi_c = -\mathcal{L}_{[\xi_a, \xi_b]}\xi_c.$$

Expand this out in terms of the Lie brackets and rearrange the terms all to the left hand side to obtain

$$[\xi_a, [\xi_b, \xi_c]] + [\xi_b, [\xi_c, \xi_a]] + [\xi_c, [\xi_a, \xi_b]] = 0$$

to confirm this. ■

Exercise 6.8.5.**complex numbers and rotations**

When we approach the problem of finding the flow lines of a vector field generating a rotation through first order differential equations, we are confronted with the fact that the eigenvalues and eigenvectors of the corresponding antisymmetric matrix are complex. The basic difference between real exponentials $e^{\lambda t}$ and the pair $\cos \omega t, \sin \omega t$ is that the linear derivative operator d/dt acting on the real exponential “eigenfunction” $e^{\lambda t}$ generates a real eigenvalue λ : $d/dt e^{\lambda t} = \lambda e^{\lambda t}$, while when acting on the other pair it exchanges them in addition to producing the coefficient ω :

$$\begin{aligned} \frac{d}{dt}(\cos \omega t, \sin \omega t) &= \omega(-\sin \omega t, \cos \omega t) \leftrightarrow \frac{d}{dt}(\cos \omega t + i \sin \omega t) = i\omega(\cos \omega t + i \sin \omega t) \\ &\leftrightarrow \frac{d}{dt}e^{i\omega t} = i\omega e^{i\omega t}. \end{aligned}$$

The eigenfunctions of the derivative involving these functions are only the complex linear combinations which define the complex exponential and the eigenvalues are purely imaginary. The rotation group is intimately connected with complex numbers.

Suppose we introduce purely imaginary angular momentum operators

$$\mathcal{L}_a = -iL_a, \quad \underline{\mathcal{S}}_a = i\underline{S}_a, \quad \mathcal{J}_a = -i\mathcal{L}_{L_a}.$$

and their corresponding “squares” which then reverse in sign compared to the original real operators

$$\mathcal{L}^2 = \delta^{ab}\mathcal{L}_a\mathcal{L}_b = -L^2, \quad \underline{\mathcal{S}}^2 = \delta^{ab}\underline{\mathcal{S}}_a\underline{\mathcal{S}}_b = -\underline{S}^2, \quad \mathcal{J}^2 = \delta^{ab}\mathcal{J}_a\mathcal{J}_b = -J^2.$$

Note that the Lie derivative dragging operator then takes the form

$$e^{-\theta n^a \mathcal{L}_{L_a}} = e^{-i\theta n^a \mathcal{L}_{\mathcal{J}_a}}.$$

a) Show that the first two have the same commutation relations

$$[\mathcal{L}_a, \mathcal{L}_b] = i\epsilon_{abc}\mathcal{L}_c, \quad [\underline{\mathcal{S}}_a, \underline{\mathcal{S}}_b] = i\epsilon_{abc}\underline{\mathcal{S}}_c.$$

b) For the Lie derivative of a (p, q) -tensor field T define

$$[\mathcal{S}_a T]^{i\dots}_{j\dots} = [\sigma^{(p,q)}(\underline{\mathcal{S}}_a)T]^{i\dots}_{j\dots} = \mathcal{S}_a^i{}_k T^{k\dots}_{j\dots} + \dots - \mathcal{S}_a^k{}_j T^{i\dots}_{k\dots} - \dots,$$

and

$$[\mathcal{L}_a T]^{i\dots}_{j\dots} = \mathcal{L}_a T^{i\dots}_{j\dots}.$$

Show that now we can add orbital angular momentum and spin angular momentum to get total angular momentum with all positive signs

$$\mathcal{J}_a = \mathcal{L}_a + \mathcal{S}_a.$$

c) Suppose we consider the function $e^{im\phi}$ in either cylindrical or spherical coordinates. Then

$$L_3 = x\partial_y - y\partial_x = \frac{\partial}{\partial\phi} \rightarrow \mathcal{L}_3 = -i\frac{\partial}{\partial\phi}.$$

Show that $e^{im\phi}$ is an eigenfunction of \mathcal{L}_3 with eigenvalue m . To consider eigenfunctions of \mathcal{L}^2 we need to wait for the expression in spherical coordinates for the Laplacian operator in the next chapter, but we already saw that $\mathcal{S}^2 = -S^2$ has positive eigenvalues $s(s+1)$ for $s = 0, 1, 2$ in Exercise 1.7.12. Part of quantum mechanical calculations involve the “addition of angular momentum” in wave function states. In fact one can show that \mathcal{J}^2 has eigenvalues $j(j+1)$ where $j = \ell - s \dots \ell + s$ corresponding to geometrical combinations from antiparallel to parallel addition of the angular and spin angular momentum states. All of this is terribly interesting but we have to put on the brakes for now. ■

Exercise 6.8.6.

gauge invariant derivative

We have seen how each tangent space undergoes a local action of the group $GL(n, \mathbb{R})$ in the sense that each tangent space undergoes a general linear transformation through an active change of frame independently at each point (the identity representation on each tangent space) and this induces an infinite-dimensional representation on each space of tensor fields of a given type. By introducing the connection 1-form matrix and the covariant derivative, we are able to define a derivative operator that is “covariant” under transformations of the frame.

Suppose that instead of the tangent space, we associate some vector space V with each point of the space, and allow some subgroup of its general linear group to act locally on those vector spaces through its identity representation. For example, consider the 1-dimensional complex vector space $V = \mathbb{C}$, for which we can choose the basis $\underline{E} = 1$ thought of as a 1×1 column matrix attached to each point of space. A complex field on our space can then be expressed

as a multiple of this basis vector $\Psi = \Psi \underline{E}$, where the inclusion of the symbolic basis vector \underline{E} anchors the values of the complex field to the points where they are evaluated.

The group of unit complex numbers $U(1)$ acts on \mathbb{C} by multiplication $\Psi \rightarrow e^{i\Lambda}\Psi$ with Lie algebra consisting of purely imaginary numbers, preserving the squared magnitude: $|\Psi|^2 = \bar{\Psi}\Psi$, useful in quantum mechanics when that represents a physically meaningful probability distribution. The “argument” θ of a complex number $\Psi = |\Psi|e^{i\theta}$ is called its phase: changing the phase does not change the magnitude. A complex function, also called a complex scalar field, is a choice of complex number at each point of a space. As a 1-dimensional complex vector space at each point, there is no reason why we have to use the same basis complex number to express all other complex numbers as multiples (we are always free to choose a new basis of any vector space, including 1-dimensional ones), and there is no reason to expect that there should be any natural relationship between the vector spaces at different points, like the tangent spaces. The only way to compare vectors at different points is to parallel transport them to the same point, true for concrete tangent vectors tied to the differential structure of our space or abstract complex vectors that are not. If we assign $\nabla_i \underline{E} = -iA_i \underline{E}$, then the covariant derivative of the scalar field thought of as anchored to our space through the basis field \underline{E} is

$$\nabla_i(\Psi) = \nabla_i(\Psi \underline{E}) = (\partial_i \Psi) \underline{E} + \Psi \nabla_i \underline{E} = (\partial_i \Psi - iA_i) \underline{E}.$$

When this correction term is nonzero, it means that we have an association of the complex numbers with our space which depends on the point somehow. Compare this to the covariant derivative of a vector field

$$\nabla_i X^j = \partial_i X^j + (\omega^j_k)_i X^k,$$

where the correction term just arises from the covariant derivative of the basis $\{e_i\}$ of the tangent space

$$\nabla_i e_j = (\omega^j_k)_i e_k.$$

Suppose on Minkowski spacetime we consider the local action of $U(1)$ on a complex scalar field Ψ by a position-dependent transformation $\Psi(x) \rightarrow e^{i\Lambda(x)}\Psi(x)$, reflecting a local change of basis of \mathbb{C} from $\underline{E} = 1$ to the unit complex number $e^{-i\Lambda(x)}$ whose choice depends on position x : $z = z1 = (e^{i\Lambda}z)e^{-i\Lambda}$. Then the transformed scalar field will yield the same probability distribution, i.e., the same physical state. Thus this local action of $U(1)$ corresponds to the freedom in the scalar field to change its phase which does not affect physical measurements. The problem is that field equations necessarily depend on derivatives of Ψ which do change under such a change of Ψ . We can fix this by introducing a “gauge covariant derivative” which transforms exactly like Ψ under this local “gauge transformation” of the phase of the field by adding the above connection-like linear transformation term to the derivative which has values in the Lie algebra of the gauge group, namely purely imaginary numbers (so the 1-form field $A = A_i dx^i$ should be real).

$$\nabla_i \Psi = (\partial_i - iA_i)\Psi.$$

The connection 1-form $-iA = -iA_i dx^i$ is like the gauge covariant derivative of the basis vector of \mathbb{C} as a complex vector space, analogous to the connection 1-form $\underline{\omega}$ for the general linear group Lie algebra-valued 1-form (namely a matrix-valued 1-form) which we add to the

partial derivative in a coordinate frame as a linear transformation of the components of a vector field arising from the derivatives of the frame vector fields.

a) Show that when simultaneously we let

$$\Psi \rightarrow e^{i\Lambda}\Psi, \quad A_i \rightarrow A_i + \partial_i\Lambda \quad (\text{namely } A \rightarrow A + d\Lambda),$$

then the phase change of the field passes right through the covariant derivative, so that conditions on the covariant derivative of the field before and after the gauge transformation are also mapped by the same phase transformation and so do not depend on a choice of gauge

$$\nabla_i\Psi \rightarrow e^{i\Lambda}\nabla_i\Psi.$$

b) Equivalently show that the product rule to the new basis vector

$$\nabla_i (e^{-i\Lambda} \underline{E}) = -i(A_i + \partial_i\Lambda) (e^{-i\Lambda} \underline{E})$$

leads to the same transformation of the 1-form A .

This freedom to add the differential of a function to the gauge potential 1-form $A = A_i dx^i$ is exactly the freedom to redefine the vector potential of the electromagnetic field $F = dA$ without changing the physical field F , but we don't yet know how to take the differential of a 1-form, which will be called the exterior derivative and studied in Chapter 11 after a sneak preview in the next Exercise. The gauge covariant derivative allows one to write down field equations for a complex scalar field that couples to the electromagnetic field in such a way that those equations are “gauge invariant,” i.e., have the same form in any choice of “gauge,” which is a particular choice of the phase of the field. Thus the electromagnetic field vector potential appears as the connection associated with this local group action on complex scalar fields. Real scalar fields are said to be associated with neutral fields that do not carry charge. We will see later that the field 2-form $F = dA$ will be the curvature associated with this connection. Thus we have a nice coupling of the phase of a complex scalar field with the freedom to choose different vector potentials to represent the electric and magnetic fields. To actually quantify the charge in terms of a quantum of charge q which is an integer multiple of the electron charge, we actually have to make a further substitution $(A, \Lambda) \rightarrow (qA, q\Lambda)$ into this geometry. ■

Exercise 6.8.7.

vector potential for electromagnetic field

For the electromagnetic 2-form field $F = \frac{1}{2}F_{ij} dx^i \wedge dx^j$ expressed in inertial coordinates on Minkowski spacetime, introduce a “vector 4-potential” 1-form field $A = A_i dx^i$ and define twice the antisymmetrized derivative of the 1-form to be its “exterior derivative,” which we set equal to the electromagnetic 2-form

$$F_{ij} = 2\partial_{[i}A_{j]} = 2A_{[j,i]} \equiv [dA]_{ij}.$$

Show that by identifying $A = -\phi dt + A_a dx^a$, $A^\sharp = -\phi \partial_t + A^a \partial_a$ with $a, b = 1, 2, 3$, this leads to the relations

$$E_a = F_{0i} = A_{0,a} - A_{a,0} = -\phi_{,a} - \partial_t A_a, \quad B^a = \epsilon^{abc} F_{bc} = \epsilon^{abc} A_{c,b} = [\text{curl } A]^a.$$

Thus ϕ is the ordinary scalar potential for the electric field, and $A_a dx^a$ is the vector potential (1-form!) for the magnetic field. The reason for introducing the vector 4-potential is that it solves half of Maxwell's equations. ■

Exercise 6.8.8.

non-Abelian gauge theories

Exercises 1.7.12 and 4.5.9 discuss the special unitary group $SU(2)$, whose Lie algebra consists of anti-Hermitian matrices $\underline{K}^\dagger = -\underline{K}$, a basis for which is provided by the Paoli matrices multiplied by i , rescaled to have the same commutation relations as the standard basis of $SO(3, \mathbb{R})$

$$\underline{E}_a = \frac{1}{2} i \underline{\sigma}_a, \quad [\underline{E}_a, \underline{E}_b] = C^c_{ab} \underline{E}_c, \quad C^c_{ab} = \epsilon_{cab},$$

and $a, b, c = 1, 2, 3$. Consider $V = \mathbb{C}^2$ and complex-vector-valued fields $\underline{\Psi} = \langle \Psi^1, \Psi^2 \rangle = \Psi^\alpha \underline{e}_\alpha \in \mathbb{C}^2$, $\alpha, \beta = 1, 2$ on Minkowski spacetime, and let the matrix group $SU(2)$ act on these fields locally in its identity representation

$$\underline{\Psi} \rightarrow \underline{U} \underline{\Psi} = e^{\theta^a \underline{E}_a} \underline{\Psi}.$$

Note that this action leaves the magnitude of the complex-vector field (field of complex vectors, not a complex “vector field”) invariant (vector field in a quite different sense than we have been using the term so far, since it is not connected to the underlying points of the space)

$$\underline{\Psi}^\dagger \underline{\Psi} \equiv \bar{\underline{\Psi}}^T \underline{\Psi} = |\Psi^1|^2 + |\Psi^2|^2$$

since $\underline{U}^\dagger \underline{U} = \underline{I}$ is the unitary condition on the matrix group. recall that the combined transpose and complex conjugate operation \dagger is called the Hermitian conjugate.

If the physical observations of a theory involving these complex fields do not depend on the “phase”, i.e., only the magnitude of the field matters, then nothing should depend on the particular choice of “gauge,” i.e., what transformation \underline{U} we apply to the field. All physical theories involve the derivatives of the fields so we need a “gauge covariant derivative” which has the same transformation behavior as $\underline{\Psi}$ itself

$$\underline{\Psi} \rightarrow \underline{U} \underline{\Psi} \quad \text{implies} \quad \nabla \underline{\Psi} \rightarrow \underline{U} \nabla \underline{\Psi} = \nabla (\underline{U} \underline{\Psi}).$$

This is the same condition we impose on the covariant derivative in terms of the component matrices of vector fields under a change of frame on the tangent spaces, so it is not surprising that the same mathematics characterizes the connection 1-forms for local changes of basis on the complex vector spaces of the fields.

Show that the following gauge invariant derivative

$$\nabla_i \underline{\Psi} = (\partial_i + A^a \underline{E}_a) \underline{\Psi}$$

is invariant provided that the gauge field 1-form matrix $\underline{A} = A^a \underline{E}_a$ transforms exactly like the connection 1-form matrix $\underline{\omega}$ of a metric connection as discussed in Section 6.3

$$\underline{A} \rightarrow \underline{U} \underline{A} \underline{U}^{-1} + \underline{U} d\underline{U}^{-1}.$$

This identifies the gauge connection with the gauge covariant derivative of the basis vectors of $V = \mathbb{C}^2$

$$\nabla_i \underline{e}_\alpha = A_i^c (\underline{E}_c)^\beta{}_\alpha \underline{e}_\beta = \Gamma^\beta{}_{i\alpha} \underline{e}_\beta.$$

Thus this leads to three 1-form fields $A_i dx^i$ which describe this geometry.

The electroweak unified field theory is based on the direct product group $SU(2) \times U(1)$ which introduces three 1-form $SU(2)$ gauge fields in addition to the $U(1)$ gauge field $A = A_i dx^i$ of electromagnetism, with a linear transformation on the combined Lie algebra coupled to the so called Higgs mechanism involving the now famous Higgs field which “gives mass” to three of these four 1-form fields (the W^\pm and Z bosons), leaving the vector potential A for the electromagnetic field massless. Covector fields are spin 1 fields which quantum mechanically correspond to particles called bosons, like the spin 1 photon which is the quantum particle corresponding to the classical electromagnetic vector potential A . All of this is beyond us, but the basic setup as a gauge covariant derivative associated with a “non-Abelian gauge field theory” is not. This geometrification of physical theories was one of the great successes of the last century. The strong interactions correspond to the group $SU(3)$ acting on \mathbb{C}^3 (the quark field space) with eight “gluon” gauge fields (3-dimensional Lie algebra) and grand unified theories to the direct product of the electroweak and strong interaction theories. [One must further introduce a charge coupling constant by replacing A by gA in the covariant derivative, where g is the basic charge unit of the theory, like the electronic charge $e > 0$ in electromagnetism, but that is a detail requiring a more in-depth treatment of this topic.] ■

Exercise 6.8.9.

angular momentum ladder operators and representation theory for $SU(2) \sim SO(3, \mathbb{R})$

When the rotation group acts on integer spin tensor fields or the more fundamental group $SU(2)$ acts on spin 1/2 complex spinor fields, we have an infinite-dimensional representation, but it can be decomposed into an infinite-direct sum of finite-dimensional representations characterized by eigenvalues of the angular momentum operators which represent the matrix generators of their isomorphic Lie algebras $so(3, \mathbb{R}) \sim su(2)$. Let’s consider the scalar field case for simplicity and consider the spherical harmonics on the unit sphere on which the rotation vector generators act. While the following matrix calculations can be done by hand, there is little to be gained by doing so. A computer algebra system makes them painless.

a) Starting with the angular momentum operators (Cartesian components of orbital angular momentum) \mathcal{L}_a satisfying $[\mathcal{L}_a, \mathcal{L}_b] = i\epsilon_{abc}\mathcal{L}_c$, introduce the ladder operators first introduced in Exercise 5.9.9

$$\mathcal{L}_{\pm} = \mathcal{L}_1 \pm i\mathcal{L}_2,$$

and derive their commutator relations from those of L_a

$$[\mathcal{L}_3, \mathcal{L}_{\pm}] = \pm\mathcal{L}_{\pm}, \quad [\mathcal{L}_+, \mathcal{L}_-] = 2\mathcal{L}_3.$$

Since all the L_a commute with L^2 , so to do these ladder operators $[\mathcal{L}^2, \mathcal{L}_{\pm}] = 0$. Applying this last equality to $Y_{\ell m}$ leads to $\mathcal{L}^2\mathcal{L}_{\pm}Y_{\ell m} = \ell(\ell+1)Y_{\ell m}$ (check it!) so the result of applying these operators leads to functions with the same eigenvalue ℓ as we started with.

b) Suppose $Y_{\ell m}$ is an eigenfunction of the operators \mathcal{L}^2 and \mathcal{L}_3 :

$$\mathcal{L}^2 Y_{\ell m} = \ell(\ell+1)Y_{\ell m}, \quad \mathcal{L}_3 Y_{\ell m} = mY_{\ell m}.$$

Then apply these commutation relations expanded out ($[A, B] = AB - BA$) to the functions $Y_{\ell m}$ in order to conclude that

$$\mathcal{L}_{\pm} Y_{\ell m} = (m \pm 1)Y_{\ell m},$$

so that (assuming the eigenfunctions are unique as is the case when you study these matters with more time than we have)

$$\mathcal{L}_{\pm} Y_{\ell m} = C(\ell, m, \pm)Y_{\ell m \pm 1}$$

which means that we have increased/decreased the eigenvalue m by 1, although we don't know the coefficient of the result.

c) Show by expanding out that

$$\mathcal{L}_- \mathcal{L}_+ = \mathcal{L}^2 - \mathcal{L}_3(\mathcal{L}_3 + 1), \quad \mathcal{L}_+ \mathcal{L}_- = \mathcal{L}^2 + \mathcal{L}_3(\mathcal{L}_3 + 1),$$

so that

$$\mathcal{L}_- \mathcal{L}_+ Y_{\ell m} = [\ell(\ell+1) - m(m+1)]Y_{\ell m}, \quad \mathcal{L}_+ \mathcal{L}_- Y_{\ell m} = [\ell(\ell+1) + m(m+1)]Y_{\ell m}.$$

Thus if we apply the first relation when $m = \ell$, the result is 0, but that can only be true if $\mathcal{L}_+ Y_{\ell \ell} = 0$ since otherwise the result should be a nonzero multiple of $Y_{\ell \ell}$. Similarly applying the second relation when $m = -\ell$, it implies $\mathcal{L}_- Y_{\ell -\ell} = 0$. (We really need to go a little deeper into this argument than we can afford here to draw these conclusions.) The result (when explained with more details) is that for each integer value of ℓ , we get $2\ell + 1$ different eigenfunctions that belong to a finite-dimensional representation of $SO(3, \mathbb{R})$ of dimension $2\ell + 1$ on which the rotation group acts as a linear transformation group, indeed when expressed in terms of the basis $\{Y_{\ell m}\}$ of this space for a given fixed value of ℓ , as a matrix group with the same matrix product relations as the rotations they represent. The ladder operators move us from each eigenfunction to the nearest neighbor in terms of its value of m , terminating at the endpoint values $\pm\ell$.

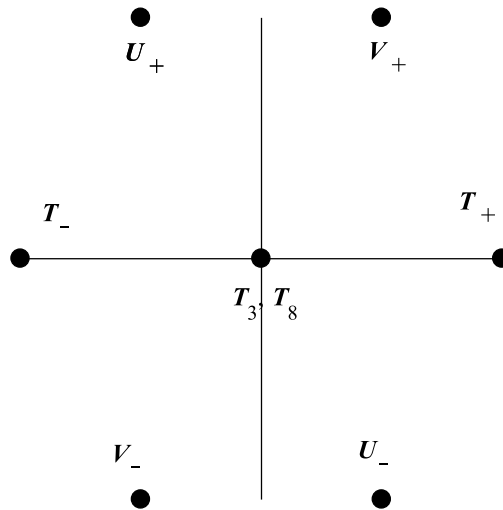


Figure 6.5: The octet diagram of the 8 Lie algebra basis matrices of the Lie algebra $su(3)$ consists of two commuting matrices at the center and three pairs of ladder operator matrices associated with the three $su(2)$ subalgebras. The horizontal axis corresponds to the usual $su(2)$ matrices.

The identity representation $\ell = 1$, $2\ell + 1 = 3$ has $m = -1, 0, 1$ allowing us to have a mental picture of the spin vector in that representation to be aligned with the z -axis ($m = 1$), aligned with the negative z -axis ($m = -1$), or aligned with neither direction ($m = 0$), thought of as in some horizontal direction. This is the fundamental representation of the Lie algebra $so(3), \mathbb{R} \sim su(2)$ and its commutation relations and the mental picture that we have of its eigenstates, whether they are functions on the sphere as in this example, or points in space as in the defining representation.

■

Remark.

The group $SU(3)$ acting on \mathbb{C}^3 (the quark color charge space of quantum chromodynamics) has three $SU(2)$ subgroups just like $SO(3, \mathbb{R})$ has three $SO(2, \mathbb{R})$ subgroups (rotations about the three independent axes, or in the three orthogonal planes) so its matrix Lie algebra consists of the union of these three subalgebras in each 2×2 submatrix of the 3×3 matrix of the Lie algebra, BUT we have to remove one because the three copies of the diagonal matrix $\sigma_3 = \langle \langle 1|0 \rangle, \langle 0, -1 \rangle \rangle$ in each of these three blocks of the matrix are not independent since there are only two independent tracefree real diagonal matrices. The bases $\underline{T}_a = \frac{I}{2} \underline{\lambda}_a$ of the Lie algebras of these three copies of $SU(2)$ expressed in terms of the 8 Gell-Mann matrices $\underline{\lambda}_a$ analogous to the original 3 Pauli matrices are here given row by row, with the diagonal matrix

last in each case

$$\begin{aligned}
 \text{1,2 block: } \lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\
 \text{1,3 block: } \lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \lambda_5 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, & & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \\
 \text{2,3 block: } \lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, & & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \\
 & & & \lambda_8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.
 \end{aligned}$$

For a basis of the tracefree diagonal matrices, λ_3 is kept and joined by the combination λ_8 which reduces to a multiple of the identity matrix on the first $SU(2)$ subalgebra. The corresponding two matrices $\underline{T}_3, \underline{T}_8$ are simultaneously diagonal so their representatives in any representation will also be so, i.e., one can classify the bases of the representation spaces by their pair of eigenvalues. The remaining 3 pairs of $SU(2)$ matrix generators can be combined into the ladder combinations which then step through the lattice of discrete values of those two eigenvalues, leading to the famous 7 point octet plane diagram of Fig. 6.5 with $\underline{T}_3, \underline{T}_8$ both at the origin and the three pairs of ladder operators located at grid points lying on a circle along three lines corresponding to the angles which are multiples $\pi/3$ along which the ladder operators move in the space of the pairs of eigenvalues of $(\underline{T}_3, \underline{T}_8)$.

An $su(3)$ Lie algebra valued-connection 1-form $A^a \underline{T}_a$ thus involves 8 coefficient fields A^a which are the gluon fields of the strong interactions which mediate the forces between the fundamental quark particles in the same way the photon field (electromagnetic field) mediates the electromagnetic forces between the electromagnetic fundamental charges. But that is theoretical physics and way beyond our scope! ▲

Exercise 6.8.10.

Gell-Mann matrices

Show that like the Pauli matrices, the Gell-Mann matrices listed above are an orthonormal set modulo a common factor of 2 under the trace inner product: $\text{Tr } \lambda_a \lambda_b = 2\delta_{ab}$. ■

6.9 Noncoordinate frames and $SO(3, \mathbb{R})$

The best example of geometry where noncoordinate frames are crucial is the rotation group itself which has its own family of geometries which are compatible with the group structure itself. Recall Exercise 1.7.10 in which we described the orientation of a rigid body like a symmetrical top with one point fixed on which it spins by an active rotation $\underline{R} = e^{\phi \underline{k}_3} e^{\theta \underline{k}_1} e^{\psi \underline{k}_3}$ of the space-fixed axes $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ to the time-dependent body fixed axes $\{\hat{e}_{1'}, \hat{e}_{2'}, \hat{e}_{3'}\}$ whose corresponding coordinates are related by $x^{i'} = R(\theta)^{-1i}{}_j x^j$, $x^i = R(\theta)^i{}_j x^{j'}$ and we calculated the components of the angular velocities of the points fixed in the body in both coordinate systems

$$\underline{\Omega}^a = \frac{\omega^a}{dt}, \quad \underline{\Omega}^{a'} = \frac{\tilde{\omega}^a}{dt},$$

where

$$\begin{aligned} \omega^1 &= \cos \psi d\theta + \sin \theta \sin \psi d\phi, & \tilde{\omega}^1 &= \cos \phi d\theta + \sin \theta \sin \phi d\psi, \\ \omega^2 &= -\sin \psi d\theta + \sin \theta \cos \psi d\phi, & \tilde{\omega}^2 &= \sin \phi d\theta - \sin \theta \cos \phi d\psi, \\ \omega^3 &= d\psi + \cos \theta d\phi, & \tilde{\omega}^3 &= d\phi + \cos \theta d\psi, \end{aligned}$$

were defined by

$$\underline{R}^{-1} d\underline{R} = \omega^a \underline{L}_a, \quad d\underline{R} \underline{R}^{-1} = \tilde{\omega}^a \underline{L}_a.$$

The isotropic combination of the first two of each set yields

$$\begin{aligned} (\omega^1)^2 + (\omega^2)^2 &= d\theta^2 + \sin^2 \theta d\phi^2, \\ (\tilde{\omega}^1)^2 + (\tilde{\omega}^2)^2 &= d\theta^2 + \sin^2 \theta d\psi^2, \end{aligned}$$

which shows clearly the metric on unit 2-spheres where θ is the polar angle down from the vertical and respectively ϕ and ψ are azimuthal angles.

In Exercise 1.7.10 followed by Exercise 4.5.7 it was shown that if we left translate the rotation group by left multiplying its matrix \underline{R} by a fixed rotation $\underline{R} \rightarrow \underline{R}_0 \underline{R}$ corresponding to a fixed rotation of the space-fixed axes $x^i \rightarrow x^i = (\underline{R}_0 \underline{R}(\theta))^i{}_j x^{j'} = R_0^i{}_j x^{j'}$, then notice that $\underline{R}^{-1} d\underline{R}$ does not change, so the 1-forms ω^a are invariant under left translation of the group into itself. Similarly the 1-forms $\tilde{\omega}^a$ are invariant under right translation of the group into itself, corresponding to a fixed rotation of the body-fixed axes. $x^{i'} \rightarrow x^{i'} = (\underline{R}(\theta) \underline{R}_0)^{-1i'}{}_j x^j = R_0^{-1i'}{}_j x^j$. Finally it was shown that the bi-invariant metric

$$ds^2 = \frac{a^2}{4} \delta_{ab} \tilde{\omega}^a \tilde{\omega}^b = \frac{a^2}{4} \delta_{ab} \omega^a \omega^b$$

corresponds to the metric on a 3-sphere of radius a in \mathbb{R}^4 through a 2-to-1 correspondence between the group $SU(2)$ whose group manifold is the unit 3-sphere and the rotation group $SO(3, \mathbb{R})$ in which antipodal points on S^3 are identified to yield the projective 3-sphere $\mathbb{P}S^3$.

The two sets of respectively left invariant and right invariant 1-forms

$$\omega^a = \omega^a{}_i d\theta^i, \quad \tilde{\omega}^a = \tilde{\omega}^a{}_i d\theta^i = R(\theta)^a{}_b \omega^b$$

which coincide at the identity matrix are each dual to a correspondingly invariant frame of vector fields obtained by simply inverting the matrix of coordinate components of these 1-forms, easily done with a computer algebra system. Define the unhatted vector fields by

$$\begin{aligned} e_1 &= \cos \psi \partial_\theta - \frac{\sin \psi}{\sin \theta} (\cos \theta \partial_\psi + \partial_\phi) , & \tilde{e}_1 &= \cos \phi \partial_\theta + \frac{\sin \phi}{\sin \theta} (\partial_\psi - \cos \theta \partial_\phi) , \\ e_2 &= -\sin \psi \partial_\theta + \frac{\cos \psi}{\sin \theta} (-\cos \theta \partial_\psi + \partial_\phi) , & \tilde{e}_2 &= \sin \phi \partial_\theta + \frac{\cos \phi}{\sin \theta} (-\partial_\psi + \cos \theta \partial_\phi) , \\ e_3 &= \partial_\psi , & \tilde{e}_3 &= \partial_\phi . \end{aligned}$$

Exercise 6.9.1.

angular momentum commutation relations

Verify the Lie bracket relations

$$[e_a, e_b] = C^c{}_{ab} e_c , \quad [\tilde{e}_a, \tilde{e}_b] = -C^c{}_{ab} \tilde{e}_c , \quad [e_a, \tilde{e}_b] = 0 , \quad C^c{}_{ab} = \epsilon_{cab} .$$

■

These 3-dimensional vector field Lie algebras each generate a 3-dimensional transformation group. The left invariant vector fields generate the right translations of the group into itself, while the right invariant vector fields generate the left translations of the group into itself. These two groups commute with each other, which is reflected in the fact that the two Lie algebras commute. The inverse map takes the above left invariant frame and maps it into the sign reversed right invariant frame, so the left and right invariant metrics describe the same abstract geometry.

Consider diagonal metric component matrices in each of these frames

$$\begin{aligned} ds_L^2 &= g_{L11}(\omega^1)^2 + g_{L22}(\omega^2)^2 + g_{L33}(\omega^3)^2 , \\ ds_R^2 &= g_{R11}(\tilde{\omega}^1)^2 + g_{R22}(\tilde{\omega}^2)^2 + g_{R33}(\tilde{\omega}^3)^2 . \end{aligned}$$

In the special cases when the first two coefficients are equal, these become

$$\begin{aligned} ds_L^2 &= g_{L11} (d\theta^2 + \sin^2 \theta d\phi^2) + g_{L33} (d\psi + \cos \theta d\phi)^2 , \\ ds_R^2 &= g_{R11} (d\theta^2 + \sin^2 \theta d\psi^2) + g_{R33} (d\phi + \cos \theta d\psi)^2 , \end{aligned}$$

These are both independent of the coordinates ψ, ϕ , making ∂_ψ and ∂_ϕ both Killing vector fields as we will learn about in Chapter 6. This will enable us to understand the dynamics of a symmetric top in Chapter 8.

Chapter 7

More on covariant derivatives

Partial differential equations govern many aspects of our real world. Partial differential equations on space or spacetime are key to quantify how fields behave as a function of position in space or in spacetime. However, these equations are not arbitrary but associated with the basic geometry of space and its symmetries. Vector analysis deals with first and second order partial differential operators that enter into these geometrically related partial differential equations that describe Newtonian gravity, electromagnetism, fluid dynamics, quantum mechanics, etc., all of which deal with fields and their differential properties examined both in orthonormal Cartesian coordinates as well as the common curvilinear coordinate systems associated with common symmetries of space. The first order vector operators of gradient, curl and divergence and the second order Laplacian are key players in this game, and the Laplacian also reconnects with the idea of total angular momentum whose description leads to the spherical harmonics which are crucial in characterizing how functions on the sphere behave under rotation. In this chapter we explore the way these operators are related to the Euclidean metric of \mathbb{R}^3 , but some aspects of these operators do not need the metric and instead have to do with the differential properties of differential forms. In chapter 11 we will extend the differential operator d from 0-forms (scalar fields) to define the exterior derivative d on the algebra of differential forms which is necessary for understanding and generalizing the line, surface and volume integrals of scalar and vector fields in space and their interrelationships with the differential operators of this chapter through the vector theorems known as Gauss's law and Stokes' theorem, as well as the line integral theorem for conservative vector fields.

7.1 Gradient, curl and divergence

In multivariable calculus the operations of gradient (“grad”), divergence (“div”) and their composition called the Laplacian (“div grad”), together with the curl, play important roles in the various integral theorems which relate line, surface and volume integrals on \mathbb{R}^3 . These turn out to all be connected together in a beautiful relationship that is hidden in a first course on the topic but which will be uncovered in Chapter 11.

The differential of a function (“scalar field”)

$$\begin{aligned} df &= f_{,i} dx^i, & f_{,i} &= \frac{\partial f}{\partial x^i} = df \left(\frac{\partial}{\partial x^i} \right), & (\text{coordinate frame}) \\ df &= f_{,i} \omega^i, & f_{,i} &= e_i f = df(e_i) & (\text{arbitrary frame}) \end{aligned}$$

is a covector field or 1-form field or simply a 1-form in the standard terminology.

In Cartesian coordinates on \mathbb{R}^n , the gradient $\vec{\nabla} f \equiv \text{grad } f$ is a vector field whose components are the corresponding partial derivatives of f

$$\begin{aligned} [\text{grad } f]^i &= [\vec{\nabla} f]^i = \delta^{ij} f_{,j}, \\ \text{grad } f &= \vec{\nabla} f = \delta^{ij} f_{,j} \frac{\partial}{\partial x^i} = df^\sharp. \end{aligned}$$

The Kronecker delta is necessary to respect index positioning and tells us that we are actually using the Euclidean metric to raise the index on the 1-form df to obtain a vector field $\vec{\nabla} f$.

The same relation can be used to evaluate the gradient in general coordinates or in an orthonormal frame with respect to the Euclidean metric or with respect to any other metric

$$\text{grad } f = \vec{\nabla} f = df^\sharp = g^{ij} f_{,j} e_i.$$

While the differential df is completely independent of a metric, the gradient only can be defined with the use of a metric.

For a function, covariant and ordinary differentiation coincide, so one can also write

$$[\vec{\nabla} f]^i = g^{ij} f_{,j} \equiv f^{;i} \equiv \nabla^i f.$$

In other words the composed operator $\vec{\nabla} = \sharp \circ \nabla$ (usually called “del”) consists of covariant differentiation followed by raising the derivative index when acting on functions.

Evaluating the differential of a function on a vector field (no metric needed) or taking the inner product of the gradient and a vector (metric required) leads to the derivative of the function along that vector field

$$\begin{aligned} df(X) &= Xf = \nabla_X f = X^i \nabla_i f & (\text{no metric required}) \\ &= X \cdot \vec{\nabla} f = g_{ij} X^i \nabla^j f. & (\text{metric required}) \end{aligned}$$

If a vector field X is tangent to a level hypersurface $f = \text{const}$ of the function f , then the derivative of f along X is zero which implies that X is orthogonal to $\vec{\nabla} f$, or turning it around,

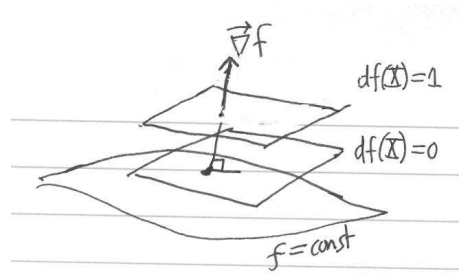


Figure 7.1: Visualizing the gradient vector with the plane representation of the differential of a function in \mathbb{R}^3 .

the gradient is orthogonal to the space of tangent vectors which are in fact tangent to the level surface of f at each point—it is a normal to the tangent plane. Without a metric one only has the differential df whose hyperplanes in the tangent space describe the linear approximation to the increment of f away from each point, and *no normal vector*!

Exercise 7.1.1.

gradient in cylindrical and spherical coordinates

Evaluate $\vec{\nabla} f$ for $f = x^2 - y^2$ on \mathbb{R}^3 in both cylindrical and spherical coordinates. ■

In Cartesian coordinates on \mathbb{R}^3 , the curl of a vector field is given by

$$\text{curl } X = \epsilon^{kij} \partial_i X_j \partial_k = \eta^{kij} \partial_i X_j \partial_k = \eta^{kij} X_{j,i} \partial_k = \eta^{kij} X_{j;i} \partial_k, \quad X_j = g_{j\ell} X^\ell.$$

Since the Christoffel symbols are symmetric in the lower indices $\Gamma^k_{[ij]} = 0$, the relation

$$\eta^{kij} X_{j;i} = \eta^{kij} (X_{j,i} + X_k \Gamma^k_{ij}) = \eta^{kij} X_{j,i}$$

means that the final two representations of the curl are equivalent in any coordinate system and reveal that the curl of a vector field is a sequence of three (four?) operations: lowering of its index to obtain a 1-form, the antisymmetric derivative of the 1-form coordinate components, and the metric dual on that pair of antisymmetric indices to give the resulting 1-form whose index raising yields a vector field. If we introduce the cross product in any coordinate frame by

$$X \times Y = \eta^i_{jk} X^j Y^k \frac{\partial}{\partial x^i},$$

then we can represent the curl as “grad cross” (usually called “del cross”)

$$\text{curl } X = \vec{\nabla} \times X,$$

where the covariant indexed covariant derivative operator $\nabla = “\vec{\nabla} \cdot”$ can be thought of as the result of taking the inner product with the contravariant indexed covariant derivative in the sense $\nabla_j X^i = g_{jk} \nabla^k X^i$.

In Cartesian coordinates on \mathbb{R}^3 , the divergence of a vector field (“del dot”) is given by

$$\operatorname{div} X = \frac{\partial X^i}{\partial x^i} = X^i_{,i} = X^i_{;i} = \vec{\nabla} \cdot X.$$

The formula involving the covariant derivative for $\operatorname{div} X$

$$\operatorname{div} X = X^i_{;i} = X^i_{,i} + \Gamma^i_{ij} X^j$$

is well-defined in any frame or coordinate system and indeed for any dimension n .

An alternative formula can be derived for the divergence of a vector field by re-expressing the contraction Γ^i_{ij} in terms of the metric. The result is somewhat simpler to use in practice. Contracting the general formula for Γ^k_{ij} gives

$$\begin{aligned} \Gamma^i_{ik} &= \frac{1}{2} g^{i\ell} [g_{\ell i,k} - g_{ik,\ell} + g_{k\ell,i}] + \frac{1}{2} (\underbrace{C^i_{ik}}_{-C^i_{ki}} + \underbrace{C^i_{ik}}_{C^i_{ik}} + \underbrace{C^i_{k i}}_{C_{kji} g^{ji}}) \\ &= \frac{1}{2} g^{i\ell} g_{\ell i,k} - \frac{1}{2} g^{i\ell} g_{ki,\ell} + \underbrace{\frac{1}{2} g^{i\ell} g_{k\ell,i}}_{g^{i\ell} g_{ki,\ell}} - \underbrace{\frac{1}{2} C^i_{ki} - \frac{1}{2} C^i_{ki}}_{-C^i_{ki}} + \underbrace{\frac{1}{2} C_{kji} g^{ji}}_{=0} \\ &= \underbrace{\frac{1}{2} g^{i\ell} g_{\ell i,k}}_{[\ln(\det \underline{g})^{1/2}]_{,k}} - C^i_{ki}, \end{aligned}$$

where we have used three facts: first, for any pair of contracted indices the order of the upper and lower indices in the contracted pair does not matter, i.e., $T^i_i = g_{ij} T^{ij} = T^i_i$, and second, when a pair of antisymmetric indices is contracted with a pair of symmetric indices $X_{ij} Y^{ij}$ the result is zero, so the contraction of the final two indices of the structure function symbol is zero, and finally we recognized the derivative of the square root of the determinant of the metric component matrix in the last line, a fact derived in Section 2.3

$$d \ln(|\det \underline{g}|^{1/2}) = \frac{1}{2} d \ln(|\det \underline{g}|) = \frac{1}{2} \operatorname{Tr} \underline{g}^{-1} d\underline{g}.$$

Thus we get the final formula

$$\Gamma^i_{ik} = [\ln(|\det \underline{g}|^{1/2})]_{,k} - C^i_{ki}.$$

Inserting this in our formula for the divergence we get

$$\begin{aligned} \operatorname{div} X &= X^i_{,i} + \underbrace{\Gamma^i_{ik} X^k}_{[\ln(|\det \underline{g}|^{1/2})]_{,k} X^k - C^i_{ki} X^k} \\ &= |\det \underline{g}|^{-1/2} [|\det \underline{g}|^{-1/2} X^i]_{,i} - C^i_{ki} X^k, \end{aligned}$$

where the final form of the right hand side is just a consequence of the product rule for differentiation.

Summarizing

$$\begin{aligned}\operatorname{div} X &= X^i{}_{;i} = |\det \underline{g}|^{-1/2} [|\det \underline{g}|^{1/2} X^i]_{,i} && \text{(coordinate frame)} \\ &= |\det \underline{g}|^{-1/2} [|\det \underline{g}|^{1/2} X^i]_{,i} - C^k{}_{ik} X^i. && \text{(arbitrary frame)}\end{aligned}$$

These formulas for the divergence operator are valid for any dimension n , while the cross product crucially depends on the dimension $n = 3$ for the dual of an antisymmetric pair of indices to result in a single index object.

Notice that the divergence operator *only* involves the metric through the factor $|\det \underline{g}|^{1/2} = \eta_{1\dots n}$ which is the only independent nonzero component of the unit volume n -form associated with the metric. It does not care about the individual metric components. Any metrics whose unit volume forms coincide will yield the same divergence operator for vector fields.

Exercise 7.1.2.

curl and div in cylindrical coordinates

In Section 5.7 and the following vector field on \mathbb{R}^3 was evaluated in cylindrical coordinates as a byproduct of evaluating ∂_ρ in Cartesian coordinates, and then in spherical coordinates in Exercise 5.7 with the result

$$X = x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} = \rho \frac{\partial}{\partial \rho} = r \sin \theta \left[\sin \theta \frac{\partial}{\partial r} + \frac{\cos \theta}{r} \frac{\partial}{\partial \theta} \right].$$

a) Check that $\operatorname{curl} X = 0$, $\operatorname{div} X = 2$ in cylindrical coordinates as is clear from the Cartesian coordinate expression.

b) If you are feeling ambitious, repeat for spherical coordinates.

■

Exercise 7.1.3.

more curl and div in cylindrical coordinates

Consider the function from Exercise 5.8.3

$$f = xy = \rho^2 \sin \phi \cos \phi = \frac{1}{2} \rho^2 \sin 2\phi = \frac{1}{2} r^2 \sin^2 \theta \sin 2\phi,$$

expressed respectively in Cartesian, cylindrical and spherical coordinates. Then in Cartesian coordinates it is obvious that

$$\begin{aligned}df &= y dx + x dy = X^\flat, \\ \vec{\nabla} f &= [df]^\sharp = y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} = X,\end{aligned}$$

where the vector field X was transformed to cylindrical and spherical coordinates in Example 5.7.1 and in spherical coordinates in Example 5.8.1

$$X = \rho \sin 2\phi \frac{\partial}{\partial \rho} + \cos 2\phi \frac{\partial}{\partial \phi} = r \sin \theta \sin 2\phi \left(\sin \theta \frac{\partial}{\partial r} + \frac{\cos \theta}{r} \frac{\partial}{\partial \theta} \right) + \cos 2\phi \frac{\partial}{\partial \phi},$$

$$X^\flat = \rho \sin 2\phi d\rho + \rho^2 \cos 2\phi d\phi = \sin \theta \sin 2\phi (r \sin \theta dr + r^2 \cos \theta d\theta) + r^2 \sin^2 \theta \cos 2\phi d\phi.$$

Lowering its indices just rescales each component by the corresponding diagonal metric component.

a) This vector field clearly has zero divergence and curl from the easy Cartesian coordinate evaluation. Compute df and $\text{grad } f = \vec{\nabla} f$ in cylindrical coordinates and confirm these evaluations.

b) If you are feeling ambitious, repeat for spherical coordinates. If not, wait for easy formulas in an exercise in the next section. ■

Exercise 7.1.4.

still more curl and div in cylindrical coordinates

Repeat the previous exercise for $f = x^2 - y^2$. ■

Exercise 7.1.5.

still more curl and div in cylindrical and spherical coordinates

Consider the following vector field expressed in Cartesian and cylindrical/spherical coordinates

$$x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} = \frac{\partial}{\partial \phi}.$$

This vector field clearly has 0 divergence and $\text{curl } X = 2 \partial/\partial z$ from the easy Cartesian coordinate evaluation. Confirm this in cylindrical and spherical coordinates. ■

7.2 Second covariant derivatives and the Laplacian

The notation for a second covariant derivative

$$T^i_{j;k\ell} \equiv T^i_{j;k;\ell} \equiv [\nabla\nabla T]^i_{j k \ell}$$

is always abbreviated to $T^i_{j;k\ell}$. In other words the semi-colon is used to separate the additional covariant derivative indices from the original tensor indices, no matter how many extra derivative indices are added.

For a function $\nabla f = df = f_{;i} \omega^i = f_{,i} \omega^i$ is the first covariant derivative and

$$\nabla\nabla f = f_{;ij} \omega^i \otimes \omega^j$$

is the second covariant derivative. The same notation is extended to the comma for repeated ordinary differentiation: $f_{,i,j} \equiv f_{,ij}$.

The Laplacian of a function is defined in Cartesian coordinates on \mathbb{R}^n by

$$\nabla^2 f = \vec{\nabla} \cdot \vec{\nabla} f = \operatorname{div} \operatorname{grad} f = \delta^{ij} f_{,ij} \quad \left(= \frac{\partial^2 f}{\partial (x^1)^2} + \cdots + \frac{\partial^2 f}{\partial (x^n)^2} \right).$$

Therefore in any frame or coordinate system one has

$$\nabla^2 f = \operatorname{div} \operatorname{grad} f = g^{ij} f_{;ij} = (g^{ij} f_{;i})_{;j}$$

since both the metric and inverse metric are covariant constant: $g^{ij}_{;k} = 0$. For this reason, raising the first derivative index and then differentiating again is equivalent to differentiating twice and then contracting with the inverse metric.

Using the formula for the divergence, letting $X = \vec{\nabla} f = \operatorname{grad} f$, we get

$$\nabla^2 f = \operatorname{div} \operatorname{grad} f = (\det \underline{g})^{-1/2} [(\det \underline{g})^{-1/2} g^{ij} f_{,j}]_{,i} - \underbrace{C^k_{ik} g^{kl} f_{,l}}_{\text{vanishes for coordinate frame}}$$

Exercise 7.2.1.

harmonic coordinates

Suppose we apply the previous formula to the coordinate functions themselves in their coordinate frame

$$\nabla^2 x^k = \operatorname{div} \operatorname{grad} x^k = (\det \underline{g})^{-1/2} [(\det \underline{g})^{-1/2} g^{ij} x^k_{,j}]_{,i}.$$

This can be simplified with the relations $x^k_{,j} = \delta^k_j$. On the other hand one can simply evaluate the covariant derivative formula for a covector field $X_k = f_{;k} = f_{,k}$

$$\begin{aligned} X_{j;i} &= X_{j,i} - X_m \Gamma^m_{ij} \xrightarrow{g^{ij}} X^i_{;i} = X_{i,i} - X_m \Gamma^{mi}_i, \\ \nabla^2 f &= f_{;i}{}^{;i} = f_{,i}{}^{,i} - f_{,m} \Gamma^{mi}_i, \end{aligned}$$

and apply this formula directly, and comparison leads to a formula we derived in a previous problem for the contraction of the final two indices of the components of the connection. Thus one arrives at two versions of the formula for the Laplacian of the coordinates themselves

$$\nabla^2 x^k = -|\det \underline{g}|^{-1/2} (|\det \underline{g}|^{1/2} g^{ki})_{,i} = -\Gamma^{ki}_{i}.$$

These turn out to be important in general relativity. Coordinates for which the Laplacian of the coordinates themselves vanishes are called harmonic coordinates. The Cartesian coordinates on \mathbb{R}^n are such coordinates, with covariant constant frame vector fields. On curved spaces where covariant constant coordinate frames are not possible, harmonic coordinates can be used to get as close to Cartesian-like coordinates as possible, with respect to some of their properties. ■

Exercise 7.2.2.

harmonic function

On \mathbb{R}^3 we compute

$$\nabla^2(x^2 - y^2) = \frac{\partial}{\partial x}(2x) - \frac{\partial}{\partial y}(2y) = 2 - 2 = 0,$$

and then transform the function to cylindrical and spherical coordinates

$$f = x^2 - y^2 = \rho^2(\cos^2 \phi - \sin^2 \phi) = \rho^2 \cos 2\phi = r^2 \sin^2 \theta \cos 2\phi.$$

Confirm that $\nabla^2 f = 0$ in cylindrical and spherical coordinates. ■

Exercise 7.2.3.

grad, curl and div in cylindrical and spherical coordinates

Suppose $\{x^i\}$ are orthogonal coordinates on \mathbb{R}^3 so the metric, inverse metric and unit volume 3-form are of the form

$$\begin{aligned} g &= (h_1)^2 dx^1 \otimes dx^1 + (h_2)^2 dx^2 \otimes dx^2 + (h_3)^2 dx^3 \otimes dx^3, \\ g^{-1} &= (h_1)^{-2} \frac{\partial}{\partial x^1} \otimes \frac{\partial}{\partial x^1} + (h_2)^{-2} \frac{\partial}{\partial x^2} \otimes \frac{\partial}{\partial x^2} + (h_3)^{-2} \frac{\partial}{\partial x^3} \otimes \frac{\partial}{\partial x^3}, \\ \eta &= h_1 h_2 h_3 dx^1 \wedge dx^2 \wedge dx^3, \end{aligned}$$

where $h_1, h_2, h_3 > 0$ are three positive scale factors needed to normalize the coordinate frame vector fields. Let $e_i = \partial/\partial x^i$ and $e_{\hat{i}} = (h_i)^{-1} \partial/\partial x^i$ be the coordinate frame and its associated normalized orthonormal frame, with $\omega^i = dx^i$ and $\omega^{\hat{i}} = h_i \omega^i$ (no sum on i). Let $X^{\hat{i}} = h_i X^i$ (no sum on i) denote the orthonormal components of a vector field.

Verify the following formulas for the gradient, curl, divergence and Laplacian operators

$$\begin{aligned}\vec{\nabla} f &= \frac{1}{h_1} \frac{\partial f}{\partial x^1} e_{\hat{1}} + \frac{1}{h_2} \frac{\partial f}{\partial x^2} e_{\hat{2}} + \frac{1}{h_3} \frac{\partial f}{\partial x^3} e_{\hat{3}}, \\ \text{curl } X &= \frac{1}{h_1 h_2 h_3} \epsilon^{ijk} \frac{\partial}{\partial x^j} (h_k X^{\hat{k}}) e_i, \\ \text{div } X &= \frac{1}{h_1 h_2 h_3} \left[\frac{\partial}{\partial x^1} (X^{\hat{1}} h_2 h_3) + \frac{\partial}{\partial x^2} (X^{\hat{2}} h_3 h_1) + \frac{\partial}{\partial x^3} (X^{\hat{3}} h_1 h_2) \right], \\ \nabla^2 f &= \frac{1}{h_1 h_2 h_3} \left[\frac{\partial}{\partial x^1} \left(\frac{h_2 h_3}{h_1} \frac{\partial f}{\partial x^1} \right) + \frac{\partial}{\partial x^2} \left(\frac{h_3 h_1}{h_2} \frac{\partial f}{\partial x^2} \right) + \frac{\partial}{\partial x^3} \left(\frac{h_1 h_2}{h_3} \frac{\partial f}{\partial x^3} \right) \right].\end{aligned}$$

■

The gradient, curl and divergence are all first order differential operators which appear to be very different from each other, but in Chapter 11, we will see how they are actually very closely related by the simple idea of the “exterior derivative” d which generalizes the differential of a function (0-forms) to an operator on any p -forms or “differential forms,” namely the antisymmetric covariant tensor fields of various ranks.

Exercise 7.2.4.

Laplacian and angular momentum

a) Using the formula

$$(\det \underline{g})^{1/2} = \begin{cases} \rho, & \text{(cylindrical coordinates)} \\ r^2 \sin \theta, & \text{(spherical coordinates)} \end{cases}$$

verify the divergence formula

$$\begin{aligned}\nabla^2 &= \rho^{-1} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2} && \text{(cylindrical coordinates)} \\ &= r^{-2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2}. && \text{(spherical coordinates)}\end{aligned}$$

b) In cylindrical coordinates show that any function of the form $f = k \ln \rho$ satisfies the Laplace equation $\nabla^2 f = 0$. [This is important for inverse square force fields like electromagnetism or gravity due to an infinite line or axially symmetric distribution of the source: charge or mass.]

c) In spherical coordinates show that any function of the form $f = k/r$ satisfies the Laplace equation $\nabla^2 f = 0$. [This is important for inverse square force fields due to a point or spherically symmetric distribution of the source.]

d) In each case show that $\vec{\nabla} f$ is a purely radial vector field as defined by the radial coordinate in the respective coordinate system whose magnitude $|\vec{\nabla} f|$ is respectively inversely proportional to the first and second powers of that radial variable.

These vector fields describe the electric or gravitational fields due to concentration of charge or mass respectively along the vertical axis $\rho = 0$ or at the origin $r = 0$ where the magnitude of the vector field goes infinite. In the latter case we get the inverse square force field associated with a point charge or point mass.

e) Show that in spherical coordinates, radial functions of the form r^n for integer powers n are eigenfunctions of the rescaled Laplacian with eigenvalue $n(n+1)$

$$r^2 \nabla^2 r^n = \begin{cases} n(n+1)r^n, & n \neq -1, \\ 0 & n = -1. \end{cases}$$

f) In Exercise 5.4.6 the squared angular momentum operator was defined by

$$\nabla^2 = \frac{L^2}{r^2} + \frac{D_r(r^2 D_r)}{r^2},$$

where $D_r = r^{-1} x^i \partial_i = \partial_r$. Thus we can identify the purely angular part of the Laplacian

$$-\mathcal{L}^2 = L^2 = \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2},$$

which obviously commutes with $L_3 = \partial_\phi$ or $\mathcal{L}_3 = -i\partial_\phi$, which has eigenfunctions

$$\mathcal{L}_3 e^{im\phi} = m e^{im\phi}.$$

For these functions to be continuous in the periodic coordinate ϕ where $\phi + 2\pi n$ for any integer n must represent the same point, they must be periodic in the azimuthal angle, which requires that m be an integer

$$\cos(2\pi m) + i \sin(2\pi m) = e^{2\pi m i} = e^0 = 1.$$

Ignoring the radial derivative terms in the Laplacian and setting $r = 1$ in the remaining terms yields the corresponding operator on the unit sphere, which is the Laplacian of the intrinsic geometry of that sphere, namely $-\mathcal{L}^2$. If we look for eigenfunctions of \mathcal{L}^2 of the form

$$Y_{\ell m}(\theta, \phi) = N_{\ell m} P_\ell(\cos \theta) e^{im\phi}$$

with eigenvalues $\ell(\ell+1)$ and normalization constants $N_{\ell m}$, show that the functions $P_\ell(\mu)$ where $\mu = \cos \theta$ must satisfy the Legendre function condition

$$\frac{d}{d\mu} \left((1 - \mu^2) \frac{d}{d\mu} P_\ell(\mu) \right) = \left(\ell(\ell+1) - \frac{m^2}{1 - \mu^2} \right) P_\ell(\mu).$$

The functions $Y_{\ell m}(\theta, \phi)$ on the unit sphere are called the spherical harmonics, and one finds for each nonnegative integer ℓ , the $2\ell+1$ values $m = -\ell \dots \ell$. The integer condition on the eigenvalues comes from the requirement that these functions be regular along the axis $\theta = 0, \pi$ or $\mu = \cos \theta = \pm 1$ (let's just accept this fact).

g) Solutions of the Laplace equation $\nabla^2 f = 0$ are called harmonics, already encountered in the multipole moment discussion of Section 2.5, where we saw that functions of the form

$$r^{-(2N-1)} \left(x^{i_1} \dots x^{i_{N-1}} - \frac{r^2}{3} \delta^{(i_1 i_2} x^{i_3} \dots x^{i_{N-1}} \right) = r^{-N} \mathcal{Y}^{i_1 \dots i_{N-1}}$$

are harmonic functions, so that tracefree symmetric tensor combinations of the tensor products of the Cartesian coordinates, rescaled to remove their dependence on r , must be eigenfunctions of the Laplacian on the unit sphere. These are the Cartesian harmonics. If instead we look for product function solutions of this equation in the form of a product of functions of the individual coordinates $r^n Y_{\ell m}(\theta, \phi)$, show that one finds $n(n+1) = \ell(\ell+1)$ using the spherical coordinate decomposition of the Laplacian. Then show that this quadratic relationship has two solutions: $n = \ell, -(\ell+1)$ or equivalently $\ell = n, -(n+1)$. This means essentially that the independent Cartesian harmonics for a given $N = (\ell+1)$ are simply a different basis of the same eigenspace of \mathcal{L}^2 with eigenvalue $\ell(\ell+1)$.

Since the Laplace equation is linear, we can get real solutions by taking the real and imaginary parts of these solutions, and expand a general solution as an infinite series in powers of r and the spherical harmonics. This can be used to expand the potential for a static electric field or gravitational field with boundary conditions on a sphere or generated by an isolated charge distribution with prescribed multipole moments. The nonpositive powers r^n of r with $n = -\ell(\ell+1)$ are regular at infinity. For the Schroedinger equation describing the complex wave function of an electron orbiting a nucleus, one adds a spherically symmetric inverse radius potential function to the Laplace equation, leading to new radial wave functions of r called generalized Laguerre polynomials replacing the power functions of r . ■

Remark.

We have talked about the vector space structure of the infinite-dimensional space of real or complex functions (scalar fields) or tensor fields over \mathbb{R}^3 . The rotation group acts on this space by dragging along as a group of linear transformations of the fields so we get an infinite-dimensional representation of both the rotation group and its Lie algebra, the latter of which are represented by the sign-reversed Lie derivatives by the corresponding generating vector fields. When we discussed decomposing the representation of the rotation group on the finite-dimensional space of $\binom{0}{2}$ -tensors over \mathbb{R}^3 , we were able to find an orthogonal direct sum of so called irreducible representations, each characterized by a single value s of the squared spin operator. The one missing element of our discussion at the infinite-dimensional level is an inner product on the space of fields. For a pair of complex scalar fields Ψ and Φ , allowing for the most general case needed for describing quantum mechanical wave functions, we can integrate the product $\bar{\Phi}\Psi$ over the space

$$\langle \Phi, \Psi \rangle = \int \int \int \bar{\Phi} \Psi \, dV = \overline{\langle \Psi, \Phi \rangle}.$$

Fields which are “square integrable” have a finite value for the self-inner product

$$\langle \Psi, \Psi \rangle = \int \int \int |\Psi|^2 dV < \infty.$$

and are said to belong to the Hilbert (vector) space of square-integrable functions over space. Since rotating a field does not change its integral, the rotation group acts on this space as a symmetry group of its inner product geometry and one can decompose this action into an orthogonal direct sum of representations with respect to this geometry. If we restrict ourselves to the unit sphere, we consider the space of square-integrable differentiable functions on that sphere. The spherical harmonics $Y_{\ell m}$ are an orthonormal basis of this space adapted to the decomposition of the rotation group action into an orthogonal direct sum characterized by the values of the spin parameter ℓ . The normalization factor spoken of above is chosen to make these functions have unit “norm” (length). ▲

Exercise 7.2.5.

angular momentum and Cartesian coordinate functions

The position vector on the unit sphere $\langle x^1, x^2, x^3 \rangle = \langle \sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta \rangle$ is rotated to another point on the sphere by any rotation \underline{R} : $x^i \rightarrow R^i_j x^j$ (the identity representation of the matrix group on \mathbb{R}^3), which means that the three component functions of this vector are rotated among themselves in the spin $\ell = 1$ representation. Indeed in Exercise 1.7.12 we showed that every 3-vector had the eigenvalue $-\ell(\ell + 1) = -2$ with $\ell = 1$ of the squared spin matrix \underline{S}^2 and hence the eigenvalue $\ell(\ell + 1) = 2$ of \underline{S}^2 , where $\underline{S}_a = i\underline{L}_a$. To get a standard basis of the representation in terms of eigenfunctions of the operator $\mathcal{L}_3 = -iL_3 = -i\partial/\partial\phi$, we just need to find the eigenvectors of its corresponding matrix.

a) Confirm that \mathcal{L}^2 expressed in terms of spherical coordinates in Exercise 7.2.4 has eigenvalue 2 acting on these three functions.

b) Show that

$$\mathcal{L}_3 \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = -\underline{\mathcal{S}}_3 \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix}$$

so that we can find the eigenvalues and eigenvectors of the corresponding matrix and then transform the coordinates to the new components in this (complex) basis of eigenvectors. Here is the result for the eigenvalues and eigenfunctions of \mathcal{L}_3 and the eigenvector of the corresponding matrix $-\underline{\mathcal{S}}_3$ to save time

$$\begin{aligned} m = \pm 1 : \quad x^{\pm 1} &= x^1 \pm ix^2, & e_{\pm 1} &= e_1 \mp ie_2, \\ m = 0 : \quad x^0 &= x^3 & e_0 &= e_3. \end{aligned}$$

Express these in terms of the spherical coordinates (and purely imaginary exponentials) and the use a computer algebra system to evaluate the following integrals

$$\langle x^{m'}, x^m \rangle = \int_0^{2\pi} \int_0^\pi \overline{x^{m'}} x^m d\theta d\phi$$

for $(m', m) = (1, 1), (-1, 1), (-1, -1)$ to verify that they are orthogonal and evaluate their “norms” (the length with this inner product, the square root of the self-inner product)

$$||x^m|| = \sqrt{\langle x^m, x^m \rangle}.$$

Divide each by its norm (“normalize” this orthogonal basis of functions). Denote the corresponding normalized basis of eigenfunctions by $\mathcal{Y}_{\ell m}$. Compare them with the $\ell = 1$ spherical harmonics $Y_{\ell m}$ you find on the web. If they agree modulo sign, you did this correctly. One can “change the phase” of the harmonics by multiplying them by any constant unit complex number $e_{\ell m}^\Phi$ since this leaves the inner products invariant. The spherical harmonics are usually defined with an extra sign $(-1)^m$ called the CondonShortley phase to make the ladder operator formulas simpler.

c) Check that the differential equation for the generalized Legendre functions is correct by evaluating it on the two θ factor functions appearing in these $Y_{\ell m}$ functions.

d) To visualize the spherical harmonics, we can take do a 45 degree rotation among the complex pairs $Y_{\ell m}, Y_{\ell -m}$, where $\bar{Y}_{\ell m} = (-1)^m Y_{\ell -m}$, to obtain a real orthonormal basis in the same way we can take the real and imaginary parts of a complex eigenvector of a rotation matrix to get an orthogonal basis of the plane of the rotation. The only complication is the sign convention used to fix the phase of the spherical harmonics, which is not universally agreed upon. This rotation takes the form

$$\langle y_{\ell m}, y_{\ell -m} \rangle = \langle \sqrt{2} \operatorname{Re} Y_{\ell m}, \sqrt{2} \operatorname{Im} Y_{\ell -m} \rangle \left\langle \frac{1}{\sqrt{2}}(Y_{\ell m} + \bar{Y}_{\ell m}), \frac{1}{i\sqrt{2}}(Y_{\ell -m} - Y_{\ell -m}) \right\rangle.$$

The usual spherical harmonics one finds listed in reference discussions of their properties are

$$\langle y_{11}, y_{10}, y_{1-1} \rangle = \sqrt{\frac{3}{4\pi}} \langle -x, z, -y \rangle$$

expressed in spherical coordinates on the unit sphere, showing the relationship between the Cartesian and spherical harmonics for $\ell = 1$. Use a computer algebra system to plot these by plotting the radial coordinate r versus (θ, ϕ) in spherical coordinates using the absolute value of these real harmonics for r . If you have the patience to separate the plots for positive and negative values, you can color them differently and you will then have reproduced the three double lobe graphics you find on the web. ■

Remark.

The Cartesian components of the position vector on the unit sphere coincide with the components of the rotationally symmetric radial unit vector field $\partial/\partial r$, whose Lie derivatives with respect to the rotation generators therefore vanish

$$0 = \mathcal{L}_{L_a} \frac{\partial}{\partial r} = -\mathcal{J}_a \frac{\partial}{\partial r} = -(\mathcal{L}_3 + \mathcal{S}_3) \left(\frac{x^i}{r} \frac{\partial}{\partial x^i} \right).$$

This explains the relation $(\mathcal{L}_3 + \mathcal{S}_3)\langle x^1, x^2, x^3 \rangle = 0$ found above.

As already mentioned above the relation between the components of the position vector on the sphere, namely the Cartesian coordinate functions, and the $\ell = 1$ spherical harmonics generalizes to tensor products of the position vector, leading to the Cartesian harmonics, which are a basis for polynomials in the Cartesian coordinates. The five $\ell = 2$ spherical harmonics correspond to the spin 2 tracefree symmetric coefficient matrices used to form quadratic functions from the tensor product $x^i x^j$, investigated in Exercise 1.6.9. The trace inner product orthogonality discussed there in fact corresponds to the orthogonality on the spaces of spherical harmonics. In short the decomposition of the representations of the rotation group on the space of harmonic functions over \mathbb{R}^3 which are regular at infinity, when restricted to the sphere, mirrors exactly the subspace of symmetric tracefree tensors over \mathbb{R}^3 , and the natural inner products on these two very different spaces correspond one to the other in this relationship.

▲

7.3 Spherical coordinate orthonormal frame

The orthonormal frame associated with the orthogonal spherical coordinate frame is related to the orthonormal Cartesian coordinate frame by a rotation

$$(e_{\hat{r}} \ e_{\hat{\theta}} \ e_{\hat{\phi}}) = (e_x \ e_y \ e_z) \underbrace{\begin{pmatrix} \sin \theta \cos \phi & \cos \theta \cos \phi & -\sin \phi \\ \sin \theta \sin \phi & \cos \theta \sin \phi & \cos \phi \\ \cos \theta & -\sin \theta & 0 \end{pmatrix}}_{\underline{\mathcal{B}} = \underline{\mathcal{A}}^{-1}}$$

The columns of the orthogonal matrix $\underline{\mathcal{B}}$ are the Cartesian coordinate frame components of the new orthonormal frame vectors and are obtained by normalizing the columns of the matrix \underline{B} which represents the Cartesian coordinate components of the spherical coordinate frame vectors.

Since $\underline{\mathcal{B}}$ is an orthogonal matrix then

$$\underline{\mathcal{A}} = \underline{\mathcal{B}}^{-1} = \underline{\mathcal{B}}^T$$

This gives the matrix needed to transform components from the Cartesian frame to the spherical one

$$(e_x \ e_y \ e_z) = (e_{\hat{r}} \ e_{\hat{\theta}} \ e_{\hat{\phi}}) \underbrace{\begin{pmatrix} \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\ \cos \theta \cos \phi & \cos \theta \sin \phi & -\sin \theta \\ -\sin \phi & \cos \phi & 0 \end{pmatrix}}_{\underline{\mathcal{B}}^{-1} = \underline{\mathcal{A}}}$$

The columns of the orthogonal matrix $\underline{\mathcal{B}}^{-1}$ are the new orthonormal frame components of the Cartesian orthonormal frame vectors and are obtained by normalizing the columns of the matrix \underline{B}^{-1} which represents the spherical coordinate components of the Cartesian coordinate frame vectors.

In Exercise 6.2.1 the connection 1-form matrix $\hat{\omega}$ for the spherical coordinate orthonormal frame was easily interpreted in terms of the rate of change of the frame rotation as one moves in the two angular directions. The matrix component along $d\phi$ simply rotates the orthonormal frame about the vertical direction at the rate $d\phi/dt$ under a translation $\phi \rightarrow \phi + t$ along an azimuthal coordinate circle, while the matrix component along $d\theta$ generates a rotation about $e_{\hat{\phi}}$ at the rate $d\theta/dt$ when applied to a translation $\theta \rightarrow \theta + t$ along a polar coordinate circle. A finite rotation by these 1-parameter subgroups corresponds to a finite translation along the coordinate lines. Thus incrementing ϕ by $\Delta\phi$ rotates the frame by that increment about the vertical axis, while incrementing θ by $\Delta\theta$ rotates the frame by that increment about the fixed azimuthal axis to the r - θ plane. Thus we can start with the Cartesian frame on the positive vertical axis at some radius $r > 0$, and angles $(\theta, \phi) = (0, 0)$ and move in the angular coordinate plane either first to $(0, \phi)$ and then to (θ, ϕ) or in the opposite order: first to $(\theta, 0)$ and then to (θ, ϕ) . The first order is simpler although we stay fixed on the vertical axis under a translation by the azimuthal angle.

Thus we can understand the matrix $\underline{\mathcal{B}} = \langle e_{\hat{r}}, e_{\hat{\theta}}, e_{\hat{\phi}} \rangle$ of as resulting from the following sequence of simpler transformations, using an obvious shorthand for the trig functions: $C_\phi =$

$\cos \phi$, etc. As already explored in Exercise 5.8.2, we start with a point with $z = r$ on the positive z -axis and rotate both the point and the Cartesian frame at that point to a general position with coordinates (r, θ, ϕ) in two moves.

$$\begin{aligned} (e_{\hat{r}} \ e_{\hat{\theta}} \ e_{\hat{\phi}})|_{(r,\theta,\phi)} &= (e_x \ e_y \ e_z)|_{(r,\theta=0,\phi=0)} \begin{pmatrix} C_\phi & -S_\phi & 0 \\ S_\phi & C_\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} C_\theta & 0 & S_\theta \\ 0 & 1 & 0 \\ -S_\theta & 0 & C_\theta \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \\ &\equiv (e_x \ e_y \ e_z)|_{(r,\theta=0,\phi=0)} \underline{R}_3(\phi) \underline{R}_1(\theta) \underline{P}. \end{aligned}$$

The first matrix on the left rotates the initial vectors $(e_x, e_y, e_z = e_{\hat{r}})$ at a point on the positive z -axis by an angle ϕ about the e_z -axis to $(e_{\hat{\theta}}, e_{\hat{\phi}}, e_{\hat{r}})$ still on that axis. The next matrix factor then rotates the point and frame by the angle θ away from the z -axis in the half-plane for the coordinate value ϕ to its final location. However, to have a right handed frame, a further permutation of the frame vectors is required to get the order $(e_{\hat{r}}, e_{\hat{\theta}}, e_{\hat{\phi}})$ where $e_{\hat{r}} \times e_{\hat{\theta}} = e_{\hat{\phi}}$.

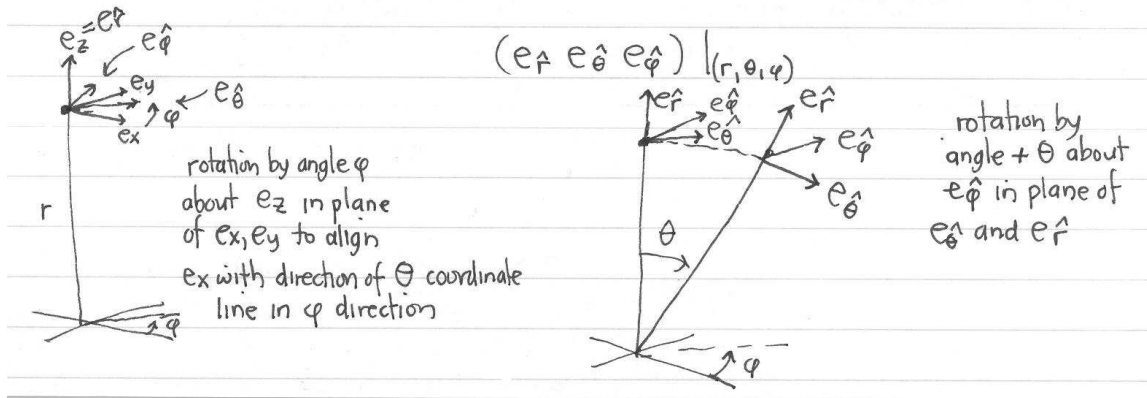


Figure 7.2: Getting from the Cartesian coordinate frame at a point on the positive z -axis to the spherical coordinate frame at a general point.

Now it is a straightforward problem to evaluate the connection 1-form matrix in the spherical orthonormal frame using this product representation (even by hand, but a computer algebra

system is better). Using the abbreviations $C_u = \cos u$, $S_u = \sin u$ one finds

$$\begin{aligned}
 \underline{\hat{\omega}} &= \underline{\mathcal{B}}^{-1} d\underline{\mathcal{B}} = (\Gamma^{\hat{i}}_{\hat{k}\hat{j}} \omega^{\hat{k}}) \\
 &= \begin{pmatrix} S_\theta C_\phi & S_\theta C_\phi & C_\theta \\ C_\theta C_\phi & C_\theta S_\phi & -S_\theta \\ -S_\phi & C_\phi & 0 \end{pmatrix} \left\{ \begin{pmatrix} C_\theta S_\phi & S_\theta C_\phi & 0 \\ C_\theta C_\phi & C_\theta S_\phi & 0 \\ -S_\phi & C_\phi & 0 \end{pmatrix} d\theta + \begin{pmatrix} C_\theta S_\phi & -C_\theta S_\phi & -C_\phi \\ S_\theta C_\phi & C_\theta C_\phi & -S_\phi \\ 0 & 0 & 0 \end{pmatrix} d\phi \right\} \\
 &= \dots = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} d\theta + \begin{pmatrix} 0 & 0 & -S_\theta \\ 0 & 0 & -C_\theta \\ S_\theta & C_\theta & 0 \end{pmatrix} d\phi \\
 &= \begin{pmatrix} 0 & -r^{-1} & 0 \\ r^{-1} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \omega^{\hat{\theta}} + \begin{pmatrix} 0 & 0 & -r^{-1} \\ 0 & 0 & -r^{-1} \cot \theta \\ r^{-1} & r^{-1} \cot \theta & 0 \end{pmatrix} \omega^{\hat{\phi}} \\
 &= \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} d\theta + \begin{pmatrix} 0 & 0 & -\sin \theta \\ 0 & 0 & -\cos \theta \\ \sin \theta & \cos \theta & 0 \end{pmatrix} d\phi,
 \end{aligned}$$

so that

$$\begin{aligned}
 \Gamma^{\hat{1}}_{\hat{\theta}\hat{2}} &= \Gamma^{\hat{r}}_{\hat{\theta}\hat{\theta}} = -r^{-1}, & \Gamma^{\hat{1}}_{\hat{\phi}\hat{3}} &= \Gamma^{\hat{r}}_{\hat{\phi}\hat{\phi}} = -r^{-1}, \\
 \Gamma^{\hat{2}}_{\hat{\theta}\hat{1}} &= \Gamma^{\hat{\theta}}_{\hat{\theta}\hat{r}} = r^{-1}, & \Gamma^{\hat{3}}_{\hat{\phi}\hat{1}} &= \Gamma^{\hat{\phi}}_{\hat{\phi}\hat{r}} = r^{-1},
 \end{aligned}$$

and finally

$$\begin{aligned}
 \Gamma^{\hat{2}}_{\hat{\phi}\hat{3}} &= \Gamma^{\hat{\theta}}_{\hat{\phi}\hat{\phi}} = -r^{-1} \cot \theta, \\
 \Gamma^{\hat{3}}_{\hat{\phi}\hat{2}} &= \Gamma^{\hat{\phi}}_{\hat{\phi}\hat{\theta}} = r^{-1} \cot \theta.
 \end{aligned}$$

This last expression gives the six nonzero orthonormal components of the connection 1-forms found directly in Exercise 6.2.1 using a computer algebra system.

Exercise 7.3.1.

matrix product representation of orthonormal frame

- Check that the matrix product of these three factor matrices is $\underline{\mathcal{B}}$.
- Fill in the dots in the subsequent step by step evaluation of $\underline{\mathcal{B}}^{-1} d\underline{\mathcal{B}}$ using this product representation, evaluating and simplifying the matrix products and re-expressing the coordinate differentials in terms of the orthonormal 1-forms.
- Check the resulting components of the connection by reading them off from the penultimate matrix.

■

Exercise 7.3.2.

spherical coordinate orthonormal frame connection vector

- As first explored in Exercise 1.2.4, the sign-reversed dual vector of a 3×3 antisymmetric matrix in an orthonormal frame

$$\Omega^i_j x^j = -\epsilon_{ijk} [* \Omega]^k x^j = [\vec{\Omega} \times \vec{x}]^i = [* \Omega]^k [\underline{k}_k]^i_j x^j$$

gives the axis of the rotation it generates, with the angle given by its magnitude in the direction around the axis given by the right hand rule provided the orthonormal frame in which it is expressed is right-handed. Recall that the three matrices \underline{k}_k with components $[\underline{k}_k]^i_j = \epsilon_{ikj}$ are the natural basis of the Lie algebra of rotations. The previous calculation can be expressed in terms of these matrices as follows

$$\begin{aligned}\underline{\hat{\omega}} &= \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} d\theta + \left[\cos \theta \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ k_0 & 1 & 0 \end{pmatrix} - \sin \theta \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \right] d\phi \\ &= [\cos \theta \underline{k}_1 - \sin \theta \underline{k}_2] d\phi + \underline{k}_3 d\theta \equiv \Omega^{\hat{k}} \underline{k}_k,\end{aligned}$$

which defines a vector-valued connection 1-form by taking the sign-reversed dual of the connection components on their antisymmetric pair of indices

$$\Omega = \Omega^{\hat{k}} e_{\hat{k}} = -\frac{1}{2} \eta^{\hat{k}\hat{i}\hat{j}} \hat{\omega}_{\hat{i}\hat{j}} e_{\hat{k}} = -\frac{1}{2} e_{\hat{i}} \otimes \eta^{\hat{i}\hat{k}\hat{j}} \omega_{\hat{k}}^{\hat{j}}.$$

The coefficients of the basis rotation matrices in the square brackets are the components of the vector

$$\langle \cos \theta, -\sin \theta, 0 \rangle.$$

Compare this with the final column of the matrix \mathcal{B}^{-1} at the beginning of this section to show that these are the orthonormal spherical components of the unit vector field $e_{\hat{z}}$. Conclude from this that

$$\Omega = (\cos \theta e_{\hat{r}} - \sin \theta e_{\hat{\theta}}) \otimes d\phi + e_{\hat{\phi}} \otimes d\theta = e_{\hat{z}} \otimes d\phi + e_{\hat{\phi}} \otimes d\theta,$$

This shows concisely how the orthonormal vectors are affected by changes in the angular directions. Increasing the polar angle θ while holding ϕ fixed rotates $e_{\hat{r}}$ and $e_{\hat{\theta}}$ in their plane about $e_{\hat{\phi}}$ by the increment in that angle, while increasing the azimuthal angle ϕ holding θ fixed rotates all of the frame vectors about the vertical axis by the increment in that angle.

This same approach to interpreting the connection 1-forms will work for any orthonormal frame (of Euclidean signature).

■

We can also derive expressions for the components of the connection in the orthonormal spherical frame from the metric formula for the components of the connection in a frame. For an orthonormal frame $\{e_i\}$, then $g_{i\hat{j}} = g(e_i, e_{\hat{j}}) = \delta_{ij}$, i.e., the components of the metric are constants so the metric component derivative terms in the formula vanish, leaving only the structure function terms

$$\Gamma^{\hat{i}}_{\hat{j}\hat{k}} = \frac{1}{2} (C^{\hat{i}}_{\hat{j}\hat{k}} + C^{\hat{i}}_{\hat{j}\hat{k}} + C^{\hat{i}}_{\hat{k}\hat{j}}) = \frac{1}{2} (C^{\hat{i}}_{\hat{j}\hat{k}} - C^{\hat{j}}_{\hat{k}\hat{i}} + C^{\hat{k}}_{\hat{i}\hat{j}}),$$

where the final formula holds since index shifting is trivial in an orthonormal frame. The structure functions were evaluated above in an exercise. The results should have been

$$C^{\hat{\theta}}_{\hat{r}\hat{\theta}} = -\frac{1}{r} = -C^{\hat{\theta}}_{\hat{\theta}\hat{r}}, \quad C^{\hat{\phi}}_{\hat{r}\hat{\phi}} = -\frac{1}{r} = -C^{\hat{\phi}}_{\hat{\phi}\hat{r}}, \quad C^{\hat{\phi}}_{\hat{\theta}\hat{\phi}} = -\frac{1}{r} \cot \theta = -C^{\hat{\phi}}_{\hat{\phi}\hat{\theta}}.$$

The nonzero components of the connection must have indices which are at most a permutation of the index positions on the nonzero structure functions. Forgetting for a moment that we know which six components of the connection are nonzero, we can use the following reasoning to avoid evaluating the formula for many components which turn out to be zero.

We saw above that the covariant constancy of the metric implies the relation $g_{ij,k} = \Gamma_{jki} + \Gamma_{ikj}$. For an orthonormal frame, $g_{ij} = \delta_{ij}$ and $g_{ij,k} = 0$ (constant components) so

$$\Gamma_{jki} = -\Gamma_{ikj} ,$$

i.e., the components of the connection are antisymmetric in their outer indices. This remains true when we raise the index since the metric component matrix is the identity matrix and explains why the matrix $\hat{\omega} = (\hat{\Gamma}_{\hat{k}\hat{j}}^{\hat{i}} \omega^{\hat{k}})$ evaluated above for the spherical orthonormal frame is antisymmetric—its matrix indices are the outer pair of indices on the components of the covariant derivative.

Thus the outer pair of indices (\hat{i}, \hat{j}) on $\Gamma_{\hat{k}\hat{j}}^{\hat{i}}$ must be distinct for this quantity to be nonzero, and antisymmetry tells us its value for one index ordering of the pair in terms of the other ordering. Given the three independent nonzero structure functions (six nonzero components equal in pairs by antisymmetry), there are only three independent nonzero components of the connection that we need to write down. Using the antisymmetry of the structure functions in the lower index pair, we get

$$\begin{aligned} C_{\hat{r}\hat{\theta}}^{\hat{\theta}} &= -C_{\hat{\theta}\hat{r}}^{\hat{\theta}} \longrightarrow \Gamma_{\hat{\theta}\hat{r}}^{\hat{\theta}} = -\Gamma_{\hat{\theta}\hat{\theta}}^{\hat{r}} \\ &= \frac{1}{2}(C_{\hat{\theta}\hat{r}}^{\hat{\theta}} - C_{\hat{\theta}\hat{r}}^{\hat{\theta}} + C_{\hat{r}\hat{\theta}}^{\hat{\theta}}) = \frac{1}{2}(C_{\hat{\theta}\hat{r}}^{\hat{\theta}} + C_{\hat{r}\hat{\theta}}^{\hat{\theta}} + C_{\hat{\theta}\hat{\theta}}^{\hat{r}}) = C_{\hat{\theta}\hat{r}}^{\hat{\theta}} \\ &= r^{-1} \\ C_{\hat{r}\hat{\phi}}^{\hat{\phi}} &= -C_{\hat{\phi}\hat{r}}^{\hat{\phi}} \longrightarrow \Gamma_{\hat{\phi}\hat{r}}^{\hat{\phi}} = -\Gamma_{\hat{\phi}\hat{\phi}}^{\hat{r}} \\ &= \frac{1}{2}(C_{\hat{\phi}\hat{r}}^{\hat{\phi}} - C_{\hat{\phi}\hat{r}}^{\hat{\phi}} + C_{\hat{r}\hat{\phi}}^{\hat{\phi}}) = \frac{1}{2}(C_{\hat{\phi}\hat{r}}^{\hat{\phi}} + C_{\hat{\phi}\hat{r}}^{\hat{\phi}} + C_{\hat{\phi}\hat{\phi}}^{\hat{r}}) = C_{\hat{\phi}\hat{r}}^{\hat{\phi}} \\ &= r^{-1} \\ C_{\hat{\theta}\hat{\phi}}^{\hat{\phi}} &= -C_{\hat{\phi}\hat{\theta}}^{\hat{\phi}} \longrightarrow \Gamma_{\hat{\phi}\hat{\theta}}^{\hat{\phi}} = -\Gamma_{\hat{\phi}\hat{\phi}}^{\hat{\theta}} \\ &= \frac{1}{2}(C_{\hat{\phi}\hat{\theta}}^{\hat{\phi}} - C_{\hat{\phi}\hat{\theta}}^{\hat{\phi}} + C_{\hat{\theta}\hat{\phi}}^{\hat{\phi}}) = \frac{1}{2}(C_{\hat{\phi}\hat{\theta}}^{\hat{\phi}} + C_{\hat{\phi}\hat{\theta}}^{\hat{\phi}} + C_{\hat{\phi}\hat{\phi}}^{\hat{\theta}}) = C_{\hat{\phi}\hat{\theta}}^{\hat{\phi}} \\ &= r^{-1} \cot \theta , \end{aligned}$$

and we do not have to waste time verifying that the remaining components are zero.

What is the significance of the antisymmetry property of the matrix $\underline{\omega} = (\Gamma_{kj}^i \omega^k)$ in an orthonormal frame? Well, from the definition

$$\nabla_{e_k} e_i = \Gamma_{ki}^j e_j ,$$

we can calculate the derivative of the frame vector fields along a vector field X using the fact

that it is linear in X (i.e., $\nabla_X Y^i = Y^i{}_{;j} X^j = X^j \nabla_{e_j} Y^i$), which in this context says

$$\nabla_X e_i = \nabla_{X^k e_k} e_i = X^k \nabla_{e_k} e_i = \Gamma^j{}_{ki} \underbrace{X^k}_{\omega^k(X)} \underbrace{e_j}_{\omega^j_i(X)}$$

so

$$\nabla_X e_i = \omega^j_i(X) e_j.$$

Summarizing this calculation, the value of the matrix-valued 1-form $\underline{\omega}$ on X gives the matrix of the linear transformation of the frame vectors which describes their covariant derivative in that direction, i.e., how they change relative to a Cartesian frame as we move in that direction. The fact that this matrix is antisymmetric tells us that in 3-dimensions it can be represented by the cross-product of a vector and that it generates a rotation.

What is the interpretation of the connection 1-form matrix $\underline{\omega} = \underline{A} d\underline{A}^{-1}$ for the spherical coordinate frame which is not orthonormal? Doing the following calculation

$$\begin{aligned} \underline{\omega} &= \begin{pmatrix} S_\theta C_\phi & S_\theta S_\phi & C_\theta \\ r^{-1} C_\theta C_\phi & r^{-1} C_\theta S_\phi & -r^{-1} S_\theta \\ -r^{-1} \frac{S_\phi}{S_\theta} & -r^{-1} \frac{C_\phi}{S_\theta} & 0 \end{pmatrix} d \begin{pmatrix} S_\theta C_\phi & r C_\theta C_\phi & -r S_\theta S_\phi \\ S_\theta S_\phi & r C_\theta S_\phi & r S_\theta C_\phi \\ C_\theta & -r S_\theta & 0 \end{pmatrix} \\ &= \dots = \begin{pmatrix} 0 & 0 & 0 \\ 0 & r^{-1} & 0 \\ 0 & 0 & r^{-1} \end{pmatrix} dr + \begin{pmatrix} 0 & -r & 0 \\ r^{-1} & 0 & 0 \\ 0 & 0 & \cot \theta \end{pmatrix} d\theta + \begin{pmatrix} 0 & 0 & -r \sin^2 \theta \\ 0 & 0 & -\sin \theta \cos \theta \\ r^{-1} & \cot \theta & 0 \end{pmatrix} d\phi, \end{aligned}$$

we can read off the following connection coordinate component formulas

$$\begin{aligned} \Gamma^\theta_{r\theta} &= \Gamma^\phi_{r\phi} = r^{-1}, \\ \Gamma^r_{\theta\theta} &= -r, \quad \Gamma^\theta_{\theta r} = r^{-1}, \quad \Gamma^\phi_{\theta\phi} = \cot \theta, \\ \Gamma^r_{\phi\phi} &= -r \sin^2 \theta, \quad \Gamma^\theta_{\phi\phi} = -\sin \theta \cos \theta, \\ \Gamma^\phi_{\phi r} &= r^{-1}, \quad \Gamma^\phi_{\phi\theta} = \cot \theta. \end{aligned}$$

Exercise 7.3.3.

spherical coordinate connection 1-forms

Check these. A computer algebra system makes this less painful. ■

The appearance of r in the θ and ϕ components of the 1-form $\underline{\omega}$ just takes into account the fact that for fixed r , e_θ and e_ϕ are not unit vectors, so the existing nonzero components of the covariant derivative in the associated orthonormal frame are simply rescaled by factors of r , except for the additional component $\Gamma^\phi_{\theta\phi} = \cot \theta = \omega^\phi(\nabla_{e_\theta} e_\phi)$. This describes the change in the length of e_ϕ as we change θ . Similarly the extra components $\Gamma^\theta_{r\theta}$ and $\Gamma^\phi_{r\phi}$ describe the change in the length of e_θ and e_ϕ as we change r .

7.4 Rotations and derivatives

RECONSIDER THIS SECTION IN VIEW OF PREVIOUS ROTATION MATRIX DISCUSSIONS

We review the properties of orthonormal bases and rotations that have been discussed at length above. This allows us to interpret the connection 1-form matrix in terms of a differential angular velocity.

For the Euclidean inner product on \mathbb{R}^n , the components of the inner product are just

$$g(e_i, e_j) = e_i \cdot e_j = \delta_{ij}$$

in the standard basis or in any orthonormal basis. If

$$\bar{e}_i = A^{-1j}{}_i e_j = B^j{}_i e_j \quad , \quad e_i = A^j{}_i \bar{e}_j = B^{-1j}{}_i \bar{e}_j$$

is a transformation relating any two orthonormal bases, where \underline{B} actively transforms the starting basis, while its inverse \underline{A} transforms indices in the corresponding passive coordinate transformation, then the inner products of the basis vectors transform in the following way

$$\delta_{ij} = A^{-1m}{}_i A^{-1n}{}_j \delta_{mn} \quad \text{or} \quad \delta_{ij} = A^m{}_i A^n{}_j \delta_{mn}$$

or in matrix form

$$\begin{aligned} \delta_{ij} &= A^m{}_i \delta_{mn} A^n{}_j \\ \underline{I} &= \underline{A}^T \underline{I} \underline{A} = \underline{A}^T \underline{A} \end{aligned}$$

This just says that the transpose of \underline{A} is its inverse

$$\underline{A}^T = \underline{A}^{-1} = \underline{B}.$$

This condition characterizes the matrices of linear transformations between orthonormal bases. Such matrices are called orthogonal, and consist of a set of either rows or columns which form an orthonormal set of vectors with respect to the usual dot product on \mathbb{R}^n . They represent rotations and reflections of \mathbb{R}^n into itself.

Note that from the product rule for determinants, taking the determinant of the equation $\underline{A}^T \underline{A} = \underline{I}$

$$1 = \det(\underline{A}^T \underline{A}) = \det(\underline{A}^T) \det \underline{A} = (\det \underline{A})^2,$$

so $\det \underline{A} = \pm 1$. Those with $\det \underline{A} = 1$ represent rotations, while those with $\det \underline{A} = -1$ consist of a rotation plus a reflection.

Suppose \underline{A} depends on a parameter λ so we get a family of orthogonal matrices. Then

$$\begin{aligned} &\frac{d}{d\lambda} [\underline{A}^T \underline{A} = \underline{I}] \\ &\underbrace{\frac{d\underline{A}^T}{d\lambda} \underline{A}}_{\left[\underline{A}^T \frac{d\underline{A}}{d\lambda} \right]^T} + \underline{A}^T \frac{d\underline{A}}{d\lambda} = 0 \end{aligned}$$

using the obvious properties of the transpose $(\underline{A} \ \underline{B})^T = \underline{B}^T \ \underline{A}^T$, $(\underline{A}^T)^T = \underline{A}$. Next using the orthogonal condition $\underline{A}^T = \underline{A}^{-1}$, this becomes

$$\underline{A}^{-1} \frac{d\underline{A}}{d\lambda} + \left[\underline{A}^{-1} \frac{d\underline{A}}{d\lambda} \right]^T = 0.$$

This just says that the matrix

$$\underline{P} \equiv \underline{A}^{-1} \frac{d\underline{A}}{d\lambda} = -\underline{P}^T$$

is antisymmetric: $P^i_j = -P^j_i$ but to respect index positioning we should really rewrite this as: $\delta_{ik} P^k_j = -\delta_{jk} P^k_i$ or finally with index lowering notation $P_{ij} = -P_{ji}$.

The same thing is true if we take the differential

$$\underline{A}^{-1} d\underline{A} = \underline{A}^{-1} \frac{d\underline{A}}{d\lambda} d\lambda$$

instead of the derivative. This explains why the derivative of the spherical coordinate frame orthogonal transformation matrix

$$\underline{\hat{\omega}} = \underline{\mathcal{A}} \ d\underline{\mathcal{A}}^{-1} = (\underline{\mathcal{A}}^{-1})^{-1} \ d(\underline{\mathcal{A}}^{-1})$$

is antisymmetric, namely with correct index positioning

$$\omega_{\hat{i}\hat{j}} = -\omega_{\hat{j}\hat{i}}.$$

The connection 1-form matrix $\underline{\omega}(X)$ evaluated on a given vector field X tells us the rate of change of the rotation which the orthonormal frame undergoes as we move in the direction of X .

However, the rate of change of a rotation can be described by an angular velocity which is more helpful in visualizing its geometry. To understand this suppose a point of \mathbb{R}^3 with initial Cartesian coordinates $x^i(0)$ undergoes an active rotation

$$\underbrace{x^i(t)}_{\text{position at } t} = A^i_j(t) \underbrace{x^j(0)}_{\text{position at } t=0}.$$

Then using the consequence $(d\underline{A}/dt)\underline{A}^{-1} = -\underline{A}(d\underline{A}^{-1}/dt)$ of differentiating $\underline{A} \underline{A}^{-1} = \underline{I}$ one finds

$$\frac{dx^i}{dt}(t) = \frac{dA^i_j(t)}{dt} \underbrace{x^j(0)}_{A^{-1j}_k(t)x^k(t)} = \frac{dA^i_j(t)}{dt} A^{-1j}_k(t) x^k(t) = - \underbrace{A^i_j(t) \frac{dA^{-1j}_k(t)}{dt}}_{\equiv P^i_k(t)} x^k(t),$$

but since \underline{P} is antisymmetric, it can be represented by its dual

$$P^i_k(t) = \epsilon_{ikm} \Omega^m(t)$$

ignoring the index positioning since we are working in an orthonormal basis so

$$\frac{dx^i}{dt}(t) = \underbrace{-\epsilon_{ikm}}_{\epsilon_{imk}} \Omega^m(t) x^k(t) = \underbrace{\epsilon_{imk} \Omega^m(t) x^k(t)}_{[\vec{\Omega}(t) \times \vec{x}(t)]^i}$$

or in vector notation

$$\frac{d\vec{x}}{dt}(t) = \vec{\Omega}(t) \times \vec{x}(t).$$

The vector $\vec{\Omega}(t) = \|\vec{\Omega}(t)\| \hat{n}(t)$ is the angular velocity vector, which corresponds to an instantaneous rotation about the instantaneous axis $n(t)$ (its direction) with an angular rate of change $\|\vec{\Omega}(t)\|$ (its magnitude), where the sense of the rotation about the axis is determined by the right hand rule.

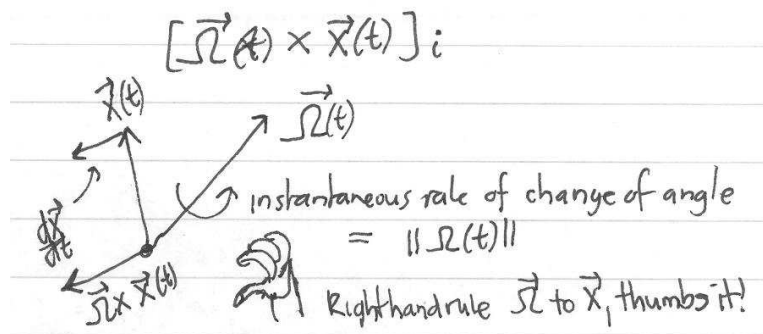


Figure 7.3: The right hand rule from $\vec{\Omega}$ to $\vec{x}(t)$ gives the direction in which $\vec{x}(t)$ is instantaneously rotating. [add vector n to new diagram]

However, this little detour into the rotation group is all we can spare time for at this point. It helps explain why the rate of change of an orthonormal frame must be described by an antisymmetric matrix, which is what the connection 1-form matrix $\underline{\omega}$ represents. Letting the vector $\vec{\Omega}(X)$ be the dual of the antisymmetric matrix $\underline{\omega}(X)$, the rate of change of the frame vectors along X is described by this angular velocity acting on them by cross product multiplication.

Okay, so you didn't do rotation and angular velocity in your physics courses, or maybe you never understood the right hand rule, or maybe you're just not patient enough to read this stuff about derivatives of orthogonal matrices and duals of antisymmetric matrices—OKAY, it doesn't matter. The cross-product and right hand rule only work in 3 dimensions where a pair of antisymmetric indices can be swapped for a single index by the duality operation. In any other dimension, you are stuck with a 2-plane in which a rotation takes place, so it is enough to look at rotations of \mathbb{R}^2 to understand how they work.

By trigonometry

$$\begin{aligned} \bar{e}_1 &= \cos \theta e_1 + \sin \theta e_2, \\ \bar{e}_2 &= -\sin \theta e_1 + \cos \theta e_2, \end{aligned}$$

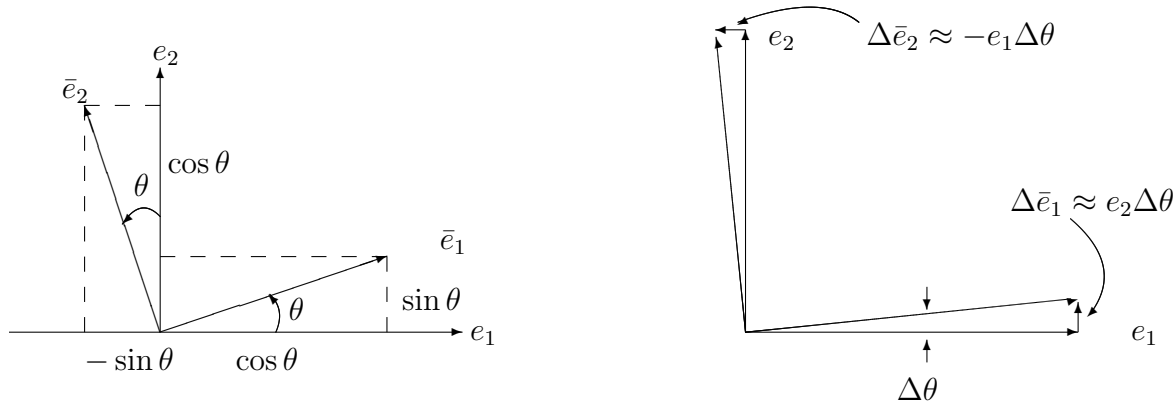


Figure 7.4: A finite and infinitesimal rotation in the plane.

or

$$\begin{aligned}
 (\bar{e}_1 \quad \bar{e}_2) &= (e_1 \quad e_2) \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \\
 \frac{d}{d\theta}(\bar{e}_1 \quad \bar{e}_2) &= (e_1 \quad e_2) \begin{pmatrix} -\sin \theta & -\cos \theta \\ \cos \theta & -\sin \theta \end{pmatrix}, \\
 \left. \frac{d}{d\theta} \right|_{\theta=0} (\bar{e}_1 \quad \bar{e}_2) &= (e_1 \quad e_2) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\
 \longrightarrow (\Delta \bar{e}_1, \Delta \bar{e}_2) &\approx (e_1 \quad e_2) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \Delta \theta.
 \end{aligned}$$

The interpretation of this is that as you begin to rotate the basis vectors through a small angle $\Delta \theta$, the vector \bar{e}_1 begins to rotate toward e_2 and \bar{e}_2 towards $-e_1$ as shown in Fig. 7.4, explaining the antisymmetry of the matrix $\underline{B} = d\underline{A}/d\theta|_{\theta=0}$.

Now look at

$$\underline{\hat{\omega}} = \underline{\mathcal{A}} d\underline{\mathcal{A}}^{-1} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} d\theta + \begin{pmatrix} 0 & 0 & -\sin \theta \\ 0 & 0 & -\cos \theta \\ \sin \theta & \cos \theta & 0 \end{pmatrix} d\phi$$

which tells us how the spherical orthonormal frame vectors begin to change as we make small increments $\Delta \theta$ and $\Delta \phi$ in the angular variables, or alternately, tells us the rate at which these frame vectors are rotating as we change the angular coordinates. The fact that these 1-forms have no component along dr means that they don't rotate as we change r , i.e., as we move radially, and that is exactly right.

If we hold ϕ fixed and increase θ , $e_{\hat{\phi}}$ remains fixed while $(e_{\hat{r}}, e_{\hat{\theta}})$ rotate by *exactly* the increment of θ in their 2-plane in the usual counterclockwise sense, so the 2×2 part of the matrix with $\hat{r}, \hat{\theta}$ indices is exactly the matrix of our two dimensional discussion.

If we hold θ fixed and increase ϕ , what happens depends on the value of θ . For $\theta = \pi/2$ we are in the x - y plane and $e_{\hat{\theta}}$ remains equal to $-e_z$ as we change ϕ but $(e_{\hat{r}}, e_{\hat{\phi}})$ undergoes the same 2-dimensional rotation by exactly the increment in ϕ .

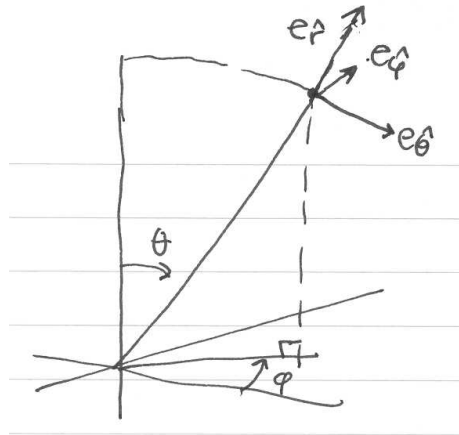
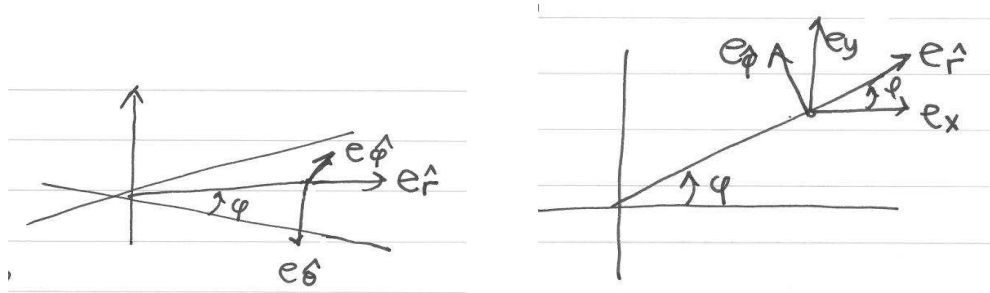


Figure 7.5: The spherical coordinate unit vectors.

Figure 7.6: The spherical coordinate unit vectors in the x - y plane. $e_{\hat{r}}$ and $e_{\hat{\phi}}$ are obtained from the Cartesian unit vectors by a rotation by angle ϕ .

This is just what

$$\hat{\omega}|_{\theta=\pi/2, d\theta=0} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} d\phi$$

describes.

At the other extreme $\theta \rightarrow 0$ we approach the z -axis, where $e_{\hat{r}} \approx e_{\hat{z}}$ remains fixed and $(e_{\hat{\theta}}, e_{\hat{\phi}})$ rotate by exactly the increment in ϕ which is what describes

$$\hat{\omega}|_{\theta \approx 0, d\theta=0} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} d\phi$$

Are you convinced?

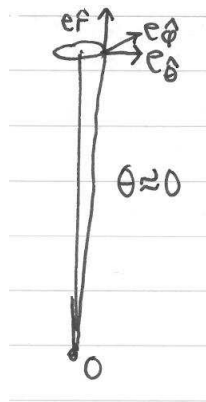


Figure 7.7: The spherical coordinate unit vectors near the z -axis rotate by exactly the increment in ϕ .

Chapter 8

Parallel transport and geodesics

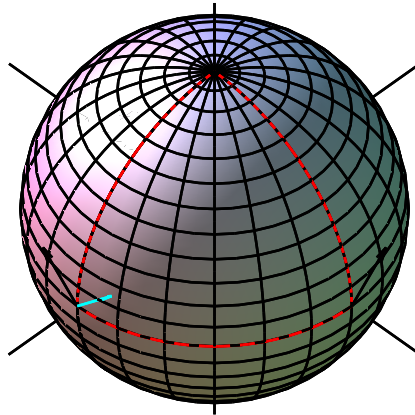


Figure 8.1: A vector (cyan) pointing to the Northeast at the Equator of the sphere at the positive x -axis is moved East along the Equator maintaining its 45° angle with the Equator till it gets to the positive y -axis, then moved up the longitude to the North Pole where it points along the $\phi = 180^\circ + 45^\circ$ longitude line. Bring the vector down the $\phi = 0^\circ$ longitude line back to the positive x -axis, where it now points Northwest (black). It has rotated by 90° due to the curvature of the sphere. If we had moved along such a great circle triangle with an increment $\Delta\phi$ between the longitude lines instead of 90° , the vector would rotate by that angle relative to its original direction.

The covariant derivative on \mathbb{R}^n is a local way of encoding the global flatness that allows one to identify all of its tangent spaces with itself. Given a tangent vector at one point of space, there is a unique tangent vector at every other point of space that is “the same,” meaning having the same length and direction. Having the same length is no big deal. It is the correlation of all the directions which is the real trick. A flat space has a global parallelism that enables a vector to be translated all around on any path keeping its direction “constant” and still arrive at a given point with the same direction. On a curved space, this is not possible. Think about taking a vector that is tangent to the unit sphere and moving it around the unit sphere so that it remains tangent to the sphere as illustrated in Fig. 8.1. Even though you try to keep its direction fixed along the surface as you move, different paths between two points lead to final vectors whose directions are generally rotated with respect to each other.

To investigate this we need to be able to transport tangent vectors along curves keeping them “covariant constant.” In general coordinates on \mathbb{R}^n , this will enable us to keep the direction (and length) of a vector constant along a curve even though its components are forced to change continuously to compensate for the changing directions and lengths of the coordinate frame vectors. In curved spaces, this will lead to a way of measuring curvature.

8.1 Covariant differentiation along a curve and parallel transport

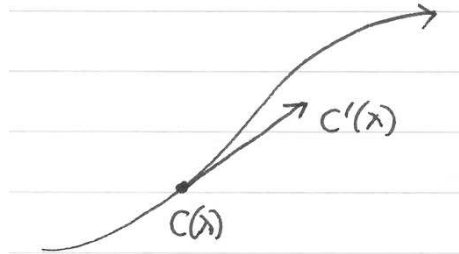


Figure 8.2: A parametrized curve and its tangent vector.

Suppose we have a parametrized curve $c(\lambda)$ in \mathbb{R}^n with tangent vector $c'(\lambda)$. The components of the tangent vector with respect to the Cartesian coordinate frame vectors are the ordinary derivatives of the composition of the Cartesian functions with that tangent vector

$$c(\lambda) = (c^1(\lambda), \dots, c^n(\lambda)),$$

$$c'(\lambda) = c^{i'}(\lambda) \left. \frac{\partial}{\partial x^i} \right|_{c(\lambda)}, \quad c^i(\lambda) = x^i \circ c(\lambda).$$

The directional derivative along the tangent vector of a real-valued function f is then

$$\nabla_{c'(\lambda)} f = c'(\lambda) f = c^{i'}(\lambda) \left. \frac{\partial f}{\partial x^i} \right|_{c(\lambda)} = \frac{d}{d\lambda} [f \circ c(\lambda)],$$

namely just the derivative of the function composed with the parametrized curve, as follows from the chain rule for this composed function.

In a general coordinate system $\{\bar{x}^i\}$ define analogously

$$\bar{c}^i(\lambda) = \bar{x}^i \circ c(\lambda) \quad (\text{evaluate coordinate functions on curve})$$

so that

$$c'(\lambda) = \bar{c}^{i'}(\lambda) \left. \frac{\partial}{\partial \bar{x}^i} \right|_{c(\lambda)}.$$

In sloppy notation typically used in physics, the composition with c would be suppressed and these coordinate functions and components along the curve would just be designated by $x^i(\lambda)$, $\bar{x}^i(\lambda)$ and $x^{i'}(\lambda)$, $\bar{x}^{i'}(\lambda)$, and the directional derivative of a function

$$\nabla_{c'(\lambda)} f = \frac{dx^i(\lambda)}{d\lambda} \left. \frac{\partial f}{\partial x^i} \right|_{c(\lambda)} = \frac{d\bar{x}^i(\lambda)}{d\lambda} \left. \frac{\partial f}{\partial \bar{x}^i} \right|_{c(\lambda)}.$$

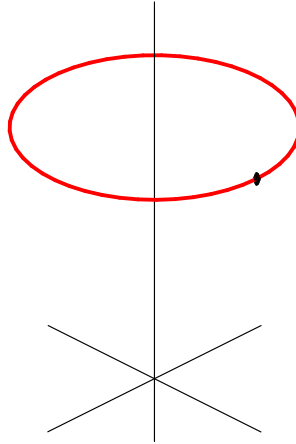


Figure 8.3: A parametrized circle about the vertical axis in a horizontal plane. **[add tangent vectors, transported vectors, angle lambda]**

The equivalence of the two expressions for the directional derivative follows from the chain rule applied to the change of variables, which became the transformation law for the components of the vector field

$$\bar{c}^i{}'(\lambda) = \left. \frac{\partial \bar{x}^i}{\partial x^j} \right|_{c(\lambda)} c^j{}'(\lambda),$$

which implies

$$\bar{c}^i{}'(\lambda) \left. \frac{\partial}{\partial \bar{x}^i} \right|_{c(\lambda)} = c^j{}'(\lambda) \left. \frac{\partial \bar{x}^i}{\partial x^j} \right|_{c(\lambda)} \left. \frac{\partial}{\partial \bar{x}^i} \right|_{c(\lambda)} = c^j{}'(\lambda) \left. \frac{\partial}{\partial x^j} \right|_{c(\lambda)}.$$

This is the foundation of our interpretation of tangent vectors as first order derivative operators.

In multivariable calculus we usually let t be the parameter variable for a parametrized curve in the plane or in space, a variable which we can think of as the time in a physical problem in which the curve represents a path in a space of some variables as a function of time, but here we use the Greek letter lambda to allow any interpretation of the parametrization of the curve. Since the prime is also used to denote the derivative of a function of a single variable, it is now particularly useful to let the overbar denote a new system of coordinates instead of putting a prime on the indices, which would be confusing if both were used together.

Example 8.1.1. On \mathbb{R}^3 consider a parametrized curve $c(\lambda)$ shown in Fig. 8.3 representing one revolution around a circle of radius $\rho_0 = r_0 \sin \theta_0$ centered on the z axis lying in a horizontal plane of constant $z = z_0 = r_0 \cos \theta_0$, parametrized by the angle λ of revolution from the positive x -direction in the counterclockwise direction as seen from above: $\lambda : 0 \rightarrow 2\pi$. This circle is a coordinate line of the ϕ -coordinate in both cylindrical and spherical coordinates (with physicist rather than calculus conventions for the coordinates), and it starts and ends at the point $(x, y, z) = (r_0 \sin \theta_0, 0, r_0 \cos \theta_0)$ above the x -axis. We can describe this (two parameter family of) parametrized curve(s) in all three coordinate systems.

Cartesian: $(x^1, x^2, x^3) = (x, y, z)$

$$\begin{aligned} x &= r_0 \sin \theta_0 \cos \lambda \equiv c^1(\lambda) & dx/d\lambda &= -r_0 \sin \theta_0 \sin \lambda = c^{1'}(\lambda) \\ y &= r_0 \sin \theta_0 \sin \lambda \equiv c^2(\lambda) & dy/d\lambda &= r_0 \sin \theta_0 \cos \lambda = c^{2'}(\lambda) \\ z &= r_0 \cos \theta_0 \equiv c^3(\lambda) & dz/d\lambda &= 0 = c^{3'}(\lambda) \end{aligned}$$

Cylindrical: $(\bar{x}^1, \bar{x}^2, \bar{x}^3) = (\rho, \phi, z)$

$$\begin{aligned} \rho &= r_0 \sin \theta_0 \equiv \bar{c}^1(\lambda) \equiv c^\rho(\lambda) & d\rho/d\lambda &= 0 = \bar{c}^{1'}(\lambda) = c^{\rho'}(\lambda) \\ \phi &= \lambda \equiv \bar{c}^2(\lambda) \equiv c^\phi(\lambda) & d\phi/d\lambda &= 1 = \bar{c}^{2'}(\lambda) = c^{\phi'}(\lambda) \\ z &= r_0 \cos \theta_0 \equiv \bar{c}^3(\lambda) \equiv c^z(\lambda) & dz/d\lambda &= 0 = \bar{c}^{3'}(\lambda) = c^{z'}(\lambda) \end{aligned}$$

Spherical: $(\bar{x}^1, \bar{x}^2, \bar{x}^3) = (r, \theta, \phi)$

$$\begin{aligned} r &= r_0 \equiv \bar{c}^1(\lambda) = c^r(\lambda) & dr/d\lambda &= 0 = \bar{c}^{1'}(\lambda) = c^{r'}(\lambda) \\ \theta &= \theta_0 \equiv \bar{c}^2(\lambda) = c^\theta(\lambda) & d\theta/d\lambda &= 0 = \bar{c}^{2'}(\lambda) = c^{\theta'}(\lambda) \\ \phi &= \lambda \equiv \bar{c}^3(\lambda) = c^\phi(\lambda) & d\phi/d\lambda &= 1 = \bar{c}^{3'}(\lambda) = c^{\phi'}(\lambda) \end{aligned}$$

The tangent vector is then

$$c'(\lambda) = r_0 \sin \theta_0 \underbrace{\left(-\sin \lambda \frac{\partial}{\partial x} \Big|_{c(\lambda)} + \cos \lambda \frac{\partial}{\partial y} \Big|_{c(\lambda)} \right)}_{\text{Cartesian}} = \underbrace{\frac{\partial}{\partial \phi} \Big|_{c(\lambda)}}_{\text{cylindrical}} = \underbrace{\frac{\partial}{\partial \phi} \Big|_{c(\lambda)}}_{\text{spherical}}.$$

□

Since the covariant derivative of a function by a tangent vector is the ordinary derivative of the function by the tangent vector, this establishes a relation between that covariant derivative of the function as a function of all the coordinates to the derivative of the same function only along the curve. We can extend this operation to tensor fields to measure their change along the curve with respect to covariant constant tensor fields. An explicit preliminary example helps make this more concrete.

Example 8.1.2. Consider the function from Exercise 7.1.3

$$f = xy = \frac{1}{2}\rho^2 \sin 2\phi = \frac{1}{2}r^2 \sin^2 \theta \sin 2\phi.$$

We calculate its derivative along the horizontal circle of the previous example in all three

coordinate systems, obtaining the same result in three different ways

$$\begin{aligned}
\nabla_{c'(\lambda)} f &= r_0 \sin \theta_0 \left(-\sin \lambda \frac{\partial}{\partial x} + \cos \lambda \frac{\partial}{\partial y} \right) \Big|_{c(\lambda)} (xy) \\
&= r_0 \sin \theta_0 (-\sin \lambda y + \cos \lambda x) \Big|_{c(\lambda)} \\
&= r_0 \sin \theta_0 (-\sin \lambda [r_0 \sin \theta_0 \sin \lambda] + \cos \lambda [r_0 \sin \theta_0 \cos \lambda]) \\
&= r_0^2 \sin^2 \theta_0 (\cos^2 \lambda - \sin^2 \lambda) = r_0^2 \sin^2 \theta_0 \cos 2\lambda \\
\text{or} \quad &= \frac{\partial}{\partial \phi} \left(\frac{1}{2} \rho^2 \sin 2\phi \right) \Big|_{c(\lambda)} = \frac{1}{2} (r_0^2 \sin \theta_0)^2 2 \cos 2\phi \Big|_{c(\lambda)} = r_0^2 \sin^2 \theta_0 \cos 2\lambda \\
\text{or} \quad &= \frac{\partial}{\partial \phi} \left(\frac{1}{2} r^2 \sin^2 \theta \sin 2\phi \right) \Big|_{c(\lambda)} = r_0^2 \sin^2 \theta_0 \cos 2\lambda .
\end{aligned}$$

□

Exercise 8.1.1.

directional derivative along a curve

Evaluate $\nabla_{c'(\lambda)} f$ for the function $f = x^2 - y^2$ of Exercise 7.1.4 along the same parametrized curve.

■

Now suppose Y is a vector field, i.e., defined everywhere, not just along the curve. Then in general coordinates the covariant derivative of Y along the tangent vector to the curve is

$$\begin{aligned}
\overline{[\nabla_{c'(\lambda)} Y] \circ c(\lambda)}^i &= \bar{Y}^i_{;j} \circ c(\lambda) \bar{c}^j{}'(\lambda) = (\bar{Y}^i_{;j} + \bar{\Gamma}^i_{jk} \bar{Y}^k) \circ c(\lambda) \bar{c}^j{}'(\lambda) \\
&= \bar{Y}^i_{;j} \circ c(\lambda) \bar{c}^j{}'(\lambda) + \bar{\Gamma}^i_{jk} \circ c(\lambda) \bar{Y}^k \circ c(\lambda) \bar{c}^j{}'(\lambda) \\
&= \frac{d}{d\lambda} [\bar{Y}^i \circ c(\lambda)] + \underbrace{\bar{\Gamma}^i_{jk} \circ c(\lambda) \bar{c}^j{}'(\lambda)}_{\bar{\omega}^i_k \circ c(\lambda) (c'(\lambda))} \bar{Y}^k \circ c(\lambda) \\
&\equiv \left[\frac{D}{d\lambda} (Y \circ c(\lambda)) \right]^i .
\end{aligned}$$

where the correction term is given by the value of the connection 1-form matrix on the tangent vector describing the rate of change of the frame along the curve. Usually the dependence on the curve $c(\lambda)$ is suppressed in sloppy notation but understood implicitly to make the expressions look less busy. One writes sloppily

$$\frac{D\bar{Y}^i}{d\lambda} = \frac{d\bar{Y}^i}{d\lambda} + \bar{\Gamma}^i_{jk} \frac{d\bar{x}^j}{dt} \bar{Y}^k ,$$

or ignoring the choice of frame to express this equation, as

$$\frac{DY}{d\lambda} = \nabla_{c'} Y .$$

Notice that at a given point on the curve, all we need to know about the vector field being differentiated in order to evaluate its covariant derivative along the curve are the values of its component functions along the curve *and nowhere else*. Their values at all points of \mathbb{R}^n off the curve are irrelevant. This enables us to extend the idea of covariant differentiation to a vector function which is only defined along a curve and nowhere else. The covariant derivative along the parametrized curve indicated by the notation $D/d\lambda$ is sometimes referred to as the intrinsic derivative or the “total covariant derivative” along the curve.

Example 8.1.3. Continuing the previous two examples, consider the covariant derivative of the vector field from Example 5.7.1

$$X = y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} = \rho \sin 2\phi \frac{\partial}{\partial \rho} + \cos 2\phi \frac{\partial}{\partial \phi}$$

along the horizontal circle $c(\lambda)$ given above. In Cartesian coordinates $D/d\lambda$ just reduces to $d/d\lambda$ of the components of X along $c(\lambda)$, i.e.,

$$\begin{aligned} \frac{D}{d\lambda} (X \circ c(\lambda)) &= \frac{d[y \circ c(\lambda)]}{d\lambda} \frac{\partial}{\partial x} \Big|_{c(\lambda)} + \frac{d[x \circ c(\lambda)]}{d\lambda} \frac{\partial}{\partial y} \Big|_{c(\lambda)} \\ &= \frac{d[r_0 \sin \theta_0 \sin \lambda]}{d\lambda} \frac{\partial}{\partial x} \Big|_{c(\lambda)} + \frac{d[r_0 \sin \theta_0 \cos \lambda]}{d\lambda} \frac{\partial}{\partial y} \Big|_{c(\lambda)} \\ &= r_0 \sin \theta_0 \cos \lambda \frac{\partial}{\partial x} \Big|_{c(\lambda)} - r_0 \sin \theta_0 \sin \lambda \frac{\partial}{\partial y} \Big|_{c(\lambda)} \\ &= \left(x \frac{\partial}{\partial x} - y \frac{\partial}{\partial y} \right) \Big|_{c(\lambda)}. \end{aligned}$$

In cylindrical coordinates

$$\bar{c}^{i'}(\lambda) = \bar{\delta}^i_2, \quad \text{or} \quad (c^{\rho'}(\lambda), c^{\phi'}(\lambda), c^{z'}(\lambda)) = (0, 1, 0),$$

so

$$\bar{\omega}^i_k(c'(\lambda)) = \bar{\Gamma}^i_{\phi k} \circ c(\lambda) c^{\phi'}(\lambda)$$

and

$$\begin{aligned} X^\rho &= \rho \sin 2\phi, & X^\rho \circ c(\lambda) &= r_0 \sin \theta_0 \sin 2\lambda, \\ X^\phi &= \cos 2\phi, & X^\phi \circ c(\lambda) &= \cos 2\lambda, \end{aligned}$$

so the covariant derivative of X along c is (refer to the connection components listed in Exercise 6.2.1)

$$\begin{aligned} \frac{DX^\rho}{d\lambda} &= \frac{dX^\rho}{d\lambda} + \underbrace{\Gamma^\rho_{\phi\phi}}_{-\rho} X^\phi = 2r_0 \sin \theta_0 \cos 2\lambda - r_0 \sin \theta_0 \cos 2\lambda = r_0 \sin \theta_0 \cos 2\lambda, \\ \frac{DX^\phi}{d\lambda} &= \frac{dX^\phi}{d\lambda} + \underbrace{\Gamma^\phi_{\phi\rho}}_{\rho^{-1}} X^\rho = -\sin 2\lambda + \frac{1}{r_0 \sin \theta_0} (r_0 \sin \theta_0 \sin 2\lambda) = -\sin 2\lambda, \\ \frac{DX^z}{d\lambda} &= 0, \end{aligned}$$

so

$$\left. \frac{DX}{d\lambda} \right|_{c(\lambda)} = r_0 \sin \theta_0 \cos 2\lambda \left. \frac{\partial}{\partial \rho} \right|_{c(\lambda)} - \sin 2\lambda \left. \frac{\partial}{\partial \phi} \right|_{c(\lambda)} .$$

□

Exercise 8.1.2.

directional derivative along curve in cylindrical coordinates

Verify that this cylindrical coordinate result is the value along this curve of the Cartesian expression $DX/d\lambda|_{c(\lambda)} = (x\partial/\partial x - y\partial/\partial y)|_{c(\lambda)}$ following Example 5.7.1.??

■

One can do this for any tensor field T , defining its covariant derivative along a parametrized curve by the analogous formula which follows from evaluating $\nabla_{c'(\lambda)} T$, or in terms of components

$$\begin{aligned} \overline{[\nabla_{c'(\lambda)} T]}^{i\dots}_{j\dots} &= \bar{T}^{i\dots}_{j\dots;k} \circ c(\lambda) \bar{c}^k{}'(\lambda) \\ &= \bar{T}^{i\dots}_{j\dots;k} \circ c(\lambda) \bar{c}^k{}'(\lambda) + \bar{\Gamma}^i_{k\ell} \circ c(\lambda) \bar{c}^k{}'(\lambda) \bar{T}^{\ell\dots}_{j\dots} \circ c(\lambda) + \dots \\ &\quad - \bar{\Gamma}^\ell_{kj} \circ c(\lambda) \bar{c}^k{}'(\lambda) \bar{T}^{i\dots}_{\ell\dots} \circ c(\lambda) - \dots , \end{aligned}$$

and using the chain rule to re-express the first term representing the ordinary derivative of the components along the curve

$$\begin{aligned} \overline{\left[\frac{DT \circ c(\lambda)}{d\lambda} \right]}^{i\dots}_{j\dots} &= \frac{d}{d\lambda} (\bar{T}^{i\dots}_{j\dots} \circ c(\lambda)) + \bar{\Gamma}^i_{k\ell} \circ c(\lambda) \bar{c}^k{}'(\lambda) \bar{T}^{\ell\dots}_{j\dots} + \dots \\ &\quad - \bar{\Gamma}^\ell_{kj} \circ c(\lambda) \bar{c}^k{}'(\lambda) \bar{T}^{i\dots}_{\ell\dots} - \dots . \end{aligned}$$

This formula defines the intrinsic derivative or total covariant derivative of the tensor field along the parametrized curve, but it only requires the values of the components of the tensor field along the curve for its evaluation. It is usually written simply $DT/d\lambda$, suppressing reference to the curve, which makes the formulas look more complicated. Thus in sloppy notation

$$\frac{D}{d\lambda} \bar{T}^{i\dots}_{j\dots} = \frac{d}{d\lambda} \bar{T}^{i\dots}_{j\dots} + \bar{\Gamma}^i_{k\ell} \bar{T}^{\ell\dots}_{j\dots} + \dots - \bar{\Gamma}^\ell_{kj} \bar{T}^{i\dots}_{\ell\dots} - \dots .$$

The metric is covariant constant so

$$\frac{D\bar{g}_{ij}}{d\lambda} = \overline{\left[\frac{Dg \circ c(\lambda)}{d\lambda} \right]}_{ij} = \overline{[\nabla_{c'(t)} g]}_{ij} = 0 ,$$

i.e., the intrinsic derivative of the metric along any curve is zero.

If Y is any covariant constant vector field, $\nabla Y = 0$, then it too will have vanishing covariant derivative along any curve

$$0 = \overline{\left[\frac{DY \circ c(\lambda)}{d\lambda} \right]}^i = \frac{d\bar{Y}^i \circ c(\lambda)}{d\lambda} + \bar{\Gamma}^i_{jk} \circ c(\lambda) \bar{c}^j{}'(\lambda) \bar{Y}^k \circ c(\lambda) .$$

Notice that everything in this system of first order linear ordinary differential equations for the functions $\bar{Y}^i \circ c(\lambda)$ is a function of λ alone. Suppose we just specify arbitrary particular values at $\lambda = 0$

$$\bar{Y}^i(c(0)) \equiv \bar{Y}_{(0)}^i.$$

These are initial conditions for this system of differential equations. Together they represent an initial value problem which is guaranteed to have a unique solution $\bar{Y}^i(\lambda)$ as covered in any first course in differential equations. We have therefore succeeded in defining a vector $Y(\lambda) = \bar{Y}^i(\lambda)\partial/\partial\bar{x}^i|_{c(\lambda)}$ along the curve which doesn't change its components with respect to a Cartesian frame as we move along the curve and is just the composition of the original vector field with the curve: $Y(\lambda) = Y \circ c(\lambda)$. This process describes the “*parallel transport*” of the initial tangent vector along the curve, sometimes called “parallel translation” and can be used in any space provided a metric is available, even if covariant constant vector fields are not.

Example 8.1.4. Continuing the previous example, the equations $DY/d\lambda = 0$ expressed in cylindrical coordinates are explicitly

$$\begin{cases} \frac{dY^\rho}{d\lambda} - r_0 \sin \theta_0 Y^\phi = 0 \\ \frac{dY^\phi}{d\lambda} + \frac{1}{r_0 \sin \theta_0} Y^\rho = 0 \\ \frac{dY^z}{d\lambda} = 0 \end{cases} \leftrightarrow \begin{cases} \frac{dY^\rho}{d\lambda} = (r_0 \sin \theta_0 Y^\phi) \\ \frac{d}{d\lambda}(r_0 \sin \theta_0 Y^\phi) = -Y^\rho \\ \frac{dY^z}{d\lambda} = 0 \end{cases} \leftrightarrow \begin{cases} \frac{du^1}{d\lambda} = u^2 \\ \frac{du^2}{d\lambda} = -u^1 \\ \frac{du^3}{d\lambda} = 0 \end{cases},$$

where we have introduced the orthonormal component combinations

$$(Y^{\hat{\rho}}, Y^{\hat{\phi}}, Y^{\hat{z}}) = (Y^\rho, \rho Y^\phi, Y^z) = (u^1, u^2, u^3)$$

along the curve where $\rho = r_0 \sin \theta_0$ (suppressing the dependence on λ) which simplifies the equations. This is a constant coefficient system of first order linear differential equations which is studied in every first course in differential equations. The last variable $u^3 = u_0^3$ is just a constant, and the first two satisfy the matrix differential equation

$$\frac{d}{d\lambda} \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} u^1 \\ u^2 \end{pmatrix}.$$

We verify that its solution

$$\begin{pmatrix} u^1 \\ u^2 \end{pmatrix} = \begin{pmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{pmatrix} \begin{pmatrix} u_0^1 \\ u_0^2 \end{pmatrix}$$

satisfies the system of differential equations

$$\begin{aligned} \frac{d}{d\lambda} \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} &= \begin{pmatrix} -\sin \lambda & \cos \lambda \\ -\cos \lambda & -\sin \lambda \end{pmatrix} \begin{pmatrix} u_0^1 \\ u_0^2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{pmatrix} \begin{pmatrix} u_0^1 \\ u_0^2 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} u^1 \\ u^2 \end{pmatrix}. \end{aligned}$$

This solution represents a clockwise rotation by angle $\lambda = \phi$ of the components in the $e_{\hat{r}}-e_{\hat{\phi}}$ horizontal plane in the tangent space, which exactly compensates the counterclockwise rotation of those axes along the circle as one rotates around it as the azimuthal angle ϕ increases. At $\lambda = 0$, the orthonormal cylindrical coordinate frame at the point $c(0)$ above the x -axis coincides with the Cartesian coordinate frame there.

All together we have explicitly

$$\begin{aligned} u^1 &= \cos \lambda u_0^1 + \sin \lambda u_0^2 \\ u^2 &= -\sin \lambda u_0^1 + \cos \lambda u_0^2 \\ u^3 &= u_0^3. \end{aligned}$$

which we can convert back to Y^ρ, Y^ϕ, Y^z to get the solution for initial conditions

$$(Y_{(0)}^\rho, r_0 \sin \theta_0 Y_{(0)}^\phi, Y_{(0)}^z) = (u_0^1, u_0^2, u_0^3)$$

corresponding to the initial tangent vector

$$Y(0) = u_0^1 \frac{\partial}{\partial x} \Big|_{c(0)} + u_0^2 \frac{\partial}{\partial y} \Big|_{c(0)} + u_0^3 \frac{\partial}{\partial z} \Big|_{c(0)}.$$

Of course we know $Y(\lambda)$ along the curve $c(\lambda)$ just represents the components of the constant vector field

$$Y = u_0^1 \frac{\partial}{\partial x} + u_0^2 \frac{\partial}{\partial y} + u_0^3 \frac{\partial}{\partial z},$$

and so is independent of the parametrized curve which takes us from our initial to our final point.

This example gives you an idea how parallel transport process can be done explicitly when the differential equations are manageable.

□

Suppose we have two points P and Q and two curves $c(\lambda)$ and $\zeta(\lambda)$ with $c(0) = \zeta(0) = P$ and $c(\lambda_1) = \zeta(\lambda_1) = Q$. If we transport a tangent vector $Y_{(0)}$ at P along each curve to Q , of course we'll get the same result. This *path independence* of parallel transport on \mathbb{R}^n is a feature of its *flat geometry*.

“Parallel transport” is called “parallel” because at each point on the curve we move the vector to the next tangent space so that it remains parallel to itself (and also has constant length). This operation provides a “connection” between any two tangent spaces to points connected to each other by a simple curve. Every vector in the first tangent space can be transported along the curve to the tangent space at the second point (indeed any other point on the curve), establishing a vector space isomorphism between them. (This mapping also preserves lengths and angles so it maps the Euclidean geometry of the first onto that of the second.) For this reason a covariant derivative on a space is often called a “connection,” and Γ_{jk}^i are called the “components of the connection.”

Inner products (and so all lengths and relative angles) of tangent vectors are preserved under this operation since the metric is covariant constant

$$g_{ij;k} = 0 \longrightarrow \frac{D}{d\lambda} g_{ij} = [\nabla_{c'(\lambda)} g]_{ij} = g_{ij;k} c^{k'}(\lambda) = 0,$$

$$\frac{d}{d\lambda} (g_{ij} X^i Y^j) = \frac{D}{d\lambda} (g_{ij} X^i Y^j) = \underbrace{\left(\frac{Dg_{ij}}{d\lambda} \right)}_{=0} X^i Y^j + g_{ij} \frac{DX^i}{d\lambda} Y^j + g_{ij} X^i \frac{DY^j}{d\lambda}.$$

Thus if X and Y are parallel transported along the curve

$$\frac{DX^i}{d\lambda} = 0 = \frac{DY^i}{d\lambda},$$

we get

$$\frac{d}{d\lambda} (g_{ij} X^i Y^j) = 0,$$

i.e., their inner product is a constant along the curve. For the case $X = Y$, this shows that lengths are preserved, and hence the inner products of unit vectors, i.e., angles are also constant under this transport. All of these properties are obvious if we work in terms of Cartesian coordinates where parallel transport amounts to defining a tensor along a curve whose Cartesian coordinate frame components are constants, but they are not obvious working in a non-Cartesian coordinate system or in a general frame where everything depends on position.

Everybody's favorite example of a curved surface is a sphere in \mathbb{R}^3 , which we can take to be a coordinate sphere $r = r_0$ in spherical coordinates, leaving the angles θ and ϕ to serve as coordinates on the 2-dimensional space of points on that sphere. Their coordinate lines are the mesh of lines of longitude and latitude, although for the latter we measure the latitude as an angle from the equator instead of the North Pole.

As illustrated in Fig. 8.5, suppose we take a unit vector at the North Pole tangent to the sphere and making an angle 30° with the $\phi = 0$ coordinate line. If we move this vector down the line of longitude $\phi = 0$, so that the vector remains tangent to the sphere, but maintains a 30° angle with respect to the direction in which we are moving, then of course the direction of the vector has to change in order to remain in the 2-dimensional tangent plane to the sphere at each point, but apart from this necessary rotation no further unnecessary rotation occurs. Keep on going around the equator as shown and then come back up to the North Pole, where it will now be at an angle of 30° to the line of longitude $\phi = \phi_0$. *But* it will be rotated by an angle ϕ_0 with respect to its initial direction!

This is the manifestation of curvature. If you transport a vector around a closed curve in a curved space, it will undergo a rotation in general. Of course on some curves it may not—for example on any great circle, take the equator to be explicit, the above exercise will not change the initial vector upon completion of one revolution.

Exercise 8.1.3.

covariant derivative in spherical coordinates

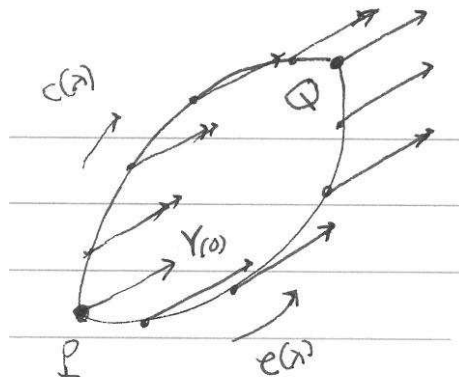


Figure 8.4: Parallel transport is path independent in a flat geometry.

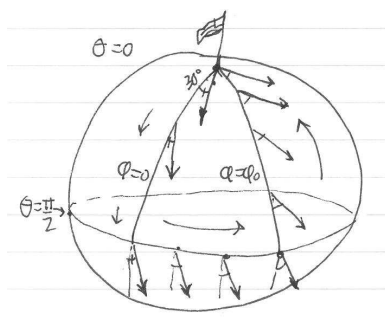


Figure 8.5: Parallel transport on a curved geometry like the sphere depends on the path. Around a closed loop the final vector is rotated with respect to the initial vector.

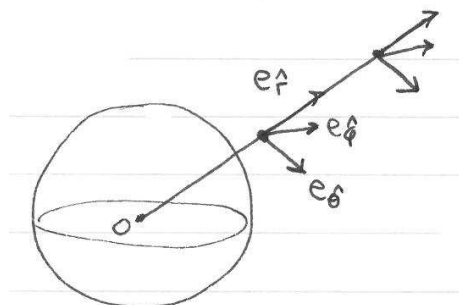


Figure 8.6: $e_{\hat{r}}$ is clearly parallel transported along its own straight coordinate lines, and $e_{\hat{\theta}}$, $e_{\hat{\phi}}$ are also parallel transported along those lines. This also is clear since their Cartesian components do not depend on r and hence do not change as r changes.

It is clear from the geometry of spherical coordinates that the orthonormal frame $\{e_{\hat{r}}, e_{\hat{\theta}}, e_{\hat{\phi}}\}$ is parallel transported along the r coordinate lines. These curves may be parametrized in spherical coordinates $\{\bar{x}^i\} = \{r, \theta, \phi\}$ by

$$\begin{aligned} r &= c^r(\lambda) = \lambda & c^{r'}(\lambda) &= 1 \\ \theta &= c^\theta(\lambda) = \theta_0 & c^{\theta'}(\lambda) &= 0 \\ \phi &= c^\phi(\lambda) = \phi_0 & c^{\phi'}(\lambda) &= 0, \end{aligned}$$

and their tangent vectors are

$$c'(\lambda) = \bar{c}^{i'}(\lambda) \frac{\partial}{\partial \bar{x}^i} \Big|_{c(\lambda)} = \frac{\partial}{\partial r} \Big|_{c(\lambda)} = e_{\hat{r}}|_{c(\lambda)}.$$

Since $e_{\hat{r}}$ is itself the unit tangent vector, the covariant derivative along this curve expressed in the orthonormal frame is

$$\frac{DX^{\hat{i}}}{d\lambda} = \frac{dX^{\hat{i}}}{d\lambda} + \Gamma^{\hat{i}}_{\hat{k}\hat{j}} c^{\hat{k}} X^{\hat{j}} = \frac{dX^{\hat{i}}}{d\lambda} + \Gamma^{\hat{i}}_{\hat{r}\hat{j}} X^{\hat{j}}.$$

Letting X be one of the vectors $\{e_{\hat{r}}, e_{\hat{\theta}}, e_{\hat{\phi}}\}$ results in constant components $X^{\hat{i}}$ (for example $[e_{\hat{r}}]^{\hat{i}} = \delta^{\hat{i}}_{\hat{r}}$), so the term $dX^{\hat{i}}/d\lambda$ vanishes.

The components $\Gamma^{\hat{i}}_{\hat{j}\hat{k}}$ were given previously in section 7.3

$$\begin{aligned} \Gamma^{\hat{r}}_{\hat{\theta}\hat{\theta}} &= -r^{-1} = -\Gamma^{\hat{\theta}}_{\hat{\theta}\hat{r}}, \\ \Gamma^{\hat{r}}_{\hat{\phi}\hat{\phi}} &= -r^{-1} = -\Gamma^{\hat{\phi}}_{\hat{\phi}\hat{r}}, \\ \Gamma^{\hat{\theta}}_{\hat{\phi}\hat{\phi}} &= -r^{-1} \cot \theta = -\Gamma^{\hat{\phi}}_{\hat{\phi}\hat{\theta}}. \end{aligned}$$

Thus for example

$$\frac{D}{d\lambda} [e_{\hat{r}}]^{\hat{i}} = \frac{d}{d\lambda} [e_{\hat{r}}]^{\hat{i}} + \Gamma^{\hat{i}}_{\hat{r}\hat{j}} [e_{\hat{r}}]^{\hat{j}} = \Gamma^{\hat{i}}_{\hat{r}\hat{r}} = 0 \longrightarrow \frac{D}{d\lambda} [e_{\hat{r}}] = 0$$

says that $e_{\hat{r}}$ is parallel transported along this curve. Verify that $e_{\hat{\theta}}$ is also parallel transported along r in the following two different ways.

- First use the orthonormal frame as in the radial unit vector example.
- Next use the coordinate frame, where $c'(\lambda) = e_{\hat{r}} = e_r$ and

$$[e_{\hat{\theta}}]^r = 0, \quad [e_{\hat{\theta}}]^\theta = r^{-1}, \quad [e_{\hat{\theta}}]^\phi = 0,$$

and

$$\begin{aligned} \Gamma^\theta_{r\theta} &= \Gamma^\phi_{r\phi} = r^{-1}, & \Gamma^r_{\theta\theta} &= -r, & \Gamma^\theta_{\theta r} &= r^{-1}, \\ \Gamma^r_{\phi\phi} &= -r \sin^2 \theta, & \Gamma^\phi_{\phi r} &= r^{-1}, \\ \Gamma^\theta_{\phi\phi} &= -\cos \theta \sin \theta, & \Gamma^\phi_{\phi\theta} &= \cot \theta, \end{aligned}$$

and one has the formula

$$\frac{DX^i}{d\lambda} = \frac{dX^i}{d\lambda} + \Gamma^i_{rj} X^j.$$

■

8.2 Parallel transport within coordinate surfaces in space

Cylindrical and spherical coordinates are constructed using some interesting surfaces whose intrinsic geometry is worth studying: cylinders, cones and spheres. Cylinders and cones are flat in the sense that they can be cut open and laid flat on a plane, but spheres are truly curved intrinsically since flattening them out requires a deformation. Since these surfaces are coordinate surfaces in one of these coordinate systems, it is very easy to restrict our attention to a particular surface where the corresponding coordinate $\mathcal{X} = x^m$ is constant. Then by setting $d\mathcal{X} = 0$ in the metric line element, we get the line element of the metric on the surface (called the “induced metric”), and to consider parallel transport on this surface, we can simply consider vectors with no \mathcal{X} component and ignore the components of the connection Γ^i_{jk} which have an \mathcal{X} index. In this way we can study a cylinder $\rho = \rho_0$ in cylindrical coordinates, or a sphere $r = r_0$ or cone $\theta = \theta_0$ in spherical coordinates. As an instructive example, we consider such a sphere of radius $r = r_0$.

It is convenient to work in the orthonormal frame

$$e_{\hat{\theta}} = \frac{1}{r_0} \frac{\partial}{\partial \theta}, \quad e_{\hat{\phi}} = \frac{1}{\sin \theta_0 r_0} \frac{\partial}{\partial \phi}$$

on the sphere since parallel transport results in a rotation, easily described in that frame. Consider one revolution of a ϕ -coordinate line $\theta = \theta_0$, already used above as an example curve for parallel transport of general vectors in space around a circle in the flat space geometry. Now we want to consider instead how tangent vectors in the tangent planes to the sphere can be transported around the circle, always remaining within those tangent planes, using the components of the connection that only involve the two angular coordinates θ, ϕ . The only nonzero orthonormal components of the connection

$$\Gamma^{\hat{\theta}}_{\hat{\phi}\hat{\phi}} = -r_0^{-1} \cot \theta_0 = -\Gamma^{\hat{\phi}}_{\hat{\phi}\hat{\theta}}$$

correspond to the two nonzero orthonormal components of the connection ϕ component of the connection 1-form matrix restricted to the sphere

$$\omega_{\hat{\phi}} = (\Gamma^{\hat{i}}_{\hat{\phi}\hat{j}}) = r_0^{-1} \cot \theta_0 \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

which generates rotations in the θ - ϕ tangent plane in spherical coordinates as one moves along the ϕ -coordinate lines.

The parametrized curve representing the ϕ -coordinate line in this sphere is

$$c(\lambda) : \quad \theta = \theta_0, \quad \phi = \lambda, \quad 0 \leq \lambda \leq 2\pi,$$

with tangent vector

$$c'(\lambda) = \left. \frac{\partial}{\partial \phi} \right|_{c(\lambda)} = r_0 \sin \theta_0 e_{\hat{\phi}}|_{c(\lambda)}.$$

For a tangent vector with orthonormal components $\langle Y^{\hat{\theta}}, Y^{\hat{\phi}} \rangle$ along this curve the parallel transport equations are

$$\frac{dY^{\hat{i}}}{d\lambda} = -\Gamma^{\hat{i}}_{\hat{k}\hat{j}} c'(\lambda)^{\hat{k}} Y^{\hat{j}} = -r_0 \sin \theta_0 \Gamma^{\hat{i}}_{\hat{\phi}\hat{j}} Y^{\hat{j}}.$$

Exercise 8.2.1.

parallel transport along lines of latitude

Show that these parallel transport equations along the ϕ coordinate circles take the matrix form

$$\frac{d}{d\lambda} \begin{pmatrix} Y^{\hat{\theta}} \\ Y^{\hat{\phi}} \end{pmatrix} = -\cos \theta_0 \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} Y^{\hat{\theta}} \\ Y^{\hat{\phi}} \end{pmatrix},$$

which implies that the angular velocity of the parallel transported vector with respect to the spherical frame is $-\cos \theta_0$. The solution of these equations is just an active rotation by the angle $\Phi = -(\cos \theta_0) \lambda$ (where in turn $\lambda = \phi$)

$$\begin{pmatrix} Y^{\hat{\theta}} \\ Y^{\hat{\phi}} \end{pmatrix} = \begin{pmatrix} \cos \Phi & -\sin \Phi \\ \sin \Phi & \cos \Phi \end{pmatrix} \begin{pmatrix} Y^{\hat{\theta}}(0) \\ Y^{\hat{\phi}}(0) \end{pmatrix}.$$

After one revolution from $\lambda = 0$ to $\lambda = 2\pi$, we return to our starting point with the final values

$$\begin{pmatrix} Y^{\hat{\theta}}(2\pi) \\ Y^{\hat{\phi}}(2\pi) \end{pmatrix} = \begin{pmatrix} \cos \Delta\Phi & -\sin \Delta\Phi \\ \sin \Delta\Phi & \cos \Delta\Phi \end{pmatrix} \begin{pmatrix} Y^{\hat{\theta}}(0) \\ Y^{\hat{\phi}}(0) \end{pmatrix}$$

where the net rotation relative to this orthonormal frame is

$$\Delta\Phi = -2\pi \cos \theta_0,$$

which is a clockwise rotation in the upper hemisphere. In a subsequent exercise we will use a tangent cone to the sphere at a ϕ coordinate circle to derive this result without solving any differential equation. ■

Exercise 8.2.2.

parallel combed hair on the sphere

Take the unit vector $e_{\hat{\phi}}$ along the line of longitude $\phi = 0$ (orthonormal components $\langle 0, 1 \rangle$) and parallel transport it along the line of latitude one revolution of the circle. According to the preceding exercise this generates a vector field on the sphere with orthonormal components $\langle \sin(\phi \cos \theta), \cos(\phi \cos \theta) \rangle$, namely

$$V(\theta, \phi) = \sin(\phi \cos \theta) e_{\hat{\theta}} + \cos(\phi \cos \theta) e_{\hat{\phi}}.$$

a) Show that its Cartesian components as a column matrix are

$$V(\theta, \phi) = \begin{pmatrix} \sin(\phi \cos \theta) \cos \theta \cos \phi - \cos(\phi \cos \theta) \sin \phi \\ \sin(\phi \cos \theta) \cos \theta \sin \phi + \cos(\phi \cos \theta) \cos \phi \\ - \sin(\phi \cos \theta) \sin \theta \end{pmatrix}.$$

b) Use a computer algebra system to plot this field of unit vectors on an equally spaced grid of points on a sphere of radius 4 (so that the arrows are relatively small compared to the sphere itself. This gives a nice visualization of the parallel transport and its effect on the initial direction at $\phi = 0$ after one revolution about the vertical axis.

c) To mark the points where the rotation has progressed through a multiple of $\pi/2$, plot on your “sphere with hair” (the previous plot) the space curves at which $\phi \cos \theta = j\pi/2$ for $j = 1, 2, 3$. This partitions the sphere into zones of comparable rotation intervals. ■

At the equator $\theta = \pi/2$ where $\cos \theta_0 = 0$, parallel transported vectors have constant components in the orthonormal frame, which means that this frame is itself parallel transported around the equator in the internal geometry of the sphere. Of course these vectors must rotate relative to the corresponding vectors parallel transported in the geometry of the surrounding space in order to remain within the tangent planes to the sphere, but no further rotation occurs within those tangent planes. As we approach the North Pole at small polar angles $0 < \theta_0 \ll 1$ then $\Delta\Phi$ approaches -2π since the orthonormal frame itself makes one complete rotation around the vertical axis compared to the parallel transported vectors which are approximately aligned with the horizontal tangent plane at the North Pole and approximately try to maintain their direction fixed in that tangent plane in the limit of very small polar angle. $e_{\hat{\theta}}$ is approximately pointing radially outward in that limiting tangent plane, while as always $e_{\hat{\phi}}$ is tangential to the ϕ coordinate circle. These both rotate one complete revolution 2π in the counterclockwise direction about the vertical axis in this limit, so the parallel transported vectors compensate by rotating by -2π relative to that frame. The parallel transported vector falls short of one complete backwards (clockwise) revolution by the small positive angle

$$\Delta\Phi + 2\pi = 2\pi(1 - \cos \theta_0),$$

which is the net angle forward in the counterclockwise direction that the final vector makes with its initial value. This is a prograde rotation in the sense that each time the vector returns to its original location going around the circle, it rotates by a small amount in the same direction as one traces out the circle. As one increases the polar angle to $\pi/2$, this forward counterclockwise rotation grows to nearly 2π , so that near the Equator the comparison of final and initial values leads to a small clockwise rotation with respect to the orthonormal frame, which then vanishes at the Equator itself.

These two limiting cases fit with our intuition about what happens at the Equator and North Pole, which is satisfying. In between the formula must interpolate between the two behaviors, which is what the cosine factor does for us. This result can be calculated without solving any differential equations or using any fancy differential geometry by considering the

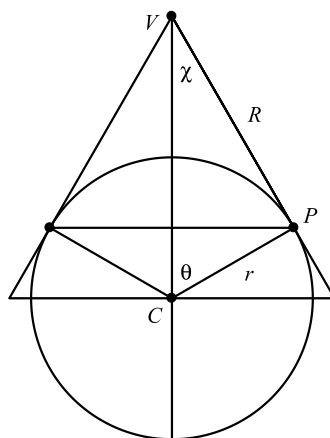


Figure 8.7: A plane cross-section of the cone tangent to a polar circle on a sphere of radius r through the vertical symmetry axis of the cone. The right triangle VPC with vertices equal to the center C of the sphere, the vertex V of the cone, and the point P of contact makes the conical opening angle χ and the polar angle θ complementary angles: $\chi + \theta = \pi/2$, which in turn makes χ equal to the “latitude” polar angle measured up from the Equator. Note $\cos \theta = \sin \chi$.

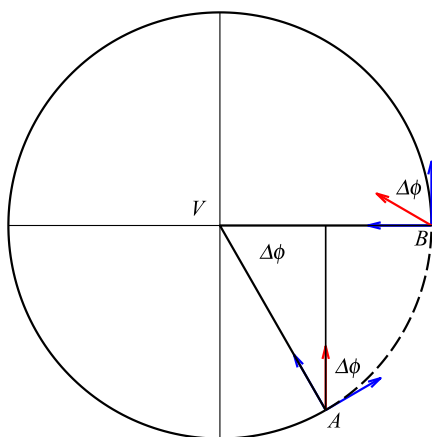


Figure 8.8: A cone can be created from a circle by removing a sector of angle $\Delta\phi$ called the defect angle and pulling the two edges of the sector together to identify them. A fixed vertical vector (red) in the plane of the circle, after removal of the sector, has its value from the point A rotated counterclockwise by the defect angle relative to its original value at B . Thus parallel transport of a vector counterclockwise around such a circle in the cone leads to a counterclockwise rotation of its direction by the defect angle relative to its original value after one revolution.

tangent cone to the sphere whose intersection with the sphere is the ϕ coordinate circle. One only needs high school trigonometry.

Exercise 8.2.3.

Tangent cone to a sphere

a) By comparing the partial circumference of the circle of radius R in Fig. 8.8 with the circumference of the polar circle at angle θ on the cone of half-opening angle χ in Fig. 8.7, which are equal, show that the defect angle is related χ by

$$\Delta\phi = 2\pi(1 - \sin\chi).$$

b) Since the polar angle and the opening angle are complementary angles in Fig. 8.7, this means that $\sin\chi = \cos\theta$ so for the tangent cone to a polar circle on the sphere, we then have

$$\Delta\phi = 2\pi(1 - \cos\theta).$$

Imagine the upward pointing vertical vector at point A in Fig. 8.8 moving around the circle to point B , relative to the orthonormal frame associated with polar coordinates in that plane. It rotates in the clockwise direction as moves around the circle counterclockwise. Convince yourself that we must subtract 2π from $\Delta\phi$ to get the net angle that the parallel transported vector has rotated with respect that orthonormal frame on the sphere. This net rotation angle is therefore

$$\Delta\Phi = \Delta\phi - 2\pi = -2\pi \cos\theta.$$

To confirm the first equality of this pair here, if the defect angle $\Delta\phi$ is an acute angle as in Fig. 8.8, then the parallel transported vector falls short of a complete rotation backwards with respect to the spherical orthonormal frame by this acute defect angle, which is what this formula implies. Thus we have recovered the previous result obtained by solving the parallel transport differential equations, but with simple geometry, a satisfying accomplishment. This works because the geometry in the limiting strip about the circle on the sphere is indistinguishable from the flat geometry of the tangent cone in an infinitesimal neighborhood of the circle. ■

We can also consider parallel transport along the θ coordinate circles, but these are great circles, and as we will see next, they are what we will call geodesics along which the tangent vector is parallel transported and other parallel transported vectors maintain their inner products with the tangent vector and with each other. Since $e_{\hat{\theta}}$ is the unit tangent to those coordinate circles, it and its orthogonal partner $e_{\hat{\phi}}$ are both parallel transported along them. This explains the vanishing of the θ -component of the connection 1-form matrix

$$\omega_{\hat{\theta}} = (\Gamma^i_{\hat{\theta}\hat{j}}) = 0.$$

In the original spherical coordinate system in space, the vanishing of the r component of the connection 1-form matrix similarly implies that the spherical orthonormal frame is parallel transported along the r -coordinate lines.

In cylindrical coordinates the orthonormal frame associated with the coordinate system only undergoes a rotation along the ϕ -coordinate circles, with the radial and azimuthal unit vectors rotating in the horizontal plane of constant z . This corresponds to the vanishing of all but the ϕ -component of the connection 1-form matrix (see Exercise [6.2.1](#))

$$\omega_{\hat{\phi}} = (\Gamma^{\hat{i}}_{\hat{\phi}\hat{j}}) = \rho^{-1} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} .$$

8.3 Geodesics

“Straight lines,” “autoparallel curves,” “geodesics”!

How can we characterize the straight lines of \mathbb{R}^n ? The tangent vector “follows its nose” as it moves along such a line. The tangent vector itself is parallel transported along the line, i.e., its intrinsic derivative along the line vanishes

$$\frac{Dc'(\lambda)}{d\lambda} = 0,$$

which in any coordinate system takes the component form (dropping the barred notation which distinguished non-Cartesian coordinates)

$$\begin{aligned} 0 &= \frac{Dc^{i'}(\lambda)}{d\lambda} = \frac{dc^{i'}(\lambda)}{d\lambda} + \Gamma^i_{kj} \circ c(\lambda) c^{k'}(\lambda) c^{j'}(\lambda) \\ &= \frac{d^2c^i(\lambda)}{d\lambda^2} + \Gamma^i_{kj} \circ c(\lambda) c^{k'}(\lambda) c^{j'}(\lambda) \\ &= \frac{d^2c^i(\lambda)}{d\lambda^2} + \Gamma^i_{kj} \circ c(\lambda) \frac{dc^k(\lambda)}{d\lambda} \frac{dc^j(\lambda)}{d\lambda} \end{aligned}$$

or in sloppy notation where $c^i(\lambda)$ becomes just x^i

$$\frac{D^2x^i}{d\lambda^2} = \frac{d^2x^i}{d\lambda^2} + \Gamma^i_{kj} \frac{dx^k}{d\lambda} \frac{dx^j}{d\lambda} = 0,$$

or in terms of the tangent vector $u^i = dx^i/d\lambda$

$$\frac{Du_i}{d\lambda} = \frac{du^i}{d\lambda} + \Gamma^i_{kj} u^k u^j = 0.$$

In the Cartesian coordinates where $\Gamma^i_{kj} = 0$, this reduces to

$$\frac{d^2x^i}{d\lambda^2} = 0$$

with solution

$$x^i = a^i \lambda + b^i$$

or more precisely $c^i(\lambda) = a^i \lambda + b^i$. The Cartesian coordinates are linear functions of the parameter, while the Cartesian components of the tangent vector are constants $c^{i'}(\lambda) = a^i$.

Example 8.3.1. We saw above that the tangent vector $c'(\lambda) = e_r$ to the r coordinate lines in their natural parametrization $\lambda = r$ is parallel transported along them. This is clear since these coordinate lines are straight half-lines. Since $dx^k/d\lambda = \delta^k_r$ holds for these curves in spherical coordinates, one can write

$$\frac{D^2x^k}{d\lambda^2} = \frac{d^2x^k}{d\lambda^2} + \Gamma^i_{rr} = \Gamma^i_{rr} = 0$$

in this “sloppy notation.”

□

A parametrized curve whose tangent vector is parallel transported along the curve is called a *geodesic*. The straight lines in \mathbb{R}^n are geodesics of the Euclidean metric on that space. The same “geodesic equations” characterize the geodesics for any metric

$$\frac{D^2 x^i}{d\lambda^2} = \frac{d^2 x^i}{d\lambda^2} + \Gamma^i_{kj} \frac{dx^k}{d\lambda} \frac{dx^j}{d\lambda} = 0.$$

Exercise 8.3.1.**covariant geodesic equation**

a) One can express the geodesic equation in terms of the contravariant or covariant tangent vector $u^i = dx^i/d\lambda$. Show that

$$\frac{du^i}{d\lambda} + \Gamma^i_{jk} u^j u^k = 0, \quad \frac{du_i}{d\lambda} - \Gamma_{jki} u^j u^k = 0.$$

b) Show that in a coordinate frame one has the simpler formula $\Gamma_{(jk)i} = \frac{1}{2}g_{jk,i}$ so that

$$\frac{du_i}{d\lambda} - g_{jk,i} u^j u^k = 0,$$

which can be written

$$\frac{d}{d\lambda} \left(\frac{\partial}{\partial u^i} \left(\frac{1}{2} g_{jk} u^j u^k \right) \right) - \frac{\partial}{\partial x^i} \left(\frac{1}{2} g_{jk} u^j u^k \right).$$

If we introduce the kinetic energy function $T(x, u) = \frac{1}{2} g_{jk} u^j u^k$ as an explicit function of x^i and u^i so we can use partial derivatives with respect to these two variables, this becomes

$$\frac{d}{d\lambda} \left(\frac{\partial T(x, u)}{\partial u^i} \right) - \frac{\partial T(x, u)}{\partial x^i} = 0$$

■

Exercise 8.3.2.**geodesic coordinate lines**

Using the cylindrical coordinate expressions for the components of the covariant derivative, verify that the ρ and z coordinate lines are geodesics (i.e., straight lines).

■

A “curve” is a set of points with no parametrization, described geometrically or as a solution of a set of equations among the coordinate functions, like our example circles: $x^2 + y^2 = r_0^2, z = r_0 \cos \theta_0$. A curve is a geodesic if it admits a parametrization which is a geodesic as a parametrized curve according to our previous definition. Solving the geodesic conditions

led to straight lines in space parametrized *linearly* in Cartesian coordinates, but nonlinear parametrizations can also be used.

Suppose $c(\lambda)$ is a parametrized curve whose tangent vector $c'(\lambda)$ is parallel transported along the curve, namely

$$\frac{Dc'(\lambda)}{d\lambda} = 0,$$

and suppose we consider a new parametrization

$$\mathcal{C}(\lambda) = c(f(\lambda))$$

of the curve, where $f(\lambda)$ is a real valued function of λ . Then

$$\mathcal{C}'(\lambda) = f'(\lambda)c'(f(\lambda))$$

is the new tangent vector, multiplied by the function $f'(\lambda)$ according to the chain rule. Its covariant derivative along the parametrized curve $\mathcal{C}(\lambda)$ is

$$\frac{D\mathcal{C}^{i'}(\lambda)}{d\lambda} = \underbrace{\frac{d\mathcal{C}^{i'}(\lambda)}{d\lambda}}_{(f'(\lambda)c^{i'}(f(\lambda)))' = f''(\lambda) + (f'(\lambda))^2 c^{i''}(f(\lambda))} + \Gamma_{kj}^i \circ \mathcal{C}(\lambda) \underbrace{\mathcal{C}^{k'}(\lambda)\mathcal{C}^{j'}(\lambda)}_{f'(\lambda)^2 c^{k'}(\lambda)c^{j'}(\lambda)}$$

so

$$\begin{aligned} \frac{D\mathcal{C}^{i'}(\lambda)}{d\lambda} &= [f'(\lambda)]^2 \underbrace{[c^{i''}(f(\lambda)) + \Gamma_{kj}^i c^{k'}(\lambda)c^{j'}(f(\lambda))]}_{\left[\frac{Dc^{i'}}{d\lambda}\right](f(\lambda)) = 0} + f''(\lambda) c^{i'}(f(\lambda)) \\ &= f''(\lambda) \frac{\mathcal{C}^{i'}(\lambda)}{f'(\lambda)} = \left(\frac{f''(\lambda)}{f'(\lambda)}\right) \mathcal{C}^{i'}(\lambda). \end{aligned}$$

Instead of the covariant derivative of the tangent vector being zero, it is instead proportional to the tangent vector. In other words, the tangent vector to the curve still “follows its nose” by maintaining its direction parallel to the curve, but also changes its length as it moves along the curve following the arbitrary parametrization. Thinking of this as the path of a point particle in \mathbb{R}^n as a function of the time λ , where the tangent vector has the interpretation of the velocity of the particle in motion, instead of having constant velocity, it has variable velocity tracing out the same path in space. The path is still the same, i.e., it is independent of the way it is traced out in time.

A parametrization of a geodesic curve for which

$$\frac{Dc'(\lambda)}{d\lambda} = 0$$

is called an *affine parametrization* of the geodesic. For such a parametrization, all three factors in the inner product $g(c'(\lambda), c'(\lambda))$ are covariant constant along the curve

$$\begin{aligned} \frac{D}{d\lambda} |c'(\lambda)|^2 &= \frac{D}{d\lambda} (g_{ij} c^{i'} c^{j'}) \\ &= \left(\frac{Dg_{ij}}{d\lambda}\right) c^{i'} c^{j'} + g_{ij} \frac{Dc^{i'}}{d\lambda} c^{j'} + g_{ij} c^{i'} \frac{Dc^{j'}}{d\lambda} = 0 \end{aligned}$$

so the length $|c'(\lambda)|$ of the tangent vector is constant

$$\frac{D}{d\lambda}|c'(\lambda)| = \frac{D}{d\lambda}(|c'(\lambda)|^2)^{1/2} = \frac{1}{2}(|c'(\lambda)|^2)^{-1/2} \frac{D}{d\lambda}(|c'(\lambda)|^2) = 0.$$

We already know this since parallel transport preserves length

$$\frac{D}{d\lambda}|c'(\lambda)| = \frac{d}{d\lambda}|c'(\lambda)| = 0.$$

If we introduce the arclength parametrization as one does in multivariable calculus

$$\frac{ds}{d\lambda} = |c'(\lambda)|,$$

this just says

$$\frac{d^2s}{d\lambda^2} = \frac{d}{d\lambda}|c'(\lambda)| = 0,$$

i.e., the arclength and the parameter λ are linear related. If we use the arclength itself as a parameter we must have

$$\frac{ds}{d\lambda} = |c'(\lambda)| = 1,$$

so the tangent vector is a unit vector, as we recall from multivariable calculus. Given a particular arclength parametrization of the curve by an arclength s measured from some reference point on the geodesic, any affine parametrization must be linearly related: $\lambda = as + b$. The constant factor a rescales the arclength (or reverses it if negative) while the term b changes the zero of the arclength function along the curve, i.e., changes the reference point from which the arclength is measured.

It is convenient to represent this constant scale factor as $a = (2\mathcal{E})^{1/2}$ as will be explained below, namely

$$\frac{d\lambda}{ds} = (2\mathcal{E})^{1/2}.$$

The condition of constancy of the square of the length of the tangent vector for an affinely parametrized geodesic then takes the sloppy notation form

$$|c'(\lambda)|^2 = g_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} = 2\mathcal{E} \quad \text{or} \quad \frac{1}{2}|c'(\lambda)|^2 = \frac{1}{2}g_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} = \mathcal{E}.$$

Choosing initial data for the system of differential equations for which the constant $2\mathcal{E} = 1$ corresponds to choosing an arclength parametrization.

While our point of departure for this discussion is \mathbb{R}^n where the geodesics are just straight lines, nothing prevents us from applying this discussion to curved spaces with a given metric and associated covariant derivative. However, whether we are using general coordinates on \mathbb{R}^n whose associated frame vector fields are not covariant constant, or whether we are working with any coordinates on a curved space with a metric and its associated covariant derivative, the calculations are essentially the same. In each case we have to solve nontrivial systems of

ordinary differential equations to determine the geodesics in that coordinate system. To be able to represent them explicitly in the given coordinate system, we must get explicit solutions of those systems of differential equations, which will only be possible for such systems which are simple enough that solutions can be found in closed form. Otherwise we have to solve them by numerical integration, in which case the choice of coordinate system is not really a consideration.

Conserved momentum, symmetries and Killing vector fields

Quantities which are constant along a parametrized curve are said to be “conserved” in the language of physics when applied to curves in space as a function of time. Conservation laws of energy and momentum are key to solving many physical problems.

Suppose ξ is a Killing vector field of a metric g_{ij} , therefore satisfying $\xi_{(i;j)} = 0$, and consider the intrinsic derivative of its inner product with the tangent $u^i = dx^i/d\lambda$ to an affine parametrized geodesic $c(\lambda)$, for which $Du^i/d\lambda = 0$. Then

$$\frac{D}{d\lambda}(\xi_i u^i) = \frac{D\xi_i}{d\lambda} u^i + \xi_i \frac{Du^i}{d\lambda} = \xi_{i;j} u^j u^i = \xi_{(i;j)} u^j u^i = 0.$$

The quantity

$$P(\xi) = \xi_i u^i = \xi_i \frac{dx^i}{d\lambda}$$

is referred to as the conserved momentum associated with the Killing vector ξ .

For Euclidean space \mathbb{R}^3 and Cartesian coordinates x^i , the translation Killing vector fields are $p_i \partial_i$ while the rotation Killing vector fields are $L_i = \epsilon_{ijk} x^j \partial_k$. Then

$$P(p_i) = u_i = \delta_{ik} \frac{dx^k}{d\lambda} = \frac{dx^i}{d\lambda}$$

are the components of linear momentum which agree with our idea of momentum as the product of the mass and the velocity if we choose a parametrization where $\lambda = t/m$ and t is the time. Similarly

$$P(L_i) = \epsilon_{ijk} x^j u^k = \epsilon_{ijk} x^j \frac{dx^k}{d\lambda} = [\vec{x} \times \vec{P}(p_i)]_i.$$

Finally our use of the word momentum operators for the vector field generators of the symmetries of Euclidean space connects with what we know from high school physics.

A common situation is that a Killing vector field arises from the circumstance that the metric components are independent of a particular coordinate, say the first one

$$\xi = \partial_1, \quad g_{ij,1} = 0 = \xi^i_{;j} = 0 \quad \rightarrow \quad \mathcal{L}_\xi g_{ij} = 0.$$

Then the corresponding momentum is just the covariant component of the tangent vector along this coordinate

$$P(\xi) = \partial_1 \cdot u = \delta^i_1 g_{ij} u^j = g_{1j} u^j = u_1 = g_{1j} \frac{dx^j}{d\lambda}.$$

Suppose instead of geodesic equations we have equations for the parametrized curve of the form

$$\frac{Du^i}{d\lambda} = F^i,$$

where $F = F^i \partial_i$ is a force field. Then the above relation becomes

$$\frac{D}{d\lambda}(\xi_i u^i) = \xi_i F^i,$$

which means the component of momentum along the Killing vector field is still conserved (constant along the curve) if the force field is orthogonal to the Killing vector field $F^i \xi_i = 0$. If the force field arises as the gradient of a potential function

$$F^\flat = -dU \leftrightarrow F_i = -U_{,i} \leftrightarrow F^i = -g^{ij} U_{,j},$$

then

$$\frac{D}{d\lambda}(\xi_i u^i) = \xi_i F^i = \xi^i F_i = -\xi^i U_{,i} = -\xi U = -\mathcal{L}_\xi U.$$

If the potential is invariant under the 1-parameter family of transformations generated by ξ , then this is zero and the momentum along ξ is conserved.

We can also evaluate how the length of the tangent vector changes

$$\frac{D}{d\lambda}(u_i u^i) = 2 \frac{Du^i}{d\lambda} u_i = 2u_i F^i,$$

which requires that the tangent be orthogonal to the force field to be conserved.

Exercise 8.3.3.

Lorentz force

For the Lorentz force law on a particle of mass m and charge q in an electromagnetic field F along a world line in Minkowski spacetime parametrized by its proper time

$$\frac{Du^i}{d\tau} = \frac{q}{m} F^i{}_j u^j,$$

the antisymmetry of the electromagnetic field guarantees that u remains a unit vector since $F_{ij} u^i u^j = 0$. However, the condition that we have conserved “momenta” arising from Killing vectors ξ of the Lorentz metric requires further that

$$\xi_i F^i{}_j u^j = 0.$$

a) Suppose we have a uniform electric field along the z -direction so that only $F^0{}_1 = E_1$ is nonzero. Show that the linear momentum components $P(p_2), P(p_3)$ are conserved, but not $E = -P(p_0)$ unless the motion is orthogonal to the electric field. Show that the component $P(L_3)$ of angular momentum is conserved.

b) Suppose we have a uniform magnetic field along the z -direction so that only $F^1{}_2$ is nonzero. Show that the linear momentum components $E = -P(p_0), P(p_3)$ are conserved. ■

8.4 Surfaces of revolution

As already noted in Section 6.5 and as briefly discussed in Appendix D, for any parametrized surface in \mathbb{R}^3 we can evaluate the metric on the surface by simply substituting the parametrization for the Cartesian coordinates into the Euclidean metric $g = \delta_{ij} dx^i \otimes dx^j$ tensor expression and expanding. In slightly different notation and language, the same result can be accomplished by substituting the parametrization for the Cartesian coordinates into the Euclidean squared differential of arclength $ds^2 = \delta_{ij} dx^i dx^j$, called the “line element,” yielding the line element the surface. This is exactly what we do to get the differential of arclength on a parametrized curve. Sloppily the line element is also referred to as the metric, since it contains the same information as the metric. Given this metric or line element on the surface, we can then evaluate the components of the connection and then the geodesic equations to study the geodesics of the intrinsic geometry of the surface. The easiest class of such surfaces to investigate are surfaces of revolution and indeed many interesting surfaces we commonly deal with are surfaces of revolution, including the flat plane, the cylinder and the sphere, as well as the hyperboloids or pseudospheres of 3-dimensional Minkowski spacetime. The fact that such surfaces have a 1-parameter group of symmetries makes it easier to derive some analytic results about their geometry. This common feature also applies to the screw-symmetric surfaces where the rotation symmetry is replaced by a screw-rotation symmetry. We will consider this generalization next.

Suppose we take a surface of revolution about the vertical axis in \mathbb{R}^3 , starting from some plane curve in any plane passing through the vertical axis and revolving it completely around this axis to sweep out the surface. This surface is most simply expressed in cylindrical coordinates with $z = Z(\rho)$ or $\rho = R(z)$ or even $(\rho, z) = (R(u), Z(u))$ describing the plane curve in the ρ - z plane used in the construction as an arbitrary parametrized curve, the most general situation, leading to the most general case of a parametrized surface in which both ρ and z are functions of a third variable

$$\begin{aligned} (i) \quad & \langle x, y, z \rangle = \vec{r}(\rho, \phi) = \langle \rho \cos \phi, \rho \sin \phi, Z(\rho) \rangle \\ (ii) \quad & \langle x, y, z \rangle = \vec{r}(z, \phi) = \langle R(z) \cos \phi, R(z) \sin \phi, z \rangle \\ (iii) \quad & \langle x, y, z \rangle = \vec{r}(u, \phi) = \langle R(u) \cos \phi, R(u) \sin \phi, Z(u) \rangle. \end{aligned}$$

We keep the azimuthal coordinate ϕ as one of the two parameters of the parametrized surface since it is adapted to the rotational symmetry, which is realized as translations in that variable. The azimuthal coordinate derivative is a Killing vector field generating those symmetries, so the component of an affinely parametrized geodesic on the surface along this Killing vector field, namely the covariant ϕ coordinate component of the tangent vector, is a constant along the geodesic.

The cylinder and sphere are surfaces of revolution and some of the terminology that is natural for a sphere can be extended to general surfaces of revolution. In terms of the general surface parametrization by (u, ϕ) , which serve as coordinates on the surface, the coordinate grid is automatically orthogonal, consisting of horizontal plane circles of revolution for the ϕ coordinate lines naturally called the “parallels” (lines of latitude on the sphere) and the u coordinate lines consisting of the vertical plane cross-sectional curves at constant ϕ , naturally

called the “meridians.” The radius $R(u)$ of the parallels, or ρ itself when $u = \rho$, plays an important role in the geometry of the surface.

By symmetry the meridians are geodesics since clearly the tangent to a meridian cannot rotate to one direction or the other to stay self-parallel within the surface: the surface is reflection-symmetric about the plane containing the meridian. It takes a little bit more convincing to see that parallels for which the azimuthal radius $R(u)$ has an extremum as a function of u are also geodesics, like the equator of a sphere whose symmetry axis coincides with the z -axis. We will confirm this explicitly with the geodesic equations below.

Exercise 8.4.1.

metric for a surface of revolution

a) For a surface of revolution in these adapted parametrizations, evaluate the differentials of the Cartesian coordinates $d\vec{r} = (\partial\vec{r}/\partial u^i) du^i$ in terms of the two parameters $(u^1, u^2) = (\rho, \phi), (\phi, z), (u, v)$ respectively and substitute the results into the Euclidean metric $g = dx \otimes dx + dy \otimes dy + dz \otimes dz$. Then expand the products and collect terms to show that the metric takes the form

$$\begin{aligned} g &\stackrel{(i)}{=} (1 + Z'(\rho)^2) d\rho \otimes d\rho + \rho^2 d\phi \otimes d\phi \\ &\stackrel{(ii)}{=} (1 + R'(\rho)^2) dz \otimes dz + R(\rho)^2 d\phi \otimes d\phi \\ &\stackrel{(iii)}{=} (R'(u)^2 + Z'(u)^2) du \otimes du + R(u)^2 d\phi \otimes d\phi = g_{uu} du \otimes du + g_{\theta\theta} d\theta \otimes d\theta. \end{aligned}$$

b) By introducing an arclength coordinate r along the meridians in the vertical half-planes of constant ϕ

$$\begin{aligned} dr &\stackrel{(i)}{=} (1 + Z'(\rho)^2)^{1/2} d\rho \\ &\stackrel{(ii)}{=} (1 + R'(z)^2)^{1/2} dz \\ &\stackrel{(iii)}{=} (R'(u)^2 + Z'(u)^2)^{1/2} du = g_{uu}^{1/2} du \end{aligned}$$

and letting $\theta = \phi$ to compare this to polar coordinates in the flat plane, one finds the common form

$$g = dr \otimes dr + R(r)^2 d\theta \otimes d\theta,$$

provided one can re-express $\rho = R(r)$ in terms of the new coordinate r . Verify this. The final result makes perfect sense since the r and θ coordinate lines on the surface are orthogonal and one needs the “azimuthal radius” ρ of the θ coordinate circle to convert increments of the azimuthal angle into an arclength: $R(r)d\theta = ds_\theta$.

c) For the nonzero metric components $g_{rr} = 1, g_{\theta\theta} = R(r)^2$, show that the only nonvanishing components of the covariant derivative are

$$\Gamma^r_{\theta\theta} = -R'(r)R(r), \quad \Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} = \frac{R'(r)}{R(r)}.$$

d) Show that the affinely parametrized geodesic equations are

$$\begin{aligned}\frac{d^2 r}{d\lambda^2} + \Gamma^r_{\theta\theta} \left(\frac{d\theta}{d\lambda} \right)^2 &= \frac{d^2 r}{d\lambda^2} - R'(r)R(r) \left(\frac{d\theta}{d\lambda} \right)^2 = 0, \\ \frac{d^2 \theta}{d\lambda^2} + 2\Gamma^{\theta}_{r\theta} \frac{dr}{d\lambda} \frac{d\theta}{d\lambda} &= \frac{d^2 \theta}{d\lambda^2} + 2 \frac{R'(r)}{R(r)} \frac{dr}{d\lambda} \frac{d\theta}{d\lambda} \\ &= R(r)^{-2} \frac{d}{d\lambda} \left(R(r)^2 \frac{d\theta}{d\lambda} \right) = 0.\end{aligned}$$

Verify the final equality, which shows that the quantity $\ell = R(r)^2 d\theta/d\lambda$ is a constant along the geodesic. The next section will interpret this as a constant of the motion for the moving particle approach to interpreting parametrized geodesic curves.

Note that if we eliminate $d\theta/d\lambda$ from the r equation using the constancy of ℓ we find

$$\frac{d^2 r}{d\lambda^2} = R'(r)R(r) \left(\frac{d\theta}{d\lambda} \right)^2 = \frac{\ell^2 R'(r)}{R(r)^3} = -\frac{d}{dr} \left(\frac{\ell^2}{2R(r)^2} \right).$$

This allows us to think of the quantity in parentheses as a potential for the effective radial force that causes that coordinate to be accelerated. We will develop this further in the next section.

e) Now consider an extremal point on a meridian: $R'(r_0) = 0$. Show that geodesic equations are both easily satisfied for the parallel $r = r_0, \theta = \theta_0 + \lambda$, which has $\theta''(\lambda) = 0$.

f) When $g_{rr} \neq 1$, show that only the first geodesic equation changes by an additional term

$$\frac{d^2 r}{d\lambda^2} - \frac{R'(r)R(r)}{g_{rr}(r)} \left(\frac{d\theta}{d\lambda} \right)^2 + \frac{1}{2} (\ln g_{rr}(r))' \left(\frac{dr}{d\lambda} \right)^2 = 0,$$

or equivalently

$$g_{rr}(r)^{1/2} \frac{d}{d\lambda} \left(g_{rr}(r)^{1/2} \frac{dr}{d\lambda} \right) - R'(r)R(r) \left(\frac{d\theta}{d\lambda} \right)^2 = 0.$$

■

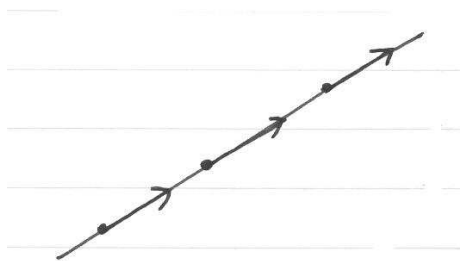


Figure 8.9: A “straight line” tangent vector follows its nose so that it maintains its fixed direction.

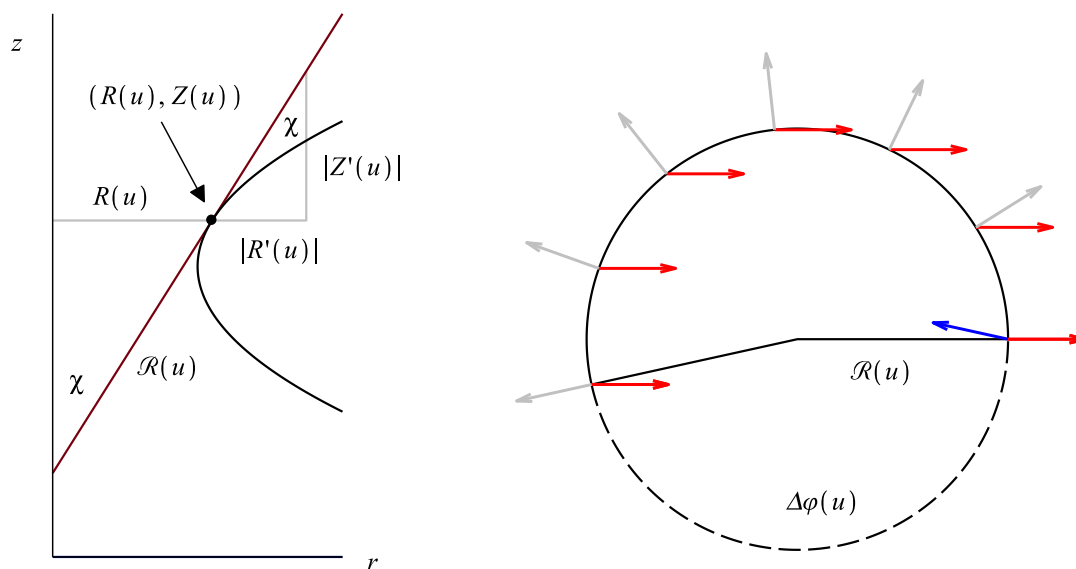


Figure 8.10: **Left:** The parametrized meridian is rotated around the z -axis to form a surface of revolution. The tangent line to this cross-section rotates into a cone tangent to the point of tangency of the tangent line, with vertex on the z -axis. **Right:** Opening up the tangent cone to a flat plane, leads to a defect angle $\Delta\phi$. A constant vector initially along the radial direction on the horizontal axis rotates backwards with respect to the radial direction as we follow the circle around counterclockwise but when one folds the flattened cone back into its original shape, by joining the two edges shown in the figure, this vector has advanced counterclockwise by exactly the defect angle with respect to the initial radial direction.

Exercise 8.4.2.**tangent cone to surface of revolution**

The tangent line to a cross-sectional meridian at a given parallel rotates into a tangent cone to the surface of revolution there. The tangent cone has the same parallel transport properties along the parallel in the flat geometry of the cone as in the curved geometry of the surface of revolution, so we can easily compute the rotation of a tangent vector along such a parallel under parallel transport without solving any transport differential equations.

a) From Fig. 8.10 (left) we can determine the hypotenuse $\mathcal{R}(u)$ of the conical right triangle shown, with χ as the opening angle of the cone, by using similar triangles with the right triangle whose hypotenuse is instead the tangent vector to the parametrized meridian tangent line: $\tan \chi = |R'(u)/Z'(u)|$. One then has determined the vertical leg $\mathcal{Z}(u)$ of this triangle and the opening angle $\chi = \arcsin(R(u)/\mathcal{R}(u))$ of the cone. Show that

$$\sin \chi = \frac{|R'(u)|}{(R'(u)^2 + Z'(u)^2)^{1/2}}, \quad \mathcal{R}(u) = \frac{R(u)}{\sin \chi} = \frac{g_{uu}^{1/2}}{|(\ln R(u))'|} \equiv \frac{1}{\kappa(u)},$$

where the significance of the curvature

$$\kappa(u) = \frac{|(\ln R(u))'|}{g_{uu}^{1/2}} \geq 0$$

of the azimuthal circle will become clear below, making its reciprocal

$$\mathcal{R}(u) = \frac{1}{\kappa(u)} \geq R(u)$$

the radius of curvature of the circle, necessarily larger than the actual radius of the circle within \mathbb{R}^3 due to the tilting of the plane of that circle relative to the tangent plane to the surface of revolution which stretches out the circle of best fit in the tangent plane compared to the actual circle. Note that the absolute value sign is necessary when $R'(u) < 0$ as occurs in the lower part of the profile curve of Fig. 8.10 if u increases in the upwards direction along it. In general the ratio of these two radii (circumferential radius to intrinsic radius) is

$$\frac{R(u)}{\mathcal{R}(u)} = \sin \chi = \frac{|R'(u)|}{g_{uu}^{1/2}} \leq 1.$$

This defines the opening angle of the tangent cone in terms of the intrinsic metric quantities.

b) From Fig. 8.10 (right) show that the defect angle of the cone is given by

$$\frac{\Delta\varphi}{2\pi} = 1 - \frac{R(u)}{\mathcal{R}(u)} = 1 - \sin \chi.$$

This is the total angle forward that a constant vector in the plane of the flattened out cone rotates during one revolution of the azimuthal circle. This gives the angle a vector is rotated during parallel transport around this circle.

Notice that for case of a horizontal plane where $Z(r) = 0$, $R(r) = r$, one finds $\mathcal{R}(r) = R(r)$ so that $\sin \chi = 1$ and $\Delta\varphi = 0$. In the flat plane, no rotation takes place under parallel transport around a closed loop. On the other hand at an extremum of the azimuthal radius $R'(u) = 0$, then the curvature $\kappa(u)$ vanishes and the tangent cone opens up into a cylinder, along which parallel transport is simple: horizontal vectors stay horizontal, vertical vectors stay vertical. In particular the unit tangent to the azimuthal circle remains horizontal, so it transported parallel to itself, making such a circle a geodesic as already noted.

c) Introduce the orthonormal frame and evaluate their Lie bracket commutator

$$e_{\hat{u}} = \frac{1}{g_{uu}^{1/2}} \frac{\partial}{\partial u}, \quad e_{\hat{\theta}} = \frac{1}{R(u)} \frac{\partial}{\partial \theta},$$

$$[e_{\hat{u}}, e_{\hat{\theta}}] = -\operatorname{sgn}(R'(u)) \kappa(u) e_{\hat{\theta}} \quad \rightarrow \quad C^{\hat{\theta}}_{\hat{u}\hat{\theta}} = -\operatorname{sgn}(R'(u)) \kappa(u),$$

and then use the frame formula for the connection components to verify that the only nonzero components of the connection in this orthonormal frame are

$$\Gamma^{\hat{u}}_{\hat{\theta}\hat{\theta}} = -\Gamma^{\hat{\theta}}_{\hat{\theta}\hat{u}} = C^{\hat{\theta}}_{\hat{u}\hat{\theta}}.$$

Show that letting $\epsilon = \operatorname{sgn}(R'(u))$, these translate into the relations

$$\nabla_{e_{\hat{\theta}}} e_{\hat{\theta}} = -\epsilon \kappa(u) e_{\hat{u}}, \quad \nabla_{e_{\hat{\theta}}} e_{\hat{u}} = \epsilon \kappa(u) e_{\hat{\theta}}.$$

d) Verify that the orthonormal frame connection 1-form is

$$\hat{\omega} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \epsilon \kappa(u) \omega^{\hat{\theta}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \epsilon \kappa(u) R(u) d\theta = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \frac{R'(u)}{g_{uu}^{1/2}} d\theta.$$

e) Since $\hat{T} = e_{\hat{\theta}}$ is the unit tangent along the azimuthal circles, we can re-interpret $\nabla_{e_{\hat{\theta}}} = D/ds$ as the arclength derivative of fields along those circles, enabling us to rewrite these covariant derivative equations in the following form

$$\frac{De_{\hat{\theta}}}{ds} = -\epsilon \kappa(u) e_{\hat{u}}, \quad \frac{De_{\hat{u}}}{ds} = \epsilon \kappa(u) e_{\hat{\theta}},$$

or since $ds = R(u) d\theta$ along these circles,

$$\frac{De_{\hat{\theta}}}{d\theta} = -\frac{d\Phi}{d\theta} e_{\hat{u}}, \quad \frac{De_{\hat{u}}}{d\theta} = \frac{d\Phi}{d\theta} e_{\hat{\theta}}, \quad \frac{d\Phi}{d\theta} = \epsilon \kappa(u) R(u) = \frac{R'(u)}{g_{uu}^{1/2}}.$$

This allows us to identify $\hat{N} = -\epsilon e_{\hat{u}}$ as the unit normal within the surface and therefore $\kappa(u)$ is the covariant curvature of these circles (the length of the arclength derivative of the unit tangent vector, exactly as in multivariable calculus except for the use of the covariant derivative here, see Appendix C) and determines the angular rate of rotation along the curve (with respect to arclength) of the orthonormal frame vectors within the intrinsic surface geometry. Its associated radius of curvature $\mathcal{R}(u)$ determines the osculating circle in the tangent plane in the same way,

a distance $\mathcal{R}(u)$ along the normal. What is interesting here is that the center of the osculating circle is the vertex of the tangent cone. Of course in the extrinsic geometry of the surface within the enveloping space R^3 , the radius of curvature of the azimuthal circle is just its actual radius $R(u)$. This points to an important difference between the intrinsic and extrinsic geometry of surfaces. We will study the extrinsic geometry later on.

f) We cannot leave these calculations without finishing them off. Show that a general parallel transported vector in the surface is described by

$$\begin{pmatrix} X^{\hat{u}} \\ X^{\hat{\theta}} \end{pmatrix} = \begin{pmatrix} \cos \Phi & \sin \Phi \\ -\sin \Phi & \cos \Phi \end{pmatrix} \begin{pmatrix} X_0^{\hat{u}} \\ X_0^{\hat{\theta}} \end{pmatrix} = \begin{pmatrix} \cos(\Omega \theta) & \sin(\Omega \theta) \\ -\sin(\Omega \theta) & \cos(\Omega \theta) \end{pmatrix} \begin{pmatrix} X_0^{\hat{u}} \\ X_0^{\hat{\theta}} \end{pmatrix}$$

where the “theta angular velocity” Ω is given by

$$\Omega = \frac{d\Phi}{d\theta} = \epsilon \kappa(u) R(u) = \epsilon \frac{R(u)}{\mathcal{R}(u)} = \epsilon \sin \chi = \frac{R'(u)}{g_{uu}^{1/2}},$$

where each of its different representations is useful in different contexts. Notice that this rotation of the components of the vector corresponds to an active rotation by the angle $-\Phi$ corresponding to a theta angular velocity $-\Omega$ compared to the active rotation of the orthonormal frame by the angle Φ with theta angular velocity Ω relative to parallel transported axes. ■

Remark.

parallel transport around symmetry circles

The failure of parallel transport around a parallel of a surface of revolution to return a vector to its original direction has a simple interpretation—there is a mismatch between the osculating circle in the enveloping geometry of \mathbb{R}^3 and the intrinsic osculating circle which can only be defined in the tangent plane to the point of the circle, and which is necessarily bigger due to the stretching which occurs as one tilts the original horizontal circle to align it with the tangent plane and yet still fit the curvature of the the original circle. Vectors must rotate backwards by the angle one moves forward around the circle to stay parallel to their original direction, but it is the radius of curvature which determines this backwards rotation angle through the usual relation of angle to arclength $\Phi = s/\mathcal{R}$. The circumference of the closed circular parallel is instead determined by its actual radius $R(u)$ in space as $s = 2\pi R(u)$ so the final backwards rotation angle after one revolution is $2\pi R(u)/\mathcal{R}(u) \leq 2\pi$. The amount by which it comes up short is

$$\Delta\varphi = 2\pi - 2\pi \frac{R(u)}{\mathcal{R}(u)} = 2\pi \left(1 - \frac{R(u)}{\mathcal{R}(u)} \right).$$

But this is exactly the amount by which the vector moves forward with respect to its original direction. This is illustrated for an azimuthal coordinate circle on a sphere in Fig. 8.22. The previous discussion may be interpreted in terms of the osculating circle in the tangent plane rolling without slipping around the azimuthal circle. ▲

Exercise 8.4.3.**planes, cylinders and cones**

The simplest surfaces of revolution come from rotating a straight line in a plane containing the vertical axis of rotation. If the line is also vertical we get a cylinder with $R(z) = R_0$ (case (ii)), but if the line is horizontal $Z(\rho) = Z_0$ (case (i)) we get an ordinary plane, while if the line is oblique (neither vertical nor horizontal), we get a cone. Using the geometry of the tangent cone of Fig. 8.10, this corresponds to $Z(\rho) = \rho \cot \chi$ (case (i)). To compare with polar coordinates in the plane, let us revert back to the notation $(\rho, \phi) \rightarrow (r, \theta)$.

a) Use Exercise 8.4.1 to evaluate the metric in each of these cases. Of course in the second case we just get the usual polar coordinate expression in the plane $ds^2 = dr^2 + r^2 d\theta^2$, while in the second case we can introduce the rescaled coordinate $X = R_0 \theta$ to get orthonormal coordinates (X, z) on that intrinsically flat surface. For the conical case show that we get

$$ds^2 = \csc^2 \chi dr^2 + r^2 d\theta^2,$$

where χ is the opening angle of the cone.

b) Introduce an arclength radial coordinate by $\tilde{r} = \csc \chi r$ and show that the resulting form of the metric is

$$ds^2 = d\tilde{r}^2 + \tilde{r}^2 \sin^2 \chi d\theta^2 \rightarrow R(\tilde{r}) = \tilde{r} \sin \chi.$$

c) Use Exercise 8.4.2 to evaluate the curvature and radius of curvature of the parallels. Show that you get

$$\mathcal{R}(\tilde{r}) = \tilde{r} = \frac{r}{\sin \chi}.$$

This shows the relationship between the intrinsic radius of curvature $\mathcal{R}(\tilde{r})$ of the parallel circle and its actual radius $R(\tilde{r}) = r$.

■

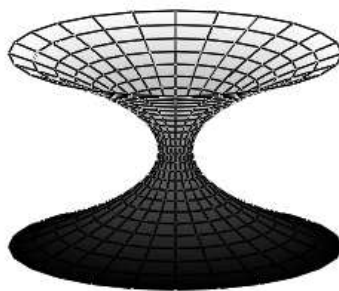


Figure 8.11: A parabola of revolution but revolved about an axis perpendicular to the symmetry axis, and on the convex side of the vertex of that parabola, which is the classic “wormhole surface.”

Exercise 8.4.4.**black hole embedding surfaces**

An interesting 2-surface geometry that can be realized as a surface in R^3 is the curved equatorial plane in a nonrotating black hole spacetime. In spherical coordinates restricted to the plane $\theta = \pi/2$ just as in ordinary Euclidean space spherical coordinates, again using θ for the azimuthal angle to compare with the situation in polar coordinates in the flat plane, the metric is

$$ds^2 = \frac{1}{1 - \frac{2m}{r}} dr^2 + r^2 d\theta^2, \quad m > 0.$$

The radial coordinate r is no longer an arclength coordinate but is tied to the circumference of azimuthal circles by the usual flat space one of the plane in polar coordinates: $C = 2\pi r$. This explicit metric is all that is needed to investigate the geodesics in its intrinsic geometry, but it is pretty easy to create a surface of revolution in R^3 which has this geometry, allowing us to visualize better its properties.

a) Introduce an arclength coordinate r^* measuring arclength from the “horizon” $r = 2m$ by solving the initial value problem using a computer algebra system

$$\frac{dr^*}{dr} = \frac{1}{\left(1 - \frac{2m}{r}\right)}^{1/2}, \quad r^*(2m) = 0, \quad r > 2m.$$

The expression you find clearly cannot be inverted to express r as a function of r^* , so we cannot explicitly introduce an arclength radial coordinate for this surface of revolution.

b) Suppose we let (r, θ, z) be cylindrical coordinates in R^3 and look for a surface graph $z = Z(r)$ such that substituting this into the metric allows us to identify the result with the above expression

$$ds^2 = dr^2 + r^2 d\theta^2 + dZ(r)^2 = (1 + Z'(r)^2) dr^2 + r^2 d\theta^2,$$

leading to the differential equation

$$1 + Z'(r)^2 = \frac{1}{1 - 2m/r} \rightarrow Z'(r) = \left(\frac{1}{1 - 2m/r} - 1 \right)^{1/2} = \left(\frac{2m}{r - 2m} \right)^{1/2},$$

choosing the positive root for the derivative. This is integrated by a trivial u -substitution for the antiderivative, discarding an additive constant (equivalent to the initial condition $Z(2m) = 0$) to yield

$$Z = \sqrt{8m(r - 2m)}.$$

Verify this.

Notice this can easily be extended to

$$Z^2 = 8m(r - 2m),$$

yielding the entire smooth surface of revolution whose cross-section is a parabola rotated from its usual position (graph of the square function) by 90 degrees. This is the “wormhole” surface of the popular literature. This is exactly the surface depicted in Figs. 8.10 and 8.11.

c) Use the results of Exercise 8.4.2 to show that the intrinsic radius of curvature of the azimuthal circle is

$$\mathcal{R}(r) = \frac{r}{(1 - 2m/r)^{1/2}}$$

and then evaluate the formula for the total rotation $\Delta\varphi$ under parallel transport after one azimuthal revolution found there. Finally consider the approximation for $m/r \ll 1$ by Taylor expanding your result to first order in this quantity to show that

$$\frac{\Delta\varphi}{2\pi} \approx \frac{m}{r}, \quad \frac{m}{r} \ll 1.$$

This is just 2/3 of the so called geodetic precession effect for a gyroscope in Earth orbit measured by the GP-B satellite experiment which took place in a satellite in a polar orbit $h = 650$ km above the Earth’s surface: $r = R_{\text{earth}} + h$. To evaluate this using the mass and radius of the Earth we need to put back in the factors of the speed of light c and Newton’s constant G which have been set to 1 in our geometrical units:

$$\frac{\Delta\varphi_{(\text{geodetic})}}{2\pi} \approx \frac{3Gm}{2c^2r}.$$

Note that this ratio has the same units as

$$\frac{Gmm_s/r}{m_sc^2} = \frac{-U_{\text{pot}}}{E_{\text{rest}}},$$

which is the dimensionless ratio of the potential energy of the satellite in the Earth’s Newtonian gravitational field to the satellite’s rest energy, so the reappearance of G and c which we had set to 1 makes sense without going through the detailed unit analysis.

This is the amount of precession angle that occurs during one revolution, i.e., one orbit of the Earth. To get the precession rate you have to divide this by the orbital time, called the period. The Newtonian speed of a circular orbit balances the gravitational and centrifugal forces (the satellite mass cancels out!), leading to its period

$$\frac{Gm}{r^2} = \frac{v^2}{r} \rightarrow v = \left(\frac{Gm}{r}\right)^{1/2} \rightarrow \Delta t = \frac{2\pi r}{v} = 2\pi \left(\frac{r^3}{Gm}\right)^{1/2}.$$

Thus

$$\frac{\Delta\varphi_{(\text{geodetic})}}{\Delta t} \approx \frac{3Gm}{2c^2r^2} \left(\frac{Gm}{r}\right)^{1/2}.$$

If you plug in the numbers you can easily find on the internet for Newton’s constant G , the speed of light c and the mass m of the Earth and convert the units of this value from radians per second to arcseconds per year, do you get the number 6.6 arcsec/yr that you can also find on the internet (although it takes a bit more work to locate)?

**Exercise 8.4.5.****parallel transport along circles**

To make the previous exercise more concrete, use a computer algebra system to make a plot similar to the one illustrating parallel transport along the lines of latitude (parallels) of the sphere suggested in Exercise 8.2.2 but now for the 2-dimensional situation of the parallel transport of the original frame at $\theta = 0$ around one revolution of the azimuthal coordinate circle, plotted on the r - θ coordinate plane. Show this using the exact formulas this the circles $r/m = 2, 3, 4, 5$ near the horizon $r = 2m$ of the black hole. First evaluate the theta angular velocity Ω exactly. What is the net precession angle $\Delta\varphi/(2\pi)$ after one revolution?

a) Then plot the two vector fields

$$\begin{pmatrix} e_{\hat{r}}^{(\parallel)} & e_{\hat{\theta}}^{(\parallel)} \end{pmatrix} = (e_{\hat{r}} \ e_{\hat{\theta}}) \begin{pmatrix} \cos(\Omega\theta) & \sin(\Omega\theta) \\ -\sin(\Omega\theta) & \cos(\Omega\theta) \end{pmatrix}, \quad 0 \leq \theta \leq 2\pi$$

in a ring of radius $r/m = 5$ together with the 4 circles $r/m = 2, 3, 4, 5$ (the first to indicate the horizon) but first divide the vector fields by 2 so they do not overlap with each other as they rotate. What seems to be happening at the limiting radius $r/m = 2$?

b) Just plot the initial and final vectors along the axis $\theta = 0$ for unit intervals from $r/m = 2$ to $r/m = 12$ to see how the precession angle after one revolution decreases with distance from the horizon. Repeat for 10 divisions of the interval $2 \leq r/m \leq 3$ dividing your vector fields by 20 so they do not overlap to study the limiting behavior near the horizon at $r/m = 2$.

c) In the embedding diagram, the circle $r = 2m$ is called the throat of the worm hole. According to our general considerations for a surface of revolution, it is an extremum of the distance from the axis of rotation, so it should be a geodesic. Does that explain this limiting behavior at $r/m = 2$?

d) If you are feeling ambitious, plot the vector fields of part a) on the embedding surface of revolution for $2 \leq r/m \leq 10$ at unit ring intervals.



8.5 Parametrized curves as “motion of point particles” and the geodesic motion approach

Tracing out a given parametrized curve is most naturally imagined as a process occurring in time. In fact in computer algebra systems, one can easily animate such curves in time using the parameter as a time variable with some constant rate converting the parameter intervals into time intervals. This visualization aid is very useful in understanding the behavior of geodesics and nongeodesic paths alike, so it is worth taking seriously. One imagines the parametrized curve as the motion of a point particle in the space under consideration, interpreting $\lambda = t$ as the time. The parametrized curve $c(\lambda)$ specifies the “position” $x^i(\lambda)$ of the particle at time λ as expressed in the given coordinate system, the tangent vector $c'(\lambda) = v(\lambda)$ is the “velocity” with coordinate components $v^i(\lambda) = x^{i'}(\lambda)$ and its magnitude or length is the speed $|v(\lambda)| = |c'(\lambda)|$.

In ordinary space \mathbb{R}^3 , if the point particle has mass m , Newton’s force law in Cartesian coordinates states that mass times acceleration equals applied force

$$m \frac{d^2 x^i}{d\lambda^2} = F^i$$

or dividing through by the mass, acceleration equals specific force, namely force per unit mass

$$\frac{d^2 x^i}{d\lambda^2} = \frac{F^i}{m}.$$

These are called the “equations of motion” for the particle. Since the ordinary derivative in Cartesian coordinates translates into the covariant derivative in any coordinate system, then in a general coordinate system this becomes

$$\frac{D^2 x^i}{d\lambda^2} = \frac{F^i}{m} = \mathcal{F}^i,$$

so the covariant derivative of the tangent vector along the curve $Dc'(\lambda)/d\lambda = Dv(\lambda)/d\lambda$ is the “acceleration” in this context and the vector field \mathcal{F} is the specific force, namely the force per unit mass. Since no mass is relevant to our particle motion analogy, all the usual physics quantities that involve mass as a factor will be divided through by the mass to become the specific such quantities, equivalent to setting the mass m equal to 1 everywhere.

A so-called conservative force field is one which can be represented as the gradient of a function $F^i = -U^{;i}$ or in terms of specific quantities $\mathcal{F}^i = -\mathcal{U}^{;i}$, leading to the equations of motion

$$\frac{D^2 x^i}{d\lambda^2} = -\mathcal{U}^{;i}.$$

One can introduce a total energy function consisting of a kinetic energy term (half the mass times the square of the speed) and a potential energy term, with the corresponding specific quantities in the same relation, introducing the specific energy $\mathcal{E} = E/m$

$$\begin{aligned} E &= K + U, & K &= \frac{1}{2} m g_{ij} x^{i'} x^{j'}, \\ \mathcal{E} &= \mathcal{K} + \mathcal{U}, & \mathcal{K} &= \frac{1}{2} g_{ij} x^{i'} x^{j'}. \end{aligned}$$

This is useful since the energy or specific energy are both constant along a geodesic (are “constants of the motion”) as a simple calculation shows, explaining as well the factor of $\frac{1}{2}$ in its definition. By working with specific quantities, no mention of mass is required, appropriate for purely geometric problems like solving geodesic equations of motion, while allowing useful intuition about particle motion to be brought into the discussion.

To evaluate the derivative of the energy function along a geodesic, one must use the chain rule to differentiate functions of position along it: $D/d\lambda = (dx^i/d\lambda)\nabla_i = x^{i'}\nabla_i$, which for scalars reduces to $Df/d\lambda = df/d\lambda = f_{,i}x^{i'} = f_{;i}x^{i'}$. Thus using the covariant constancy of the metric, one has

$$\begin{aligned}\frac{d\mathcal{E}}{d\lambda} &= \frac{D\mathcal{E}}{d\lambda} = \frac{1}{2}\frac{Dg_{ij}}{d\lambda}x^{i'}x^{j'} + \frac{1}{2}g_{ij}\frac{Dx^{i'}}{d\lambda}x^{j'} + \frac{1}{2}g_{ij}x^{i'}\frac{Dx^{j'}}{d\lambda} + \mathcal{U}_{,j}x^{j'} \\ &= g_{ij}\left(\frac{Dx^{i'}}{d\lambda} - \mathcal{F}^i\right)x^{j'} = g_{ij}\left(\frac{D^2x^i}{d\lambda^2} - \mathcal{F}^i\right)x^{j'} = 0.\end{aligned}$$

Furthermore, this can be rewritten to describe the rate of change of the speed

$$\frac{D}{d\lambda}(g_{ij}x^{i'}x^{j'})^{1/2} = \frac{1}{2}(g_{ij}x^{i'}x^{j'})^{-1/2}2\mathcal{F}_jx^{j'} = (g_{ij}x^{i'}x^{j'})^{-1/2}\mathcal{F}_jx^{j'},$$

which is zero only when the velocity is orthogonal to the force: $\mathcal{F}_jx^{j'} = 0$. Thus the speed is constant in the force-free case of geodesic motion, while the velocity is covariant constant.

In the case of ordinary space \mathbb{R}^3 , one more particle motion concept is needed: the angular momentum $\vec{L} = m\vec{r} \times \vec{v}$ or specific angular momentum $\vec{L}/m = \vec{r} \times \vec{v}$ in the notation of multivariable calculus. The component of specific angular momentum about the z -axis is simply expressed in cylindrical coordinates as $\ell = \rho v^{\hat{\phi}} = \rho^2 d\phi/d\lambda$. For motion in force fields which are rotationally symmetric about the z -axis, we know from high school physics that this angular momentum is a constant.

Remark.

In physics notation time derivatives are always denoted with an overdot instead of a prime superscript

$$\dot{f} = f'.$$

This would be useful for us here since then we would not have to look closely to distinguish between primed coordinate superscripts and a prime indicating a derivative of a component: $x^{i'}$ or $(x^i)' = x^{i'}$ but let's just stick with the more familiar prime derivative symbol. \blacktriangle

Exercise 8.5.1.

geodesics on a surface of revolution

a) Suppose we take a surface of revolution about the vertical axis in \mathbb{R}^3 and consider the

affinely parametrized geodesic equations derived above.

$$\begin{aligned}\frac{d^2 r}{d\lambda^2} + \Gamma^r_{\theta\theta} \left(\frac{d\theta}{d\lambda} \right)^2 &= \frac{d^2 r}{d\lambda^2} - R'(r)R(r) \left(\frac{d\theta}{d\lambda} \right)^2 = 0, \\ \frac{d^2 \theta}{d\lambda^2} + 2\Gamma^\theta_{r\theta} \frac{dr}{d\lambda} \frac{d\theta}{d\lambda} &= \frac{d^2 \theta}{d\lambda^2} + 2 \frac{R'(r)}{R(r)} \frac{dr}{d\lambda} \frac{d\theta}{d\lambda} \\ &= R(r)^{-2} \frac{d}{d\lambda} \left(R(r)^2 \frac{d\theta}{d\lambda} \right) = 0.\end{aligned}$$

From the angular equation we therefore conclude that the specific angular momentum $\ell = R(r)^2 d\theta/d\lambda$ is a constant for solutions of the geodesic equations. Substituting this result into the radial equation of motion leads to

$$\frac{d^2 r}{d\lambda^2} = \ell^2 \frac{R'(r)}{R(r)^3} = -\frac{dU(r)}{d\lambda} = -U(r)_{,r} \frac{dr}{d\lambda} = 0.$$

where

$$U(r) = \frac{\ell^2}{2R(r)^2}$$

is the so called “centrifugal potential,” whose sign-reversed derivative gives the specific force (centripetal acceleration) in the radial direction for radial motion due to the angular motion. This is an example of an “effective potential” not for a real applied force, but for an effective force due to another degree of freedom, in this case the angular motion.

b) The length of the tangent vector remains constant under parallel transport, so the specific energy $\mathcal{E} \geq 0$ is a constant for solutions of the geodesic equations

$$\frac{1}{2}|c'(\lambda)|^2 = \frac{1}{2} \left(\frac{dr}{d\lambda} \right)^2 + \frac{1}{2} R(r)^2 \left(\frac{d\theta}{d\lambda} \right)^2 = \mathcal{E} = \frac{1}{2} \left(\frac{ds}{d\lambda} \right)^2.$$

One can introduce an arclength parametrization by

$$\frac{ds}{d\lambda} = (2\mathcal{E})^{1/2}$$

or simply by choosing the constant $2\mathcal{E} = 1$ by choosing a unit tangent vector (unit speed motion) as initial data for the differential equations.

Eliminate the angular velocity from this equation to find

$$\frac{1}{2}|c'(\lambda)|^2 = \frac{1}{2} \left(\frac{dr}{d\lambda} \right)^2 + \frac{\ell^2}{2R(r)^2} = \mathcal{E},$$

where the term in the kinetic energy representing the contribution from the angular motion now acts like an “effective potential” for the radial motion. This equation can be directly integrated in principle (or numerically integrated in practice), thus sidestepping integrating the second-order differential equation for r

$$\lambda = \int \frac{dr}{(2\mathcal{E} - \ell^2/R(r)^2)^{1/2}}.$$

If in turn this relation can be inverted to give $r = r(\lambda)$, then one can integrate the angular momentum relation $d\theta/d\lambda = \ell/R(r)^2$

$$\theta = \int \frac{\ell}{R(r(\lambda))^2} d\lambda,$$

which gives the solution formally up to “quadrature,” an old-fashioned word for doing these two integrals.

c) If we are only interested in the path of a geodesic, we can instead use θ as a non-affine parameter by a chain rule re-expression of the derivative in the radial first order equation by substituting

$$\frac{dr}{d\lambda} = \frac{dr}{d\theta} \frac{d\theta}{d\lambda} = \frac{dr}{d\theta} \frac{\ell}{R(r)^2}$$

into

$$\mathcal{E} = \frac{1}{2} \left(\frac{dr}{d\lambda} \right)^2 + \frac{\ell^2}{2R(r)^2} = \frac{1}{2} \left(\frac{dr}{d\theta} \frac{\ell}{R(r)^2} \right)^2 + \frac{\ell^2}{2R(r)^2} = \frac{1}{2} \left(\frac{dr}{d\theta} \right)^2 \frac{\ell^2}{R(r)^4} + \frac{\ell^2}{2R(r)^2}.$$

This can be used to solve for $d\theta/dr$ and integrate to find $\theta = \theta(r)$ and if possible, inverted to get r as a function of θ . Solve this energy equation for $d\theta/dr$, modulo sign, and express θ as function of r using an integral formula as above.

In the case of zero specific angular momentum $\ell = 0$, i.e., of purely radial motion for constant θ , this just tells us instead that $dr/d\lambda$ is zero, so r is linearly related to an affine parameter, and is therefore an affine parameter itself. In fact we know it is an arclength parameter for purely radial motion, and by choosing $2\mathcal{E} = 1$, the parameter λ measures increments of r . ■

Remark

To interpret the parameters (\mathcal{E}, ℓ) in terms of initial data, we need to consider the tangent vector, called the velocity in the physics language

$$v = \frac{dr}{d\lambda} e_r + \frac{d\theta}{d\lambda} e_\theta = \frac{dr}{d\lambda} e_{\hat{r}} + R \frac{d\theta}{d\lambda} e_{\hat{\theta}} = v^{\hat{r}} e_{\hat{r}} + v^{\hat{\theta}} e_{\hat{\theta}}.$$

The speed is just the length of the velocity vector, namely the square root of twice the (specific) energy: $|v| = (2\mathcal{E})^{1/2}$. If we introduce the usual polar angle in the $e_{\hat{r}}-e_{\hat{\theta}}$ tangent plane measured counterclockwise from the positive $e_{\hat{r}}$ -axis, call it β , then

$$v = \frac{dr}{d\lambda} e_r + R \frac{d\theta}{d\lambda} e_{\hat{\theta}} = (2\mathcal{E})^{1/2} (\cos \beta e_{\hat{r}} + \sin \beta e_{\hat{\theta}}),$$

so that comparing formulas

$$\left(\frac{dr}{d\lambda}, R \frac{d\theta}{d\lambda} \right) = ((2\mathcal{E})^{1/2} \cos \beta, (2\mathcal{E})^{1/2} \sin \beta).$$

Recalling that $d\theta/d\lambda = \ell/R^2$, then we obtain

$$\frac{d\theta}{d\lambda} = \frac{(2\mathcal{E})^{1/2}}{R} \sin \beta \quad \text{and} \quad \ell = R^2 \frac{d\theta}{d\lambda} = R(2\mathcal{E})^{1/2} \sin \beta,$$

and hence

$$(v^{\hat{r}}, v^{\hat{\theta}}) = \left(\frac{\ell}{R} \coth \beta, \frac{\ell}{R} \right) \quad \mathcal{E} = \frac{\ell^2}{2R^2 \sin^2 \beta}.$$

Note that the combination

$$R_{\text{ext}} = \frac{|\ell|}{(2\mathcal{E})^{1/2}} = R(\lambda) |\sin \beta(\lambda)|$$

is a constant length which reduces to the distance from the z -axis of a point on a geodesic where a radial turning point occurs, namely where $dr/d\lambda = 0 = \cos \beta$ and $|\sin \beta| = 1$.

Thus once one picks the energy which determines the affine parametrization for a geodesic (the choice $\mathcal{E} = 1/2$ is the arclength parametrization), then picking the initial angle β completely determines the initial data for the second order geodesic differential equations at a given starting point (r_0, θ_0) , namely the initial values of the two components of the velocity $(dr/d\lambda, d\theta/d\lambda)$. This also determines both the initial value of $dr/d\lambda$ and the constant value of ℓ needed for the radial motion problem, ignoring the angular motion. Alternatively one can determine the choice of parametrization for “nonradial” motion by fixing $\ell \neq 0$, say $|\ell| = 1$, and then determining the energy parameter \mathcal{E} from the initial values of (R, β) through the above relations, as illustrated in figure 8.12. This allows us to use a single potential plot with a variable energy level to describe all possible motions.

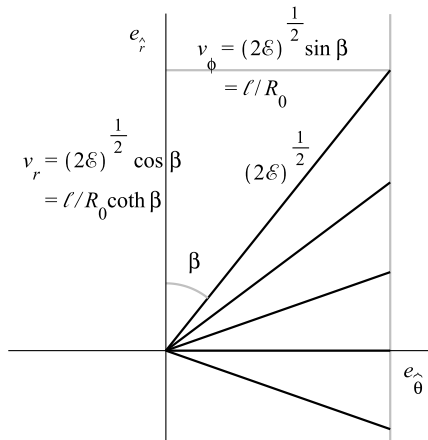


Figure 8.12: Holding $R_0 = R(r(\lambda_0))$ and ℓ fixed while considering different initial velocities, the initial velocity has a constant azimuthal (horizontal) orthonormal component but a variable speed $(2\mathcal{E})^{1/2}$ and radial (vertical) orthonormal component. On the other hand, fixing the speed and varying the angular momentum would change both the radial and azimuthal component and the graph of the effective potential.

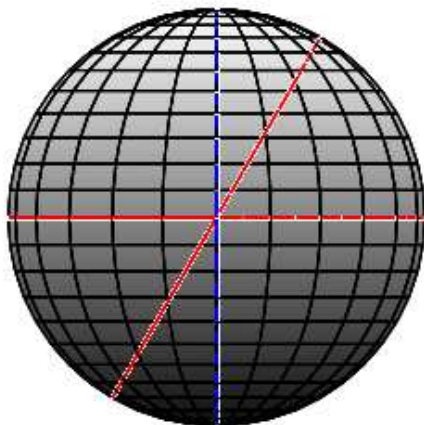


Figure 8.13: A great circle through the equator makes an angle with the vertical at those intersection points which equals the polar angle of the nearest point on that great circle to one of the poles, or its complement, modulo absolute value since $0 \leq \theta \leq \pi$. This is clear from this edge on view looking down the diameter connecting the two intersection points with the equator.

Remark.

Suppose we consider a great circle on the unit sphere in spherical coordinates passing through the equator at polar angle $\theta = \pi/2$, where $R = \sin \theta$. If we look edge on down the diameter which connects antipodal points on this great circle on the equator, it is clear that the absolute value of the angle β_0 from the vertical at the equator of a tangent vector to this curve equals the extremal polar angle closest to one of the poles on that great circle. This is exactly what the conserved angular momentum quantity tells us since the product $\sin \theta \sin \beta$ is constant. At the equator $(\theta, \beta) = (\pi/2, \beta_0)$, while at the extremal point $(\theta, \beta) = (\theta_{\text{ext}}, \pi/2)$, so $\sin \theta = \sin \beta_0$. Since $0 \leq \theta_{\text{ext}} \leq \pi$, one must have $\theta_{\text{ext}} = |\beta_0|$ or $\pi - |\beta_0|$.

▲

Exercise 8.5.2.

surface of revolution meridian arclength

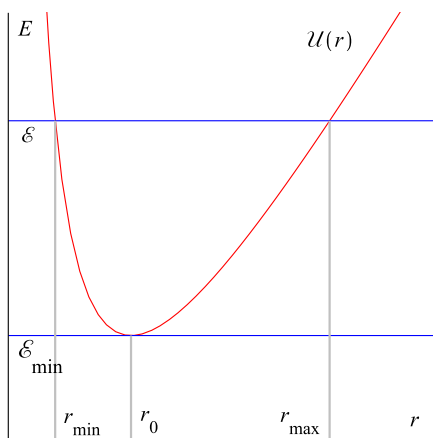


Figure 8.14: When a 1-dimensional potential energy function has a “potential well,” namely an interval with a concave up graph surrounding a minimum, then the minimum point like at r_0 is a stable equilibrium solution $r = r_0$ of the equation of motion, while energy levels above that minimum energy have “turning points” r_{\min}, r_{\max} where the energy level intersects the graph of the potential $U(r_{\min}) = \mathcal{E} = U(r_{\max})$ and the kinetic energy must be zero, corresponding to points where the motion must be momentarily stopped.

For practice in evaluating arclengths of curves in the x - y plane, calculus books use very special parametrized curves or graphs of functions $y = f(x)$ parametrized by $x = t, y = f(t)$ such that the length of the tangent vector is a perfect square and the integration can actually be performed. For some of these examples the resulting relationship between the arclength function s with respect to some arbitrary reference point on the curve can be inverted in terms of the curve parameter so that one can actually find an arclength parametrization of the curve. Revolving these special example curves (re-expressed in terms of the r - z plane) around the vertical axis, we get actual examples of surfaces where the preceding problem can be worked explicitly. Find an interesting example in such a book and construct a surface of revolution where one can express $R(r)$ explicitly and apply the preceding general theory. ■

8.6 The Euclidean plane and the Kepler problem

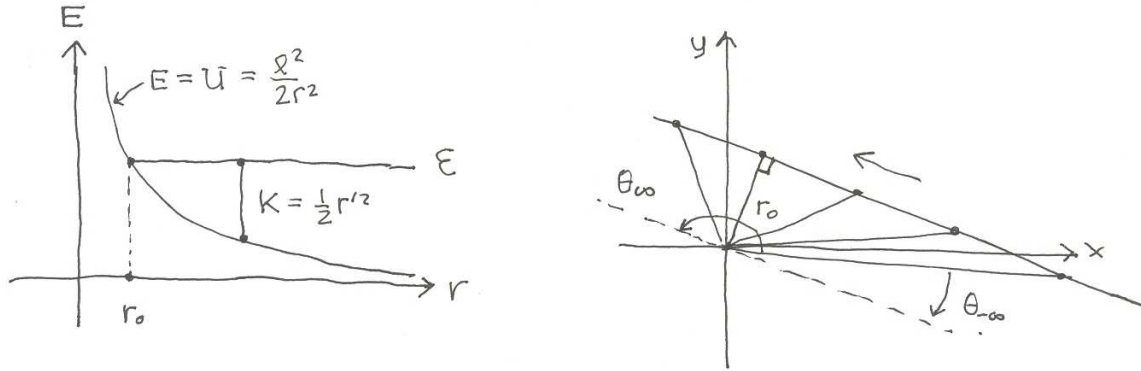


Figure 8.15: A straight line in the plane not passing through the origin seen in polar coordinates, using energy considerations. [Change r_0 to r_p in new figure.]

Continuing the particle motion analogy for the simpler case of the flat Euclidean plane in polar coordinates is very instructive since we know that the geodesics are just straight lines. The specific energy equation

$$\frac{1}{2}|c'(\lambda)|^2 = \frac{1}{2} \left(\frac{dr}{d\lambda} \right)^2 + \frac{\ell^2}{2R(r)^2} = \mathcal{E}$$

leads to a very useful visualization of the radial motion, thinking of the problem in terms of the motion of a point in space with respect to the time $t = \lambda$. The first term is the specific kinetic energy $\mathcal{K} = \frac{1}{2} (dr/d\lambda)^2$ and the second term the specific potential energy $\mathcal{U} = \frac{1}{2} \ell^2 / R(r)^2$. Their sum is a constant along a geodesic, namely the specific energy. If we graph the potential function and the constant of energy, then only values of r are allowed where the potential falls below the “energy level” so that the kinetic energy is nonnegative. Where they intersect, the kinetic energy and hence the radial derivative itself must be zero, which corresponds to a turning point of the radial motion where the radial motion must reverse direction, unless it is an equilibrium solution at a critical point of the potential where the radius is fixed by the equation of motion.

For example, take the case of the flat plane as a surface of revolution. Just rotate the line $z = 0$ in the ρ - z plane around the z -axis in cylindrical coordinates (ρ, ϕ, z) , and let $(r, \theta) = (\rho, \phi)$, where $R(r) = r$ and (r, θ) are the usual polar coordinates. The first order geodesic equations are

$$\frac{1}{2} \left(\frac{dr}{d\lambda} \right)^2 + \frac{\ell^2}{2r^2} = \mathcal{E}, \quad \frac{d\theta}{d\lambda} = \frac{\ell}{r^2}.$$

Figure 8.15 illustrates the situation for a straight line not passing through the origin, seen from the point of view of polar coordinates. The turning point r_p (“p” for perihelion) is defined

by setting the radial velocity $dr/d\lambda$ to zero, so that $\frac{1}{2}\ell^2/r_p^2 = \mathcal{E}$ or $r_p = |\ell|/(2\mathcal{E})^{1/2}$. This is just the distance of closest approach to the origin, and in the physics language is the “orbit perihelion.” The geodesic starts at $\lambda \rightarrow -\infty$ at the angle $\theta_{-\infty}$ and monotonically increments by $\pm\pi$ to its limiting final value θ_{∞} while the radius decreases from ∞ to its turning point and then increases back out to infinity as $\lambda \rightarrow \infty$. We can envision the radial motion taking place in the r - E plane as a point coming in from infinity moving to the left along the line of constant energy, decreasing its radial speed as the kinetic energy shrinks, until its radial motion stops at the turning point and then reverses and then it moves back out (symmetrically in λ with respect to the turning point) to infinity again. Since $d\theta/d\lambda = \ell/r^2$, the angular velocity has its largest values near the turning point where the angle is changing the most rapidly as one moves along the straight line in an affine parametrization that is proportional to its arclength.

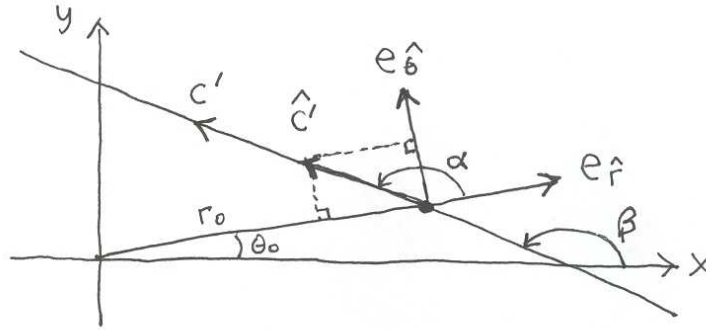


Figure 8.16: Initial data for the geodesic represented in the preceding figure, making an angle β_1 with respect to the horizontal axis. The initial direction can be specified by the angle $\alpha - > \beta_0$ made by the tangent vector c' with respect to the orthonormal frame associated with the polar coordinates.

To specify initial data for a nonradial geodesic like the one illustrated in figure 8.15, one needs an initial position and initial tangent vector $(r_0, \theta_0, r'_0, \theta'_0)$, or equivalently $(r_0, \theta_0, \beta_0, \mathcal{E})$. To interpret this geometrically, see figure 8.16. Since we have defined the “specific energy” by setting $\mathcal{E} = \frac{1}{2}|c'|^2$ (remember the kinetic energy is $E = \frac{1}{2}mv^2$, so $E/m = \frac{1}{2}v^2$ thinking of c' as the velocity), then turning this around: $|c'| = (2\mathcal{E})^{1/2}$. Thus we can define the initial unit tangent vector to be

$$\hat{c}'|_0 = \cos \beta_0 e_{\hat{r}} + \sin \beta_0 e_{\hat{\theta}} = \cos \beta_0 e_r + \frac{\sin \beta_0}{r_0} e_{\theta},$$

which implies

$$c'|_0 = (2\mathcal{E})^{1/2}(\cos \beta_0 e_r + \frac{\sin \beta_0}{r_0} e_{\theta}) = r'_0 e_r + \theta'_0 e_{\theta},$$

and finally the specific angular momentum is

$$\ell = r_0^2 \theta'_0 = r_0 (2\mathcal{E})^{1/2} \sin \beta_0.$$

Thus if we fix the specific angular momentum instead of requiring arclength parametrization (which would instead fix $2\mathcal{E} = 1$), then letting $\sin \beta_0 \rightarrow 0$ means $\mathcal{E} \rightarrow \infty$. This allows us to plot a single fixed potential and vary the energy (i.e., initial angle β_0) to characterize the different kinds of nonradial geodesics which are possible, instead of fixing the energy and plotting a whole family of potentials (by varying instead the coefficient ℓ^2).

Of course we know explicitly how to express the geodesics of the Euclidean plane which pass through an arbitrary point (x_0, y_0) . They are straight lines which may be parametrized by the angle β_1 they make with respect to the direction unit vector e_1 in the tangent space at that point, measured in the usual counterclockwise direction

$$\begin{aligned} c(s, x_0, y_0, \beta_1) &= \langle x_0, y_0 \rangle + s \langle \cos \beta_1, \sin \beta_1 \rangle = \langle x_0 + s \cos \beta_1, y_0 + s \sin \beta_1 \rangle, \\ c'(s, x_0, y_0, \beta_1) &= \langle \cos \beta_1, \sin \beta_1 \rangle, \quad |c'(s, x_0, y_0, \beta_1)| = 1. \end{aligned}$$

Clearly these families of straight lines are not very compatible with a coordinate system built on concentric circles about the origin unless the point in question is the origin itself, so using polar coordinates to describe the geodesics of the plane is a doomed exercise. However, it is an instructive stepping stone to a more interesting problem of motion in the plane under the influence of a central force, i.e., a force field which is radially directed and spherically symmetric about a point, e.g., the gravitational force on a “point particle” like the earth around the sun, or the classical electric force of an electron orbiting the nucleus.

Substituting these expressions for x and y in the polar coordinate map (i.e., $\Phi \circ c(s, x_0, y_0, \beta_1)$ in precise notation), one finds a pretty complicated relationship

$$r = [(x_0 + s \cos \beta_1)^2 + (y_0 + s \sin \beta_1)^2]^{1/2}, \quad \tan \theta = \frac{y_0 + s \sin \beta_1}{x_0 + s \cos \beta_1}.$$

These represent the general solution of the geodesic equations expressed in polar coordinates, where invoking the “square peg in a round hole” saying might be in order.

Exercise 8.6.1.

plane geodesics in polar coordinates

To simplify the formulas, we express these geodesics for the nonradial motion case in terms of an arclength parametrization $s = (2\mathcal{E})^{1/2}\lambda$ with $s = 0$ at the initial data point (r_0, θ_0) with initial angle β_0 . The first order equations are then

$$\left(\frac{dr}{ds}\right)^2 + \frac{r_p^2}{r^2} = 1, \quad \frac{d\theta}{ds} = \frac{(\operatorname{sgn} \ell) r_p}{r^2},$$

recalling the definition of the radius of closest approach to the origin and conserved quantity

$$r_p = |\ell/(2\mathcal{E})^{1/2}| = r |\sin \beta| = r_0 |\sin \beta_0|.$$

Introduce also $s_p = r_0 \cos(\pi - \beta_0) = -r_0 \cos \beta_0$, which satisfies $r_0^2 = r_p^2 + s_p^2$. Notice that the sign of $d\theta/ds$ cannot change for nonradial geodesics, but that of dr/ds must change since

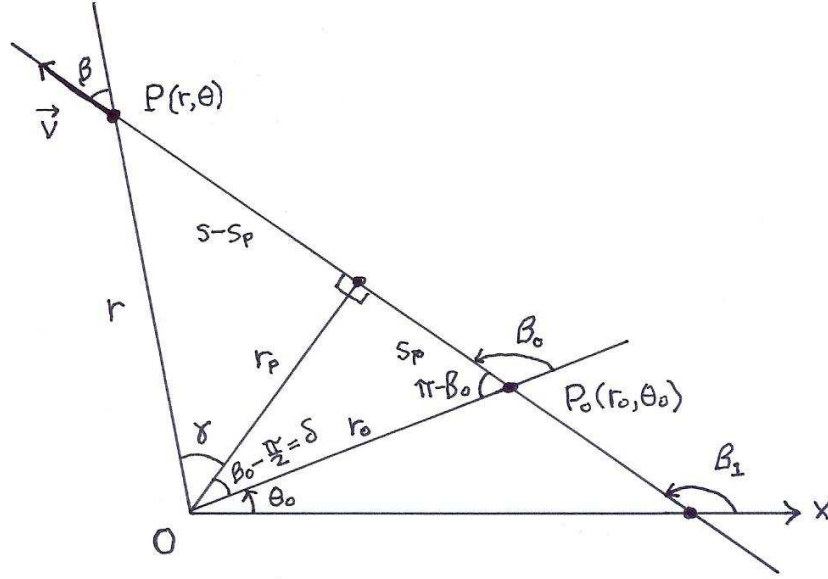


Figure 8.17: The geometry of straight lines in polar coordinates as a geodesic solution. This shows the case where $\beta_0 = \beta_1 - \theta_0$ is a positive obtuse angle and the point $P(r, \theta)$ is on the other side of the point of closest approach to the origin from the initial point $P(r_0, \theta_0)$ so that $s_p = r_0 \cos(\pi - \beta_0) > 0$.

it always decreases as one comes in from infinity and then increases after passing the point of closest approach to the origin.

a) Solve the radial equation to find

$$\frac{dr}{ds} = \text{sgn} \left(\frac{dr}{ds} \right) \sqrt{1 - \frac{r_p^2}{r^2}},$$

and then solve this for ds and integrate, imposing the initial condition $s = 0$ at $r = r_0$ to find

$$s = \text{sgn} \left(\frac{dr}{ds} \right) \int_{r_0}^r \frac{u}{\sqrt{u^2 - r_p^2}} du + s_p = \text{sgn} \left(\frac{dr}{ds} \right) \sqrt{r^2 - r_p^2} + \text{sgn}(s_p) \sqrt{r_0^2 - r_p^2},$$

which is equivalent to the Pythagorean relation

$$(s - s_p)^2 + r_p^2 = r^2.$$

Note that the sign $\text{sgn}(dr/ds)$ here changes when s passes the value s_p corresponding to the point of closest approach to the origin, while s_p can have any sign.

b) Re-express the previous result for r^2 using the definition of s_p to find

$$r^2 = s^2 + r_0^2 + 2r_0 s \cos \beta_0.$$

c) Use a computer algebra system to integrate the resulting angular equation

$$\frac{d\theta}{ds} = \frac{\operatorname{sgn}(\ell)r_p}{s^2 + r_0^2 + 2r_0s \cos \beta_0}$$

with the initial condition $\theta = \theta_0$ when $s = 0$ to find

$$\theta - \theta_0 = \operatorname{sgn}(\ell) \left[\arctan \left(\frac{s + r_0 \cos \beta_0}{r_0 |\sin \beta_0|} \right) - \arctan \left(\frac{\cos \beta_0}{|\sin \beta_0|} \right) \right] \equiv \operatorname{sgn}(\ell)[\gamma + \delta].$$

Convince yourself that for $0 \leq \beta_0 \leq \pi$, the second angle δ inside the square brackets equals the complementary angle $\delta = \pi/2 - |\beta_0|$.

d) Now refer to Fig. 8.17. See that the radial solution for r as a function of s is just the law of cosines for the triangle OP_0P applied to the angle $\angle OP_0P$, while the angular solution for $\theta - \theta_0$ is exactly the sum of the two angles γ and δ shown there

$$\gamma = \arctan \left(\frac{s - s_p}{r_0} \right), \quad \delta = \arctan \left(\frac{s_p}{r_0} \right).$$

e) Substitute $(x_0, y_0) = r_0(\cos \theta_0, \sin \theta_0)$ into the Cartesian expression for r^2 given originally for the straight line, and show that it reduces to the law of cosines result given above.

f) In a similar way, try to reconcile the Cartesian expression for θ with the expression resulting from the integration above. Warning: I have not yet done this myself. It may be tricky. ■

Exercise 8.6.2.

orbit equation for plane geodesics in polar coordinates

Suppose we wish only to determine the path of a geodesic as in Exercise 8.4.1 with the azimuthal radius function $R(r) = r$, expressing r as a function of θ along the geodesic, i.e., using θ as a nonaffine parameter. What would the result look like? Starting from the known solution for an arclength parametrized geodesic in polar coordinates, use a computer algebra system to invert the relationship between θ and s and then use it to replace s in the expression for r and simplify the result. You will see that it is not very simple. ■

Kepler's problem

These manipulations are typical of elementary mechanics in the physics world, and while perhaps a bit pointless here in the case of the flat plane, we are only a step away from being able to solve Kepler's problem of determining the orbits of the planets using the same approach. A radially inward pointing specific force which is proportional to the inverse square of the distance

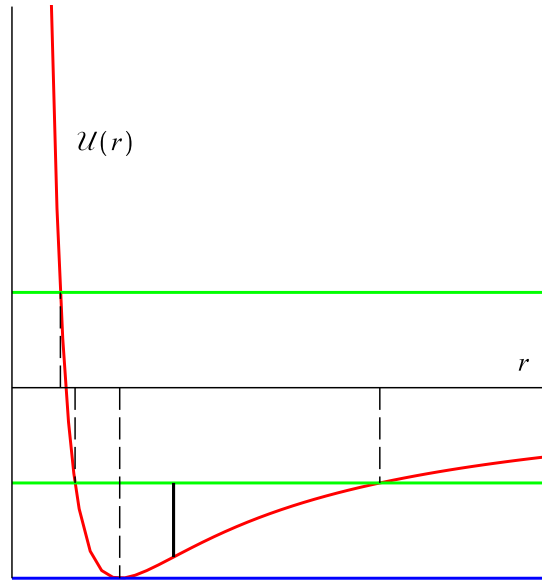


Figure 8.18: The effective potential for the radial motion in the Kepler problem for the variable r . The r -axis (black) describes the zero energy solutions (parabolas), while the minimum energy (blue) at the bottom of the negative energy potential well describes the circular orbit. The two green lines describe the positive energy solutions (hyperbolas) and the negative energy solutions (ellipses). Turning points are indicated by the dashed vertical lines. The solid vertical line represents the kinetic energy for the negative energy orbit.

from the origin in the plane is described by a potential inversely proportional to that distance with a negative constant of proportionality in each case

$$\mathcal{F}^r = -\frac{k}{r^2} = -\frac{d\mathcal{U}}{dr}, \quad \mathcal{U} = -\frac{k}{r}, \quad k > 0.$$

The new energy equation is then

$$\frac{1}{2} \left(\frac{dr}{d\lambda} \right)^2 + \frac{\ell^2}{2r^2} - \frac{k}{r} = \mathcal{E}, \quad \mathcal{U}_{(\text{eff})} = \frac{\ell^2}{2r^2} - \frac{k}{r},$$

where $\mathcal{U}_{(\text{eff})}$ is an effective potential for the radial motion. Note that now the angular momentum is an essential parameter in determining the shape of the total radial potential, while in the geodesic case it only scaled the single potential term, so that simultaneously scaling the energy could compensate for that first scaling. The extra real potential term decouples the scaling freedom in the affine parameter from the scaling of the total radial potential.

The following approach can be used to study any rotationally symmetric conservative force field in the plane. A simple toy example that often appears in classical mechanics textbooks is the 2-dimensional harmonic oscillator, which actually pops up in the Kepler problem with an obvious change of variable (reciprocal radius). One can also explore other powers for the radial potential.

To get the orbits for the Kepler problem, we use the nonaffine parameter θ and re-express the energy equation for the new radial variable $u = 1/r$

$$\mathcal{E} = \frac{1}{2} \left(\frac{dr}{d\theta} \right)^2 \frac{\ell^2}{r^4} + \frac{\ell^2}{2r^2} - \frac{k}{r} = \frac{\ell^2}{2} \left[\left(\frac{du}{d\theta} \right)^2 + u^2 - \frac{2k}{\ell^2} u \right],$$

or

$$\left(\frac{du}{d\theta} \right)^2 = -u^2 + \frac{2\mathcal{E} + 2ku}{\ell^2} = -(u - A)^2 + A^2 \left(1 + \frac{2\mathcal{E}\ell^2}{k^2} \right),$$

where $A = k/\ell^2$ comes from completing the square on the quadratic expression. This has the easy solution

$$\frac{1}{r} = u = A + B \cos(\theta - \theta_0) = A(1 + e \cos(\theta - \theta_0)), \quad (A, B, e) = \left(\frac{k}{\ell^2}, \frac{k}{\ell^2} e, \left(1 + \frac{2\mathcal{E}\ell^2}{k^2} \right)^{1/2} \right)$$

representing a conic section of eccentricity e and semi-latus rectum $p = 1/A$. For the elliptical case of bound orbits $\mathcal{E} < 0$, the major and minor semi-axes are $a = p/(1 - e^2)$, $b = p/\sqrt{1 - e^2}$, while the perihelion and aphelion are $r_{\min} = a(1 - e)$, $r_{\max} = a(1 + e)$.

Exercise 8.6.3.

quadratic potential motion

a) Verify this solution by substituting it into the differential equation and simplifying it to the fundamental trig identity.

b) Derive this solution using a computer algebra system. You will get different results if you input literally the above differential equation, or if you solve it for one of the two roots for $du/d\theta$ and solve that, or if you first solve for $d\theta/du$ and integrate. In each case the computer algebra system tries to express the solution in terms of the tangent (or arctangent in the latter case) instead of the cosine (or the arccosine) which is much simpler. ■

This form of the solution is nearly obvious if we think about the new form of the potential.

$$\frac{1}{2} \left(\frac{du}{d\theta} \right)^2 + \mathcal{V}(u) = \frac{\mathcal{E}}{\ell^2}, \quad \mathcal{V}(u) = \frac{1}{2} u^2 - \frac{ku}{\ell^2} = \frac{1}{2} u \left(u - \frac{2k}{\ell^2} \right).$$

This is just a quadratic potential displaced from the origin describing a linear Hooke's law force with a displaced origin, so its solutions are just oscillations about the minimum point $u = k/\ell^2$, which leads to the general form of the solution given above. Fig. 8.19 shows the same situation as the previous figure.

For negative energy $\mathcal{E} < 0$, then the eccentricity e is a proper fraction and the orbit is an ellipse with major and minor axes

$$a = \frac{p}{(1 - e^2)} = \frac{k}{2\mathcal{E}}, \quad b = \frac{p}{(1 - e^2)^{1/2}} = \frac{\ell}{(2\mathcal{E})^{1/2}},$$

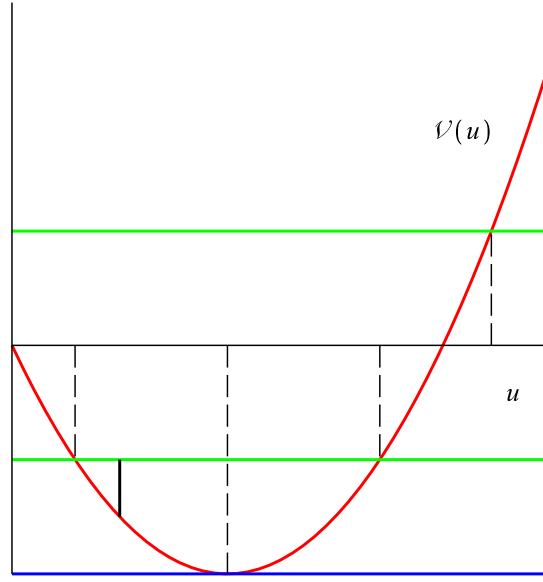


Figure 8.19: The effective potential for the radial motion in the Kepler problem for the reciprocal radial variable $u = 1/r$. The r -axis (black) describes the zero energy solutions (parabolas), while the minimum energy (blue) at the bottom of the negative energy potential well describes the circular orbit. The two green lines describe the positive energy solutions (hyperbolas) and the negative energy solutions (ellipses). Turning points are indicated by the dashed vertical lines.

and perihelion and aphelion

$$r_{(\min)} = a(1 - e), \quad r_{(\max)} = a(1 + e).$$

If $\mathcal{E} = 0$, the eccentricity $e = 1$ and the orbit is a parabola with minimum radius

$$r_{(\min)} = \frac{p}{2} = \frac{\ell^2}{k}.$$

For positive energy $\mathcal{E} > 0$, then the eccentricity e is an improper fraction and the orbit is a hyperbola with minimum radius and orthogonal semi-axis

$$r_{(\min)} = \frac{p}{(1 + e)} = a(1 - e) = \frac{\ell^2}{k}, \quad a = \frac{p}{(e^2 - 1)} = \frac{k}{2\mathcal{E}}.$$

To get relation between the orbit and the affine parameter (time), one must integrate the energy equation, solved for $d\lambda/dr$

$$\frac{dr}{d\lambda} = \left(2\mathcal{E} + \frac{2k}{r} - \frac{\ell^2}{r^2} \right)^{1/2}$$

and integrated

$$\lambda - \lambda_0 = \int \frac{r dr}{(2\mathcal{E}r^2 + 2kr - \ell^2)^{1/2}}.$$

Exercise 8.6.4.
radius versus time

a) Evaluate this integral with a computer algebra system. Find the period T for an elliptical orbit by doubling the definite integral from $r_{(\min)}$ to $r_{(\max)}$.

b) If this is not successful try plan B:

Supply the missing steps in completing the square and factoring the quadratic expression in the numerator of the expression for the square of the speed:

$$\begin{aligned} \left(\frac{dr}{d\lambda}\right)^2 &= -|2\mathcal{E}| + \frac{2k}{r} - \frac{\ell^2}{r^2} \\ &= \frac{|2\mathcal{E}|}{r^2} \left(-r^2 + \frac{2k}{|2\mathcal{E}|}r - \frac{\ell^2}{|2\mathcal{E}|}\right) \\ &= \frac{|2\mathcal{E}|}{r^2} (-(r-a)^2 + a^2e^2) = \frac{|2\mathcal{E}|}{r^2} ((r_{(\max)} - r)(r - r_{(\min)})) , \end{aligned}$$

where $a = k/|2\mathcal{E}|$, $e = (1 - |2\mathcal{E}|\ell^2/k^2)^{1/2}$. Then solve for $d\lambda$ and integrate from $r_{(\min)}$ to $r_{(\max)}$ and double to get the period:

$$\begin{aligned} T &= \frac{2}{|2\mathcal{E}|^{1/2}} \int_{r_{(\min)}}^{r_{(\max)}} \frac{r dr}{\sqrt{(r_{(\max)} - r)(r - r_{(\min)})}} \stackrel{\text{easy}}{=} \frac{2}{|2\mathcal{E}|^{1/2}} \frac{\pi}{2} (r_{(\max)} + r_{(\min)}) \\ &= \frac{2\pi a}{|2\mathcal{E}|^{1/2}} = \frac{2\pi k}{|2\mathcal{E}|^{3/2}}. \end{aligned}$$

With all this help, the computer algebra system easily spits out the easy integral with the assumptions $r_{(\max)} > r, r > r_{(\min)}, r_{(\min)} > 0$. Notice that the result is independent of the angular momentum, only depending on the energy. This means that at a given initial radius, all the elliptical orbits have the same period as the circular orbit with the Kepler speed $v_K = (k/r)^{1/2}$, namely $T = 2\pi r/v_K = 2\pi a^{3/2}/k^{1/2}$. ■

Exercise 8.6.5.
black hole orbits

The general relativistic orbits corresponding to the Kepler problem are described by geodesics in the Schwarzschild spacetime representing a nonrotating black hole. Only a small change in our nonrelativistic equations is required to adapt them to their relativistic versions. Read chapter 25 of *Gravitation* by Misner, Thorne and Wheeler to investigate this problem, or wait until Section 8.12 where we will derive the relativistic equations. The affinely parametrized geodesics in the equatorial plane of a spherical coordinate system in which the point source mass M of

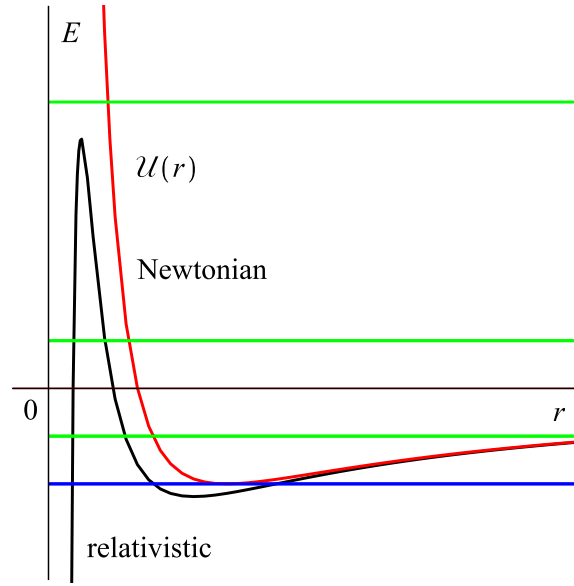


Figure 8.20: The effective potential for the relativistic Kepler problem, compared with the Newtonian potential. The additional inverse third power term in the attractive relativistic gravitational potential overcomes the repulsive inverse square centrifugal potential term at small radii, leading to orbits which fall into the center if the energy is high enough to pass over the barrier. The maximum of the relativistic potential represents an unstable circular orbit, while the minimum represents a stable circular orbit.

the gravitational field is located at the origin, reverting back to polar coordinate notation in that plane, are described by the same angular equation and a radial energy equation with one factor slipped into the formula

$$\frac{d\theta}{d\lambda} = \frac{\ell}{r^2}, \quad \frac{1}{2} \left(\frac{dr}{d\lambda} \right)^2 + \frac{1}{2} \left(1 - \frac{2M}{r} \right) \left(1 + \frac{\ell^2}{r^2} \right) = \frac{1}{2} \mathcal{E}^2.$$

This factor multiplying the effective potential term associated with the angular momentum (“the centrifugal potential” associated with the repulsive centrifugal force) has the value 1 at large radii compared to the “Schwarzschild radius” $r = 2M$, but goes to 0 as one approaches this radius from larger values, thus cutting off the repulsive effects of the centrifugal potential. Thus very close to the black hole, the attractive gravitational force overcomes the centrifugal force and pulls test particles (whose speed tops out at the speed of light $c = 1$) into the hole no matter what their angular momentum is.

a) Expand out the radial energy equation and rearrange terms to find the result

$$\frac{1}{2} \left(\frac{dr}{d\lambda} \right)^2 + \frac{\ell^2}{2r^2} - \frac{M}{r} - \frac{M\ell^2}{r^3} = \frac{1}{2} (\mathcal{E}^2 - 1).$$

Thus an additional inverse third power term in the potential is the only change needed to find the orbits, and such a term simply causes the precession of the Newtonian orbits around the

central mass. This corresponds to an effective radial force which increases the attraction to the center for very small radii, thus overcoming the centrifugal potential barrier of the corresponding Kepler problem

$$F_r = -\frac{M}{r^2} - \frac{3M\ell^2}{r^4}.$$

b) Re-express the radial equation in terms of the variable $u = M/r$ and the ratio $\tilde{\ell} = \ell/M$ and reparametrize the orbits by the azimuthal angular variable θ to find

$$\left(\frac{du}{d\theta}\right)^2 = \frac{\mathcal{E}^2 - (1 - 2u)(1 + \tilde{\ell}^2 u^2)}{\tilde{\ell}^2} = -u^2 + \frac{\mathcal{E}^2 - 1 + 2u + 2\tilde{\ell}^2 u^3}{\tilde{\ell}^2}.$$

This is almost the same as the Newtonian expression except that the Newtonian specific energy corresponds to the function $\frac{1}{2}(\mathcal{E}^2 - 1)$ of the relativistic specific energy \mathcal{E} and there is an extra cubic term in u responsible for the famous precession of the corresponding Newtonian conic section orbits.



8.7 2-spheres, pseudospheres and other conics of revolution

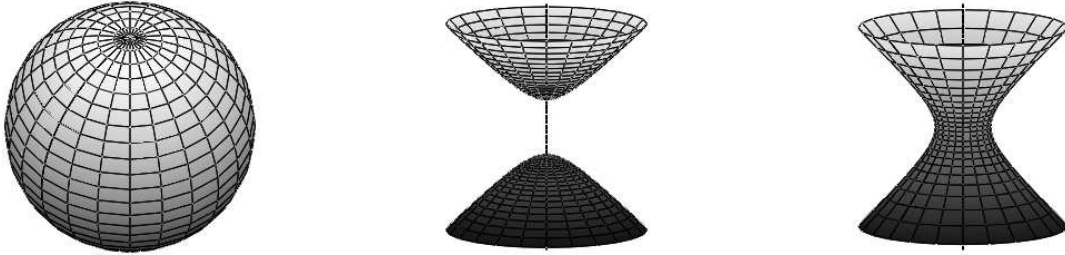


Figure 8.21: The 2-sphere in Euclidean 3-space and the two pseudospheres in 3-dimensional Minkowski spacetime: hyperboloids of two and one sheets, which are spacelike and timelike surfaces respectively, corresponding to all unit timelike and spacelike separations respectively from the origin of inertial coordinates at the center of the hyperboloids.

The sphere is our most familiar surface in ordinary Euclidean space which is both intrinsically and extrinsically curved as well as maximally symmetric under the 3-parameter group of rotations which leave its geometry invariant, including translations between any two points on the sphere (homogeneity symmetry) and a rotation about every point on the sphere (isotropy symmetry). We normally study spheres centered at the origin of our Cartesian coordinate system, where they occur as the level surfaces of the distance function from the origin.

The pseudospheres in 3-dimensional Minkowski spacetime have similar properties in their Lorentzian geometry, but come in two different types. The single sheeted hyperboloid consists of all points a fixed spacelike separation from the origin of the inertial coordinates, which can be viewed as a world sheet of a circle as it accelerates away outward from the origin in the 2-dimensional space cross-sections. The double sheeted hyperboloid consists of all points at a fixed timelike separation from the origin in the future or in the past. These pseudospheres have a rotational symmetry about the time axis and two boost symmetries which allow motion along the radial direction within these surfaces.

These surfaces can be studied easily using the spherical/pseudospherical coordinates which parametrize the family of such surfaces of different radii/pseudoradii

$$\begin{array}{lll}
 x = r \sin \theta \cos \phi, & x = \tau \cosh \chi \cos \phi, & x = \ell \sinh \chi \cos \phi, \\
 y = r \sin \theta \sin \phi, & y = \tau \cosh \chi \sin \phi, & x = \ell \sinh \chi \sin \phi, \\
 z = r \cos \theta, & t = \tau \sinh \chi, & t = \ell \cosh \chi,
 \end{array}$$

where $r \geq 0$, $\ell \geq 0$ but $\tau \in \mathbb{R}$ allows positive and negative values of τ to describe the future and past pseudospheres respectively, while $0 \leq \theta \leq \pi$, $0 \leq \phi \leq 2\pi$ (or $-\pi < \phi \leq \pi$ of convenient) and finally the hyperbolic angle or “rapidity” $\chi \geq 0$ for the two sheeted hyperboloids, but $\chi \in \mathbb{R}$ for the one sheeted hyperboloid.

spheres

The Euclidean metric in spherical coordinates is

$$\begin{aligned} g &= dx \otimes dx + dy \otimes dy + dz \otimes dz \\ &= dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2 \theta d\phi \otimes d\phi \\ &= \omega^{\hat{r}} \otimes \omega^{\hat{r}} + \omega^{\hat{\theta}} \otimes \omega^{\hat{\theta}} + \omega^{\hat{\phi}} \otimes \omega^{\hat{\phi}} \end{aligned}$$

with coordinate frame and dual frame

$$e_r = \frac{\partial}{\partial r}, \quad e_\theta = \frac{\partial}{\partial \theta}, \quad e_\phi = \frac{\partial}{\partial \phi}, \quad \omega^r = dr, \quad \omega^\theta = d\theta, \quad \omega^\phi = d\phi$$

and associated orthonormal frame and dual frame

$$e_{\hat{r}} = \frac{\partial}{\partial r}, \quad e_{\hat{\theta}} = \frac{1}{r} \frac{\partial}{\partial \theta}, \quad e_{\hat{\phi}} = \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi}, \quad \omega^{\hat{r}} = dr, \quad \omega^{\hat{\theta}} = r d\theta, \quad \omega^{\hat{\phi}} = r \sin \theta d\phi.$$

The coordinate surface $r = r_0$ is a sphere of radius r_0 on which $\{\theta, \phi\}$ serve as local coordinates with coordinate singularity at the poles $\theta = 0, \pi$ where ϕ is undefined. We can live with this problem as long as we are careful.

The 2-sphere is a space in its own right and we can use all the machinery we have developed for \mathbb{R}^n in general coordinate systems to study it. We can also picture the 2-dimensional tangent spaces to the 2-sphere as subspaces of the full 3-dimensional tangent space of \mathbb{R}^3 at each point. The sphere has “intrinsic” or internal geometry of a 2-dimensional nature, plus “extrinsic” or external geometry that has to do with how it sits inside the larger space, i.e., how it bends. For example, a cylinder locally has the same flat 2-dimensional geometry as a plane (cut it along a seam and roll it out flat), but clearly it is bent as a subspace of \mathbb{R}^3 . To study the intrinsic geometry we simply use the 2-dimensional coordinate system and calculate as though we were studying \mathbb{R}^2 in non-Cartesian coordinates.

Setting $r = r_0$ and $dr = 0$ in the full metric gives us the “induced metric” on the sphere of radius r_0 , which tells us the inner products of the frame vectors e_θ, e_ϕ or $e_{\hat{\theta}}, e_{\hat{\phi}}$

$$\begin{aligned} {}^{(2)}g &= g|_{r=r_0, dr=0} = r_0^2 d\theta \otimes d\theta + r_0^2 \sin^2 \theta d\phi \otimes d\phi \\ &= r_0^2 \underbrace{[d\theta \otimes d\theta + \sin^2 \theta d\phi \otimes d\phi]}_{\text{metric for unit sphere } r=1}, \end{aligned}$$

where the factor r_0^2 scales up lengths in the geometry by the scale factor r_0 compared to the unit sphere. The sphere is a surface of revolution, and by defining $(r, \theta, R(r)) = (r_0 \theta, \phi, r_0 \sin(r/r_0))$, the metric on the sphere of radius r_0 takes the standard form

$${}^{(2)}g = dr \otimes dr + R(r)^2 d\theta \otimes d\theta \equiv g_{\alpha\beta} du^\alpha du^\beta$$

that we can compare with polar coordinates in the plane, remembering of course that now r refers to the rescaled polar angle on the sphere in order to make this comparison. To refer to indexed expressions in the 2-dimensional context, Greek indices will be understood to range from 1 to 2 (namely $u^1 = r$ and $u^2 = \theta$) only to distinguish them from the Latin indices which run from 1 to 3 here. Notice that the azimuthal radius function $R(r)$ has the limiting value $\lim_{r \rightarrow 0} R(r) = r$, so for values of r much less than r_0 (near its North Pole, where it intersects the positive z -axis), the metric is almost the metric of the flat plane expressed in polar coordinates.

On the 2-sphere without reference to the flat 3-dimensional Cartesian coordinates, we could use the general formula which defines the components of the unique covariant derivative for which this 2-dimensional metric $^{(2)}g$ is covariant constant

$$\Gamma^\alpha_{\beta\gamma} = \frac{1}{2}g^{\alpha\delta}(g_{\delta\beta,\gamma} - g_{\beta\gamma,\delta} + g_{\gamma\delta,\beta} + C_{\delta\beta\gamma} - C_{\beta\gamma\delta} + C_{\gamma\delta\beta}), \quad \alpha, \beta = 1, 2$$

applied either in the coordinate frame or in the associated orthonormal frame. However, we have already evaluated all of these components in the 3-dimensional context of spherical coordinates, so we will continue to use the angular coordinates symbols $(u^1, u^2) = (\theta, \phi)$ to specialize the full spherical coordinate formulas to these coordinate surfaces. All we have to do is confine our attention to the components of the covariant derivative in spherical coordinates with no radial indices and set $r = r_0$ in their expressions, which leads to the only nonvanishing coordinate components

$$\Gamma^\theta_{\phi\phi} = -\cos\theta \sin\theta, \quad \Gamma^\phi_{\phi\theta} = \cot\theta = \Gamma^\phi_{\theta\phi}$$

in the coordinate frame $\{e_\theta, e_\phi\}$, while

$$\Gamma^{\hat{\theta}}_{\hat{\phi}\hat{\phi}} = -r_0^{-1} \cot\theta = -\Gamma^{\hat{\phi}}_{\hat{\phi}\hat{\theta}} \equiv \epsilon \kappa \quad (\text{where } \epsilon = \text{sgn}(\cos\theta))$$

are the only nonvanishing components with respect to the associated orthonormal frame $\{e_{\hat{\theta}}, e_{\hat{\phi}}\}$. These corresponding to the relations

$$\nabla_{e_{\hat{\phi}}} e_{\hat{\phi}} = -\epsilon \kappa e_{\hat{\theta}}, \quad \nabla_{e_{\hat{\phi}}} e_{\hat{\theta}} = \epsilon \kappa e_{\hat{\phi}},$$

where

$$\kappa = \epsilon \frac{R'(r)}{R(r)} = \epsilon \frac{\cot(r/r_0)}{r_0} = \epsilon \frac{\cot\theta}{r_0} \leftrightarrow \mathcal{R} = \frac{1}{\kappa} = \frac{r_0}{|\cot\theta|} = \frac{R(r_0\theta)}{|\cos\theta|}$$

are the intrinsic curvature and intrinsic radius of curvature of the azimuthal circles explored in Exercise 8.4.2. Note that the equator at $\theta = \pi/2$ has zero curvature since it is a great circle and therefore a geodesic. Also the unit normal $\epsilon e_{\hat{\theta}}$ points towards the North Pole in the Northern Hemisphere $\theta < \pi/2$, but towards the South Pole in the Southern Hemisphere $\theta > \pi/2$. As shown in Fig. 8.22, the division of the actual radius R of a ϕ coordinate circle by the cosine corresponds exactly to the vertical projection of that radial vector from the z -axis to the coordinate circle onto the tangent plane to the sphere there, stretching that radius from its actual radius in the limit near the North Pole to an infinite radius at the Equator. This exactly captures the parallel transport around one loop of the coordinate circle. If one rolls this “intrinsic osculating circle” in the tangent plane around the coordinate circle without slipping,

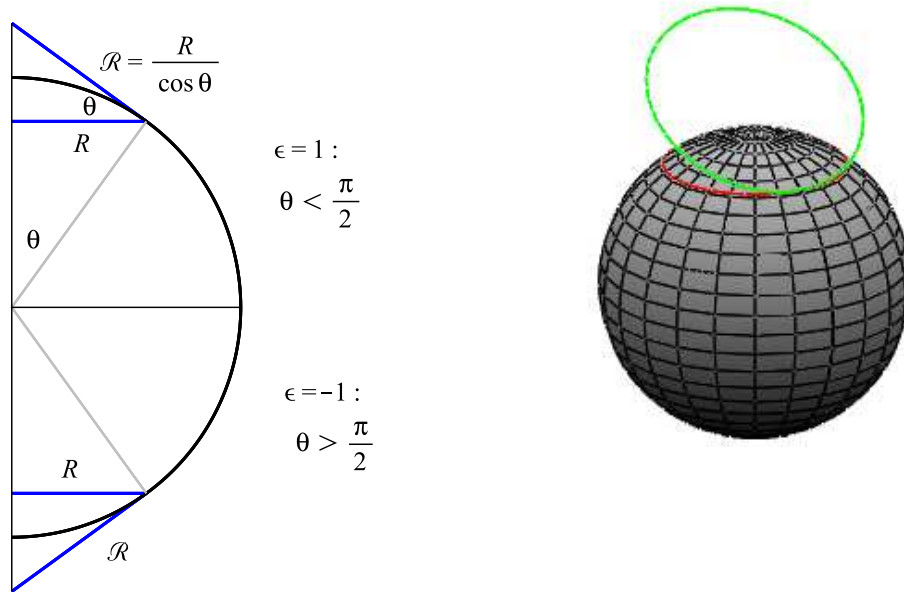


Figure 8.22: The intrinsic radius of curvature \mathcal{R} of the azimuthal coordinate circles of constant θ is stretched from the actual radius R of the circle by its vertical projection onto the tangent plane. The unit normal to those circles points towards the North Pole in the Northern Hemisphere and towards the South Pole in the Southern Hemisphere, while the Equator is a geodesic with infinite radius of curvature bridging the two cases. The mismatch of these two radii leads to the net rotation of a vector under parallel transport around one loop of these coordinate circles.

since its radius is longer than the coordinate circle, the angle it has reached on the intrinsic osculating circle is smaller than 2π by the ratio R/\mathcal{R} , but a parallel transported tangent vector has to rotate in the opposite direction from the orthonormal coordinate frame to compensate for its rotation and by an angle equal to the change of angle on the intrinsic osculating circle relative to the orthonormal coordinate frame, which means it comes up short in the backwards direction, and thus the parallel transported vector has moved forward by the difference with 2π .

Exercise 8.7.1.

intrinsic osculating circle

We can easily parametrize the intrinsic osculating circle shown in Fig. 8.22 at the point $\vec{r}(r, \theta, \phi)$ locating the tangent plane under consideration. If we use the notation $\underline{e}_{\hat{\theta}}$ to denote the 3-vector in the enveloping space \mathbb{R}^3 corresponding to the tangent vector $e_{\hat{\theta}}$ in the tangent

plane, then the center of this intrinsic osculating circle is at the point

$$\underline{C}(r, \theta, \phi) = \vec{r}(r, \theta, \phi) - \epsilon \mathcal{R}(r, \theta, \phi) \underline{e}_{\hat{\theta}},$$

while the osculating circle itself can be parametrized in the standard multivariable calculus way in terms of the radius of curvature \mathcal{R} , the unit normal $\hat{N} = \epsilon e_{\hat{\theta}}$, and the unit tangent $\hat{T} = e_{\hat{\phi}}$, namely

$$\underline{OC}(r, \theta, \phi, t) = \underline{C}(r, \theta, \phi) + \mathcal{R}(r, \theta, \phi)(\cos t \epsilon \underline{e}_{\hat{\theta}} + \sin t \underline{e}_{\hat{\phi}}).$$

Use a computer algebra system to make a plot of the unit sphere and the intrinsic osculating circle at the point $(\theta, \phi) = (\pi/6, 0)$. Rotate it around and zoom in to see the pixels merge with the coordinate circle. This is a perfect example of the fact that we can use any quadratically parametrized curve in \mathbb{R}^3 to approximate the curvature of a given parametrized curve. ■

The nonzero components of the connection correspond to the antisymmetric connection 1-form matrix

$$\underline{\hat{\omega}} = (\omega^{\hat{\alpha}}_{\hat{\phi}\hat{\theta}}) \omega^{\hat{\phi}} = \frac{1}{r_0 \sin \theta} \begin{pmatrix} 0 & -\cos \theta \\ \cos \theta & 0 \end{pmatrix} (r_0 \sin \theta d\phi) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cos \theta d\phi.$$

This just tells us that as we move along the ϕ coordinate lines, the orthonormal frame vectors begin to rotate with respect to parallel transported vectors on the sphere. Near the North Pole $\theta \approx 0$ as shown in Fig. 8.23, this rotation is nearly a rotation by the angle ϕ . On the other hand on the equator $\theta = \pi/2$, the connection 1-form matrix vanishes, and both $e_{\hat{\theta}}$ and $e_{\hat{\phi}}$ are parallel transported along the equator. The fact that the connection 1-forms have no 1-form components along θ tells us that these frames are also parallel transported along the θ coordinate lines.

For very small θ , this picture looks just like the polar coordinate frame in the plane $z = r_0$ tangent to the sphere at the North Pole. You can see that the orthonormal frame rotates by (approximately) the angle ϕ as we move around the ϕ coordinate circle from $\phi = 0$, which is exactly what the antisymmetric matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

describes the rate of change of as we begin moving from any particular ϕ value along the ϕ direction. Parallel transported vectors around this circle, however, try to maintain their direction with respect to the enveloping \mathbb{R}^3 as best they can while still remaining tangent to the sphere.

In general, the components of the covariant derivative tell us how to parallel transport vectors along the $\{\theta, \phi\}$ coordinate lines. In particular the θ coordinate circles (great circles through the poles) and the single ϕ coordinate circle $\theta = \pi/2$ (the equator) are all great circles which we know to be geodesics on the sphere.

The equation

$$\nabla e_{\hat{\theta}} e_{\hat{\alpha}} = \Gamma^{\hat{r}}_{\hat{\theta}\hat{\alpha}} e_{\hat{r}} = 0$$

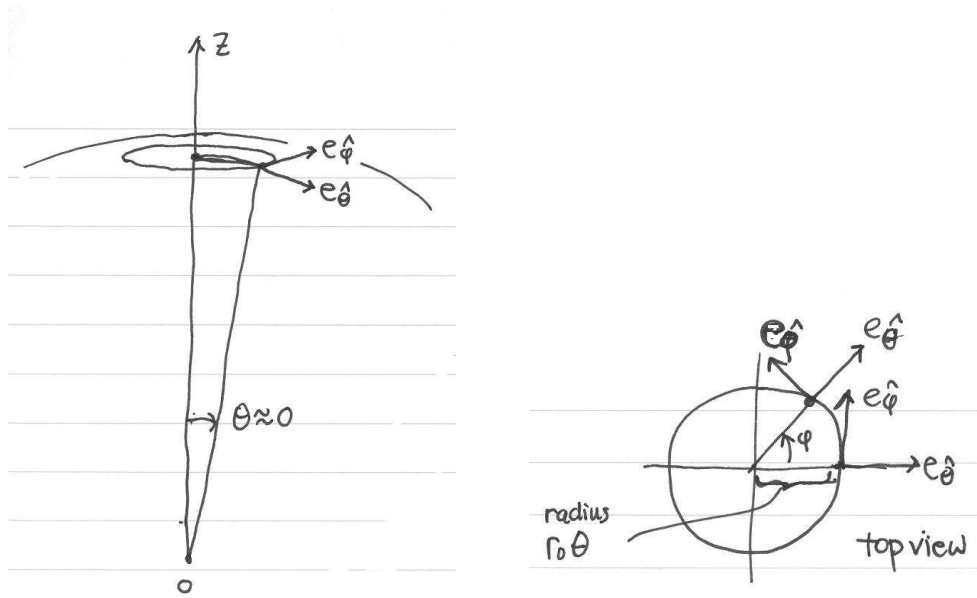


Figure 8.23: Visualizing the covariant derivatives of $e_{\hat{\theta}}$ and $e_{\hat{\phi}}$ along the ϕ coordinate lines near the vertical axis.

tells us that the orthonormal frame is covariant constant along the θ coordinate circles where $e_{\hat{\theta}}$ is a unit tangent vector, so these are also geodesics. The equations

$$\nabla_{e_{\hat{\phi}}} e_{\hat{\theta}} = r_0^{-1} \cot \theta e_{\hat{\phi}}, \quad \nabla_{e_{\hat{\phi}}} e_{\hat{\phi}} = r_0^{-1} \cot \theta e_{\hat{\theta}}$$

tell us the same thing is true for the orthonormal frame along the equator where $\cot \theta = \cot(\pi/2) = 0$, and since $e_{\hat{\phi}}$ is a unit tangent along the equator, $De_{\hat{\phi}}/d\lambda = 0$ confirming that it too is a geodesic.

Thus suppose

$$Y_{(0)} = Y_{(0)}^{\hat{\theta}} e_{\hat{\theta}}|_{\theta=0}$$

is some tangent vector at the North Pole. If we parallel transport it down a line of longitude, its orthonormal components remain constant, i.e., it maintains its length and its angle with the line of longitude.

The same remains true as we move along the equator, and then back up along a line of longitude to the North Pole again, resulting in the final tangent vector $Y_{(f)}$ which has rotated by the increment in ϕ between the two lines of longitude, exactly as we described from intuition.

Notice that this tells us that the 2-sphere cannot admit a covariant constant vector field, since if it did, it would coincide with its parallel transport along every such loop and thus the final value would have to equal the original value of the North Pole, which we have just shown will not happen in general.

Exercise 8.7.2. geodesics on the cylinder

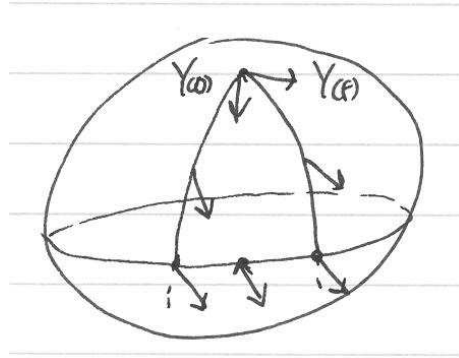


Figure 8.24: Visualizing parallel transport around a particular closed loop on the sphere.

Suppose we do an analogous discussion with a cylinder $\rho = \rho_0$ in cylindrical coordinates $\{\rho, \phi, z\}$ on \mathbb{R}^3 . Then $\{\phi, z\}$ are local coordinates on this 2-dimensional space.

- What are the nonvanishing components of the connection in the coordinate and associated orthonormal frame?
- Do covariant constant vector fields exist?
- Does a covariant constant orthonormal frame exist?
- Write out the geodesic equations

$$\frac{D^2\phi}{d\lambda^2} = 0, \quad \frac{D^2z}{d\lambda^2} = 0.$$

Can you solve these equations? Can you explain the solutions?

If we define $X = r_0\phi$, we obtain orthonormal coordinates (X, z) whose coordinate frame vector fields are covariant constant. These are just Cartesian coordinates on the strip of the flat plane obtained if we cut the cylinder along the line $\phi = \pi$ and flatten it out: $r_0\pi \leq X \leq r_0\pi$, but we must identify the edges of this strip. If we also identified two horizontal lines cutting off this strip into a rectangle, we would obtain the “flat torus.”

Exercise 8.7.3. geodesics on the unit sphere

Investigate the geodesic equations on the unit 2-sphere as a surface of revolution following the example of Exercise 8.4.1, in terms of which $(r, \theta) = (\theta, \phi)$. Figure 8.25 shows the relevant energy plot corresponding to a geodesic starting at the equator at $(\theta, \phi) = (\pi/2, 0)$, rising to a minimum value θ_1 and then falling to a maximum value $\theta_2 = \pi - \theta_1$ in the “radial coordinate” (with respect to the North Pole) θ , while the azimuthal angular coordinate ϕ undergoes one full revolution by an increment of 2π . From the point of view of particle motion, the geodesics are “bound orbits” (orbit = path) trapped in the “potential well.” The point at the bottom of the well, when the energy is $\mathcal{E} = 1/2$, is a stable equilibrium solution representing moving around the equator at fixed $\theta = \pi/2$.

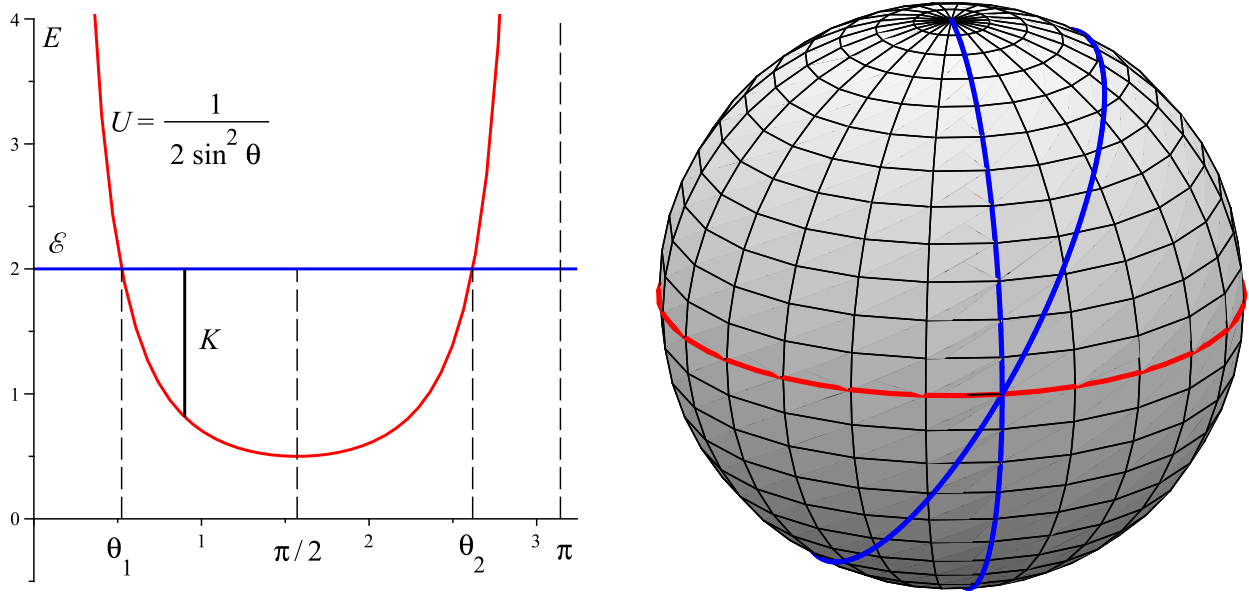


Figure 8.25: A great circle line on the unit sphere not passing through the North Pole, studied using energy considerations, for unit specific angular momentum in the given affine parametrization: $d\phi/d\lambda = 1/\sin^2 \theta$. This geodesic crosses the equator at $\phi = 0$, making an angle β with the vertical direction which is just the polar angle of the point on the great circle closest to the North Pole.

What angle should the initial unit tangent vector make with the vertical in order that the minimum and maximum angles be $(\theta_1, \theta_2) = (\pi/6, 5\pi/6)$ as shown in the potential figure? One can figure this out just by looking down the x -axis so the initial point is at the origin of the projection onto the y - z plane, which is parallel to the tangent plane at the initial point. How can one use the energy equation to find the same result?

■

Exercise 8.7.4. geodesics on the unit sphere: orbit equation

Computer algebra systems still need help sometimes to integrate a simple differential equation that can be done with a bit of change of variable manipulation that we no longer teach but which is occasionally necessary to remember. Use the approach of part c) of Exercise 8.5.1 to find the geodesic orbit equation $\phi = \phi(\theta)$ for the unit sphere and then invert it in the following steps.

- Show that the energy equation re-expressed using θ as the independent variable and

introducing the constant $c = \ell/(2\mathcal{E})$ becomes

$$\begin{aligned} \left(\frac{d\theta}{d\phi}\right)^2 &= \sin^2 \theta (c^{-2} \sin^2 \theta - 1) \leftrightarrow \left(\frac{d\phi}{d\theta}\right)^2 = \csc^4 \theta (c^{-2} - \csc^2 \theta) \\ \rightarrow \phi &= \int \frac{\csc^2 \theta d\theta}{\sqrt{c^{-2} - \csc^2 \theta}}. \end{aligned}$$

b) The numerator is just $du = d \cot \theta = -\csc^2 \theta d\theta$ so it suggests the variable substitution $u = \cot \theta$, for which $\csc^2 \theta = 1 + u^2$. Make this change to arrive at

$$\phi - \phi_e = \int \frac{-du}{\sqrt{a^2 - u^2}} = \arccos(u/a) = \arccos(\cot \theta/a),$$

where $a^2 = c^{-2} - 1 > 0$. Beware if your computer algebra system delivers a more complicated (but equivalent) result involving the arctan. We are ignoring signs here until the final result, inverted to yield

$$\cot \theta = \sqrt{c^{-2} - 1} \cos(\phi - \phi_e) = \cot \theta_e \cos(\phi - \phi_e).$$

The value $\phi = \phi_e$ is the azimuthal coordinate where $\cot \theta_e = a$ which means $\csc \theta_e = c^{-1} = \sqrt{2\mathcal{E}}/\ell = 1/R_e$ which implies that the denominator of the above integrals after the first are zero, and so the reciprocal integrand $d\theta/d\phi = 0$ vanishes, which is at a turning point of the “radial” motion along the meridians at which θ has an extreme value $\theta_e, \pi - \theta_e$. Midway between these lies the equator at $\theta = \theta_0 = \pi/2$, where $\phi = \phi_0 = \phi_e \pm \pi/2, \phi_e \pm 3\pi/2$. Thus if we choose initial conditions at the equator $(\theta, \phi) = (\pi/2, 0)$, we get the two solutions

$$\theta = \operatorname{arccot}(\cot \theta_e \cos(\phi \mp \pi/2)),$$

where $\operatorname{sgn} \cot \theta_e = \operatorname{sgn} \ell$ leads to the four possible paths from the initial data point having the same energy and same absolute value of the angular momentum. (This solution is courtesy of the DAMPT at Cambridge University, and obscure differential geometry texts one might find with much greater effort in some library without being able to use Google.)

c) If $\langle d\theta/d\lambda, d\phi/d\lambda \rangle = \mathcal{E} \langle \cos \beta, \sin \beta \rangle$, express θ_e in terms of the initial value of β_0 at $(\theta, \phi) = (\pi/2, 0)$.

d) Is there any hope of integrating the integral for λ versus θ ? (I have not checked this out yet.)

■

Exercise 8.7.5.

orthonormal coordinate frame connection 1-form matrix by transformation

For the 2-sphere of radius r_0 , the nonvanishing coordinate components of the metric are

$$g_{\theta\theta} = r_0^2, \quad g_{\phi\phi} = r_0^2 \sin^2 \theta$$

and the nonvanishing components of the connection are

$$\Gamma_{\phi\phi}^{\theta} = -\cos \theta \sin \theta, \quad \Gamma_{\phi\theta}^{\phi} = \cot \theta = \Gamma_{\theta\phi}^{\phi},$$

the associated connection 1-form matrix is

$$\underline{\omega} = \begin{pmatrix} \omega^\theta_\theta & \omega^\theta_\phi \\ \omega^\phi_\theta & \omega^\phi_\phi \end{pmatrix} = \begin{pmatrix} 0 & \Gamma^\theta_{\phi\phi} d\phi \\ \Gamma^\phi_{\phi\theta} d\phi & \Gamma^\phi_{\theta\phi} d\theta \end{pmatrix} = \begin{pmatrix} 0 & -\cos\theta \sin\theta d\phi \\ \cot\theta d\phi & \cot\theta d\theta \end{pmatrix}.$$

The normalization of the orthogonal coordinate frame is accomplished by the diagonal rescaling matrix change of frame

$$\underline{A} = \begin{pmatrix} g_{\theta\theta}^{1/2} & 0 \\ 0 & g_{\phi\phi}^{1/2} \end{pmatrix}, \quad (e_{\hat{\theta}} \ e_{\hat{\phi}}) = (e_\theta \ e_\phi) \underline{A}^{-1}, \quad (\omega^{\hat{\theta}} \ \omega^{\hat{\phi}}) = (\omega^\theta \ \omega^\phi) \underline{A}.$$

Use the formula for the differential of the inverse of a matrix stated in Exercise 2.3.6

$$\hat{\omega} = \underline{A} (\underline{\omega} + d\underline{A}^{-1} \underline{A}) \underline{A}^{-1} = \underline{A} \underline{\omega} \underline{A}^{-1} + \underline{A} d\underline{A}^{-1}$$

to evaluate the components of the connection 1-form matrix in the corresponding orthonormal frame, thus reproducing the result stated in the text which was obtained by considering the θ - ϕ components of the connection 1-form matrix for the orthonormal frame associated with all three spherical coordinates. ■

Minkowski geometry

If we change the metric by making the vertical z -axis into the vertical time t -axis through a switch in sign of the self-dot product of vertical vectors, the spheres of constant distance from the origin are mapped to the pseudospheres of constant spacetime interval from the origin, which can be either a spacelike separation or a timelike separation or a null separation, corresponding to the one sheeted hyperboloids, the two-sheeted hyperboloids and the double cone respectively. The null cone has a degenerate induced metric since it contains a null direction and a spacelike direction, which has interesting implications for light rays emanating from a point source.

However, one can use the 3-dimensional Lorentz group (one rotation about the time axis, 2 independent boosts along the two spacelike axes) to understand all the geodesics in terms of more basic ones oriented more simply with respect to the inertial coordinate axes. Any plane through the origin cuts these surfaces in a curve which is a geodesic of the surface in the same way that the planes through the origin in Euclidean space cut spheres in great circles, and the equator (a parallel) and prime meridian (a meridian) provide two simply oriented such geodesics with respect to the Cartesian coordinate system.

For the timelike pseudospheres (hyperboloid of two sheets) inside the light cone, any meridian is a geodesic and they are all equivalent under rotations about the time axis. Boosting these maps them into geodesics which have a minimum distance from the time axis, which corresponds to the effects of the centrifugal potential barrier around that axis in the motion point of view. For the spacelike pseudospheres (hyperboloid of one sheet), the horizontal circular cross-section (“equator”) is a geodesic, and boosting this produces an elliptical-like shape which in the local rest frame of the boost looks like a circle. In fact it is an ellipse in the

Euclidean geometry since the boost scales the lengths in the direction of the boost, which appears to stretch out the circle in the direction of the boost when viewed in the Euclidean geometry, but this actually contracts the circle in that direction in the Lorentzian geometry. This describes all the spacelike geodesics on those spacelike surfaces. The timelike geodesics on the other hand are all essentially like one of the meridians, which pass through the equator and extend out to infinity in both directions, while not making a complete revolution of the time axis. They are in fact hyperbolas. There are two special null hyperbolas corresponding to the two unique null directions at each point of the timelike surface.

unit spacelike pseudosphere

The pseudospheres inside the light cone of \mathbb{M}^3 have a positive-definite induced metric, and are said to be spacelike surfaces since their tangent vectors along the surface are all spacelike. The unit such pseudosphere can be handled nearly exactly as the unit sphere with the change from trigonometric to hyperbolic functions. With the pseudospherical coordinates

$$x = \tau \cosh \chi \cos \phi, \quad y = \tau \cosh \chi \sin \phi, \quad t = \tau \sinh \chi,$$

one evaluates the Minkowski metric inside the light cone to be

$$\begin{aligned} g &= dx \otimes dx + dy \otimes dy - dt \otimes dt \\ &= -d\tau \otimes d\tau + \tau^2 d\chi \otimes d\chi + \tau^2 \sinh^2 \chi d\phi \otimes d\phi \\ &= -\omega^{\hat{\tau}} \otimes \omega^{\hat{\tau}} + \omega^{\hat{\chi}} \otimes \omega^{\hat{\chi}} + \omega^{\hat{\phi}} \otimes \omega^{\hat{\phi}} \end{aligned}$$

with coordinate frame and dual frame

$$e_{\tau} = \frac{\partial}{\partial \tau}, \quad e_{\chi} = \frac{\partial}{\partial \chi}, \quad e_{\phi} = \frac{\partial}{\partial \phi}, \quad \omega^{\tau} = d\tau, \quad \omega^{\chi} = d\chi, \quad \omega^{\phi} = d\phi$$

and associated orthonormal frame and dual frame

$$e_{\hat{\tau}} = \frac{\partial}{\partial \tau}, \quad e_{\hat{\chi}} = \frac{1}{\tau} \frac{\partial}{\partial \chi}, \quad e_{\hat{\phi}} = \frac{1}{\tau \sinh \chi} \frac{\partial}{\partial \phi}, \quad \omega^{\hat{\tau}} = d\tau, \quad \omega^{\hat{\chi}} = \tau d\chi, \quad \omega^{\hat{\phi}} = \tau \sinh \chi d\phi.$$

The pseudospheres are coordinate surfaces $\tau = \tau_0$. We can study the unit pseudosphere $\tau = 1, d\tau = 0$ with metric

$$g^{(2)} = d\chi \otimes d\chi + \sinh^2 \chi d\phi \otimes d\phi$$

where (r, θ) in the surface of revolution discussion becomes (χ, ϕ) here, with $R(\chi) = \sinh \chi$ and an effective potential

$$U_{\text{eff}} = \frac{\ell^2}{2 \sinh^2 \chi}$$

which goes to zero as $\chi \rightarrow \infty$. One can analyze this to obtain an orbit equation for initial data at some nonzero value χ_0 as done in Exercise 8.7.4 for the unit sphere. Let's spare ourselves this calculation.

Exercise 8.7.6.**connection in spacelike pseudospherical coordinates**

Like the unit sphere, all the meridians emanating from the time axis at $\chi = 0$ are geodesics. Similar to that case, one can rotate the inertial coordinate frame on the vertical time axis to align it with the ϕ direction and then boost it to a general χ value to get the full 3-dimensional orthonormal frame matrix $\underline{\mathcal{B}}$ and connection 1-form matrix $\underline{\hat{\omega}} = \underline{\mathcal{B}}^{-1}d\underline{\mathcal{B}}$, from which one can see that the meridians for the pseudospheres are indeed geodesics. Do this. ■

Exercise 8.7.7.**connection on the spacelike pseudosphere**

a) Specialize the previous result to evaluate the orthonormal frame matrix $\underline{\mathcal{B}}^{(2)}$ and obviously antisymmetric connection 1-form matrix $\underline{\hat{\omega}}^{(2)} = \underline{\mathcal{B}}^{(2)-1}d\underline{\mathcal{B}}^{(2)}$ for the unit spacelike pseudosphere.

b) Interpret this result in terms of the intrinsic curvature of the parallels. We can introduce an osculating circle in the tangent space to points on these parallels just as we did for the unit sphere. Use a computer algebra system to show the osculating circle at $(\chi, \phi) = (\text{arccosh } 2, 0)$.

c) Evaluate directly the unnormalized coordinate frame matrix \underline{B} and connection 1-form matrix $\underline{\omega}$ for the full frame, then specialize to the unit spacelike pseudosphere and read off the nonzero connection components, or use the formulas already developed for an arbitrary surface of revolution. ■

unit timelike pseudosphere

The pseudospheres outside the light cone of \mathbb{M}^3 have a Lorentzian induced metric with a timelike direction along the meridians, and are said to be timelike surfaces since their tangent vectors along the surface are both timelike and spacelike. The unit such pseudosphere can be handled like the previous case but taking into account this crucial sign change. With the pseudospherical coordinates ($\ell \geq 0$)

$$x = \ell \sinh \chi \cos \phi, \quad y = \ell \sinh \chi \sin \phi, \quad t = \ell \cosh \chi,$$

one evaluates the Minkowski metric outside the light cone to be

$$\begin{aligned} g &= dx \otimes dx + dy \otimes dy - dt \otimes dt \\ &= d\ell \otimes d\ell - \ell^2 d\chi \otimes d\chi + \ell^2 \cosh^2 \chi d\phi \otimes d\phi \\ &= \omega^{\hat{\ell}} \otimes \omega^{\hat{\ell}} - \omega^{\hat{\chi}} \otimes \omega^{\hat{\chi}} + \omega^{\hat{\phi}} \otimes \omega^{\hat{\phi}} \end{aligned}$$

with coordinate frame and dual frame

$$e_{\ell} = \frac{\partial}{\partial \ell}, \quad e_{\chi} = \frac{\partial}{\partial \chi}, \quad e_{\phi} = \frac{\partial}{\partial \phi}, \quad \omega^{\ell} = d\ell, \quad \omega^{\chi} = d\chi, \quad \omega^{\phi} = d\phi$$

and associated orthonormal frame and dual frame

$$e_{\hat{t}} = \frac{\partial}{\partial \ell}, \quad e_{\hat{\chi}} = \frac{1}{\ell} \frac{\partial}{\partial \chi}, \quad e_{\hat{\phi}} = \frac{1}{\ell \cosh \chi} \frac{\partial}{\partial \phi}, \quad \omega^{\hat{t}} = d\ell, \quad \omega^{\hat{\chi}} = \ell d\chi, \quad \omega^{\hat{\phi}} = \ell \cosh \chi d\phi.$$

Exercise 8.7.8.

connection in timelike pseudospherical coordinates

a) Like the previous case, all the meridians are geodesics. Similarly one can rotate the inertial coordinate frame on the positive x axis to align it with the ϕ direction and then boost it to a general χ value to get the full 3-dimensional orthonormal frame matrix $\underline{\mathcal{B}}$ and connection 1-form matrix $\underline{\hat{\omega}} = \underline{\mathcal{B}}^{-1} d\underline{\mathcal{B}}$, from which one can see that the meridians for the pseudospheres are indeed geodesics. Represent $\underline{\mathcal{B}}$ in terms of the product of these two symmetry operations and a permutation as in the unit sphere case.

b) Interpret your result for the connection 1-form matrix as generating a boost along the meridians and a rotation about the time axis. Interpret the part of the connection 1-form along $d\chi$ for the full spacetime connection as an intrinsic curvature for the meridians. What about the intrinsic curvature of the parallels?

c) The meridians are timelike world lines. For the unit pseudosphere $\ell = 1$, χ is a proper time coordinate along these world lines. Evaluate their spacetime velocity (unit tangent $u = e_{\hat{\chi}}$) and acceleration ($a = \nabla_{e_{\hat{\chi}}} e_{\hat{\chi}}$). Show that they undergo unit acceleration corresponding to their unit curvature, with a spacelike normal vector. Thus one can interpret the unit timelike pseudosphere as the world tube (history) of a circle which accelerates outward from the origin of the spatial inertial coordinates (x, y) . One can introduce an osculating hyperbola for points on the meridians as discussed in Appendix C. ■

Exercise 8.7.9.

geodesics on the unit timelike pseudosphere

The energy equation now takes the form

$$\frac{1}{2} \left(\frac{d\chi}{d\lambda} \right)^2 - \frac{\cosh^2 \chi}{2} \left(\frac{d\phi}{d\lambda} \right)^2 = -\mathcal{E},$$

reversing its overall sign to make the first term positive. Now we have a potential well shown in Fig. 8.26 which traps geodesics with negative energy $\mathcal{E} < 0$ corresponding to closed spacelike geodesics (which are related by a boost to the circular cross-section $\chi = 0$)

$$U_{\text{eff}} = -\frac{\ell^2}{2 \cosh^2 \chi}$$

and also nonclosed timelike geodesics for positive energies $\mathcal{E} > 0$, and finally null geodesics for $\mathcal{E} = 0$. Are there any games we can play here? Perhaps life is too short for this.

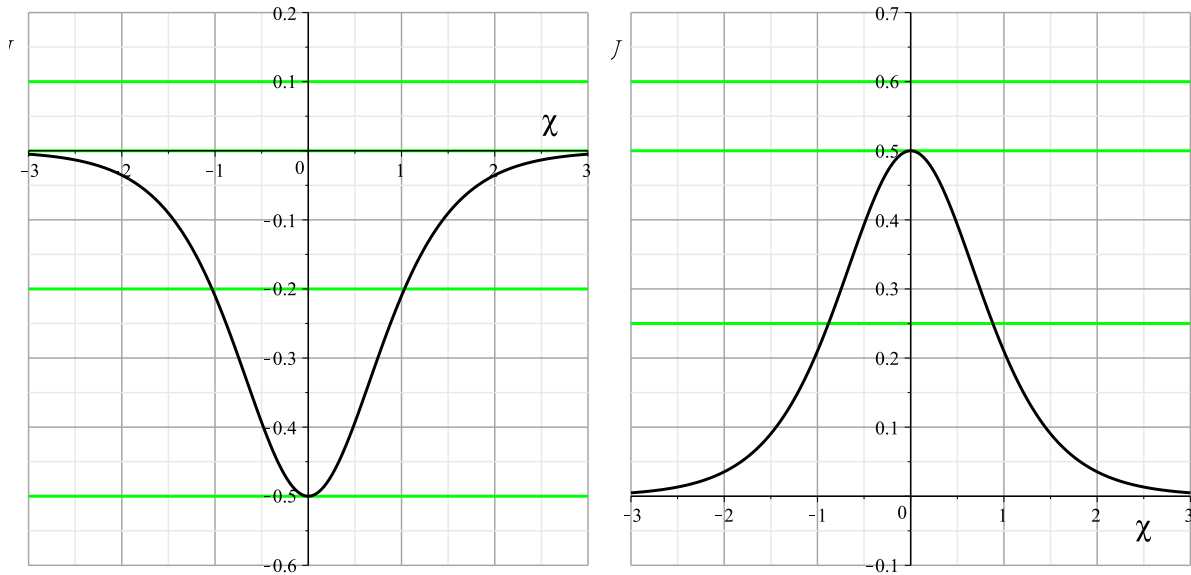


Figure 8.26: **Left:** The centrifugal potential for a unit timelike pseudosphere of one sheet which has an stable equilibrium at the equator. Spacelike geodesics have negative energy in this sign-reversed potential well picture, while timelike geodesics have positive energy and the null geodesics have zero energy.

Right: The corresponding potential for the unit hyperboloid of one sheet in Euclidean space with the same parametrization. Now the equatorial circle is an unstable equilibrium at $\mathcal{E} = 1/2$, and those with less energy do not cross the equatorial circle, while those which have higher energy cross into negative χ values. A single pair of incoming or outgoing geodesics correspond to the energy $\mathcal{E} = 1/2$ which wrap an infinite number of times around the hyperboloid.

**Exercise 8.7.10.****geodesics on the unit hyperboloid of one sheet**

We can compare the previous case with the geodesics on the unit hyperboloid of one sheet in Euclidean space: $x^2 + y^2 - z^2 = \rho^2 - z^2 = 1$. It has the same parametrization but we instead use the Euclidean inner product. Call the circle $z = 0$ the throat of the hyperboloid, which is a geodesic.

a) Using the same surface parametrization as above but renaming t to z , evaluate the induced metric (again using θ for ϕ)

$$ds^2 = \cosh(2\chi) d\chi^2 + \cosh^2 \chi d\theta^2.$$

The energy equation now takes the form

$$\frac{\cosh(2\chi)}{2} \left(\frac{d\chi}{d\lambda} \right)^2 + \frac{\cosh^2 \chi}{2} \left(\frac{d\phi}{d\lambda} \right)^2 = \mathcal{E},$$

with centrifugal potential

$$U_{\text{eff}} = \frac{\ell^2}{2 \cosh^2 \chi}.$$

The change of $g_{\chi\chi}$ from 1 to $\cosh(2\chi) = \cosh^2 \chi + \sinh^2 \chi$ means that an increment of χ corresponds to much greater arclength along the meridian hyperbolas in the Euclidean case, where equal increments in that variable lead to longer and longer segments of the hyperbola in the Euclidean geometry.

b) The two infinite wrapping geodesics with the equator $\chi = 0$ as their asymptote correspond to the conserved quantity $\ell/\sqrt{2\mathcal{E}} = 1 = R_{\min} = R(\chi) \sin \beta$, where $R(\chi) = \cosh \chi$. This offers a test of the numerical accuracy of the computer algebra system solution of the geodesics. Consider initial data at $(x, y, z) = (2, 0, \sqrt{3})$ with $\sin \beta = 1/2$ or $\beta = 5\pi/6$ so that this condition is satisfied. These geodesics will wrap counterclockwise around the vertical axis an infinite number of times as they approach the throat of the hyperboloid at $z = 0$. Since the throat is an unstable equilibrium for the radial motion, there should be a tendency for small numerical errors to effectively change the value of this conserved quantity, so that the numerical solution will either pass through the throat or be reflected away from it. Test this out with a geodesic numerical solver template.

c) **Boomerang game.** By trial and error find an angle which makes a geodesic from the initial point $(x, y, z) = (2, 0, \sqrt{3})$ wrap around the symmetry axis multiple times before returning to the initial point.

d) Investigate the easier instability by starting on the throat with $(uv) = (0, 0)$ and $\beta = 1.0000000001\pi/2$ and 20 digit accuracy. How many loops around the throat occur before the geodesic is flung out suddenly?

**Exercise 8.7.11.****parabola of revolution geodesics**

While we are examining conics of revolution, we might as well consider parabolas of revolution, which come in two varieties depending on whether we rotate the parabola around its symmetry axis or around an axis perpendicular to its symmetry axis

$$\begin{aligned}\langle x, y, z \rangle &= \langle v \cos u, v \sin u, v^2 \rangle, & (\text{circular paraboloid}) \\ \langle x, y, z \rangle &= \langle v^2 \cos u, v^2 \sin u, v \rangle. & (\text{wormhole})\end{aligned}$$

The latter surface pops up as the intrinsic geometry of the equatorial plane of a nonrotating black hole and will be discussed in a subsequent section.

Investigate the geodesics on a circular paraboloid. Can one loop multiple times around the symmetry axis?

**Exercise 8.7.12.****geodesics on ellipse of revolution**

Finally consider an ellipse of revolution, or circular ellipsoid parametrized in Cartesian coordinates by

$$\langle x, y, z \rangle = \langle a \sin v \cos u, a \sin v \sin u, b \cos v \rangle,$$

or in cylindrical coordinates by

$$\rho = a \cos v, \quad \phi = u.$$

Investigate its geodesics. Is there anything interesting you can say about or do with them? Are their special obvious solutions of these geodesic equations determined by reflection symmetry?

Start by deriving its metric

$$ds^2 = a^2 \cos^2 v \, du^2 + (a^2 + (b^2 - a^2) \cos^2 v) dv^2.$$



8.8 The torus

The mathematical torus is the surface that we are all familiar with from eating doughnuts, except in the mathematical context we probably don't want to be reminded of dough, so the alternative spelling donut might be more appropriate. This issue is a moot point if we simply call it a torus.

The torus is a surface of revolution not typically discussed in multivariable calculus, so it makes a good surface to examine after planes, cones, spheres and cylinders. Pseudospheres are definitely not mentioned! Tori also have a hole in them, which makes them interesting for other reasons, and their family is described by two independent parameters so that one can actually change the shape, not possible with spheres or cylinders, although in the latter case if one considers a closed finite cylinder (including its circular disk end caps) with a given height, one can change the shape. Furthermore, the tori exhibit both positive and negative curvature as we will see later.

Like spheres and cylinders on which the orthogonal spherical and cylindrical coordinates are built, tori can also be realized as coordinate surfaces in \mathbb{R}^3 as the foundation of toroidal coordinates, which are obtained using bi-polar coordinates (ξ, ϕ) in the ρ - z coordinate plane of cylindrical coordinates

$$\begin{aligned} x &= (a + b \cos \chi) \cos \phi &= \frac{c \sinh \zeta_0}{\cosh \zeta_0 - \cos \xi} \cos \phi, \\ y &= (a + b \cos \chi) \sin \phi &= \frac{c \sinh \zeta_0}{\cosh \zeta_0 - \cos \xi} \sin \phi, \\ z &= b \sin \chi &= \frac{c \sin \xi}{\cosh \zeta_0 - \cos \xi}. \end{aligned}$$

or directly in terms of cylindrical coordinates,

$$\begin{aligned} \rho &= a + b \cos \chi &= \frac{c \sinh \zeta_0}{\cosh \zeta_0 - \cos \xi}, \\ z &= b \sin \chi &= \frac{c \sin \xi}{\cosh \zeta_0 - \cos \xi}. \end{aligned}$$

The two parameters (a, b) of the standard parametrization of the torus and (c, ζ_0) of the alternative toroidal coordinate parametrization change the size and shape of the torus. The variable ϕ measures the angle by which the circular cross-section is revolved about the symmetry axis, while the extra variable χ or ξ describes those circular cross-sections. Clearly the above unfriendly expressions in bi-polar coordinates suggest that we analyze the torus directly through its induced metric without reference to an adapted coordinate system in which it is a coordinate surface.

The standard surface parametrization follows from the obvious construction of a torus. See Fig. 8.27. One takes a circle of radius $b > 0$ in the ρ - z plane (in cylindrical coordinates, for any angle ϕ) centered at the point a on the ρ -axis and rotates it around the z -axis. For $a > b$ one has a smooth torus with a hole in the middle. In the case $a = b$, the donut hole pinches to a point.

If instead $0 < a < b$, one has a “conical singularity” on the z -axis pointing towards the origin where a limiting tangent cone approximates the surface at those points, which smooths out at $a = 0$ (by opening up into a tangent plane) when a sphere of radius b is obtained by rotating a half-circle about the vertical axis. Instead for $-b < a < 0$ one has conical singularities pointing away from the origin. Finally when $a = -b$ the circle disappears from the ρ - z half-plane so the game is over.

ϕ is the usual azimuthal angular coordinate common to cylindrical and spherical coordinates with range $0 \leq \phi < 2\pi$ or alternatively $-\pi < \phi \leq \pi$, measuring the angle about the vertical axis. χ is an angular coordinate parametrizing the vertical circle being revolved around the vertical axis in the usual way in the counterclockwise direction from the positive cylindrical coordinate ρ direction, with the same range as ϕ .

The torus has an inner equator ($\chi = \pi$) of radius $a - b$ and an outer equator ($\chi = 0$) of radius $a + b$ and a Northern Circle ($\chi = \pi/2$) and a Southern Circle ($\chi = -\pi/2$) instead of North and South Poles like the sphere, both of radius b . The vertical circles of radius a in the ρ - z plane being revolved are parametrized by χ for fixed ϕ . The arclength coordinate $r = b\chi$ measures arclength around these circles starting from the outer equator. Each point on these circles undergoes a circle of revolution about the vertical axis of radius

$$\rho = a + b \cos \chi = a + b \cos(r/b) = a + b \sin(\pi/2 - r/b) = R(r)$$

The Northern and Southern Circles divide the torus into an outer half $-\pi/2 \leq \chi \leq \pi/2$ or equivalently $\cos \chi > 0$, where the coordinate lines are concave inward with respect to the interior of the torus, and an inner half $\cos \chi < 0$, where the ϕ coordinate lines are concave outward with respect to the interior of the torus, while the χ coordinate remains concave inward.

The only natural candidate for a unit torus analogous to the unit sphere corresponds to the parameter values $(a, b) = (2, 1)$ in which a unit circle is revolved around the axis with a unit radius for the inner equator. For the values $(a, b) = (1, 1)$, the torus is pinched so that the hole has collapsed to a point, not a typical torus configuration.

Exercise 8.8.1. toroidal coordinates

The hyperbolic functions are needed to make an orthogonal change of coordinates in the ρ - z half-plane based on two mutually orthogonal families of non-concentric circles which cover that half-plane (2-dimensional bipolar coordinates)

$$\rho = \frac{c \sinh \zeta}{\cosh \zeta - \cos \xi}, \quad z = \frac{c \sin \xi}{\cosh \zeta - \cos \xi}.$$

Solve the expression for z/ρ for $\sin \xi$ and solve the first relation for $\cos \xi$ and use the identity $\cos^2 \xi + \sin^2 \xi = 1$ to eliminate ξ and obtain the equation of a family of circles parametrized by ζ . Clear denominators to obtain the quadratic equation of these circles and then complete the square to find their center a_ζ and radius b_ζ , obtaining the result $(a_\zeta, b_\zeta) = (c \coth \zeta, c \operatorname{csch} \zeta) = (c/\sinh \zeta)(\cosh \zeta, 1)$.

Then repeat the process to instead eliminate ζ by solving for $\cosh \zeta$ and $\sinh \zeta$ and using the identity $\cosh^2 \zeta - \sinh^2 \zeta = 1$. Clear denominators and complete the square to identify the

z -intercept and z -value of the center of the circle on the z -axis represented by the resulting family of circles parametrized by ξ .

Google toroidal coordinates to see what the orthogonal coordinate grid looks like. Fortunately it is not necessary to have a new coordinate system on all of space just to understand the geometry of a single torus—it is enough to consider the parametrization of the surface by the two variables χ and ϕ to get its metric. This has already been done for general surfaces of revolution, so we need only specialize that discussion for the new azimuthal metric scale function.

■

Exercise 8.8.2.

toroidal coordinates for the torus

As a corollary of the previous problem, for the case $\zeta = \zeta_0$ which describes a circle, express the parameters (a, b) for the torus in terms of the parameters (c, ζ_0) .

■

Exercise 8.8.3.

toroidal coordinate metric

a) Replacing ζ_0 by ζ in the toroidal coordinate parametrization of the torus, derive the Euclidean metric expressed in these coordinates. Google toroidal coordinates to check your result. A quick route to this result uses a computer algebra system to evaluate the Jacobian matrix $\underline{J} = (\partial x^i / \partial x^{j'})$ of the coordinate transformation and then evaluating $\underline{J}^T \underline{J}$.

b) Repeat for the less well known right handed local orthogonal coordinate system (β, u, v) involving concentric circles centered on a point $r = a > 0$ on the r -axis in the r - z plane

$$(x, y, z) = ((a + \beta \cos v) \cos u, (a + \beta \cos v) \sin u, \beta \sin v).$$

This system has no problem for $0 < \beta < a$, but begins to have self-intersections starting at $\beta = a$ as the ring tori turn into spindle tori. One finds this coordinate system associated with the Tokamac magnetic confinement of controlled fusion, but it is not one coordinate systems on the famous list of separable coordinate systems.

■

Exercise 8.8.4.

surface of revolution connection

a) Following the general approach for a surface of revolution, evaluate the differentials of the Cartesian coordinates given in the standard parametrization and substitute them into the Euclidean metric, expanding the result and collecting terms to derive the metric on the torus

$$\begin{aligned} g &= \underbrace{b^2 d\chi \otimes d\chi}_{(b d\chi) \otimes (b d\chi)} + \underbrace{(a + b \cos \chi)^2}_{R(r)^2} d\phi \otimes d\phi \\ &= dr \otimes dr + R(r)^2 d\phi \otimes d\phi, \end{aligned}$$

where we have used the simple constant scaling property $b d\chi = d(b\chi) = dr$ so that $b^2 d\chi \otimes d\chi = (b d\chi) \otimes (b d\chi) = dr \otimes dr$. Note that this result for the metric should be obvious since r measures arclength, explaining the component $g_{rr} = 1$, while $\rho = R(r)$ is the radius of the ϕ -coordinate circle needed to convert the angle ϕ into an arclength: $ds_\phi = R(r) d\phi$, and geometrically the surface of revolution construction makes it clear that the two coordinates are orthogonal.

b) Evaluate the formulas developed for a general surface of revolution for the nonzero values of the connection components

$$\begin{aligned}\Gamma^r_{\phi\phi} &= -\frac{1}{2}(R(r)^2)_{,r} = -R(r)R'(r) = (a + b \cos \chi) \sin \chi, \\ \Gamma^\phi_{r\phi} &= \Gamma^\phi_{\phi r} = \frac{R'(r)}{R(r)} = -\frac{\sin \chi}{a + b \cos \chi}.\end{aligned}$$

c) Continuing to apply the surface of revolution formulas, evaluate the connection 1-forms in the related orthonormal frame $e_{\hat{r}} = e_r$, $e_{\hat{\phi}} = R(r)^{-1}e_\phi$ using the product rule while differentiating the rescaled azimuthal frame vector field and combine them into the connection 1-form matrix $\hat{\omega} = (\omega^{\hat{i}}_{\hat{j}})$, which should be antisymmetric. For example, more concretely one has for one of these component calculations

$$\nabla_{e_{\hat{r}}} e_{\hat{\phi}} = \nabla_{e_r} \left(\frac{e_\phi}{R} \right) = -\frac{R'}{R^2} e_\phi + \frac{1}{R} \underbrace{\nabla_{e_r} e_\phi}_{\Gamma^\phi_{r\phi} e_\phi} = \underbrace{\left(-\frac{R'}{R} + \Gamma^\phi_{r\phi} \right)}_{=0 = \Gamma^{\hat{\phi}}_{\hat{r}\hat{\phi}}} \underbrace{\frac{e_\phi}{R}}_{e_{\hat{\phi}}} = 0,$$

which shows that the connection 1-form $\hat{\omega}^{\hat{\phi}}_{\hat{\phi}} = \Gamma^{\hat{\phi}}_{\hat{r}\hat{\phi}} \omega^{\hat{\phi}}_{\hat{\phi}}$ is zero as it should be. Similarly you can evaluate $\nabla_{e_{\hat{\phi}}} e_{\hat{r}}$ and $\nabla_{e_{\hat{\phi}}} e_{\hat{\phi}}$ and show that the corresponding connection 1-forms they determine are related by a minus sign as they should be. Note that we already know that $\nabla_{e_{\hat{r}}} e_{\hat{r}} = \nabla_{e_r} e_r = 0$.

You should find the result

$$\hat{\omega} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \sin \chi d\phi.$$

Remark

This can be directly compared with the connection 1-form matrix on the 2-sphere, and would be identical if one measured the angle χ down from the Northern circle: $\cos(\pi/2 - \chi) = \sin \chi$, at least for the interval $-\pi/2 \leq \chi \leq \pi/2$ on the outer half of the torus, while switching sign on the inner half. First notice that when $\chi = 0, \pi$ on the outer or inner equator, this matrix is zero, indicating that as you move away from a point on an equator, the orthonormal frame does not rotate to first order compared to parallel transport. When $\chi = \pm\pi/2$, one has the maximum value of the rate of change of the rotation. Compare this with the rotation of the cylindrical coordinate orthonormal frame which describes the rotation of the outward horizontal radial direction and the counterclockwise azimuthal direction as one increases ϕ . Here the azimuthal angular coordinate is identical to the cylindrical coordinate, while the increasing radial direction along the Southern rim of the torus at $\chi = -\pi/2$ is horizontal and outward so the result is

exactly the same as in cylindrical coordinates. However, at the Northern Circle at $\chi = \pi/2$, the radial direction is inward (see the figure keeping $r = b\chi$ in mind) and the sign change at the Northern Circle of this matrix reflects the flip in the radial direction compared to the Southern Circle and the usual cylindrical coordinate situation. ■

Exercise 8.8.5.

geodesics on the unit torus

Now specialize to the unit torus $(a, b) = (2, 1)$ shown in Fig. 8.27. Use the analysis of Exercise 8.4.1 for a surface of revolution to study the geodesics with initial data at $r = 0, \phi = 0$. To adopt the notation of that section which included the flat plane in polar coordinates, one must let $\phi \rightarrow \theta$. In other words, just write out the energy equation, the equation of motion for r , and the first order equation for $\theta = \phi$ representing conservation of angular momentum. Verify that the case $\ell = 0$, which corresponds to constant ϕ , i.e., a radial (r) coordinate circle, is a geodesic, which we can refer to as a “radial geodesic.” For convenience, for all the “nonradial” geodesics (nonconstant ϕ), we can set the specific angular momentum equal to 1 so that different initial angles correspond to different specific energies. What is the potential $U(r)$ then explicitly? What is its value at the initial point $(r, \phi) = (0, 0)$ on the outer equator? What is its value on the inner equator?

e) Using the relations of the remark following Exercise 8.4.1 with unit specific angular momentum $\ell = 1$, show that the geodesic with energy $\mathcal{E} = 1/2$ which rises over the Northern Circle and just grazes the inner equator before rising up again corresponds to the initial angle $\beta = \arcsin(1/3) = \arctan(1/\sqrt{8}) = \arccos(\sqrt{8}/3) \approx 19.74^\circ$. What energy corresponds to the seven loop geodesic shown in figure 8.28?

f) Investigate the periodic orbits numerically by trial and error. Find the first few angles admitting 1 or more loops, or at least one such periodic orbit with a low number of loops through hole, returning to the starting point.

g) Think of something fun to study with this problem if you have time. ■

A fun game is to use a numerical geodesic plotter to find by trial and error the special initial angles β starting at the outer equator that lead to closed geodesics on a particular torus like the unit torus. Because of the obvious reflection symmetries $\xi \rightarrow -\xi, \phi \rightarrow -\phi$, it is enough to consider initial angles $0 \leq \beta \leq \pi/2$ to classify these orbits. To get an idea what we are facing, it pays to first look at the corresponding problem for the so-called flat torus.

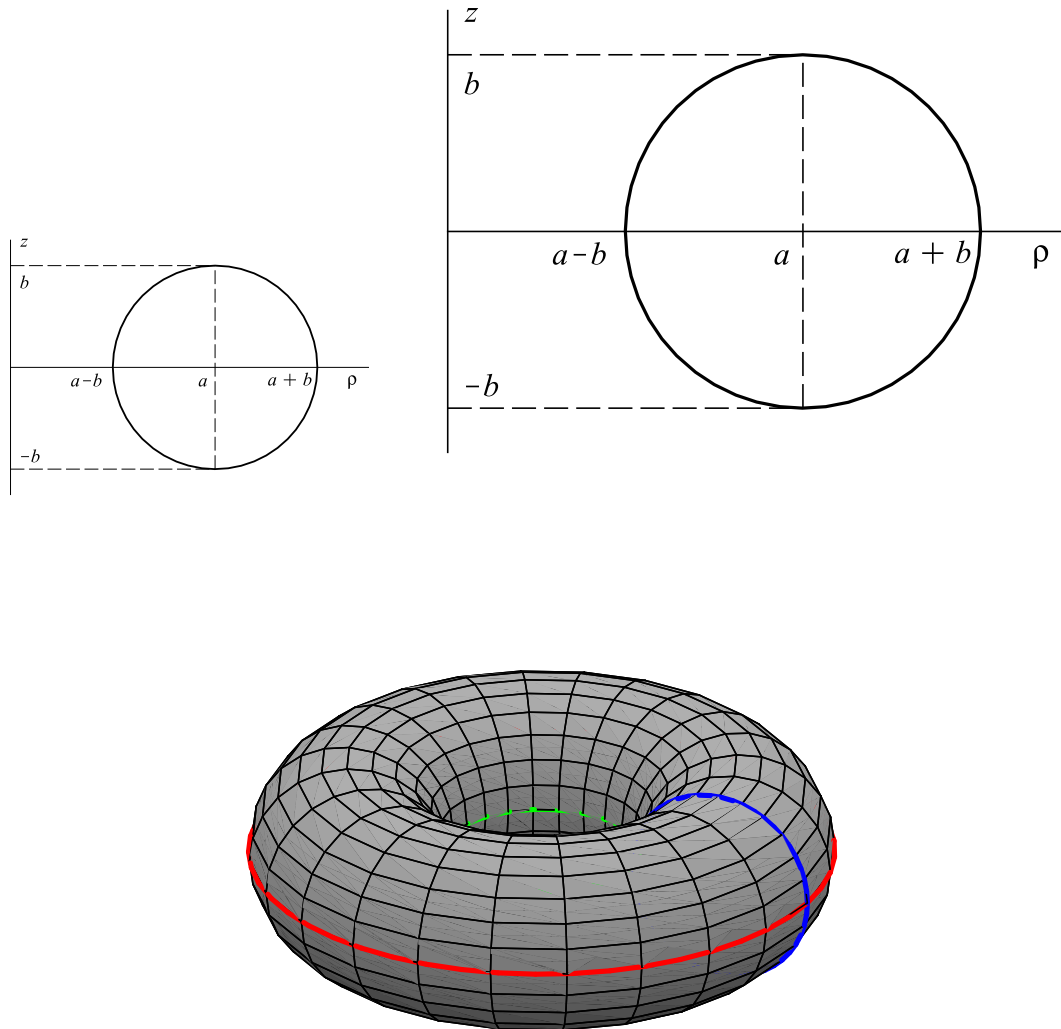


Figure 8.27: The torus is a surface of revolution obtained by rotating about the z -axis a circle in the ρ - z plane with center off the axis. Illustrated here is the case $(a, b) = (2, 1)$ of a unit circle which is revolved around the axis, with an inner equator of unit radius. The outer and inner equators are shown together with the “Greenwich line of longitude” or “Prime Meridian” $\phi = 0$. The Northern and Southern Circles correspond to the North and South Poles on the sphere.

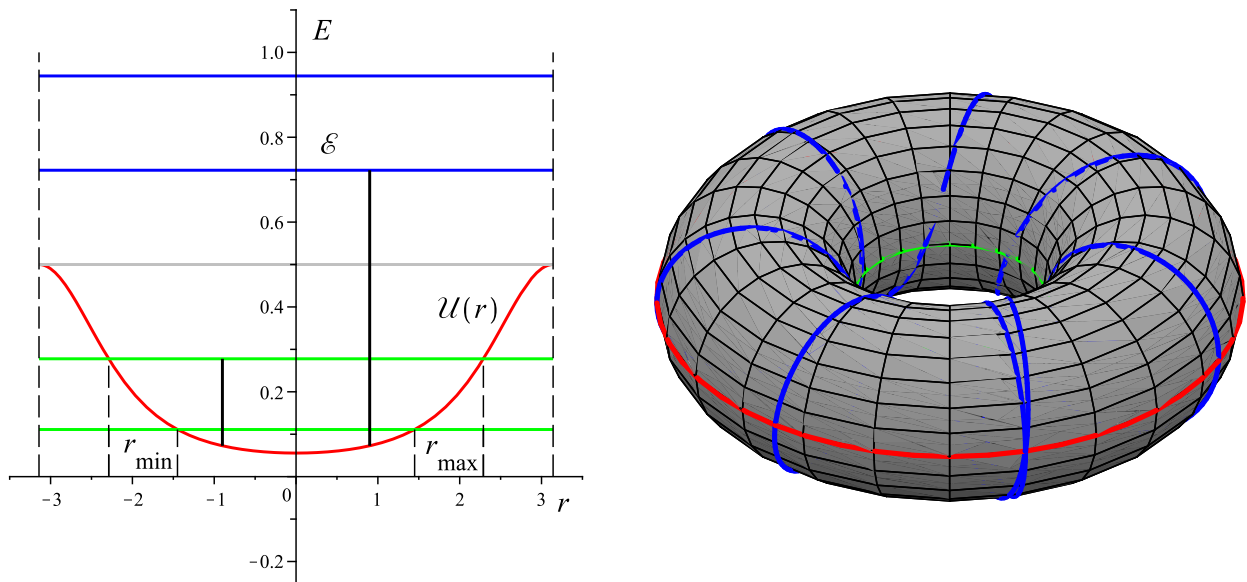


Figure 8.28: The unit specific angular momentum potential (i.e., for $\ell = 1$) for “nonradial” torus geodesics starting on the outer equator and an experimentally found periodic orbit with 7 loops, corresponding to an initial angle of $\beta \approx 0.119$. If the energy is less than $1/2$, the geodesics do not make it through the hole. Of course the potential is periodic as well, and we have only shown one period here: $-\pi \leq r \leq \pi$. For energy larger than $1/2$, the angle ϕ is unbounded.

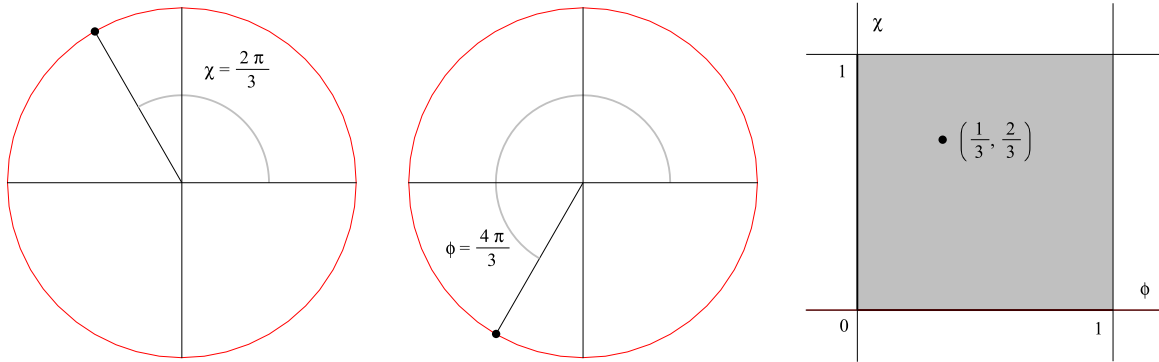


Figure 8.29: The flat torus represents a pair of points on two unit circles, mapped onto the flat Euclidean plane by assigning any pair of corresponding angles (measured in units of the circumference 2π (“revolutions”). Angles outside the interval $[0, 2\pi)$ in radians, or $[0, 1)$ in revolutions lead to points outside the fundamental unit square $[0, 1) \times [0, 1)$ and can be re-identified with a point in that square by adding to each angle an appropriate integer multiple of the circumference (2π radians = 1 revolution).

Example 8.8.1. As illustrated in Fig. 8.29, consider the cross-product set $S_1 \times S_1$ consisting of all pairs of angles (χ, ϕ) with each angle identified modulo multiples of 2π , describing a pair of points, one on each of two unit circles. By measuring angles in units of the circumference 2π , we get a unique correspondence between all points in the interior of the fundamental unit square plus two edges: $0 \leq \chi < 1$, $0 \leq \phi < 1$. Thus the unit Cartesian coordinate grid divides up the plane into an infinite number of repeated unit squares which are identified with each other in this way. The straight lines on this plane then project onto a sequence of line segments crossing the unit square all with the same slope, obtained by resetting each angle appropriately as it reaches the edge of the fundamental unit square. For example, if we consider a straight line through the origin with positive slope, when that slope is irrational, it will never return to the origin, while if it is rational p/q , after a length $L = (p^2 + q^2)^{1/2}$, it will have made p crossings of the unit square in the vertical direction and q crossings in the horizontal direction. Figure 8.30 illustrates this for slope $3/2$.

Thus the flat geometry of the infinite Euclidean plane is inherited by this flat torus and the straight lines are its geodesics, some of which are closed, returning to their starting point after a finite length and then retracing their path, and others which are infinite, never returning to their starting point. The shortest such closed geodesics have length 1: the horizontal and vertical lines. If one considers all geodesics emanating from the origin with positive slope, the closed ones occur in pairs with rational slopes p/q or q/p (corresponding to a reflection symmetry across the diagonal line with slope 1, in turn representing the interchange of the two angles) and length $L = (p^2 + q^2)^{1/2}$, and hence can be partially ordered by their lengths. (Can the same squared integer be represented in two different ways as the sum of the square of a pair of integers $L^2 = p^2 + q^2 = r^2 + s^2$ so that more than two such closed geodesics have the same length? This is a question for number theory.) Slopes p/q in which p and q have

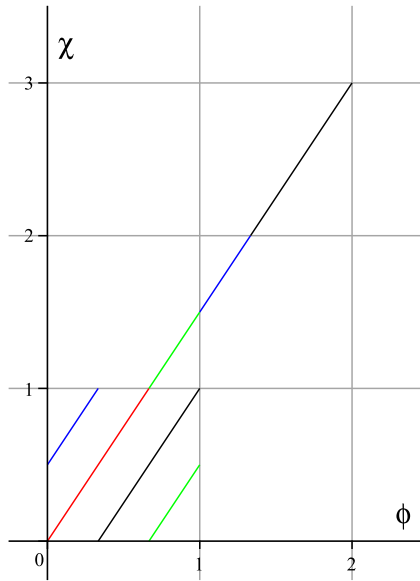


Figure 8.30: The flat torus is obtained by identifying points in the plane differing by a fixed pair of numbers in the respective Cartesian coordinates, here chosen to the $(1, 1)$. Relabeling (x, y) as a pair of angles (χ, ϕ) measured in units of 2π , any point in the plane can be re-identified with the corresponding location in the fundamental unit square $0 \leq \chi < 1$, $0 \leq \phi < 1$. Here the line segment from the origin $(0, 0)$ to $(2, 3)$ is shown projected in this way into this unit square, corresponding to the polar angle $\arctan(3/2) \approx 56.3^\circ$ measured from the horizontal axis, or $\beta = \arctan(2/3) \approx 33.7^\circ$ from the vertical.

common factors correspond to retracings of a closed geodesic in which those common factors are removed and so can be ignored. Fig. 8.31 illustrates this for low values of the periods. Note that each lattice point representing a closed geodesic corresponds to an initial polar angle β measured from the vertical axis.

□

Now return to the unit torus and its geodesic problem, examining the initial values of geodesics starting at $(\chi = r, \phi) = (0, 0)$, making an angle β with the positive vertical axis as seen from within \mathbb{R}^3 . Because of the reflection symmetries (up-down, sideways), we can assume $0 \leq \beta \leq \pi/2$. The geodesics are divided into 6 disjoint sets which can be categorized. Excluding the r -coordinate circle geodesic with $\ell = 0$, we can set $\ell = 1$ and classify the remaining 5 sets by their energy: the outer equator $r = 0$ with energy $\mathcal{E} = 1/18$, those geodesics which do not reach the inner equator with energy $1/18 < \mathcal{E} < 1/2$, the geodesic asymptotic to the inner equator with energy $\mathcal{E} = 1/2$, the inner equator itself $r = \pi$ with energy $\mathcal{E} = 1/2$ which does not correspond to any initial point on the outer equator, and finally those geodesics with energy $\mathcal{E} > 1/2$ which cross the inner equator and hence “pass through the donut hole” an infinite number of times if they are not closed and at some point begin retracing their own path. Similarly the geodesics which do not reach the inner equator with energy $1/18 < \mathcal{E} < 1/2$ circle

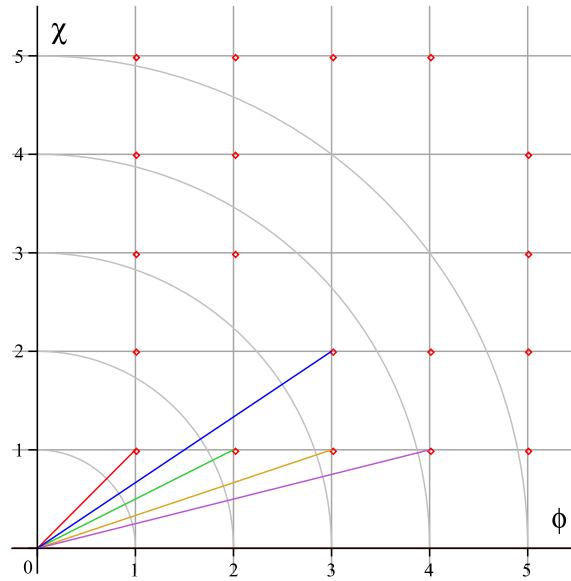


Figure 8.31: The integer pair lattice points (p, q) in the plane correspond to closed geodesics of the flat torus, and can be partially ordered by their length $(p^2 + q^2)^{1/2}$, shared by both pairs (p, q) and (q, p) . Here we show only the cases (p, q) with $p \geq q$ (and no common factors) up to a radius 5 and the corresponding geodesic line segments before re-identification to the fundamental unit square. The lengths and lattice points of the geodesic segments shown are: $\sqrt{2} : (1, 1)$, $\sqrt{5} : (2, 1)$, $\sqrt{10} : (3, 1)$, $\sqrt{13} : (3, 2)$, $\sqrt{17} : (4, 1)$, which defines an interesting infinite sequence 2, 5, 10, 13, 17, 25, 26, 29, 34, 39, ... (more number theory).

the torus an infinite number of times if they are not closed.

The closed geodesics can be described by the pair of wave numbers associated with the separate periods of the individual angular variables: let non-negative integer pairs (m, n) be the number of oscillations in the vertical and horizontal directions respectively during one complete period of the geodesic. For example, $(1, 0)$ describes the r -circle geodesics and $(0, 1)$ the outer equator geodesic. However, there are two kinds of closed geodesics corresponding to $(1, 1)$: those which pass through the hole and those which do not. The same is true of the remaining positive integer pairs. This roughly speaking doubles the number of closed geodesics relative to the same initial value problem for the flat torus. For each pair (p, q) with fixed p , the closed geodesics with energies approaching $1/2$ from above or below can have q values going off to infinity by getting closer and closer to that transitional energy value.

8.9 Geodesics as extremal curves: a peek at the calculus of variations

We have defined geodesics as autoparallel curves. If you imagine a very small toy car moving on a surface with its steering wheel locked so it can only go straight ahead, it would follow a geodesic on that surface. However, the other way of characterizing geodesics is that they are locally the shortest path between two points in a space with a positive-definite metric, or in general, the extremal path in the more general case of an indefinite metric like a Lorentz metric, where the paths actually maximize the arclength.

Optimization problems, or max-min problems for short, are touched upon in multivariable calculus at least for two independent variables. For example, suppose $\vec{r} = \vec{r}(u, v)$ is a parametrized surface in \mathbb{R}^3 , an example of which would be simply the graph of a function of two variables: $z = f(x, y)$, so that $\langle x, y, z \rangle = \langle x, y, f(x, y) \rangle = \langle u, v, f(u, v) \rangle$, for those who never make it to parametrized surfaces. Given some fixed point $P_0(x_0, y_0, z_0)$ not on that surface, one can ask what point $P(x, y, z)$ on the surface is closest to the given point? We try to extremize the distance function $D(u, v) \geq 0$ defined by $D(u, v)^2 = (x_0 - x(u, v))^2 + (y_0 - y(u, v))^2 + (z_0 - z(u, v))^2$. To solve this problem one looks for the critical points of the function D , but since $\partial h(D)/\partial u = h'(D)\partial D/\partial u$, etc., the set of critical points of any function of the distance contain the critical points of D itself, so usually one extremizes the square of the distance to simplify the calculation. Once the critical points are found, one can use the second derivative test to test them for a minimum if there is more than one critical point. At a critical point where the partial derivatives vanish, the differential of the distance function is zero

$$dD(x_0, y_0) = \frac{\partial D(x_0, y_0)}{\partial u} du + \frac{\partial D(x_0, y_0)}{\partial v} dv = 0.$$

The differential represents the first order change in the function, so critical points are points where this first order change vanishes.

Now consider two points P_1 and P_2 on a sphere and ask what differentiable curve $c(\lambda)$ between the two given points minimizes the arclength function? The space of unknowns is now an infinite-dimensional space of all differentiable parametrized curves from P_1 to P_2 and the “objective function” (function to be extremized) is a “functional” (function of functions) on this space, namely the arclength of the curve c

$$A_1(c) = \int_{\lambda_1}^{\lambda_2} ds = \int_{\lambda_1}^{\lambda_2} \frac{ds}{d\lambda} d\lambda = \int_{\lambda_1}^{\lambda_2} |c'(\lambda)| d\lambda.$$

This arclength is clearly independent of the parametrization, symbolically represented by the cancelling of the differential $d\lambda$ in the expression. Alternatively we can extremize the integral of the square of the length of the tangent vector instead of the length itself, but this functional depends on the choice of parametrization

$$A_2(c) = \int_{\lambda_1}^{\lambda_2} \frac{1}{2} |c'(\lambda)|^2 d\lambda = \int_{\lambda_1}^{\lambda_2} \frac{1}{2} \left(\frac{ds}{d\lambda} \right)^2 d\lambda = \int_{\lambda_1}^{\lambda_2} \frac{1}{2} \left(\frac{ds}{d\lambda} \right) ds,$$

and the factor of a half is included so when we differentiate the square using the power rule, the twos cancel. This parametrization dependence is symbolized here by the additional factor of $ds/d\lambda$ in the integrand with respect to the differential of arclength, a factor which depends on the parametrization. Physicists call these functionals “action integrals,” for reasons we need not go into here. Once we have a way to determine the “critical points” of these functionals, which are curves between the two fixed points, we will see that the same critical points result from each such choice. However, notice that at least in the first case where the arclength function is independent of changes of parametrization, for those directions in the space of parametrized curves which represent the same path between the two given points, the action function does not change in value, so there is additional freedom in changing the parametrized curve that does not correspond to a physical change in the path.

The integrand of such an action integral is called a “Lagrangian”

$$A(c) = \int_{\lambda_1}^{\lambda_2} L(c(\lambda), c'(\lambda)) d\lambda.$$

This Lagrangian function is a function of both the position on the curve (explicitly through the metric, if its components in the coordinate system are not constants, but also implicitly since each tangent vector depends on its location) and of the tangent vector along the curve, or in physics language, it is a function on the space of position and velocity. Mathematically speaking, it is a function of the tangent vectors along the curve, i.e., on the space of all tangent spaces to the space in which the curves are living. This is called the tangent bundle, or “velocity phase space” in the physics language. However, a vocabulary lesson will not help us with our immediate problem, so let’s get back to the details. In the sloppy notation when expressed in terms of a fixed coordinate system x^i on the space in question, like the sphere, for example, the two Lagrangian functions are respectively

$$L_1(c, c') = \left(g_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} \right)^{1/2}, \quad L_2(c, c') = \frac{1}{2} g_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda}.$$

To find the critical points of the action, we find its first order variation by considering variations of the curve c whose endpoints are fixed, so that the variation at the endpoints must vanish. To avoid confusion with the ordinary differentials like dx^i for the independent variables of an ordinary max-min problem, we use a delta $\delta x^i(\lambda)$ for the variation of the coordinate functions along the curve, but if we let (λ_1, λ_2) be the fixed values of the parameters at the fixed endpoints of the curve, then we must have $\delta x^i(\lambda_1) = 0 = \delta x^i(\lambda_2)$. Once we figure out how to calculate the first order variation of the action, we must require that the coefficient of each independent variation $\delta x^i(\lambda)$ for each value of λ in the interval $\lambda_1 < \lambda < \lambda_2$, just as we required the coefficient of each independent differential to vanish in the ordinary max-min problem.

To make this more concrete, one can consider a one-parameter family of curves c_σ , i.e., $c_\sigma(\lambda) = c(\lambda, \sigma)$, or in terms of the coordinate representation: $x^i(\lambda, \sigma)$ so that the velocity (tangent vector) is

$$\frac{dx^i}{d\lambda}(\lambda, \sigma) = \frac{\partial x^i(\lambda, \sigma)}{\partial \lambda}$$

If we have a curve in this family which extremizes the action, it will also extremize the action on this family alone, which makes this into a finite-dimensional problem on more familiar ground. The action becomes a function of a single variable so its derivative with respect to σ must vanish, or equivalently, its differential with respect to σ . Letting $\delta\sigma$ denote the differential in this context, then since partial derivatives commute we have for the variation of the velocity

$$\delta x^i = \frac{\partial x^i}{\partial \sigma} \delta \sigma, \quad \delta \left(\frac{dx^i}{d\lambda} \right) = \frac{\partial}{\partial \sigma} \left(\frac{\partial x^i}{\partial \lambda} \right) \delta \sigma = \frac{\partial}{\partial \lambda} \left(\frac{\partial x^i}{\partial \sigma} \right) \delta \sigma = \frac{\partial}{\partial \lambda} \delta x^i = \frac{d}{d\lambda} \delta x^i,$$

where in the last equality we return to the ordinary derivative notation thinking of σ as a parameter in the functions rather than another independent variable. Now that we understand the variation of our fundamental variables, we can consider the variation of a function of those variables like the Lagrangian, just using the chain rule

$$\begin{aligned} \delta L &= \frac{\partial L}{\partial x^i} \delta x^i + \frac{\partial L}{\partial(dx^i/d\lambda)} \delta \left(\frac{dx^i}{d\lambda} \right) && \text{chain rule} \\ &= \frac{\partial L}{\partial x^i} \delta x^i + \frac{\partial L}{\partial(dx^i/d\lambda)} \frac{d}{d\lambda} \delta x^i && \text{variation of velocity} \\ &= \frac{\partial L}{\partial x^i} \delta x^i + \frac{d}{d\lambda} \left(\frac{\partial L}{\partial(dx^i/d\lambda)} \delta x^i \right) - \frac{d}{d\lambda} \left(\frac{\partial L}{\partial(dx^i/d\lambda)} \right) \delta x^i && \text{integration by parts} \\ &= \left(\frac{\partial L}{\partial x^i} - \frac{d}{d\lambda} \left(\frac{\partial L}{\partial(dx^i/d\lambda)} \right) \right) \delta x^i + \frac{d}{d\lambda} \left(\frac{\partial L}{\partial(dx^i/d\lambda)} \delta x^i \right). && \text{regroup} \end{aligned}$$

The integration by parts refers to using the product derivative rule under the integral sign. Since the limits of integration are not varied, the variation of the action integral is the integral of the variation of the Lagrangian

$$\begin{aligned} \delta A_c &= \int_{\lambda_1}^{\lambda_2} \delta L d\lambda = \int_{\lambda_1}^{\lambda_2} \left[\left(\frac{\partial L}{\partial x^i} - \frac{d}{d\lambda} \left(\frac{\partial L}{\partial(dx^i/d\lambda)} \right) \right) \delta x^i + \frac{d}{d\lambda} \left(\frac{\partial L}{\partial(dx^i/d\lambda)} \delta x^i \right) \right] d\lambda \\ &= \int_{\lambda_1}^{\lambda_2} \delta L d\lambda = \int_{\lambda_1}^{\lambda_2} \left(\frac{\partial L}{\partial x^i} - \frac{d}{d\lambda} \left(\frac{\partial L}{\partial(dx^i/d\lambda)} \right) \right) \delta x^i d\lambda + \underbrace{\frac{\partial L}{\partial(dx^i/d\lambda)} \delta x^i \Big|_{\lambda_1}^{\lambda_2}}_{=0}. \end{aligned}$$

The integral of the derivative term evaluates to the endpoints of the curve where the variation is zero (it is crucial that we insist on $\delta x^i(\lambda_1) = 0 = \delta x^i(\lambda_2)$), so that term which results from integration by parts vanishes. This is the only reason for still teaching integration by parts in a beginning calculus course, since we clearly don't need to teach integration techniques any more at that level apart from changing the variable of integration. This is not a very good reason to continue doing so.

Finally the conclusion. If the variation of this integral is to be zero no matter what the variation $\delta x^i(\lambda)$ is at each λ , apart from the endpoints where it must vanish, the coefficient of this variation must be zero. These are called the Lagrangian equations for the problem

$$\frac{\partial L}{\partial x^i} - \frac{d}{d\lambda} \left(\frac{\partial L}{\partial(dx^i/d\lambda)} \right) = 0.$$

By introducing the so called canonical momentum $p_i = \partial L / \partial \dot{x}^i$, this can be written

$$\frac{dp_i}{d\lambda} = \frac{\partial L}{\partial x^i}.$$

If we think of the curves in this problem as the paths of a point particle moving on the space in question, then we can enlarge our problem to include a potential force acting on the point particle by adding a term to the second action depending only on position and not velocity

$$L = \frac{1}{2} g_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} - \mathcal{U}$$

so that the canonical momentum is unchanged

$$p_i = \frac{\partial L}{\partial (dx^i/d\lambda)} = \frac{\partial}{\partial (dx^i/d\lambda)} \left(\frac{1}{2} g_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} \right) = g_{ij} \frac{dx^j}{d\lambda}$$

and just equals the covariant components of the tangent vector / velocity. Then we get

$$\begin{aligned} \frac{\partial L}{\partial x^i} &= g_{ij,k} \frac{dx^k}{d\lambda} \frac{dx^j}{d\lambda} - \underbrace{\frac{\partial \mathcal{U}}{\partial x^i}}_{= -\mathcal{U}_{,i} = \mathcal{F}_i}, \\ &= -\mathcal{U}_{,i} = \mathcal{F}_i \end{aligned}$$

using the chain rule $df/d\lambda = f_{,k} dx^k/d\lambda$, so the Lagrangian equations are

$$\frac{dp_i}{d\lambda} - \frac{\partial L}{\partial x^i} = \frac{dp_i}{d\lambda} - g_{ij,k} \frac{dx^k}{d\lambda} \frac{dx^j}{d\lambda} - \mathcal{F}_i \equiv \frac{Dp_i}{d\lambda} - \mathcal{F}_i = 0,$$

where the final step just recognizes the covariant form of the covariant derivative geodesic condition as discussed in Exercise 8.3.1. This just says that the covariant derivative of the canonical (specific) momentum equals the covariant component of the (specific) applied force. Note that when the Lagrangian is independent of a particular coordinate x^k , namely $\partial L / \partial x^k = 0$, then the corresponding component of the canonical momentum is constant: $dp_k/d\lambda = 0$.

This simple observation is at the heart of the more general Noether's theorem, a fundamental result for both theoretical physics and the calculus of variations due to Emmy Noether, a woman who taught at Bryn Mawr College after being shut out of a position in the old boy European academic network. When the coordinate is an angular coordinate like the azimuthal coordinate in polar, cylindrical or spherical coordinates, the corresponding canonical momentum component is a component of angular momentum, which is a constant of the motion when the Lagrangian does not depend explicitly on that coordinate. This explains the existence of the constant angular momentum ℓ in our 2-dimensional problem for the geodesics of surfaces of revolution. We have already interpreted this constant in terms of the inner product of the tangent vector to the geodesic with a Killing vector field, a quantity which is constant along geodesics.

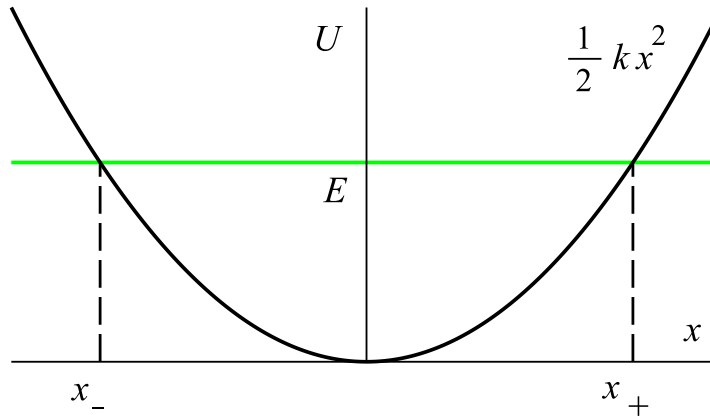


Figure 8.32: The simple quadratic harmonic oscillator potential governs harmonic motion.

Exercise 8.9.1.

Lagrangian equations for geodesics

a) For a metric on a surface which admits an orthogonal coordinate system in which the metric only depends on one variable, like a surface of revolution or a screw-symmetric surface,

$$L = \frac{Z(r)^2}{2} \left(\frac{dr}{d\lambda} \right)^2 + \frac{R(r)^2}{2} \left(\frac{d\theta}{d\lambda} \right)^2,$$

derive the two Lagrangian equations of motion.

b) Show that the equation for θ gives the conservation of the momentum conjugate to that variable.

■

Exercise 8.9.2.

simple harmonic oscillator

a) One of the simplest forces for 1-dimensional motion is the linear Hook's law force exhibited by a spring in its linear regime

$$F = -kx = -\frac{dU}{dx}, \quad U = \frac{1}{2}kx^2$$

acting on a mass m , for which the Lagrangian is

$$L = T - U = \frac{1}{2}m \left(\frac{dx}{dt} \right)^2 + \frac{1}{2}kx^2.$$

This is the simplest potential well shape, just the graph of a parabola at the origin, shown in Fig. 8.32.

b) Show that the energy function $E = T + U$ is a constant for solutions of the equations of motion, and the turning points x_{\pm} of the motion for any positive energy E can be expressed in terms of that energy by solving the quadratic energy constraint with zero dx/dt for its two roots.

c) Show that the general solution is an arbitrary linear combination of sines and cosines of frequency $\omega = \sqrt{k/m}$, namely

$$x = c_1 \cos \omega t + c_2 \sin \omega t = A \cos(\omega t - \delta),$$

where

$$A = (c_1^2 + c_2^2)^{1/2}, \quad \tan \delta = \frac{c_2}{c_1}$$

are called the amplitude and phase shift of the oscillation. Evaluate the energy for these solutions in terms of these quantities. How is the amplitude related to the turning points?

d) When the variable $x = \ell \theta$ in the Lagrangian L is interpreted as the small displacement by an angle θ from vertical equilibrium of a pendulum of mass m and length ℓ hanging in a constant gravitational field, the restoring force of gravity has the form

$$F = -mgx = -mg\ell \theta = -mg\ell \frac{d}{d\theta} \left(\frac{1}{2}\theta^2 \right),$$

this Lagrangian describes a linear pendulum. We need only further set $k = mg/L$ get the corresponding Lagrangian

$$L = \frac{1}{2}m\ell^2 \left(\frac{d\theta}{dt} \right)^2 + \frac{1}{2}mg\ell \theta^2.$$

What is the frequency ω and the period $T = 2\pi/\omega$ of this motion in terms of g and ℓ ? ■

Exercise 8.9.3.

principle of least action for a charged particle

Since we know that the Minkowski spacetime interval is maximized along a straight line, and the Lagrangian principle is often known as the principle of least action (Hamilton's principle), by reversing the sign of the arclength we can make it a minimum action, though negative. Consider a point particle of mass m and charge q moving along an affinely parametrized timelike world line $x^i(\lambda)$ in an electromagnetic field with inertial components $F_{ij} = 2A_{[j,i]} = A_{j,i} - A_{i,j}$. The 1-form $A = A_i dx^i$ is vector 4-potential for the electromagnetic field introduced in Exercise

6.8.7, and is needed to add an interaction term to the free particle Lagrangian to take into account the Lorentz force law exerted on it by the electromagnetic field. This interaction can be described by the following obviously parametrization independent action

$$I = - \int_{\lambda_1}^{\lambda_2} m d\tau + \int_{\lambda_1}^{\lambda_2} q A_i dx^i,$$

since the second term is a “line integral” which does not depend on the parametrization. Here by reversing the generic line element differential we get a real differential of proper time when we take the square root

$$d\tau^2 = -ds^2 = -\eta_{ij} dx^i dx^j \geq 0 \rightarrow d\tau = (-\eta_{ij} dx^i dx^j)^{1/2} = \left(-\eta_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} \right)^{1/2} d\lambda,$$

where $(\eta_{ij} = \text{diag}(-1, 1, 1, 1))$ is the Minkowski metric component matrix. Since $A_i dx^i = A_i (dx^i/d\lambda) d\lambda$, this gives us the Lagrangian

$$L = -m \left(-\eta_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} \right)^{1/2} + q A_i \frac{dx^i}{d\lambda},$$

a) In order to evaluate the Lagrange derivative of the first term note that the canonical momentum is

$$p_k = \frac{\partial L}{\partial(dx^k/d\lambda)} = \frac{m}{d\tau/d\lambda} \eta_{kj} \frac{dx^j}{d\lambda} + q A_k = m \eta_{kj} \frac{dx^j}{d\tau} + q A_k.$$

b) At the next step we will need to take the derivative of the vector potential along the curve, which is a chain rule application

$$\frac{dA_i}{d\lambda} = A_{i,j} \frac{dx^j}{d\lambda}.$$

Use this to combine the two terms involving the derivative of the vector 4-potential into the electromagnetic 2-form to obtain

$$\frac{d}{d\lambda} \left(\frac{\partial L}{\partial(dx^k/d\lambda)} \right) - \frac{\partial L}{\partial x^k} = m \frac{d}{d\lambda} \left(\frac{1}{d\tau/d\lambda} \frac{dx_i}{d\lambda} \right) - q F_{ij} \frac{dx^j}{d\lambda} = 0.$$

c) Since this analysis is independent of the parametrization, we can choose either $\lambda = \tau$ or $\lambda = t$. In the former case $dx^i/d\tau = u^i$ is the unit 4-velocity and this becomes the Lorentz force law

$$m \frac{du^i}{d\tau} = q F^i_j u^j.$$

In the latter case show instead that

$$\frac{dx^i}{d\lambda} = \frac{dx^i}{dt} = \left(\frac{dt}{d\tau} \right)^{-1} \frac{dx^i}{d\tau} = \gamma^{-1} \frac{dx^i}{d\tau},$$

which rescales the proper time derivative by the gamma factor

$$\gamma = \frac{dt}{d\tau} = \left(-\eta_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt} \right)^{-1/2} = (1 - \delta_{ij} v^i v^j)^{-1/2}$$

here expressed in terms of the 3-velocity $v^i = \gamma^{-1} u^i = dx^i/dt$. The equations of motion are then

$$m \frac{d}{dt} \left(\gamma \frac{dx^i}{dt} \right) = q F^i{}_j \frac{dx^j}{dt} = q(E^i + [v \times B]^i).$$

d) Notice that the interaction term in the coordinate time parametrization is

$$q A_i \frac{dx^i}{dt} = q(-\phi + A_a v^a),$$

which reduces to minus the potential energy $q\phi$ of the charge in a conservative electric field when the vector 3-potential vanishes. Consider the quadratic kinetic energy Lagrangian which is only valid for affinely parametrized curves, keeping the same interaction term but replacing the previous negative kinetic energy function by half the mass times the self-inner product of the velocity

$$L^{(2)} = \frac{1}{2} m \eta_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} + q A_i \frac{dx^i}{d\lambda}.$$

Show that this also produces the correct equations of motion provided that the parameter is linearly related to the proper time along the world line. ■

Exercise 8.9.4.

spherical pendulum: gravity as geometry

The spherical pendulum generalizes a 1-dimensional pendulum to allow the mass m to move in two independent directions under the influence of gravity, namely on the surface of a sphere at a fixed radius r_0 from the point at which the pendulum mass is hung.

a) The Lagrangian is the kinetic energy function for motion on a sphere, while the potential energy is mg times the height above the lowest point on the sphere (the South Pole)

$$\begin{aligned} I &= \int T - U dt = \int \frac{1}{2} m r_0^2 \left(\left(\frac{d\theta}{dt} \right)^2 + \sin^2 \theta \left(\frac{d\phi}{dt} \right)^2 \right) - mg(1 - \cos \theta) dt \\ &= \frac{mgr_0}{\omega_0} \int \frac{1}{2} \left(\left(\frac{d\theta}{dT} \right)^2 + \sin^2 \theta \left(\frac{d\phi}{dT} \right)^2 \right) - (1 - \cos \theta) dT. \end{aligned}$$

Show that the second expression can be obtained by introducing the dimensionless time $T = \omega_0 t$, where $\omega_0 = \sqrt{g/r_0}$ is the frequency of small oscillations about the South Pole, familiar from high school physics, yielding the recognizable period $2\pi/\omega_0 = 2\pi\sqrt{r_0/g}$. By an appropriate choice of the original time units we can simply set $mgr_0/\omega_0 = 1$ to make matters simpler, and

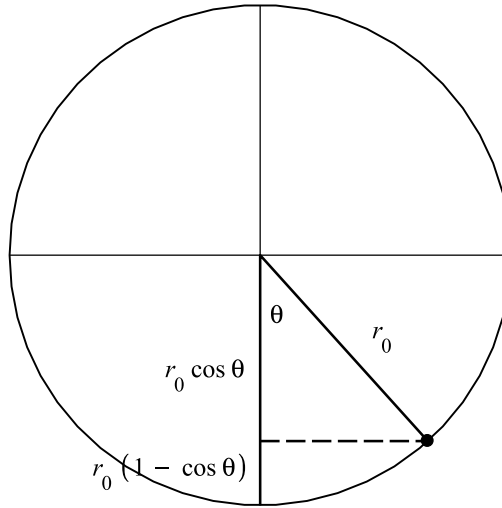


Figure 8.33: The spherical pendulum geometry. The height of the mass from the lowest point on the sphere at the South Pole determines the potential for the restoring gravitational force. Motion along the meridians describes the ordinary 1-dimensional pendulum. Note that the polar angle is being measured from the South Pole here, but the expression for the metric on the sphere remains the same.

an overall constant factor in the Lagrangian does not affect the equations of motion anyway (though it does affect the definition of the canonical momenta!).

The energy then becomes

$$\mathcal{E} = \frac{1}{2} \left(\left(\frac{d\theta}{dT} \right)^2 + \sin^2 \theta \left(\frac{d\phi}{dT} \right)^2 \right) + (1 - \cos \theta).$$

When the mass is at rest (zero kinetic energy), the energy is in fact equal to the potential,

$$\mathcal{E} = (1 - \cos \theta_{\mathcal{E}}),$$

which is the height above the South Pole on the unit sphere. This is the maximum height to which the mass can rise with this energy. We can also introduce the conserved angular momentum

$$\ell = \frac{\partial L}{\partial(\partial\phi/\partial T)} = \sin^2 \theta \frac{d\phi}{dT}$$

and re-express it as

$$\mathcal{E} = \frac{1}{2} \left(\frac{d\theta}{dT} \right)^2 + \frac{\ell^2}{2 \sin^2 \theta} + (1 - \cos \theta) = \frac{1}{2} \left(\frac{d\theta}{dT} \right)^2 + \mathcal{U}_{\text{eff}},$$

giving an effective potential \mathcal{U}_{eff} for the radial (polar angle) motion along the meridians. Show that for a given angular momentum, this potential has a minimum at θ_c determined by

$$\frac{\ell^2}{2 \sin^2 \theta_c} = \frac{\sin^2 \theta_c}{2 \cos \theta_c}.$$

If in addition $d\theta/dT = 0$, this is a circular orbit, and the energy equation then implies

$$\frac{\sin^2 \theta_c}{2 \cos \theta_c} + (1 - \cos \theta_c) = (1 - \cos \theta_{\mathcal{E}}).$$

Show that by converting this entirely to $\mu = \cos \theta_c$, it becomes a quadratic equation in μ for which the plus root is relevant(?), determining the critical angle as a function of the energy \mathcal{E} or equivalently of $\theta_{\mathcal{E}}$. One can then evaluate the angular momentum for that circular orbit.

c) Adding a constant to the Lagrangian also does not affect the equations of motion so add the energy to it in these units

$$I_{\mathcal{E}} = \int \frac{1}{2} \left(\left(\frac{d\theta}{dT} \right)^2 + \sin^2 \theta \left(\frac{d\phi}{dT} \right)^2 \right) + [\mathcal{E} - (1 - \cos \theta)] dT.$$

Now let's change the parametrization of the solution curves to make the potential term constant

$$d\lambda = [\mathcal{E} - (1 - \cos \theta)] dT = (\cos \theta - \cos \theta_{\mathcal{E}}) dT, \quad \cos \theta > \cos \theta_{\mathcal{E}}.$$

This change is valid for all solutions with this energy except those which are at rest at the maximum allowed height (corresponding to the extreme value θ_e of the polar angle) where the related rate between the time and the new parameter vanishes, making the interval in λ shrink as that radius is approached by a moving mass. Show that this transforms the new action to

$$I_{\mathcal{E}} = \int \frac{1}{2} (\cos \theta - \cos \theta_{\mathcal{E}}) \left(\left(\frac{d\theta}{d\lambda} \right)^2 + \sin^2 \theta \left(\frac{d\phi}{d\lambda} \right)^2 \right) + 1 d\lambda.$$

This is a purely kinetic Lagrangian for a new metric on the unit sphere which is “conformally rescaled”

$$g_{\mathcal{E}} = (\cos \theta - \cos \theta_{\mathcal{E}}) g, \quad \cos \theta > \cos \theta_{\mathcal{E}}.$$

The equations of motion for the spherical pendulum with energy \mathcal{E} are the geodesic equations for this deformed sphere, which shortens up the circumferential radius $R(\theta) = \sin \theta$ of the parallels on the sphere by this new factor to become $R_{\mathcal{E}}(\theta) = (\cos \theta - \cos \theta_{\mathcal{E}}) \sin \theta$, squeezing the walls of the sphere inward until they meet the symmetry axis in the embedded surface which has this new metric as its induced metric. This conversion of a conservative force motion problem into a geodesic problem is associated with the key words “Jacobi metric” or “Maupertuis’ principle.”

d) Referring to the discussion of Exercise ??, figure out how to embed the new metric as a surface in \mathbb{R}^3 . TO DO. The deformed sphere should compress towards the symmetry axis so that the circular orbit for a given energy is at the place where the new profile curve has a vertical tangent line. Since constant factors only change the parametrization, one can use instead the conformal factor $(\cos \theta - \cos \theta_{\mathcal{E}})/(1 - \cos \theta_{\mathcal{E}})$ which equals 1 at $\theta = 0$, so that the new surface has the same radius of curvature as the original unit sphere for a better comparison.

e) If this proves too difficult, one can try the Jacobi metric technique for the 2-dimensional rotationally symmetric harmonic oscillator with potential $U = \frac{1}{2}kr^2$ to warm up.

in progress...



The boundary value problem for geodesics

The Lagrangian approach shows that the autoparallel condition implies that the arclength of a geodesic between two fixed points is extremized among all nearby curves. However, there may be multiple geodesics of different length between these points. For example, on a sphere there are two directions one can travel along a great circle containing two points, in the short or long direction. Among all those curves near the long direction, the geodesic will be shortest, but of course the short direction around the same great circle will be shorter. For antipodal points there are an infinite number of great circles connecting the two points, all of the same length. In fact this is an example of focusing of the geodesics by positive curvature. For surfaces which are not of constant curvature like the sphere, we will see how the intrinsic curvature of a surface leads to a focusing length for the family of geodesics which emanate from a given initial point. This is important in modern day astrophysics since the gravitational field acts as a focusing lens for light rays that can reveal extremely useful information about matter in between the emitter and receiver. Gravitational lensing is indeed a whole industry these days.

The problem of finding the shortest distance geodesic between two fixed points for a positive-definite metric is a boundary value problem for the system of differential equations. Such boundary value problems are rarely discussed, perhaps because they are not so easy to discuss. The problem is one of aiming to hit a target. If one fires a rifle at a very distant target, one has to compensate for the falling bullet under the influence of gravity and the drift due to wind. One can calculate how to compensate for gravity, but without knowing the detailed wind profile, one has to make some test shots to basically measure the wind effects on the motion in order to aim with some hope of success. For two points on a surface, one can simply shoot geodesics at different initial angles from one point, using a binary search trial and error method to zone in on a geodesic that passes closer and closer to the target point.

The search for closed geodesics on the torus is exactly this problem, as is the boomerang game for the wormhole-like parabolas of revolution suggested in Exercise 8.7.10. Whenever one has a surface of revolution with an unstable equilibrium circular geodesic, one can use it as a sling shot to fling an incoming grazing geodesic off to any desired target point on the same or opposite sides of that circle. This makes for more interesting target shooting, since one cannot correlate the initial direction with a guess to head for the target point.

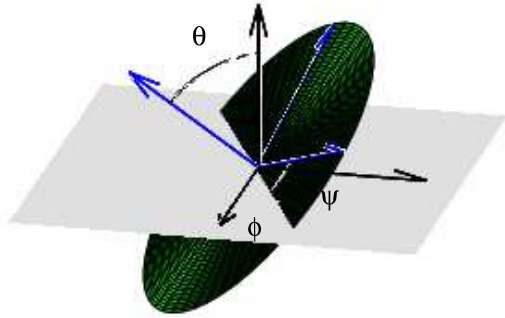


Figure 8.34: The Euler angles for an active rotation of the standard unit vectors (black) in their usual orientation looking down from the first octant, first by an angle ψ about the z -axis, then by a rotation by the polar angle θ about the x -axis, which tilts the circular disk shown, and then by an azimuthal angle ϕ about the z -axis again, which rotates that circular disk about the vertical axis. The new polar axis about which the angle ψ occurs has a direction with polar coordinates $(\theta, \phi - \pi/2)$.

8.10 The rigid body example and $SO(3, \mathbb{R})$

The problem of a symmetric top is a fun illustration of intertwining the group theoretical mathematics of the rotation group with Riemannian geometry and dynamics, and is easily handled by our same approach of an effective potential for 1-dimensional motion after using the symmetries to reduce the system of differential equations to that case. Recall Exercise 1.7.10 in which we described the orientation of a rigid body like a symmetrical top with one point fixed on which it spins by an active rotation $\underline{R} = e^{\phi \underline{k}_3} e^{\theta \underline{k}_1} e^{\psi \underline{k}_3}$ of the space-fixed axes $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ to the time-dependent body fixed axes $\{\hat{e}_{1'}, \hat{e}_{2'}, \hat{e}_{3'}\}$ whose corresponding coordinates are related by $x^{i'} = R(\theta)^{-1i}{}_j x^j$, $x^i = R(\theta)^i{}_j x^{j'}$ and we calculated the components of the angular velocities of the points fixed in the body in both coordinate systems

$$\underline{\Omega}^a = \frac{\omega^a}{dt}, \quad \underline{\Omega}^{a'} = \frac{\tilde{\omega}^a}{dt},$$

where

$$\begin{aligned}(\underline{\omega}^1, \underline{\omega}^2, \underline{\omega}^3) &= (\cos \psi \theta + \sin \theta \sin \psi d\phi, -\sin \psi d\theta + \sin \theta \cos \psi d\phi, d\psi + \cos \theta d\phi), \\(\underline{\tilde{\omega}}^1, \underline{\tilde{\omega}}^2, \underline{\tilde{\omega}}^3) &= (\cos \phi d\theta + \sin \theta \sin \phi d\psi, \sin \phi d\theta - \sin \theta \cos \phi d\psi, d\phi + \cos \theta d\psi)\end{aligned}$$

were defined by

$$\underline{R}^{-1} d\underline{R} = \omega^a \underline{L}_a, \quad d\underline{R} \underline{R}^{-1} = \tilde{\omega}^a \underline{L}_a.$$

If we adopt the physics notation of using an overdot for the time derivative instead of the calculus prime, we get

$$\begin{aligned}(\underline{\Omega}^1, \underline{\Omega}^2, \underline{\Omega}^3) &= (\cos \psi \dot{\theta} + \sin \theta \sin \psi \dot{\phi}, -\sin \psi \dot{\theta} + \sin \theta \cos \psi \dot{\phi}, \dot{\psi} + \cos \theta \dot{\phi}), \\(\underline{\tilde{\Omega}}^1, \underline{\tilde{\Omega}}^2, \underline{\tilde{\Omega}}^3) &= (\cos \phi \dot{\theta} + \sin \theta \sin \phi \dot{\psi}, \sin \phi \dot{\theta} - \sin \theta \cos \phi \dot{\psi}, \dot{\phi} + \cos \theta \dot{\psi}).\end{aligned}$$

In Exercise 1.7.10 followed by Exercise 4.5.7 it was shown that if we left translate the rotation group by left multiplying its matrix \underline{R} by a fixed rotation $\underline{R} \rightarrow \underline{R}_0 \underline{R}$ corresponding to a fixed rotation of the space-fixed axes, then notice that $\underline{R}^{-1} d\underline{R}$ does not change, so the 1-forms ω^a are invariant under left translation of the group into itself. Similarly the 1-forms $\tilde{\omega}^a$ are invariant under right translation of the group into itself, corresponding to a fixed rotation of the body-fixed axes. $x^{i'} \rightarrow x^{i'} = (\underline{R}(\theta) \underline{R}_0)^{-1i}{}_j x^j = R_0^{-1i}{}_j x^{j'}$. Finally it was shown that the bi-invariant metric

$$ds^2 = \frac{a^2}{4} \delta_{ab} \tilde{\omega}^a \tilde{\omega}^b = \frac{a^2}{4} \delta_{ab} \omega^a \omega^b$$

corresponds to the metric on a 3-sphere of radius a in \mathbb{R}^4 , investigated in Exercise 4.5.9 and discussed in Section 6.9.

In Section 2.5 the Cartesian components of the moment of inertia tensor I_{ab} were described and the kinetic energy function introduced for a rigid body of mass M and volume V .

$$T = \frac{1}{2} I_{ab} \Omega^a \Omega^b = \frac{1}{2} I_{a'b'} \Omega^{a'} \Omega^{b'}.$$

The matrix of body-fixed components of the moment of inertia tensor is constant and diagonal if the body-fixed axes are chosen to be the principal axes of that tensor, in which case the kinetic energy corresponds to the length of the tangent vector $\underline{R}'(t)$ to the curve $\underline{R}(t)$ in the rotation group with respect to the time-independent right invariant metric

$$ds_I^2 = \frac{1}{2} I_{a'b'} \tilde{\omega}^a \tilde{\omega}^b.$$

The Lagrangian equations for the kinetic energy Lagrangian describes the free motion of a rigid body about its center of mass, provided the moment of inertia tensor is evaluated about its center of mass. For a spherical body of mass M and radius a , this was evaluated in Exercise 2.5.3 to be $I_{1'1'} = I_{2'2'} = I_{3'3'} = 2Ma^2/5 \equiv I_M$. Thus the motion of this simple body follows the geodesics of the conformally related bi-invariant metric of the unit 3-sphere. The conjugate momenta to the body-fixed components of the angular momentum define the body-fixed components of the angular momentum which differ only by the multiplicative constant I_M , and these are constants by Exercise ??.

8.11 The screw-symmetric helical tube

Okay, I thought the torus was a lot of fun, but who knew that lurking around the corner was an even more interesting surface that contains the family of tori as a special case? Some crazy architect in the UK spent a whole lot of time describing 90 different pasta shapes as parametrized surfaces in Mathematica and published their results in a neat coffee table book in 2011: *Pasta by Design* (by George L. Legendre, not so crazy after all!). As a big pasta fan, I resisted buying this book until I was looking for a way to give a popular talk in the spring of 2012 using some analogy with autoparallel curves on surfaces to give an intuitive idea of how general relativity works, as well as tying the mathematical tools of general relativity to their Italian origins to reflect my life of annual academic commuting to Rome. The cavatappi pasta jumped out at me since it had some helical geodesics like the world lines of circular orbits in spacetime, and the rest just fell into place. Of course I got carried away with that surface too, but will only give a short introduction here. Since this is just a matter of adding one more parameter to the torus family, everything we did so far generalizes to this surface, although it is not a closed compact surface like the torus. Furthermore, the screw symmetry which generalizes the more familiar rotational symmetry is enough different that we can rethink some things we took for granted in the former case.

Starting from the usual Cartesian coordinates (x, y, z) in Euclidean space where $ds^2 = dx^2 + dy^2 + dz^2$, a tubular surface built from a helix can be represented by the following parametrization

$$x = (a + b \cos \chi) \cos \phi, \quad y = (a + b \cos \chi) \sin \phi, \quad z = c \phi + b \sin \chi, \quad (8.1)$$

or equivalently as $\rho = a + b \cos \chi, z = c \phi + b \sin \chi$ in the usual cylindrical coordinates (ρ, ϕ, z) related to the Cartesian ones by $x = \rho \cos \phi, y = \rho \sin \phi, z = z$. It is assumed that $a > b > 0$, and for convenience, that $c > 0$. The “central” helical curve corresponds to setting the radius of the vertical circular cross-section of the tube to zero: $b = 0$, where a is the radius of the cylinder containing that helix, while c is its inclination parameter, with “coiling” angle of inclination $\arctan(c/a)$. Setting $c = 0$ reduces this surface to a torus.

Fig. 8.35 illustrates the construction and one complete revolution of a helical tubular surface with its inner and outer equators marked off. The grid shown in the computer rendition of the surface consists of the constant ϕ circles which result from the intersection of the torus with vertical planes through the symmetry axis (the meridians) and the constant χ helices (the “parallels,” parallel intended in a generalized sense). The Northern ($\chi = \pi/2$) and Southern ($\chi = -\pi/2$) Polar helices correspond to the Northern and Southern Polar circles on the torus which in turn generalize the North and South Poles on the sphere. The radial arc length coordinate $r = b \chi$ and the corresponding angle χ are measured upwards from the outer equator.

Substituting the differentials of these coordinates into the Euclidean metric $ds^2 = dx^2 + dy^2 + dz^2$ to evaluate the induced metric on the surface, one finds easily

$$ds^2 = ((a + b \cos \chi)^2 + c^2) d\phi^2 + 2cb \cos \chi d\phi d\chi + b^2 d\chi^2. \quad (8.2)$$

This is independent of ϕ , which means that it is invariant under translations of ϕ along the family of parallels, termed helical symmetry. The vector field $\partial/\partial\phi$ on the surface which generates

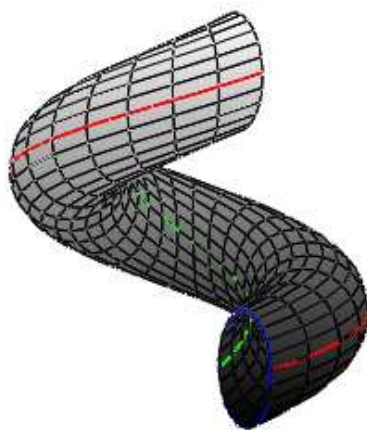
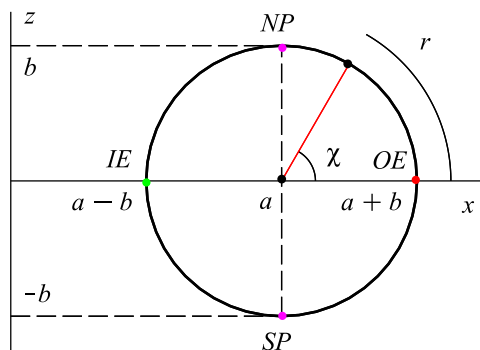


Figure 8.35: A vertical half-plane cross-sectional circle of the helical tubular surface built around a helix through the center of this circle whose axis of symmetry is the z -axis. This circle in the x - z plane is simultaneously rotated around this axis while being translated upwards along that axis ($c > 0$), so that the right hand rule wrapping fingers around the helix in the direction in which it is rising (right figure) puts the thumb up. One can also consider a left-handed helix with $c < 0$. Illustrated in the second figure is one turn $0 \leq u \leq 2\pi$ of the “unit tube” case $(a, b) = (2, 1)$ of a unit circle which is revolved and translated around and along the z -axis, with an inner equator always a unit distance from the axis. The inclination angle of the helix is taken to be that of one version of the smooth cavatappi pasta shape: $\arctan(4/5) \approx 21.80^\circ$. The outer ($\chi = 0$, red) and inner ($\chi = \pm\pi$, green) equators are shown together with the “prime meridian” ($\theta = 0$, blue).

these translations is said to be a Killing vector field of this metric. For motion along geodesic curves within the surface, the component of its affinely parametrized geodesic tangent along the Killing vector field remains constant, a “constant of the motion,” or a conserved momentum associated with this symmetry group. This Killing vector field is just the restriction to the surface of the Killing vector field of the Euclidean metric $\xi = y \partial/\partial x - x \partial/\partial y + c \partial/\partial z = \partial/\partial \phi + c \partial/\partial z$ expressed in either Cartesian or cylindrical coordinates, which generates a rotation about a “screw axis” by an amount which is proportional to a simultaneous translation along the direction of that axis. In cylindrical coordinates these corkscrew rotations are $\rho \rightarrow \rho, \phi \rightarrow \phi + t, z \rightarrow z + ct$. Within the surface, these are just translations in ϕ , so the Killing vector field expressed in the surface coordinates, namely the intrinsic Killing field, is just $\chi = \partial/\partial \phi$.

If we introduce the arclength radial coordinate and cylindrical radius function for the helical center curve

$$r = b\chi, \quad R = a + b \cos(r/b), \quad (8.3)$$

the metric can be written in the following form as well as a second form obtained by completing the square on the differential $d\phi$, for which we give two versions in order to highlight the conserved momentum combination we shall see emerging for geodesic motion

$$\begin{aligned} ds^2 &= (R^2 + c^2) d\phi^2 + 2c \cos(r/b) d\phi dr + dr^2 \\ &= (R^2 + c^2) \left(d\phi + \frac{c \cos(r/b)}{R^2 + c^2} dr \right)^2 + \left(\frac{R^2 + c^2 \sin^2(r/b)}{R^2 + c^2} \right) dr^2 \\ &= \frac{((R^2 + c^2)d\phi + c \cos(r/b) dr)^2}{R^2 + c^2} + \left(\frac{R^2 + c^2 \sin^2(r/b)}{R^2 + c^2} \right) dr^2. \end{aligned} \quad (8.4)$$

Completing the square adapts the metric to the orthogonal decomposition of the tangent space with respect to the intrinsic Killing vector field $\xi = \partial/\partial \phi$. Notice that this metric is invariant under reflections through the origin of the coordinates: $(r, \phi) \rightarrow (-r, -\phi)$. This orthogonal form of the metric allows us to easily read off the metric determinant as the product of the diagonal metric components in this orthogonal 1-form basis

$$\det(g)^{1/2} = \sqrt{R^2 + c^2 \sin^2(r/b)} = \sqrt{(a + b \cos(r/b))^2 + c^2 \sin^2(r/b)}.$$

The differential of surface area can be integrated over one revolution of the surface to provide the surface area of this portion of the surface

$$S = \int \int dS = \int_0^{2\pi} \int_0^{2\pi b} \det(g)^{1/2} dr d\phi = 2\pi \int_0^{2\pi b} \det(g)^{1/2} dr,$$

but the result is an extremely long formula infested with elliptic functions that is of little use to reproduce.

If $(r(\lambda), \phi(\lambda))$ is an affinely parametrized geodesic of this metric on the surface, with tangent

$$U = \frac{dr}{d\lambda} \partial_r + \frac{d\phi}{d\lambda} \partial_\phi = U^r \partial_r + U^\phi \partial_\phi, \quad (8.5)$$

then its orthogonal decomposition is

$$U = U^{\hat{r}} e_{\hat{r}} + U^{\hat{\phi}} e_{\hat{\phi}} \quad (8.6)$$

expressed in terms of its components

$$\begin{aligned} U^{\hat{r}} &= \left(\frac{R^2 + c^2 \sin^2(r/b)}{R^2 + c^2} \right)^{1/2} \frac{dr}{d\lambda}, \\ U^{\hat{\phi}} &= (R^2 + c^2)^{1/2} \left(\frac{d\phi}{d\lambda} + \frac{c \cos(r/b)}{R^2 + c^2} \frac{dr}{d\lambda} \right) \end{aligned} \quad (8.7)$$

with respect to the orthonormal frame

$$\begin{aligned} e_{\hat{r}} &= \left(\frac{R^2 + c^2 \sin^2(r/b)}{R^2 + c^2} \right)^{-1/2} \left(\partial_r + \frac{c \cos(r/b)}{R^2 + c^2} \partial_{\phi} \right), \\ e_{\hat{\phi}} &= (R^2 + c^2)^{-1/2} \frac{\partial}{\partial \phi} \end{aligned} \quad (8.8)$$

whose dual frame is

$$\omega^{\hat{r}} = \left(\frac{R^2 + c^2 \sin^2(r/b)}{R^2 + c^2} \right)^{1/2} dr, \quad \omega^{\hat{\phi}} = (R^2 + c^2)^{1/2} \left(d\phi + \frac{c \cos(r/b)}{R^2 + c^2} dr \right). \quad (8.9)$$

The component of the tangent vector along the Killing vector field is a conserved screw-angular momentum, i.e., a constant along the geodesic

$$\ell = \frac{\partial}{\partial \phi} \cdot U = (R^2 + c^2) \left(\frac{d\phi}{d\lambda} \right) + c \cos(r/b) \left(\frac{dr}{d\lambda} \right) = (R^2 + c^2)^{1/2} U^{\hat{\phi}}, \quad (8.10)$$

which is exactly the combination occurring in the last form of the metric adapted to the orthogonal decomposition with respect to the radial direction along the meridians. In an affine parametrization, the square of the length of the tangent is also a constant, which we will call twice the energy

$$\left(\frac{ds}{d\lambda} \right)^2 = \frac{\ell^2}{R^2 + c^2} + \left(\frac{R^2 + c^2 \sin^2(r/b)}{R^2 + c^2} \right) \left(\frac{dr}{d\lambda} \right)^2 = 2E. \quad (8.11)$$

The length of the tangent vector in an affine parametrization of a geodesic is a constant along that geodesic, with $E = \frac{1}{2}$ for an arclength parametrization in which this length is 1.

If we interpret this problem as geodesic motion in the surface in the physics language of motion in space where the affine parameter λ plays the role of the time (and U is then called the velocity vector and U^r, U^{ϕ} the velocities), E and ℓ are called “constants of the motion.” Since both ℓ and E are constants of the motion, we obtain a single constraint on the square of the “radial velocity” $dr/d\lambda$, or equivalently the orthonormal component $U^{\hat{r}}$, which is of the form

$$\frac{1}{2} (U^{\hat{r}})^2 + V = E, \quad (8.12)$$

where

$$V = \frac{\ell^2}{R^2 + c^2} = \frac{\ell^2}{(a + b \cos(r/b))^2 + c^2} \quad (8.13)$$

acts as an “effective potential” for the radial motion alone. Note that it is an even function of the radial variable r . Qualitatively, this potential leads to the same kind of radial motion as for the special case $c = 0$ of a ring torus thoroughly discussed previously. One can fix the potential for $\ell \neq 0$ by choosing ℓ to have a particular value, and then using the freedom of the affine parametrization to vary the energy E corresponding to energy levels in the graph of the effective potential to describe the allowed interval of radial motion where $E \geq V$. When $E = V$, turning points of the motion occur. These intervals are symmetric about $r = 0$, the outer equator and the corresponding geodesics may be called bound orbits in analogy with the torus problem and more generally 1-dimensional motion in a potential well. If there are no zeros, the motion is unbound with unbounded values of the radial coordinate corresponding to an infinite number of crossings of the inner equator. Local minima of the potential then correspond to stable equilibria, while local maxima correspond to unstable equilibria. The inner and outer equators are themselves geodesics which correspond respectively to a stable and unstable equilibrium.

Note that solving the energy equation for $dr/d\lambda = f(r)$, the problem is reduced to a quadrature

$$\lambda = \int_{r_0}^r f(t) dt \quad (8.14)$$

which must be evaluated numerically, after which the screw-momentum equation can be rewritten for $\phi = \phi(r)$

$$\ell = F(r, dr/d\lambda, d\phi/d\lambda) = F(r, f(r)^{-1}, f(r)d\phi/dr) \quad (8.15)$$

and solved for $d\phi/dr$ and integrated formally to give ϕ as a function of r (only amenable to numerical integration), but this gives no overview of the classification of the orbits which result from the process. The effective potential instead gives a simple visual description of that classification and enables particularly interesting orbits to be studied.

The second order geodesic equations are

$$\begin{aligned} & \mathcal{D} \frac{d^2 r}{d\lambda^2} + \frac{c^2}{b} \cos \frac{r}{b} \sin \frac{r}{b} \left(\frac{dr}{d\lambda} \right)^2 \\ & + 2c \cos \frac{r}{b} \left(a + b \cos \left(\frac{r}{b} \right) \right) \sin \frac{r}{b} \frac{dr}{d\lambda} \frac{d\phi}{d\lambda} \\ & + \left(\left(a + b \cos \frac{r}{b} \right)^2 + c^2 \right) \left(a + b \cos \frac{r}{b} \right) \sin \frac{r}{b} \left(\frac{d\phi}{d\lambda} \right)^2 = 0, \\ & \mathcal{D} \frac{d^2 \phi}{d\lambda^2} - \frac{c}{b} \sin \frac{r}{b} \left(\frac{dr}{d\lambda} \right)^2 + 2 \left(a + b \cos \frac{r}{b} \right) \sin \frac{r}{b} \frac{dr}{d\lambda} \frac{d\phi}{d\lambda} \\ & + c \cos \frac{r}{b} \left(a + b \cos \frac{r}{b} \right) \sin \frac{r}{b} \left(\frac{d\phi}{d\lambda} \right)^2 = 0, \\ & \mathcal{D} = \left(\left(a + b \cos \frac{r}{b} \right)^2 + c^2 \sin^2 \frac{r}{b} \right)^2. \end{aligned} \quad (8.16)$$

Notice that $(dr/d\lambda, d\phi/d\lambda) \rightarrow (-dr/d\lambda, -d\phi/d\lambda)$ is a symmetry of this system which maps $\ell \rightarrow -\ell$, and so its initial data at a given initial position is reflection symmetric about the origin in the initial tangent space.

Exercise 8.11.1.
cavatappo 2.0

If we are going to screw-rotate a circle around the z -axis, why should we start with a vertical circle as in the previous construction? No reason other than a clever pasta architect chose that orientation of the circle. Let's give ourselves a better choice by enlarging our family of surfaces by a tilt back angle for the initial circle that we start the construction with.

The geometry of this construction is made clear if we introduce the unit vectors along the coordinate lines of the cylindrical coordinate system

$$\hat{\rho} = \langle \cos \phi, \sin \phi, 0 \rangle, \quad \hat{\phi} = \langle -\sin \phi, \cos \phi, 0 \rangle, \quad \hat{z} = \langle 0, 0, 1 \rangle.$$

Our initial circle has its horizontal diameter along the positive x -axis but we tilt its initially vertical plane around that axis by an angle ψ , i.e., using the right hand rule with the thumb pointed along the positive x -axis, and the fingers pointing in the direction of the rotation about that axis.

The new parametrized surface using the new notation $(u^1, u^2) = (u, v) = (\phi, \chi)$ for the surface coordinates is

$$\begin{aligned} \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= (a + b \cos(v)) \hat{\rho} + b \sin(v) \left(-\sin(\psi) \hat{\phi} + \cos(\psi) \hat{z} \right) + cu \hat{z} \\ &= (a + b \cos(v)) \begin{pmatrix} \cos(u) \\ \sin(u) \\ 0 \end{pmatrix} \\ &\quad + b \sin(v) \left(-\sin(\psi) \begin{pmatrix} -\sin(u) \\ \cos(u) \\ 0 \end{pmatrix} + \cos(\psi) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right) + cu \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} (a + b \cos(v)) \cos(u) + b \sin(\psi) \sin(v) \sin(u) \\ (a + b \cos(v)) \sin(u) - b \sin(\psi) \sin(v) \cos(u) \\ b \cos(\psi) \sin(v) + cu \end{pmatrix} \tag{8.17} \\ &\equiv \vec{r}(u, v), \tag{8.18} \end{aligned}$$

with $\psi = 0$ reducing this to the previous family. The central helix resulting from the screw-rotation of the center of the initial circle is the curve $b = 0$, with inclination angle $\eta = \arctan(c/a)$ up from the horizontal. We can extend the meridian and parallel terminology to this case in an obvious way associated with this (u, v) coordinate grid, with the special parallels of the inner and outer equators $v = 0, \pi$ and the northern and southern polar helices $v = \pm\pi/2$, and the prime meridian $u = 0$. We call this new improved version of the smooth cavatappo surface the cavatappo 2.0 surface.

a) Evaluate the two tangent vectors along the coordinate lines

$$\vec{r}_1(u, v) = \frac{\partial \vec{r}}{\partial u}(u, v), \quad \vec{r}_2(u, v) = \frac{\partial \vec{r}}{\partial v}(u, v).$$

and their dot products $g_{ij} = \vec{r}_i \cdot \vec{r}_j$ and thus the metric on the surface $ds^2 = g_{ij} du^i du^j$.

b) Show that this metric is much simpler if we choose the tilt back angle $\psi = \eta$ equal to the angle of inclination of the central helix. Call this family the orthogonally tilted cavatappo 2.0 surface. [Hint: express $\cos \eta, \sin \eta$ in terms of the inclination parameters a, c and replace them in the surface parametrization.] What is true about $g_{12} = g_{uv}$ for the orthogonal case that is not true in general? The following explicit parametrization defines the orthogonally tilted cavatappo surface

$$\vec{r}(u, v) = \begin{pmatrix} (a + b \cos(v)) \cos(u) + \frac{bc}{\sqrt{a^2 + c^2}} \sin(u) \\ (a + b \cos(v)) \sin(u) - \frac{bc}{\sqrt{a^2 + c^2}} \cos(u) \\ + \frac{ab}{\sqrt{a^2 + c^2}} \sin(v) + cu \end{pmatrix}. \quad (8.19)$$

Notice that $a = 0$ reduces this to a cylinder of radius a but with a twisting grid, while $c = 0$ reduces this to a torus.

c) Find an orthogonal grid of new parallels which are orthogonal to the meridians by completing the square on dv in the metric to write the metric as a sum of squares and let the one coming from the completed square define the differential of a new radial coordinate $dv_\perp = dv + \dots$ and choose $v_\perp = 0$ at the origin of the original coordinates. Note, however, that if one re-expresses the metric in terms of this new coordinate, it will depend on both coordinates, which makes it less useful for describing the geometry. However, the associated orthonormal frame is useful in describing directions in the tangent space.

d) The metric coefficients of the orthogonal 1-forms obtained in the previous step multiply together to yield the metric determinant. Verify this by evaluating the determinant of the original metric matrix and compare the two expressions. Evaluate the total surface area of the surface for one revolution

$$S = \int \int dS = \int_0^{1\pi} \int_0^{1\pi} \det(g)^{1/2} dv du$$

and show that it is the product of the circumference $C = 2\pi b$ of the orthogonal circular cross-sections and the arclength of the central helix for one revolution

$$L = \int ds = \int_0^{2\pi} g_{uu}^{1/2}|_{b=0} = 2\pi \sqrt{a^2 + c^2}.$$

This generalizes the theorem of Pappus example of the torus surface area to that of a screw-symmetric surface.

e) Explore geodesics on this surface using a computer algebra system. What is the effective potential for this problem?



Exercise 8.11.2.

the Lorentz cavatappo 2.0 surface

We can reconsider the previous orthogonality for the initial circle in the Minkowski geometry on \mathbb{R}^3 in which the z -axis becomes the t -axis. For the case of a timelike central curve, which is then a world line in the spacetime, it makes sense to start with a circle in the horizontal plane and tilt it up with a Lorentz transformation boost, which can then be specialized to the direction orthogonal to the central helical world line of the circle used in the construction, which is the local rest space of the observer following that world line. We then get a timelike tubular world sheet, which is the history of a circular loop moving in spacetime. This is in fact a toy model for string theory, with the horizontal plane cross-sections of the world sheet being the “closed string” as seen at that moment of the observer’s time.

To see the geometry of the construction we need the same cylindrical coordinate unit vectors except we rename z to t and so $\hat{t} = \langle 0, 0, 1 \rangle$. Then tilting the plane of the circle up from the horizontal by the hyperbolic rotation of the angular unit vector $\hat{\phi}$ in its plane with \hat{t} by the hyperbolic angle ι , we get

$$\begin{aligned}
 \begin{pmatrix} x \\ y \\ t \end{pmatrix} &= (a + b \cos(v)) \hat{\rho} + b \sin(v) \left(\cosh(\iota) \hat{\phi} + \sinh(\iota) \hat{t} \right) + cu \hat{t} \\
 &= (a + b \cos(v)) \begin{pmatrix} \cos(u) \\ \sin(u) \\ 0 \end{pmatrix} \\
 &\quad + b \sin(v) \left(\cosh(\iota) \begin{pmatrix} -\sin(u) \\ \cos(u) \\ 0 \end{pmatrix} + \sinh(\iota) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right) + cu \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\
 &= \begin{pmatrix} (a + b \cos(v)) \cos(u) - b \cosh(\iota) \sin(v) \sin(u) \\ (a + b \cos(v)) \sin(u) + b \cosh(\iota) \sin(v) \cos(u) \\ b \sinh(\iota) \sin(v) + cu \end{pmatrix}. \tag{8.20}
 \end{aligned}$$

a) Evaluate the tangent vectors to the grid and their inner products to evaluate the surface metric $ds^2 = g_{ij} du^i du^j$.

b) Express $\cosh \iota$ and $\sinh \iota$ in terms of the tilt parameters a, c such that the tilt angle of the surface equals the hyperbolic inclination angle of the central helix, and replace them in the above expression to obtain the timelike orthogonally tilted cavatappo surface

$$\begin{pmatrix} x \\ y \\ t \end{pmatrix} = \begin{pmatrix} (a + b \cos(v)) \cos(u) - \frac{bc}{\sqrt{c^2 - a^2}} \sin(v) \sin(u) \\ (a + b \cos(v)) \sin(u) + \frac{bc}{\sqrt{c^2 - a^2}} \sin(v) \cos(u) \\ \frac{ab}{\sqrt{c^2 - a^2}} \sin(v) + cu \end{pmatrix}. \tag{8.21}$$

Show how this tilt condition simplifies the surface metric considerably.

c) Find an orthogonal grid of new parallels which are orthogonal to the meridians by completing the square on dv in the metric to write the metric as a difference of squares and let the one coming from the completed square define the differential of a new radial coordinate $dv_{\perp} = dv + \dots$ and choose $v_{\perp} = 0$ at the origin of the original coordinates. The same reservations as in the previous problem apply to re-expressing the metric in terms of this new coordinate.

d) The metric coefficients of the orthogonal 1-forms obtained in the previous step multiply together to yield the metric determinant. Verify this by evaluating the determinant of the original metric matrix and compare the two expressions. Evaluate the total surface area of the surface for one revolution

$$S = \int \int dS = \int_0^{1\pi} \int_0^{1\pi} \det(g)^{1/2} dv du$$

and show that it is the product of the circumference $C = 2\pi b$ of the orthogonal circular cross-sections and the arclength of the central helix for one revolution (proper period of the motion)

$$T_o = \int ds = \int_0^{2\pi} g_{uu}^{1/2}|_{b=0} = 2\pi\sqrt{c^2 - a^2}.$$

This generalizes the previous theorem of Pappus example to the Lorentzian case.

e) Explore geodesics on this surface using a computer algebra system. What is the effective potential for this problem?

■

Remark.

Suppose one considers a spacelike central helix, which corresponds to tachyonic motion at a speed greater than that of light. The orthogonal direction to a spacelike unit tangent vector is timelike, and a circle in a timelike plane does not have any physical interpretation since it corresponds to a closed timelike curve in spacetime, and its equation is not compatible with hyperbolic rotations as well. Thus there is no analog of the orthogonally tilted cavatappo surface in this case. ▲

Exercise 8.11.3.

tilted helical surfaces

We can generalize the helical surfaces of Exercise 8.11.1 by letting $(R(v), Z(v))$ be a parametrized profile curve in the tilted back profile plane taking the place of the profile circle of the generalized cavatappo surface. The general form of such a parametrized surface using the new

notation $(u^1, u^2) = (u, v) = (\phi, \chi)$ for the surface coordinates is

$$\begin{aligned}
 \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= R(v) \hat{\rho} + Z(v) \left(-\sin(\psi) \hat{\phi} + \cos(\psi) \hat{z} \right) + cu \hat{z} \\
 &= R(v) \begin{pmatrix} \cos(u) \\ \sin(u) \\ 0 \end{pmatrix} \\
 &\quad + Z(v) \left(-\sin(\psi) \begin{pmatrix} -\sin(u) \\ \cos(u) \\ 0 \end{pmatrix} + \cos(\psi) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right) + cu \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\
 &= \begin{pmatrix} R(v) \cos(u) + Z(v) \sin(\psi) \sin(u) \\ R(v) \sin(u) - Z(v) \sin(\psi) \cos(u) \\ Z(v) \cos(\psi) + cu \end{pmatrix} \tag{8.22} \\
 &\equiv \vec{r}(u, v). \tag{8.23}
 \end{aligned}$$

We can extend the meridian and parallel terminology to this case in an obvious way associated with this (u, v) coordinate grid.

a) Use a computer algebra system to evaluate the two tangent vectors along the coordinate lines

$$\vec{r}_1(u, v) = \frac{\partial \vec{r}}{\partial u}(u, v), \quad \vec{r}_2(u, v) = \frac{\partial \vec{r}}{\partial v}(u, v).$$

and their dot products $g_{ij} = \vec{r}_i \cdot \vec{r}_j$ and thus the metric on the surface $ds^2 = g_{ij} du^i du^j$.

b) Evaluate the geodesic equations, which are really long and ugly and not manageable expressions. Show that the simultaneous conditions $v'(t) = 0, v = v_0, Z(v_0) = 0, R'(v_0) = 0$ reduce them to setting the second derivatives of both variables to zero, which means that extremals of $R(v)$ on the ρ -axis ($Z(v_0) = 0$) are geodesics, along which u is an affine parameter.

This explains why the inner and outer equators of the generalized cavatappo surfaces are geodesics and shows how a computer algebra system is essential to derive this result. ■

Exercise 8.11.4.

helicoids

If we take a straight line segment in the initial vertical plane ($\psi = 0$) with slope m as the profile curve for a generalized helical surface, we get a helicoid strip

$$R(v) = a + v, \quad Z(v) = cu + mv, \quad a \leq v \leq b,$$

namely

$$\vec{r}(u, v) = \langle (a + v) \cos u, (a + v) \sin u, cu + mv \rangle.$$

a) Evaluate the Gaussian curvature of this surface with a computer algebra system, easily showing that it is a negative curvature surface.

b) Evaluate the geodesic equations, easily showing that the profile curves, as expected, are geodesics.

c) Letting $a = 0$, show that all other geodesics spiral outward from the axis of symmetry in both directions due to the conservation of the screw angular momentum which gives them a minimal value of $R(v)$.

■

Exercise 8.11.5.

cyclides

Another way of arriving at the cavatappo 2.0 surface is by parametrizing a circle of fixed radius in the normal plane to the central helix using the Frenet-Serret frame along the curve. In general for any space curve (the “spine”) one can define a tubular surface in exactly this way, and allowing for a variable circular radius leads to a “canal surface”.

a) Given the helix (see Appendix C)

$$\vec{r}(\phi) = \langle a \cos(\phi), a \sin \phi, c\phi \rangle,$$

with unit tangent $\vec{T}(\phi)$, unit normal $\vec{N}(\phi)$, and unit binormal $\vec{B}(\phi)$, show that

$$\vec{R}(\phi, \theta) = \vec{r}(\phi) + b \left(-\cos \theta \vec{N}(\phi) + \sin \theta \vec{B}(\phi) \right)$$

gives the previous parametrization of the cavatappo 2.0 surface with $(u, v) = (\phi, \theta)$. [CHECK??]

b) Define a cyclide from the spine curve circle

$$\vec{r}(\phi) = \langle a \cos \phi, a \sin \phi, 0 \rangle$$

by

$$\vec{R}(\theta, \phi) = \vec{r}(\phi) + (b + c \cos(\phi)) (\cos(\theta) \langle \cos(\phi), \sin(\phi), 0 \rangle + \sin(\theta) \langle 0, 0, 1 \rangle)$$

Investigate its geodesics for $(a, b, c) = (1, 0.4, 0.3)$. This is a torus with broken symmetry, so one is stuck with both degrees of freedom in the problem. One has meridians $\phi = \phi_0$ but no parallels as in a surface of revolution. Woever, one should be able to classify the geodesics starting from one of two geodesic circle intersection points with the outer equator geodesic, geodesics because of the reflection symmetries across the x - z plane $y = 0$ and the x - y plane $z = 0$, namely $\phi = 0, \pi$, $\theta = 0$. There should be radially bound and unbound geodesics separated by the unstable inner equator geodesic $\theta = \pi$, as in the rotationally symmetric ring torus.

This “Dupin cycloid” shape with a small opening was used by the Italian physicist Tullio Regge for a modern sofa design actually marketed as the Detechma sofa in the 1970s.

■

8.12 The Schwarzschild equatorial plane geometry

A relatively simple geometry governs the orbits of planets around the sun in the general theory of relativity. Since these orbits are planar orbits, we only need 2 space dimensions and of course the time dimension to describe how things move in this geometry, so a 3-manifold. We won't worry about how this particular "Schwarzschild metric" arises from the Einstein equations. We will just use it to play with its geodesics on a very elementary level. Any plane through the origin in spherical coordinates would do for describing the plane of an orbit, but we will choose the equatorial plane $\theta = \pi/2$ of the polar angle down from the vertical axis in spherical coordinates (r, θ, ϕ) , but since we have been doing so many 2-dimensional surfaces using (r, θ) , again we will let θ denote the azimuthal coordinate instead, so that we are just using polar coordinates in the 2-plane of the orbit, plus the time coordinate t . The metric is

$$ds^2 = -(1 - 2m/r) dt^2 + (1 - 2m/r)^{-1} dr^2 + r^2 d\theta^2, \quad r > 2m. \quad (8.24)$$

We have already encountered the planar orbit part of the metric in the black hole embedding discussion, which is also limited to $r > 2m$. The value $2m$ corresponds to the so called horizon of the black hole represented by this metric, inside of which light cannot escape because the gravitational field is too strong. A black hole is a vacuum spacetime analogous to a point mass solution of Newtonian gravitation equations. Spherical mass distributions confined within a radius $R > 2m$ have a gravitational field described by the Schwarzschild metric outside that radius.

Along a timelike geodesic world line where the arclength represents the proper time, we have

$$-d\tau^2 = ds^2 = -(1 - 2m/r) dt^2 + (1 - 2m/r)^{-1} dr^2 + r^2 d\theta^2, \quad r > 2m, \quad (8.25)$$

or

$$-1 = -(1 - 2m/r) \left(\frac{dt}{d\tau} \right)^2 + (1 - 2m/r)^{-1} \left(\frac{dr}{d\tau} \right)^2 + r^2 \left(\frac{d\theta}{d\tau} \right)^2, \quad (8.26)$$

or more generally in an affine parametrization

$$-\mu^2 = -\left(1 - \frac{2m}{r}\right) \left(\frac{dt}{d\lambda} \right)^2 + \left(1 - \frac{2m}{r}\right)^{-1} \left(\frac{dr}{d\lambda} \right)^2 + r^2 \left(\frac{d\theta}{d\lambda} \right)^2. \quad (8.27)$$

There are two symmetry coordinates t and θ on which the metric does not depend, each with its conserved momentum, just the covariant components of the tangent vector along those coordinates (see Section 8.3)

$$\mathcal{E} \equiv -p_t = \left(1 - \frac{2m}{r}\right) \left(\frac{dt}{d\lambda} \right), \quad \ell \equiv p_\theta = r^2 \left(\frac{d\theta}{d\lambda} \right) \quad (8.28)$$

in terms of which previous constant becomes

$$-\mu^2 = -\left(1 - \frac{2m}{r}\right)^{-1} \mathcal{E}^2 + \left(1 - \frac{2m}{r}\right)^{-1} \left(\frac{dr}{d\lambda} \right)^2 + \frac{\ell^2}{r^2}. \quad (8.29)$$

Rearranging this equation one finds

$$\frac{1}{2} \left(\frac{dr}{d\lambda} \right)^2 + \underbrace{\frac{1}{2} \left(1 - \frac{2m}{r} \right) \left(\mu^2 + \frac{\ell^2}{r^2} \right)}_{V(r/m) \equiv \mathcal{V}^2/2} = \frac{1}{2} \mathcal{E}^2 \quad (8.30)$$

so that setting $\mu = 1$, one obtains the starting point for the study of the radial motion carried out in Exercise 8.12.1. Setting $\mu = 0$ extends this to the case of null geodesics, the paths of light rays.

The orbit equation for the variable $u = u/r$, with $\tilde{\ell} = \ell/m$ of that exercise then corresponds to

$$\left(\frac{du}{d\theta} \right)^2 = \frac{\mathcal{E}^2 - (1 - 2u)(\mu^2 + \tilde{\ell}^2 u^2)}{\tilde{\ell}^2} = \underbrace{\frac{\mathcal{E}^2 - \mu^2 + \mu^2 u}{\tilde{\ell}^2}}_{\xrightarrow{\mu \rightarrow 0} \tilde{b}^{-2}} - (1 - 2u)u^2. \quad (8.31)$$

For large radii $u \ll 1$, one can ignore the cubic term $2u^3$ in this differential equation and this reduces to the Newtonian gravity case discussed before Exercise 8.12.1.

A more involved approximation shows how to lowest order the cubic term leads to the precession of the perihelion (minimum radius point of the orbit) of the conic section orbits. This nicely describes the anomalous precession of the perihelion of the orbit of Mercury. The case $\mu = 0$ for null orbits describes instead the deflection of starlight by the Sun, two of the classic tests of general relativity which brought world wide fame to Albert Einstein.

Consider the timelike geodesics with $\mu = 1$ making $\lambda = \tau$ the proper time, and introduce the dimensionless variables: $\tilde{r} = r/m$, $\tilde{\ell} = \ell/m$. The graph of the dimensionless potential for the radial motion is a 1-parameter family of curves

$$V(\tilde{r}) = \frac{1}{2} \left(1 - \frac{2}{\tilde{r}} \right) \left(1 + \frac{\tilde{\ell}^2}{\tilde{r}^2} \right) \rightarrow \frac{1}{2} \text{ as } \tilde{r} \rightarrow \infty.$$

Normally the geodesics are discussed in terms of the relativistic potential $\mathcal{V} = \sqrt{2V}$ which approaches 1 at large radii, but we continue with the equivalent discussion using the potential V as in the nonrelativistic case.

Exercise 8.12.1.

nonrotating black hole orbits

a) Use a computer algebra system to plot V versus \tilde{r} for $0 \leq \tilde{r} \leq 25$, for the values $\tilde{\ell} = 3, 2\sqrt{3}, \sqrt{3} + 2, 4, 4.3685, 4.6937$, narrowing your vertical viewing window to $0.4 \leq V \leq 0.6$ to get an idea of this 1-parameter family of potential curves.

b) The critical points $V'(\tilde{r}_0) = 0$ of V represent the circular orbits at constant radius $\tilde{r} = \tilde{r}_0$. Find the critical points \tilde{r}_{\pm} of V as a function of $\tilde{\ell}$ by solving a quadratic equation for \tilde{r} : $\tilde{r}_{-} \leq \tilde{r}_{+}$. Show that $\tilde{\ell}_{\min} = 2\sqrt{3}$ is the minimum value of the angular momentum parameter for which a critical point exists, for which $\tilde{r}_{-} = \tilde{r}_{+} = 6$, and two distinct roots exist for larger values. Show that $3 \leq \tilde{r}_{-} \leq 6$, $\tilde{r}_{+} \geq 6$. [Hint. Expand \tilde{r}_{-} about $1/\tilde{\ell} = 0$ to find that its high angular

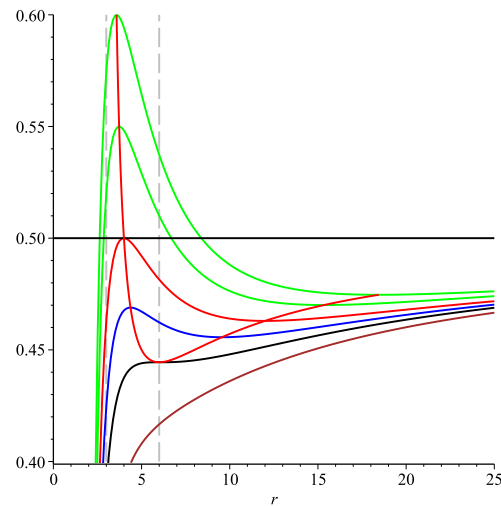


Figure 8.36: Selected values of the family of potential functions $V(r)$ for the radial motion of geodesics in the equatorial plane of the Schwarzschild spacetime, plotted versus the dimensionless radius $\tilde{r} = r/m$. The upper two curves have a centrifugal barrier for hyperbolic-like orbits with relativistic energy parameter $\mathcal{E}^2 > 1$ corresponding to energy levels in this diagram of $\frac{1}{2}\mathcal{E}^2 > \frac{1}{2}$, so that for sufficiently high energy orbits are captured by the black hole. The next lower curve corresponding to the critical level $\mathcal{E} = 1$ (horizontal line at the value 0.5) corresponds to the case in which the unstable circular orbit at the peak of the centrifugal barrier has the same energy as the parabolic-like orbit. Below this some of the bound elliptic-like orbits are also captured until the local maximum and local minimum of the potential come together at $\tilde{r} = 6$ and disappear, below which there are no more circular orbits.

momentum limit $\ell \rightarrow \infty$ is 6.] The smaller root describes the unstable equilibrium at the peak of the potential near the black hole, while the larger root describes the stable equilibrium at the minimum points which exist farther from the whole. The circular orbit at $\tilde{r} = 6$ is called the “last stable circular orbit.”

c) Plot the two parametrized curves $(\tilde{r}_-(\ell), V(\tilde{r}_-(\ell)))$ and $(\tilde{r}_+(\ell), V(\tilde{r}_+(\ell)))$ together with the previous plot of the selected potential profile curves as shown in Fig. 8.36, representing respectively the unstable and stable circular orbits. Include the vertical lines at $\tilde{r} = 3, 6$ as shown.

d) Show that $V(\tilde{r}_+) < 1/2$ corresponding to $\mathcal{E} < 1$. Show that $\ell = 4$ corresponds to $\tilde{r}_- = 4$ and $V(\tilde{r}_-) = 1/2$ or $\mathcal{E} = 1$, and that orbits with $\mathcal{E} \geq 1$ are unbound. Those unbound orbits with $\ell > 4$ and $\mathcal{E}^2/2 > V(\tilde{r}_-)$ overcome the centrifugal potential energy barrier to fall into the black hole, while those with smaller energy are reflected by this barrier.

e) Use a computer algebra system to write down the three proper time parametrized geodesic equations and impose the conditions $r = r_0$, $dr/d\tau = 0 = d^2r/d\tau^2$ and $d^2\theta/d\tau^2 = 0 = d^2t/d\tau^2$ for uniformly rotating circular geodesics, satisfying the time and angular equations. Show that the remaining radial equation determines the coordinate velocity

$$\omega = \frac{d\theta}{dt} = \frac{d\theta/d\tau}{dt/d\tau} = \pm \left(\frac{m}{r^3} \right)^{1/2} \equiv \pm \omega_K.$$

This is the same as the Newtonian Keplerian angular velocity for circular orbits.

f) Derive the alternative radial equation from the potential by

$$\frac{d^2r}{d\tau^2} = -\frac{\partial V}{\partial r},$$

and set the right hand side to zero for circular geodesics and solve for ℓ^2 to find

$$\ell = \pm \left(\frac{mr^2}{r-3m} \right)^{1/2} \equiv \pm \ell_K.$$

Back substitute this into the energy equation for circular geodesics to find

$$\mathcal{E} = \frac{(r-2m)}{(r(r-3m))^{1/2}} \equiv \mathcal{E}_K.$$

Then show that the signed angular speed is

$$\nu = r \frac{d\theta}{d\tau} = \frac{\ell}{r} = \pm \underbrace{\left(\frac{m}{r-3m} \right)^{1/2}}_{\equiv \nu_K}$$

while the corresponding gamma factor is

$$\gamma = (1 - \nu^2)^{-1/2} = \left(\frac{r-2m}{r-3m} \right)^{1/2} \equiv \gamma_K.$$

Finally show that

$$\left. \frac{dt}{d\tau} \right|_K = \left(1 - \frac{2m}{r} \right)^{-1/2} \gamma_K$$

and that the proper time angular velocity is

$$\left. \frac{d\theta}{d\tau} \right|_K = \frac{d\theta}{dt} \frac{dt}{d\tau} = \pm \left(1 - \frac{2m}{r} \right)^{-1/2} \gamma_K \omega_K = \pm \left(1 - \frac{3m}{r} \right)^{-1/2} \left(\frac{m}{r^3} \right)^{1/2}.$$

■

Exercise 8.12.2.

parallel transport along a circular geodesic orbit: Frenet-Serret frame along a timelike helix

For the Schwarzschild equatorial plane spacetime, introduce the orthonormal frame

$$e_t = \left(1 - \frac{2m}{r} \right)^{-1/2} \frac{\partial}{\partial t}, \quad e_r = \left(1 - \frac{2m}{r} \right) \frac{\partial}{\partial r}, \quad e_\theta = \frac{1}{r} \frac{\partial}{\partial \theta}.$$

Introduce the 4-velocity to a timelike helical world line representing a circular orbit in this spacetime

$$U = \frac{\partial t}{\partial \tau} \frac{\partial}{\partial t} + \frac{\partial \theta}{\partial \tau} \frac{\partial}{\partial \theta} = \gamma(e_t + \nu e_\theta), \quad \gamma = (1 - \nu^2)^{-1/2},$$

and let e_Θ be the orthogonal vector in the t - θ plane of the tangent space

$$e_\Theta = \gamma(\nu e_t + e_\theta).$$

The three vectors U , e_r , e_Θ are a Frenet-Serret frame along the helix.

a) Use a computer algebra system to derive the connection components in this frame, assuming $\nu = \nu(r)$ is a function only of r . The U derivatives of this frame define the curvature κ and torsion Ω of the helix in spacetime

$$\begin{aligned} \frac{DU}{d\tau} &= \nabla_U U = \Gamma^r_{UU} e_r \equiv \kappa e_r, \\ \frac{De_r}{d\tau} &= \nabla_U e_r = \Gamma^U_{Ur} U + \Gamma^\Theta_{Ur} e_\Theta \equiv \kappa U + \Omega e_\Theta, \\ \frac{De_\Theta}{d\tau} &= \nabla_U e_\Theta = -\Omega e_r. \end{aligned}$$

Show that the formulas for these two quantities can be written

$$\kappa = \left(1 - \frac{2m}{r} \right)^{1/2} \frac{\gamma_K^2}{r} (\nu^2 - \nu_K^2), \quad \Omega = - \left(1 - \frac{2m}{r} \right)^{1/2} \frac{\gamma^2}{\gamma_K^2} \frac{\nu}{r}.$$

The curvature κ is the magnitude of the acceleration, zero for a geodesic. The torsion Ω is proper angular velocity of the axes e_r, e_Θ which span the local rest space of the world line, called the

Fermi-Walker rotation, measured with respect to axes in the local rest space which are said to undergo Fermi-Walker transport. Along an accelerated helix Fermi-Walker transported vectors rotate with the opposite angular velocity with respect to these axes. Along a geodesic helix, for which

$$\Omega_K = - \left(1 - \frac{2m}{r}\right)^{1/2} \frac{\nu_K}{r} = \mp \left(\frac{m}{r^3}\right)^{1/2},$$

the Fermi-Walker transport reduces to parallel transport. These Fermi-Walker/parallel transported vectors are

$$e_1 = \cos(\Omega\tau) e_r - \sin(\Omega\tau) e_\Theta, \quad e_2 = \sin(\Omega\tau) e_r + \cos(\Omega\tau) e_\Theta.$$

b) The proper angular velocity of the geodesic orbit and the Fermi-Walker angular velocity have opposite signs. Their sum represents the net angular velocity of the parallel transported axes with respect to Cartesian axes at infinite radius

$$\Omega_{(\text{net})} \equiv \left. \frac{d\theta}{d\tau} \right|_K + \Omega_K = \mp \left(\frac{m}{r^3}\right)^{1/2} \left(1 - \left(1 - \frac{3m}{r}\right)^{-1/2}\right).$$

The Cartesian axes at infinite radius are represented locally in the local rest space along the helix by

$$e_x = \cos(\omega t) e_r - \sin(\omega t) e_\Theta, \quad e_y = \sin(\omega t) e_r + \cos(\omega t) e_\Theta.$$

By expanding the formula for $\Omega_{(\text{net})}$ to first order in the quantity $3m/r$, show that one finds that for large radius

$$\Omega_{(\text{net})} \approx \mp \frac{3}{2} \frac{m}{r} \left(\frac{m}{r^3}\right)^{1/2}.$$

c) Two thirds of this result was obtained using the tangent cone to the embedding surface of the constant time slice of this spacetime in Exercise 8.4.4 evaluated for a circle at fixed radius and time t . It is not surprising that motion in two of the three dimensions results in 2/3 of the result for motion in time as well. We can get this result directly by looking at the connection coefficients in the original orthonormal frame e_θ, e_r, e_t which in this order is a Frenet-Serret frame for the spacelike circle of fixed radius where e_θ is the unit tangent

$$\begin{aligned} \frac{De_\theta}{ds} &= \nabla_{e_\theta} e_\theta = \Gamma_{\theta\theta}^r e_r \equiv \kappa e_r, \\ \frac{De_r}{ds} &= \nabla_{e_\theta} e_r = \Gamma_{\theta r}^\theta e_\theta + \Gamma_{\theta r}^r e_r \equiv -\kappa e_\theta + \tau e_t, \\ \frac{De_t}{ds} &= \nabla_{e_\theta} e_t = -\tau e_r. \end{aligned}$$

Examining the connection coefficients in this frame, one sees that torsion τ vanishes and the curvature is

$$\kappa = \left(1 - \frac{2m}{r}\right)^{1/2} \frac{1}{r} \equiv \frac{1}{\mathcal{R}(r)},$$

which at large radius agrees with the flat space curvature of the circle. The radius of curvature $\mathcal{R}(r)$ was found by other means in Exercise 8.4.4. Here κ is the arclength rate of change of the angle of rotation of the parallel transported axes with respect to the given frame, so adding the oppositely signed orbital angular velocity leads at large radii to the net angular velocity of the parallel transported axes

$$\frac{d\theta}{ds} - \kappa = \frac{1}{r} - \frac{1}{r} \left(1 - \frac{2m}{r}\right)^{1/2} \rightarrow -\frac{m}{r^2},$$

so that multiplying by the proper velocity $ds/d\tau = \nu_K \rightarrow (m/r)^{1/2}$ of the circular orbit as in that exercise to convert to the proper time rate of change leads to the final result of that exercise for counter-clockwise motion

$$\left(\frac{d\theta}{ds} - \kappa\right) \frac{ds}{d\tau} \rightarrow -\frac{m}{r^2} \left(\frac{m}{r}\right)^{1/2} = -\frac{m}{r} \left(\frac{m}{r^3}\right)^{1/2},$$

which is 2/3 the correct result. ■

Chapter 9

Intrinsic curvature

\mathbb{R}^n with its Euclidean metric is flat. The 2-sphere is not. How do we describe this mathematically? We need to introduce a quantity that will be called curvature and show how it agrees with our vague intuitive idea of curvature. We will introduce the “Riemann” curvature tensor in several steps. Parallel transport of a tangent vector along a curve preserves its constant length, and its angle with any other vectors which are also parallel transported, but there is no guarantee that if we transport it around a curve and come back to its initial location that its final direction will coincide with its initial direction. Because its length is preserved, the final transported vector has to be related to its starting value by some rotation, and when this rotation is nontrivial, it is a sign that curvature is present. If we shrink the loop to an “infinitesimal loop” we get an infinitesimal rotation, which is described by an antisymmetric matrix.

If we consider infinitesimal loops in a 2-plane, we need a 2-form to pick out the 2-plane, and we need to associate with that 2-plane an antisymmetric matrix to describe the rotation of the tangent space that results. As we will discover, an antisymmetric matrix-valued 2-form is the object we need to do the job, which will lead to a $\binom{1}{3}$ -tensor field R^i_{jmn} , namely a matrix valued 2-form, antisymmetric in mn , leaving a $\binom{1}{1}$ -tensor after the full tensor is evaluated in these arguments on a pair of vectors which determine the 2-plane. But we are getting ahead of ourselves. Let’s go slowly.

Before we embark on our journey to understand this parallel transport problem, we must also recognize that the cylinder is curved. However, it can be cut along a line parallel to its symmetry axis and rolled out flat onto a plane, so its intrinsic geometry is flat. The cylinder is instead extrinsically curved within the larger space in which it finds itself. We will study this other aspect of curvature in the next chapter. We cannot flatten out a sphere, which instead is truly curved intrinsically.

We begin by introducing a tensor which vanishes identically in flat Euclidean space and then interpret this “curvature” tensor as a measure of how it characterizes the failure of parallel transport around a loop to bring a tangent vector back to its original value in a truly curved space. This same tensor also characterizes the focusing/defocusing properties of nearby geodesics. In flat Euclidean space a pair of nearby straight line geodesics, if initially parallel, remain so, but in a curved geometry positive curvature causes them to converge, while negative

curvature causes them to diverge from each other. If one imagines the geodesics to be analogous to light rays (which is literally realized in general relativity), then curvature causes the geometry to act like a lens in its affect on families of geodesics, thus providing another way of measuring curvature.

9.1 Calculating the curvature tensor and its symmetries

We progress slowly starting with a Cartesian coordinate calculation on \mathbb{R}^n with its flat Euclidean metric, then repeating the calculation in a coordinate frame on \mathbb{R}^n with any metric, and then finally we repeat the calculation in a general frame in such a space. We examine how a second covariant derivative along two vector fields performed in different orders compares to the corresponding derivative by the Lie bracket of those two fields. With hindsight this is an almost obvious thing to do since in flat space in Cartesian coordinates, the covariant derivative along a coordinate frame vector field $X = \delta^j_k \partial_j = \partial_k$ for some particular k reduces to the Lie derivative along that vector field,

$$[\mathcal{L}_X Z]^i = Z^i_{,j} X^j - X^i_{,j} Z^j = Z^i_{,j} \delta^j_k = Z^i_{,k}$$

and the Lie derivative for vector fields satisfies the Jacobi identity, which can be written in the form

$$\mathcal{L}_X \mathcal{L}_Y Z - \mathcal{L}_Y \mathcal{L}_X Z \equiv [\mathcal{L}_X, \mathcal{L}_Y] Z = \mathcal{L}_{[X,Y]} Z.$$

Thus the same identity must hold for the covariant derivative if $X = \partial_i$ and $Y = \partial_j$ are just coordinate frame vector fields, so we see what happens for general vector fields. Amazingly it still holds!

(1) Calculation in Cartesian coordinates on \mathbb{R}^n

On \mathbb{R}^n in Cartesian coordinates, covariant and ordinary differentiation coincide, so

$$[\nabla_Y Z]^i = Z^i_{,j} Y^j = Z^i_{,j} Y^j,$$

hence

$$\begin{aligned} (\nabla_X \nabla_Y - \nabla_Y \nabla_X) Z^i &= [\nabla_Y Z]^i_{,k} X^k - [\nabla_X Z]^i_{,k} Y^k \\ &= (Z^i_{,j} Y^j)_{,k} X^k - (Z^i_{,j} X^j)_{,k} Y^k \\ &= Z^i_{,jk} Y^j X^k - Z^i_{,jk} X^j Y^k \\ &\quad + Z^i_{,j} Y^j_{,k} X^k - Z^i_{,j} X^j_{,k} Y^k \\ &= \underbrace{[Z^i_{,jk} - Z^i_{,kj}]}_{0 \text{ since partial derivatives commute}} X^k Y^j + Z^i_{,j} \underbrace{[Y^j_{,k} X^k - X^j_{,k} Y^k]}_{[X,Y]^j} \\ &\quad \underbrace{\hspace{10em}}_{[\nabla_{[X,Y]} Z]^i} \end{aligned}$$

Thus

$$\underbrace{[\nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]}]}_{\text{second order differential operator}} Z = 0.$$

In other words this second order differential operator on vector fields is identically zero on \mathbb{R}^n . This will be true no matter what coordinates we use to express the operator, since it represents a tensor field independent of how we choose to evaluate it.

(2) Calculation in arbitrary coordinates for any metric

We now repeat this calculation for arbitrary coordinates on any space with a metric tensor with coordinate component matrix (g_{ij}) and the associated metric connection components $\Gamma^i_{jk} = \{\Gamma^i_{jk}\}$, now including the components of the connection in the formulas. Recall the formula

$$[\nabla_Y Z]^i = (Z^i_{,j} + \Gamma^i_{jm} Z^m) Y^j,$$

so we can iterate it, expanding into 6 grouped terms using the product rule and second partial derivative, which we regroup

$$\begin{aligned} [\nabla_X \nabla_Y Z]^i &= ([\nabla_Y Z]^i_{,k} + \Gamma^i_{km} [\nabla_Y Z]^m) X^k \\ &= \underbrace{\{[(Z^i_{,j} + \Gamma^i_{jm} Z^m) Y^j]_{,k} + \Gamma^i_{km} (Z^m_{,j} + \Gamma^m_{jp} Z^p) Y^j\}}_{\text{expand using product rule}} X^k \\ &= \underbrace{\{[Z^i_{,jk} + \Gamma^i_{jm} Z^m_{,k} + \Gamma^i_{jm,k} Z^m] Y^j}_{(1)} + \underbrace{(Z^i_{,j} + \Gamma^i_{jm} Z^m) Y^j_{,k}}_{(4)} \\ &\quad + \underbrace{\Gamma^i_{km} (Z^m_{,j} + \Gamma^m_{jp} Z^p) Y^j}_{(5)}\}_{(6)} X^k \\ &= \underbrace{[Z^i_{,jk} + \Gamma^i_{jm} Z^m_{,k} + \Gamma^i_{km} Z^m_{,j} + \Gamma^i_{jm,k} + \Gamma^i_{km} \Gamma^m_{jp} Z^p]}_{(1)} X^k Y^j + \underbrace{Z^i_{,j} Y^j_{,k}}_{(4)} X^k \\ &= [\dots]_{jk}^i X^k Y^j + \underbrace{Z^i_{,j} Y^j_{,k}}_{(4)} X^k. \end{aligned}$$

Observe that term (1) is symmetric in (j, k) , while the pair (2) plus (5) together are symmetric in those indices, so if we switch these indices in the formula and subtract, all these terms will cancel out. Finally in term (6) for standardization, we can replace the dummy indices (m, p) by (n, m)

Now if we switch X and Y in these formulas

$$\begin{aligned} [\nabla_X \nabla_Y Z]^i &= [\dots]_{jk}^i X^k Y^j, \\ [\nabla_Y \nabla_X Z]^i &= [\dots]_{jk}^i Y^k X^j = [\dots]_{kj}^i X^k Y^j \end{aligned}$$

then subtracting cancels out the pairs of terms (1), (2) and (5) since they are symmetric in these indices (j, k) , leaving the remaining terms, with the pair of terms (4) combining to form the derivative of Z by the Lie bracket $[X, Y]$

$$\begin{aligned} [\nabla_X \nabla_Y - \nabla_Y \nabla_X] Z^i &= [\Gamma^i_{jm,k} - \Gamma^i_{km,j} + \Gamma^i_{kn} \Gamma^n_{jm} - \Gamma^i_{jn} \Gamma^n_{km}] X^k Y^j Z^m \\ &\quad + \underbrace{Z^i_{,j} [Y^j_{,k} X^k - X^j_{,k} Y^k]}_{(4)} \\ &= [\nabla_{[X, Y]} Z]^i \end{aligned}$$

Thus subtracting the last term from the right hand side we get the final result

$$(\nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X, Y]}) Z^i = R^i_{mkj} X^k Y^j Z^m,$$

where

$$R^i_{mkj} = \Gamma^i_{jm,k} - \Gamma^i_{km,j} + \Gamma^i_{kn}\Gamma^n_{jm} - \Gamma^i_{jn}\Gamma^n_{km}.$$

This operator is actually linear in X, Y , and Z , i.e., it defines the components of a tensor field

$$R = R^i_{mkj} e_i \otimes \omega^m \otimes \omega^k \otimes \omega^j$$

which is explicitly antisymmetric in its last pair of indices. It is called the Riemann curvature tensor and the previous calculation shows that in \mathbb{R}^n with the Euclidean metric, this curvature tensor is identically zero since it is obviously zero in any Cartesian coordinate system.

(3) Calculation in an arbitrary frame

The components of this Riemann curvature tensor in an arbitrary frame are

$$R^i_{mkj} = R(\omega^i, e_m, e_k, e_j)$$

or

$$[\nabla_{e_i} \nabla_{e_j} - \nabla_{e_j} \nabla_{e_i} - \nabla_{[e_i, e_j]}] e_k = R^{\ell}_{kij} e_{\ell}$$

but by definition of the connection components

$$\nabla_{e_j} e_k = \Gamma^{\ell}_{jk} e_{\ell}$$

we can iterate this derivative using the product rule (the connection coefficients are just scalars)

$$\begin{aligned} \nabla_{e_i} \nabla_{e_j} e_k &= \nabla_{e_i} (\Gamma^{\ell}_{jk} e_{\ell}) = (\nabla_{e_i} \Gamma^{\ell}_{jk}) e_{\ell} + \Gamma^{\ell}_{jk} \underbrace{\nabla_{e_i} e_{\ell}}_{\Gamma^m_{i\ell} e_m} \\ &= (\Gamma^{\ell}_{jk,i} + \Gamma^{\ell}_{im} \Gamma^m_{jk}) e_{\ell} \end{aligned}$$

and by linearity in the differentiating vector field

$$\nabla_{[e_i, e_j]} e_k = \nabla_{C^m_{ij} e_m} e_k = C^m_{ij} \nabla_{e_m} e_k = C^m_{ij} \Gamma^{\ell}_{mk} e_{\ell}$$

so substituting these two relations into the following combination yields

$$\begin{aligned} &[\nabla_{e_i} \nabla_{e_j} - \nabla_{e_j} \nabla_{e_i} - \nabla_{[e_i, e_j]}] e_k \\ &= \underbrace{(\Gamma^{\ell}_{jk,i} - \Gamma^{\ell}_{ik,j} - C^m_{ij} \Gamma^{\ell}_{mk} + \Gamma^{\ell}_{im} \Gamma^m_{jk} - \Gamma^{\ell}_{jm} \Gamma^m_{ik})}_{R^{\ell}_{kil}} e_{\ell}. \end{aligned}$$

The only difference in this formula compared to the coordinate frame formula is the extra structure function term which appears compared to the coordinate frame calculation where the Lie brackets of the coordinate frame vectors vanish.

If we evaluate this same formula in a coordinate frame where $C^m_{ij} = 0$, we can re-interpret it in terms of the antisymmetrized second covariant derivative of the vector field instead of the commutator of successive covariant directional derivatives of the vector field

$$(\nabla_i \nabla_j - \nabla_j \nabla_i) Z^k = \nabla_i \wedge \nabla_j Z^k = Z^k_{;ji} - Z^k_{;ij} = -2Z^k_{;[j i]} = R^k_{mij} Z^m.$$

In any frame the components of the curvature tensor can be packaged as a matrix-valued 2-form called the curvature 2-form

$$\underline{\Omega} = (\Omega^i_j), \text{ where } \Omega^i_j = \frac{1}{2} R^i_{jmn} \omega^m \wedge \omega^n,$$

which in turn are merely the matrix of components of the $\binom{1}{1}$ -tensor valued 2-form

$$R^i_{jmn} e_i \otimes \omega^j \otimes \omega^m \otimes \omega^n = \frac{1}{2} R^i_{jmn} e_i \otimes \omega^j \otimes \omega^m \wedge \omega^n,$$

which emphasizes the different roles played by the first and second pair of indices of the Riemann curvature tensor.

We can lower the first index on the curvature tensor to obtain a completely covariant $\binom{0}{4}$ -tensor with components $R_{mnij} = g_{m\ell} R^\ell_{nij}$. One can show (Exercises below) that this “covariant Riemann tensor” is antisymmetric in its first pair of indices as well as its second, and it is symmetric under the interchange of these two pairs

$$\begin{aligned} R_{nmij} &= -R_{mnij}, & (\text{first pair antisymmetry}) \\ R_{mnji} &= -R_{mnij}, & (\text{second pair antisymmetry}) \\ R_{nmij} &= R_{ijmn}. & (\text{pair interchange symmetry}) \end{aligned}$$

For a 2-dimensional space, it has therefore at most one independent nonzero component R_{1212} . In an orthogonal frame this means that R^1_{212} is the single independent component.

Exercise 9.1.1.

Riemann, Ricci and Einstein in 3 dimensions

For a 3-dimensional space these three symmetries make the covariant curvature tensor equivalent to a symmetric bilinear function on the 3-dimensional space of 2-vectors, which has only $3(3+1)/2 = 6$ independent components, the same number of independent components as a symmetric second rank tensor.

Since each of the two pairs of antisymmetric indices is equivalent to a single index by the dual operation, taking the “double dual” of the covariant Riemann tensor leads to a symmetric $\binom{0}{2}$ -tensor $H_{ij} = H_{ji}$

$$[*R^*]^i_j = \left(\frac{1}{2} \eta^i_{kl} \right) R^{kl}_{mn} \left(\frac{1}{2} \eta^j_{mn} \right) \equiv -H^i_j$$

which is easily inverted by taking the double dual of the result

$$R^{ij}_{kl} = -\eta^{ij}_m H^m_n \eta^n_{kl} = -\eta^{ijm} \eta_{kln} H^n_m.$$

To manipulate this relationship, introduce the Ricci $\binom{0}{2}$ -tensor and scalar curvature by the contractions

$$R_{ij} = R^k_{ikj}, \quad R = R^i_i = g^{ij} R_{ij} = R^{ik}_{ki} = -R^{ik}_{ik},$$

and the symmetric Einstein $\binom{0}{2}$ -tensor

$$G_{ij} = R_{ij} - \frac{1}{2} R g_{ij}.$$

Using the pair interchange symmetry to rewrite the above relationship

$$[*R^*]^i_j = \frac{1}{4} \eta^i_{kl} \eta^j_{mn} R^{kl}_{mn} = \frac{1}{4} \eta^{ikl} \eta_{jmn} R_{kl}^{mn} = \frac{1}{4} \eta^{ikl} \eta_{jmn} R^{mn}_{kl},$$

and then expanding the double eta using the identity (recall M is the number of minus signs in the diagonal matrix of inner products of the vectors in an orthonormal frame)

$$\begin{aligned} (-1)^M \eta^{ikl} \eta_{jmn} &= \delta^{ikl}_{jmn} = \epsilon^{ijl} \epsilon_{jmn} \\ &= \delta^i_j \delta^k_m \delta^\ell_n + \delta^i_m \delta^k_n \delta^\ell_j + \delta^i_n \delta^k_j \delta^\ell_m - \delta^i_j \delta^k_n \delta^\ell_m - \delta^i_m \delta^k_j \delta^\ell_n - \delta^i_n \delta^k_m \delta^\ell_j \\ &= \delta^i_j \delta^{kl}_{mn} + \delta^i_m \delta^k_n \delta^\ell_j + \delta^i_n \delta^k_j \delta^\ell_m - \delta^i_m \delta^k_j \delta^\ell_n - \delta^i_n \delta^k_m \delta^\ell_j, \end{aligned}$$

show that the sign-reversed double dual of the Riemann tensor is just the Einstein tensor apart from another possible sign

$$[*R^*]^i_j = \left(\frac{1}{2} \eta^i_{kl} \right) R^{kl}_{mn} \left(\frac{1}{2} \eta^j_{mn} \right) = (-1)^{M+1} G^i_j.$$

■

Exercise 9.1.2.

Riemann, Ricci and Einstein in 2 dimensions

a) Now that we have done the harder calculation, do the same double dual derivation for $n = 2$ to see that the Riemann curvature tensor can be expressed entirely in terms of the scalar curvature by

$$R^{ij}_{kl} = \frac{R}{2} \delta^{ij}_{kl}$$

which implies

$$R^i_j = \frac{R}{2} \delta^i_j.$$

Note that the dual of a pair of antisymmetric indices has no indices—it is a scalar.

b) Show that this implies that the Einstein tensor in 2 dimensions vanishes identically.

The Gaussian curvature of a 2-metric that we will learn about in the next chapter is just half the scalar curvature, provided the sign conventions are appropriate to make the scalar curvature positive for spheres (in any dimension).

**Exercise 9.1.3.****covariant components of Riemann**

a) Starting from the formula for the coordinate components of the curvature tensor

$$R^i{}_{jmn} = \Gamma^i{}_{nj,m} - \Gamma^i{}_{mj,n} + \Gamma^i{}_{m\ell}\Gamma^\ell{}_{nj} - \Gamma^i{}_{n\ell}\Gamma^\ell{}_{mj},$$

and the index-lowered components of the connection

$$\Gamma_{ijk} = g_{i\ell}\Gamma^\ell{}_{jk} = \frac{1}{2}(g_{ij,k} + g_{kj,i} - g_{ik,j}), \quad \Gamma^i{}_{jk} = g^{i\ell}\Gamma_{\ell jk},$$

use the inverse matrix derivative formula of Exercise 2.3.6 re-expressed in terms of them

$$g^{ij}{}_{,k} = -g^{im}g^{jn}g_{mn,k} = -g^{im}g^{jn}(\Gamma_{mk,n} + \Gamma_{nk,m})$$

to complete the following derivation of the fully covariant coordinate components of the curvature tensor expressed directly in terms of the second order derivatives of the metric components by expanding the product rule and using the previous identity, then cancelling some terms (the “...” terms below), and then finally re-expressing the derivatives of the index-lowered connection components in terms of the second derivatives of the metric

$$\begin{aligned} R_{ijmn} &= g_{ik}R^k{}_{jmn} = g_{ik} \{ (g^{kp}\Gamma_{pnj})_{,m} - (g^{kp}\Gamma_{pmj})_{,n} \} + \Gamma_{im\ell}\Gamma^\ell{}_{nj} - \Gamma_{in\ell}\Gamma^\ell{}_{mj} \\ &= \Gamma_{inj,m} + \dots - \Gamma_{imj,n} - \dots + \Gamma_{im\ell}\Gamma^\ell{}_{nj} - \Gamma_{in\ell}\Gamma^\ell{}_{mj} \\ &= \Gamma_{inj,m} - \Gamma_{imj,n} + \Gamma_{lim}\Gamma^\ell{}_{nj} - \Gamma_{lin}\Gamma^\ell{}_{mj} \\ &= \frac{1}{2}(g_{im,jn} - g_{in,jm} + g_{jn,im} - g_{jm,in}) + g_{k\ell}(\Gamma^k{}_{im}\Gamma^\ell{}_{jn} - \Gamma^\ell{}_{in}\Gamma^k{}_{jm}). \end{aligned}$$

b) Using this formula, verify the symmetry property under pair interchange $R_{mni j} = R_{ijmn}$. Note that from the antisymmetry in the last pair of indices of the curvature tensor, this immediately implies antisymmetry in the first pair as a consequence.



For an orthonormal frame in which $g_{ij} = \delta_{ij}$, raising the first index of the antisymmetric pair (i, j) in R_{ijmn} back to its original position does not change the components, so we obtain an antisymmetric matrix valued 2-form as announced at the beginning of this chapter. However, we still need to see how this tensor we have found from a derivation unmotivated by any idea of curvature translates into a description of curvature in terms of parallel transport around loops.

Exercise 9.1.4.**symmetries of Riemann**

a) The Riemann curvature tensor has another index symmetry which is less obvious. Using the coordinate formula for its components, evaluate the following cyclic sum and show that

its 3 pairs of derivative terms and its three pair of product terms cancel in pairs due to the symmetry of the components of the connection $\Gamma^i_{jk} = \Gamma^i_{kj}$

$$R^i_{jkl} + R^i_{k\ell j} + R^i_{\ell jk} = 0, \quad (\text{cyclic symmetry} = \text{“Bianchi identity of the first kind”})$$

or equivalently (since it is already antisymmetric on the last pair of indices)

$$R^i_{[jkl]} = 0.$$

This reflects a deeper symmetry associated with the symmetry of the metric connection that is best left to a more sophisticated study.

b) For $n = 2$ and $n = 3$ this additional symmetry is not independent of the previous ones since the four indices of the covariant curvature tensor cannot be distinct. Consider the identity $R_{ijkl} + R_{iklj} + R_{iljk} = 0$ for $(ijkl) = (1212)$ to convince yourself of this in the first case, and for $(ijkl) = (3123)$ in the second case.

c) For $n = 4$ one can finally have 4 distinct index values and this cyclic symmetry imposes additional conditions on the number of independent components of the curvature tensor. Since there is only one set of 4 distinct index values, regardless of which one is in the fixed position, the other symmetries imply that there is only one additional condition on the number of independent components. Convince yourself of this. A 2-vector has $4(3)/2 = 6$ independent components and a symmetric bilinear function on 2-vectors is equivalent to a symmetric 6×6 matrix which has $6(7)/2 = 21$ independent components. This cyclic symmetry subtracts one more to make only 20 independent components of the Riemann curvature tensor. $4(5)/2 = 10$ of these are packaged in the symmetric Ricci or Einstein tensors, leaving an additional 10 independent components which determine the so called Weyl curvature tensor in 4 dimensions. This is a topic for a higher level course and is essential to Einstein’s general theory of relativity for describing the gravitational field. The Ricci/Einstein tensors are tied to the presence of matter in spacetime by Einstein’s equations, while the Weyl curvature is analogous to the radiative part of the electromagnetic field, namely the part which carries gravitational waves.

d) If there were no symmetries, the curvature tensor would have n^4 independent components since it has 4 indices, each one of which can take n values. Imposing symmetries must reduce that number, but chances are the number of constraints are polynomials of no higher degree, so the final result is probably a 4th degree polynomial.

Note that if we think of R_{ijmn} as a symmetric bilinear function on the space of 2-vectors, which has dimension $N = n(n-1)/2$ the number of independent components of that symmetric matrix on the N -dimensional vector space is

$$\frac{1}{2}N(N+1) = \frac{1}{8}n(n-1)(n^2 - n + 2) \equiv CnoB(n),$$

only taking into account the three index pair interchange symmetries without taking into account the Bianchi identity constraint. Now we have to take into account the Bianchi identity which only kicks in at $n = 4$ as we saw above (i.e., the number of independent conditions should be a 4th degree polynomial which vanishes for $n = 1, 2, 3$). It is easy to find the formula

$$C(n) = \frac{1}{2}n^2(n^2 - 1)$$

counting the number of independent components of the Riemann tensor by searching the web. Use a computer algebra system to factor the difference

$$C(n) - CnoB(n) = CB(n)$$

which defines the number of independent algebraic constraints on the components imposed by the Bianchi identities (of the first kind). Are you surprised by this simple result? Re-express this in terms of factorials. There is a simple mathematical symbol for this result involving n and 4. What is it? Does this offer some clue perhaps to how to untangle the overlapping symmetry conditions on the independent components of Riemann?

■

Exercise 9.1.5.

curvature of planes, cylinders, spheres

For the following 2-dimensional surfaces in \mathbb{R}^3 :

- 1) in cylindrical coordinates $\{\rho, \phi\}$ on the plane $z = 0$ where $^{(2)}g = d\rho \otimes d\rho + \rho^2 d\phi \otimes d\phi$,
- 2) in cylindrical coordinates $\{\phi, z\}$ on the cylinder $\rho = r_0$ where $^{(2)}g = r_0^2 d\phi \otimes d\phi + dz \otimes dz$,
- 3) in spherical coordinates $\{\theta, \phi\}$ on the sphere $r = r_0$, where $^{(2)}g = r_0^2(d\theta \otimes d\theta + \sin^2 \theta d\phi \otimes d\phi)$,
 - a) calculate the single independent component R^1_{212} of the curvature tensor in the orthogonal coordinate frame, and then raise its index to obtain $R^{12}_{12} = g^{22} R^1_{212}$,
 - b) and then repeat the calculation in the associated orthonormal frame to get $R^{\hat{1}}_{\hat{2}\hat{1}\hat{2}} = R^{\hat{1}\hat{2}}_{\hat{1}\hat{2}} = R^{12}_{12}$, showing that the two results are related to each other by the appropriate metric factor scalings.

It should be no surprise that the first two surfaces lead to zero curvature, while the sphere has a constant curvature reflecting the fact that all points are equivalent on the sphere as far as the geometry is concerned; this constant curvature component turns out to be r_0^{-2} , directly analogous to the curvature r_0^{-1} for a circle of radius r_0 .

b) If you want another example doable by hand, repeat this calculation in the plane for parabolic coordinates, which are also orthogonal.

c) You can easily check your work with a computer algebra system. These hand calculations for very simple metrics are useful to get a flavor of curvature evaluation, but computers are essential for any serious evaluations. Computing curvature tensors from a metric by hand is like long division, who needs it?

■

Exercise 9.1.6.

curvature of an ellipsoid of revolution

Consider an ellipsoid of revolution resulting from revolving an ellipse with semiaxes a and b in the x - z plane around the z -axis

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} + \frac{z^2}{b^2} = 1,$$

parametrized by a simple extension of the usual parametrization of a sphere by the polar and azimuthal angles, although for the ellipsoid the polar angle is no longer interpretable as in the spherical case $a = b$

$$\vec{r}(\theta, \phi) = \langle a \cos \theta \cos \phi, a \cos \theta \sin \phi, b \sin \theta \rangle.$$

a) Derive the expression for the metric

$$ds^2 = (a^2 + (b^2 - a^2) \sin^2 \theta) d\theta^2 + a^2 \cos^2 \theta d\phi^2.$$

b) Use a computer algebra system to derive the result

$$R^{\theta\phi}{}_{\theta\phi} = \frac{a^2}{(a^2 + (b^2 - a^2) \sin^2 \theta)^2}.$$

Notice that for the sphere $a = b$ this reduces to the correct value $1/a^2$. This is also the value at the North and South Poles $\sin \theta = 0$ where by rotational symmetry the sphere of radius a resulting from the revolving the osculating circle of best fit is a quadratic approximation of best fit there. ■

Exercise 9.1.7.

curvature of an elliptical paraboloid

Consider the metric on the surface $z = \frac{1}{2}(9x^2 + 4y^2)$ parametrized by

$$\vec{r}(\rho, \phi) = \langle 2\rho \cos \phi, 3\rho \sin \phi, 18\rho^2 \rangle$$

studied in Exercise 1.6.12.

a) Evaluate the differential $d\vec{r} = \langle dx, dy, dz \rangle$ in terms of the parametrization variables (ρ, ϕ) , and then substitute these expressions into the Euclidean metric to obtain the metric on the surface

$$\begin{aligned} {}^{(2)}g &= dx(\rho, \phi) \otimes dx(\rho, \phi) + dy(\rho, \phi) \otimes dy(\rho, \phi) + dz(\rho, \phi) \otimes dz(\rho, \phi) \\ &= g_{\rho\rho} d\rho \otimes d\rho + g_{\rho\phi} (d\rho \otimes d\phi + d\phi \otimes d\rho) + g_{\phi\phi} d\phi \otimes d\phi. \end{aligned}$$

These coordinates are not orthogonal so one has to do a bit more work to evaluate the general formulas for the components of the connection and curvature.

b) Identify the metric matrix \underline{g} and find its inverse \underline{g}^{-1} and the corresponding inverse metric

$${}^{(2)}g^{-1} = g^{\rho\rho} \frac{\partial}{\partial \rho} \otimes \frac{\partial}{\partial \rho} + g^{\rho\phi} \left(\frac{\partial}{\partial \rho} \otimes \frac{\partial}{\partial \phi} + \frac{\partial}{\partial \phi} \otimes \frac{\partial}{\partial \rho} \right) + g^{\phi\phi} \frac{\partial}{\partial \phi} \otimes \frac{\partial}{\partial \phi}.$$

c) Use a computer algebra system to evaluate the components of the connection.

d) Use a computer algebra system to evaluate the Ricci scalar, which in our convention should be positive (the opposite of Maple's sign convention). Its value at the origin is $2(4)(9) = 72$.

■

Exercise 9.1.8.**curvature of surface of revolution**

a) For a surface of revolution we calculated the metric and components of the connection in Exercise 8.4.1. Show that the single independent component of the curvature tensor is

$$R^r_{\theta r \theta} = -R(r)R''(r), \quad R^{r\theta}_{r\theta} = -\frac{R''(r)}{R(r)}.$$

b) Evaluate this for the torus using Section 8.8 and confirm the formula

$$\frac{1}{2}R = R^{r\theta}_{r\theta} = R^{\hat{r}\hat{\theta}}_{\hat{\theta}\hat{r}} = \frac{\cos(r/b)}{b(a + b \cos(r/b))}.$$

Notice that for $a = 0$ this clearly reduces to the correct expression for a sphere of radius b , namely the Gaussian curvature $1/b^2$.

c) For a surface of constant curvature

$$R^{r\theta}_{r\theta} = -\frac{R''(r)}{R(r)} = R_0,$$

leading to the differential equation

$$R''(r) + R_0 R(r) = 0.$$

Write down the general solutions for the three cases $R_0 > 0, = 0, < 0$. If we want a surface of revolution which intersects its symmetry axis to be regular at that intersection point, its limiting tangent plane must be orthogonal to that axis there, so that in the limit approaching that point, the metric should approach the form of polar coordinates in that plane, which imposes 2 conditions on the function $R(r)$: $R(0) = 0$ and $\lim_{r \rightarrow 0} R(r)/r = 1$. With these extra conditions, what are the three unique solutions to this differential equation?

■

Exercise 9.1.9.**curvature in cylindrical coordinates**

Verify that curvature tensor of Euclidean R^3 vanishes in cylindrical coordinates. Note that at most $9 = 3 \times 3$ independent components $\{R^2_{3ij}, R^3_{1ij}, R^1_{2ij}\}_{ij=\{23,31,12\}}$ need to be checked because the coordinates are orthogonal, so the antisymmetry on the first index pair when lowered means that the diagonal index pairs vanish and only these three index combinations remain independent. In fact because of the pair interchange symmetry, only 6 independent components need to be checked—we can think of $R_{ijkl} = R_{klij}$ as defining a symmetric matrix on the 3-dimensional vector space of antisymmetric matrices, which only has 6 independent components.

**Exercise 9.1.10.****curvature from integrability conditions**

Although we have motivated curvature by parallel transport, historically the idea of parallel transport only came decades after the curvature tensor itself was discovered by looking at the integrability conditions for the solution of the partial differential equations that follow from the existence of a covariant constant vector field.

Set the coordinate components of the covariant derivative of a vector field to zero and solve it for the partial derivatives

$$\nabla_j Z^k = \partial_j Z^k + \Gamma^k_{jm} Z^m = 0 \quad \rightarrow \quad \partial_j Z^k = -\Gamma^k_{jm} Z^m$$

and then differentiate it again

$$\partial_i \partial_j Z^k = \dots$$

and evaluate the commutator of the second partial derivatives to obtain the result

$$0 = \partial_i \partial_j Z^k - \partial_j \partial_i Z^k = -R^k_{mij} Z^m.$$

The vanishing of the mixed partials acting on the vector field Z is a necessary condition for a covariant constant vector field to exist. Thus Z must be a 0 eigenvalue eigenvector of the linear transformation-valued curvature 2-form $\underline{\Omega}$: $\underline{\Omega} Z = \underline{0}$. For a covariant constant frame to exist, this implies that all the components of the curvature tensor must be zero. This idea of integrability conditions for the existence of solutions of partial differential equations is fundamental.

**Exercise 9.1.11.****Jacobi and Bianchi identities**

Most component geometric formulas have a noncomponent, i.e., frame independent counterpart. For example, the simple symmetry of the coordinate components of the metric connection $\Gamma^k_{[ij]} = 0$ was seen in Exercise 6.6.4 to be expressible as the vanishing of the torsion tensor defined without reference to components

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y] = 0.$$

This translates this antisymmetric combination of the covariant directional derivative into a Lie bracket

$$\nabla_X Y - \nabla_Y X = [X, Y].$$

The Lie bracket in turn satisfies the Jacobi identity

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0.$$

Replace Y in the covariant derivative equation by $[Y, Z]$

$$\nabla_X[Y, Z] - \nabla_{[Y, Z]}X = [X, [Y, Z]],$$

and then in the first term replace the vector field $[Y, Z]$ being differentiated by its covariant derivative representation to obtain a three term left hand side involving two second covariant derivative terms. Now form the cyclic sum of this equation and use the Jacobi identity to set the right hand side to 0. Regroup the nine terms into three curvature tensor terms to obtain

$$R(X, Y)Z + R(Y, Z)X + R(Z, X)Y = 0,$$

where another frequent variation of curvature tensor notation is the following

$$R(X, Y)Z = \left([\nabla_X, \nabla_Y] - \nabla_{[X, Y]} \right) Z = R^i_{jmn} X^m Y^n Z^j e_i.$$

Express this cyclic curvature identity in terms of its components to see that the Bianchi identity of the first kind is really just a consequence of the Jacobi identity and the symmetry of the metric connection (vanishing torsion). In Chapter 11 we will see that this derivation can be translated from the Jacobi identity for vector fields to the vanishing second exterior derivative identity for 1-forms.

■

9.2 Interpretation of curvature

Curvature is a notion inversely related to radius of curvature. A circle is more curved if it has a small radius, less curved or “flatter” if it has a larger radius. In fact in multivariable calculus one learns that the curvature κ of a circle is the reciprocal of its radius r_0 : $\kappa = 1/r_0$. A straight line corresponds to the limit $r_0 \rightarrow \infty$ in which a circle (fixing one point on its circumference) flattens out to a straight line and has zero curvature.

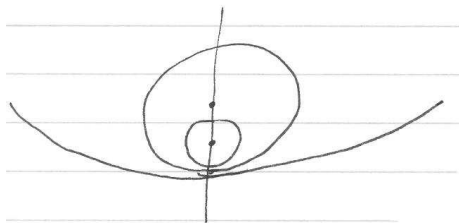


Figure 9.1: Smaller circles curve more, and have larger curvature. Larger circles curve less, and have smaller curvature.

For an arbitrary curve $c(\lambda)$, at each point we can determine a circle of best fit to the curve (the osculating circle) in the plane of the tangent vector $c'(\lambda)$ and its derivative $c''(\lambda)$ (the osculating plane) having radius $\rho = 1/\kappa$ in terms of the curvature κ at the given point on the curve. All of this is done in both \mathbb{R}^2 and \mathbb{R}^3 in a typical multivariable calculus course.

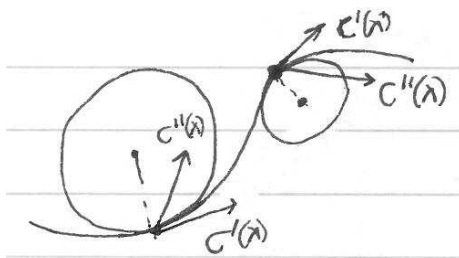


Figure 9.2: Measuring curvature with a circle of best fit at each point.

To handle curvature for surfaces in \mathbb{R}^3 , it turns out we can always find two circles of best fit at right angles to each other which best describe how the surface curves at each point. These circles may lie on the same side or on opposites of the tangent plane, as with an ellipsoid (upper figure) or a saddle surface (lower figure) shown in Fig. 9.3. The so-called Gaussian curvature of the surface is taken to be

$$\kappa = \pm \left(\frac{1}{r_1} \right) \left(\frac{1}{r_2} \right) : \begin{cases} + & \text{if same side,} \\ - & \text{if opposite sides.} \end{cases}$$

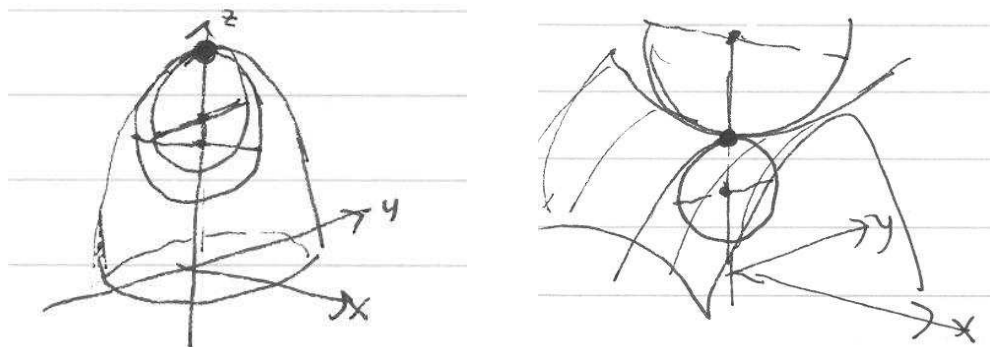


Figure 9.3: Two circles of best fit at right angles on the same side of the surface (positive curvature), or on the opposite sides (negative curvature).

Increasing either radius flattens out the surface at the point. A cylinder is an example where one of the two radii has become infinite, so one of these two circles of best fit reduces to a straight line, leading to zero curvature.

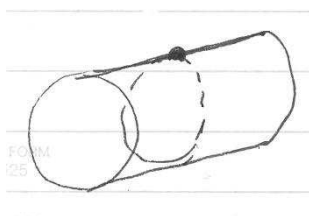


Figure 9.4: One of these two circles flattens out to a line for a cylinder (zero curvature).

Since all directions along a sphere at a given point are equivalent (and all points are equivalent in terms of the geometry of the sphere), and the circles of best fit are just great circles through that point, both radii are equal to the radius of the sphere itself at every point, leading to a constant curvature $\kappa = 1/r_0^2$. Notice that curvature has dimensions of inverse length in the 1-dimensional case and of inverse area in the 2-dimensional case.

The 2-dimensional curvature concept generalizes to a notion of curvature in higher-dimensional spaces, in a way similar to the generalization from a curve to a surface. For the surface in \mathbb{R}^3 , we can consider each plane cross-section of the surface by a plane containing the normal line to the surface, yielding a cross-sectional curve through the given point whose curvature can be evaluated, so one gets a curvature for each direction in the tangent plane associated with the plane cross-section curve in that direction. From that family of plane cross-sections one can show that a pair of orthogonal directions in the tangent plane capture all the curvature information. For higher dimensional spaces we will again consider all 2-plane cross-sections of the full tangent space corresponding to 2-surfaces through the given point and transfer our intuition about the curvature of a surface in \mathbb{R}^3 to that situation.

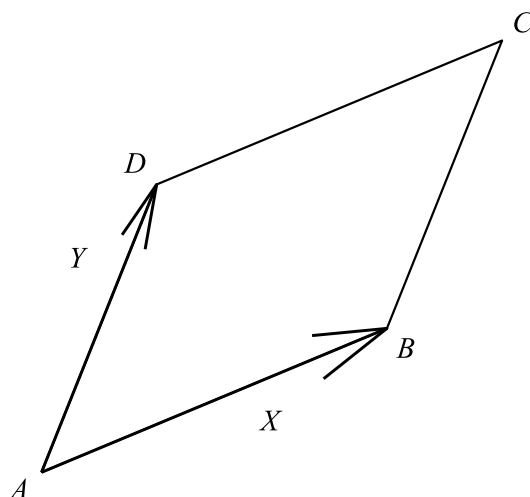


Figure 9.5: The oriented parallelogram formed by an ordered pair of two tangent vectors (X, Y) in the tangent space and the counterclockwise loop around that parallelogram starting at the origin A , moving first along X to its tip at B , then to the tip of $X + Y$ at C , then to the tip of Y at D , and finally back to the origin along Y .

However, all of these notions of curvature so far describe the bending of curves and surfaces and are associated with what is called extrinsic curvature, because they depend on how these spaces fit into a larger space. Intrinsic curvature instead refers only to the effects of curvature which are measurable within a given space. The failure of parallel transport to be independent of the path is an obvious consequence of such intrinsic curvature effects and is the basis of a limiting procedure (the zoom of calculus which leads to all differential structure) which can be used to measure curvature using a tensor, the so called Riemann curvature tensor. It too relies on the notion of selecting a 2-surface to measure curvature effects, but within the space. To pick out a limiting 2-plane in the tangent space to a 2-surface we will need a 2-form. Then associated with the failure of parallel transport of vectors around a loop in that surface to return the vector to its original value, we need to associate a linear transformation with that 2-surface. The curvature tensor must therefore have 2 indices playing the role of a linear transformation (1 up, 1 down), and 2 indices associated with a 2-form (2 down) for a total of 4 indices (1 up, 3 down).

To preview the mathematical procedure which connects the curvature tensor to parallel transport around a loop, for a given point in some curved space of dimension 3 or greater with a positive definite metric, pick a 2-plane in the tangent space at the given point. This can be done by specifying two linearly independent tangent vectors X and Y . The 2-vector $X \wedge Y$ contains both orientation information as well as length information. Its length using the 2-vector inner product which avoids overcounting gives the area of the parallelogram formed

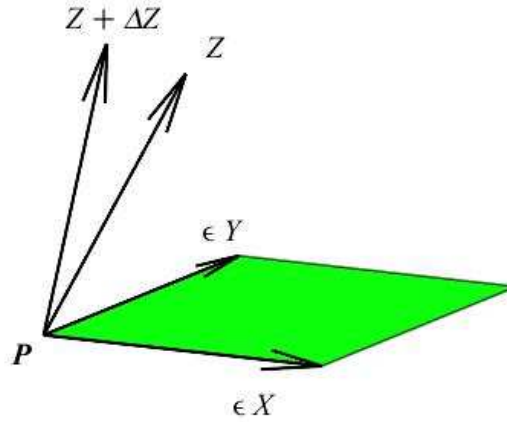


Figure 9.6: Parallel transporting a tangent vector Z around an “infinitesimal” parallelogram formed by two tangent vectors $\epsilon X, \epsilon Y$ in a limiting process leads to an infinitesimal rotation of Z to $Z + \Delta Z$.

by the two vectors (just the length of their cross-product in multivariable calculus in \mathbb{R}^3). Now multiply X and Y by a small enough positive number ϵ that we can identify them with directed curve segments in the space itself, i.e., the part of the tangent space they occupy is so small that we can use it as a good approximation to part of the space itself. The parallelogram in the tangent space may be interpreted as a closed curve in the space itself beginning and ending at the point P of our discussion.

Take any tangent vector Z at P and parallel transport it around the loop. Its length must remain constant so at most it can rotate relative to its original value. The difference is a small rotation which we have seen is described by an antisymmetric matrix in an orthonormal frame. This difference is approximately

$$\Delta Z^i \approx -R^i_{jmn}(\epsilon X)^m(\epsilon Y)^n Z^j$$

This approximation gets better as $\epsilon \rightarrow 0$. Note that its value is proportional to ϵ^2 , or more precisely, to the area of the parallelogram formed by ϵX and ϵY .

The antisymmetry in the *last* pair of indices means that only the components of $X \wedge Y$, not $X \otimes Y$, contribute to this value. The antisymmetry of the first pair of indices, when both lowered, means that in an orthonormal frame the mixed indices are also antisymmetric and hence the matrix

$$R^i_{jmn}(\epsilon X)^m(\epsilon Y)^n$$

represents a small rotation, which when contracted with Z^j , rotates it by this small amount to produce the increment in Z .

If we fix the indices (mn) on R^i_{jmn} , we are looking at the subspace of the tangent space spanned by the frame vectors e_m and e_n or $\partial/\partial x^m$ and $\partial/\partial x^n$ in coordinate frame. The remaining 2 indices describe the small rotation (in the sense of the increment of a vector under the rotation) associated with that 2-plane of directions. We can basically think of the curvature tensor as a linear-transformation-valued function on 2-planes in the tangent space. As long as we have a metric with Euclidean signature as we have assumed in these discussions, the linear transformations describe the rate of change of rotations.

In 2-dimensions, the choice of 2-plane is fixed (the whole 2-dimensional tangent space) and a single number characterizes the rotation that occurs transporting a vector in a small parallelogram in that 2-plane, explaining why the curvature tensor has only 1 independent component R^1_{212} .

So far we have only made claims about the curvature tensor and this parallel transport process. We must show how the limiting process of parallel transport around a smaller and smaller loop yields the above curvature tensor formula, which gives curvature its interpretation. In doing so we will only be using the intrinsic geometry of the space, while our notions of curvature for curves and surfaces in ordinary space comes from the way they bend in space, i.e., how their tangent directions and normal directions rotate as we move from point to point on the surface within the enveloping space. This is the extrinsic geometry of the curve or surface which relates to how it fits into the larger space. Instead when we talk about how tangent vectors within our given space behave without reference to some enveloping larger space, we are dealing with the intrinsic geometry of that space. However, when a space does occur within a larger space (keep in mind the example of curves and surfaces in \mathbb{R}^3), the curvature of the full space can be constructed from the intrinsic curvature of the subspace together with its extrinsic curvature. In particular when the full space is flat (zero curvature!), the extrinsic and intrinsic curvature must exactly cancel each other out when combined, so the extrinsic curvature is hard wired directly to the intrinsic curvature and our initial intuitive remarks about how curvature is measured for curves and surfaces (an extrinsic curvature discussion) then transfers to how we measure the intrinsic curvature. On a sphere in ordinary space, for example, we talked about the radii of curvature of great circles through a given point, which is an extrinsic notion, but we must tie that to how we transport tangent vectors around so that they remain tangent to the sphere, an intrinsic notion. The relationship between extrinsic and intrinsic curvature will be explored in the next chapter.

9.3 The limiting loop parallel transport calculation of curvature

Although consulting various derivations in the literature makes this calculation of the parallel transport of a vector around a loop in the limit in which it shrinks to a point seem a bit complicated, the coordinate calculation is nothing more than the limiting case of Green's theorem in the plane \mathbb{R}^2 in which one finds that the “third component of the curl” of a vector field in the plane is the limiting value of its line integral around a coordinate rectangle as that rectangle shrinks to a point. Let's first revisit that calculation and then bootstrap it up to our application.

A line integral of a vector field, really a 1-form $X = X_1 dx^1 + X_2 dx^2$ with coordinate components $\langle X_1 | X_2 \rangle$ along a parametrized curve $c : x^k = x^k(\lambda)$, $k = 1, 2$, $\lambda_1 \leq \lambda \leq \lambda_2$ in the plane can be thought of as the solution of a first-order differential equation for a scalar quantity $Z = Z(\lambda)$ defined along that curve

$$\begin{aligned} \frac{dZ}{d\lambda} &= X_k \frac{dx^k}{d\lambda}, \\ \underbrace{\int_c \frac{dZ}{d\lambda} d\lambda}_{\Delta Z|_c} &= \int_c X_k \frac{dx^k}{d\lambda} d\lambda = \int_c X_k dx^k, \end{aligned}$$

where the increment of Z along the curve is $\Delta Z|_c = Z(\lambda_2) - Z(\lambda_1)$. Notice that the necessary index positioning tells us we are really integrating a covector field or 1-form here. We will discuss all of these vector integral topics in Chapter 11. The net change in the quantity Z from the beginning of a directed curve segment to its end is given by this line integral. For a simple closed loop curve c , the net change around the loop is the line integral $\oint_c X_k dx^k$. To relate this to the “curl” of the “vector field” (i.e., 1-form), we consider essentially the application of Green's theorem to a shrinking Cartesian coordinate rectangle in the plane, but here we will do the whole calculation without referring to Green's theorem.

First we set up the piecewise continuous parametrized curve c on a coordinate rectangle $ABCD$ as illustrated in Fig. 11.1

$$c = \begin{cases} A \rightarrow B : & x^2 = x_0^2, x^1 : x_0^1 \rightarrow x_0^1 + \Delta x^1, \lambda = x^1, \\ B \rightarrow C : & x^1 = x_0^1 + \Delta x^1, x^2 : x_0^2 \rightarrow x_0^2 + \Delta x^2, \lambda = x^2, \\ C \rightarrow D : & x^2 = x_0^2 + \Delta x^2, x^1 : x_0^1 + \Delta x^1 \rightarrow x_0^1, \lambda = x^1, \\ D \rightarrow A : & x^1 = x_0^1, x^2 : x_0^2 + \Delta x^2 \rightarrow x_0^2, \lambda = x^2. \end{cases}$$

Next we evaluate the four contributions to the line integral around the loop, but considering

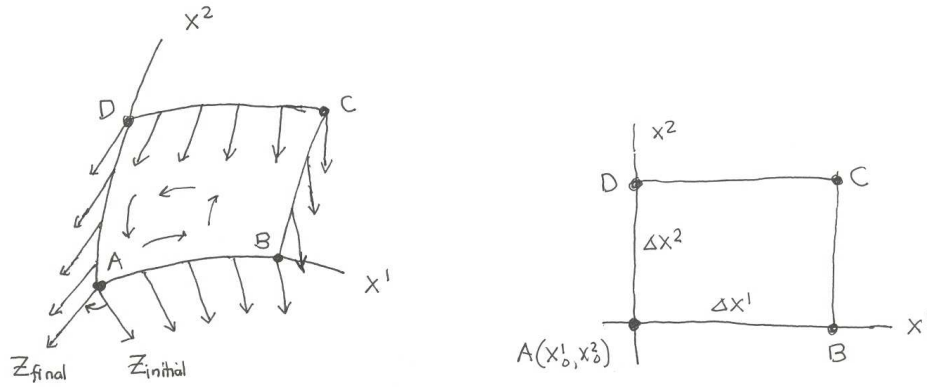


Figure 9.7: Left: Parallel transporting a tangent vector around a coordinate rectangle leads to a rotation between the initial and final directions. Right: In the coordinate space in the x^1 - x^2 plane, this rectangle is simple.

opposite sides of the parallelogram in pairs. The first pair is

$$\begin{aligned}\Delta Z|_{A \rightarrow B} &= \int_{x_0^1}^{x_0^1 + \Delta x^1} X_1|_{x^2=x_0^2} dx^1, \\ \Delta Z|_{C \rightarrow D} &= \int_{x_0^1 + \Delta x^1}^{x_0^1} X_1|_{x^2=x_0^2 + \Delta x^2} dx^1 = - \int_{x_0^1}^{x_0^1 + \Delta x^1} X_1|_{x^2=x_0^2 + \Delta x^2} dx^1,\end{aligned}$$

and their sum alone is

$$\begin{aligned}\Delta Z|_{A \rightarrow B} + \Delta Z|_{C \rightarrow D} &= \int_{x_0^1}^{x_0^1 + \Delta x^1} X_1|_{x^2=x_0^2} dx^1 - \int_{x_0^1}^{x_0^1 + \Delta x^1} X_1|_{x^2=x_0^2 + \Delta x^2} dx^1 \\ &= - \int_{x_0^1}^{x_0^1 + \Delta x^1} \left(X_1|_{x^2=x_0^2 + \Delta x^2} - X_1|_{x^2=x_0^2} \right) dx^1 \\ &\approx - \int_{x_0^1}^{x_0^1 + \Delta x^1} \left(\frac{\partial X_1}{\partial x^2} \right) \Big|_{x^2=x_0^2 + \Delta x^2} \Delta x^2 dx^1,\end{aligned}$$

where this approximation is valid in the limit $\Delta x^2 \rightarrow 0$ as long as X_2 is differentiable as will be assumed. The contribution from the other two sides has the opposite sign since they are traced out in the opposite direction from the first pair (check this!), so that combining the two pairs one finds

$$\begin{aligned}\Delta Z|_c &= \Delta Z|_{A \rightarrow B} + \Delta Z|_{C \rightarrow D} + \Delta Z|_{B \rightarrow C} + \Delta Z|_{D \rightarrow A} \\ &\approx \int_{x_0^2}^{x_0^2 + \Delta x^2} \left(\frac{\partial X_2}{\partial x^1} \right) \Big|_{x^1=x_0^1 + \Delta x^1} \Delta x^1 dx^2 - \int_{x_0^2}^{x_0^2 + \Delta x^2} \left(\frac{\partial X_1}{\partial x^2} \right) \Big|_{x^2=x_0^2 + \Delta x^2} \Delta x^2 dx^1.\end{aligned}$$

In the limit $(\Delta x^1, \Delta x^2) \rightarrow (0, 0)$ assuming the integrands of these integrals are continuous, one can approximate each of their values by the value of the integrand at any point of the interval

(this is just the fundamental theorem of calculus!) and just multiply by the increment in the integrating variable. In particular we can choose these values at the left endpoints of each coordinate interval

$$\begin{aligned}\Delta Z|_c &\approx \left(\frac{\partial X_2}{\partial x^1} \right) \Big|_{x^1=x_0^1, x^2=x_0^2} \Delta x^1 \Delta x^2 - \left(\frac{\partial X_1}{\partial x^2} \right) \Big|_{x^1=x_0^1, x^2=x_0^2} \Delta x^2 \Delta x^1 \\ &= \left(\frac{\partial X_2}{\partial x^1} - \frac{\partial X_1}{\partial x^2} \right) \Big|_{x^1=x_0^1, x^2=x_0^2} \Delta x^1 \Delta x^2.\end{aligned}$$

Thus the “third component of the curl” of a vector field in the plane represents the limiting ratio of its line integral around the loop divided by the area of the loop in this multivariable calculus derivation. It immediately generalizes to the interpretation of all three components of the curl in space as well in a equivalent limiting application of Stoke’s theorem.

The parallel transport equations

$$\frac{dZ^i}{d\lambda} = -\Gamma^i_{jk} \frac{dx^j}{d\lambda} Z^k \equiv X^i_k \frac{dx^j}{d\lambda}, \quad X^i_k \equiv -\Gamma^i_{jk} Z^k,$$

for each fixed value of the index i , are exactly where the previous derivation starts. Suppose we consider a coordinate rectangle loop in the first two coordinates (x^1, x^2) exactly as above in any coordinate system $\{x^i\}$ on an n -dimensional space with metric g_{ij} . The net change in these components around the loop is therefore just

$$\Delta Z^i|_c \approx \left(\frac{\partial X^i_2}{\partial x^1} - \frac{\partial X^i_1}{\partial x^2} \right) \Big|_{x^1=x_0^1, x^2=x_0^2} \Delta x^1 \Delta x^2.$$

Continuing further by evaluating these partial derivatives. First consider the partial derivative term along ∂_1 , which comes from the

$$\begin{aligned}\left(\frac{\partial X^i_2}{\partial x^1} \right) \Big|_{x^1=x_0^1, x^2=x_0^2} &= - \left(\frac{\partial}{\partial x^1} (\Gamma^i_{2k} Z^k) \right) \Big|_{x^1=x_0^1, x^2=x_0^2} \\ &= - \left(\frac{\partial \Gamma^i_{2k}}{\partial x^1} Z^k + \Gamma^i_{2k} \frac{dZ^k}{dx^1} \right) \Big|_{x^1=x_0^1, x^2=x_0^2} \\ &= -(\partial_1 \Gamma^i_{2k} Z^k - \Gamma^i_{2k} \Gamma^k_{1m} Z^m) \Big|_{x^1=x_0^1, x^2=x_0^2},\end{aligned}$$

using the original differential equation in the form

$$\frac{dZ^i}{dx^1} = -\Gamma^i_{1m} Z^m$$

for the rate of change along the x^1 sides of the parallelogram. Note that we can’t partial differentiate Z since it is only defined along the curve by its differential equation along that curve.

Then evaluating the difference of this and the expression with the indices 1 and 2 interchanged, and then exchanging the dummy indices (k, m) in order to factor out Z^k , yields

$$\begin{aligned}\Delta Z^i|_c &\approx [(\Gamma_{1k,2}^i - \Gamma_{2k,1}^i) Z^k + \Gamma_{2k}^i \Gamma_{1m}^k Z^m - \Gamma_{1k}^i \Gamma_{2m}^k Z^m]|_{x^1=x_0^1, x^2=x_0^2} \Delta x^1 \Delta x^2 \\ &\approx [(\Gamma_{1k,2}^i - \Gamma_{2k,1}^i + \Gamma_{2m}^i \Gamma_{1k}^m - \Gamma_{1m}^i \Gamma_{2k}^m) Z^k]|_{x^1=x_0^1, x^2=x_0^2} \Delta x^1 \Delta x^2 \\ &\approx -R^i{}_{k12} Z^k|_{x^1=x_0^1, x^2=x_0^2} \Delta x^1 \Delta x^2.\end{aligned}$$

Thus

$$\begin{aligned}\lim_{(\Delta x^1, \Delta x^2) \rightarrow (0,0)} \frac{\Delta Z^i|_c}{\Delta x^1 \Delta x^2} &= -R^i{}_{k12} Z^k|_{x^1=x_0^1, x^2=x_0^2} \\ &= -R^i{}_{kmn} Z^k X^m Y^n|_{x^1=x_0^1, x^2=x_0^2} \\ &= -\Omega^i{}_k(X, Y) Z^k|_{x^1=x_0^1, x^2=x_0^2},\end{aligned}$$

where $X = \partial_1, Y = \partial_2$ are the coordinate frame vector fields and

$$\Omega^i{}_k = R^i{}_{kmn} dx^m \wedge dx^n$$

is the matrix of curvature 2-forms, whose evaluation on the pair (X, Y) which determine the plane in the tangent space corresponding to the plane of the shrinking closed loop yields $\Omega^i{}_k(X, Y) = R^i{}_{kmn} X^m Y^n = R^i{}_{k12}$. This matrix, when its first index is lowered, is anti-symmetric, and represents an “infinitesimal rotation,” or more precisely the instantaneous rate of change of a rotation, which describes how the tangent vector starts to rotate under such a parallel transport process once the increment factor $\Delta x^1 \Delta x^2$ is factored in

$$Z \rightarrow Z - R(X, Y) Z \Delta x^1 \Delta x^2,$$

where

$$[R(X, Y) Z]^i \equiv R^i{}_{jmn} X^m Y^n Z^j \equiv [(\nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X, Y]}) Z]^i,$$

although the last term on the right hand side of the last equality does not contribute to the commuting coordinate frame vector fields.

This result is in fact true for any pair of commuting vector fields $[X, Y] = 0$, for a parallelogram formed by using the flow lines of the two vector fields. The sides of these parallelograms correspond to the increments $\Delta \lambda_1$ and $\Delta \lambda_2$ which refer to the natural parameters along those flow lines as in the case when the vector fields are actually coordinate frame derivatives. To see how this derivation generalizes to two vector fields with a nonvanishing commutator, one needs the geometric interpretation of the Lie bracket.

The limiting loop parallel transport frame curvature calculation

We saw that an additional term involving the commutators of the frame vector fields appears in the component formula for the curvature tensor in a general frame compared with a coordinate frame. Suppose we extend the previous calculation to any two vector fields X and Y whose

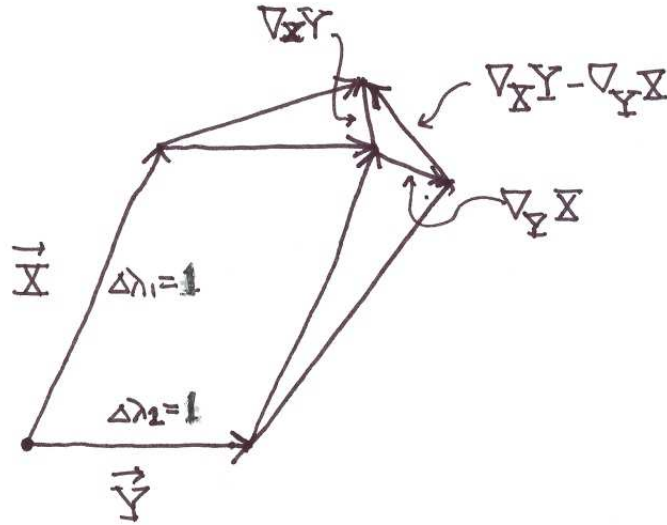


Figure 9.9: Parallel transporting the values of Y and X a unit parameter distance along the flow lines of X and Y respectively from a common starting point and comparing the final tangent vectors there (those which appear parallel in the figure) with the actual values of those vector fields there (which appear slightly changed), one finds a gap between the endpoints, thus forming an “open quadrilateral” which is closed by the difference vector $\nabla_X Y - \nabla_Y X$.
 [TO DO: replace X and Y by ϵX and ϵY in caption and in figure.]

parameter distance along their flow line through that point, forming a closed parallelogram at lowest order in the tangent space in the approximation that we can identify the nearby space with this portion of the tangent space. If we then compare the parallel transported vector with the actual value of the vector field at the new point, their difference will equal the covariant derivative of the one with respect to the other in the limit region near the origin of the tangent space.

[FIX]

To see this, adopting a sloppy but suggestive notation, we use the fact that the covariant derivative along X measures the difference between a vector field and its parallel transported vector along the flow lines of X , etc.

$$Y(x + X) \approx Y_{||}(x + X) + [\nabla_X Y](x), \quad X(x + Y) \approx X_{||}(x + Y) + [\nabla_Y X](x).$$

The closer of the quadrilateral here is just the difference of the covariant derivatives $\nabla_X Y -$

$\nabla_Y X$. Let's evaluate this quantity in a coordinate frame

$$\begin{aligned}
 [\nabla_X Y - \nabla_Y X]^i &= (Y^i_{,j} + \Gamma^i_{kj} Y^k) X^j - (X^i_{,j} + \Gamma^i_{kj} X^k) Y^j \\
 &= (Y^i_{,j} - X^i_{,j}) + \Gamma^i_{kj} Y^k X^j - \Gamma^i_{jk} X^j Y^k \\
 &= [X, Y]^i - \underbrace{(\Gamma^i_{jk} - \Gamma^i_{kj})}_{0} X^j Y^k = [X, Y]^i,
 \end{aligned}$$

since the antisymmetric part of the connection components in a coordinate frame vanish identically for the metric covariant derivative. The connection associated with a metric is called a symmetric connection for this reason, and this condition is equivalent to the vanishing of the so called torsion tensor

$$T(X, Y) \equiv \nabla_X Y - \nabla_Y X - [X, Y] = T^i_{jk} X^j Y^k.$$

Returning to our calculation, we conclude that the Lie bracket of X and Y is the closer of this parallel transport quadrilateral.

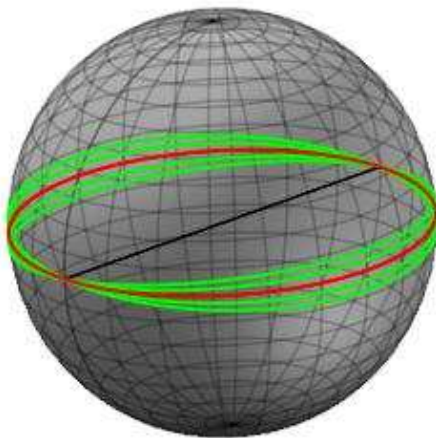


Figure 9.10: Great circles near the equator of the sphere passing through common antipodal points of the equator. After leaving one such point they refocus at the second point after a distance of half the circumference of the sphere.

9.4 Positive curvature focusing of geodesics and negative curvature defocusing

The most symmetric curved surface we know is a part of our daily lives: the sphere. If we take any point on the sphere and send out geodesics in all directions, any pair of these great circles are forced to move closer to each other with respect to their original tangent lines in space at the initial point as one moves farther from that initial point, due to the constraint that they remain within the surface but try not to veer left or right within the surface until finally they meet each other at the antipodal point on the other side of the sphere. Because of the high symmetry of the sphere, all the geodesics emanating from any point all “refocus” at the same point an equal distance along each one. We can think of the geodesic circles about the original point, which consists of all points an equal distance along each great circle from that starting point, and as we move along the family of geodesics at unit speed in the arclength, these circles expand outward and then contract until they shrink to a point at the antipodal point. If we think of flashing a light at the initial point which moves outward in the surface at unit speed, then these geodesic circles are the wave fronts and the geodesics themselves the light rays which move orthogonally to the wavefront. The positive curvature of the sphere acts

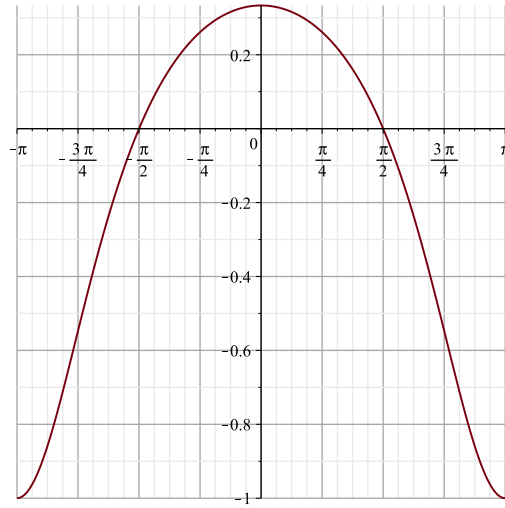


Figure 9.11: The Gaussian curvature versus polar angle on the unit torus $(a, b) = (2, 1)$.

as a lens to refocus the light rays within a distance equal to the circumference of the sphere.

For any metric on a space, the family of geodesics emanating from a given point in all directions is called the geodesic spray, as if we sprayed water in all directions and the droplets kept moving in their original directions in the space without slowing down. As this spray moves through regions of positive curvature, they are drawn closer together, i.e., are focused, and the opposite situation holds for negative curvature, they move farther apart. The torus is a simple surface which interpolates between maximum positive curvature at the outer equator and maximum negative curvature (a minimum of the curvature which is negative) at the inner horizon. Those geodesics which start out at the outer equator but fail to cross the inner equator are reflected by the centrifugal potential barrier and sent back to the outer equator where they meet their counterparts in the opposite upper/lower hemitorus: geodesics with initial tangent vectors at equal angles above and below the outer equator meet again a certain distance along the outer equator, and the same but shorter distance along each other. They are all refocused together in pairs but the greater their initial separation angle, the farther along the outer equator is their meeting point. However, as one decreases this angle to zero, the refocusing point on the equator reaches a limiting point at a certain minimal distance from the starting point which is a function of the curvature. Consider a point on the sphere itself and consider shooting a geodesic at a small angle from the equator. No matter how small the angle it does not come back to the equator until half the circumference of the equator is reached: πr_0 is the minimal refocusing length along that great circle and hence any great circle, since the length is the same along both great circles.

Recall the result of Exercise 9.1.8 for half the curvature scalar which in the next chapter we will see equals the Gaussian curvature

$$\frac{1}{2}R = R^{r\theta}_{r\theta} = R^{\hat{r}\hat{\theta}}_{\hat{r}\hat{\theta}} = \frac{\cos(r/b)}{b(a + b\cos(r/b))}.$$

This has the same sign as the cosine $\cos(r/b) = \cos \chi$ of the angle measured around the profile circle from the outer equator. It is positive on the outer hemitorus and negative on the inner hemitorus, with reflection symmetry between the upper and lower hemitori across either the inner or outer equator, and is extremized at the equators themselves. The polar circles of zero curvature separate the two regions of focusing by positive curvature and defocusing by negative curvature, with the extreme lensing effect of each type occurring at the equators. Fig. 9.12 dramatically shows how the positive curvature region in the outer hemitorus focuses geodesics quickest at the outer equator where a minimum distance exists before three geodesics leaving an initial point on the outer equator at equal angles above and below that equator together with the equatorial direction itself cross again on the equator, all equidistant from the point of departure. The pair of orbits which skirt the polar circles mark the boundary inside which all the geodesics between them intersect with their neighbors in the same upper/lower hemitorus. Those bound geodesics (don't cross the inner equator) which are outside this pair cross the polar circles where they are increasingly distanced from each other until the cross back over in their return to the outer equator.

Gravitational lensing on spacetime is a manifestation of this phenomenon. Anyone who has some interest in popular expositions of our understanding of the universe knows the name of Hawking. In his pre-celebrity days, Stephen Hawking and his collaborator Roger Penrose at the time proved their singularity theorems which imply that general relativity always focuses light rays as long as the material source of the gravitational field obeys a certain positive energy condition, and it never defocuses them under that condition. Because of the rapidly increasing sophistication of astrophysical instrumentation and clever data analysis, this has become an important tool of modern theoretical astrophysics. Inflation, however, is caused by a quantum field which does not obey the positive energy condition, allowing that phase of the universe to be an exception to this behavior.

We can quantify this lensing effect of geodesics by considering a coordinate system adapted to an affinely parametrized geodesic curve $c(\lambda)$ extended to a 1-parameter family $C(\lambda, \sigma)$ of nearby geodesics which initially remain “near” the original geodesic $c(\lambda)$, with $C(\lambda, 0) = c(\lambda)$. This represents a 2-surface in our space and we can then introduce the tangents to this surface

$$u^i(\lambda, \sigma) = \frac{\partial C^i}{\partial \lambda}(\lambda, \sigma), \quad \eta^i(\lambda, \sigma) = \frac{\partial C^i}{\partial \sigma}(\lambda, \sigma)$$

so that along the original geodesic we have its tangent u and a “connecting vector” η along that geodesic

$$u^i(\lambda) = \left. \frac{\partial C^i(\lambda, \sigma)}{\partial \lambda} \right|_{\sigma=0}, \quad \eta^i(\lambda) = \left. \frac{\partial C^i(\lambda, \sigma)}{\partial \sigma} \right|_{\sigma=0}$$

with the geodesic condition holding for u

$$\frac{Du^i}{d\lambda}(\lambda) = 0.$$

We can imagine the tip of $\sigma\eta(\lambda)$ for small values of σ as tracing out the path of a nearby geodesic. Since u and η come from a coordinate grid in the surface, their commutator vanishes

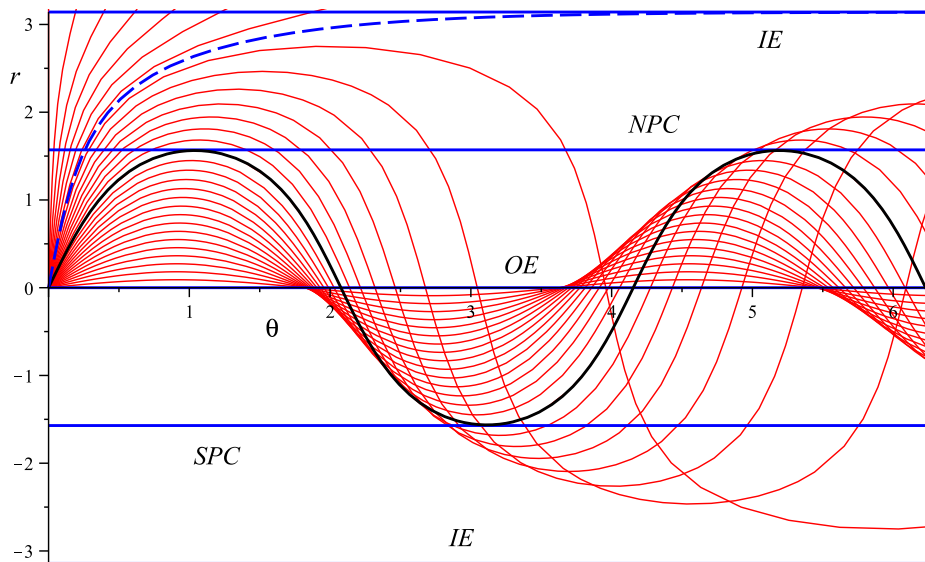


Figure 9.12: The spray of geodesics leaving the origin of coordinates in the first quadrant on the unit ring torus $(a, b) = (2, 1)$ for $-\pi \leq r \leq \pi$, $0 \leq \theta \leq 2\pi$ (one azimuthal revolution). As one increases the angle initial tangent vectors make with the meridians at the outer equator, first one passes through the unbound geodesics which cross the inner equator at $r = \pi$ (up to the thick dashed curve which asymptotically approaches the inner equator), then the bound geodesics with a smaller and smaller turning point radius $r_{(\max)}$. Looking at where these geodesics cross the Northern Polar Circle at $r = \pi/2$, starting at $\theta = 0$ the crossing point moves to the right through unbound geodesics and then into the bound geodesics where their turning point lowers until it reaches that circle (the thick black geodesic, very close to the $[3, 2, 0]$ closed geodesic which is slightly higher), after which the geodesics which reach that circle overshoot it first, crossing over and then returning to that circle as their second crossing point moves to the right and the turning point rises. Note also the half wavelength $\Delta\theta = \pi/\sqrt{3} \approx 1.81$ of the small oscillations about the outer equator. For $0 < \theta < \Delta\theta$ only the outer equator itself from this spray reaches points on that outer equator, but for $\Delta\theta < \theta < 2\Delta\theta$ on the outer equator, a second member of this family reaches the outer equator (with shorter length), while for $2\Delta\theta < \theta < 3\Delta\theta$ a third member of this family reaches the outer equator. Of course for $\pi < \theta < 2\pi$, shorter geodesics arrive from the opposite azimuthal direction.

$[u, \eta] = 0$ and the symmetry condition on the connection means that on that surface

$$\nabla_u \eta - \nabla_\eta u = [u, \eta] = 0.$$

This equation makes sense since both vector fields are defined on the surface along which we are differentiating, and these covariant derivatives equal the corresponding covariant derivatives along their flow lines ($\nabla_u \rightarrow D/d\lambda, \nabla_v \rightarrow D/d\sigma$) so

$$\frac{Du}{d\sigma} - \frac{D\eta}{d\lambda} = 0.$$

If we apply the definition of the curvature tensor with the commuting vector fields u, η on the 2-surface, we find (using the geodesic condition on u)

$$\begin{aligned} R^i_{jmn} u^j u^m \eta^n &= (\nabla_u \nabla_\eta - \nabla_\eta \nabla_u) u^i \\ &= \left(\frac{D}{d\lambda} \frac{D}{d\sigma} - \frac{D}{d\sigma} \frac{D}{d\lambda} \right) u^i \\ &= \frac{D^2 \eta^i}{d\lambda^2} \end{aligned}$$

Evaluating this at $\sigma = 0$ along the original geodesic, we obtain the Jacobi equation of geodesic deviation

$$\frac{D^2 \eta^i}{d\lambda^2} = R^i_{jmn} u^j u^m \eta^n = -R^i_{jmn} u^j \eta^m u^n.$$

This is particularly simple for a surface where there are no extra dimensions and only one independent component of the curvature tensor. Note that only the component of η orthogonal to u contributes to the right hand side since $R^i_{jmn} u^j u^m u^n = 0$ due to the antisymmetry in the last index pair, while the right hand side only has a component orthogonal to u since $u_i R^i_{jmn} u^j u^m \eta^n = 0$ due to the antisymmetry in the first index pair. We can then orthogonally decompose $\eta = \zeta \hat{N} + \alpha \hat{T}$ along the unit tangent $\hat{T} = \hat{u}$ along the geodesic and the intrinsic unit normal $\hat{N} = \hat{\eta}$ within the surface. Both of these unit vectors are parallel transported along the geodesic

$$0 = \frac{D\hat{T}}{d\lambda} = \frac{D\hat{N}}{d\lambda}$$

so we get

$$\frac{D^2 \alpha}{d\lambda^2} \hat{T} + \frac{D^2 \zeta}{d\lambda^2} \hat{N} = -\zeta R^i_{jmn} u^j \hat{N}^m u^n e_i = -\text{sgn}(\hat{N}) \left(\hat{N}_i \zeta R^i_{jmn} u^j \hat{N}^m u^n \right) \hat{N},$$

where we have projected the right hand side along the unit normal whose sign is $\text{sgn}(\hat{N}) = \hat{N}_i \hat{N}^i$ in the event that our metric is not positive-definite. The tangential component of this equation $D^2 \alpha / d\lambda = 0$ requires α to be a linear function of the affine parameter, while the normal component of the equation reduces to the scalar equation

$$\frac{d^2 \zeta}{d\lambda^2} + \zeta \text{sgn}(\hat{N}) \left(\hat{N}_i R^i_{jmn} u^j \hat{N}^m u^n \right) = 0.$$

If we specialize this to an arclength parametrization $\lambda = s$ so that $u = \hat{u} = \hat{T}$, then it becomes

$$\frac{d^2\zeta}{ds^2} + \zeta \operatorname{sgn}(\hat{N}) \left(\hat{N}_i R^i_{jmn} \hat{T}^j \hat{N}^m \hat{T}^n \right) = 0.$$

In the orthonormal frame $(e_{\hat{1}}, e_{\hat{2}}) = (\hat{T}, \hat{N})$ arising from an unnormalized frame (e_1, e_2) , the curvature component expression

$$\operatorname{sgn}(\hat{N}) R_{\hat{2}\hat{1}\hat{2}\hat{1}} = R^{\hat{2}}_{\hat{1}\hat{2}\hat{1}} = \operatorname{sgn}(\hat{T}) R^{\hat{2}\hat{1}}_{\hat{2}\hat{1}} = \operatorname{sgn}(\hat{T}) R^{21}_{21} \equiv \operatorname{sgn}(?) R_{\text{gauss}}$$

is just the Gaussian curvature K of the surface up to a sign which is positive in the positive-definite case. Thus apart from a sign, the Gaussian curvature coefficient plays the role of a squared frequency when it is positive leading to oscillations in the solutions which is easy to analyze when the geodesic is a line of curvature along which this coefficient is constant. This means that a geodesic departing at a small angle from the base geodesic c on which this analysis is based will reach a maximum displacement from that geodesic and return to cross it. Thus positive curvature (ignoring the complications of the extra signs in the nonpositive-definite case) leads to a refocusing effect on the geodesic spray from the given geodesic, causing them to reconverge on that geodesic at what is called a Jacobi point. On the other hand a negative curvature spreads out the nearby geodesics exponentially, defocusing them.

The geodesic deviation equation

$$\frac{d^2\zeta}{ds^2} + K\zeta = 0$$

has constant K along a geodesic line of curvature and when $K > 0$ has oscillating solutions

$$\zeta = A \cos(\Omega s) + B \sin(\Omega s) = A \cos(2\pi s/\Lambda) + B \sin(2\pi s/\Lambda), \quad \Omega = \sqrt{K}, \quad \Lambda = 2\pi/\Omega.$$

This has a convergence arclength of half a wavelength $L = \Lambda/2 = \pi/\sqrt{K}$.

Exercise 9.4.1.

meridians on ellipsoid of revolution

An ellipsoid of revolution was introduced in Exercise 9.1.6. Here we adopt the parametrization

$$\langle x, y, z \rangle = \langle a \sin \theta \cos \phi, a \sin \theta \sin \phi, b \cos \theta \rangle,$$

leading to its metric

$$ds^2 = a^2 \cos^2 \theta d\phi^2 + (a^2 + (b^2 - a^2) \cos^2 \theta) d\theta^2,$$

namely

$$g_{\theta\theta} = \frac{b^2}{a^2 + (b^2 - a^2) \cos^2 \theta}, \quad g_{\phi\phi} = a^2 \cos^2 \theta.$$

a) Evaluate its nonzero connection components

$$\Gamma^\phi_{\phi\theta} = \Gamma^\theta_{\theta\phi} = -\tan \theta, \quad \Gamma^\theta_{\phi\phi} = \frac{a^2 \cos \theta \sin \theta}{a^2 + (b^2 - a^2) \cos^2 \theta}, \quad \Gamma^\theta_{\theta\theta} = \frac{(a^2 - b^2) \cos \theta \sin \theta}{a^2 + (b^2 - a^2) \cos^2 \theta}.$$

b) Evaluate the nonzero curvature component

$$R^\phi_{\theta\phi\theta} = \frac{b^2}{(a^2 + (b^2 - a^2)\cos^2\theta)^2}.$$

c) The meridians $\phi = \phi_0, \theta = \theta(\lambda)$ are geodesics so nearby meridians satisfy the geodesic deviation equation. If we assume λ is an arclength parametrization then

$$u^i = \frac{dx^i}{d\lambda} = \delta^i_\theta \frac{d\theta}{d\lambda} = \delta^i_\theta g_{\theta\theta}^{-1/2}.$$

Why?

d) Let $\eta^i = \eta_0 \delta^i_\phi$ or $\eta = \eta_0 \partial/\partial\phi$ be a separation vector between nearby meridians. Using the chain rule $df/d\lambda = f_{,\theta} d\theta/d\lambda$, show that

$$\frac{D\eta^i}{d\lambda} = \delta^i_\phi \Gamma^\phi_{\theta\phi} g_{\theta\theta}^{-1/2}.$$

e) Show that this separation vector satisfies

$$\frac{D^2\eta^i}{d\lambda^2} = -\delta^i_\phi R^\phi_{\theta\phi\theta} g^{\theta\theta} \eta_0.$$

Show that this is equivalent to being a solution of the geodesic deviation equation. ■

Exercise 9.4.2.

minimum convergence length on the ellipsoid of revolution

The equator $\theta = \pi/2$ is a geodesic along which the curvature is constant with value $K_{gauss} = b^2/a^4$, so that $\omega = b/a^2$ and $L = \pi a^2/b$. The condition that this be less than half the circumference of the equator (like a sphere) is

$$\frac{\pi a^2}{b} < \pi a \rightarrow \frac{b}{a} < 1,$$

so oblate ellipsoids $b/a < 1$ have a shorter convergence length, while prolate ellipsoids $b/a > 1$ have a longer value. For a simple closed geodesic (which has no self-intersections), at least two full oscillations about the equator must take place during one azimuthal revolution, which requires $b/a < 1/2$.

a) Use a computer algebra system to find the angle of inclination from the equator at $(\theta, \phi) = (0, 0)$ of the simple closed geodesic which makes exactly two oscillations about the equator during one azimuthal revolution for $b/a = 0.45$ using trial and error. Exactly one such geodesic exists for $1/3 < b/a < 1/2$. Repeat for $b/a = 0.4$. Raising the departure angle from the equator lengthens the focal length so for each ellipsoid in this interval, sufficiently increasing the focal length on the equator to half its circumference leads to this closed geodesic.

b) For $1/4 < b/a < 1/3$ there is a simple closed geodesic with three oscillations away from the equator during one full azimuthal revolution. Find it for $b/a = 0.3$. It appears that for $1/(m-1) < b/a < 1/m$ there exist simple closed geodesics making m oscillations during one azimuthal revolution, except for $n = 1$ where increasing the departure angle to $\pi/2$ leads only to the meridian geodesic at this limiting value.

One can extend this analysis to closed geodesics which make m oscillations about the equator during n azimuthal oscillations, as for the torus analysis. Similar considerations apply to any geodesic with constant positive curvature, like the concave inward extremal parallels of a surface of revolution. ■

Exercise 9.4.3.

minimum convergence length on the torus

a) Evaluate the convergence arclength L for the outer equator of the torus which is a line of constant curvature

$$R_{\text{gauss}} = \frac{\cos(r/b)}{b(a + b \cos(r/b))} \Big|_{r=0} = \frac{1}{b(a + b)}$$

which is the maximum value of the curvature function, and hence leads to the minimum convergence length. What fraction of the total circumference of the outer equator does this represent in general and for the unit torus $(a, b) = (2, 1)$?

b) Show that L reduces to half the circumference of a sphere in the degenerate case $a = 0$.

c) What is the condition on the ratio $a/b = c + 1$ for which the outer equator circumference $2\pi(a + b)$ is an integer multiple of the convergence length L ? Show that it is simply that $c + 2 = (a + b)/b = a/b + 1 = n^2$ is a perfect square. $n = 1$ corresponds to the sphere $a = 0$, while $n = 2$ corresponds to the ratio $a/b = 3$, which allows one complete oscillation between antipodal points of the outer equator, and $n = 3$ corresponds to the ratio $a/b = 8$. ■

Exercise 9.4.4.

minimum convergence length on the cavatappo surfaces

Redo the previous exercise for the cavatappo surfaces 1.0 and 2.0. How does the vertical stretching out of the torus to create these surfaces affect the minimum convergence length? ■

Exercise 9.4.5.

curvature of elliptic paraboloid

a) Use a computer algebra system to evaluate the curvature scalar for the elliptic paraboloid of Exercise 1.6.12

$$z = g(x, y) = \frac{1}{2} (9x^2 + 4y^2) .$$

using the parametrization

$$\langle x, y, z \rangle = \langle 2\rho \cos \phi, 3\rho \sin \phi, 18\rho^2 \rangle.$$

b) Show that the value of half the curvature scalar at the vertex $\rho = 0$ equals the determinant of the matrix of second derivatives of the original function $g(x, y)$ with respect to the Cartesian coordinates.

■

Exercise 9.4.6.

curvature of pseudospheres compared to corresponding hyperboloids in \mathbb{R}^3

a) Use a computer algebra system to evaluate the curvature scalar for the pseudospheres of 3-dimensional Minkowski spacetime discussed in Section 8.7.

b) Compare the constant negative curvature for the spacelike pseudospheres inside the light cone to the corresponding positive curvature hyperbolas of revolution in \mathbb{R}^3 . What is the extremal curvature in the Euclidean case (maximal absolute value of the curvature), which clearly has to occur at the vertex on the axis of revolution of the surface?

c) Compare the constant curvature for the timelike pseudospheres outside the light cone with the corresponding negative curvature hyperbolas of revolution \mathbb{R}^3 . What is the extremal curvature of the latter surfaces (maximal absolute value of the curvature), which clearly has to exist on the throat of the surface? Can you figure out the correct sign for the geodesic deviation equation on the pseudospherical surfaces of this type?

d) What can you say about the minimal geodesic convergence length for the positive curvature hyperboloid?

■

Chapter 10

Extrinsic curvature

10.1 The extrinsic curvature tensor

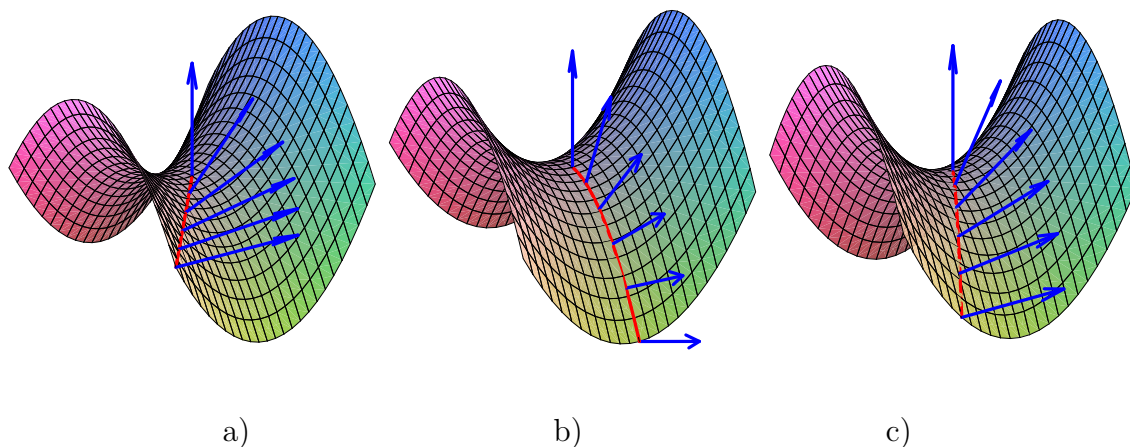


Figure 10.1: Two independent factors cause the surface normal to rotate as one moves along a curve: the rotation of the normal to the curve itself and the further sideways rotation due to the tilting of the surface along the curve. In a) only the sideways rotation along the straight line vertical cross-sectional curve is present. In b) along a line of curvature, the curve normal coincides with the surface normal so only the first rotation is present. Moving along a general cross-sectional curve like c), the normal must rotate because the curve itself is bending and because the surface is tilting underneath the curve with respect to the horizontal direction.

In multivariable calculus we encounter our first quantitative measure of curvature generalizing the inverse relationship between the radius of a circle and how tightly it is curved—we define the curvature of a circle to be the reciprocal of its radius to get started, and then we generalize this to other curves using the machinery of the unit tangent and unit normal to the curve, reviewed in Appendix C. Given a parametrized curve

$$x = x(t), \quad y = y(t), \quad z = z(t) \leftrightarrow \vec{r} = \vec{r}(t),$$

and its tangent vector $\vec{r}'(t)$, we can introduce a differential element of arclength ds along it

$$\frac{ds}{dt}(t) = |\vec{r}'(t)|$$

but in general we cannot integrate this exactly to yield an arclength function along the curve so that we can reparametrize it by the arclength. However, we can use the chain rule to evaluate arclength derivatives along the curve without reparametrization. The unit tangent is the first arclength derivative of the position vector of the curve, and the second derivative of the position vector (first derivative of the unit tangent) determines the unit normal as its direction and the nonnegative curvature as its length. The curvature actually measures how fast the unit tangent and orthogonal unit normal rotate in their plane relative to vectors which are momentarily not

rotating as one moves along the curve, and so is a kind of angular velocity, a fact not mentioned in multivariable calculus. The curving of the curve is quantified by how its direction, the unit tangent, rotates. We visualize this by introducing the osculating circle which is the circle of best fit to the curve at each point, using the reciprocal of the curvature as its radius, the radius of curvature of the curve, and its center a distance equal to that radius along the unit normal vector.

Surfaces are first encountered in multivariable calculus as graphs of functions of two variables $z = f(x, y)$ realized as surfaces in \mathbb{R}^3 . However, these are very restrictive so to be able to handle any kind of surface efficiently, we must follow the example of parametrized curves. A parametrized surface is handled just like a parametrized curve but with two parameters as reviewed in Appendix D

$$x = x(u, v), \quad y = y(u, v), \quad z = z(u, v) \leftrightarrow \vec{r} = \vec{r}(u, v) = r(u^1, u^2),$$

where we can use the indexed parameters when needed for indexed formulas. We can think of this 2-parameter set of points as a 1-parameter family of curves in two different ways. By holding v fixed we get a curve parametrized by u , and by holding u fixed we get a curve parametrized by v . Together these two families of curves form a grid on the surface which is used by computer algebra systems to give some 3- d perspective to the screen representations of surfaces in 3- d graphics. An actual grid is taken by showing equally spaced curves in each of the parameter intervals used to graph the surface. In other words one takes the more familiar rectangular grid in the u - v parameter plane and maps it onto the surface in \mathbb{R}^3 . One can then apply all the machinery to analyze the two families of curves, each with their own tangent vector, unit tangent, unit normal, curvature and osculating circle. From this information we would like some measure of the curvature of the surface they form. This is the idea of the extrinsic curvature of a surface.

Each unit normal to each of the grid lines rotates in general but this does not mean the surface is bending unless the tangent plane formed by the span of the two tangent vectors changes orientation. (Any nonrectangular parametrization of a flat plane would have a curved grid, but the surface would still be flat.) This is only reflected in the changing direction of the normal to the tangent plane. A unit normal is created by taking the cross-product of the two tangent vectors and normalizing it to a unit vector \vec{n} . The ordering of the two parameters picks out a unique such unit normal through this cross product of the tangent vectors in that order, and so associates what is called an outer orientation to the surface, just a choice of one side or the other of the tangent plane.

The extrinsic curvature of a truly “curved surface” in space (\mathbb{R}^3) is characterized by the fact that as one moves about on the surface, the surface bends, which can be qualitatively measured by the rotation of the direction of the normal to the surface. This is what bending means. A changing unit normal can only rotate since its length is constant. But if the tip of the normal rotates, it moves initially in the direction orthogonal to itself, which is to say, along a direction tangent to the surface, just like the radial unit normal to an ordinary sphere which rotates in the direction tangential to the sphere as one moves about on that sphere. Figure 10.1 illustrates the situation. Since one can move in any direction in the tangent plane to the surface to move along the surface, the initial arclength rate of change of the normal in that

direction is another vector in the tangent plane. In other words to measure the bending of the surface at a point, we need a function from tangent vectors in the tangent plane to new tangent vectors in the tangent plane, giving the arclength derivative of the unit normal as a function of the tangent vector to each curve moving away from that point within the surface. This turns out to be a linear map from vectors to vectors in the 2-dimensional tangent space to the surface, namely a mixed second rank tensor. This is called the extrinsic curvature tensor, and lives in the 2-dimensional tangent space to the surface.

If n is a unit normal vector field defined over a surface, and X is a tangent vector belonging to the tangent plane to the surface (for short a “surface vector”: $n \cdot X = 0$) at some point, then it makes sense to differentiate the normal along X since n is only defined on the surface itself. A simple calculation then backs up the previous claim about how the normal is allowed to change

$$n \cdot n = 1 \xrightarrow{\nabla_X} (\nabla_X n) \cdot n + n \cdot (\nabla_X n) = 0 \rightarrow n \cdot (\nabla_X n) = 0.$$

This result states that the derivative of the normal along the surface must be orthogonal to the normal and hence it lies in the subspace of the tangent space corresponding to the tangent plane to the surface. Thus $\nabla_X n$ is again a surface vector, so

$$X \rightarrow \nabla_X n \equiv -K(X) \equiv S(X)$$

defines a linear transformation of the subspace of the full tangent space which is tangent to the surface. This linear transformation can be identified with a $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ -tensor in the tangent space which takes surface vectors to surface vectors, and it can be extended by linearity to any input vector which is not necessarily a surface vector by setting $K(n) = 0$. K is called the extrinsic curvature tensor (useful in general relativity), while its sign-reversal S is called the shape operator, since as we will see it describes the shape of the surface within the enveloping space in which it sits. Expressed in components we have

$$K(X)^i = K^i_j X^j.$$

The overall sign of the extrinsic curvature is not really significant except in relation to the chosen normal, since changing the sign of the normal vector simply reverses the side of the tangent plane on which it lies and changes the sign of the extrinsic curvature as well. Any geometry associated with the extrinsic curvature tensor must be understood to be related to the choice of one of the two possible continuous unit normal vector fields defined on the surface, assuming that such a continuous choice is possible. Orientable surfaces have this property.

If Y is another surface vector, then we can dot it into this expression using the metric

$$-Y \cdot \nabla_X n = Y \cdot K(X) = Y^k g_{ki} K^i_j X^j \equiv K_{kj} Y^k X^j = K^b(Y, X),$$

which leads to the value of the index lowered $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$ -tensor on the two vectors. However, if Y is a surface vector, it is orthogonal to the normal

$$Y \cdot n = 0,$$

so if one differentiates this condition, one finds

$$0 = \nabla_X(Y \cdot n) = (\nabla_X Y) \cdot n + Y \cdot \nabla_X n = n \cdot (\nabla_X Y) - Y \cdot K(X) = n \cdot (\nabla_X Y) - K^b(Y, X),$$

and therefore

$$K^b(Y, X) = n \cdot (\nabla_X Y) = -S^b(Y, X).$$

This says that the full evaluation of the covariant form of the extrinsic curvature tensor on two surface vectors is the normal component of the derivative of one surface vector along the other.

In fact this covariant extrinsic curvature tensor is a symmetric tensor

$$K^b(X, Y) = K^b(Y, X) \leftrightarrow K_{ij} = K_{ji},$$

whose sign-reversal $S^b = -K^b$ is called the second fundamental form of the surface in classical differential geometry, the first fundamental form being the metric tensor of the surface. These two tensors are both symmetric covariant tensors, which are called quadratic forms in the old language, just another name for a quadratic function of a vector variable. The intrinsic metric of the surface determines its intrinsic geometry, while the extrinsic curvature tensor determines its extrinsic geometry within the larger space, so they are both fundamental to the geometry of the surface in the context of the larger space. Since a symmetric 2×2 matrix has three independent components, these six functions associated with these two symmetric tensors on the surface determine the local geometrical properties of the surface. The sign difference between the extrinsic curvature and the second fundamental form is a matter of convention related to converging versus diverging world lines in general relativity (the gravitational field is attractive).

To show this symmetry property of the extrinsic curvature tensor easily, consider two vector fields X and Y defined throughout space whose values on the surface are surface vectors, and whose Lie bracket $[X, Y]$ on the surface is also a surface vector. If X and Y truly are surface vectors, they can be represented in terms of coordinates on the surface, and in those coordinates one can compute their Lie bracket, which will be a linear combination of the surface coordinate frame vectors, and hence will also be a surface vector. In fact the condition that two vector fields which are linearly independent at each point be tangent to 2-surfaces is that their Lie bracket be expressible as a linear combination of the two vector fields themselves on those 2-surfaces for exactly this reason.

The condition that the metric connection be symmetric was shown to be

$$\nabla_X Y - \nabla_Y X = [X, Y] \leftrightarrow \Gamma^k_{ij} = \Gamma^k_{ji} \text{ (coordinate components only)}.$$

Now dot this equation with n to get the covariant extrinsic curvature on the left hand side

$$K^b(Y, X) - K^b(X, Y) = n \cdot \nabla_X Y - n \cdot \nabla_Y X = n \cdot [X, Y] = 0,$$

which vanishes if $[X, Y]$ is also a surface vector as assumed.

Is it a problem to have assumed that X , Y , and $[X, Y]$ are vector fields defined throughout space whose values on the surface belong to the subspace of the tangent space along the surface,

rather than only being defined on the surface itself? No. Suppose we take any function $x^{1'}$ on space for which the surface in question is a level surface of this function $x^{1'} = x_0^{1'}$, and let $x^{2'}, x^{3'}$ be any two other independent differentiable functions such that the Jacobian matrix $\partial x^{i'}/\partial x^j$ has a nonzero derivative everywhere on the surface. $\{x^{i'}\}$ can then be taken as a new system of coordinates which are adapted to this surface, in which any vector field which has a vanishing first component on the surface is a surface vector there. Obviously the Lie bracket of any two such vector fields is also a surface vector on the surface. Basically such vector fields on the whole space extend vector fields defined only on the surface so that when their values on the surface are considered, directions off the surface are irrelevant.

Note that

$$K^b(X, X) = -X \cdot \nabla_X n = n \cdot \nabla_X X = -S^b(X, X).$$

Suppose we let $X = \hat{T}(t)$ be the unit tangent to a curve $c(t)$ in the surface which is the cross-section of the surface by a plane through the normal line to the surface at $c(t_0)$. This is by definition a plane curve whose unit normal vector $\hat{N}(t)$ (the direction of $\hat{T}'(t_0)$) therefore lies in the same plane, and lies on the side of the tangent line in that plane on which the curve itself lies (as long as the curvature is nonzero). Call such a curve at this point a normal cross-sectional curve there.

Then modulo the sign, by construction the unit normal to this normal cross-sectional curve is the normal to the surface at the given point: $\hat{N}(t_0) = \pm n|_{c(t_0)}$, where the sign here depends on which side of the surface the choice of surface normal lies in comparison to the curve's normal whose direction is determined. Thus the covariant extrinsic curvature component along the unit tangent

$$K^b(\hat{T}(t_0), \hat{T}(t_0)) = n|_{c(t_0)} \cdot \nabla_{\hat{T}(t_0)} \hat{T}(t_0) = \pm \hat{N}(t_0) \cdot \nabla_{\hat{T}(t_0)} \hat{T}(t_0) = \pm \kappa(t_0)$$

is just the curvature of the curve modulo the sign since $\nabla_{\hat{T}(t)} \hat{T}(t) = \kappa(t) \hat{N}(t)$ equals the arclength derivative of the unit tangent along the curve. Thus the extrinsic curvature tensor is a machine which produces the curvature of such normal cross-sectional curves along any direction within the surface at each point of the surface.

As a symmetric second rank tensor represented by a 2×2 symmetric matrix in an orthonormal frame, the extrinsic curvature can always be diagonalized by an orthogonal transformation (rotation) by expressing it in an orthogonal frame of eigenvectors, which are called the principal directions of the extrinsic curvature, while the eigenvalues (reversed in sign) are called the principal curvatures k_1 and k_2 . Since the overall sign of these eigenvalues is reversed by changing the sign of the normal vector field, only their relative sign has any significance. Let $r_i = 1/|k_i|$ define corresponding radii of curvature. However, for the sphere of radius r_0 where we want curvature to be positive for an outward normal (see Fig. 10.2 a), corresponding to a bending away of the nearby normals, it is the sign-reversed extrinsic curvature which has the repeated eigenvalue $k_1 = k_2 = 1/r_0$, so it makes sense to call the eigenvalues of the sign-reversed extrinsic curvature the principal curvatures. The eigenvectors (tangent vectors in the tangent plane to the surface) are called the principal directions of curvature and are always orthogonal. Each has an osculating circle associated with the normal plane cross-section of the surface (the intersection with the surface of a plane through the normal line) in that direction which are the circles of best fit to the surface in these two orthogonal directions.

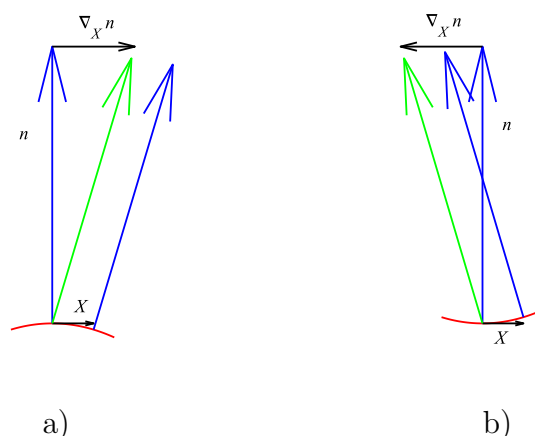


Figure 10.2: To compare the normal at a nearby new point (approximately at the tip of X in the tangent plane) with the normal at that starting point, one must parallel transport the new normal back to the original point and then take the difference. This difference then approaches the covariant derivative of the normal in the limit of small displacement. In figure a) where n is antiparallel to \hat{N} , as one moves along a principal direction of curvature, when the principal curvature along that direction is positive (but a negative eigenvalue of the extrinsic curvature) the normal rotates towards the direction of motion, i.e., away from itself. In figure b) where n is parallel to \hat{N} , when that principal curvature is negative (but a positive eigenvalue of the extrinsic curvature), the normal vector rotates backwards, i.e., towards itself. On the other hand when the extrinsic curvature is positive as in case b), the normal lines converge, useful in describing the paths of self-gravitating particles in spacetime which tend to attract each other. Interestingly enough for a spacelike surface in 3-dimensional Minkowski spacetime, the normals in figures (a) and (b) behave oppositely (for the concave up case as you move from the vertical normal at the center the normals tilt farther away from that vertical normal rather than back towards it.)

Points for which the two eigenvalues are equal are called umbilic points, and all directions at such points are principal directions. Points may be classified by the relative signs of the two eigenvalues. If they are both the same (opposite) sign, a point is called elliptic (hyperbolic), while if one is zero but the other is not, the point is called parabolic. For elliptic points the centers of curvature of the two osculating circles for cross-sectional curves along the two principal directions are on the same side of the surface (like for a sphere or ellipsoid), but on the opposite side for hyperbolic points (like a saddle). For a cylinder which bends only along one direction (perpendicular to its axis of symmetry), all points are parabolic.

At nonumbilic points, lines of curvature are defined as curves whose tangent is a principal directions of curvature, i.e., an eigenvector of the extrinsic curvature at each point along the curve. For a surface with isolated umbilic points, no special care is required to implement this concept, but on a surface like a sphere where all directions are principal directions, one loses these special curves. It might be interesting as an example to see what happens in the case

of a family say of triaxial ellipsoids (unequal principal axes) whose limit is a sphere. These complications do not seem to be discussed in the introductory discussions one finds available electronically.

The trace and determinant of the component matrix \underline{K} of the extrinsic curvature tensor K are invariants, since under a change of frame the matrix transforms according to $\underline{K}' = \underline{A} \underline{K} \underline{A}^{-1}$, and $\text{Tr } \underline{K} = \text{Tr } \underline{K}'$ and $\det \underline{K} = \det \underline{K}'$. One can therefore evaluate them in a frame of eigenvectors in terms of the eigenvalues. The trace of the extrinsic curvature tensor reversed in sign $-\text{Tr } \underline{K} = -K^i_i = k_1 + k_2$ is the sum of the principal curvatures, called the mean curvature, while its determinant $\det \underline{K} = k_1 k_2$ is their product (independent of the arbitrary overall sign of the extrinsic curvature), called the Gaussian curvature. We will see later that this is very closely related to the Riemannian curvature. When the Gaussian curvature is positive (elliptic points), the surface is said to be positively curved and the surface locally lies on one side of its tangent plane. When the Gaussian curvature is negative (hyperbolic points), the surface is said to be negatively curved, and the tangent plane must cut the surface (like in the case of a saddle) since some normals to normal cross-sectional curves lie on one side of the plane and others lie on the opposite side. When the Gaussian curvature is zero (parabolic points), the surface is just said to be flat. Of course here we are assuming Euclidean geometry. In the geometry of 3-dimensional Minkowski spacetime, these statements change.

The sign of the principal curvatures indicates whether the surface is bending up or down with respect to the normal direction. A positive principal curvature means that the normal is rotating away from the current normal direction as one moves along the associated principal direction (bending down) while a negative one means that it is rotating towards the current normal direction (bending up). Figure 10.2 shows the two cases. A zero value of a principal curvature means that the normal is not rotating along that direction. Another way of understanding the sign of the principal curvatures is to imagine following the normal lines to the surface in the direction of the normal vector. Positive curvature means that these normal lines are spreading apart from each other, while negative curvature means that they are converging towards each other. The extrinsic curvatures, terminology used in the application of differential geometry to general relativity, the relativistic theory of gravity, have the opposite sign compared to the principal curvatures so that positive values correspond to convergence of the normal lines, which is useful in applications to gravitational theory where the presence of mass leads to convergence of the geodesic paths of test particles because of the attractive nature of gravity.

One way of comparing all the normals to a surface in space is simply to parallel transport them all to the same point in space (say the origin of \mathbb{R}^3), with their initial points at the origin of the tangent space, so that their tips all lie on the unit sphere in that tangent space. This enables one to interpret the normal vector field as a map from the surface to the unit sphere, which is called the Gauss map. For a sphere of any radius about the origin, for example, this Gauss map just projects a point on that sphere to the corresponding point on the unit sphere obtained by just dividing the position vector by its length. For a vertical cylinder whose symmetry axis is the z -axis, and whose normals are therefore all horizontal pointing in all possible horizontal directions, the Gauss map sends all the normals to the equator of the unit sphere, tracing out the entire equator. For a more irregular surface like the saddle of figure 10.1, the Gauss map is more complicated. Figure 10.3 shows the image of the Gauss map for the three curves shown

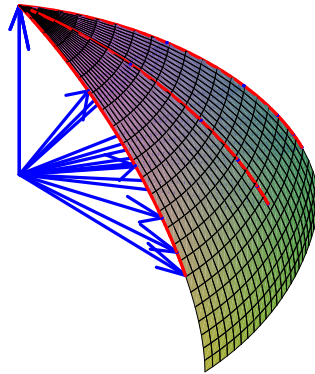


Figure 10.3: If one plots the normals with their tails all at the origin, one visualizes the Gauss map from the surface to the unit sphere. Here the images of the three curves from figure 10.1 are shown in the part of the unit sphere $0 \leq \theta \leq \pi/2$, $\pi/2 \leq \phi \leq 3\pi/4$ above the first half of the second quadrant of the horizontal plane.

in figure 10.1.

10.2 Spheres, cylinders and cones: some useful concrete examples

Rather than launching into an abstract discussion leading to a formula for the components of the extrinsic curvature in a coordinate system adapted to a general surface, let's get explicit. Spheres and cylinders in \mathbb{R}^3 are good examples of surfaces which respectively have nontrivial intrinsic and extrinsic curvature (spheres) or are extrinsically curved but intrinsically flat (cylinders). Because these surfaces reside within a flat 3-space, the intrinsic and extrinsic curvatures are locked together in a way that can be quantified with a simple calculation.

In both cylindrical coordinates $\{\rho, \phi, z\}$ and spherical coordinates $\{r, \theta, \phi\}$, the first coordinate is adapted to these surfaces: cylinders of radius ρ with the z -axis as a symmetry axis and spheres of radius r centered at the origin. Holding these fixed, the remaining two coordinates serve to describe the corresponding surface.

The orthonormal frame $\{e_{\hat{r}}, e_{\hat{\theta}}, e_{\hat{\phi}}\}$ is adapted to the family of concentric spheres of radius r , for which $n = e_{\hat{r}}$ is the unit normal vector field, which happens to be covariant constant in its own direction. $\{e_{\hat{\theta}}, e_{\hat{\phi}}\}$ provide an orthonormal frame for the vector fields which are tangent to each sphere in the sense that they belong to the subspace of the tangent space containing the angular directions along the sphere. From the definition of the extrinsic curvature tensor, we can re-interpret the covariant derivative relations

$$\begin{aligned}\nabla_{e_{\hat{\theta}}} e_{\hat{r}} &= \underbrace{\Gamma^{\hat{\theta}}_{\hat{\theta}\hat{r}}}_{-K^{\hat{\theta}}_{\hat{\theta}} = r^{-1}} e_{\hat{\theta}} + \underbrace{\Gamma^{\hat{\phi}}_{\hat{\theta}\hat{r}}}_{-K^{\hat{\phi}}_{\hat{\theta}} = 0} e_{\hat{\phi}}, \\ \nabla_{e_{\hat{\phi}}} e_{\hat{r}} &= \underbrace{\Gamma^{\hat{\theta}}_{\hat{\phi}\hat{r}}}_{-K^{\hat{\theta}}_{\hat{\phi}} = 0} e_{\hat{\theta}} + \underbrace{\Gamma^{\hat{\phi}}_{\hat{\phi}\hat{r}}}_{-K^{\hat{\phi}}_{\hat{\phi}} = r^{-1}} e_{\hat{\phi}}.\end{aligned}$$

reading off the values of the connection components found in section 7.3. The extrinsic curvature tensor is therefore proportional to the identity tensor on the subspace

$$-K = r^{-1} Id_{(2)} \leftrightarrow -K^{\hat{i}}_{\hat{j}} = r^{-1} \delta^i_j, \quad i, j = 2, 3,$$

which means that the covariant form of the extrinsic curvature tensor is proportional to the metric of the sphere

$$-K^b = r^{-1}(\omega^{\hat{\theta}} \otimes \omega^{\hat{\theta}} + \omega^{\hat{\phi}} \otimes \omega^{\hat{\phi}}).$$

This means that the spherical orthonormal frame vectors tangent to the sphere are eigenvectors of the extrinsic curvature tensor with eigenvalues both equal to the sign-reversed reciprocal radius of the sphere. The principal curvatures are therefore both equal to $r^{-1} = k_1 = k_2$ and the sphere is an umbilic surface consisting entirely of umbilic points. Since the single independent eigenvalue is degenerate, all directions in the subspace of the tangent space along the sphere are equivalent: the extrinsic curvature is invariant under rotations of the tangent space about the radial direction. The Gaussian curvature $k_1 k_2 = r^{-2}$ is positive and constant.

Notice that the integral of the Gaussian curvature over the sphere is just this constant $1/r^2$ times the surface area $A = 4\pi r^2$ of the sphere, namely just 4π . This is not a coincidence but reflects something much deeper (associated with the Gauss-Bonnet Theorem).

The connection 1-form matrix can be then written

$$\begin{aligned}\underline{\hat{\omega}} &= -K^{\hat{\theta}}_{\hat{\theta}}\omega^{\hat{\theta}} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - K^{\hat{\phi}}_{\hat{\phi}}\omega^{\hat{\phi}} \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} + \omega^{\hat{\theta}}_{\hat{\phi}} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \\ &= d\theta \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \sin\theta d\phi \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} + \cos\theta d\phi \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix},\end{aligned}$$

showing its decomposition into the θ - ϕ block describing the intrinsic covariant derivative on the spheres and the remaining extrinsic part describing the rotation of the normal as one moves along the sphere. The two rotations of the adapted orthonormal frame rotate surface vectors as the surface tilts to keep them in the surface, while the intrinsic rotation simply rotates about the normal direction within the surface. This is indeed the case for more general situations of a p -surface within an n -dimensional space in which adapted coordinates pick out this p -surface and an orthonormal frame is adapted to the corresponding orthogonal decomposition of its tangent spaces. The connection 1-form matrix generates rotations/pseudorotations of the adapted orthonormal frame consisting of one part with leaves those elements of the frame orthogonal to the surface fixed (intrinsic rotations) and those which rotate those extrinsic frame vectors, which describe a generalized extrinsic curvature with an extra index to take into account the multiple normal directions.

If we repeat this for the cylinders of cylindrical coordinates, then $e_{\hat{\rho}}$ is the unit normal vector field (also covariant constant along its own direction) which only rotates if one moves horizontally around the cylinders, but remains constant along the vertical direction

$$\begin{aligned}\nabla_{e_{\hat{\phi}}}e_{\hat{\rho}} &= \underbrace{\Gamma^{\hat{\phi}}_{\hat{\phi}\hat{\rho}}}_{-K^{\hat{\phi}}_{\hat{\phi}} = \rho^{-1}} e_{\hat{\phi}} + \underbrace{\Gamma^{\hat{\phi}}_{\hat{\phi}\hat{z}}}_{-K^{\hat{z}}_{\hat{\phi}} = 0} e_{\hat{z}}, \\ \nabla_{e_{\hat{z}}}e_{\hat{\rho}} &= \underbrace{\Gamma^{\hat{\phi}}_{\hat{z}\hat{\rho}}}_{-K^{\hat{\phi}}_{\hat{z}} = 0} e_{\hat{\phi}} + \underbrace{\Gamma^{\hat{z}}_{\hat{z}\hat{\rho}}}_{-K^{\hat{z}}_{\hat{z}} = 0} e_{\hat{z}},\end{aligned}$$

so

$$S = -K = \rho^{-1}e_{\hat{\phi}} \otimes \omega^{\hat{\phi}}.$$

The two frame vectors $e_{\hat{\phi}}, e_{\hat{z}}$ spanning the cylinder directions are eigenvectors of the extrinsic curvature tensor but with distinct eigenvalues $-\rho^{-1}$ and 0, corresponding to the bending of the surface only in the angular direction, with the radius of curvature equal to the radius of the circular cross-section. The Gaussian curvature is zero, which is a consequence of the fact that its intrinsic geometry is flat, as we will see later.

The connection 1-form matrix here consists only of the single extrinsic curvature term since the intrinsic cylindrical geometry is flat

$$\underline{\hat{\omega}} = \underbrace{-K^{\hat{\phi}}}_{d\phi} \omega^{\hat{\phi}} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The two frame vectors $e_{\hat{\phi}}, e_{\hat{z}}$ are principal directions, and their coordinate lines are (orthogonal) lines of curvature. Notice that they are also geodesics.

Cones: a useful cautionary example

Most of what we do is local, but global issues are important. The coordinate surfaces of the polar angle θ in spherical coordinates measured down from the positive z -axis are half cones, which are intrinsically flat but not differentiable at the vertex. This single problem point leads to very interesting global curvature effects. Like unwrapping a cylinder by cutting it along a coordinate line of z and flattening it out into a strip of a plane, a cone can be cut along a ray of the azimuthal ϕ coordinate line and flattening it out into a plane with a wedge missing. This missing wedge leads to a nontrivial rotation of any vector parallel transported around a loop that contains the vertex, even though on every loop that does not contain the vertex if one parallel transports a vector around it, the vector returns to its original position.

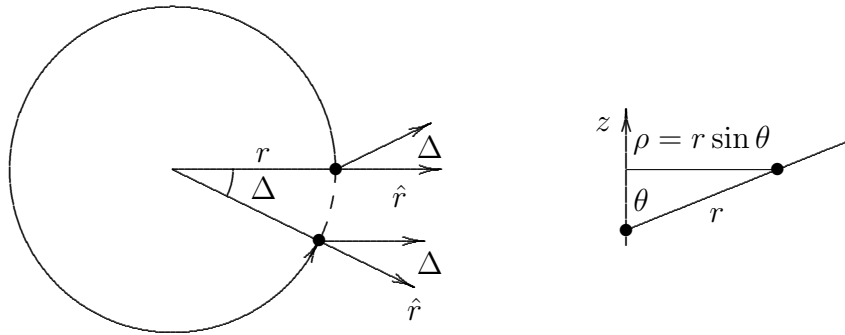


Figure 10.4: Parallel transport around a circle centered on the vertex of a cone of the polar angle θ in spherical coordinates, which has an angular “defect” $\Delta = 2\pi(1 - \sin \theta)$.

Figure 10.4 nicely illustrates this feature of the conical geometry. The circumference of a ϕ -coordinate line is $C = 2\pi\rho = 2\pi r \sin \theta$, but when the cone is cut along the $\phi = 0$ coordinate line and flattened out to a plane, the resulting circle has larger circumference $2\pi r$. The difference defines the defect angle

$$2\pi r - 2\pi r \sin \theta = r\Delta \rightarrow \Delta = 2\pi(1 - \sin \theta).$$

An angle $\theta = \pi/6$ leads to $\Delta = \pi$, for example, while an angle $\theta \approx 48.6$ degrees leads to $\Delta = \pi/2$ and an angle $\theta \approx 66.4$ degrees leads to $\Delta = \pi/6$.

When the radial unit vector \hat{r} is parallel transported from the cut around the ϕ coordinate circle in the clockwise direction, remaining horizontal in the flattened cone, it will have rotated forward by the defect angle when it returns to the cut. Any circle in the flattened out cone which does not contain the vertex will have no such rotation associated with it. Thus although the cone is locally flat, it has a global property of curvature due to the single bad point at its vertex. Such global issues can have important physical consequences.

This feature of parallel transport was investigated in Exercise 8.4.2 for the tangent cone to a surface of revolution, but holds also for the cone geometry itself. This is ironic, it is locally flat yet exhibits global curvature because of this topological defect revealed by the “singularity” at the vertex where the surface is no longer differentiable. In general curvature holonomy discusses the effect of curvature on parallel transport around closed curves which results in a rotation of the initial vector at the start to the final vector at the finish at the same point. Since the local geometry of the cone is flat, as long as our closed loop curve does not encircle the vertex singularity, there is no rotation, but if the loop does go around the vertex, there is a rotation by a fixed angle that was calculated in that exercise, independent of the radius.

10.3 Extrinsic curvature as a quadratic approximation

The covariant extrinsic curvature or shape tensor is often introduced concretely using a quadratic approximation to the surface in the same way that the curvature of a curve is related to a quadratic approximation to the curve relative to its tangent line, which is realized explicitly through the osculating circle. This is relatively easy to describe in the language of multivariable calculus in which all the tangent vectors in space are dealt with using their Cartesian component vectors. Covariant differentiation of these tangent vectors along another tangent vector then reduces to the ordinary directional derivative along those tangent vectors using partial differentiation.

When the surface is a single surface parametrized by two parameters as described in Appendix D as a parametrized surface, the manipulations are simpler and we will describe them now. This includes the previous cases in which the surface occurs as part of a larger coordinate system on the space, where two of the three coordinates parametrize the coordinate surfaces of the third coordinate. We will handle implicitly defined surfaces in a subsequent section.

So we begin with the parametrization

$$\vec{r}(u^1, u^2) = \langle x^1(u^1, u^2), x^2(u^1, u^2), x^3(u^1, u^2) \rangle,$$

and the first partial derivatives give us the Cartesian component vectors of the coordinate frame on the surface, and of the right handed normal to the surface and its unit normal direction

$$\begin{aligned}\vec{r}_1(u^1, u^2) &\equiv \frac{\partial \vec{x}}{\partial u^1}(u^1, u^2) \sim \frac{\partial}{\partial u^1}, \\ \vec{r}_2(u^1, u^2) &\equiv \frac{\partial \vec{x}}{\partial u^2}(u^1, u^2) \sim \frac{\partial}{\partial u^2}, \\ \vec{n}(u^1, u^2) &\equiv \vec{r}_1(u^1, u^2) \times \vec{r}_2(u^1, u^2), \\ \hat{n}(u^1, u^2) &\equiv \frac{\vec{n}(u^1, u^2)}{|\vec{n}(u^1, u^2)|}.\end{aligned}$$

These determine the linear approximation to the surface at a given point $\vec{r}_0 = \vec{r}(u_0^1, u_0^2)$, whose graph is the tangent plane

$$\vec{r}(u^1, u^2) = \vec{r}(u_0^1, u_0^2) + \vec{r}_1(u^1, u^2)\Delta u^1 + \vec{r}_2(u^1, u^2)\Delta u^2,$$

which for sufficiently small increments in the parameters we can identify with a surface in the tangent space, namely as the zero-value surface of a covector, just by replacing the increments by the corresponding differentials

$$\vec{r}_1(u_0^1, u_0^2)du^1 + \vec{r}_2(u_0^1, u_0^2)du^2 = 0.$$

In fact we can drop the 0 subscripts to apply this to any point on the surface, which is just the differential of the vector-valued function \vec{r}

$$d\vec{r}(u^1, u^2) = \vec{r}_1(u^1, u^2)du^1 + \vec{r}_2(u^1, u^2)du^2 = \vec{r}_a(u^1, u^2)du^a,$$

where $a, b = 1, 2$ allows us to use index summation formulas. The self-inner product of this vector differential is the metric line element or squared differential of arclength

$$ds^2 = d\vec{r}(u^1, u^2) \cdot d\vec{r}(u^1, u^2) = g_{ab} du^a du^b, \quad g_{ab} = \vec{r}_a \cdot \vec{r}_b.$$

In the old terminology of the subject this is called the first fundamental form of the surface.

Next we introduce the second derivatives which correspond to the covariant derivatives of the coordinate frame vector fields along those same vector fields

$$\begin{aligned} \vec{r}_{11}(u^1, u^2) &\equiv \frac{\partial^2 \vec{x}}{\partial (u^1)^2}(u^1, u^2) \sim \nabla \frac{\partial}{\partial u^1} \frac{\partial}{\partial u^1}, \\ \vec{r}_{22}(u^1, u^2) &\equiv \frac{\partial^2 \vec{x}}{\partial (u^2)^2}(u^1, u^2) \sim \nabla \frac{\partial}{\partial u^2} \frac{\partial}{\partial u^2}, \\ \vec{r}_{12}(u^1, u^2) &\equiv \frac{\partial^2 \vec{x}}{\partial u^2 \partial u^1}(u^1, u^2) \sim \nabla \frac{\partial}{\partial u^2} \frac{\partial}{\partial u^1}, \\ \vec{r}_{21}(u^1, u^2) &\equiv \vec{r}_{12}(u^1, u^2). \end{aligned}$$

yielding a symmetric matrix of such component vectors (using the obvious abbreviation $\nabla_1 = \nabla_{\partial_1}$, etc.)

$$\begin{pmatrix} \vec{r}_{11} & \vec{r}_{12} \\ \vec{r}_{12} & \vec{r}_{22} \end{pmatrix} \xrightarrow{\hat{n} \cdot} \begin{pmatrix} \hat{n} \cdot \vec{r}_{11} & \hat{n} \cdot \vec{r}_{12} \\ \hat{n} \cdot \vec{r}_{12} & \hat{n} \cdot \vec{r}_{22} \end{pmatrix} \sim \begin{pmatrix} \hat{n} \cdot \nabla_1 \partial_1 & \hat{n} \cdot \nabla_2 \partial_1 \\ \hat{n} \cdot \nabla_1 \partial_2 & \hat{n} \cdot \nabla_2 \partial_2 \end{pmatrix}$$

where the order of partial derivatives does not matter (for differentiable fields as we assume). When the unit normal is dotted into this matrix of vectors, we get by definition the sign-reversed shape tensor or the extrinsic curvature since $\hat{n} \cdot \nabla_X Y = S^b(X, Y)$, so

$$\begin{pmatrix} \hat{n} \cdot \vec{r}_{11} & \hat{n} \cdot \vec{r}_{12} \\ \hat{n} \cdot \vec{r}_{12} & \hat{n} \cdot \vec{r}_{22} \end{pmatrix} = - \begin{pmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{pmatrix} = \begin{pmatrix} K_{11} & K_{12} \\ K_{12} & K_{22} \end{pmatrix}.$$

This implies that the original vectors are related by

$$\nabla_{\partial_a} \partial_b = \frac{S_{ab}}{(\hat{n} \cdot \hat{n})} \hat{n} + {}^{(2)}\Gamma^c_{ab} \partial_c$$

so that the inner product with the normal removes the self-inner product sign, and the rest must be a surface vector, which in fact introduces the connection components of the metric connection of the surface itself

$${}^{(2)}\Gamma^c_{ab} \partial_c = {}^{(2)}\nabla_{\partial_a} \partial_b.$$

In other words the surface covariant derivative is simply the orthogonal projection into its tangent plane of the full space covariant derivative.

Now if we consider the quadratic Taylor polynomial approximation to the surface

$$\begin{aligned} \vec{r}(u^2, u^2) &= \vec{r}(u_0^1, u_0^2) + \vec{r}_1(u^1, u^2) \Delta u^1 + \vec{r}_2(u^1, u^2) \Delta u^1 \\ &+ \frac{1}{2} (\vec{r}_{11}(u^1, u^2) (\Delta u^1)^2 + 2\vec{r}_{21}(u^1, u^2) \Delta u^1 \Delta u^1 + \vec{r}_{22}(u^1, u^2) (\Delta u^2)^2), \end{aligned}$$

then the quadratic terms in the normal component of the difference vector with the fixed point at the origin of the tangent plane, with the notational change $\Delta u^a \rightarrow du^a$, leads to a second quadratic form

$$\hat{n} \cdot (\vec{r}(u^2, u^2) - \vec{r}(u_0^1, u_0^2)) = -\frac{1}{2} S_{ab}(u_0^1, u_0^2) du^a du^b = \frac{1}{2} K_{ab}(u_0^1, u_0^2) du^a du^b,$$

Removing the factor of two and the zero subscript from the right hand side of this equation, this defines what is called the second fundamental form of the surface in the old terminology, with a plus sign or a minus sign depending on your whim—both signs are found in the literature.

Exercise 10.3.1.

Taylor approximation to the sphere

Suppose we consider a sphere of radius a with its South pole at the origin, solving for the lower hemisphere function:

$$x^2 + y^2 + (z - a)^2 = a^2 \rightarrow z = a \pm \sqrt{a^2 - x^2 - y^2} \equiv f_{\pm}(x, y)$$

a) Use a computer algebra system to evaluate the Taylor quadratic polynomial approximation at the origin of the function f_- whose graph is the lower hemisphere, where the tangent space is the $z = 0$ plane, the unit normal is $\langle 0, 0, 1 \rangle$ and the quadratic approximation gives directly z versus (x, y) . Identify the covariant shape tensor from the quadratic coefficients.

b) Repeat for the upper hemisphere at the North pole and look at the Taylor quadratic polynomial approximation for $f_+ - 2a$.

■

Exercise 10.3.2.

monkeysaddle degeneracy

a) The monkey saddle surface $z = xy^2 - x^3$ is the graph of a third degree polynomial which is its own Taylor polynomial at the origin with only third degree terms, so the shape tensor at the origin vanishes. Plot this surface together with its tangent plane to see that it has three ridges which peek up above the horizontal tangent plane $z = 0$ at the origin, no matter how small the plotting window, as long as the surface can be distinguished from its tangent plane.

b) Use a computer algebra system to evaluate the matrix of components of the shape tensor using the normal projected second derivatives of the parametrized position vector $\vec{r}(u^1, u^2) = \langle u^1, u^2, u^1(u^2)^2 - (u^1)^3 \rangle$. Show that its eigenvalues are always of opposite signs everywhere away from the origin where its principal curvatures are nonzero and principal curvature directions are well-defined. The origin is a degenerate point of the surface where the shape tensor vanishes so its limiting intersection of the surface with its tangent plane there does not have to be a single point or 1 line or 2 crossed lines through the origin. At all other points, this intersection is described by a pair of crossed lines, since the principal curvatures are of opposite signs and the zero values of the covariant shape tensor in the surface tangent space result in such a pair.

c) Evaluate and plot the Gaussian curvature, which is always negative except at the origin. From your plot you can see 4 critical points of this function which are local minima. Find them

and locate them on the surface by plotting a small sphere at those points. See how you can color your surface using the Gaussian curvature and plot the surface this way. ■

Exercise 10.3.3.

Gaussian curvature of a surface of revolution

a) Consider a surface of revolution with cylindrical coordinates $\rho = R(u)$ and $z = Z(u)$ describing the profile curve

$$\vec{r}(u, v) = \langle R(u) \cos v, R(u) \sin v, Z(u) \rangle.$$

We evaluated the line element for the surface in Exercise 8.4.1

$$ds^2 = \underbrace{(R'(u)^2 + Z'(u)^2)}_{g_{rr}(u)} du \otimes du + R(u)^2 dv \otimes dv$$

Use a computer algebra system to evaluate the shape tensor and show that the Gaussian curvature is given by

$$\begin{aligned} K_{gauss} &= \frac{g'_{rr}(u)R'(u) - 2g_{rr}(u)R''(u)}{2g_{rr}(u)R(u)} \\ &= \frac{Z'(u)/R(u)}{(R'(u)^2 + Z'(u)^2)^2} \begin{vmatrix} R'(u) & Z'(u) \\ R''(u) & Z''(u) \end{vmatrix}. \end{aligned}$$

b) When $g_{rr}(u) = 1$ this simplifies to $-R''(u)/R(u)$. The same thing is true at a parallel which is an extremum of the azimuthal radius: $R'(u_0) = 0$. Then

$$K_{gauss}(u_0) = -R''(u_0)/R(u_0).$$

c) Can we generalize this easily to a screw-rotation symmetry surface starting from a similar profile curve at $\phi = 0$? ■

Exercise 10.3.4.

Gaussian curvature of a helicoid

Evaluate the shape tensor and its eigenvector structure and the Gaussian curvature and for the helicoid

$$\vec{r}(u, v) = \langle u \cos(v), u \sin(v), cv \rangle, \quad c \neq 0$$

and play with its properties. Consider its geodesics if it moves you. ■

10.4 Total curvature: intrinsic plus extrinsic curvature

To see how intrinsic and extrinsic curvature are locked together in a flat enveloping space, we will do a simple calculation that is appropriate to the cylinders or spheres of our corresponding coordinate systems. In both cases the coordinates are an example of what are called Gaussian normal coordinates, where the first coordinate measures the arclength along its coordinate lines which are orthogonal to its coordinate surfaces. Such a metric has the form

$$\begin{aligned} g &= g_{ij} dx^i dx^j = \epsilon dx^1 \otimes dx^1 + g_{ab} dx^a \otimes dx^b, \\ g^{-1} &= \epsilon \frac{\partial}{\partial x^1} \otimes \frac{\partial}{\partial x^1} + g^{ab} \frac{\partial}{\partial x^a} \otimes \frac{\partial}{\partial x^b}, \\ \det g &= \det {}^{(2)}g, \end{aligned}$$

where the index range $a, b, c = 2, 3$ will be understood in this section and the 2-metric on the coordinate surfaces $x^1 = x_0^1$ is

$${}^{(2)}g = g_{ab} dx^a \otimes dx^b, \quad {}^{(2)}g^{-1} = g^{ab} \frac{\partial}{\partial x^a} \otimes \frac{\partial}{\partial x^b},$$

while $\epsilon = \pm 1$ allows us to extend this to the Lorentz case.

In fact the first coordinate (a radial coordinate in both cases) is an arclength coordinate whose coordinate lines are straight lines, namely geodesics, while the remaining two coordinates describe a family of nested surfaces orthogonal to those coordinate lines which are the coordinate surfaces for the first coordinate. These surfaces are said to be geodesically parallel since they are separated by the same arclength along each such coordinate line. The same construction is important in 3-dimensional toy cosmological models where the first coordinate is a timelike one, and its coordinate surfaces are spacelike slices of the spacetime, as well as in flat Minkowski spacetime in pseudospherical or pseudocylindrical coordinates. Instead of a coordinate singularity at the zero value of the radial coordinate (the origin in spherical coordinates, the z -axis in cylindrical coordinates) where the coordinate surface shrinks to lesser dimension (a point and a line respectively), one has a “big bang” singularity in a cosmological model where the universe begins. In the spacetime context, such a coordinate system is called a synchronous reference system since it relies on a global synchronization of time, which elapses uniformly along the time coordinate lines.

Define the extrinsic curvature component matrix by

$$K_{ab} = -\frac{1}{2} g_{ab,1}, \quad K^a{}_b = g^{ac} K_{cb}.$$

Then evaluating the component formula for the coordinate components of the connection

$$\Gamma_{ijk} = \frac{1}{2} (g_{ij,k} - g_{jk,i} + g_{ki,j}), \quad \Gamma^i{}_{jk} = g^{i\ell} \Gamma_{\ell jk}$$

under the coordinate conditions $g_{11} = \epsilon = \pm 1, g_{1a} = 0$ leads to some simplification. Since $g_{11,i} = 0 = g_{1a,0}$ there are no lowered connection components Γ_{ijk} which are nonzero with

more than one index 1, and since $g_{1a} = 0$, that remains true of the connection components themselves.

$$\Gamma_{1ab} = -\frac{1}{2}g_{ab,1} = K_{ab}, \quad \Gamma_{a1b} = \Gamma_{ab1} = -K_{ab}, \quad \Gamma_{abc} = {}^{(2)}\Gamma_{abc} \equiv \frac{1}{2}(g_{ab,c} - g_{bc,a} + g_{ca,b}),$$

where ${}^{(2)}\Gamma_{bc}^a$ indicates the same formula in terms of ${}^{(2)}g_{ab}$ (the intrinsic connection components). Raising the index leads to the following nonzero components of the connection, which vanish if the index 1 is repeated

$$\begin{aligned} \Gamma^1_{ab} &= -\frac{\epsilon}{2}g_{ab,1} \equiv \epsilon K_{ab}, \quad \Gamma^a_{1b} = \Gamma^a_{b1} = -K^a_b, \\ \Gamma^a_{bc} &= {}^{(2)}\Gamma^a_{bc} \equiv \frac{1}{2}g^{ad}(g_{db,c} - g_{bc,d} + g_{cd,b}). \end{aligned}$$

The unit normal vector field $n = \partial/\partial x^1$ has components $n^i = \delta^i_1$ so that $n^i n_i = g_{ij} n^i n^j = g_{11} = \epsilon$. Along the surface directions, its covariant derivative is

$$\nabla_a n^i = n^i_{,a} + \Gamma^i_{aj} n^j = \Gamma^i_{aj} \delta^j_1 = \Gamma^i_{a1} = \delta^i_b \Gamma^b_{a1} = -\delta^i_b K^b_a$$

and dotting this with a surface vector field

$$e_b \cdot \nabla_{e_a} n = -K_{ba}.$$

This confirms our identification of the extrinsic curvature tensor above.

Exercise 10.4.1.

extrinsic curvature as a connection component

Show that the above formulas imply the other extrinsic curvature relation

$$n \cdot \nabla_{e_a} e_b = K_{ab}.$$

■

Next consider the formula for the coordinate components of the vanishing Riemann curvature tensor $R^i_{jkl} = 0$ on the flat 3-space. These can only have zero, one, or two index values equal to 1 since those with three indices all the same are easily seen to be 0 because of the coordinate conditions. Those with one or two index values equal to 1 will involve the extrinsic curvature. Those with none correspond to the intrinsic components of the Riemann curvature tensor plus extrinsic curvature terms from the contracted index sum in the product terms, namely those curvature components along the x^1 coordinate surface.

Exercise 10.4.2.

decomposition of curvature on a family of surfaces

a) Verify the above relations for the decomposition of the components of the connection.

b) Derive the formula

$$R^a_{bcd} = {}^{(2)}R^a_{bcd} - \epsilon K^a_c K_{bd} + \epsilon K^a_d K_{bc}$$

or equivalently

$$R^{ab}_{cd} = {}^{(2)}R^{ab}_{cd} - \epsilon K^a_c K^b_d + \epsilon K^a_d K^b_c$$

starting from the formula

$$R^a_{bcd} = \Gamma^a_{db,c} - \Gamma^a_{cb,d} + \Gamma^a_{ci} \Gamma^i_{db} - \Gamma^a_{di} \Gamma^i_{cb}.$$

Note the sum over $i = 1, 2, 3$ which must be separated into its $i = 1$ term and the sum over $e = 2, 3$.

c) Verify the formula

$$R^1_{a1b} = -\epsilon (K_{ab,1} + K_{ac} K^c_b).$$

d) Verify the formula

$$R^1_{bcd} = -2\epsilon \nabla_{[c} K_{d]b}$$

e) Show that

$$G^1_a = R^1_a = R^{i1}_{ia} = -R^{ib}_{ba} = -R^1_{bca} g^{bc} = -\nabla_b (K^b_a - K^c_c \delta^b_a).$$

f) Show that

$$\begin{aligned} G^1_1 &= R^1_1 - \frac{1}{2}(R^1_1 + R^b_b) = \frac{1}{2}(R^1_1 - R^b_b) = \frac{1}{2}(R^{1b}_{1b} - (R^{1b}_{1b} + R^{ba}_{ba})) \\ &= -\frac{1}{2}({}^{(2)}R + \epsilon(K^b_a K^a_b - K^a_a K^b_b)). \end{aligned}$$

These formulas hold for 4-dimensional spacetime as well where $a, b, c = 2, 3, 4$. ■

Now since the total curvature is zero in the case of both Euclidean \mathbb{R}^3 and 3-dimensional Minkowski spacetime, we have the relation

$${}^{(2)}R^a_{bcd} = \epsilon(K^a_c K_{bd} - K^a_d K_{bc}).$$

This has only one independent component

$${}^{(2)}R^{23}_{23} = \epsilon(K^2_2 K^3_3 - K^2_3 K^3_2) = \epsilon \det \underline{K} = \epsilon k_1 k_2,$$

which is the Gaussian curvature, where k_1, k_2 are the two principal curvatures (eigenvalues of the sign reversed extrinsic curvature). The surface Riemann tensor is therefore representable as $R^{ab}_{cd} = \epsilon k_1 k_2 \delta^{ab}_{cd}$, i.e., is proportional to the generalized Kronecker delta on the surface with two indices.

Exercise 10.4.3.**spaces of constant curvature**

A space of constant curvature has the simplest possible curvature tensor

$$R^{ab}{}_{cd} = R_0 \delta^{ab}{}_{cd},$$

where R_0 is a constant. The simplest extrinsic curvature is completely isotropic

$$K^a{}_b = \frac{\mu}{a^2} \delta^a{}_b, \quad \mu = \pm 1.$$

Show that the vanishing of the total curvature tensor then implies

$$R_0 = \frac{\epsilon\mu}{a^2}.$$

■

Exercise 10.4.4.**spherical coordinates with a signature change**

In Section 10.2 the connection 1-form matrix in the orthonormalized spherical coordinate frame was split into its extrinsic and intrinsic curvature parts. Suppose we change the signature of the spherical coordinate metric on \mathbb{R}^3 by making the radial coordinate r timelike and renaming it to the time coordinate t

$$ds^2 = -dt^2 + t^2(d\theta^2 + \sin^2\theta d\phi^2).$$

a) Trace how the sign change of $g_{11} = g_{tt}$ affects the orthonormal frame connection 1-form matrix compared to the original metric.

b) If we let $g^{(2)} = g_{ab}dx^a \otimes dx^b$ for $a, b = 2, 3$ be the “spatial metric” on the space slices of this spacetime which represent constant time surfaces, then show that the nonzero components of the extrinsic curvature tensor can be expressed as

$$K_{ab} = -\frac{1}{2} \frac{\partial g_{ab}}{\partial t} = -\frac{1}{t} g_{ab}.$$

This negative multiple of the space metric describes the expansion of the spheres with time, i.e., the intrinsic distance between points fixed in the sphere grows with time. A positive extrinsic curvature of this type would describe a contraction of the spheres with time, or a collapsing cosmology. This could be accomplished simply by changing the sign of the time coordinate $t \rightarrow -t$ so that the big bang singularity is to the future of all negative times, representing a big crunch.

c) Use a computer algebra system to evaluate the Einstein tensor for this no longer flat geometry and show that it takes the form

$$G_{ij} = t^{-2} n_i n_j, \quad n_i = \delta^t{}_i,$$

where $n = \partial_t$ is the “outward” unit normal to the spheres of constant time t . This may be interpreted as the energy-momentum tensor for a so called dust matter with density proportional to t^{-2} and spacetime velocity n . Notice that this “cosmological model” has a “big bang” “curvature singularity” at $t = 0$ where the components of the curvature tensor and the dust matter density go infinite.

d) How does this discussion change if we replace t^2 in the metric by $\sin^2 t$? This represents a cyclic cosmology which expands from a big bang singularity to a maximum radius and then recollapses to a big crunch singularity. ■

Exercise 10.4.5.

extrinsic curvature of pseudospheres

Consider timelike pseudo-spherical coordinates (τ, χ, ϕ) defined by replacing the polar angle of spherical coordinates on \mathbb{R}^3 down from the vertical axis by the hyperbolic angle down from the time axis $(t, r) = (\tau \cosh \chi, \tau \sinh \chi)$ in 3-dimensional Minkowski spacetime with metric

$$ds^2 = -d\tau^2 + dx^2 + dy^2 = -d\tau^2 + dr^2 + r^2 d\phi^2 = -d\tau^2 + \tau^2(d\chi^2 + \sinh^2 \chi d\phi^2).$$

a) Derive the last equality in this expression for the Evaluate the mixed extrinsic curvature tensor of the pseudo-spherical coordinate surfaces of the timelike pseudo-radial coordinate τ , namely the hyperbolas of revolution $\tau^2 = t^2 - x^2 - y^2$.

b) Use the above relationship to show that these are surfaces of constant negative curvature $R^\theta_{\chi\phi} = -1/\tau^2$.

Notice that in the Lorentzian geometry of 3-dimensional Minkowski spacetime, for a future timelike pseudo-sphere the diagram 10.2 has the reverse behavior—namely, moving the position vector from the origin away from the vertical along the concave up surface, the unit normal tilts further away too (since it is aligned with the position vector) ...

c) Repeat for spacelike pseudo-spherical coordinates $(t, r) = \ell(\sinh \chi, \cosh \chi)$.
(in progress) ■

Exercise 10.4.6.

curvature of hyperbolic paraboloid

Consider the saddle surface $z = x^2 - y^2$ in \mathbb{R}^3 pictured in Fig. 10.1, a hyperbolic paraboloid. This surface may be parametrized by the Cartesian coordinates in the plane $z = 0$ as

$$\vec{r}(u, v) = \langle u, v, u^2 - v^2 \rangle,$$

namely

$$x = u, \quad y = v, \quad z = u^2 - v^2.$$

This surface may also be described in cylindrical coordinates by the condition

$$z = r^2(\cos^2 \phi - \sin^2 \phi) = r^2 \cos 2\phi,$$

and by the alternative parametrization

$$\vec{R}(t, \theta) = \vec{r}(t \cos \theta, t \sin \theta) = \langle t \cos \theta, t \sin \theta, t^2 \cos(2\theta) \rangle$$

by polar coordinates in the plane $z = 0$, which is equivalent to

$$r = t, \quad \phi = \theta, \quad z = t^2 \cos 2\theta.$$

Let's suppress the explicit dependence on θ by writing $R(t)$ to reinterpret this as a 1-parameter family of curves where t parametrizes each curve in the family and θ parametrizes the family of curves. These are all normal cross-sectional curves of the surface at the origin, so are useful for relating the curve curvature to the surface curvature at that point.

a) Evaluate the tangent $\vec{R}'(t)$ and its derivative $\vec{R}''(t)$, the unit tangent $\hat{T}(t) = \vec{R}'(t)/|\vec{R}'(t)|$, the unit normal $\hat{N}(t) = \hat{T}'(t)/|\hat{T}'(t)|$, and the curvature $\kappa(t) = |\vec{R}'(t) \times \vec{R}''(t)|/|\vec{R}'(t)|^3$ of this family of parametrized curves, each of which is a plane curve parabola except for the degenerate cases at $\theta = \pm\pi/4, \pm3\pi/4$ which are horizontal straight lines in the x - y plane. Note that it is most efficient to use the vector quotient rule

$$\left(\frac{\vec{F}(t)}{g(t)} \right)' = \frac{g(t)\vec{F}'(t) - \vec{F}(t)g'(t)}{g(t)^2}$$

to evaluate $\hat{T}'(t)$, combining fractions and keeping common factors of the components of the resulting vector factored out. This factoring makes it simple to normalize $\hat{T}'(t)$.

Evaluate the unit binormal $\hat{B}(t) = \vec{R}'(t) \times \vec{R}''(t)/|\vec{R}'(t) \times \vec{R}''(t)|$. It should be horizontal and independent of t since these are plane curves lying in vertical planes. Note that it is easier to evaluate $\hat{N}(t) = \hat{B}(t) \times \hat{T}(t)$ than by differentiation of $\hat{T}(t)$ if you are quotient rule challenged.

Confirm that $\hat{N}(0) = \langle 0, 0, 1 \rangle$. Evaluate $\kappa(0)$ as a function of θ . What are its values for $\theta = 0, \pi/4, \arctan(2), \pi/2$? Figure 10.1 shows the last three of these curves.

b) Calculate the two tangent vectors with respect to u and v , namely the partial derivatives of the position vector

$$\vec{r}_1(u, v) = \frac{\partial \vec{r}}{\partial u}(u, v), \quad \vec{r}_2(u, v) = \frac{\partial \vec{r}}{\partial v}(u, v).$$

Evaluate the matrix \underline{g} of their inner products $g_{ij} = \vec{r}_i \cdot \vec{r}_j$ and use it to express the metric of the surface

$$^{(2)}g = g_{11}du \otimes du + g_{12}du \otimes dv + g_{21}dv \otimes du + g_{22}dv \otimes dv.$$

Show that $g_{12} = 0$ so that these two vectors are orthogonal only when $uv = 0$, which is where the surface intersects the (vertical) x - z or y - z planes. Evaluate the inverse matrix \underline{g}^{-1} and $\det \underline{g}$. Note that one could also evaluate the metric just by re-expressing $dx \otimes dx + dy \otimes dy + dz \otimes dz$ in terms of (u, v) and simplifying.

c) Evaluate the surface normal vector $\vec{N}(u, v) = \vec{r}_1(u, v) \times \vec{r}_2(u, v) = |\vec{N}(u, v)|\hat{n}(u, v)$ and its length $|\vec{N}(u, v)|$ and direction $\hat{n}(u, v)$. The length should just be $(\det \underline{g})^{1/2}$, which is the single independent component of the unit area 2-form $^{(2)}\eta = \eta_{uv}du \wedge dv$. Evaluate numerically

the double integral of the length for the parameter range $0 \leq u \leq 1, 0 \leq v \leq 1$. Later we will see that this is the surface area of this surface over the unit rectangle in the first quadrant.

d) Evaluate the (u, v) coordinate components of the covariant extrinsic curvature tensor

$$-K_{ij} = \vec{r}_i \cdot \partial_j \hat{n}, \quad i, j = 1, 2,$$

and then raise its first index to get the mixed extrinsic curvature tensor

$$K^i{}_j = g^{ik} K_{kj}, \quad \underline{K} = \underline{g}^{-1}(K_{ij}).$$

Show that this mixed tensor is diagonal only when $uv = 0$. Evaluate it at the origin where it is diagonal and evaluate it on the unit tangents $\hat{T} \cdot K(\hat{T}) = K_{ij} T^i T^j$ to the four vertical cross-sectional curves of part a). Confirm that you get the curvatures of those curves computed in part a), plus a sign indicating whether the center of curvature of the curve is on the same side or opposite side of the tangent line.

e) Although formulas are easily written down for the eigenvalues and eigenvectors of the mixed extrinsic curvature matrix $\underline{K} = (g^{ik} K_{kj})$, $i, j = 1, 2$ using the quadratic formula or a computer algebra system, they are not very much fun. Consider the point $(u, v) = (1, 1)$ and evaluate this matrix and find its eigenvalues and eigenvectors. The principal curvatures are the sign-reversed eigenvalues.

The 2-component eigenvectors $\underline{U} = \langle U^1, U^2 \rangle$ of this 2×2 matrix \underline{K} are components with respect to the basis \vec{r}_1, \vec{r}_2 of the tangent plane to this surface, so to visualize them in space, one must reform those linear combinations of the basis vectors to get the corresponding 3-vectors $U^1 \vec{r}_1 + U^2 \vec{r}_2$. The two independent eigenvectors should be orthogonal. Divide them by their lengths to get orthonormal vectors. Check that their dot product is zero.

Then plot the two normalized eigenvectors in a 3d plot with the surface with equal units displayed on the axes. Rotate the plot until you convince yourself that the picture looks right, i.e., the normal plane cross-sections through those two directions are orthogonal, and the two vectors lie in the tangent plane to the surface. What are the two radii of curvature along these directions? Do their eigenvalue signs look right, i.e., is the negative curvature (positive eigenvalue of \underline{K}) cross-section concave down with respect to the normal, and the positive curvature one (negative eigenvalue of \underline{K}) concave up?

f) Use a computer algebra system to evaluate the single independent component of the Riemann curvature tensor for the surface ${}^{(2)}R^{uv}_{uv}$. Show that it equals the Gaussian curvature. ■

Exercise 10.4.7.

curvature of surfaces of revolution

For surfaces of revolution for which the metric takes the form (allowing for a Lorentzian signature)

$$g = \epsilon dr \otimes dr + R(r)^2 d\theta \otimes d\theta, \quad \epsilon = \pm 1,$$

i.e., with the nonzero metric components $g_{rr} = \epsilon$, $g_{\theta\theta} = R(r)^2$, we showed that the only nonvanishing components of the connection are

$$\Gamma^r_{\theta\theta} = -\epsilon R'(r)R(r), \quad \Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} = \frac{R'(r)}{R(r)}.$$

a) By direct evaluation of the formula for the curvature tensor (summing over $k = 1, 2$)

$$\begin{aligned} R^r_{\theta r\theta} &= \Gamma^r_{\theta\theta,r} - \Gamma^r_{r\theta,\theta} + \Gamma^r_{rk}\Gamma^k_{\theta\theta} - \Gamma^r_{\theta k}\Gamma^k_{r\theta} = \Gamma^r_{\theta\theta,r} - \Gamma^r_{\theta\theta}\Gamma^\theta_{r\theta} \\ &= (-\epsilon R'(r)R(r))' - (-\epsilon R(r)'R(r))(R'(r)/R(r)) = -\epsilon R''(r)R(r), \end{aligned}$$

show that its single independent component can be expressed in the form

$$R^{r\theta}_{\theta\theta} = R^r_{\theta r\theta}g^{\theta\theta} = -\epsilon \frac{R''(r)}{R(r)}.$$

In other words, check these steps.

b) For the flat plane with $R(r) = r$, confirm the vanishing of the curvature tensor.

c) For a sphere of radius r_0 in Euclidean space ($\epsilon = 1$) with $(r, \theta) = (r_0\theta, \phi)$ to fit into this naming scheme, one has $R(r) = r_0 \sin(r/r_0)$. Verify that $R^{r\theta}_{\theta\theta} = 1/r_0^2$.

d) For the torus, translating the variable names appropriately, evaluate this curvature component to find

$$R^{r\phi}_{r\phi} = \frac{\cos \chi}{b(a + b \cos \chi)}.$$

Notice that for the special case $a = 0$ of a sphere of radius b , this reduces to $1/b^2$ as it should. Notice that $\cos \xi > 0$ describes the outer half of the torus, which has positive curvature, while $\cos \xi < 0$ describes the inner half of the torus, which has negative curvature.

e) For a general surface of revolution in Euclidean space $\epsilon = 1$, use Exercise 8.5.3 to evaluate the connection 1-form matrix for the associated orthonormal frame $e_{\hat{r}}, e_{\hat{\theta}}$ starting from the coordinate connection 1-form matrix and rescaling transformation

$$\underline{\omega} = \begin{pmatrix} 0 & -R'(r)R(r) d\theta \\ \frac{R'(r)}{R(r)} d\theta & \frac{R'(r)}{R(r)} d\theta \end{pmatrix}, \quad \underline{A} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{R(r)} \end{pmatrix}.$$

The result should be

$$\underline{\hat{\omega}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} R'(r) d\theta.$$

Show that this agrees with the result for polar coordinates in the plane, and the results for the torus and 2-sphere obtained previously.

f) For a timelike pseudo-sphere of radius τ_0 in 3-dimensional Minkowski spacetime ($\epsilon = -1$) with $r = \tau_0 \chi$ to fit into this naming scheme, one has $R(r) = \tau_0 \sinh(r/\tau_0)$. Verify that $R^{\chi\theta}_{\chi\theta} = -1/\tau_0^2$, confirming the result of a previous exercise. ■

Exercise 10.4.8.**torus extrinsic curvature**

Now that we have evaluated the intrinsic curvature for the torus, let's attack the extrinsic curvature in the same approach taken with the hyperbolic paraboloid saddle surface.

a) Let $(u, v) = (\chi, \phi) = (u^1, u^2)$ rename our variables so that the parametrized torus becomes

$$\vec{r}(u, v) = \langle (a + b \cos u) \cos v, (a + b \cos u) \sin v, b \sin u \rangle.$$

Evaluate the two tangents to the grid $\vec{r}_a(u, v) = \partial \vec{r}(u, v) / \partial u^a$ for $a, b = 1, 2$ and the outward normal $\vec{N}(u, v) = \vec{r}_1(u, v) \times \vec{r}_2(u, v)$ and its magnitude $|\vec{N}(u, v)|$, which should equal the square root of the metric matrix determinant. Let $\hat{n}(u, v) = \vec{N}(u, v) / |\vec{N}(u, v)|$ be the outward unit normal.

b) Use a computer algebra system to evaluate and simplify first the sign-reversed covariant extrinsic curvature tensor $-K_{ab} = \hat{n} \cdot \partial \vec{r}_b / \partial u^a$ following the saddle surface example, which will be rotationally symmetric and hence not depend on the azimuthal angle v , and then matrix multiply this matrix $(-K_{ab})$ on the left by the inverse metric matrix to obtain the sign-reversed mixed extrinsic curvature matrix

$$(-K^a_b) = (-g^{ac} K_{cb}).$$

c) Evaluate its determinant, the Gaussian curvature, confirming that it equals the single independent $R^{\chi\phi}_{\xi\phi}$ of the intrinsic curvature tensor evaluated above, modulo the variable renaming. Find its eigenvectors and confirm that one is tangent to the meridians and the other is orthogonal to those meridians along the parallels. Thus the parameter grid is formed from the two families of lines of curvature. ■

Exercise 10.4.9.**extrinsic curvature of the cavatappo 2.0 surface**

Repeat the previous problem for the orthogonally tilted cavatappo surface, where the unit normal one obtains is the outward normal. Show that the sign-reversed extrinsic coordinate matrix is

$$(-K^i_j) = \begin{pmatrix} \frac{\cos(\eta) \cos(v)}{N} & 0 \\ -\frac{c}{bN} & \frac{1}{b} \end{pmatrix}$$

where

$$N = \frac{a^2 + c^2 + ab \cos(v)}{\sqrt{a^2 + c^2}} = |\det(g)|^{1/2} / b$$

and its eigenvalues are

$$k_1 = \frac{1}{b}, \quad k_2 = \frac{\cos(\eta) \cos(v)}{N}$$

along the respective eigenvectors with components $\langle 0, 1 \rangle$ and $\langle 1, \sin(\eta) \rangle$, namely along the circular meridians and their orthogonal trajectories. Thus the orthogonal grid is formed from the lines of curvature of this surface.

**Exercise 10.4.10.****Lorentz cavatappo surface**

Repeat the previous problem for the orthogonally tilted Lorentz cavatappo surface, but reverse the sign of the unit normal from this recipe in order to get an outward normal. Show that the coordinate matrix $(-K^i_j)$ is

$$(-K^i_j) = \begin{pmatrix} \frac{\sinh(\beta) \cos(v)}{N} & 0 \\ -\frac{N}{c} & \frac{1}{b} \end{pmatrix}$$

where

$$N = \frac{c^2 - a^2 - ab \cos(v)}{\sqrt{c^2 - a^2}} = |\det(g)|^{1/2}/b$$

and its eigenvalues are

$$k_1 = -\frac{1}{b}, \quad k_2 = \frac{\sinh(\beta) \cos(v)}{N}$$

along the respective eigenvectors with components $\langle 0, 1 \rangle$ and $\langle 1, -\sinh(\beta) \rangle$, namely along the circular meridians and their orthogonal trajectories. Thus the orthogonal grid is formed from the lines of curvature of this surface.



10.5 Tube/tubular surfaces

The torus and helical tube surfaces we have studied above turn out to fall into a mathematical category called swept surfaces (as I learned only after having studied their geodesics and having evaluated their extrinsic curvatures). One takes a profile curve and uses it to sweep it along a trajectory curve to generate a surface, allowing the profile curve to be rotated and rescaled as one sweeps it along the trajectory curve. The helical tube surface used to model cavatappo pasta takes a circle in the normal plane to a helix and rotates it about the normal line to the vertical plane to obtain the profile curve which is then swept along the helix, always remaining vertical. The general family of such helical surfaces allows the circular profile curve to be rotated by any angle about the normal line, but leaving the circular profile in the normal plane itself results in the orthogonally tilted cavatappo surface, an example of what is called a tube surface or tubular surface. This is exactly how one thinks of tubes in everyday life—having a fixed circular cross-section orthogonal to the direction of the tube at each point. The torus is a special case of these helical tube surfaces in which the helical trajectory degenerates to a circle. The circular profile curves of a tubular surface are called meridians, as in the case of surfaces of revolution.

By choosing a circle about the origin of fixed radius in the normal plane to any trajectory curve, one gets a general tubular surface. As long as the radius of the circle is small compared to the radius of curvature of the trajectory curve, one has a surface which appears to just be a “thickening” of the original trajectory curve, like a ring torus with a small transverse radius. When the radius of the profile circle is comparable to the radius of curvature of the trajectory curve, the resulting surface can start to have self-intersections and have more complicated structure (like the horn and spindle tori), due to the intersections of the normal lines along the evolute of the trajectory curve, which is at a distance equal to the radius of curvature of the trajectory curve along its normal line.

The extrinsic curvature of a tubular surface is a fascinating example to study since one can understand the geometry of the principal curvatures based on the individual curvatures of the trajectory curve and the meridians. One principal curvature is equal to the reciprocal of the fixed radius of those meridians (its obvious curvature as a circle!) and the other principle curvature is along the family of orthogonal curves, a new family of parallels, and this curvature interpolating between the reciprocals of the radius of two concentric circles in the osculating plane at the two ends of the diameter along the normal line of the profile circle. Amazingly this interpolation can be seen geometrically, including the location of the center of the corresponding osculating circle associated with the second principle curvature.

Suppose we have a parametrized space curve $\vec{r}(t)$, with first and second derivatives $\vec{r}'(t)$ and $\vec{r}''(t)$, from which we can calculate the curvature $\kappa(t)$ and radius of curvature $\mathcal{R}(t) = 1/\kappa(t)$, and the unit tangent $\hat{T}(t)$, unit normal $\hat{N}(t)$ and unit binormal $\hat{B}(t)$. The line through $\vec{r}(t)$ along the unit normal is the normal line, while the plane spanned by the two normal vectors is the normal plane. The plane of $\vec{r}'(t)$ and $\vec{r}''(t)$ or equivalently by $\hat{T}(t)$ and $\hat{N}(t)$ is the osculating plane, or velocity-acceleration plane in the physics language. The osculating circle has its center $\vec{C}(t)$ a distance equal to $\mathcal{R}(t)$ along $\hat{N}(t)$ on the normal line within the osculating

plane

$$\vec{C}(t) = \vec{r}(t) + \mathcal{R}(t) \hat{N}(t).$$

It only makes sense to use a computer algebra system to evaluate these more complicated combination formulas!

Now take a circle of fixed radius b about the origin of the normal plane (namely $\vec{r}(t)$), where the angle Θ is the usual polar angle in the plane of the ordered pair of vectors $\{\hat{N}(t), \hat{T}(t)\}$

$$\vec{r}(t, \Theta) = \vec{r}(t) + b \left(\cos \Theta \vec{N}(t) + \sin \Theta \vec{T}(t) \right).$$

This tubular surface has a natural grid associated with the (t, Θ) parametrization/coordinatization. The Θ circles are the meridians, while the t lines are a natural family of parallels, as in the surface of revolution terminology, but not orthogonal to the meridians in general.

For any given t , the strip from $t - \delta t$ to $t + \delta t$ for small increment δt is approximately like a strip from a torus with central radius \mathcal{R} and cross-sectional radius b (assumed to be less than \mathcal{R} for all t), with inner radius $\mathcal{R}(t) - b$ and outer radius $\mathcal{R}(t) + b$. See Fig. 10.5. Indeed the osculating circles for $\Theta = 0, \pi$ where the normal line intersects the surface lie in the osculating plane at $\vec{r}(t)$, and are concentric circles about the same center $\vec{C}(t)$ of those same radii, circles which coincide with the inner and outer equators for the corresponding approximating torus strip. Because these circles are orthogonal to the cross-sectional circles, they must be lines of curvature so the principal curvatures are just the reciprocals of their radii, namely $|k_1| = 1/b$ and $|k_2| = 1/(\mathcal{R}(t) - b)$ at $\Theta = 0$ and $|k_1| = 1/b$ and $|k_2| = 1/(\mathcal{R}(t) + b)$ at $\Theta = \pi$. If we agree to determine the sign of the principal curvature as $+$ if its osculating circle is on the same side as that normal and $-$ if on the opposite side, since the circles are on opposite sides of the tangent plane for $\Theta = 0$, we must assign $k_2 = -1/(\mathcal{R}(t) - b)$ at $\Theta = 0$ but $k_2 = 1/(\mathcal{R}(t) + b)$ at $\Theta = \pi$, with $k_1 = 1/b$. For other values of Θ , we need to interpolate the curvature from its minimum negative value through zero to its maximum positive value.

So let's start calculating. Define the two tangents

$$\begin{aligned} \vec{r}_1(t, \Theta) = \frac{\partial \vec{r}}{\partial t}(t, \Theta) &= |\vec{r}'(t)| \left[(1 - b \kappa(t) \cos \Theta) \hat{T}(t) + b \tau(t) \hat{E}_2(t, \Theta) \right], \\ \vec{r}_2(t, \Theta) = \frac{\partial \vec{r}}{\partial \Theta}(t, \Theta) &= \vec{E}_2(t, \Theta) = b \hat{E}_2(t, \Theta) = b \left(\cos \Theta \hat{B}(t) - \sin \Theta \hat{N}(t) \right). \end{aligned}$$

Notice that

$$\vec{E}_1(t, \Theta) = |\vec{r}'(t)| (1 - b \kappa(t) \cos \Theta) \hat{T}(t) = \frac{\partial \vec{r}}{\partial t}(t, \Theta) - \tau(t) \vec{E}_2(t, \Theta)$$

is orthogonal to $\vec{E}_2(t, \Theta)$, so that $\{\vec{E}_1(t, \Theta), \vec{E}_2(t, \Theta)\}$ is a natural orthogonal frame with associated orthonormal frame $\{\hat{E}_1(t, \Theta), \hat{E}_2(t, \Theta)\}$.

Interpreting (t, Θ) as coordinates on the surface we have

$$\partial_t \leftrightarrow \vec{r}_1(t, \Theta), \quad \partial_\Theta \leftrightarrow \vec{r}_2(t, \Theta),$$

and the combination

$$|\vec{r}'(t)|^{-1} \partial_t - \tau(t) \partial_\Theta \leftrightarrow \vec{E}_1(t, \Theta)$$

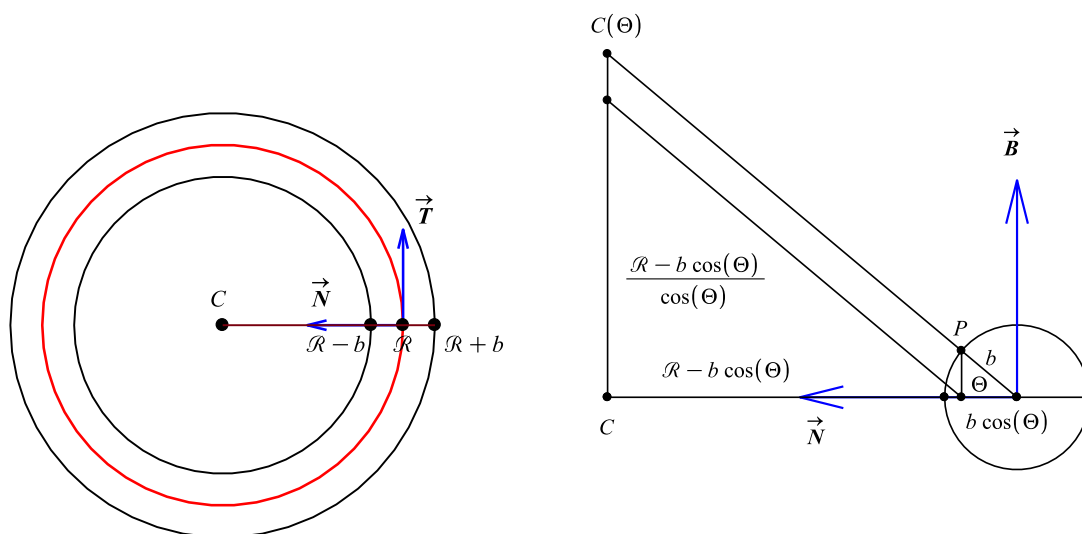


Figure 10.5: **Left:** The osculating plane of a circle of radius \mathcal{R} (shown in red) with its unit tangent and unit normal vectors at a given point: the osculating circle coincides with the original circle. Shown also is the normal line to the center C with the osculating circles of the two parallels of a tubeplot of meridian radius b centered on that circle for $\Theta = 0, \pi$, with respective radii $\mathcal{R} - b$ and $\mathcal{R} + b$. This is exactly the situation for a point on a general tube plot for the concentric osculating circles in the osculating plane of the central trajectory curve for the two parallels which intersect that osculating plane. **Right:** The corresponding normal plane to a point on the central trajectory curve of a tubular surface of radius b , slightly magnified compared to the left view. The center $C(\Theta)$ of the osculating circle to a point P on a meridian for $-\pi/2 < \Theta < \pi/2$ lie along the line through the center C of the osculating circle of the central trajectory which is parallel to the binormal. The distance $\mathcal{R} - b \cos \Theta$ of P from this axis is then stretched by division by $\cos \Theta$ as the corresponding line segment rotates from the horizontal up to the angle Θ keeping its endpoint on that axis. These osculating circles are on the opposite side of the tangent plane at P from the meridian circle, making $k_1 < 0$. For $\pi/2 < \Theta < \pi$ the diagram moves below the normal line, and the osculating circle is on the same side of the tangent plane, making $k_1 > 0$. For angles near the poles $\Theta = \pm\pi$, the centers $C(\Theta)$ move out to infinity.

so that

$$(1 - b\kappa(t)\cos\Theta)^{-1}(|\vec{r}'(t)|^{-1}\partial_t - \tau(t)\partial_\Theta) \leftrightarrow \hat{E}_1(t, \Theta) = \hat{T}(t),$$

while

$$b^{-1}\partial_\Theta \leftrightarrow \hat{E}_2(t, \Theta).$$

Next the matrix of inner products of these frame vectors, which determines the metric line element, letting $dS = |\vec{r}'(t)|dt$ be the differential of arclength along the trajectory curve

$$\begin{aligned} ds^2 &= [(1 - b\kappa(t)\cos\Theta)^2 + b^2\tau(t, \Theta)^2] dS^2 + 2b^2\tau(t)^2 dS d\Theta + b^2 d\Theta^2 \\ &= (1 - b\kappa(t)\cos\Theta)^2 dS^2 + b^2 (d\Theta + \tau(t)dS)^2. \end{aligned}$$

The orthogonal representation corresponds to the above-mentioned orthogonal frame, whose orthonormal dual frame is therefore

$$\hat{W}^1(t, \Theta) = (1 - b\kappa(t)\cos\Theta) dS, \quad \hat{W}^2(t, \Theta) = b(d\Theta + \tau(t)dS).$$

An outward normal vector to the tubular surface is

$$\vec{N}(t, \Theta) = \vec{r}_1(t, \Theta) \times \vec{r}_2(t, \Theta) = b(1 - b\kappa(t)\cos\Theta)\hat{N}(t, \Theta)$$

where the unit outward normal is

$$\hat{N}(t, \Theta) = (\cos\Theta\hat{N}(t) + \sin\Theta\hat{B}(t)),$$

assuming that $b/\mathcal{R} < 1$.

Next we calculate the derivatives of the unit normal using this correspondence and the Frenet-Serret relations of Appendix C

$$\begin{aligned} |\vec{r}'(t)|^{-1}\partial_t\vec{N}(t, \Theta) &= |\vec{r}'(t)|^{-1}(\cos\Theta\hat{N}'(t) + \sin\Theta\hat{B}'(t)) \\ &= \cos\Theta(-\kappa(t)\hat{T}(t) + \tau(t)\hat{B}(t)) + \sin\Theta(-\tau(t)\hat{N}(t)) \\ &= -\kappa(t)\cos\Theta\hat{T}(t) + \tau(t)\hat{E}_2(t, \Theta) \end{aligned}$$

and

$$\partial_\Theta\vec{N}(t, \Theta) = -\sin\Theta\vec{N}(t) + \cos\Theta\hat{B}(t) = \hat{E}_2(t, \Theta).$$

Thus we have

$$\begin{aligned} \nabla_{\hat{T}(t)}\vec{N}(t, \Theta) &= (1 - b\kappa(t)\cos\Theta)^{-1}(|\vec{r}'(t)|^{-1}\partial_t - \tau(t)\partial_\Theta)\vec{N}(t, \Theta) \\ &= \frac{-\kappa(t)\cos\Theta}{1 - b\kappa(t)\cos\Theta}\hat{T}(t) = k_1(t, \Theta)\hat{T}(t), \end{aligned}$$

and

$$\nabla_{\hat{E}_2(t, \Theta)}\vec{N}(t, \Theta) = \frac{1}{b}\partial_\Theta\vec{N}(t, \Theta) = \frac{1}{b}\hat{E}_2(t, \Theta) = k_2(t, \Theta)\hat{E}_2(t, \Theta).$$

This last principal curvature is no surprise since it is just the curvature of the meridian circle of radius $b = \mathcal{R}_2 = 1/k_2$. The first principal curvature can be rewritten in terms of the corresponding radius of curvature of the new parallel curves along \hat{T} in the surface

$$\mathcal{R}_1(t, \Theta) = |1/k_1(t, \Theta)| = \frac{\mathcal{R} - b \cos \Theta}{|\cos \Theta|},$$

which has a nice geometrical interpretation illustrated in Fig. 10.5. As one increases Θ from 0 to π , the curvature smoothly interpolates between $k_1(t, 0) < 0$ and $k_1(t, \pi) > 0$ through the value 0 at the binormal line.

Exercise 10.5.1.

tilted cavatappo surface curvature

a) Show that the result in Exercise 10.4.9 for the curvature k_2 of the orthogonal parallels follows from the tubular surface formula using the values of the curvature and torsion of the Euclidean helix given in Appendix C.

b) Repeat for the result of Exercise 10.4.10 applied to the Lorentz orthogonally tilted timelike cavatappo surface using the values for the curvature and torsion of the Lorentz timelike helix to see how the signs in the formula change in the Lorentz case.

■

10.6 Surface geodesics studied from the outside

Our approach to geodesics in surfaces embedded in flat spaces or spacetimes with a globally flat connection has been to simply calculate the surface metric through a parametrization of that surface which provides us with local coordinates on that surface in terms of which we can then work entirely intrinsically, leaving behind the larger space in which it lives. By adapting coordinates on the whole space to a family of surfaces which contain the surface of interest, we have paved the way to see how that surface fits into the larger space directly in terms of the components of the connection of the adapted coordinates, separating them into extrinsic curvature terms and intrinsic connection terms. Geodesics arise from the geodesic equations on the whole space by ignoring the derivatives along the normal direction which are necessary to keep the curve in the surface, while keeping the remaining derivative terms which ensure that the tangent to the surface geodesic does not further change direction within the surface. In the adapted coordinates the extra coordinates held fixed then give us the path in the coordinate system of the larger space which allows the geodesics to be visualized as paths in the larger space. Even with the surface parametrization approach, at the end we map the surface geodesics in terms of the surface coordinates back into the Cartesian coordinates of the enveloping flat space so we can visualize then in the context of the surface embedding in that larger space.

However, suppose we do not have either an adapted set of coordinates to describe the surface or even a parametrization of the surface which enables us to work in an intrinsic coordinate system. We can still handle the surface geodesics as constrained motion within the larger space using only the original Cartesian coordinates. The motion in space is accelerated since that normal component of the covariant derivative remains nonzero to constrain the motion to the surface. As long as one can compute the normal to the surface, one can evaluate that normal component of the acceleration and study the equations of motion that correspond to it. One only needs to describe the surface as the level surface of a function on the space to get its normal via the gradient and then we are in business. This opens up the geodesic game to a much wider class of surfaces than we are able to handle explicitly through a parametrization.

The ideas are relatively simple. Suppose we have an implicit surface $f(x^1, x^2, x^3) = 0$ in \mathbb{R}^3 , with its dot product geometry. We can extend this to the Lorentz case later. Then the gradient vector $\vec{n} = \nabla f$ provides us with a normal to the surface. Suppose we have an affinely parametrized geodesic $c(\lambda)$ of the surface geometry, given by $c^i = x^i(c(\lambda))$ which we can sloppily denote by $x^i(\lambda)$ in common practice, or as $\vec{r}(\lambda)$ in vector form. Then we can also write compactly $f(\vec{r}(\lambda)) = 0$, so that by the chain rule, the tangent to the curve is orthogonal to the gradient

$$\vec{r}'(\lambda) \cdot \vec{n}(\vec{r}(\lambda)) = 0.$$

The condition that this curve be a geodesic of the surface is that its “acceleration vector”

$$\vec{a}(\lambda) \equiv \vec{r}''(\lambda) = \nabla_{\vec{r}'(\lambda)} \vec{r}'(\lambda) = \frac{D^2 \vec{r}(\lambda)}{d\lambda^2}$$

be orthogonal to the surface so that its tangent vector does not rotate left or right within the surface with respect to its forward direction, but only rotates in the plane of \vec{n} and itself in order to remain within the surface. This means that the surface normal is aligned with the unit

normal to the curve, but we don't know in advance whether they are parallel or antiparallel, so we can simply choose the surface unit normal as the curve's normal direction and allow the curve's curvature to have either sign instead of determining the direction of the unit normal to be in the direction of $\hat{T}'(\lambda)$ where $\hat{T}(\lambda) = \vec{r}'(\lambda)/|\vec{r}'(\lambda)|$ is the unit tangent. For a surface geodesic the length of this tangent is a constant, interpreted as the speed if we view tracing out the geodesic as motion along a path in space where λ is identified with the time

$$v(\lambda)^2 \equiv \vec{T}(\lambda) \cdot \vec{T}(\lambda) = \vec{r}'(\lambda) \cdot \vec{r}'(\lambda) = \left(\frac{ds(\lambda)}{d\lambda} \right)^2.$$

We can then easily generate a right handed but unnormalized orthogonal frame for any curve in the surface aligned with both the surface normal and the curve's tangent by defining

$$\begin{aligned} \vec{T}(\lambda) &= \vec{r}'(\lambda), \\ \vec{\mathcal{B}}(\lambda) &= \vec{T}(\lambda) \times \vec{n}(\vec{r}(\lambda)) = \vec{n}(\vec{r}(\lambda)) \times \vec{r}'(\lambda), \\ \vec{\mathcal{N}}(\lambda) &= \vec{\mathcal{B}}(\lambda) \times \vec{T}(\lambda) = -\vec{T}(\lambda) \times (\vec{T}(\lambda) \times \vec{n}(\vec{r}(\lambda))) = (\vec{T}(\lambda) \cdot \vec{T}(\lambda)) \vec{n}(\vec{r}(\lambda)), \end{aligned}$$

where the final equality follows from the double cross product identity and the orthogonality of \vec{T} and \vec{n} . The second two vectors in this frame have self-dot products

$$\begin{aligned} \vec{\mathcal{B}}(\lambda) \cdot \vec{\mathcal{B}}(\lambda) &= \vec{\mathcal{B}}(\lambda) \cdot \vec{\mathcal{B}}(\lambda) \vec{T}(\lambda) \cdot \vec{T}(\lambda), \\ \vec{\mathcal{N}}(\lambda) \cdot \vec{\mathcal{N}}(\lambda) &= \vec{n}(\vec{r}(\lambda)) \cdot \vec{n}(\vec{r}(\lambda)) \vec{T}(\lambda) \cdot \vec{T}(\lambda), \end{aligned}$$

using the quadruple scalar product identity in the first case. Using these we can easily normalize the orthogonal frame when needed to obtain $(\hat{T}, \hat{\mathcal{N}}, \hat{\mathcal{B}})$. In general the curve's own Frenet-Serret frame will be related to this one by an additional rotation in the curve's normal plane

$$(\hat{T} \quad \hat{\mathcal{N}} \quad \hat{\mathcal{B}}) = (\hat{T} \quad \hat{\mathcal{N}} \quad \hat{\mathcal{B}}) \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \chi & -\sin \chi \\ 0 & \sin \chi & \cos \chi \end{pmatrix} = (\hat{T} \quad \hat{\mathcal{N}} \quad \hat{\mathcal{B}}) e^{\chi L_1}.$$

This frame is defined for any curve in the surface but of course it is aligned with the Frenet-Serret frame only for surface geodesics. The vector $\vec{\mathcal{B}}$ places the role of the curve's normal within the surface modulo the sign of its direction. The tangential component of the acceleration within the surface is simply its component along $\vec{\mathcal{B}}$, which defines the surface curvature when the arclength derivative is used. Using the chain rule to convert the lambda derivatives to arclength derivatives we have

$$\kappa_s(\lambda) \equiv a_{\hat{\mathcal{B}}}(\lambda) = \hat{\mathcal{B}}(\lambda) \cdot \frac{D^2 \vec{r}}{ds^2}(\lambda) = \hat{\mathcal{B}}(\lambda) \cdot \frac{\vec{r}''(\lambda)}{\vec{r}'(\lambda) \cdot \vec{r}'(\lambda)}.$$

This must be zero for a surface geodesic along which $\hat{\mathcal{B}}$ is parallel transported in the intrinsic surface geometry to join \hat{T} in forming the adapted orthonormal 2-frame which satisfies the Frenet-Serret degenerate frame relations (both the unit tangent and this unit normal are parallel transported along the curve).

The extrinsic curvature of the surface reversed in sign is the shape operator, which is the projection of the following tensor on the surface into the tangent plane

$$\mathcal{S}_{ij} = \nabla_i \hat{n}_j.$$

Since we are only going to evaluate this along the surface, we do not need to distinguish the this tensor from its projection. When evaluated on surface tangent vectors the normalization factor (inverse length) of \vec{n} passes through the derivative

$$\mathcal{S}(\vec{r}'(\lambda), \vec{r}'(\lambda)) = \left(\vec{r}'(\lambda) \cdot \frac{(\nabla \vec{n})(\vec{r}(\lambda))}{|\vec{n}(\vec{r}(\lambda))|} \right) \cdot \vec{r}'(\lambda)$$

Its value on the unit tangent produces the component of the acceleration normal to the surface

$$-a_{\hat{n}} = \mathcal{S}(\vec{r}', \vec{r}') = -\hat{n} \cdot \left(\nabla_{\vec{T}} \vec{T} \right).$$

Thus the acceleration vector is

$$\begin{aligned} \vec{T}' = \vec{r}'' &= a_{\hat{n}} \hat{n} + a_{\hat{\mathcal{B}}} \hat{\mathcal{B}} = -\mathcal{S}(\vec{r}', \vec{r}') \hat{n} + \kappa_s (\vec{r}' \cdot \vec{r}') \hat{\mathcal{B}} \\ &= -\frac{(\nabla \vec{n})(\vec{r}', \vec{r}') \vec{\mathcal{N}}}{\vec{\mathcal{B}} \cdot \vec{\mathcal{B}}} + \kappa_s (\vec{r}' \cdot \vec{r}') \hat{\mathcal{B}} \\ &\equiv \kappa (\vec{r}' \cdot \vec{r}') \hat{N}, \end{aligned}$$

where the $\vec{\mathcal{B}}$ self-dot product factor normalizes both for arclength and the unit surface normal and the last line defines the space curvature and unit normal of the curve. The tip of the tangent vector (whose length is constant) can only rotate, and the extrinsic curvature determines its rotation rate in its osculating plane, while the surface curvature determines its rotation rate in its normal plane. Comparing the previous relation with the angle χ between the actual unit normal and the surface normal already introduced above

$$\hat{N} = \cos \chi \hat{\mathcal{N}} + \sin \chi \hat{\mathcal{B}},$$

then the previous equations show that

$$\tan \chi = -\frac{\kappa_s}{\mathcal{S}(\hat{T}, \hat{T})} = \frac{\kappa_s}{K(\hat{T}, \hat{T})},$$

which vanishes for a surface geodesic where the two normals are aligned. For a general surface curve, the relative rotation of the surface normal relative to the curve normal is determined by the ratio of the intrinsic curvature of the curve to its extrinsic curvature $-\mathcal{S}(\hat{T}, \hat{T}) = K(\hat{T}, \hat{T})$.

This partially quantifies the remarks at the beginning of this chapter regarding the two contributions to the rotation of the surface normal along a curve within the surface: the normal rotates around the unit tangent direction due to the tilting of the surface underneath the curve while rotating in the plane of the normal and unit tangent in order to stay orthogonal to the surface as the unit tangent itself rotates in that plane. The horizontal straight line in Fig. 10.1

a) is also a surface geodesic since it does not have to even bend to stay in the saddle surface, so the surface normal only rotates about its fixed direction as the surface tilts sideways underneath the line as one moves along it. The parabola in Fig. 10.1 c) on the contrary is a geodesic by reflection symmetry about the vertical cross-sectional plane in which it lies, and along it the surface normal only rotates to stay orthogonal to the surface while the surface tangent line in the curve's normal plane remains horizontal.

The affinely parametrized surface geodesics have zero intrinsic curvature, so they are determined by the equations

$$\vec{r}'' = -\frac{(\nabla \vec{n})(\vec{r}', \vec{r}') \vec{N}}{\vec{B} \cdot \vec{B}},$$

for which the “energy”

$$\frac{1}{2} \vec{r}' \cdot \vec{r}' = \frac{1}{2} \left(\frac{ds}{d\lambda} \right)^2 = \mathcal{E}$$

is constant, and equals 1/2 for an arclength parametrization. One can easily numerically integrate this second order system of differential equations directly to describe the geodesics as paths in space, even for a parametrized surface. One can specify the initial data for the tangent or velocity vector in terms of a single angle with respect to a suitably chosen initial direction in the surface at the initial location, and one is guaranteed a unique solution.

Exercise 10.6.1.

shape operator insensitive to length of normal

- Evaluate $\nabla_i \hat{n}_j$ in terms of $\vec{n} = |\vec{n}| \hat{n}$.
- Show that for \vec{X}, \vec{Y} orthogonal to \vec{n} , $\mathcal{S}(\vec{X}, \vec{Y}) = \nabla \vec{n}(\vec{X}, \vec{Y}) / |\vec{n}|$.

■

Exercise 10.6.2.

geodesics on the sphere and ellipsoid

- Whenever one does numerical solution of differential equations, it is important to have a test case where the analytic solution is known to check its accuracy. Apply this approach to the unit sphere $x^2 + y^2 + z^2 = 1$ and examine the solutions of the initial value problem at the point $(1, 0, 0)$ along the positive x -axis on the equator using arclength parametrization initial data (unit velocity at an angle β with respect to the vertical). The geodesics should return to the initial point at $\lambda = 2\pi$.

- Extend this to an ellipsoid

$$\frac{x^2}{9} + \frac{y^2}{4} + \frac{z^2}{1} = 1$$

already studied in Exercise 8.7.12 using an obvious parametrization. Test it for initial data at one of the intersection points with the coordinate axes with initial tangent lying in a coordinate plane, since by reflection symmetry about the coordinate planes, the coordinate plane intersection curves (ellipses) are geodesics. Then try a geodesic at 45 degrees to the coordinate

axes at such initial data points. Remember that the final value of λ (starting at $\lambda = 0$ is the actual arclength of the geodesic. ■

Exercise 10.6.3.

geodesics on the approximate gyroid

For a real challenge consider one cube of the triply periodic approximate gyroid surface

$$f(x, y, z) = \cos(x) \sin(y) + \cos(y) \sin(z) + \cos(z) \sin(x) = 0.$$

$$-\pi \leq x, y, z \leq \pi.$$

This is not only invariant under translations of each of the coordinates by multiples of 2π but also under permutations of the coordinates.

a) Plot this with a computer algebra system and rotate around to see its profile in the y - z plane. It appears to have an approximate reflection symmetry about the line $z = y - \pi/2$, which is a plane clearly intersecting one of the tubes in this lattice in a loop. Find a parametrization of this loop and find the lowest point P on this loop to use as an initial data point for geodesics that initially move along this loop.

b) Use implicit differentiation to determine the tangent plane at the origin, through which the surface passes. Because of the permutation symmetry the direction of its normal there is not surprising. Repeat for the point P where the tangent plane is parallel to the x -axis, and given this, its normal direction is also not surprising. Find a right handed orthonormal frame \vec{EO}_i at the origin and \vec{EP}_i at P containing the upward unit normal as the third vector and a horizontal unit vector as the first vector, which fixes both frames uniquely. Plot each frame on the surface with a square piece of the tangent plane, for example: $\{t_1 \vec{EO}_1 + t_2 \vec{EO}_2 \mid -1 \leq t_1, t_2 \leq 1\}$. Express unit tangent vectors in each of these tangent planes in terms of a polar angle β measured from the first frame vector. This provides two useful locations for initial data for arclength parametrized geodesic equations.

c) Notice that the origin seems to be a monkey saddle, that is with three equally spaced ridges which rise above the tangent plane, and 3 equally spaced valleys that fall below the tangent plane. Evaluate the extrinsic curvature at the origin and diagonalize it to find the principal curvature directions and plot unit vectors along them in your previous plot. What is the angle between your original axes and these new orthogonal directions?

d) The tube point P appears to be an ordinary simple saddle point with pair of aligned ridges rising up above the tangent plane. Repeat c) for this case.

e) Explore the geodesic spray from each of these two initial data points. Try to find a geodesic that leaves the second initial data point and loops around back into the general vicinity of its point of departure by trial and error starting from initial data along the horizontal tangent vector to the loop.

f) Use a computer algebra system to study the geodesics on this surface using these two initial data points. Play. See if you can find anything interesting. ■

Exercise 10.6.4.**rotation of the surface normal compared to the Frenet-Serret frame**

a) Use a computer algebra system to reproduce Fig. 10.1 for the surface $z = x^2 - y^2$ and the image of the three curves on that surface: a) $y = x$, b) $x = 0$ and c) $y = 2x$.

b) Include the Frenet-Serret frame for each curve in the plot, and animate it and the surface normal along the curve for each of the 3 curves to show the relative rotation of the surface normal in the normal plane to the curve.

■

Exercise 10.6.5.**relative rotation of Frenet-Serret frame and surface adapted frame**

Recall from Appendix C the Frenet-Serret relations in Cartesian coordinates, adjusted for the lambda derivative $d/d\lambda = (ds/d\lambda) d/ds = (\vec{r}' \cdot \vec{r}')^{1/2} d/ds$, where the speed factor $v = (\vec{r}' \cdot \vec{r}')^{1/2}$ is constant for an affinely parametrized curve

$$\begin{pmatrix} \hat{T}' & \hat{N}' & \hat{B}' \end{pmatrix} = (\vec{r}' \cdot \vec{r}')^{1/2} \begin{pmatrix} \hat{T} & \hat{N} & \hat{B} \end{pmatrix} \begin{pmatrix} 0 & -\kappa & 0 \\ \kappa & 0 & -\tau \\ 0 & \tau & 0 \end{pmatrix}.$$

a) Show that the corresponding relations for the surface adapted frame along the curve are

$$\begin{aligned} \begin{pmatrix} \hat{T}' & \hat{N}' & \hat{B}' \end{pmatrix} &= (\vec{r}' \cdot \vec{r}')^{1/2} \begin{pmatrix} \hat{T} & \hat{N} & \hat{B} \end{pmatrix} \begin{pmatrix} 0 & -K(\hat{T}, \hat{T}) & -\kappa_s \\ K(\hat{T}, \hat{T}) & 0 & -\tau_s \\ \kappa_s & \tau_s & 0 \end{pmatrix} \\ &= (\vec{r}' \cdot \vec{r}')^{1/2} \begin{pmatrix} \hat{T} & \hat{N} & \hat{B} \end{pmatrix} e^{\chi L_1} \begin{pmatrix} 0 & -\kappa & 0 \\ \kappa & 0 & d\chi/ds - \tau \\ 0 & \tau - d\chi/ds & 0 \end{pmatrix} e^{-\chi L_1}. \end{aligned}$$

For a surface geodesic where $\chi = 0$ these reduce to the Frenet-Serret relations provided that the curve and surface unit normals are taken to have the same direction rather than opposite directions, in which case τ_s reduces to the torsion of the curve.

b) The upper line in the previous equation must have this form by antisymmetry and the existing definitions, leaving only the additional angular velocity τ_s to be determined. The lower line is simply the derivative of the rotational relationship between the two frames involving the relative angle χ , utilizing the Frenet-Serret relations for the derivatives of the Frenet-Serret vectors and the product rule. Can one find a formula for $d\chi/ds$ which is not too ugly? Or instead for τ_s ? The relationship between them is implied by the equality of the two right hand sides of this last multiple equation.

■

Exercise 10.6.6.**Gaussian curvature of implicitly defined surface**

Suppose our surface is defined implicitly by $f(x, y, z) = 0$. If we want to color the surface using the value of the Gaussian curvature in order to visualize how this curvature varies on the surface, we need a formula for it. There are two ways one can proceed.

a) We can assume we are at a point where the normal direction is not horizontal, which would imply a vertical tangent plane. For interesting surfaces this would only occur at isolated points in any case, which we can handle with limits from “regular points” where the normal direction is not horizontal. One can then use (x, y) as implicit coordinates on the surface understanding z to be some unknown function $z(x, y)$ of (x, y) locally around some regular point of the surface and then evaluate all the necessary derivatives of $\vec{r}(u, v) = \langle u, v, z(u, v) \rangle$ using the chain rule

$$\frac{\partial f(u, v, z(u, v))}{\partial u} = \frac{\partial f}{\partial x}(u, v, z(u, v)) + \frac{\partial f}{\partial z}(u, v, z(u, v)) \frac{\partial z}{\partial u}(u, v)$$

but since $f(u, v, z(u, v)) = 0$ we get

$$\frac{\partial z}{\partial u}(u, v) = -\frac{(\partial f / \partial x)(u, v, z(u, v))}{(\partial f / \partial z)(u, v, z(u, v))}.$$

Thus we get a basis of the tangent plane to the surface

$$E_a(\vec{r}(u^1, u^2)) = \frac{\partial \vec{r}(u^1, u^2)}{\partial u^a},$$

and then a right-handed normal

$$\vec{n}(u, v) = E_1(\vec{r}(u^1, u^2)) \times E_2(\vec{r}(u^1, u^2)).$$

Next we can calculate its derivatives as in the case of a parametrized surface and continue until we can evaluate the shape tensor and its eigenvector structure.

b) We can use projection to get a 3 matrix which has one zero eigenvalue. Given the unit normal \hat{n}_i we can raise the index to get the vector (so we need to insert the inverse metric matrix in a Minkowski spacetime), and then project the two indices with the projection tensor

$$P(\hat{n})^i_j = \delta^i_j - \frac{\hat{n}^i \hat{n}_j}{\hat{n} \cdot \hat{n}},$$

which kills normal vectors

$$P(\hat{n})^i_j n^j = 0,$$

so that in matrix form we get the components of the mixed shape tensor

$$\underline{S} = \underline{P} \underline{S} \underline{P}.$$

Its eigenvalues and eigenvectors will include the 0 value corresponding to the unit normal eigenvector, and then the two principal curvatures and the corresponding principal curvature directions. Since this is a 3×3 matrix, we can get exact formulas for these quantities, from which we can get the Gaussian curvature from the product of the two remaining eigenvalues.

c) Try these approaches with a computer algebra system applied to the approximate gyroid surface. Plot the surface in its fundamental cube using color by Gaussian curvature. ■

Chapter 11

Differential forms: integration and differentiation

Any multivariable calculus course at the very minimum treats line integrals of scalars and vector fields, and at least deals with Green's theorem in the plane which allows an interpretation as either Gauss's law or Stoke's theorem, at least one of which might have been encountered in elementary physics. Surface integrals don't make it into our current syllabus at my university, but the idea of the flux of field lines through a surface seems to be implanted along the way again in some exposure to physics for many students. Here we will see that line, flux and volume integrals in ordinary space over curves, surfaces and solid regions bounded by surfaces generalize nicely to the process of integration over p -surfaces in \mathbb{R}^n of all possible dimensions between 1 and n : curves, surfaces, \dots , hypersurfaces, and open regions enclosed by hypersurfaces. All of these activities generalize the natural pairing of covectors and vectors to produce a scalar, and hence do not require any metric to define or evaluate. However, the physical interpretation of the results so obtained depends crucially on reinterpreting the process in terms of a metric.

The integral theorems of multivariable calculus are directly related to differential properties of scalar and vector fields through the differential, gradient, curl and divergence, all of which are unified into the single concept of the exterior derivative of differential forms and their metric relatives. Quite apart from its role in integration theory, this differential operator turns out to be very efficient in representing the various geometric conditions satisfied by the connection and curvature of a metric. This will be free bonus from our development of the mathematics of integration theory for differential forms.

11.1 Changing the variable in a single variable integral

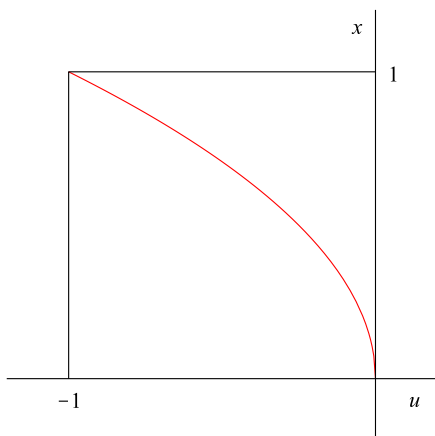


Figure 11.1: The transformation $u = -x^2$, $x = (-u)^{1/2}$ maps increasing u to decreasing x and vice versa.

In changing the variable of a definite integral of a function of a single variable, the process works by simple substitution, replacing the upper and lower limits by the corresponding values of the new variable. For example, suppose $x = g(u)$ is a monotonic change of variable with $g'(u)$ nowhere vanishing on the interval $[x_1, x_2]$ with $x_1 < x_2$ and let $u_1 = g(x_1)$, $u_2 = g(x_2)$. Then either $u_1 < u_2$ or $u_1 > u_2$, but if one wants the new integral to have ordered limits of integration, we need only replace the relative rate of change of the two variables in the transformed integrand by its absolute value

$$\int_{x_1}^{x_2} f(x) dx = \int_{u_1}^{u_2} f(g(u)) \frac{dx}{du} du = \int_{u_{\min}}^{u_{\max}} f(g(u)) \left| \frac{dx}{du} \right| du,$$

where $u_{\min} = \min(u_1, u_2)$ and $u_{\max} = \max(u_1, u_2)$ are the maximum and minimum values of this pair. Thus the interval of integration $[x_1, x_2]$ on the x -axis corresponds to the interval $[u_{\min}, u_{\max}]$ on the u -axis without having to worry which way the former interval is traced out as the new variable is increased from left to right on the latter interval. The role of the derivative factor is to guarantee that the function is integrated against the differential of arclength along the x -axis, which is what dx represents.

Of course when we apply the “ u -substitution method” for single variable integrals we first identify some function of x as the new variable u but then invert the relationship to replace x everywhere in the integral by its expression in terms of u . For example, the change of variable $u = -x^2$ inverts to $x = (-u)^{1/2}$ on the interval $0 \leq x \leq 1$, and leads to a reversal of the direction of the increasing independent variable, tracing out the interval $-1 \leq u \leq 0$. Here is an explicit example

$$\int_0^1 e^{-x^2} x dx = \int_0^{-1} e^u \left(-\frac{1}{2} \right) du = \int_{-1}^0 e^u \left(\frac{1}{2} \right) du = \frac{1}{2}(1 - e^{-1}) > 0.$$

This has to be positive since we are integrating a positive integrand with ordered limits of integration.

11.2 Changing variables in multivariable integrals

These same change of variable ideas for a single variable integral extend to multivariable integrals except that the absolute value $|dx/du|$ for a single variable change is replaced by the absolute value of the Jacobian determinant $|\det(dx^i/du^j)|$ of the change of the set of independent variables in terms of which the integrand function is expressed. This guarantees that the integrand function is integrated against the differential of volume in the space of the orthogonal coordinates x^i . For parametrized p -surfaces within \mathbb{R}^n (curves, surfaces, \dots , hypersurfaces), we can define appropriate integrals simply in terms of multiple integrals on \mathbb{R}^p , thus associating differential measures with these sets (arclength, surface area, \dots , hypersurface area).

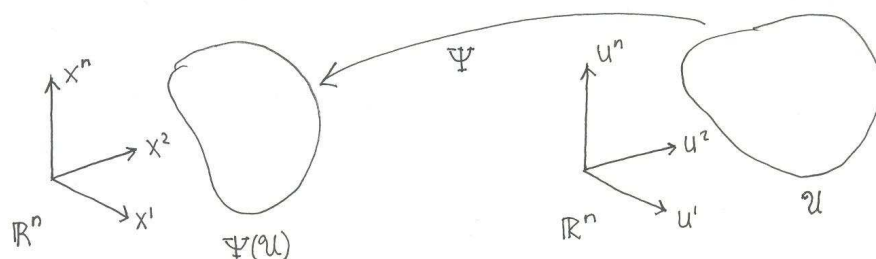


Figure 11.2: A n -dimensional parametrized region in \mathbb{R}^n . Ψ maps from the parameter or coordinate space to the physical space.

As pictured suggestively in Fig. 11.2, we generalize the map $g : \mathbb{R} \rightarrow \mathbb{R}$ associated with the change of variable $x = g(u)$ for the definite integral substitution to an invertible map $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ from a closed region U onto its image $\Psi(U)$, i.e., $x^i = \Psi^i(u)$. More precisely, let $\{u^i\}$ be the standard Cartesian coordinates on the domain \mathbb{R}^n (the coordinate space) and denote the standard Cartesian coordinates on the image space by x^i (the physical space), so that $\Psi^i = x^i \circ \Psi$ are the component functions of this map, expressing the coordinates x^i in terms of the parameters u^i , or in more appropriate language, expressing the old Cartesian coordinates in terms of the new coordinates. The cylindrical and spherical coordinate parametrization maps from the coordinate space into \mathbb{R}^3 are good examples of this. For example, for the former coordinates one has

$$\begin{aligned} x &= \rho \cos \phi & \Psi^1(u) &= u^1 \cos u^2 \\ y &= \rho \sin \phi & \Psi^2(u) &= u^1 \sin u^2 \\ z &= z & \Psi^3(u) &= u^3. \end{aligned} \quad \longleftrightarrow$$

Then the integral of some real valued function f on the image $\Psi(U)$ “in real space” can be expressed as an integral over U (in the parameter or coordinate space) by

$$\int \cdots \int_{\Psi(U)} f \underbrace{dx^1 dx^2 \cdots dx^n}_{dV_x} = \int \cdots \int_U f \circ \Psi \underbrace{\left| \det \left(\frac{\partial \Psi^i}{\partial u^j} \right) \right|}_{\text{correction factor}} \underbrace{du^1 du^2 \cdots du^n}_{dV_u}.$$

Besides re-expressing the function f in terms of the new variables to obtain $f \circ \Psi$, a correction factor takes into account the change in the volume element against which the function is being integrated. This correction factor is the absolute value of the determinant of the Jacobian matrix $\partial x^i / \partial u^j$ of partial derivatives of the image space coordinates with respect to the domain space coordinates and gives the differential volume of an n -parallelepiped whose sides are the n vectors $(\partial x^i / \partial u^j) du^j$ in the tangent space of physical space. The integral of the function f with respect to the differential of volume dV_x in physical space is thus faithfully re-expressed in terms of the new variables.

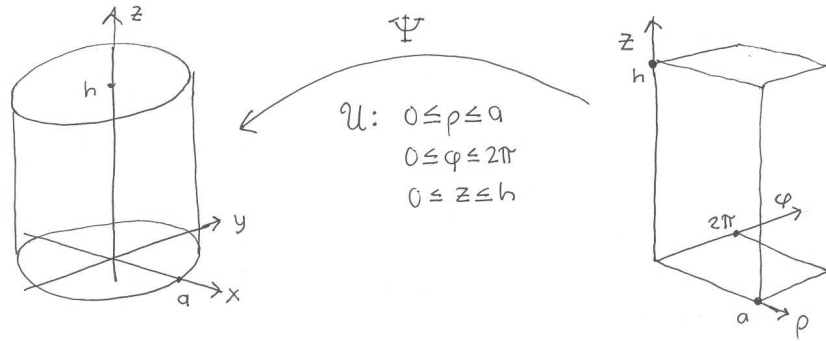


Figure 11.3: Integrating over a parametrized cylinder in \mathbb{R}^3 , corresponding to a rectangular box in the parameter space.

Changing to cylindrical coordinates to evaluate an integral over a cylindrical region is a familiar example of this. For example, consider an integral over the solid cylinder of radius a and height h parametrized by the variable ranges in cylindrical coordinates $U : 0 \leq \rho \leq a, 0 \leq \phi \leq 2\pi, 0 \leq z \leq h$ illustrated in Fig. 11.3

$$\begin{aligned} \iiint_{\Psi(U)} \underbrace{[x^2 - y^2]}_f dx dy dz &= \iiint_U \underbrace{[\rho^2 \cos 2\phi]}_{f \circ \Psi} \underbrace{\rho}_{\left| \det \left(\frac{\partial \Psi^i}{\partial u^j} \right) \right|} d\rho d\phi dz \\ &= \int_0^{2\pi} \int_0^a \int_0^h \rho^3 \cos 2\phi dz d\rho d\phi. \end{aligned}$$

For the present example, the correction factor can be evaluated by appealing to a geometric argument about how to compute the differential volume in cylindrical coordinates without ever mentioning the Jacobian determinant, so this may often be skipped in multivariable calculus.

The absolute value sign in the correction factor guarantees that the integral of a positive function of $\Psi(U)$ results in integrating a positive function on U . This correction factor is essentially all we need to know to describe integration of a p -form field on a parametrized p -surface in an n -dimensional space.

Example 11.2.1. Use double integrals to determine the obvious centroid $(-\frac{5}{2}, 5)$ of the region R enclosed by the parallelogram in the x - y plane shown in Fig. 1.1 whose sides are formed by the lines

$$y = x, \quad y = x + 15, \quad y = -\frac{1}{2}x, \quad y = -\frac{1}{2}x + \frac{15}{2}.$$

To accomplish this we introduce the linear change of variables

$$x = x' - 2y', \quad y = x' + y',$$

with the absolute value Jacobian determinant $|\det(\partial x^i / \partial x^{j'})| = 3$, in terms of which the parallelogram has edge lines described by

$$x' = 0, \quad x' = 5, \quad y' = 0, \quad y' = 5.$$

Thus for any function $f(x, y)$ in the plane

$$\int_R f(x, y) dA = \int_0^5 \int_0^5 f(x' - 2y', x' + y') 3dy' dx'.$$

Thus

$$A_x = \int_0^5 \int_0^5 (x' - 2y') 3dy' dx', \quad A_y = \int_0^5 \int_0^5 (x' + y') 3dy' dx', \quad A = \int_0^5 \int_0^5 (1) 3dy' dx',$$

and

$$\langle \bar{x}, \bar{y} \rangle = \frac{1}{A} \langle A_x, A_y \rangle = \langle -\frac{5}{2}, 5 \rangle.$$

□

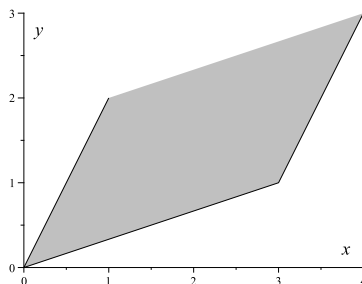


Figure 11.4: A parallelogram in the plane.

Exercise 11.2.1.

integration in the plane over a parallelogram

Repeat this Exercise with the parallelogram in the plane formed by the vectors $\langle 1, 2 \rangle, \langle 3, 1 \rangle$ of Exercise 1.6.3????, where the obvious centroid position vector is the average of these two vectors: $\langle 2, \frac{3}{2} \rangle$.

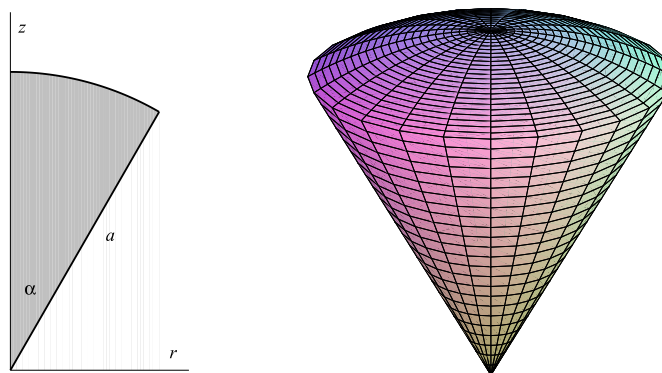


Figure 11.5: A snow cone region of space with vertex at the origin, shown with its r - z half plane cross-section.

■

Exercise 11.2.2.

snow cone centroid integration

Calculate the familiar volume correction factors for cylindrical and spherical coordinates by evaluating the absolute value of the determinant of the Jacobian matrix $(\partial x^i / \partial \bar{x}^j)$ given in Sections 5.7 and 5.8 for these coordinate systems.

Use either coordinate system to evaluate the z coordinate of the centroid of the snow cone topped by a sphere of radius a centered at the origin, with a conical base with vertex at the origin and opening angle α measured from the upwards vertical direction. Since we are integrating z , which is one of the cylindrical coordinates, it suggests using those coordinates, but the bounding surfaces do not correspond to constant values of that coordinate, which nevertheless are described by simple conditions on z as a function of r . On the other hand the two snow cone boundary surfaces each do correspond to constant values of one of the spherical coordinates, but the integrand must be re-expressed as a function of those coordinates (albeit, a simple one). Thus it is not obvious which choice is simplest before iterating and evaluating the integral. Either one is not so difficult to evaluate.

■

11.3 Parametrized p -surfaces and pushing forward the coordinate grid and tangent vectors

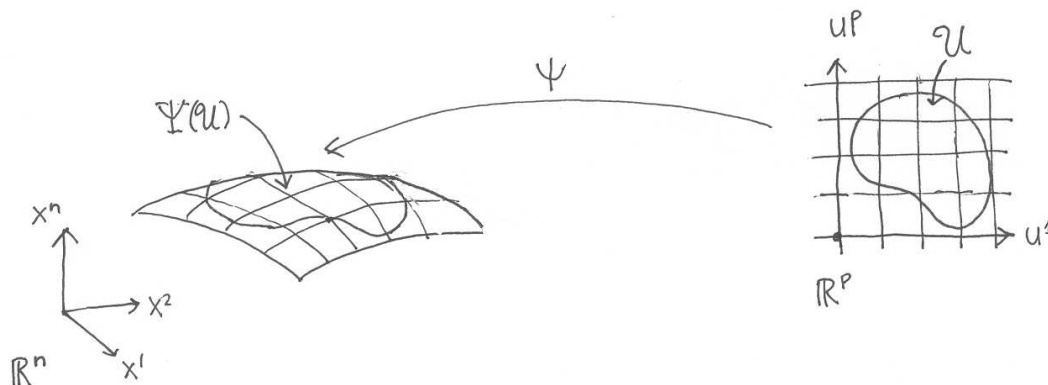


Figure 11.6: A parametrized p -surface in \mathbb{R}^n or some n -dimensional space is a function $\Psi : \mathbb{R}^p \rightarrow \mathbb{R}^n$. The coordinate grid on the parameter space transfers to a grid on the image space.

Suppose we have a 1-1 map Ψ from a closed region U of the “parameter space” \mathbb{R}^p into \mathbb{R}^n or some n -dimensional space with local coordinates $\{x^i\}$. Let $\Psi^i = x^i \circ \Psi$ be those coordinates expressed as functions of the “parameters” (u^1, \dots, u^p) . Let $\alpha, \beta, \dots = 1, 2, \dots, p$ denote the indices for the parameters.

For example, the above parameter map Ψ for cylindrical coordinates represents \mathbb{R}^3 as a parametrized 3-surface or “3-space.” On the other hand fixing the radial coordinate

$$\begin{aligned} x &= a \cos \phi & \Psi^1(u) &= a \cos u^1 \\ y &= a \sin \phi & \Psi^2(u) &= a \sin u^1 \\ z &= z & \Psi^3(u) &= u^2 \end{aligned} \quad \longleftrightarrow$$

leads to a parametrized 2-surface representing an infinite cylinder of radius a , or fixing z as well leads to a parametrized 1-surface or curve

$$\begin{aligned} x &= a \cos \phi & \Psi^1(u) &= a \cos u^1 \\ y &= a \sin \phi & \Psi^2(u) &= a \sin u^1 \\ z &= h & \Psi^3(u) &= h. \end{aligned} \quad \longleftrightarrow$$

In general, fixing any $n - p$ coordinates in the parametrization map associated with local coordinates on \mathbb{R}^n leads to a parametrized p -surface in which the p parameters correspond to p of these coordinates. The two angular coordinates $\{\theta, \phi\}$ of spherical coordinates in \mathbb{R}^3 parametrize the radial coordinate spheres, for example.

For a given parametrized p -surface, fixing all the parameters but one, say u^α , yields a parametrized curve which is the image of the u^α coordinate line on \mathbb{R}^p under the parametrization

map Ψ . Varying the other parameters moves this curve around. One can do this for each of the parameters in turn. This corresponds to mapping to coordinate grid of \mathbb{R}^p onto the image surface in the n -dimensional space (see figure 11.6). The tangent vector at a point $\Psi(u)$ on this p -surface

$$E_\alpha(u) = \frac{\partial \Psi^i(u)}{\partial u^\alpha} \frac{\partial}{\partial x^i} \Big|_{\Psi(u)} = \text{“} \frac{\partial}{\partial u^\alpha} \text{”}$$

is the tangent vector to the parametrized curve corresponding to the u^α coordinate line in \mathbb{R}^p . See figure 11.7. In this way both the coordinate grid and their tangents $\partial/\partial u^i$ are “pushed forward” by the parametrization map Ψ from the parameter space to the image p -surface. The quotation marks around $\partial/\partial u^\alpha$ are meant in the sense that when we differentiate functions on \mathbb{R}^n by these tangent vectors, once we re-express the function in terms of the coordinates u^α on the p -surface, we just take the ordinary partial derivatives by those coordinates.

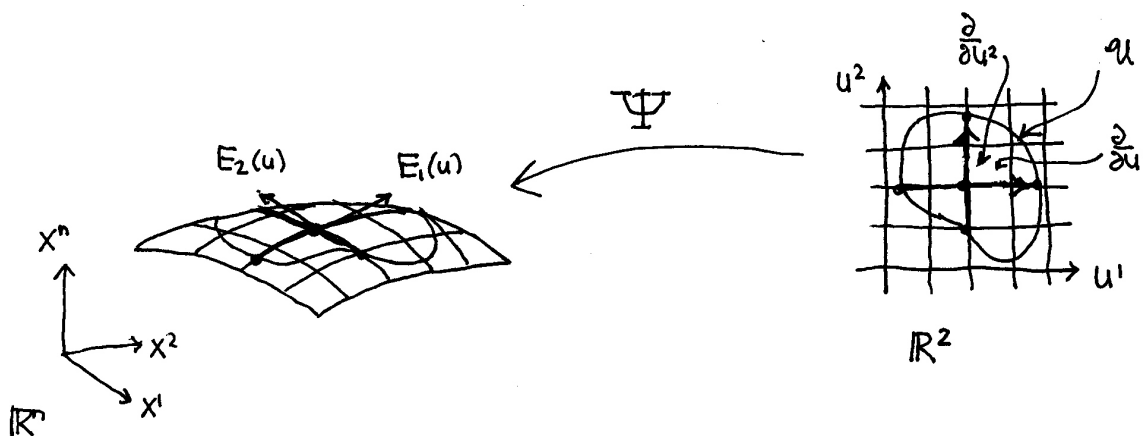


Figure 11.7: A parametrized 2-surface in \mathbb{R}^n is oriented by the parametrization. Both the coordinate grid and their tangents $\partial/\partial u^i$ are pushed forward by the parametrization map Ψ from the parameter space to the image surface.

The tangent p -plane to the parametrized p -surface is spanned by the p -vectors $\{E_\alpha(u)\}$, assuming that they are linearly independent so that its dimension is actually p . This is a condition we must place on the parametrized p -surface. If $\{E_\alpha(u)\}$ are linearly independent, then the p -vector

$$E_1(u) \wedge \cdots \wedge E_p(u) \quad \text{with components} \quad [E_1(u) \wedge \cdots \wedge E_p(u)]^{i_1 \cdots i_p} \\ = p! \, E_1(u)^{[i_1} \cdots E_p(u)^{i_p]}$$

must be nonzero everywhere. This determines the orientation of the tangent p -plane. For $p = 1$ this condition reduces to the requirement that the tangent vector to a parametrized curve not vanish anywhere, which should ring a bell from defining line integrals in multivariable calculus.

Since the map Ψ is 1-1, for each image point on the surface, there is only one point in \mathcal{U} that is mapped onto it by Ψ . Thus if $X(u) = X(u)^\alpha \partial/\partial u^\alpha$ is a vector field on the parameter space, it pushes forward to a surface vector field on the p -surface in a natural way

$$X(u) = X(u)^\alpha \frac{\partial}{\partial u^\alpha} \rightarrow X(u)^\alpha E_\alpha(u).$$

The distinction between $\partial/\partial u^\alpha$ and E_α is that one must first re-express a function on \mathbb{R}^n in terms of the parametrization before differentiating by the former vector field, while the second one is free to operate on the function directly. The results are the same for a given function.

We can clearly extend this to tensor products of such vector fields, and to general contravariant tensor fields. In fact for any parametrized curve $c(t)$ in \mathcal{U} , composition with the map Ψ pushes it forward (towards the image of the map) to a curve $\Psi \circ c(t)$ on the p -surface. By construction, this pushes forward its tangent as a curve in \mathcal{U} to the tangent to the image curve in $\Psi(\mathcal{U})$.

Exercise 11.3.1.

differential of surface area

- a) Evaluate $E_1 \wedge E_2$ for the parametrized cylinders $\rho = \rho_0$ in cylindrical coordinates.
- b) Repeat for the parametrized spheres $r = r_0$ in spherical coordinates.
- c) Repeat for the parametrized elliptical paraboloids of Exercise 1.6.12.

■

11.4 Pulling back functions, covariant tensors and differential forms

Suppose one has any function F on \mathbb{R}^n , which can be thought of as a 0-form. Then its composition $\Psi^*F = F \circ \Psi$ with the parametrization map Ψ naturally defines a function on the parameter space \mathbb{R}^p called the pull back of this function since it pulls the function in the opposite direction of the map itself, which goes instead from \mathbb{R}^p to \mathbb{R}^n . The expression for this function is obtained just by substituting into F the expressions for the coordinates $x^i(u)$ as functions of the parameters. This same operation can be applied to any covariant tensor. For example, the coordinate differentials dx^i pull back to 1-forms Ψ^*dx^i on the parameter space by the chain rule

$$dx^i(u) = \frac{\partial x^i(u)}{\partial u^\alpha} du^\alpha,$$

while any 1-form pulls back similarly

$$f = f_i dx^i \rightarrow \Psi^*f = f_i \circ \Psi \Psi^*dx^i = f_i \circ \Psi \frac{\partial x^i}{\partial u^\alpha} du^\alpha.$$

If one has a metric tensor on \mathbb{R}^n , it too pulls back to a metric on the parameter space with describes the “induced metric” on the p -surface

$$g = g_{ij} dx^i \otimes dx^j \rightarrow \Psi^*g = g_{ij} \circ \Psi \frac{\partial x^i}{\partial u^\alpha} \frac{\partial x^j}{\partial u^\beta} du^\alpha \otimes du^\beta = g_{\alpha\beta} du^\alpha \otimes du^\beta.$$

Its components can be thought of as the inner products of the image vectors

$$g_{\alpha\beta} = \Psi^*g \left(\frac{\partial}{\partial u^\alpha}, \frac{\partial}{\partial u^\beta} \right) = g(E_\alpha, E_\beta).$$

This same “pull back” operation works also for $p = n$, when instead this is interpreted as a change of coordinates. For example, the parameter map associated with cylindrical (or spherical) coordinates on \mathbb{R}^3 pulls back the Euclidean metric on \mathbb{R}^3 to the coordinate expression for the metric on the coordinate space. By interpreting those parameters as coordinate functions on the original physical space, we interpret this as just a way of re-expressing the same tensor on physical space.

Suppose

$$T = \frac{1}{p!} T_{i_1 \dots i_p} dx^{i_1 \dots i_p} = T_{i_1 \dots i_p} dx^{|i_1 \dots i_p|}$$

is a p -form field on our n -dimensional space, usually called a “differential p -form” or if the “degree” p is not made explicit, a “differential form.” Simply expressing this differential form in terms of the parameters on the parametrized p -surface by substituting the parametrized expressions $\Psi^i(u) = x^i \circ \Psi$ for the Cartesian coordinates x^i into it leads to a differential p -form

on the parameter space called the pull back of the p -form T

$$\begin{aligned}
\Psi^*T &\equiv \frac{1}{p!} \underbrace{T_{i_1 \dots i_p} \circ \Psi}_{\text{express components in terms of parameters}} d(x^{i_1} \circ \Psi) \wedge \dots \wedge d(x^{i_p} \circ \Psi) \\
&= \frac{1}{p!} \underbrace{T_{i_1 \dots i_p} \circ \Psi}_{\text{antisym}} \underbrace{\frac{\partial \Psi^{i_1}}{\partial u^{\alpha_1}} \dots \frac{\partial \Psi^{i_p}}{\partial u^{\alpha_p}}}_{\text{only antisym part contributes}} du^{\alpha_1} \wedge \dots \wedge du^{\alpha_p} \\
&= \frac{1}{p!} T_{i_1 \dots i_p} \circ \Psi \frac{\partial \Psi^{i_1}}{\partial u^{\alpha_1}} \dots \frac{\partial \Psi^{i_p}}{\partial u^{\alpha_p}} \epsilon^{\alpha_1 \dots \alpha_p} du^{1 \dots p} \\
&= \frac{1}{p!} T_{i_1 \dots i_p} \circ \Psi \frac{\partial \Psi^{i_1}}{\partial u^1} \dots \frac{\partial \Psi^{i_p}}{\partial u^p} p! du^{1 \dots p} \\
&= \frac{1}{p!} T_{i_1 \dots i_p} \circ \Psi [E_1(u) \wedge \dots \wedge E_p(u)]^{i_1 \dots i_p} du^{1 \dots p}.
\end{aligned}$$

The single independent component of this p -form on \mathbb{R}^p , function is the natural contraction of the p -covector T on the p -vector $E_1(u) \wedge \dots \wedge E_p(u)$ at each point of the parametrized p -surface. It is a function on \mathbb{R}^p .

Suppose we first define the integral of a p -form $fu^{1 \dots p}$ on an open region $\mathcal{U} \subset \mathbb{R}^p$ to be the ordinary multivariable iterated integral of its single independent component function

$$\int_{\mathcal{U}} f du^{1 \dots p} = \int_{\mathcal{U}} f du^1 \wedge \dots \wedge du^p \equiv \int \dots \int_{\mathcal{U}} f du^1 du^2 \dots du^p,$$

where the right hand side is symbolic for any particular choice of the successive partial integrations of some iteration of the integral ordered in any way in terms of the p variables u^α .

We can then define the integral of a p -form on a parametrized p -surface as just the integral on \mathbb{R}^p of the pulled back p -form

$$\int_{\Psi(\mathcal{U})} T \equiv \int_{\mathcal{U}} \Psi^*T = \int \dots \int_{\mathcal{U}} T_{[i_1 \dots i_p]} \circ \Psi(u) [E_1(u) \wedge \dots \wedge E_p(u)]^{i_1 \dots i_p} du^1 du^2 \dots du^p.$$

To summarize, we substitute the parametrization into T , expand it out to get a coefficient function times $du^{1 \dots p}$ and we just integrate that coefficient function on \mathcal{U} in the ordinary sense. The coefficient function is the natural contraction of the p -form T with the p -vector $E_1(u) \wedge \dots \wedge E_p(u)$ of the parametrization, divided by $p!$ to avoid over counting.

In the case $p = n$, we can let Ψ be the identity map so that $u^i = x^i$ and $[E_1(u) \wedge \dots \wedge E_n(u)]^{i_1 \dots i_n} = \delta_{1 \dots n}^{i_1 \dots i_n}$, and this reduces to

$$\int_{\Psi(\mathcal{U})} T = \int \dots \int_{\mathcal{U}} T_{1 \dots n}(u) du^1 du^2 \dots du^n.$$

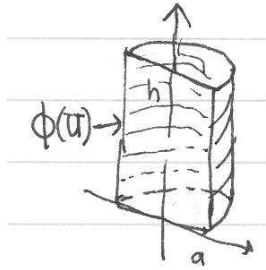


Figure 11.8: A parametrized half cylinder.

Example 11.4.1. Suppose we consider our old friend

$$*X^b = * (ydx + xdy) = ydy \wedge dz + xdz \wedge dx = (ydy - xdx) \wedge dz \equiv T$$

and integrate it over the parametrized half cylinder surface

$$x = a \cos u^1, \quad y = a \sin u^1, \quad z = u^2.$$

$$\begin{aligned} \int_{\Psi(\mathcal{U})} T &= \int_u [(a \sin u^1) d(a \sin u^1) - (a \cos u^1) d(a \cos u^1)] \wedge du^2 \\ &= \int_u [a^2 \sin u^1 \cos u^1 + a^2 \sin u^1 \cos u^1] du^1 \wedge du^2 \\ &= \int \int_u a^2 \sin 2u^1 du^1 du^2 = \int_0^h \int_0^\pi a^2 \sin 2u^1 du^1 du^2 \\ &= a^2 \underbrace{\int_0^h du^2}_h \underbrace{\int_0^\pi \sin 2u^1 du^1}_{-\frac{1}{2} \cos 2u^1 \big|_0^\pi} = -\frac{a^2 h}{2} [\cos 2\pi - \cos 0] = 0. \end{aligned}$$

In this context the 2-form

$$\Psi^*T = a^2 \sin 2u^1 du^1 \wedge du^2$$

is equivalent to re-expressing T in cylindrical coordinates

$$\begin{aligned} T &= (ydy - xdx) \wedge dz = [\rho \sin \phi d(\rho \sin \phi) - \rho \cos \phi d(\rho \cos \phi)] \wedge dz \\ &= [2\rho^2 \sin \phi \cos \phi d\phi + \rho(\sin^2 \phi - \cos^2 \phi) d\rho] \wedge dz \\ &= \rho^2 \sin 2\phi d\phi \wedge dz - \rho \cos 2\phi d\rho \wedge dz \end{aligned}$$

and setting ρ to a and $d\rho$ to 0

$$\underbrace{T_{\rho=a, d\rho=0}}_{\text{"restriction of } T \text{ to surface } \rho = a"} = a^2 \sin 2\phi d\phi \wedge dz \quad \longleftrightarrow \quad \Psi^*T = a^2 \sin 2u^1 du^1 \wedge du^2.$$

□

Exercise 11.4.1.**surface integral on a sphere**

Repeat the above discussion for $T = (ydy - xdx) \wedge dz$ on the part of a sphere $r = a$ in the first octant using the spherical coordinates $\{\theta, \phi\}$ as the parameters $\{u^1, u^2\}$.

$$x = a \sin u^1 \cos u^2, \quad y = a \sin u^1 \sin u^2, \quad z = a \cos u^1.$$

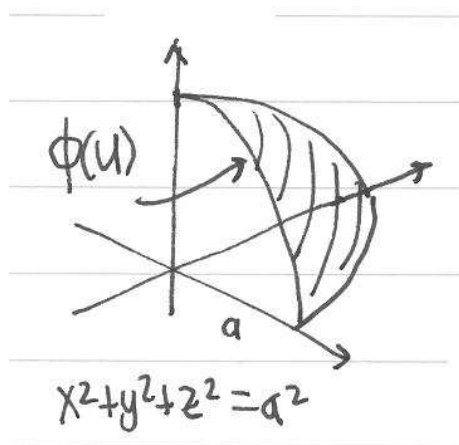


Figure 11.9: A parametrized sector of a sphere. [add axis labels, $\Psi(u)$, etc]

First evaluate $\int_{\phi(u)} T$. Then evaluate T in spherical coordinates and restrict it to the sphere by setting $r = a$, $dr = 0$. Compare with your result for Ψ^*T .

■

Exercise 11.4.2.**contraction of 2-form with 2-vector**

In the cylindrical problem, the tangent vectors to the parameter grid are

$$\begin{aligned} (E_1^i(u)) &= \langle -a \sin u^1, a \cos u^1, 0 \rangle, & (E_2^i(u)) &= \langle 0, 0, 1 \rangle, \\ E_1(u) &= -a \sin u^1 \frac{\partial}{\partial x} \Big|_{\Psi(u)} + a \cos u^1 \frac{\partial}{\partial y} \Big|_{\Psi(u)}, & E_2(u) &= \frac{\partial}{\partial z} \Big|_{\Psi(u)}. \end{aligned}$$

so

$$E_1(u) \wedge E_2(u) = a \sin u^1 \frac{\partial}{\partial z} \wedge \frac{\partial}{\partial x} \Big|_{\Psi(u)} + a \cos u^1 \frac{\partial}{\partial y} \wedge \frac{\partial}{\partial z} \Big|_{\Psi(u)}.$$

The natural contraction with T is then

$$\begin{aligned}
 & \frac{1}{2} T_{ij} \circ \Psi [E_1(u) \wedge E_2(u)]^{ij} \\
 &= \underbrace{T_{23} \circ \Psi}_{a \sin u^1} \underbrace{[E_1(u) \wedge E_2(u)]^{23}}_{a \cos u^1} + \underbrace{T_{31} \circ \Psi}_{a \cos u^1} \underbrace{[E_1(u) \wedge E_2(u)]^{31}}_{a \sin u^1} + \underbrace{T_{12} \circ \Psi}_0 \underbrace{[E_1(u) \wedge E_2(u)]^{12}}_0 \\
 &= 2a^2 \sin u^1 \cos u^1 = a^2 \sin 2u^1.
 \end{aligned}$$

Calculate $E_1(u) \wedge E_2(u)$ for the previous Exercise and its contraction with T in exactly this same way. ■

Exercise 11.4.3.
integration over a triangular surface in space

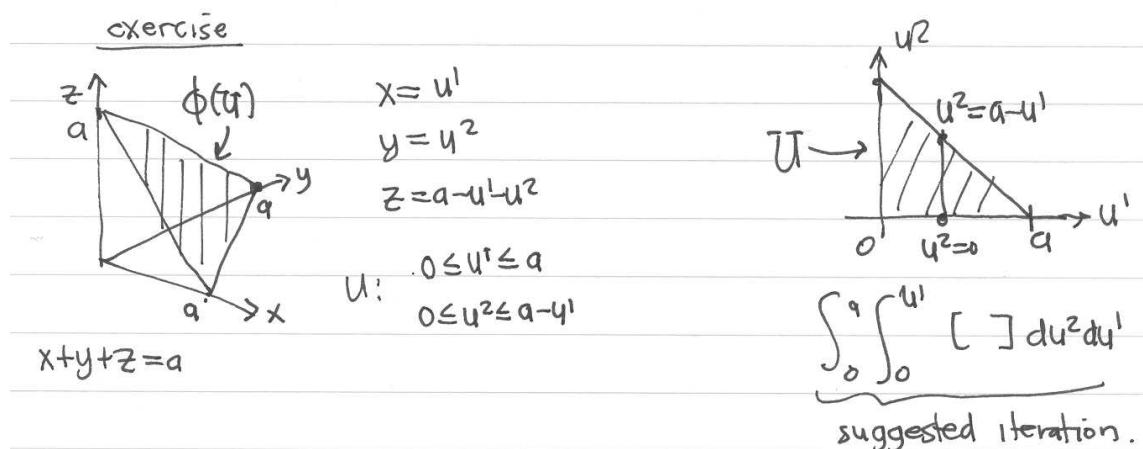


Figure 11.10: A parametrized triangle in \mathbb{R}^3 . [Correct the limit of integration in new figure]

Repeat both Exercises of the preceding page for the simpler plane surface shown in Fig. 11.10. ■

11.5 Changing the parametrization

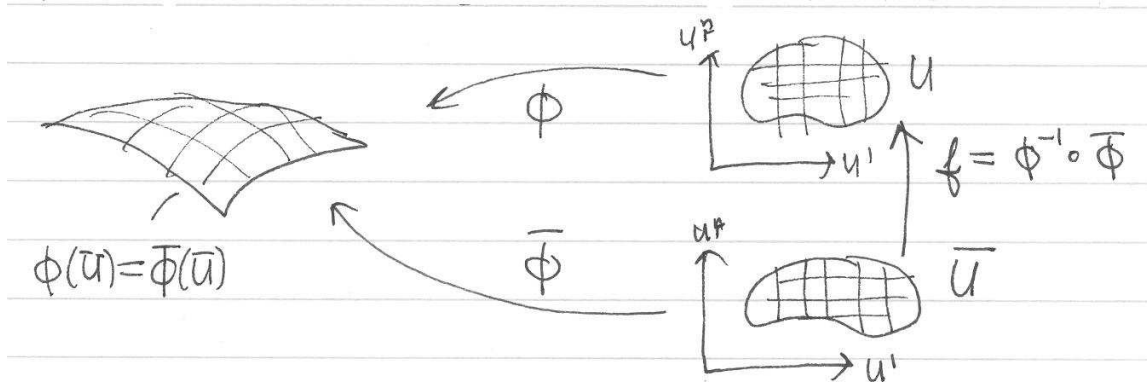


Figure 11.11: Two different parametrizations of the same p -surface. [**change** $(\Phi, \textit{script} f)$ to (Ψ, Φ) **in new figure**]

Suppose we have two different such parametrizations Ψ and $\bar{\Psi}$ of the same p -surface. What do we need in order that the integral of a p -form T on it not depend on the parametrization? Well, $\Phi = \Psi^{-1} \circ \bar{\Psi}$ is a map from \mathbb{R}^p to \mathbb{R}^p which corresponds to the relationship between the parameters which specify the same points on our p -surface.

The values for each parametrized surface as defined above are

$$\begin{aligned} \int_{\Psi(u)} T &= \int \cdots \int_U \frac{1}{p!} T_{i_1 \dots i_p} \circ \Psi(u) [E_1(u) \wedge \cdots \wedge E_p(u)]^{i_1 \dots i_p} du^1 du^2 \cdots du^p \\ \int_{\bar{\Psi}(\bar{u})} T &= \int \cdots \int_{\bar{U}} \frac{1}{p!} T_{i_1 \dots i_p} \circ \bar{\Psi}(\bar{u}) [\bar{E}_1(\bar{u}) \wedge \cdots \wedge \bar{E}_p(\bar{u})]^{i_1 \dots i_p} \underbrace{d\bar{u}^1 d\bar{u}^2 \cdots d\bar{u}^p}_{\text{dummy variables so can use any symbol}} \end{aligned}$$

But the function $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^p$ just represents a change of variable in this ordinary integral $u^\alpha = \Phi^\alpha(\bar{u})$. But by definition $\bar{\Psi} = \Psi \circ \Phi$, i.e., in components $\bar{\Psi}^i(\bar{u}) = \Psi^i(\Phi(\bar{u}))$ so

$$\bar{E}^i_{\alpha}(\bar{u}) = \frac{\partial \bar{\Psi}^i(\bar{u})}{\partial \bar{u}^\alpha} = \frac{\partial \Psi^i(\bar{u})}{\partial u^\beta} \frac{\partial \Phi^\beta}{\partial \bar{u}^\alpha} = E^i_{\beta}(u(\bar{u})) \frac{\partial \Phi^\beta(\bar{u})}{\partial \bar{u}^\alpha}$$

and hence

$$\begin{aligned} \bar{E}_1(\bar{u}) \wedge \cdots \wedge \bar{E}_p(\bar{u}) &= \left[E_{\alpha_1}(u(\bar{u})) \frac{\partial \Phi^{\alpha_1}(\bar{u})}{\partial \bar{u}^1} \right] \wedge \cdots \wedge \left[E_{\alpha_p}(u(\bar{u})) \frac{\partial \Phi^{\alpha_p}(\bar{u})}{\partial \bar{u}^p} \right] \\ &= \underbrace{E_{\alpha_1}(u(\bar{u})) \wedge \cdots \wedge E_{\alpha_p}(u(\bar{u}))}_{\epsilon_{\alpha_1 \dots \alpha_p} E_1(u(\bar{u})) \wedge \cdots \wedge E_p(u(\bar{u}))} \frac{\partial \Phi^{\alpha_1}(\bar{u})}{\partial \bar{u}^1} \cdots \frac{\partial \Phi^{\alpha_p}(\bar{u})}{\partial \bar{u}^p} \\ &= \det \left(\frac{\partial \Phi^\alpha(\bar{u})}{\partial \bar{u}^\beta} \right) E_1(u(\bar{u})) \wedge \cdots \wedge E_p(u(\bar{u})) \end{aligned}$$

since

$$\epsilon_{\alpha_1 \dots \alpha_p} \frac{\partial \Phi^{\alpha_1}(\bar{u})}{\partial \bar{u}^1} \dots \frac{\partial \Phi^{\alpha_p}(\bar{u})}{\partial \bar{u}^p} = \det \left(\frac{\partial \Phi^\alpha(\bar{u})}{\partial \bar{u}^\beta} \right)$$

so

$$\int_{\tilde{\phi}(u)} T = \int_{\bar{u}} \dots \int \frac{1}{p!} T_{i_1 \dots i_p} \circ \Psi \circ \Phi [E_1(u) \wedge \dots \wedge E_p(u)]^{i_1 \dots i_p} \circ \Phi \det \left(\frac{\partial \Phi^\alpha(\bar{u})}{\partial \bar{u}^\beta} \right) d\bar{u}^1 d\bar{u}^2 \dots d\bar{u}^p$$

But this is exactly the re-expression of the ordinary multivariable integral $\int_{\phi(u)} T$ under a change of variable except that the necessary correction factor discussed in Section 11.2

$$\left| \det \left(\frac{\partial \Phi^\alpha}{\partial \bar{u}^\beta} \right) \right|$$

is missing the absolute value sign here and hence will have the correct sign only if

$$\det \left(\frac{\partial \Phi^\alpha}{\partial \bar{u}^\beta} \right) > 0.$$

In other words, once we wish to define an integral of a p -form on a p -surface independent of the parametrization, we have to give it the additional mathematical structure of an “orientation,” which takes into account that this sign can change under a change of variable.

The p -vectors $E_1(u) \wedge \dots \wedge E_p(u)$ for all parametrizations are proportional and nonzero since they determine the same tangent p -plane at each point of the p -surface. However, the nonzero proportionality factor can be positive or negative. Each parametrization determines its own orientation for the p -plane—namely any other basis of the tangent p -plane has the same (opposite) orientation if the proportionality factor relating the new basis p -vector to the given p -vector $E_1(u) \wedge \dots \wedge E_p(u)$ of the parametrization is positive (negative). Then the integral of a p -form on an oriented p -surface is well-defined independent of the choice of parametrization. For a parametrization with the opposite orientation, one simply changes the sign of the integral to get the correct value of the oriented p -surface integral.

In order for this choice of sign to be globally consistent on the p -surface, the p -surface must allow a continuous choice of a nonzero p -vector in every tangent subspace tangent to it, in which case it is said to be orientable. The Mobius strip is a relatively famous example of a nonorientable surface where this game cannot be played globally.

For the case $p = 1$ of a curve, this just corresponds to a choice of direction for the tangent vector E_1 of any parametrization, and hence for the direction along which we move in the “positive direction” along the curve as we increase the parameter in an oriented parametrization. This can be done consistently as long as the parametrization map is 1-1 everywhere, which does not allow the tangent to vanish at any point, where the parametrized curve might retrace its own path.

For the case $p = 2$ of an ordinary surface, this is a choice of a “screw sense” for a loop in the surface. This tells us the direction one must rotate the first basis vector E_1 to go towards the second basis vector E_2 if they are to have the chosen orientation, at each tangent plane to the 2-surface. Within a 3-dimensional space, there is only one extra direction which points on

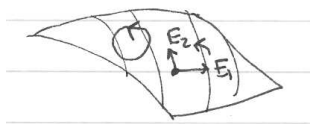
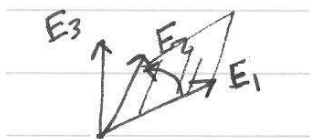


Figure 11.12: Orientation of a 2-surface.

one side or the other of the surface, so one can equivalently specify this orientation by linking it to a choice of this direction. Curling fingers of the right hand from E_1 to E_2 then requires the third direction in any frame for the whole space to lie on the side of the plane of E_1 and E_2 determined by the thumb if the basis is to be right handed, called the right hand rule. In the usual dot product geometry on \mathbb{R}^3 , this leads to the right hand rule choice of unit normal direction from the two possible choices. We will discuss this in detail in general later.

Figure 11.13: Orientation of a 3-surface determined by the right hand rule in \mathbb{R}^3 .

For $p = 3$, this is a choice of a left or right handed basis of each tangent space to the 3-surface. Curling fingers of the right hand from E_1 to E_2 then requires E_3 to lie on the side of the plane of E_1 and E_2 determined by the thumb if the basis is to be right handed. Otherwise it is left handed. Of course this has to be able to be done consistently on the p -surface: the surface must be “orientable.” This is not always possible.

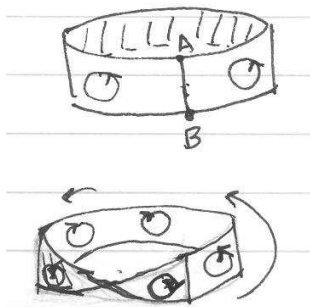


Figure 11.14: A 2-surface can be orientable or not, like the Mobius strip.

Suppose we take a cylindrical strip and cut it as shown in Fig. 11.14 and twist it once so

that A and B exchange places and then re-attach the two ends smoothly. Now as we move our orientation-indicating circle around the strip it comes out reversed after one loop. It is not possible to continuously assign an orientation to this “Möbius strip.” It is not orientable, so one cannot define the integral of a 2-form on it.

11.6 Integration and a metric

One does not need a metric to define the integral of a p -form on a p -surface in an n -dimensional space, whose end result is to produce a real number. However, when a metric is present, one can rewrite the integration process by re-expressing it in terms of the metric, which allows us to interpret the integral using the metric geometry and hence assign a “meaning” to that real number value of the integral. It is easier to illustrate this with a pair of concrete examples before describing the general situation.

Curves in a 3-dimensional flat space

Suppose we are in an $(n = 3)$ -dimensional space with a $(p = 1)$ -dimensional surface, the familiar case of curves in \mathbb{R}^3 but with any flat metric g of any signature, like 3-dimensional Minkowski spacetime \mathcal{M}^3 . Let's work in orthonormal Cartesian coordinates x^i for this metric and a curve c such that $x^i = c^i(\lambda) = x^i(c(\lambda)) = “x^i(\lambda)”$ and U is the interval $\lambda \in [a, b]$. Then with tangent and unit tangent coordinate components

$$c'^i(\lambda) = \frac{dx^i}{d\lambda}(\lambda) = \left| \frac{dx}{d\lambda}(\lambda) \right| \hat{T}^i(\lambda), \quad \left| \frac{dx}{d\lambda} \right| = \left| g_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} \right|^{1/2}, \quad \hat{T}^i \hat{T}_i = \epsilon,$$

we have

$$\begin{aligned} \int_c F &= \int_c F_i dx^i = \int_U c^*(F) = \int_a^b F_i \circ c(\lambda) \frac{dx^i}{d\lambda}(\lambda) d\lambda \\ &= \int_a^b \underbrace{F_i \circ c(\lambda) \hat{T}^i(\lambda)}_{\epsilon F^{\parallel}(\lambda)} \underbrace{\left| \frac{dx^i}{d\lambda}(\lambda) \right|}_{ds(\lambda)} d\lambda \\ &= \int_a^b F_i \circ c(\lambda) \underbrace{\hat{T}^i(\lambda) ds(\lambda)}_{ds^i(\lambda)} = “ \int_c \vec{F} \cdot d\vec{s} ”. \end{aligned}$$

This integral is interpreted as the integral of the tangential component (modulo the sign ϵ) of the corresponding vector field against the differential of arclength along the curve. The quotes indicate the sloppy notation in which we ignore the dependence on λ . This is the familiar situation of a line integral of a vector field in multivariable calculus when we are working in Euclidean 3-space, but in fact nothing of the above depended on the dimension $n = 3$ or having a flat metric, so this is the general situation for the integral of a 1-form on a directed curve for any n and any metric.

Surfaces in a 3-dimensional flat space

Suppose again we are in an $(n = 3)$ -dimensional space but with a $(p = 2)$ -dimensional surface, the familiar situation of ordinary surfaces in \mathbb{R}^3 but with any flat metric g of any signature,

like 3-dimensional Minkowski spacetime \mathcal{M}^3 . Let's work in orthonormal Cartesian coordinates x^i for this metric.

Let \hat{n} be any unit normal to this surface with nonzero length: $\hat{n}_i \hat{n}^i = \epsilon = \pm 1$ (so that this discussion excludes null surfaces where this self-inner product vanishes, a case which requires special handling). Any 2-form on the space can be represented as the dual of a 1-form, which in turn up to sign is just the dual of the 2-form but that sign is really annoying to keep track of. It is convenient to raise its index to a vector field index so that unit oriented 3-form η associated with the metric and the orientation of the whole space is totally covariant

$$\begin{aligned} F &= \frac{1}{2} F_{ij} dx^i \wedge dx^j = \frac{1}{2} (*B)_{ij} dx^i \wedge dx^j = \frac{1}{2} B^k \eta_{kij} dx^i \wedge dx^j \\ &= \frac{1}{2} B^m \delta^k_m \eta_{kij} dx^i \wedge dx^j. \end{aligned}$$

For example, in Euclidean space in Cartesian coordinates we can write this as

$$\begin{aligned} F &= B^1 dx^2 \wedge dx^3 - B^2 dx^1 \wedge dx^3 + B^3 dx^1 \wedge dx^2 \\ &= B^1 dx^2 \wedge dx^3 + B^2 dx^3 \wedge dx^1 + B^3 dx^1 \wedge dx^2. \end{aligned}$$

The alternating sign ordered index pair expression generalizes to other values of n , but the cyclic sum is more convenient for $n = 3$. That inserted Kronecker delta is important but first we need a detour to explain why.

We can orthogonally decompose any tangent vector at a point of the surface to a vector component along the unit normal to the surface and a vector component in the tangent plane to the surface. Define the projection tensor

$$P(\hat{n})^i_j \equiv -\frac{\hat{n}^i \hat{n}_j}{\hat{n}^p \hat{n}_p} + \delta^i_j = \epsilon \hat{n}^i \hat{n}_j + \delta^i_j \quad \leftrightarrow \quad \delta^i_j = \epsilon \hat{n}^i \hat{n}_j - P(\hat{n})^i_j$$

which subtracts away the normal component $X^\perp = \epsilon X^j \hat{n}_j$

$$P(\hat{n})^i_j X^j = -\epsilon \hat{n}^i \hat{n}_j X^j + \delta^i_j X^j = X^i - (\epsilon \hat{n}_j X^j) \hat{n}^i \equiv X^i - X^\perp \hat{n}^i$$

so that we can represent X as a piece along the unit normal and a piece in the tangent plane to surface which is orthogonal to the unit normal

$$X^i = X^\perp \hat{n}^i + P(\hat{n})^i_j X^j, \quad \hat{n}_i P(\hat{n})^i_j X^j = 0.$$

Indeed the surface projection tensor when contracted on either index with the unit normal gives zero. This surface projection removes the components of all tensors along the normal direction to yield a tensor which only lives over the tangent space to the surface. The metric for example projects to 2-tensor on the whole space which evaluates to zero if any vector along the normal direction is inserted into one of its vector slots

$$^{(2)}g_{ij} = g_{mn} P(\hat{n})^m_i P(\hat{n})^n_j,$$

while the unit 3-form projects to a 3-form which also gives 0 upon evaluation on any vector along the unit normal

$$^{(2)}\eta_{ijk} = \eta_{mnp} P(\hat{n})^m{}_i P(\hat{n})^n{}_j P(\hat{n})^p{}_k.$$

Next we are interested in pulling back this 2-form to the surface with a parametrization map Ψ so that we can integrate the resulting 2-form on the parameter space, so if we insert into our previous expression for $\Psi^*(F)$ the representation

$$\delta^k{}_m = \epsilon \hat{n}^k \hat{n}_m + P(\hat{n})^k{}_m$$

of the unit tensor in terms of the projection along and orthogonal to the unit normal, when we pull back $P^k(\hat{n})_m \eta_{kij}$ to the surface we will have a 3-form that only accepts a vector tangent to the surface in its first slot to be nonzero, but the other two slots will be filled by such vectors as well when pulled back to the 2-dimensional parameter space where every 3-form vanishes. Thus only the normal projection term survives which evaluates the first slot along the normal direction, thus allowing the two other slots to be filled by vectors tangent to the surface.

The 2-form $^{(2)}\eta_{ij} = n^k \epsilon_{kij}$ is evaluated on the unit normal in the first slot and therefore evaluates to the area of the parallelogram formed by two vectors X and Y tangent to the surface (such that $\eta(\hat{n}, X, Y) > 0$), since the height of the parallelepiped formed by the 3 vectors is 1, so in fact this 2-form corresponds to the unit volume 2-form on the surface, more commonly called the differential of surface area. Thus the pullback will be

$$\begin{aligned} \Psi^*(F) &= \frac{1}{2} \Psi^* (\epsilon B^m \hat{n}^k \hat{n}_m \eta_{kij} dx^i \wedge dx^j) = \frac{1}{2} \Psi^* (B^\perp {}^{(2)}\eta_{ij} dx^i \wedge dx^j) \\ &= \Psi^* (B^\perp {}^{(2)}\eta) \\ &= "B^\perp dS" = "\epsilon \vec{B} \cdot \hat{n} dS" = "\epsilon \vec{B} \cdot d\vec{S}" . \end{aligned}$$

The last two expressions are in quotes because this is what we usually call the normal component of the vector field $\vec{B} = B^\sharp$ and the differential dS of surface area on the surface, with a vector differential denoted by

$$dS^i = \hat{n}^i dS = \hat{n}^i {}^{(2)}\eta.$$

The Cartesian components of this vector-valued 2-form in a Cartesian coordinate frame in the Euclidean case are

$$d\vec{S} = \langle dy \wedge dx, dz \wedge dx, dx \wedge dy \rangle ,$$

and the pullback of F will be

$$\Psi^*(F) = \Psi^*(B^1 dy \wedge dx + B^2 dz \wedge dx + B^3 dx \wedge dy) .$$

The payoff of these manipulations is that we can interpret the integral of this 2-form on the surface as the integral of the normal component of the vector field against the differential of surface area, which in the physical application is called the flux of the vector field through that region of the surface over which the integration takes place. For a vector field in Euclidean \mathbb{R}^3 representing a stationary velocity field of a fluid, whose flow lines are the paths of the fluid, this would quantify the flow rate of the fluid through the surface. Area times velocity has units of

volume per time so this would give the amount of fluid which passes through this surface region per unit time. For the electric or magnetic fields, this gives the flux of those fields through a region of a surface, important in the interpretation of how these fields act on charges and currents.

Exercise 11.6.1.

integration of a 2-form in 3-spacetime

a) Consider the above discussion in 3-dimensional Minkowski spacetime \mathcal{M}^3 in oriented coordinates $(x^0, x^1, x^2) = (t, x, y)$, with unit volume 3-form $\eta_{012} = \epsilon_{012} = 1$ and metric

$$g = -dt \otimes dt + dx \otimes dx + dy \otimes dy,$$

and a spacelike surface $\Sigma : t = t_0$ with future-pointing unit timelike normal $\hat{n} = \partial_t$ on that surface with components $\langle 1, 0, 0 \rangle$, for which $\epsilon = -1$. We can use the parametrization map

$$(t, x, y) = \Psi(u^1, u^2) = (t_0, u^1, u^2),$$

and integrate over some region U of the parameter space such that $\Sigma = \Psi(U)$. Consider a 2-form

$$F = B^0 dx \wedge dy + B^1 dy \wedge dt + B^2 dt \wedge dx.$$

Show that

$$\int_{\Sigma} F = \int_U \Psi^*(B^0) du^1 du^2 = \left[\int_{\Sigma} B^0 dx dy \right] = \left[\int_{\Sigma} -B^i \hat{n}_i dx dy \right] = \left[\int_{\Sigma} -B^i dS_i \right],$$

where the quotes mean we are using sloppy notation where we don't distinguish between the coordinates and parameters $(x, y) = (u^1, u^2)$.

b) What happens to the signs in this calculation if we repeat it for a timelike surface $x = x_0$ with a unit normal $\hat{n} = \partial_x$?

■

Integrating over a p -surface in an n -dimensional space

The remaining case $p = 3$ for $n = 3$ is straightforward. The integral of a 3-form over a region of the space translates into the integral of its dual (modulo that annoying sign), a scalar function, with respect to the unit volume 3-form of the metric, i.e., the differential of volume on the space.

The case $p = n - 1$ of a hypersurface oriented by a chosen unit normal behaves very much like surfaces in three dimensions. The dual of an $(n - 1)$ -form is a 1-form, and the hypersurface integral is again interpreted as the integral of the normal component of the corresponding vector field with respect to the differential of hypersurface area. The way that a unit normal to a hypersurface corresponds to an orientation of the hypersurface requires a lengthy discussion of its own to handle the general case, which we postpone till a later section.

For the cases $n/2 < p < n - 1$, then $n - p < p$ so that by taking the dual of the p -form, we reduce the number of indices as in the case $p = n - 1$ and can interpret the integral in terms of the normal components of the dual (more economical) or the tangential components of the p -form itself (less economical) against the differential of p -surface volume. When $p = n/2$ for an even dimension, the dual also has p indices so they are equally index laden in interpretation. For $n = 4$ we get the first case of this type: the dual of a 2-form is a 2-form.

In progress . . . but out of time now. We will discuss these in the context of Stokes' theorem later, and eventually I will flush this out in more detail.

11.7 The exterior derivative d

Suppose we work in a coordinate frame on \mathbb{R}^n or some n -dimensional space of interest. Then for each p satisfying $0 \leq p \leq n$, we have p -forms or “differential forms of degree p ” which may be expressed in terms of the coordinate differentials as

$$\begin{aligned} T &= \frac{1}{p!} T_{i_1 \dots i_p} dx^{i_1 \dots i_p} = \frac{1}{p!} T_{i_1 \dots i_p} dx^{i_1} \wedge \dots \wedge dx^{i_p} \\ &= T_{i_1 \dots i_p} dx^{[i_1 \dots i_p]} \quad (\text{if we don't want to overcount}) \\ &= T_{i_1 \dots i_p} dx^{i_1} \otimes \dots \otimes dx^{i_p}, \quad (\text{if we want to just think of it} \\ &\quad \text{as a } \binom{0}{p} \text{-tensor field}) \end{aligned}$$

where of course $T_{i_1 \dots i_p} = T_{[i_1 \dots i_p]}$ holds.

A function f is a 0-form and its differential $df = f_{,i} dx^i = (\partial f / \partial x^i) dx^i$ is a 1-form. Thus the differential d maps 0-forms to 1-forms, the extra covariant index being the derivative index. If we start instead with a p -form, adding a derivative index to its component symbol will not yield an object which is antisymmetric in all of its indices unless we also take the antisymmetric part of this new object. We will then get a $(p+1)$ -form. Apart from a normalization constant, this is how the *exterior derivative* d is defined as an extension of the operator d which takes the differential of a function.

The actual definition is simple for $1 \leq p \leq n$

$$\begin{array}{ccc} p\text{-form} & \xrightarrow{d} & (p+1)\text{-form} \\ (1) \quad T = \frac{1}{p!} T_{i_1 \dots i_p} dx^{i_1 \dots i_p} & \xrightarrow{d} & dT \equiv \frac{1}{p!} dT_{i_1 \dots i_p} \wedge dx^{i_1 \dots i_p} . \end{array}$$

In words, take the differential of its component functions and wedge that into the coordinate frame basis p -form to obtain a $(p+1)$ -form. This is all we need in practice to evaluate dT for any p -form T , but we can develop shortcut formulas.

Its components are easily calculated by expanding the differential using the definitions $df = f_{,i} dx^i$ and $dx^j \wedge dx^{i_1 \dots i_p} = dx^{j i_1 \dots i_p}$

$$\begin{aligned} dT &= \frac{1}{p!} dT_{i_1 \dots i_p} \wedge dx^{i_1 \dots i_p} = \frac{1}{p!} T_{i_1 \dots i_p, j} dx^j \wedge dx^{i_1 \dots i_p} \\ &= \frac{1}{p!} T_{[i_1 \dots i_p, j]} dx^{j i_1 \dots i_p} \quad (\text{only the antisymmetric part contributes}) \\ &\equiv \frac{1}{(p+1)!} [dT]_{i_1 \dots i_p} dx^{j i_1 \dots i_p}, \quad (\text{definition of components of } (p+1)\text{-form}) \end{aligned}$$

Comparing the last two equalities we get

$$[dT]_{j i_1 \dots i_p} = \frac{(p+1)!}{p!} T_{[i_1 \dots i_p, j]} = (p+1) T_{[i_1 \dots i_p, j]} .$$

So the exterior derivative of a p -form has its coordinate components equal to $(p+1)$ times the antisymmetric part of their derivatives, except the extra index is added at the beginning instead of at the end as in the covariant derivative. The notation $\partial_i f \equiv f_{,i}$ is better suited to this

$$[dT]_{ji_1 \dots i_p} = (p+1) \partial_{[j} T_{i_1 \dots i_p]} .$$

The factor of $(p+1)$ is necessary to eliminate overcounting. Suppose we expand this expression using the definition of the antisymmetric part

$$\partial_{[j} T_{i_1 \dots i_p]} = \frac{1}{(p+1)!} \delta_{ji_1 \dots i_p}^{mn_1 \dots n_p} \partial_m T_{n_1 \dots n_p}$$

so that we get

$$[dT]_{ji_1 \dots i_p} = \frac{(p+1)}{(p+1)!} \delta_{ji_1 \dots i_p}^{mn_1 \dots n_p} \partial_m T_{n_1 \dots n_p} = \delta_{ji_1 \dots i_p}^{mn_1 \dots n_p} \partial_m T_{[n_1 \dots n_p]} .$$

The factorial factor disappears once we avoid overcounting in the sum over the p antisymmetric indices. Relabeling the indices we can also write this as

$$(2) \quad [dT]_{i_1 \dots i_{p+1}} = (p+1) \partial_{[i_1} T_{i_2 \dots i_{p+1}]} = \delta_{i_1 i_2 \dots i_{p+1}}^{j_1 j_2 \dots j_{p+1}} \partial_{j_1} T_{[j_2 \dots j_{p+1}]} .$$

But we can do better. In this antisymmetrization over $p+1$ indices, p of them are already antisymmetric, so the complete antisymmetrization collapses to something much simpler.

Recall that the generalized Kronecker delta may be defined as the determinant of a matrix of ordinary Kronecker deltas, which we can then expand along the first row using a cofactor expansion, the minors of which are by definition Kronecker deltas of one less order

$$\begin{aligned} \delta_{i_1 i_2 \dots i_{p+1}}^{j_1 j_2 \dots j_{p+1}} &= \begin{vmatrix} \delta_{i_1}^{j_1} & \delta_{i_2}^{j_1} & \dots & \delta_{i_{p+1}}^{j_1} \\ \delta_{i_1}^{j_2} & \delta_{i_2}^{j_2} & \dots & \delta_{i_{p+1}}^{j_2} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{i_1}^{j_{p+1}} & \delta_{i_2}^{j_{p+1}} & \dots & \delta_{i_{p+1}}^{j_{p+1}} \end{vmatrix} \\ &= \delta_{i_1}^{j_1} \delta_{i_2 \dots i_{p+1}}^{j_2 \dots j_{p+1}} - \delta_{i_2}^{j_1} \delta_{i_1 i_3 \dots i_{p+1}}^{j_2 \dots j_{p+1}} + \delta_{i_3}^{j_1} \delta_{i_1 i_2 i_4 \dots i_{p+1}}^{j_2 \dots j_{p+1}} - \dots + (-1)^p \delta_{i_{p+1}}^{j_1} \delta_{i_1 \dots i_p}^{j_2 \dots j_{p+1}} \\ &= \sum_{k=1}^{p+1} (-1)^{k-1} \delta_{i_k}^{j_1} \delta_{i_1 \dots \widehat{i_k} \dots i_{p+1}}^{j_2 \dots j_{p+1}} \\ &\equiv \sum_{k=1}^{p+1} (-1)^{k-1} \delta_{i_k}^{j_1} \delta_{i_1 \dots \widehat{i_k} \dots i_{p+1}}^{j_2 \dots j_{p+1}} , \end{aligned}$$

where $\widehat{i_k}$ means this index is omitted from the index set (a convenient abbreviation).

Using this formula for the exterior derivative gives what might be called the “alternating formula”

$$\begin{aligned}
 [dT]_{i_1 \dots i_{p+1}} &= \delta_{i_1 i_2 \dots i_{p+1}}^{j_1 j_2 \dots j_{p+1}} \partial_{j_1} T_{|j_2 \dots j_{p+1}|} = \sum_{k=1}^{p+1} (-1)^{k-1} \delta_{i_k}^{j_1} \delta_{i_1 \dots \widehat{i_k} \dots i_{p+1}}^{j_2 \dots j_{p+1}} \partial_{j_1} T_{|j_2 \dots j_{p+1}|} \\
 &= \sum_{k=1}^{p+1} (-1)^{k-1} \partial_{i_k} T_{i_1 \dots \widehat{i_k} \dots i_{p+1}} \\
 (3) \quad &= \partial_{i_1} T_{i_2 \dots i_{p+1}} - \partial_{i_2} T_{i_1 i_3 \dots i_{p+1}} + \partial_{i_3} T_{i_1 i_2 i_4 \dots i_{p+1}} + \dots + (-1)^p \partial_{i_p} T_{i_1 \dots i_{p-1} i_{p+1}}.
 \end{aligned}$$

So we have three different looking formulas (1), (2), and (3) that we can use to compute the components of the exterior derivative d of a p -form, in addition to the definition which works with the p -form itself.

Explicit formulas are simple

This has to be made more explicit to sink in, so we examine the first few cases using the first definition of the exterior derivative (not one of these three component formulas).

The case $p = 1$: $T = T_i dx^i$.

$$\begin{aligned}
 dT &= dT_i \wedge dx^i = \partial_j T_i dx^j \wedge dx^i = \partial_j T_i dx^{ji} \\
 &= \partial_{[j} T_{i]} dx^{ji} = \frac{1}{2} (2\partial_{[j} T_{i]}) dx^{ji} \equiv \frac{1}{2} [dT]_{ji} dx^{ji}
 \end{aligned}$$

so

$$[dT]_{ji} = \underbrace{2\partial_{[j} T_{i]}}_{\text{formula (2) with } p+1=2} = \underbrace{\partial_j T_i - \partial_i T_j}_{\text{formula (3)}}.$$

The case $p = 2$: $T = \frac{1}{2} T_{ij} dx^{ij}$.

$$\begin{aligned}
 dT &= \frac{1}{2} dT_{ij} \wedge dx^{ij} = \frac{1}{2} \partial_k T_{ij} dx^k \wedge dx^{ij} = \frac{1}{2} \partial_{[k} T_{ij]} dx^{kij} \\
 &= \frac{1}{3!} (3\partial_{[k} T_{ij]}) dx^{kij} \equiv \frac{1}{3!} [dT]_{kij} dx^{kij}
 \end{aligned}$$

so

$$\begin{aligned}
 [dT]_{kij} &= \underbrace{3\partial_{[k} T_{ij]}}_{\text{formula (2) with } p+1=3} = 3 \cdot \frac{1}{3!} \begin{pmatrix} \partial_k T_{ij} + \partial_i T_{jk} + \partial_j T_{ki} \\ -\partial_k T_{ji} - \partial_i T_{kj} - \partial_j T_{ik} \end{pmatrix} \\
 &= \underbrace{\partial_k T_{ij} + \partial_i T_{jk} + \partial_j T_{ki}}_{\text{cyclic sum easier to remember}} = \underbrace{\partial_k T_{ij} - \partial_i T_{kj} + \partial_j T_{ki}}_{\text{formula (3)}},
 \end{aligned}$$

where we used the antisymmetric property $T_{jk} = -T_{kj}$ a number of times. In this case the cyclic sum $\partial_k T_{ij} + \partial_i T_{jk} + \partial_j T_{ki}$ is easier to remember than the alternating sign formula (3).

This is why we use the index pairs 23, 31, 12 instead of the ordered pairs 23, 13, 12; however, there is more reason to do this as we will see below.

The case $p = n - 1$:

$$T = \frac{1}{(n-1)!} T_{i_1 \dots i_{n-1}} dx^{i_1 \dots i_{n-1}} = T_{i_1 \dots i_{n-1}} dx^{|i_1 \dots i_{n-1}|}.$$

Note that there are only n terms in this sum since in each wedge product exactly one coordinate differential is missing: $T = \sum_{j=1}^n T_{1 \dots \hat{j} \dots n} dx^{1 \dots \hat{j} \dots n}$. Then using the fact that $dx^k \wedge dx^{1 \dots \hat{j} \dots n}$ equals zero if $k \neq j$ (since two factors of dx^k will lead to zero) and equals $(-1)^{j-1} dx^{1 \dots j \dots n}$ if $k = j$ since the dx^j factor must jump over $j-1$ indices to get to its natural ordered position, one finds

$$\begin{aligned} dT &= \sum_{j=1}^n dT_{1 \dots \hat{j} \dots n} dx^{1 \dots \hat{j} \dots n} = \sum_{j=1}^n \partial_k T_{1 \dots \hat{j} \dots n} dx^k \wedge dx^{1 \dots \hat{j} \dots n} \\ &= \sum_{j=1}^n \partial_j T_{1 \dots \hat{j} \dots n} (-1)^{j-1} dx^{1 \dots j \dots n}, \end{aligned}$$

so

$$[dT]_{1 \dots n} = \sum_{j=1}^n (-1)^{j-1} \partial_j T_{1 \dots \hat{j} \dots n},$$

which is formula (3).

The case $p = n$: $T = T_{1 \dots n} dx^{1 \dots n}$.

$$dT = dT_{1 \dots n} \wedge dx^{1 \dots n} = \partial_k T_{1 \dots n} dx^k \wedge dx^{1 \dots n} = 0,$$

since p -forms are identically zero if $p > n$ since one necessarily has repeated indices.

Okay, so we've done as much as we can with formulas. How does this work in practice?

Example 11.7.1. • Recall our friend $X^\flat = ydx + xdy = df$, where $f = xy$. Then by antisymmetry

$$dX^\flat = dy \wedge dx + dx \wedge dy = (-dx \wedge dy) + dx \wedge dy = 0,$$

which shows that $d^2 f \equiv d(df) = 0$.

• Consider the 2-form $T = (x^2 + y^2)dx + (x^2 - y^2)dy$. Then

$$dT = (2xdx + 2ydy) \wedge dx + (2xdx - 2ydy) \wedge dy = 2ydy \wedge dx + 2xdx \wedge dy = 2(x - y)dx \wedge dy$$

and

$$d^2 T \equiv d(dT) = 2(dx - dy) \wedge dx \wedge dy = 0.$$

Again $d^2 T = 0$.

- Now a 3-form $T = xyz \, dy \wedge dz + (x + y)dz \wedge dx + \sin(x + z)dx \wedge dy$. Then

$$dT = yz \, dx \wedge dy \wedge dz + dy \wedge dz \wedge dx + \cos(x + z)dz \wedge dx \wedge dy = [yz + 1 + \cos(x + z)] \, dx \wedge dy \wedge dz$$

and $d^2T = d(dT) = 0$ since there is no independent coordinate differential left to be wedged into $dx \wedge dy \wedge dz$ in three dimensions. \square

Exercise 11.7.1.

exterior derivatives in cylindrical coordinates

We can also evaluate these exterior derivatives in other coordinate systems, like cylindrical coordinates. For example

$$X^\flat = \rho \sin 2\phi \, d\rho + \rho^2 \cos 2\phi \, d\phi$$

$$dX^\flat = 2\rho \cos 2\phi \, d\phi \wedge d\rho + 2\rho \cos 2\phi \, d\rho \wedge d\phi = (2\rho \cos \phi - 2\rho \cos \phi) d\rho \wedge d\phi = 0.$$

- a) Transform T and dT of the previous page to cylindrical coordinates to obtain

$$\begin{aligned} T &= (x^2 + y^2)dx + (x^2 - y^2)dy = \dots \\ &= \rho^2[\cos \phi + \cos 2\phi \sin \phi]d\rho + \rho^3[-\sin \phi + \cos 2\phi \cos \phi]d\phi, \\ dT &= 2(x - y)dx \wedge dy = \dots = 2\rho^2(\cos \phi - \sin \phi)d\rho \wedge d\phi. \end{aligned}$$

- b) Now doing the exterior derivative in cylindrical coordinates, show that this result for dT is what you actually get (using trig identities!). \blacksquare

Two facts seem to be coming to light.

1. $d^2T \equiv d(dT) = 0$ for any p -form T .
2. Our definition of dT in a particular coordinate system is actually independent of the coordinate system.

To show the first just do the exterior derivative twice

$$\begin{aligned} dT &= \frac{1}{p!} dT_{i_1 \dots i_p} \wedge dx^{i_1 \dots i_p} = \frac{1}{p!} T_{i_1 \dots i_p, j} dx^j \wedge dx^{i_1 \dots i_p} \\ d^2T &= \frac{1}{p!} d(T_{i_1 \dots i_p, j}) \wedge dx^j \wedge dx^{i_1 \dots i_p} = \frac{1}{p!} T_{i_1 \dots i_p, jk} dx^k \wedge dx^j \wedge dx^{i_1 \dots i_p} \\ &= \frac{1}{p!} T_{[i_1 \dots i_p, jk]} dx^{kj i_1 \dots i_p}. \end{aligned}$$

But partial derivatives commute so antisymmetrizing over the index set containing the symmetric pair jk symbolizing the second partial derivatives whose order doesn't matter gives zero: $d^2T = 0$.

Inessential mathematical games detour

Not so fast you say. I should have antisymmetrized after taking the first derivative and then again after taking the second derivative. How do I know that this is equivalent to just antisymmetrizing after taking the second derivative which is what I did? Well, our powerful notation automatically incorporates all the properties necessary to make things work out. At each step only the antisymmetric part contributes since the lower indices are summed against an antisymmetric upper set of indices, so it is not necessary to make explicit the antisymmetrization over the lower indices.

However, just for fun let's use the component formulas to confirm this fact. Then

$$\begin{aligned}
 [dT]_{i_1 \dots i_{p+1}} &= \frac{1}{p!} \delta_{i_1 \dots i_{p+1}}^{j_1 \dots j_{p+1}} \partial_{j_1} T_{j_2 \dots j_{p+1}}, \\
 [d(dT)]_{i_1 \dots i_{p+2}} &= \frac{1}{(p+1)!} \delta_{i_1 i_2 \dots i_{p+2}}^{j_1 j_2 \dots j_{p+2}} \partial_{j_1} [dT]_{j_2 \dots j_{p+2}} \\
 &= \frac{1}{(p+1)!} \delta_{i_1 i_2 \dots i_{p+2}}^{j_1 j_2 \dots j_{p+2}} \partial_{j_1} \left(\frac{1}{p!} \delta_{j_2 \dots j_{p+2}}^{k_2 \dots k_{p+2}} \partial_{k_2} T_{k_3 \dots k_{p+2}} \right) \\
 &= \frac{1}{(p+1)!} \left(\frac{1}{p!} \delta_{j_2 \dots j_{p+2}}^{k_2 \dots k_{p+2}} \delta_{i_1 i_2 \dots i_{p+2}}^{j_1 j_2 \dots j_{p+2}} \right) \partial_{j_1} \partial_{k_2} T_{k_3 \dots k_{p+2}} \\
 &= \frac{1}{(p+1)!} \left(\delta_{i_1 i_2 \dots i_{p+2}}^{j_1 k_2 \dots k_{p+2}} \right) \partial_{j_1} \partial_{k_2} T_{k_3 \dots k_{p+2}} \\
 &= 0,
 \end{aligned}$$

since $\partial_{[i} \partial_{j]} = 0$ (the order of partial derivatives does not matter), using the fact that the antisymmetrizer on the p indices $j_2 \dots j_{p+2}$ on the second delta factor in the parentheses has no effect since it is already antisymmetric.

Now why is d independent of the coordinate system? An ugly way to show this is to simply transform its components and show that they obey the correct transformation law. We made a big deal out of the fact that the partial derivatives of tensor components do not transform “as a tensor” which led to the covariant derivative, but antisymmetrization kills the second derivative terms which arise from the derivatives of $A^i_j = \partial \bar{x}^i / \partial x^j$, restoring the correct transformation rule.

Another inessential detour

So here we go with components. First we do a preliminary calculation involving the second derivatives

$$A^i_j = \frac{\partial \bar{x}^i}{\partial x^j} \rightarrow dA^i_j = \frac{\partial \bar{x}^i}{\partial x^\ell \partial x^j} dx^\ell = A^i_{j,\ell} dx^\ell$$

and hence

$$dA^i_j \wedge dx^j = A^i_{j,\ell} dx^\ell \wedge dx^j = A^i_{[j,\ell]} dx^\ell \wedge dx^j = 0,$$

which is zero since $f_{[j,\ell]} = \frac{1}{2}[\partial^2 f / \partial x^j \partial x^\ell - \partial^2 f / \partial x^\ell \partial x^j] = 0$ for any function.

Next we need another preliminary calculation which flips the derivative from the matrix inverse back to the matrix in their product

$$A^{-1k}_i A^i_j = \delta^k_j \rightarrow (dA^{-1k}_i) A^i_j + A^{-1k}_i dA^i_j = 0 \rightarrow (dA^{-1k}_i) A^i_j = -A^{-1k}_i dA^i_j.$$

Then we need the transformation of the components and of the differentials

$$\bar{T}_{i_1 \dots i_p} = A^{-1k_1}_{i_1} \dots A^{-1k_p}_{i_p} T_{k_1 \dots k_p}, \quad d\bar{x}^{i_1 \dots i_p} = A^{i_1}_{j_1} \dots A^{i_p}_{j_p} dx^{j_1 \dots j_p}.$$

Finally we are ready to attack the differential expressed in the new coordinates in the following calculation which starts out just with a simple substitution, followed by the product rule, followed by simplifying the first term in which the matrix products reduce to the identity, followed by substitution of the derivative formula for the inverse matrix in the next-to-last line:

$$\begin{aligned} \bar{d} T &= \frac{1}{p!} d\bar{T}_{i_1 \dots i_p} \wedge d\bar{x}^{i_1 \dots i_p} \\ &= \frac{1}{p!} d[A^{-1k_1}_{i_1} \dots A^{-1k_p}_{i_p} T_{k_1 \dots k_p}] \wedge [A^{i_1}_{j_1} \dots A^{i_p}_{j_p} dx^{j_1 \dots j_p}] \\ &= \frac{1}{p!} [A^{-1k_1}_{i_1} \dots A^{-1k_p}_{i_p}] [A^{i_1}_{j_1} \dots A^{i_p}_{j_p}] dT_{k_1 \dots k_p} \wedge dx^{j_1 \dots j_p} \\ &\quad + \frac{1}{p!} [dA^{-1k_1}_{i_1} \dots A^{-1k_p}_{i_p} + \dots + A^{-1k_1}_{i_1} \dots dA^{-1k_p}_{i_p}] A^{i_1}_{j_1} \dots A^{i_p}_{j_p} \wedge T_{k_1 \dots k_p} dx^{j_1 \dots j_p} \\ &= \frac{1}{p!} \underbrace{dT_{j_1 \dots j_p} \wedge dx^{j_1 \dots j_p}}_{= dT} \\ &\quad + \frac{1}{p!} \underbrace{[dA^{-1k_1}_{i_1} A^{i_1}_{j_1} \delta^{k_2}_{j_2} \dots \delta^{k_p}_{j_p} + \dots + \delta^{k_1}_{j_1} \dots \delta^{k_{p-1}}_{j_{p-1}} dA^{-1k_p}_{i_p} A^{i_p}_{j_p}] \wedge T_{k_1 \dots k_p} dx^{j_1 \dots j_p}}_{= \frac{1}{p!} [A^{-1k_1}_{i_1} dA^{i_1}_{j_1} T_{k_1 j_2 \dots j_p} + \dots + A^{-1k_p}_{i_p} dA^{i_p}_{j_p} T_{k_1 \dots k_p}] \wedge dx^{j_1} \wedge \dots \wedge dx^{j_p}} \\ &= dT, \end{aligned}$$

since each of these additional terms after the first vanishes because it contains a factor like $dA^i_j \wedge dx^j = 0$. Thus $\bar{d}T = dT$.

Properties of d

What properties does the exterior derivative have? Well, it is a derivative operator so it should obey sum and product rules

$$\begin{aligned}
 T + S &= \frac{1}{p!} T_{i_1 \dots i_p} dx^{i_1 \dots i_p} + \frac{1}{p!} S_{i_1 \dots i_p} dx^{i_1 \dots i_p} \\
 &= \frac{1}{p!} (T_{i_1 \dots i_p} + S_{i_1 \dots i_p}) dx^{i_1 \dots i_p} \\
 d(T + S) &= \frac{1}{p!} d(T_{i_1 \dots i_p} + S_{i_1 \dots i_p}) \wedge dx^{i_1 \dots i_p} \\
 &= \frac{1}{p!} (dT_{i_1 \dots i_p} + dS_{i_1 \dots i_p}) \wedge dx^{i_1 \dots i_p} \quad (\text{ordinary differential sum rule}) \\
 &= \frac{1}{p!} [dT_{i_1 \dots i_p} \wedge dx^{i_1 \dots i_p} + dS_{i_1 \dots i_p} \wedge dx^{i_1 \dots i_p}]
 \end{aligned}$$

so

$$(A) \quad d(T + S) = dT + dS.$$

Now if

$$T = \frac{1}{p!} T_{i_1 \dots i_p} dx^{i_1 \dots i_p}, \quad S = \frac{1}{q!} S_{j_1 \dots j_q} dx^{j_1 \dots j_q},$$

Then

$$\begin{aligned}
 T \wedge S &= \frac{1}{p!q!} T_{i_1 \dots i_p} S_{j_1 \dots j_q} dx^{i_1 \dots i_p} \wedge dx^{j_1 \dots j_q} \\
 d(T \wedge S) &= \frac{1}{p!q!} [dT_{i_1 \dots i_p} S_{j_1 \dots j_q} + T_{i_1 \dots i_p} dS_{j_1 \dots j_q}] \wedge dx^{i_1 \dots i_p} \wedge dx^{j_1 \dots j_q} \\
 &= \left(\frac{1}{p!} dT_{i_1 \dots i_p} \wedge dx^{i_1 \dots i_p} \right) \wedge \left(\frac{1}{q!} S_{j_1 \dots j_q} dx^{j_1 \dots j_q} \right) \\
 &\quad + \frac{1}{p!} T_{i_1 \dots i_p} \frac{1}{q!} \underbrace{\frac{dS_{j_1 \dots j_q} \wedge dx^{i_1 \dots i_p}}{\partial_k S_{j_1 \dots j_q} \underbrace{dx^k \wedge dx^{i_1 \dots i_p}}_{(-1)^p dx^{i_1 \dots i_p} \wedge dx^k}}}_{(-1)^p dx^{i_1 \dots i_p} \wedge \partial_k S_{j_1 \dots j_q} dx^k} \wedge dx^{j_1 \dots j_q} \\
 &\quad \underbrace{\quad}_{dx^{i_1 \dots i_p} \wedge dS_{j_1 \dots j_q}} \\
 &= (-1)^p \left(\frac{1}{p!} T_{i_1 \dots i_p} dx^{i_1 \dots i_p} \right) \wedge \left(\frac{1}{q!} dS_{j_1 \dots j_q} \wedge dx^{j_1 \dots j_q} \right) \\
 &= dT \wedge S + (-1)^p T \wedge dS,
 \end{aligned}$$

so

$$(B) \quad d(T \wedge S) = dT \wedge S + (-1)^p T \wedge dS$$

and finally

$$(C) \quad d^2 T \equiv d(dT) = 0.$$

These three properties (A), (B), and (C) uniquely characterize the exterior derivative. A final property extends the coordinate independence of this operator to any map between two spaces M and N .

Suppose $\Phi : M \rightarrow N$ is a map between two spaces like \mathbb{R}^n and \mathbb{R}^m and suppose T is a p -form on the image space N . Then Φ^*T is its pull back to M , also a p -form. It turns out that we can do the exterior derivative before or after the pull back and still get the same result

$$d(\Phi^*T) = \Phi^*(dT),$$

where the d on the left is the exterior derivative on M while the d on the right is the exterior derivative on N . This is best summarized in what is called a “commutative diagram.” Let T_M and T_N be p -forms on M and N respectively. While Φ maps forward from M to N , the pull back operation pulls back the differential forms from N to M

$$\begin{array}{ccc} M & \xrightarrow{\Phi} & N \\ & & \\ T_M & \xleftarrow{\Phi^*} & T_N \\ \downarrow d & & \downarrow d \\ dT_M & \xleftarrow{\Phi^*} & dT_N \end{array}$$

It does not matter if you first pull back T_N from N to M and then take its exterior derivative on M , or first take its exterior derivative on N and then pull back to M , you get the same result, hence the “commutativity” of the two paths from the top right to the bottom left of the diagram.

A special case of this are the parameter maps associated with a new non-Cartesian coordinate system on \mathbb{R}^n . Expressing a p -form in terms of the new coordinates is equivalent to pulling it back to the coordinate space. Computing its exterior derivative in the new coordinates yields the same result as first taking the exterior derivative in Cartesian coordinates and then re-expressing the result in the new coordinates.

Exercise 11.7.2.

exterior derivative in a frame, curvature 2-form

a) A component independent formula for the exterior derivative of a 1-form σ evaluated on a pair of vector fields X and Y is easily verified by simply expanding both sides of the formula using the coordinate component definition of the exterior derivative and the product rule

$$\begin{aligned} d\sigma(X, Y) &= X\sigma(Y) - Y\sigma(X) - \sigma([X, Y]), \\ (\sigma_{j,i} - \sigma_{i,j})X^iY^j &= X^i(\sigma_jY^j)_{,i} - Y^j(\sigma_iX^i)_{,j} - \underbrace{\sigma_k(X^iY^k_{,i} - Y^jX^k_{,j})}_{(\sigma_jX^iY^j_{,i} - \sigma_iY^jX^i_{,j})}. \end{aligned}$$

Finish this last step of expanding and canceling 4 terms on the right hand side to obtain the left hand side.

b) Given a frame $\{e_i\}$ with Lie brackets $[e_i, e_j] = C^k_{ij}e_k$ and dual frame $\{\omega^i\}$, the previous formula together with the duality relations $\omega^k(e_i) = \delta^k_i$ leads to

$$(d\omega^k)_{ij} = d\omega^k(e_i, e_j) = e_i(\omega^k(e_j)) - e_j(\omega^k(e_i)) - \omega^k([e_i, e_j]) = -C^k_{ij}$$

for the frame components of the 2-form $d\omega^k$, which means that the 2-form itself is therefore

$$d\omega^k = -\frac{1}{2}C^k_{ij}\omega^i \wedge \omega^j.$$

Use this relation together with the product rule $d(f\sigma) = df \wedge \sigma + f d\sigma$ to derive the frame component formula for the exterior derivative of a 1-form $\sigma = \sigma_k \omega^k$

$$(d\sigma)_{ij} = \sigma_{j,i} - \sigma_{i,j} - \sigma_k C^k_{ij} = 2\sigma_{[j,i]} - \sigma_k C^k_{ij}.$$

Thus additional structure function terms appear in the formula for the exterior derivative of a p -form compared to the coordinate frame formula. One can easily use the product rule for differential forms to write down a more general formula for a p -form, but we don't need it here.

c) For a metric $g = g_{ij}\omega^i \otimes \omega^j$ with connection 1-form matrix $\underline{\omega} = (\omega^i_j) = (\Gamma^i_{kj}\omega^k)$, recall the vanishing torsion $(\frac{1}{2})$ -tensor expressed the symmetry of the connection

$$\begin{aligned} T(X, Y) &= \nabla_X Y - \nabla_Y X - [X, Y] = 0, \\ T^k(e_i, e_j) &= T^k_{ij} = \Gamma^k_{[ij]} - C^k_{ij} = 0. \end{aligned}$$

This can be interpreted as a vector-valued 2-form with frame vector components

$$\Theta^k(X, Y) = \omega^k(\Theta(X, Y)) = \omega^k(\nabla_X Y) - \omega^k(\nabla_Y X) - \theta^k([X, Y]),$$

and compared to

$$d\omega^k(X, Y) = X\omega^k(Y) - Y\omega^k(X) - \theta^k([X, Y]).$$

Express the difference in terms of the connection 1-forms to show that

$$\Theta^k(X, Y) = d\omega^k(X, Y) + \omega^k_j(X)\theta^j(Y) - \omega^k_j(Y)\theta^j(X)$$

so that one has the identity (Cartan's first structural equations)

$$\Theta^k = d\omega^k + \omega^k_j \wedge \omega^j = 0.$$

d) Introduce the curvature 2-form matrix (components of the tensor-valued 2-form)

$$\underline{\Omega} = (\Omega^i_j) = \left(\frac{1}{2} R^i_{jkl} \omega^k \wedge \omega^l \right),$$

and show that the frame component formula for the curvature tensor components can be rewritten in the following way (Cartan's second structural equations)

$$\underline{\Omega} = d\underline{\omega} + \underline{\omega} \wedge \underline{\omega},$$

where the matrix product is implied in the second term, i.e.,

$$d\Omega^i_j = d\omega^i_j + \omega^i_k \wedge \omega^k_j.$$

The curvature 2-form packaging of the curvature tensor emphasizes that it is a linear transformation-valued 2-form whose 2-form arguments naturally pick out the plane spanned by the two vector field arguments, i.e., $\Omega^i_j(X, Y)Z^j$ is the limiting linear transformation of a vector Z when parallel transported around the shrinking loop parallelogram of X and Y as described in Chapter 9, once the magnitude of the 2-vector $X \wedge Y$ is divided out (and a sign factor is determined in the case negative self-inner products exist).

e) Take the exterior derivative of Cartan's first structural formula, using $d^2 = 0$ and replacing $d\omega$ using the definition of the curvature 2-form, to obtain Bianchi's first identity

$$0 = d\Theta^i + \omega^i_j \Theta^j = \Omega^i_j \wedge \omega^j = \frac{1}{2} R^i_{jmn} \omega^j \wedge \omega^m \wedge \omega^n = \frac{1}{2} R^i_{[jmn]} \omega^j \wedge \omega^m \wedge \omega^n,$$

which in component form translates to

$$3R^k_{[jmn]} = R^k_{jmn} + R^k_{mnj} + R^k_{njm} = 0.$$

f) Calculate $d\Omega$ using the previous formula for Ω together with $d^2 = 0$ and $d(\omega \wedge \omega) = d\omega \wedge \omega - \omega \wedge d\omega$ and re-express $d\omega$ using the same formula, to obtain the relation

$$d\Omega + \omega \wedge \Omega - \Omega \wedge \omega = 0,$$

$$[d\Omega^m_n]_{kij} + 3!\Gamma^m_{[i|\ell|} R^\ell_{|n|jk]} - 3!R^m_{\ell[ij} \Gamma^\ell_{k]n} = 0,$$

where the vertical bar delimiters exclude the index from the antisymmetrization. Now use the coordinate frame formula for $R^m_{nij;k} = R^m_{nij,k} + \dots$ and evaluate the following expression which has two more connection coefficient terms compared to the previous one

$$3R^m_{n[ij;k]} = 3R^m_{n[ij,k]} + \dots = [d\Omega^m_n]_{kij} + \dots$$

Finally note that because of the antisymmetrization which includes the lower two symmetric indices $\Gamma^k_{ij} = \Gamma^k_{ji}$ in those additional two terms, they vanish, leaving the identity

$$0 = 3R^m_{n[ij;k]} = R^m_{nij;k} + R^m_{njki} + R^m_{nki;j},$$

called Bianchi's identity of the second kind.

g) Define the symmetric Einstein tensor by

$$G^i_j = R^i_j - \frac{1}{2} R \delta^i_j,$$

where the symmetric Ricci tensor and scalar curvature are defined by

$$R_{ij} = R^k_{ikj}, \quad R = g^{ij} R_{ij} = R^i_i = R^{ij}_{ij}.$$

Then fill in the missing steps of the following equations

$$\begin{aligned} 0 &= 3R^{mn}{}_{[ij;k]}\delta^i{}_m\delta^j{}_n = \dots \\ &= -2\left(R^j{}_k - \frac{1}{2}R\delta^j{}_k\right)_{;j} = -2G^j{}_{k;j} \end{aligned}$$

This says the Einstein tensor has zero divergence. This turns out to be a fundamental property for Einstein's theory of general relativity. ■

Exercise 11.7.3. curvature of the 3-sphere

On the 3-sphere viewed as the group manifold of $SU(2)$ explored in terms of the rotation group $SO(4, \mathbb{R})$ in Exercises 4.5.6 and 6.7.5, where two mutually commuting almost orthonormal frames corresponding to linear combinations of the natural rotation group generating matrices were introduced with Lie bracket relations

$$[E_a, E_b] = C^a{}_{bc}E_c, \quad [E_a, \tilde{E}_b] = 0, \quad [\tilde{E}_a, \tilde{E}_b] = -C^a{}_{bc}\tilde{E}_c, \quad C^c{}_{ab} = \epsilon_{cab}.$$

If we introduce the corresponding dual frames we therefore get the relations

$$dW^c = -\frac{1}{2}C^c{}_{ab}W^b \wedge W^a, \quad d\tilde{W}^c = \frac{1}{2}C^c{}_{ab}\tilde{W}^b \wedge \tilde{W}^a,$$

In Exercise 4.5.7, we showed that the metric on the 3-sphere is

$$g = g_{ab}W^a \otimes W^b = g_{ab}\tilde{W}^a \otimes \tilde{W}^b, \quad g_{ab} = -\frac{1}{8}C^c{}_{ab}C^d{}_{bc} = \frac{1}{4}\delta_{ab}.$$

Show that

$$\Gamma^c{}_{ab} = \frac{1}{2}C^c{}_{ab} = -\tilde{\Gamma}^c{}_{ab},$$

so that the connection 1-form matrices in these two frames are

$$\omega^a{}_b = \frac{1}{2}C^a{}_{cb}W^c, \quad \tilde{\omega}^a{}_b = -\frac{1}{2}C^a{}_{cb}\tilde{W}^c.$$

Either use the above formula for the exterior derivative of a 1-form in a frame to evaluate the curvature 2-form or use the formula for the frame components of the curvature tensor directly (both of which require using the quadratic Jacobi identity satisfied by the structure constants: see Exercise 1.7.8) to obtain

$$\tilde{R}^a{}_{bcd} = R^a{}_{bcd} = \frac{1}{4}C^a{}_{be}C^e{}_{cd} \rightarrow \tilde{R}^{ab}{}_{cd} = R^{ab}{}_{cd} = \epsilon^{abe}\epsilon_{ecd} = \delta^{ab}{}_{cd}.$$

This corresponds to constant unit curvature for the 3-sphere just like the 2-sphere has unit curvature. This is of course invariant under the rotations of that sphere, which is why it is a bi-invariant metric on the Lie group $SU(2)$, invariant both under left and right translations of the group into itself. These remarks extend to the the 3-dimensional rotation group $SO(3, R)$ and its natural left and right frames discussed in Section 6.9 which are necessary to study the problem of the motion of a rigid body.

**Exercise 11.7.4.** **$SU(2)$ gauge derivative**

Recall the local action of $SU(2)$ on a \mathbb{C}^2 complex vector valued field $\underline{\Psi} = \Psi^\alpha \underline{e}_\alpha$ on Minkowski spacetime as a local gauge group action discussed in Exercise 6.8.8. The gauge covariant derivative

$$\nabla_i \underline{\Psi} = \partial_i \underline{\Psi} + A^a \underline{E}_a \underline{\Psi}, \quad \nabla_i \underline{e}_\alpha = A_i^c (\underline{E}_c)^\beta{}_\alpha \underline{e}_\beta = \Gamma^\beta{}_{i\alpha} \underline{e}_\beta.$$

with Lie algebra-valued 1-form matrix $\underline{A} = A^a \underline{E}_a$ is invariant under the gauge transformations

$$\underline{\Psi} \rightarrow \underline{U} \underline{\Psi} = e^{\theta^a \underline{E}_a} \underline{\Psi}, \quad \underline{A} \rightarrow \underline{U} \underline{A} \underline{U}^{-1} + \underline{U} d\underline{U}^{-1}.$$

We can introduced a corresponding curvature 2-form as a Lie algebra-valued object defined in a way similar to the case for a metric connection in two ways.

a) Evaluate

$$[\nabla_i, \nabla_j] \underline{\Psi} = \underline{F}_{ij} \underline{\Psi}.$$

where

b) Evaluate

$$F = d\underline{A} + \frac{1}{2} [\underline{A}, \wedge \underline{A}]$$

where the simultaneous Lie bracket and wedge product are indicated by the “ \wedge ” notation.
in progress...



11.8 The exterior derivative and a metric

star, sharp, flat, d and ∇

When we have a metric tensor field $g = g_{ij}dx^i \otimes dx^j$ on our space, we can use the lowering and raising maps \flat and \sharp to convert p -vector fields into p -forms and vice versa. These are inverse operations. We also have the metric duality map $*$ which converts p -vector fields and p -forms into $(n-p)$ -vector fields and $(n-p)$ -forms respectively, and then back again, although $*$ is not its own inverse since $** = (-1)^{\text{integer}}$, which means that it differs from the inverse by a sign factor which depends on p, n and the signature of the metric (the number of minus signs among the self inner products in an orthonormal frame). All of these operations may be used with the exterior derivative to make new differential operators.

First let Λ^p be the space of p -forms on our n -dimensional space and let $[\Lambda^p]^\sharp$ be the space of p -vector fields. Then the index shifting and duality maps may be represented as follows

$$\begin{array}{ccc} \Lambda^{n-p} & \xrightleftharpoons[\flat]{\sharp} & [\Lambda^{n-p}]^\sharp \\ * \downarrow \uparrow * & & * \downarrow \uparrow * \\ \Lambda^p & \xrightleftharpoons[\flat]{\sharp} & [\Lambda^p]^\sharp \end{array}$$

These operations commute, i.e., it doesn't matter if you first shift indices and then take the dual or first take the dual and then shift indices. For example, if $T \in \Lambda^p$ then the successive operations correspond to the following moves in the diagram

$$\begin{array}{ccc} *[T^\sharp] & = & [*T]^\sharp \\ \text{right then up} & & \text{up then right} \end{array}$$

This means we can just write $*T^\sharp$ without specifying the order in which these operations are done on T . If you need to be convinced here is the explicit calculation

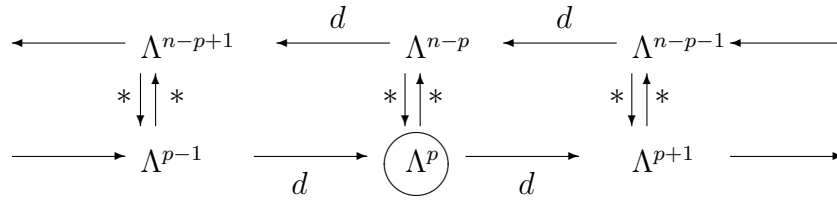
$$\begin{aligned} [T^\sharp]^{i_1 \dots i_p} &= T^{i_1 \dots i_p} \\ [*[T^\sharp]]^{i_{p+1} \dots i_n} &= \frac{1}{p!} T^{i_1 \dots i_p} \eta_{i_1 \dots i_p}^{i_{p+1} \dots i_n} = \frac{1}{p!} T_{i_1 \dots i_p} \eta^{i_1 \dots i_p i_{p+1} \dots i_n} \\ [*T]_{i_{p+1} \dots i_n} &= \frac{1}{p!} T_{i_1 \dots i_p} \eta^{i_1 \dots i_p i_{p+1} \dots i_n} \\ [[*T]^\sharp]^{i_{p+1} \dots i_n} &= \frac{1}{p!} T_{i_1 \dots i_p} \eta^{i_1 \dots i_p i_{p+1} \dots i_n} = [*[T^\sharp]]^{i_{p+1} \dots i_n}, \end{aligned}$$

so we just write

$$[*T]^\sharp = *[T^\sharp] = *T^\sharp.$$

This is just to remind ourselves of calculations done in Part I.

Now how can we mix these operations with the exterior derivative? Suppose we just look at $*$ and d alone. We could make a picture like the following, starting with a p -form and applying a succession of these two operations which change the degree of the differential form



Yuch! (or “Blech!” as the psychiatrist Lucy says when kissed by Snoopy in the Charlie Brown comic strip). Let’s forget we saw that. If we start in Λ^p we can do things like

$$\begin{aligned} {}^*d : \Lambda^p &\longrightarrow \Lambda^{n-p-1} \\ d^* : \Lambda^p &\longrightarrow \Lambda^{n-p+1} , \\ {}^*d^* : \Lambda^p &\longrightarrow \Lambda^{p-1} \end{aligned}$$

which are paths in above diagram starting at Λ^p (in that diagram I told you to forget). The last operator lowers the degree of the p -form by 1, going in the opposite direction of d :

$$\Lambda^{p-1} \xrightleftharpoons[d]{{}^*d^*} \Lambda^p .$$

We can also make second-order operators. $d^2 \equiv 0$ is of no use but

$$d^*d^* : \Lambda^p \longrightarrow \Lambda^p \quad \text{and} \quad {}^*d^*d : \Lambda^p \longrightarrow \Lambda^p$$

are two interesting second-order linear differential operators which produce p -forms from p -forms. These turn out to be related to the Laplacian (for 0-forms) and its generalization to p -forms.

By including index shifting, all of these operators can be extended to p -vector fields. First lower the indices on a p -vector field to obtain a p -form, then do various combinations of the duality operation $*$ and exterior derivative d to obtain a q -form (corresponding to moving around following the arrows in the forgotten diagram), and at the end raise the indices to go back to a q -vector field. With a little patience, we could get explicit component formulas for any of these, just by composing the component formulas for the individual operations.

One useful formula, however, re-expresses the exterior derivative of a p -form in terms of its covariant derivative. In a coordinate frame the covariant derivative is

$$\nabla_{i_1} T_{i_2 \dots i_{p+1}} = \partial_{i_1} T_{i_2 \dots i_{p+1}} - \Gamma_{i_1 i_2}^j T_{j i_3 \dots i_{p+1}} - \dots - \Gamma_{i_1 i_{p+1}}^j T_{i_1 \dots i_p j} .$$

Recall that the components of the covariant derivative are symmetric $\Gamma_{[ij]}^k = 0$ so the correction terms in the covariant derivative of a p -form go to zero under antisymmetrization

$$\nabla_{[i_1} T_{i_2 \dots i_{p+1}]} = \partial_{[i_1} T_{i_2 \dots i_{p+1}]} - \Gamma_{[i_1 i_2}^j T_{j] i_3 \dots i_{p+1}} - \dots - \Gamma_{[i_1 i_{p+1}}^j T_{i_1 \dots i_p j]} = \partial_{[i_1} T_{i_2 \dots i_{p+1}]} ,$$

recalling that the notation $\Gamma^j_{[i_1 i_2 T]_j | i_3 \dots i_{p+1}}$ means that the index j is excluded from antisymmetrization in each term. This result tells us that in the coordinate component formula for the exterior derivative, the ordinary derivative can be replaced by the covariant derivative

$$[dT]_{i_1 \dots i_{p+1}} = (p+1) \partial_{[i_1} T_{i_2 \dots i_{p+1}]} = (p+1) \nabla_{[i_1} T_{i_2 \dots i_{p+1}]},$$

but since dT and ∇T are frame-independent objects, this is true in any frame, i.e.,

$$[dT]_{i_1 \dots i_{p+1}} = (p+1) \nabla_{[i_1} T_{i_2 \dots i_{p+1}]}.$$

In a coordinate frame this can be rewritten as

$$[dT]_{i_1 \dots i_{p+1}} = (p+1) T_{[i_2 \dots i_{p+1}, i_1]} = (p+1) T_{[i_2 \dots i_{p+1}; i_1]},$$

which is a “comma to semicolon rule” valid for the exterior derivative and the metric connection in a coordinate frame.

Exercise 11.8.1.

tensor-valued differential forms

a) For a symmetric connection in a coordinate frame the exterior derivative of a p -form has components which are the antisymmetrized partial derivative of the components, or equivalently the antisymmetrized covariant derivative of those components due to the symmetry of the covariant index pair of the connection coefficients. If we have a tensor with a subset of antisymmetric indices, we can think of that tensor as a tensor-valued differential form and extend the exterior derivative to it by defining the covariant exterior derivative as the antisymmetrized covariant derivative on the antisymmetric set of indices, which produces the ordinary antisymmetrized derivative on the antisymmetric set plus all the connection coefficient terms, but those associated with the antisymmetric set of indices cancel out leaving only the extra gamma terms associated with the extra indices. Take the following example

$$\begin{aligned} (D\Omega^i_j)_{mnp} &= (p+1) \nabla_{[p} R^i_{j]mn} = (p+1) R^i_{j[mn;p]} \\ &= (p+1) (\partial_{[p} R^i_{j]mn} + \Gamma^i_{[p|k} R^k_{j]mn} - \Gamma^k_{[p|j} R^i_{k]mn} + \dots) \\ &= (d\Omega^i_j)_{pmn} + (\omega^i_k \wedge \Omega^k_j)_{pmn} - (\omega^k_j \wedge \Omega^i_k)_{pmn}. \end{aligned}$$

Check that the omitted terms “...” cancel out. Since $\underline{\Omega}$ is a 2-form, we can change the order of the wedge factors and obtain the matrix relation

$$D\underline{\Omega} = d\underline{\Omega} + \underline{\omega} \wedge \underline{\Omega} - \underline{\Omega} \wedge \underline{\omega} = 0,$$

which is zero because of the Bianchi identity of the second kind.

b) Thinking of the identity tensor as a vector-valued 1-form $Id = e_i \otimes \omega^i$ with vector component ω^i , show that Cartan’s first structural equation takes the simple form

$$\Theta^i = D\omega^i = 0,$$

while Bianchi’s first identity becomes

$$D\Theta^i = \Omega^i_j \wedge \omega^j = 0.$$

■

grad, div and curl

Returning to the previous discussion, recall that for a function f we already introduced the gradient as

$$\begin{aligned}\text{grad } f &= (df)^\sharp \equiv \vec{\nabla} f, \\ [\text{grad } f]^i &= g^{ij} \partial_j f = g^{ij} \nabla_j f \equiv \nabla^i f.\end{aligned}$$

Suppose we start with a vector field X . Then the result of the following sequence of operations

$$\begin{array}{c} *d \quad * \underbrace{X^\flat}_{\text{1-form}} \\ \underbrace{\quad}_{(n-1)\text{-form}} \\ \underbrace{\quad}_{n\text{-form}} \\ \underbrace{\quad}_{\text{function}} \end{array}$$

is a function. What is its component formula? The dual of a 1-form is an $(n-1)$ -form

$$[*X^\flat]_{i_2 \dots i_n} = X_i \eta^i_{i_2 \dots i_n} = X^i \eta_{ii_2 \dots i_n}$$

and its exterior derivative is an n -form

$$[d*X^\flat]_{i_1 i_2 \dots i_n} = n \nabla_{[i_1} (\eta_{|i_2 \dots i_n]} X^{i_1}) = n (\nabla_{[i_1} X^{i_1}) \eta_{|i_2 \dots i_n]},$$

where η can be factored out of the derivative since it is covariant constant, and the dual of this expression is a function

$$\begin{aligned} *d*X^\flat &= \frac{1}{n!} [n (\nabla_{[i_1} X^{i_1}) \eta_{|i_2 \dots i_n]}] \eta^{i_1 \dots i_n} \\ &= \frac{1}{(n-1)!} \underbrace{\eta_{ii_2 \dots i_n} \eta^{i_1 i_2 \dots i_n}}_{(-1)^M \delta^{i_1}_i} \nabla_{i_1} X^i = (-1)^M \underbrace{\nabla_i X^i}_{\text{div } X}, \end{aligned}$$

where M is the number of negative signs among the diagonal components of the metric in an orthonormal frame: $(-1)^M = \text{sgn det}(g_{ij})$. Thus we get the divergence of the vector field, apart from a possible sign when the metric has negative self-inner product values. For \mathbb{R}^n with the Euclidean metric, the sign is $+1$, so one gets exactly the divergence.

Remark.

We can get rid of the sign $(-1)^M$ by using the inverse dual map at the end instead of the dual map. Consider the following short derivation. For a p -form S with $p = n$, one has $*^{-1}S = (-1)^{M+n(p-1)} *S = (-1)^M *S$, so

$$*^{-1}d*X^\flat = (-1)^M *d*X^\flat = \text{div } X \equiv -\delta X^\flat$$

defines the divergence of a vector field, or equivalently the codifferential δX^\flat of a 1-form X^\flat . This generalizes naturally to any p -form as we will see shortly.



Suppose $n = 3$. Consider the operator

$$[*dX^\flat]^\sharp$$

for a vector field X . The result of these three operations is a vector field whose components are easily calculated

$$\begin{aligned} [dX^\flat]_{ij} &= \partial_i X_j - \partial_j X_i \\ [*dX^\flat]_k &= \frac{1}{2}(\partial_i X_j - \partial_j X_i)\eta^{ij}{}_k \\ [*dX^\flat]^k &= \frac{1}{2}(\partial_i X_j - \partial_j X_i)\eta^{ijk} = \partial_{[i} X_{j]} \eta^{ijk} = \partial_i X_j \eta^{ijk} = \eta^{kij} \partial_i X_j. \end{aligned}$$

In Cartesian coordinates on \mathbb{R}^3 , this has the expression

$$[*dX^\flat]^k = \epsilon^{kij} \frac{\partial X_j}{\partial x^i}$$

which is the expression for the curl of the vector field X . Since $[*dX^\flat]^\sharp$ is a vector field independent of the choice of coordinates, this is true independent of the coordinates

$$[*dX^\flat]^\sharp = \text{curl } X.$$

In calculus we learned that certain second order derivative combinations of grad, curl and div on \mathbb{R}^3 vanish identically. These are just consequences of the fact that $d^2T \equiv 0$ for p -forms T with $0 \leq p \leq 3$. For example

$$\begin{aligned} (p=0) \quad \text{curl grad } f &= [*d(\underbrace{\text{grad } f}^{df})]^\sharp = [*d^2f]^\sharp = 0, \\ (p=1) \quad \text{div curl } X &= [*d(\underbrace{\text{curl } X}^{*dX^\flat})]^\sharp = [*d^{**}dX^\flat]^\sharp = [*d^2X^\flat]^\sharp = 0, \end{aligned}$$

since $^{**}T = T$ for any p -form T for the Euclidean metric on \mathbb{R}^3 .

We can also consider repeating $*d$ or d^* twice

$$\begin{aligned} \text{curl curl } X &= [*d(\underbrace{\text{curl } X}^{*dX^\flat})]^\sharp = [*d^*dX^\flat]^\sharp, \\ \text{grad div } X &= [d(\text{div } X)]^\sharp = [d^*d^*X^\flat]^\sharp, \\ \text{div grad } f &= [*d^*\text{grad } f]^\flat = [*d^*df]. \end{aligned}$$

The remaining combination $d^*d^*f = 0$ is identically zero since $*f$ is a 3-form and its exterior derivative is identically zero.

While we're at it, what about the “del” notation $\vec{\nabla}f$, $\vec{\nabla} \cdot \vec{X}$, $\vec{\nabla} \times \vec{X}$? Well, on \mathbb{R}^3 we can define the cross product of two vector fields by

$$X \times Y = *(X \wedge Y).$$

The component formula is

$$\begin{aligned} [X \times Y]^i &= \frac{1}{2}[X \wedge Y]^{jk}\eta_{jk}{}^i && \text{definition of } \times \\ &= \frac{1}{2}(2X_{[i}Y_{j]})\eta_{jk}{}^i && \text{definition of wedge} \\ &= \eta^i{}_{jk}X^{[j}Y^{k]} && \text{shifting indices} \\ &= \eta^{ijk}X_jY_k && \text{drop brackets,} \end{aligned}$$

where in the final line the explicit antisymmetrization is redundant since η is antisymmetric, so only the antisymmetric part contributes anyway. In the Cartesian coordinate frame this is just the usual formula

$$[X \times Y]^i = \epsilon_{ijk}X^jY^k.$$

We've already defined $\vec{\nabla}$ as the covariant derivative operator with the derivative index raised, so $\text{grad } f = \vec{\nabla}f$ and

$$[\text{curl } X]^i = \eta^{ijk}\partial_jX_k = \eta^{ijk}\nabla_jX_k = [\nabla \times X]^i$$

while

$$\text{div } X = \nabla_iX^i = g_{ij}\nabla^iX^j = \vec{\nabla} \cdot X$$

so

$$\text{div curl } X = \vec{\nabla} \cdot (\vec{\nabla} \times X),$$

etc. These may be easily evaluated in any coordinate system now.

What about the various product rules for grad, curl, div? Most of them are disguised versions of the product rule for d

$$d(T \wedge S) = dT \wedge S + (-1)^pT \wedge dS \quad 0 \leq p, q \leq n.$$

where T is a p -form and S is a q -form.

Because of the antisymmetry condition $T \wedge S = (-1)^{pq}S \wedge T$, it is enough to look at the cases $p \leq q$, but also $p + q < 3$ since the exterior derivative of a 3-form is identically zero and $(p + q)$ -forms are zero for $p + q > 3$. This leaves

$$(p, q) \in \{(0, 0), (0, 1), (0, 2), (1, 1)\}$$

Exercise 11.8.2. grad curl div

Using the definitions of grad, curl, div in terms of d , re-express the left hand sides of the following identities and use the above product rule for d with the given values of (p, q) to rewrite them in terms of grad, div, curl (recall $f \wedge T = f T$ for zero-form f):

$$\begin{aligned} (0, 0) : \quad \text{grad}(fh) &= h \text{grad } f + f \text{grad } h, \\ (0, 1) : \quad \text{curl}(fX) &= (\text{grad } f) \times X + f \text{curl } X, \\ (0, 2) : \quad \text{div}(fX) &= (\text{grad } f) \cdot X + f \text{div } X, \\ (1, 1) : \quad \text{div}(X \times Y) &= Y \cdot (\text{curl } X) - X \cdot (\text{curl } Y). \end{aligned}$$

■

Example 11.8.1. The first two derivations of the previous Exercise are completely straight forward but the last two are a bit challenging since they need the unfamiliar identity $*(T \wedge *S) = \langle T, S \rangle$ for two p -forms, which follows from $T \wedge *S = \langle T, S \rangle \eta$ and $*\eta = 1$. Thus

$$\begin{aligned} \text{div } fX &= *d^*(fX^\flat) = *d(f^*X^\flat) = *[df \wedge *X^\flat + (-1)^0 f d^*X^\flat] \\ &= \underbrace{*[df \wedge *X^\flat]}_{\langle df, X^* \rangle} + f^*d^*X^\flat = (\text{grad } f) \cdot X + f \text{div } X, \end{aligned}$$

and finally

$$\begin{aligned} \text{div}(X \times Y) &= *d^*[(X \wedge Y)^\flat] = *d^*[(X^\flat \wedge Y^\flat)] \\ &= *d(X^\flat \wedge Y^\flat) = *[dX^\flat \wedge Y^\flat - X^\flat \wedge dY^\flat] \\ &= *[Y^\flat \wedge dX^\flat] - *[X^\flat \wedge dY^\flat] \\ &= *[Y^\flat \wedge *(dX^\flat)] - *[X^\flat \wedge *(dY^\flat)] \\ &= \langle Y^\flat, (\text{curl } X)^\flat \rangle - \langle Y^\flat, (\text{curl } X)^\flat \rangle \\ &= Y \cdot (\text{curl } X) - X \cdot (\text{curl } Y). \end{aligned}$$

□

Notice that these “vector analysis” identities which are usually provided by Cartesian coordinate component calculations like

$$\begin{aligned} \text{div}(X \times Y) &= \partial_i(\epsilon^{ijk} X_j Y_k) = \epsilon^{ijk}[(\partial_i X_j) Y_k + X_j \partial_i Y_k] \\ &= (\epsilon^{ijk} \partial_i X_j) Y_k - (\epsilon^{ijk} \partial_i Y_k) X_j = (\text{curl } X) \cdot Y - (\text{curl } Y) \cdot X \end{aligned}$$

have just been proven for *any* positive-definite metric on a 3-dimensional space in any coordinate system (since they are independent of the coordinates). Thus we can extend all of this \mathbb{R}^3 vector analysis immediately to the 3-sphere, for example.

This is the power of real mathematics as opposed to “just getting by” techniques that are usually used in applied sciences. Not impressed? Maxwell’s equations for the electromagnetic field brings you cable and satellite TV, your cell phone calls, wireless internet, computers, your

favorite radio station and all the rest of our modern life. They involve the electric and magnetic vector fields E and B and the charge density function ρ and current density vector field J

$$\begin{aligned}\operatorname{div} B &= 0, & \operatorname{div} E &= 4\pi\rho, \\ \operatorname{curl} E + \frac{\partial B}{\partial t} &= 0, & \operatorname{curl} B - \frac{\partial E}{\partial t} &= 4\pi J,\end{aligned}$$

and can be written in the simple form

$$dF = 0, \quad {}^*d^*F = 4\pi\mathcal{J}$$

by defining the electromagnetic 2-form F and 4-current density 1-form \mathcal{J} on spacetime

$$\begin{aligned}F &= (E_x dx + E_y dy + E_z dz) \wedge dt + B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy, \\ \mathcal{J} &= -\rho dt + J_x dx + J_y dy + J_z dz.\end{aligned}$$

Many of the somewhat complicated manipulations done in physics courses become very simple in this language. We don't have time to go into too much of that here, but I wanted you to get a glimpse of this idea.

Exercise 11.8.3.

Maxwell's equations in differential form

Evaluate the differential form version of Maxwell's equations and show that they are equivalent to the vector form. ■

Exercise 11.8.4.

vector potential for electromagnetic field

The Maxwell equation $dF = 0$ for the 2-form F means that at least locally it admits a "vector potential", more precisely, a 1-form potential A such that $F = dA$ so that $dF = d^2A = 0$ automatically satisfies half of Maxwell's equations. Let $A = -\phi dx^0 + A_i dx^i$ be the potential 1-form so that the corresponding vector field is $A^\sharp = \phi \partial_0 + \vec{A}$, defining the usual scalar potential ϕ and vector potential \vec{A} .

a) Evaluate $F = dA$ and compare to the electric and magnetic fields in the previous Exercise to obtain the classic relations using index notation

$$E^i = -\partial_i \phi + \partial_0 A^i = [-\operatorname{grad} \phi + \partial_0 \vec{A}]^i, \quad B^i = [\operatorname{curl} \vec{A}]^i.$$

b) Adding the differential of any function to the vector potential (1-form!), i.e., adding the spacetime gradient of any function to the vector potential vector field, leads to the same F since $d(A + d\Lambda) = dA + d^2\Lambda = dA = F$. This is called a gauge transformation of the vector potential, under which the electric and magnetic fields do not change. Thus any physical quantities should be invariant under such a gauge transformation. Show how the pair (ϕ, \vec{A}) change under such a gauge transformation. ■

the codifferential δ

Apart from an annoying sign, the operation $T \rightarrow *d^*T$ which takes a p -form T to a $(p-1)$ -form defines the “codifferential” δT . The codifferential is defined as the adjoint operator with respect to the inner product of p -forms, itself defined in terms of the duality operation

$$\langle S, T \rangle \eta = S \wedge *T = T \wedge *S.$$

Thoroughly discussed in Section 4.3, this is just the natural inner product for any tensors of a given rank scaled down by a factorial factor to avoid overcounting

$$\langle S, T \rangle = \frac{1}{p!} S^{i_1 \dots i_p} T_{i_1 \dots i_p}.$$

For a real vector space V with an inner product and a linear transformation L of the space into itself, the adjoint linear transformation L^\dagger is defined by $\langle X, LY \rangle = \langle L^\dagger X, Y \rangle$. This was explored in Exercise 4.5.13. For differential forms this will be true modulo a differential which under the integral sign can be made to vanish with appropriate boundary conditions on the region of integration, corresponding to “integration by parts” in common language.

Let S be a p -form, T a p -form, then $S \wedge *T$ is an n -form and their inner product is defined by

$$S \wedge *T = \langle S, T \rangle \eta = T \wedge *S,$$

or if R is an $(n-p)$ -form then letting $T = *^{-1}R$ so that $*T = R$, one has

$$S \wedge R = \langle S, *^{-1}R \rangle \eta.$$

Since $**$ applied to any p -form S is the identity plus a possible sign change, the inverse dual $*^{-1}$ is just $*$ multiplied by a sign factor; recall from Exercise 4.3.11 that $*^{-1}S = (-1)^{M+p(n-p)}S$.

Next let α be a $(p-1)$ -form, β a p -form, then $d\alpha$ and β are both p -forms and by the previous definition of their inner product one has

$$d\alpha \wedge *\beta = \langle d\alpha, \beta \rangle \eta = \beta \wedge *d\alpha.$$

On the other hand the wedge product $\alpha \wedge *\beta$ is an $(n-1)$ -form, so its exterior derivative is an n -form. Using the product rule for the exterior derivative, and then rewriting each of the two terms in terms of the inner product leads to

$$\begin{aligned} d(\alpha \wedge *\beta) &= d\alpha \wedge *\beta + (-1)^{p-1} \alpha \wedge d*\beta && \text{(product rule for } d) \\ &= \langle d\alpha, \beta \rangle \eta + \langle \alpha, \underbrace{(-1)^{p-1} *^{-1} d*\beta}_{\equiv -\delta\beta} \rangle \eta && \text{(convert to inner product notation)} \end{aligned}$$

leads to the definition of a $(p-1)$ -form $\delta\beta$ satisfying the identity

$$\langle d\alpha, \beta \rangle \eta = d(\alpha \wedge *\beta) + \langle \alpha, \delta\beta \rangle \eta.$$

δ is called the codifferential

$$\delta\beta = (-1)^{p *^{-1}} d*\beta. \quad (\beta \text{ a } p\text{-form})$$

Exercise 11.8.5.**codifferential versus divergence sign**

a) Note that $\delta f = 0$ for a 0-form in the same way that $d\beta = 0$ for an n -form. Show that $\delta^2 = 0$ in the same way that $d^2 = 0$.

b) Determine the annoying sign in the first definition of the codifferential when expressed only in terms of the dual and not its inverse by using the sign formula for \ast^{-1}

$$\delta\beta = (-1)^{M+n(p-1)+1} \ast d\ast\beta \quad (\beta \text{ a } p\text{-form}).$$

Note that for a 1-form X^\flat with $p = 1$ this becomes $\delta X^\flat = -(-1)^M \ast d\ast X^\flat$.

b) The metric divergence of a p -vector field X is defined by

$$\begin{aligned} \operatorname{div} X &\equiv (\operatorname{div} X^\flat)^\sharp = \frac{1}{(p-1)!} g^{-1/2} (\partial_{i_1} (g^{1/2} X^{i_1 \dots i_p})) \frac{\partial}{\partial x^{i_2}} \wedge \dots \wedge \frac{\partial}{\partial x^{i_p}} \\ &= \frac{1}{(p-1)!} X^{i_1 \dots i_p}{}_{;i_1} \frac{\partial}{\partial x^{i_2}} \wedge \dots \wedge \frac{\partial}{\partial x^{i_p}}. \end{aligned}$$

Use the coordinate formulas for the duality operations from Section 4.3 to show that this definition is related by a sign change to the codifferential

$$\delta X = -\operatorname{div} X.$$

c) The metric Laplacian may be defined for any tensor by $\nabla^2 T = \nabla^i \nabla_i T$. For a p -form the deRham Laplacian is defined by $\Delta_{\text{deR}} T = (d\delta + \delta d)T$. Show that for a scalar field or function f (a 0-form) one has

$$\Delta_{\text{deR}} f = \delta df = -\nabla^2 f.$$

d) For a tensor field, the deRham Laplacian has additional curvature terms compared to the “ordinary Laplacian” ∇^2 . For example, consider a vector field A^\sharp like the index-raised vector potential (1-form really) for the electromagnetic field tensor. The definition of the curvature tensor evaluated on coordinate frame vector fields whose commutator vanishes is

$$\left(\nabla_{\partial_i} \nabla_{\partial_j} - \nabla_{\partial_j} \nabla_{\partial_i} - \nabla_{[\partial_i, \partial_j]} \right) A^i = R^i{}_{jmn} A^m$$

which if we simplify becomes

$$A^k{}_{;ji} - A^k{}_{;ij} = R^k{}_{mij} A^m.$$

Contract this with δ^i_k to obtain the so-called Ricci identity

$$A^i{}_{;ji} - A^i{}_{;ij} = R_{jm} A^m.$$

Then consider the Maxwell equation with $dA = F$, i.e., $F_{ji} = A_{i;j} - A_{j;i}$ if we use the covariant derivative formula for the exterior derivative

$$4\pi \mathcal{J}_j = F_j{}^i{}_{;i} = A^i{}_{;ji} - A_j{}^i{}_{;i},$$

and replace the first term using the Ricci identity to obtain

$$4\pi\mathcal{J}_j = A^i{}_{;ij} + R_{jm}A^m - A_j{}^{;i}{}_{;i} = \underbrace{-\nabla^2 A_j + A_i R^i{}_j}_{\Delta_{\text{deR}}A} + \underbrace{(A^i{}_{;i})_{;j}}_{-d\delta A}.$$

It turns out that the first two terms are the deRham Laplacian, but in flat spacetime this agrees with the ordinary Laplacian, and the second term is the gradient of the divergence of the vector potential. Often this divergence is chosen to be zero to fix the freedom in the choice of vector potential while simplifying this to a normal wave equation for that field.

On the other hand sticking with the powerful index-free notation, we have by definition

$$4\pi\mathcal{J} = \delta dA = -d\delta A + \Delta_{\text{deR}}A,$$

so comparing with the previous equation confirms that its first two terms are the deRham Laplacian. ■

Exercise 11.8.6.

Maxwell's equations and the codifferential

a) Now that we have a name for this combination operation, show that half of Maxwell's equations are

$$-\delta F = 4\pi\mathcal{J}^\flat,$$

which implies $\delta\mathcal{J}^\flat = 0$. Write out this “conservation law” for charge density ρ and current density J in terms of the decomposition of the spacetime 1-form $\mathcal{J}^\flat = \rho dt + J_a dx^a$ for $a = 1, 2, 3$.

b) Since $d^2 = 0$, if we apply the exterior derivative to this Maxwell equation and re-express in terms of the deRham Laplacian $\Delta_{\text{deR}} = d\delta + \delta d$, we obtain the wave equation for the electromagnetic field tensor

$$-\Delta_{\text{deR}}F = -d\delta F = 4\pi d\mathcal{J}^\flat,$$

so that in vacuum ($\mathcal{J}^\flat = 0$), we get the source-free wave equation in Minkowski spacetime

$$\left(-\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)F_{ij} = 0.$$

c) If instead we insert $F = dA$ into the previous Maxwell equation and re-express it in terms of the deRham Laplacian, show that we get the vector potential wave equation

$$-\Delta_{\text{deR}}A + d\delta A = 4\pi d\mathcal{J}^\flat.$$

Check signs??

d) If we impose the “gauge condition” $\delta A = 0$ that the vector potential have zero divergence (called the Lorentz gauge), we get a wave equation for the vector potential 1-form whose source is the spacetime current density. This can be accomplished starting from any initial vector

potential A by doing a gauge transformation to a new vector potential $A + d\Lambda$ which should satisfy this condition

$$0 = \delta(A + d\Lambda) = \delta A + \Delta_{\text{deR}}\Lambda = \delta A + \nabla^2\Lambda,$$

which amounts to solving the Laplace equation for Λ . ■

Commutative diagrams?

Perhaps we were too quick to dismiss that “commutative diagram” earlier in this chapter illustrating the affect of combinations of the exterior derivative d and duality operation $*$ on differential forms. Consider the case $n = 3$ of \mathbb{R}^3 with the Euclidean metric. We only need functions and vector fields as our fundamental fields since the remaining nonzero p -forms and p -vector fields for $p = 2, 3$ can be represented in terms of these by combinations of index raising and duality operations. Thus our entry points into the diagram are 0-forms (the space Λ^0 of functions) and 1-vectors (the space $\Lambda^{1\sharp}$ of vector fields), shown circled in Fig. 11.15.

Say we start at Λ^0 or $\Lambda^{1\sharp}$ on the lower left of the diagram. Starting with $f \in \Lambda^0$ and moving one step right, then down, we get $\text{grad } f = (df)^\sharp$. Moving two steps right, we get $d^2f = 0$. Starting with $X \in \Lambda^{1\sharp}$ and moving up to Λ^1 , then moving two steps to the right we get $d^2X^\flat = 0$. Moving one step to the right, then two steps up, we get the vector field $\text{curl } X = (*dX^\flat)^\sharp$. Instead moving one step up from Λ^1 , one step left, then one step down, we get the function $\text{div } X = *d*X^\flat$. For second order operators, we need to keep on going further. Starting with $f \in \Lambda^0$, move over one, up one, left one, down one to get $\text{div grad } f = *d*df$, just a counterclockwise loop. Similarly starting with $X \in \Lambda^{1\sharp}$, doing one counterclockwise loop from Λ^1 and then back down to a vector field yields $\text{curl curl } X = *d*dX^\flat$. Thus while totally unnecessary for using differential forms and its interaction with a metric through duality and index shifting, it is kind of pretty in providing an underlying scheme into which the operations of grad, div and curl all fit nicely with well-defined rules for how they work together in succession.

Furthermore, we can consider the image and null space of the exterior derivative operator as a linear operator among these infinite-dimensional spaces of p -forms. A p -form in the image of d is representable as the exterior derivative of a $(p - 1)$ -form is called exact, while a p -form in the null space of d has vanishing exterior derivative and is called closed. An exact form is closed, but the reverse statement depends on the topology of the region over which the question is considered. The image space of $d : \Lambda^0 \rightarrow \Lambda^1$ consists of 1-forms whose corresponding vector fields (conservative vector fields) are representable as the gradient of a potential function, while the image of $d : \Lambda^1 \rightarrow \Lambda^2$ consists of the dual of 1-forms whose associated vector fields can be represented as the curl of another vector field, called the vector potential. The null space of $d : \Lambda^0 \rightarrow \Lambda^1$ consists of functions f for which $df = 0$, i.e., constant functions. The null space of $d : \Lambda^1 \rightarrow \Lambda^2$ consists of the dual of 1-forms (called “closed”) whose associated vector fields have zero curl. Having zero curl is a necessary condition for admitting a scalar potential, but it is not sufficient. If the latter null space is larger than the image space of the previous operator, which that null space necessarily contains because of the identity $d^2f = 0$, or $\text{curl grad } f = 0$, then it

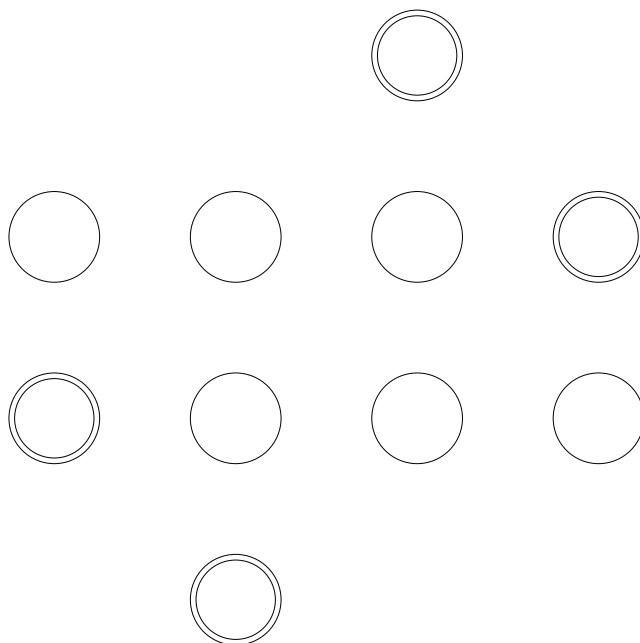


Figure 11.15: The d and $*$ operations on 0-forms (functions) and 1-forms (index-lowered vector fields) illustrated as a sequence of linear maps. Two successive d operations lead to zero, so one must insert a $*$ in between them to get a second order operator which is nonzero. The codifferential $\delta \sim *d*$ corresponds to a trip around a square path in the opposite direction from d . Their compositions $d\delta$ and δd return to the initial starting point, going around a square in the opposite directions. The deRham Laplacian is defined by $\Delta = d\delta + \delta d$ and maps p -forms to p -forms.

is not true that every curlfree vector field is a conservative vector field. This leads to topological questions about the space on which we are working, and requires global considerations that are a bit more sophisticated than the local calculations we are doing departing from the flat \mathbb{R}^n spaces. This area of mathematics falls under the keywords “de Rham cohomology.” One calls a differential form exact if it is the exterior derivative of another form, and closed if its own exterior derivative is zero. The identity $d^2 = 0$ means that every exact form is closed, but closed forms do not have to be exact. The determining factor turns out to be the topology of the space we are working on. For example, planes, spheres, torii, etc. are examples of topologically distinct 2-dimensional spaces with metrics, and their topological properties as sets of points are those properties which do not depend on their actual shape but are the same for any deformation of the space as long as points remain distinct, roughly speaking. While interesting, we are not in a position to travel in that direction from where we are now.

Exercise 11.8.7.

spacetime deRham cohomology

If we are interested in repeating this diagram stuff for Minkowski spacetime, namely \mathbb{R}^4

with the Lorentzian metric on it, we need only need $p = 0, 1, 2$ since the remaining $p = 3, 4$ can be represented as the duals of 1-forms and 2-forms. We would thus need three entry points, for functions f , vector fields X and 2-vector fields F , or we could make life simpler by just considering p -forms and not p -vectors in the diagram, limiting it to two rows by eliminating the index shifting. Repeat the above diagram for these two rows, adding on one more dimension as appropriate for \mathbb{R}^4 .

■

11.9 Induced orientation on a boundary

In multivariable calculus we see Green's theorem in the plane, which can be written either in terms of the integral of the third component of a curl or a divergence of a vector field in the plane, and its generalizations to Stokes' theorem and Gauss's law in space involving integrals of vector fields and scalars over curves, surfaces and open regions of space. All of these integrals require an orientation for the subspace over which we perform the integration: a consistent direction for the curve in a line integral and a choice of normal direction for a surface and finally the default right hand rule orientation for any parametrization of an open region of space. We need to quantify this idea of orientation in order to see how all of these activities fit into the single concept of integrating p -forms over p -surfaces and a single generalized theorem involving the exterior derivative (called Stokes' theorem) that describes all the vector theorems at once.

First coordinate adapted to boundary

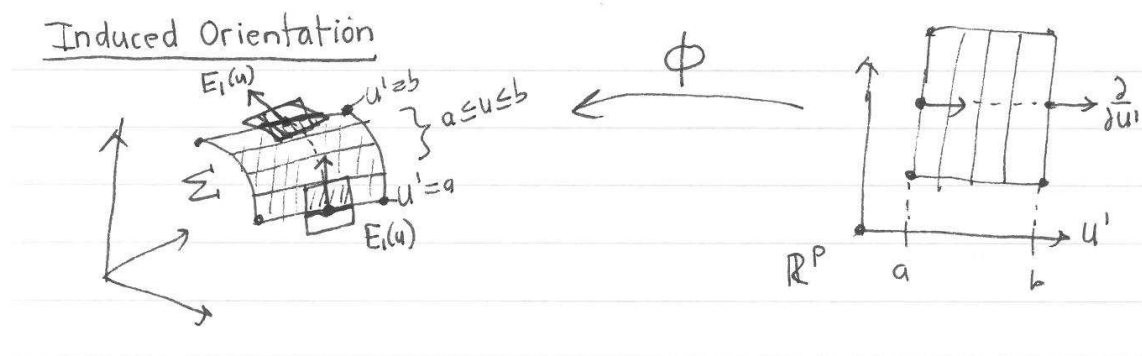


Figure 11.16: A parametrized surface with boundary and the induced orientation of the boundary.

Suppose we have a parametrized p -surface Σ in \mathbb{R}^n with a boundary

$$x^i = x^i(u^1, \dots, u^p), \quad E^i_A(u) \equiv \frac{\partial x^i}{\partial u^A}(u), \quad (A, B, \dots = 1, \dots, p),$$

part of which corresponds to constant values of the first parameter u^1 so that $a \leq u^1 \leq b$ holds on Σ itself, as illustrated in Fig. 11.16. Other parts of the boundary might correspond to a parallel discussion for other coordinates, but we choose u^1 for this derivation here for the sake of concreteness and because its tangent vector has to be first in the following wedge product to make signs come out right later. We are doing the preparation work for generalizing Green's theorem in the plane to Stokes' theorem in higher dimensional scenarios.

The p -vector $E_1(u) \wedge \dots \wedge E_p(u)$ determines the inner orientation of Σ at each point, said to be positively oriented. At the boundary, to be denoted by $\partial\Sigma$, half of the tangent p -plane

to Σ will hang off the p -surface — in fact the tangent $(p-1)$ -plane to $\partial\Sigma$ at these boundary points will cut the tangent p -plane to Σ into two halves. Half of the nonzero vectors will point inward towards interior points of Σ , while half will point outward, except for those vectors in the tangent $(p-1)$ plane subspace which are tangent to $\partial\Sigma$.

Suppose $u^1 \leq b$ describes those points of Σ near the boundary $\partial\Sigma$ located at $u^1 = b$, so that $E_1(u)$ points outward at $u^1 = b$, then the remaining parameters $\{u^2, \dots, u^p\}$ give a parametrization of $\partial\Sigma$ whose associated orientation, namely that of $E_2(u) \wedge \dots \wedge E_p(u)$, is called the induced orientation of $\partial\Sigma$, determined by the orientation of Σ (namely $E_1(u) \wedge \dots \wedge E_p(u)$). If instead $a \leq u^1$ describes points near the boundary at $u^1 = a$, so that $E_1(u)$ points inward at $u^1 = a$, then $\{u^2, \dots, u^p\}$ give an orientation for $\partial\Sigma$ (namely $E_2(u) \wedge \dots \wedge E_p(u)$) which is opposite to the induced orientation.

Another way of stating this is that if $\{E_\alpha\}_{\alpha=1, \dots, p}$ is any set of vector fields which provide a positively oriented basis for the tangent p -planes to Σ such that on $\partial\Sigma$, E_1 points outward while $E_2 \wedge \dots \wedge E_p$ describes the $(p-1)$ -dimensional subspace of the tangent space tangent to $\partial\Sigma$, then $E_2 \wedge \dots \wedge E_p$ is positively oriented with respect to the induced orientation of $\partial\Sigma$. Revisiting the previous paragraph in this light, in the above parametrization definition, $-E_1(u)$ points outward when $a \leq u^1$ describes the boundary, so $-E_2(u) \wedge \dots \wedge E_p(u)$ orients the boundary

$$\underbrace{[-E_1(u)]}_{\text{outer}} \wedge \underbrace{[-E_2(u) \wedge \dots \wedge E_p(u)]}_{\text{induced orientation for } \partial\Sigma} = \underbrace{E_1(u) \wedge \dots \wedge E_p(u)}_{\text{orientation for } \Sigma}.$$

On the other hand, when $u^1 \leq b$ describes the boundary, then $E_2(u) \wedge \dots \wedge E_p(u)$ orients the boundary

$$\underbrace{[E_1(u)]}_{\text{outer}} \wedge \underbrace{[E_2(u) \wedge \dots \wedge E_p(u)]}_{\text{induced orientation for } \partial\Sigma} = \underbrace{E_1(u) \wedge \dots \wedge E_p(u)}_{\text{orientation for } \Sigma}.$$

As illustrated in Fig. 11.17, we can even extend this to the case $p = 1$ of a curve segment Σ with its two 0-dimensional endpoints $\partial\Sigma$ on which a 0-vector (function) orientation can be induced

$$\begin{aligned} \partial\Sigma_+ : [E_1(u)] \wedge [+1] &= E_1(u), \\ \partial\Sigma_- : [-E_1(u)] \wedge [-1] &= E_1(u). \end{aligned}$$

This assigns a plus sign to the terminal point and a minus sign to the initial point of the directed curve segment. Note that in the other extreme when $p = n$, corresponding to an open set of our space \mathbb{R}^n , one can always use the orientation of the whole space on Σ for its orientation.

Figure 11.18 illustrates the situation for p -surfaces in \mathbb{R}^2 and \mathbb{R}^3 . Note that for the case $p = 2$ in \mathbb{R}^2 , for a region with a hole in it, the outer orientation of the boundary is counterclockwise, but the inner orientation is clockwise, which should be familiar from Green's theorem in the plane. Indeed the induced orientation can be represented by a directed loop symbolizing the rotation from the first to the second vector of an oriented frame, and the orientation is the direction in which such a loop flows if you bring it in contact with the boundary. The usual counterclockwise orientation of the plane for Cartesian coordinates $(x^1, x^2) = (x, y)$ is needed for Green's theorem.

Note that for the cases $n = 3$ in \mathbb{R}^3 , and $p = 2$ or $p = 3$, we can also describe the inner orientation of a surface or bounding surface respectively by a choice of any vector pointing

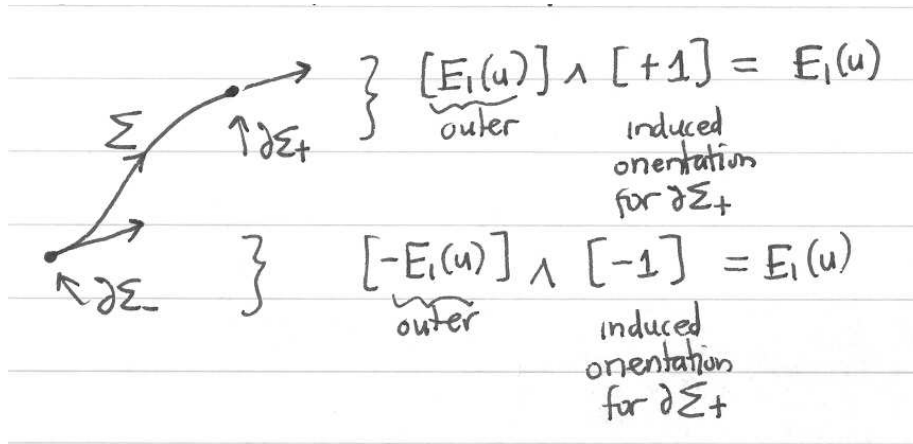


Figure 11.17: A parametrized curve with boundary points and their induced orientation.

out of the surface (i.e., not belonging to the subspace of the tangent space that is tangent to the surface) picking out one side or the other of the tangent plane to the surface and linking it to the inner orientation by the right hand rule. This is called an outer orientation for the surface, and is the way we were introduced to the orientation of ordinary surfaces in space in multivariable calculus needed for Stokes' theorem and Gauss's law, both generalizations of Green's theorem to three dimensions. For a closed surface, an outward normal picks out the inner orientation needed for the latter law. We will see how all these fit together in the general context soon.

Consider the region Σ in \mathbb{R}^3 between two spherical coordinate spheres: $r_1 \leq r \leq r_2$, as illustrated in Fig. 11.19, letting $u^1 = r$. For the outer boundary

$$\begin{aligned} x &= r_2 \sin \theta \cos \phi & 0 \leq \theta \leq \pi, \\ y &= r_2 \sin \theta \sin \phi & 0 \leq \phi \leq 2\pi, \\ z &= r_2 \cos \theta & (u^2, u^3) = (\theta, \phi), \end{aligned}$$

the radial coordinate vector field $E_1(u) = e_r$ points out of Σ , while the ordered pair (θ, ϕ) orient the outer spherical boundary (equivalent to the choice of outer normal by the right hand rule). Instead for the inner boundary

$$\begin{aligned} x &= r_1 \sin \theta \cos \phi & 0 \leq \theta \leq \pi, \\ y &= r_1 \sin \theta \sin \phi & 0 \leq \phi \leq 2\pi, \\ z &= r_1 \cos \theta & (u^2, u^3) = (\phi, \theta), \end{aligned}$$

the same vector field $E_1(u)$ points into Σ , and instead the ordered pair (ϕ, θ) orient the inner spherical boundary (equivalent to the choice of inner normal by the right hand rule). \square

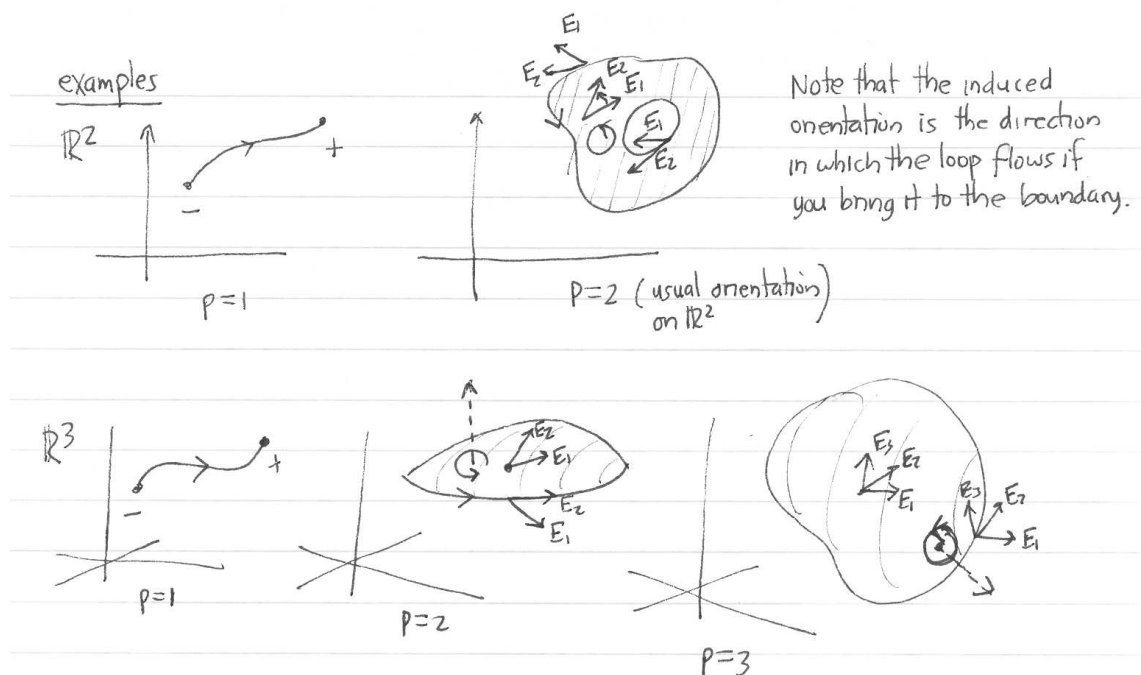


Figure 11.18: Examples of parametrized regions with boundary and the induced orientation of the boundary in \mathbb{R}^2 and \mathbb{R}^3 .

Example 11.9.2. Let Σ be a ball of radius R in \mathbb{R}^4

$$(x^1)^2 + (x^2)^2 + (x^3)^2 + (x^4)^2 \leq R^2.$$

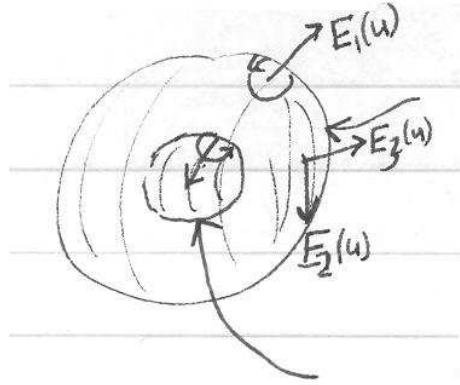
At the North Pole $(0, 0, 0, 1)$, the last Cartesian coordinate frame vector field $\partial/\partial x^4$ points out of Σ . The tangent plane $x^4 = R$ is tangent to $\partial\Sigma$ and because it takes 3 transpositions to restore the outward normal $\frac{\partial}{\partial x^4}$ to its rightful place at the end for the usual orientation $\frac{\partial}{\partial x^1} \wedge \frac{\partial}{\partial x^2} \wedge \frac{\partial}{\partial x^3} \wedge \frac{\partial}{\partial x^4}$ of the whole space, we must choose the inner orientation as follows

$$\frac{\partial}{\partial x^4} \wedge \underbrace{\left(-\frac{\partial}{\partial x^1} \wedge \frac{\partial}{\partial x^2} \wedge \frac{\partial}{\partial x^3} \right)}_{\text{inner orientation of } \partial\Sigma} = \underbrace{\frac{\partial}{\partial x^1} \wedge \frac{\partial}{\partial x^2} \wedge \frac{\partial}{\partial x^3} \wedge \frac{\partial}{\partial x^4}}_{\text{orientation of } \mathbb{R}^4 \text{ taken as orientation for } \Sigma}.$$

The induced orientation of the boundary $\partial\Sigma = S^3$ at the North Pole is the opposite of the subspace $\mathbb{R}^3 \subset \mathbb{R}^4$ ($x^4 = 0$) with its natural orientation $\partial/\partial x^1 \wedge \partial/\partial x^2 \wedge \partial/\partial x^3$. We will return to this case in detail later. \square

Induced orientation for any coordinate ordering

The key property of the induced orientation on $\partial\Sigma$ is that if the coordinate x^k (for some fixed value of k) whose interval $a \leq u^k \leq b$ delimits Σ such that it has a boundary at $u^k = b$ and/or

**Example 11.9.1.**Figure 11.19: The region between two concentric spheres at the origin in \mathbb{R}^3 and its boundary.

at $u^k = a$, then the wedge of $E_k(u)$ from the left onto the induced orientation $(p-1)$ -vector should result in the p -vector $E_i(u) \wedge \dots \wedge E_p(u)$ orienting Σ for the part of the boundary at $u^k = b$, and the opposite sign at $u^k = a$. This is accomplished by the following signs, noting that it requires $k-1$ transpositions to get $E_k(u)$ to its ordered location (note that the vector $E_k(u)$ is missing in the $(p-1)$ -vector on the right in each case)

at $u^k = b$:

$$E_k(u) \wedge [(-1)^{k-1} E_1(u) \wedge \dots \wedge E_{k-1}(u) \wedge E_{k+1}(u) \wedge \dots \wedge E_p(u)] = E_1(u) \wedge \dots \wedge E_p(u),$$

at $u^k = a$:

$$-E_k(u) \wedge [(-1)^{k-1} E_1(u) \wedge \dots \wedge E_{k-1}(u) \wedge E_{k+1}(u) \wedge \dots \wedge E_p(u)] = E_1(u) \wedge \dots \wedge E_p(u).$$

Thus the $(p-1)$ -vector in square brackets provides the induced orientation on these parts of the boundary.

Exercise 11.9.1.**snow cone surface integral**

Suppose we consider the snow cone region $0 \leq \theta \leq \alpha < \pi/2$, $0 \leq r \leq a$, $0 \leq \phi \leq 2\pi$ in spherical coordinates, oriented with the usual orientation by the ordered coordinates (r, θ, ϕ) consistent with the ordered Cartesian coordinates (x, y, z) . On the top surface $\partial\Sigma_{\text{top}} : r = a$, then (θ, ϕ) are oriented ($\partial/\partial\theta \times \partial/\partial\phi \sim \partial/\partial r$ is along the outer normal), but on the lateral side surface $\partial\Sigma_{\text{side}} : \theta = \alpha$, then (ϕ, r) are oriented ($\partial/\partial\phi \times \partial/\partial r \sim \partial/\partial\theta$ is along the outer normal).

Show that for $\alpha = \pi/6$, the surface integral of the vector field $\vec{F} = \langle 0, 0, x^2 \rangle$ over $\partial\Sigma_{\text{top}}$ is $\frac{\pi}{64}$, while the its integral over $\partial\Sigma_{\text{side}}$ has the opposite sign. The result is that the surface integral over $\partial\Sigma$ is zero. Note that we evaluate this surface integral by integrating the 2-form $*F^\flat = x^2 dx \wedge dy$.

■

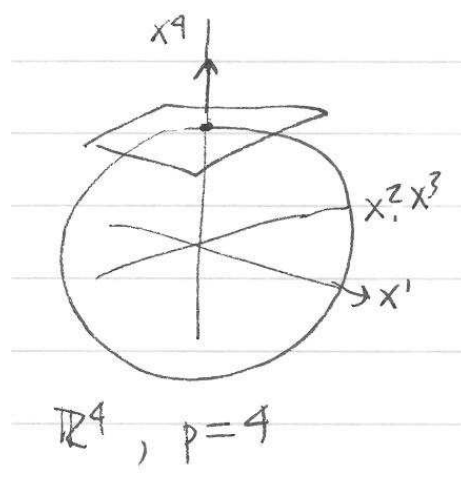


Figure 11.20: Examining the induced orientation in the tangent plane at the North pole of a sphere in \mathbb{R}^4 .

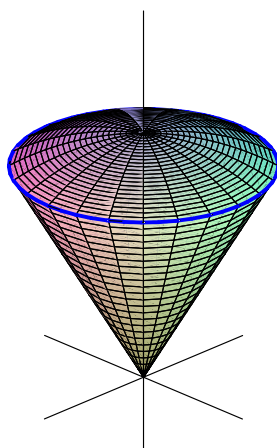


Figure 11.21: A snow cone region Σ of \mathbb{R}^3 with a boundary $\partial\Sigma$ consisting of a lateral side which is part of a cone $\theta = \alpha < \pi/2$, and a top which is part of a sphere $r = a > 0$. The induced orientation on the boundary corresponding by the right hand rule to the outward normal is respectively (θ, ϕ) and (ϕ, r) .

11.10 Stokes' theorem

Let T be a $(p-1)$ -form on \mathbb{R}^n and let B be an oriented p -surface with boundary ∂B with the induced orientation. Then the integral of T on the boundary of B equals the integral of the differential of T on B itself

$$\int_{\partial B} T = \int_B dT.$$

The proof of this is basically a generalization of the simple proof of Green's theorem in the plane found in most multivariable calculus textbooks. This theorem does not require a metric to evaluate either side of the equality. However, if one has a metric around to use, one can rewrite this theorem using the metric so that one can have a better mental picture of what it represents geometrically. This rewriting task expresses the $(p-1)$ -form as the dual of an index-lowered $(n-p+1)$ -vector field and expresses linear combinations of components and basis $(p-1)$ -forms as metric dot products. However, to understand why we went to the trouble of fixing an induced orientation on the boundary with a certain sign choice, it is instructive to look at a piece of the proof in which this sign pops up. First let's appreciate how this theorem captures both Stokes' theorem and Gauss's law in multivariable calculus.

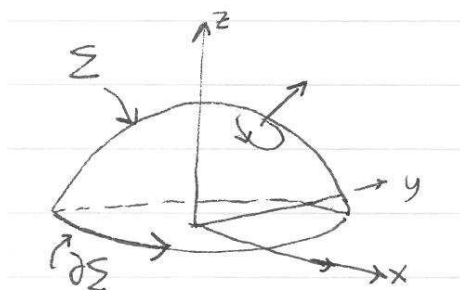


Figure 11.22: The upward normal oriented upper hemisphere and its bounding circle with the induced counterclockwise (seen from above) orientation, linked to the upward unit normal by the right hand rule.

The case $p = 2$ in \mathbb{R}^3

Let $X^\flat = X_i dx^i = g_{ij} X^j dx^i$ be our 1-form. Then using the fact that the double dual $** = 1$ is the identity operation on \mathbb{R}^3 for every p -form, the $p = 1$ version of Stokes' theorem states

$$\int_{\partial B} X^\flat = \int_B dX^\flat.$$

Then we can convert this into a form in which we introduce the usual Euclidean metric by rewriting everything explicitly for the vector field X instead of the 1-form X^\flat . First the left

hand side becomes

$$\int_{\partial B} X^b = \int_{\partial B} X_i dx^i = \int_{\partial B} X^i g_{ij} dx^j = \int_{\partial B} X \cdot d\vec{s},$$

where we define the vector differential of arclength and its index-lowered form by

$$ds^i = dx^i, \quad ds_i = g_{ij} dx^j.$$

Then the right hand side becomes

$$\int_B dX^b = \int_B \underbrace{*(dX^b)}_{[\text{curl } X]^b} = \int_B (\text{curl } X) \cdot d\vec{S},$$

using the fact that the dual of the curl is

$$*(dX^b) = *[(\text{curl } X)^b] = (\text{curl } X)^i \underbrace{\eta_{ijk} dx^{jk}/2}_{dS_i} = (\text{curl } X) \cdot d\vec{S},$$

and $dS^i = g^{ij} dS_j$ is the vector differential of surface area on B . Finally putting both sides together again leads to the metric version of the $p = 1$ Stokes' theorem

$$\int_{\partial B} X \cdot d\vec{s} = \int_B \text{curl } X \cdot d\vec{S},$$

which is the usual Stokes' theorem in \mathbb{R}^3 , equating the line integral of the vector field around the closed boundary curve to the surface integral of its curl over the surface bounded by that curve. This latter integral which is the integral of the normal component of the curl of the vector field with respect to surface area over that surface, while the line integral is interpreted as the total circulation of the vector field around the loop. Thus the normal component of the curl of the vector field is interpreted as a local circulation surface density of the vector field within the surface whose surface integral is the total circulation.

This can be further deconstructed in terms of length and direction information by introducing the unit tangent \hat{T} to the boundary curve pointing in the direction of the induced orientation and defining the scalar differential of arclength

$$d\vec{s} = \hat{T} ds, \quad \text{or} \quad dx^i = T^i ds,$$

and introducing similarly the surface unit normal \hat{n} picked out by the right hand rule from the inner orientation of the surface

$$d\vec{S} = \hat{n} dS, \quad \text{or} \quad dS^i = n^i dS.$$

With these definitions, Stokes' theorem becomes

$$\int_{\partial B} X \cdot \hat{T} d\vec{s} = \int_B (\text{curl } X) \cdot \hat{n} d\vec{S},$$

interpreting the line integral of the vector field as the integral of its tangential component with respect to the scalar differential of arclength on the directed curve, and interpreting the surface integral of the curl vector field as the integral of its normal component with respect to the scalar differential of surface area on the oriented surface.

The case $p = 3$ in \mathbb{R}^3

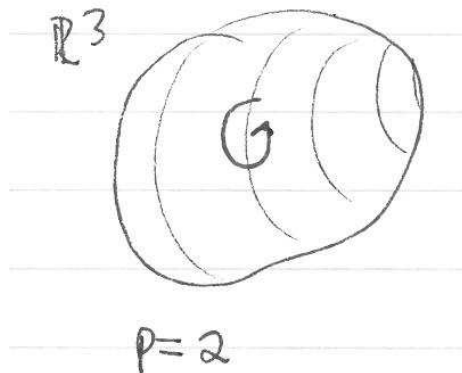


Figure 11.23: The induced orientation on the boundary of a closed surface in \mathbb{R}^3 , whose inner counterclockwise orientation as seen from outside is linked to the outer normal by the right hand rule.

Let $*X^\flat = \frac{1}{2}X^i\eta_{ijk}dx^{jk}$ be our 2-form on \mathbb{R}^3 . Then again using $** = 1$ on \mathbb{R}^3 , the $p = 2$ version of Stokes' theorem states

$$\int_{\partial B} *X^\flat = \int_B d*X^\flat.$$

Rewriting the left hand side yields

$$\int_{\partial B} *X^\flat = \int_{\partial B} \frac{1}{2}X^i\eta_{ijk}dx^{jk} = \int_{\partial B} X^i dS_i = \int_{\partial B} X \cdot d\vec{S}.$$

Rewriting the right hand side using the divergence identity $*d*X^\flat = \text{div } X$ yields

$$\int_B d*X^\flat = \int_B \underbrace{*(d*X^\flat)}_{\text{div } X} = \int_B \underbrace{*(\text{div } X)}_{(\text{div } X)\eta},$$

so using the more suggestion notation $dV = \eta$ for the oriented unit volume 3-form, Stokes' theorem becomes

$$\int_{\partial B} X \cdot d\vec{S} = \int_B (\text{div } X) dV,$$

or

$$\int_{\partial B} X \cdot \hat{n} dS = \int_B (\text{div } X) dV.$$

This is Gauss's law. Its physical interpretation is that the integral of the divergence of a vector field over a region equals the flux of the vector field through the bounding surface of that region (its normal component integrated with respect to the differential of surface area). Thus the divergence is treated like a local volume flux density whose integral yields the total flux out of the region.

Idea of the proof of Stokes' Theorem

The technical proof of Stokes' Theorem involves complications similar to defining manifolds through a system of overlapping coordinate systems (Google: integration on chains stokes' theorem), but for a first exposure to this topic, this is all overkill. I myself never had the patience to fully absorb that stuff and in practice: surprise, one never uses it anyway. Mathematicians like to know that the foundation is solid, though, and we must thank them for doing this tedious job for us. Believe me, tedious is an understatement.

The key idea of the proof is relatively simple and explains the induced orientation sign convention. Imagine a parametrized p -surface Σ in \mathbb{R}^n (parametrization map $\Psi : \mathcal{U} \subset \mathbb{R}^p \rightarrow \mathbb{R}^n$) with bounding $(p-1)$ -surface $\partial\Sigma$ that can be described as the image of the region $a \leq u^k \leq b$ in the parameter space for a particular parameter u^k , where $1 \leq k \leq p$. Let $\partial\Sigma_-$ correspond to $u^k = a$ and $\partial\Sigma_+$ correspond to $u^k = b$. Furthermore assume the coordinates x^i are adapted to Σ in the sense that the first p coordinates parametrize Σ , which is described by constant values of the remaining coordinates

$$\Psi : \quad x^1 = u^1, \dots, x^p = u^p, \quad x^{p+1} = x_0^{p+1}, \dots, x^n = x_0^n.$$

In practice any p coordinates can serve as the parameters, not just the first p coordinates. Some good examples to keep in mind are in \mathbb{R}^3 in spherical coordinates: for $p = 2$ the ring-like band region on a sphere $r = r_0, \alpha \leq \theta \leq \beta$, or for $p = 3$ the region $r_1 \leq r \leq r_2$ between two spheres, or in cylindrical coordinates: for $p = 2$ a part of the cylindrical surface $\rho = \rho_0, a \leq z \leq b$, or for $p = 3$ the cylindrical solid $0 \leq \rho \leq \rho_0, a \leq z \leq b$, which has two such separate boundaries.

We need some simplifying notation for this exercise for a single omitted basis vector or covector, using angle brackets for the omitted index since these delimiters have not yet been used with indices

$$\begin{aligned} E_{\langle k \rangle}(u) &= E_1(u) \wedge \dots \wedge E_{k-1}(u) \wedge E_{k+1}(u) \wedge \dots \wedge E_p(u), \\ du^{\langle k \rangle} &= du^1 \wedge \dots \wedge du^{k-1} \wedge du^{k+1} \wedge \dots \wedge du^p, \end{aligned}$$

which satisfy

$$\begin{aligned} E_k(u) \wedge [(-1)^{k-1} E_{\langle k \rangle}(u)] &= E_1(u) \wedge \dots \wedge E_p(u), \\ du^k \wedge [(-1)^{k-1} du^{\langle k \rangle}] &= du^1 \wedge \dots \wedge du^p = du^{1\dots p}. \end{aligned}$$

The sign comes from the $k-1$ transpositions needed to get the missing factor back to its proper ordered place. Comparing this with the induced orientation, if $E_1(u) \wedge \dots \wedge E_p(u)$ orients Σ , then $(-1)^{k-1} E_{\langle k \rangle}(u)$ determines the induced orientation of $\partial\Sigma_+$ and $-(-1)^{k-1} E_{\langle k \rangle}(u)$ determines the induced orientation of $\partial\Sigma_-$. Alternatively, the integral of the $(p-1)$ -form $(-1)^{k-1} du^{\langle k \rangle}$ converts to an ordinary iterated integral $(-1)^{k-1} du^1 \dots du^{k-1} du^{k+1} \dots du^p$ in the first case, and the negative of this in the second case.

Now suppose we have a simple $(p-1)$ -form $\omega = f dx^{\langle k \rangle}$ such that its pullback to the parameter space takes the simple form $\Psi^* \omega = f du^{\langle k \rangle}$, ignoring the dependence on x or u . In fact if we start with any $(p-1)$ -form, then pulling it back to the p -surface will cause all

other components to vanish since they will contain factors of the differentials of the coordinates which are held fixed on the surface and hence are zero (if more than one of the p coordinate differentials dx^i , $1 \leq i \leq p$ is missing from the p -form, such an unwanted differential dx^i , $p < i \leq n$ will be present in its place). Integrating the p -form $d\omega$ over the p -surface simply means setting $x^i = x_0^i$ and $dx^i = 0$ for $p < i \leq n$ and replacing x^i by u^i for $0 \leq i \leq p$, and then converting the differential form $du^{1\dots p}$ to a simple p -fold iterated integral, as follows

$$\begin{aligned}
 \Psi^* d\omega &= d(\Psi^* \omega) = (\partial f / \partial u^k) du^k \wedge du^{\langle k \rangle} = (\partial f / \partial u^k) (-1)^{k-1} du^{1\dots p}, \\
 \int_{\Sigma} d\omega &= \int_{\mathcal{U}} d(\Psi^* \omega) = \int_{\mathcal{U}} (\partial f / \partial u^k) (-1)^{k-1} du^{1\dots p} \\
 &= \underbrace{\int \cdots \int}_{p-1} \int_a^b (\partial f / \partial u^k) (-1)^{k-1} du^k du^1 \cdots du^{k-1} du^{k+1} du^p \\
 &= \underbrace{\int \cdots \int}_{p-1} [f|_{u^k=b} - f|_{u^k=a}] (-1)^{k-1} du^1 \cdots du^{k-1} du^{k+1} du^p \\
 &= \underbrace{\int \cdots \int}_{p-1} f|_{u^k=b} (-1)^{k-1} du^1 \cdots du^{k-1} du^{k+1} du^p \\
 &\quad - \underbrace{\int \cdots \int}_{p-1} f|_{u^k=a} (-1)^{k-1} du^1 \cdots du^{k-1} du^{k+1} du^p \\
 &= \int_{\partial \Psi_+} \omega + \int_{\partial \Psi_-} \omega = \int_{\partial \Psi} \omega.
 \end{aligned}$$

Here the underbrace notation is a suggestive way of symbolically representing some explicit $(p-1)$ -fold iterated integral over the allowed ranges of the remaining coordinates/parameters explicitly describing the p -surface, implicitly described by constant values of the remaining $n-p$ coordinates. The only role played by the induced orientation on the boundary is to assign the appropriate sign coefficient $(-1)^{k-1}$ in front of each such iterated integral. The sign $(-1)^{k-1}$ arises simply from the exterior derivative permutation to order the differentials to get the orientation of the surface. By doing the u^k integral first, one undoes the corresponding partial derivative leading to a difference of values at the two endpoints describing the allowed range of that coordinate/parameter for the p -surface. The two separate remaining $(p-1)$ -fold iterated integrals then correspond to the integral of the original p -form ω over the two parts of the boundary $\partial \Sigma_+$ and $\partial \Sigma_-$ with the appropriate sign which defines the induced orientation.

Remark.

For a $(p-1)$ -form α and a p -form β we defined the codifferential by the identity

$$\langle d\alpha, \beta \rangle \eta = d(\alpha \wedge {}^* \beta) + \langle \alpha, \delta \beta \rangle \eta.$$

If we promote the point-wise inner product \langle, \rangle to an inner product on the space of differential forms of each degree by integrating it over our space with respect to the unit volume n -form

$$\langle\langle S, T \rangle\rangle = \int_{\Sigma} \langle S, T \rangle \eta,$$

then integrating the previous identity and using Stokes' theorem leads to

$$\begin{aligned} \langle\langle d\alpha, \beta \rangle\rangle &= \langle\langle \alpha, \delta\beta \rangle\rangle + \int_{\Sigma} d(\alpha \wedge {}^*\beta) \\ &= \langle\langle \alpha, \delta\beta \rangle\rangle + \int_{\partial\Sigma} \alpha \wedge {}^*\beta. \end{aligned}$$

When the final term vanishes, this states that the codifferential is the adjoint of the exterior derivative with respect to this global inner product for p -forms. On a sphere or a torus, for example, there is no boundary ("the boundary vanishes"), so this term is not present. On all of $\Sigma = \mathbb{R}^3$ which we can imagine as having its boundary $\partial\Sigma$ the limiting sphere at infinity, we need to consider p -forms for which the inner products converge as integrals, so we have to restrict ourselves to those p -forms for which the self-inner product is finite (square-integrable). By suitably restricting the asymptotic dependence on the distance from the origin, this term can be made zero as well. This term arises in the variation of the action integral of Lagrangian function as a "total divergence" which can be integrated away to the boundary where the variation is fixed and hence not contribute to the calculation of the Lagrange equations.

When $p = 1$ so that α is a 0-form or function, the above identity becomes

$$\int_{\Sigma} \langle d\alpha, \beta \rangle \eta = \int_{\partial\Sigma} \alpha ({}^*\beta) + \int_{\Sigma} \alpha \delta\beta \eta.$$

in progress...



11.11 Worked examples of Stokes' theorem and Gauss's law for \mathbb{R}^3

It is important to work through some examples and compare them with the approach of ordinary multivariable calculus, where the general Stokes' theorem becomes the ordinary Stokes' theorem for a closed surface integral and its boundary but Gauss's law for a closed region of space and its closed surface boundary.

The ordinary Stokes' theorem in \mathbb{R}^3

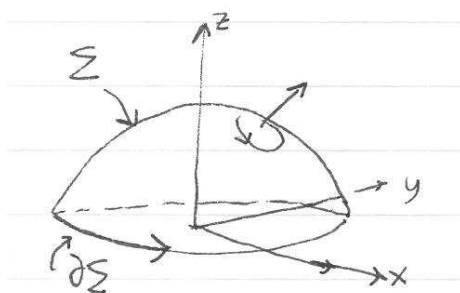


Figure 11.24: The upward normal oriented upper hemisphere and its bounding circle with the induced counterclockwise (seen from above) orientation.

We start with a closed surface and its boundary illustrated in figure 11.24: the upper hemisphere of radius a at the origin bounded by the counterclockwise directed circle (as seen from above) of radius a in the x - y plane. Stokes theorem for the integral of a 1-form X^b on the boundary is

$$\int_{\partial\Sigma} X^b = \int_{\Sigma} dX^b,$$

which will require parametrizations for both the surface and bounding curve with the correct orientations

$$\begin{aligned} \Sigma &: x^2 + y^2 + z^2 = a^2, \quad z \geq 0, && \text{oriented by upper normal} \\ \partial\Sigma &: x^2 + y^2 = a^2, \quad z = 0. && \text{induced orientation: counterclockwise from above} \end{aligned}$$

To get these compatible parametrizations, we describe Σ by setting $r = a$ in the spherical coordinate parametrization map Ψ . Call this parametrization map Ψ_{Σ} :

$$\begin{aligned} x &= a \sin \theta \cos \phi, & 0 \leq \theta \leq \pi/2, \\ y &= a \sin \theta \sin \phi, & 0 \leq \phi \leq 2\pi, \\ z &= a \cos \theta, \end{aligned}$$

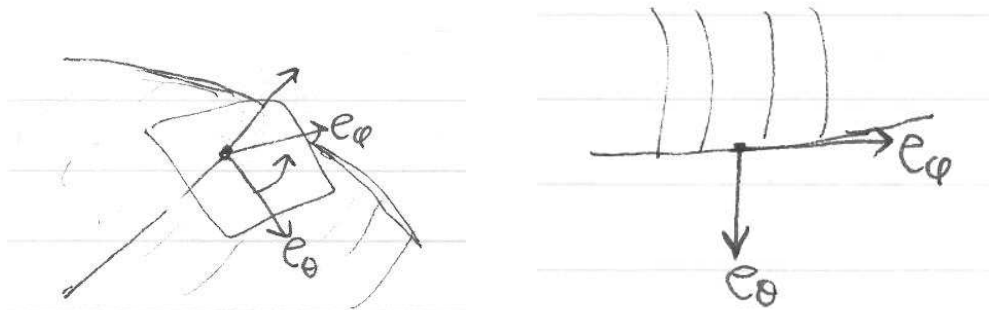


Figure 11.25: Left: Linking the inner orientation to the outward normal of the upper hemisphere by the right hand rule. Right: The right hand rule curling from e_θ to e_ϕ on the boundary of the upper hemisphere picks out the outward (upward above the x - y plane) normal.

and by additionally setting $\theta = \pi/2$ in the parametrization map to parametrize the bounding circle $\partial\Sigma$. Call this parametrization map $\Psi_{\partial\Sigma}$:

$$\begin{aligned} x &= a \cos \phi, & 0 \leq \phi \leq 2\pi, \\ y &= a \sin \phi, \\ z &= 0. \end{aligned}$$

Then the ordered pair (θ, ϕ) orient Σ with the correct orientation related to the upward normal by the right hand rule and with the correct counterclockwise orientation of the bounding circle (equator of sphere) as seen from above given by ϕ

$$e_\theta \wedge [e_\phi] = e_\theta \wedge e_\phi.$$

Now we are ready to perform an integral with a specific 1-form.

Example 11.11.1. line integral: example 1

We need a 1-form to use to verify this version of Stokes' theorem. Let's take our old friend which is an exact differential

$$\begin{aligned} X^\flat &= ydx + xdy = d(xy), \\ dX^\flat &= dy \wedge dx + dx \wedge dy = 0 = d^2(xy), \end{aligned}$$

so that the right hand side of Stokes' Theorem is identically zero. The left hand side is

$$\begin{aligned} \Phi_{\partial\Sigma}^*(X^\flat) &= (a \sin \phi) d(a \cos \phi) + (a \cos \phi) d(a \sin \phi) \\ &= -a^2 \sin^2 \phi d\phi + a^2 \cos^2 \phi d\phi = a^2 \cos 2\phi d\phi, \\ \int_{\partial\Sigma} X^\flat &= \int_0^{2\pi} a^2 \cos 2\phi d\phi = -\frac{1}{2}a^2 \sin 2\phi \Big|_0^{2\pi} = 0. \end{aligned}$$

□

Example 11.11.2. line integral: example 2

Okay, since our first trial vector field led to a zero result, let's try something more interesting by switching a sign in the first component to get a nonzero result. Take instead

$$\begin{aligned} X^b &= -ydx + xdy, \\ dX^b &= -dy \wedge dx + dx \wedge dy = 2dx \wedge dy. \end{aligned}$$

Then the left hand side of Stokes' theorem is the line integral

$$\begin{aligned} \Phi_{\partial\Sigma}^*(X^b) &= -(a \sin \phi) d(a \cos \phi) + (a \cos \phi) d(a \sin \phi) = a^2 d\phi, \\ \underbrace{\int_{\partial\Sigma} X^b}_{\oint_{\partial\Sigma} X \cdot \vec{ds}} &= \int_0^{2\pi} a^2 d\phi = 2\pi a^2. \end{aligned}$$

The right hand side of Stokes' theorem is the surface integral

$$\begin{aligned} \Phi_{\partial\Sigma}^*(dX^b) &= 2d(a \sin \theta \cos \phi) \wedge d(a \sin \theta \sin \phi) \\ &= 2a^2(\cos \theta \cos \phi d\theta - \sin \theta \sin \phi d\phi) \wedge (\cos \theta \sin \phi d\theta + \sin \theta \cos \phi d\phi) \\ &= 2a^2(\sin \theta \cos \theta \cos^2 \phi d\theta \wedge d\phi - \sin \theta \cos \theta \sin^2 \phi d\phi \wedge d\theta) \\ &= 2a^2 \sin \theta \cos \theta d\theta \wedge d\phi = a^2 \sin 2\theta d\theta \wedge d\phi, \\ \underbrace{\int_{\Sigma} dX^b}_{\int (\text{curl } X) \cdot \hat{n} dS} &= \int_0^{2\pi} \int_0^{\pi/2} a^2 \sin 2\theta d\theta d\phi = 2\pi a^2 \left(-\frac{1}{2} \cos 2\theta\right) \Big|_0^{\pi/2} = 2\pi a^2. \end{aligned}$$

□

Example 11.11.3. the multivariable calculus approach to the previous problem

The multivariable calculus approach evaluates everything in terms of vector fields

$$\int_{\Sigma} \text{curl } X \cdot \hat{n} dS = \int_{\partial\Sigma} X \cdot \hat{T} \vec{ds}$$

The previous vector field and its curl in that component notation are

$$\begin{aligned} X &= \langle -y, x, 0 \rangle, \\ \text{curl } X &= \left\langle \frac{\partial}{\partial y}(0) - \frac{\partial}{\partial z}(x), \frac{\partial}{\partial z}(-y) - \frac{\partial}{\partial x}(0), \frac{\partial}{\partial x}(x) - \frac{\partial}{\partial y}(-y) \right\rangle = \langle 0, 0, 2 \rangle. \end{aligned}$$

The unit outward normal to the hemisphere is obtained by normalizing the gradient of the radial function r^2

$$r^2 = x^2 + y^2 + z^2 = a^2 \longrightarrow n = \frac{1}{2} \vec{\nabla}(x^2 + y^2 + z^2) = \langle x, y, z \rangle, \quad \hat{n} = \frac{1}{r} \langle x, y, z \rangle,$$

$$\text{curl } X \cdot \hat{n} = \frac{2z}{r} = 2 \cos \theta.$$

Recall that the surface area differential for the sphere derived using limiting orthogonal ar-length arguments is $dS = a^2 \sin \theta d\theta d\phi$ so the vector differential is $\vec{dS} = \hat{n} dS$ and the surface integral is

$$\int_{\Sigma} \text{curl } X \cdot \hat{n} dS = \int_0^{2\pi} \int_0^{\pi/2} (2 \cos \theta)(a^2 \sin \theta) d\theta d\phi = 2\pi a^2,$$

as before.

The line integral requires parametrizing the circle which is easily done using the spherical coordinates setting $r = a$, $\theta = \pi/2$, $\phi = t$ so that

$$\begin{aligned} \vec{r}(t) &= \langle x(t), y(t), z(t) \rangle = \langle a \cos t, a \sin t, 0 \rangle, \\ \vec{r}'(t) &= \langle x'(t), y'(t), z'(t) \rangle = \langle -a \sin t, a \cos t, 0 \rangle = \langle -y(t), x(t), 0 \rangle = X(\vec{r}(t)), \\ \hat{T}(t) &= \frac{\langle -y(t), x(t), 0 \rangle}{(x(t)^2 + y(t)^2)^{1/2}} \\ ds &= a dt. \end{aligned}$$

Thus

$$X(\vec{r}(t)) \cdot \hat{T}(t) = \frac{y(t)^2 + x(t)^2}{(x(t)^2 + y(t)^2)^{1/2}} = (x(t)^2 + y(t)^2)^{1/2} = a,$$

so that

$$\int_{\partial \Sigma} X \cdot \hat{T} \vec{ds} = \int_0^{2\pi} a(a dt) = 2\pi a^2,$$

or just plugging into the parametrization

$$\begin{aligned} &= \int X(\vec{r}(t)) \cdot \vec{r}'(t) dt = \int_0^{2\pi} [(-a \sin t)(-a \sin t) + (a \cos t)(a \cos t)] dt \\ &= \int_0^{2\pi} a^2 dt = 2\pi a^2. \end{aligned}$$

The nonmetric version is clearly simpler, but the metric version gives us a physical picture of what we are integrating.

□

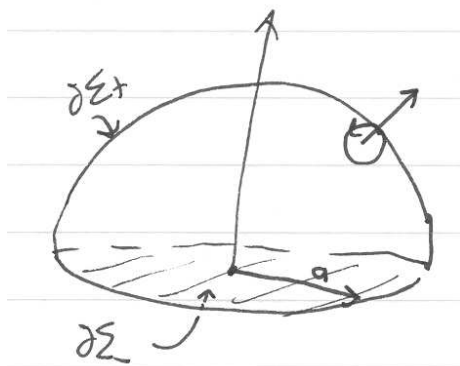


Figure 11.26: The upper hemisphere of radius a joined to the circular disk in the x - y plane that it cuts off, forming a closed surface $\partial\Sigma$ which is the boundary of the upper half ball of radius a .

A Gauss's law problem

Let Σ be the interior of the upper hemisphere of radius a at the origin, with the usual \mathbb{R}^3 orientation $dx \wedge dy \wedge dz \sim dr \wedge d\theta \wedge d\phi$. Its boundary $\partial\Sigma$ has two parts: the upper hemisphere with the upward (outer) orientation and the disk of radius a in the xy plane with the downward (outer) orientation.

In each case we can use a spherical coordinate parametrization

$$\begin{array}{lll} \Sigma : & x = r \sin \theta \cos \phi, & 0 \leq r \leq a, \\ (r, \theta, \phi) \text{ oriented} & y = r \sin \theta \sin \phi, & 0 \leq \theta \leq \pi/2, \\ & z = r \cos \theta, & 0 \leq \phi \leq 2\pi, \end{array}$$

$$\begin{array}{lll} \partial\Sigma_+ : & x = r \sin \theta \cos \phi, & 0 \leq \theta \leq \pi/2, \\ (\theta, \phi) \text{ oriented} & y = r \sin \theta \sin \phi, & 0 \leq \phi \leq 2\pi, \\ & z = r \cos \theta, & \end{array}$$

$$\begin{array}{lll} \partial\Sigma_- : & x = r \cos \phi, & 0 \leq r \leq a, \\ (\phi, r) \text{ oriented} & y = r \sin \phi, & 0 \leq \phi \leq 2\pi, \\ & z = 0. & \end{array}$$

where the latter induced orientation follows from the boundary condition $\theta \leq \pi/2$

$$e_\theta \wedge [e_\phi \wedge e_r] = e_r \wedge e_\phi \wedge e_\theta.$$

We need a 2-form to integrate on $\partial\Sigma$. Take $*X^\flat$, where

$$\begin{aligned} X &= x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z} = r \frac{\partial}{\partial r}, \\ *X^\flat &= *(xdx + ydy + zdz) = xdy \wedge dz + ydz \wedge dx + zdx \wedge dy, \\ d*X^\flat &= d(xdy \wedge dz + ydz \wedge dx + zdx \wedge dy) = 3dx \wedge dy \wedge dz, \end{aligned}$$

or in spherical coordinates

$$\begin{aligned} *X^\flat &= *(rdr) = r *dr = r *\omega^{\hat{r}} = r \omega^{\hat{\theta}\hat{\phi}} = r(r d\theta) \wedge (r \sin \theta d\phi) = r^3 \sin \theta d\theta \wedge d\phi, \\ d*X^\flat &= 3r^2 \sin \theta dr \wedge d\theta \wedge d\phi. \end{aligned}$$

Then Stokes theorem is

$$\int_{\partial\Sigma} *X^\flat = \int_{\Sigma} d(*X^\flat).$$

The right hand side (volume integral) is

$$\int_{\Sigma} 3dx \wedge dy \wedge dz = 3 \int_0^{2\pi} \int_0^{\pi/2} \int_0^a r^2 \sin \theta dr d\theta d\phi = 3(2\pi a^3/3) = 2\pi a^3,$$

while the left hand side (surface integral) is

$$\begin{aligned} \int_{\partial\Sigma} r^3 \sin \theta d\theta \wedge d\phi &= \int_{\partial\Sigma_+} r^3 \sin \theta d\theta \wedge d\phi + \int_{\partial\Sigma_-} r^3 \sin \theta d\theta \wedge d\phi \\ &= \int_0^{2\pi} \int_0^{\pi/2} a^3 \sin \theta d\theta d\phi + 0 = 2\pi a^3, \end{aligned}$$

where the second integral is zero since on the base of the hemisphere, $\theta = \pi/2$, so $d\theta$ is zero on that flat disc.

Exercise 11.11.1.

paraboloidal solid integration

Because of the axial symmetry, polar and cylindrical coordinates are appropriate to express the integrals once their integrands have been evaluated in Cartesian coordinates.

a) Evaluate both sides of Gauss's law for the vector field $\langle yz, xz, xy \rangle$ and the solid region between the bounding surfaces $z = 4 - x^2 - y^2$ and $z = 0$.

b) For each of these two bounding surfaces, this time both with the upward normal, evaluate both sides of Stoke's theorem for this vector field.

■

Exercise 11.11.2.

wedge of cylinder integration

Evaluate both sides of Gauss's law for the vector field $\langle x, y, z + z^2 \rangle$ and the solid region between the bounding surfaces $z = x + 1$ and $z = 0$ enclosed by the unit cylinder $x^2 + y^2 = 1$.



Exercise 11.11.3.
unit ball integration

Evaluate both sides of Gauss's law for the vector field $\langle 0, 0, z \rangle$ and the solid region enclosed by the unit sphere.



11.12 Examples in \mathbb{R}^4 and \mathbb{M}^4

So far the examples we have discussed are right out of any good multivariable calculus course. Unless we consider spaces with dimension larger than 3, we can't really appreciate the generality of this approach. 4-dimensional spaces are the obvious next step for exploration, and find interesting applications to stretch our 3-dimensional intuition to higher dimensions, as well as consider integration in spacetime. As in 3-dimensions, the spheres and cylinders and the pseudospheres and cylinders are a good place to start in Euclidean and Minkowski geometries.

3-spheres, 3-cylinders and 2-cylinders in \mathbb{R}^4

The 3-spheres, 3-cylinders and 2-cylinders centered at the origin in \mathbb{R}^4 are a good test case. Spherical coordinates can be introduced in any \mathbb{R}^n space by the same iterative process that leads from polar coordinates in the plane to spherical coordinates in \mathbb{R}^3 . It is enough to illustrate this for $n = 4$.

Let $R = ((x^1)^2 + (x^2)^2 + (x^3)^2 + (x^4)^2)^{1/2}$ be the radial distance function, and let $(x, y, z) = (x^1, x^2, x^3)$, (ρ, ϕ, z) and (r, θ, ϕ) be the usual Cartesian, cylindrical and spherical coordinates in the subspace $x^4 = 0$, so that $R^2 = r^2 + (x^4)^2$. Analogous to the discussion of spherical coordinates using the ρ - z plane in which the additional polar coordinate decomposition leads from polar to spherical coordinates, we can introduce polar coordinates in the r - x^4 plane measuring the angle down from the positive x^4 -axis

$$\begin{aligned} x^4 &= R \cos \chi, \quad r = R \sin \chi, \\ R &= (\delta_{ij} x^i x^j)^{1/2}, \quad \chi = \arccos(x^4/R). \end{aligned}$$

This leads to one more iteration of what we have already done in \mathbb{R}^3

$$\begin{aligned} x^1 &= R \sin \chi \sin \theta \cos \phi &= r \sin \theta \cos \phi &= \rho \cos \phi, \\ x^2 &= R \sin \chi \sin \theta \sin \phi &= r \sin \theta \sin \phi &= \rho \sin \phi, \\ x^3 &= R \sin \chi \sin \theta &= r \sin \theta &= z, \\ x^4 &= R \cos \chi &= x^4 &= x^4. \end{aligned}$$

The only thing which breaks this pattern of succession of polar coordinates is the fact that in the plane \mathbb{R}^2 , we measure the angle from the positive first axis instead of from the second as in all of the additional angles that are introduced (which only range from 0 to π instead of from 0 to 2π for the first angle), and we adopt the physics convention by switching to the notation (ρ, ϕ) from (r, θ) at the first step.

The Euclidean metric and unit volume 4-form are easily evaluated from the differentials of

the Cartesian coordinates

$$\begin{aligned}
g &= \delta_{ij} dx^i \otimes dx^j = dx^1 \otimes dx^1 + dx^2 \otimes dx^2 + dx^3 \otimes dx^3 + dx^4 \otimes dx^4 \\
&= dR \otimes dR + R^2(d\chi \otimes d\chi + \sin^2 \chi (d\theta \otimes d\theta + \sin^2 \theta d\phi \otimes d\phi)) \\
&= dx^4 \otimes dx^4 + dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2 \theta d\phi \otimes d\phi, \\
&= dx^4 \otimes dx^4 + dz \otimes dz + d\rho \otimes d\rho + \rho^2 d\phi \otimes d\phi, \\
(\det g)^{1/2} &= R^3 \sin^2 \chi \sin \theta = r^2 \sin \theta = \rho d\phi, \\
\eta &= dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4 \\
&= R^3 \sin^2 \chi \sin \theta d\chi \wedge d\phi \wedge d\theta \wedge dR \\
&= r^2 \sin \theta dr \wedge d\theta \wedge d\phi \wedge dx^4 \\
&= \rho d\rho \wedge d\phi \wedge dz \wedge dx^4.
\end{aligned}$$

The Cartesian coordinates with the ordering (x^1, x^2, x^3, x^4) establish an orientation for \mathbb{R}^4 (call it “positive”) which leads to the orderings (χ, θ, ϕ, R) of the spherical coordinates, (r, θ, ϕ, x^4) of the 3-cylindrical coordinates, and (ρ, ϕ, z, x^4) of the 2-cylindrical coordinates being associated with that positive orientation. Notice that (since $\epsilon_{1234} = -\epsilon_{4123}$)

$$dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4 = dx^4 \wedge (-dx^1 \wedge dx^2 \wedge dx^3),$$

so if one considers the closed surface of the unit cube $0 \leq x^i \leq 1$ at the top face $x^4 = 1$, the induced orientation has the opposite sign usual ordering (x^1, x^2, x^3) associated with the orientation of \mathbb{R}^3 . At the North Pole $(x^1, x^2, x^3, x^4) = (0, 0, 0, 1)$ of the unit 3-sphere $R = 1$, (χ, θ, ϕ) behave like a right handed triad in the tangent plane $x^4 = 1$ with respect to the Cartesian coordinates (x^1, x^2, x^3) or (r, θ, ϕ) , but this is opposite to the induced orientation on the closed surface $R = 1$, which is instead associated with the ordering (χ, ϕ, θ) .

Consider the position vector field (with magnitude and direction unit vector)

$$X = x^i \frac{\partial}{\partial x^i} = R \frac{\partial}{\partial R}, \quad |X| = R, \quad n = \hat{X} = \frac{x^i}{R} \frac{\partial}{\partial x^i} = \frac{\partial}{\partial R}$$

and its associated index-lowered 1-form and corresponding unit 1-form

$$\begin{aligned}
X^\flat &= x^1 dx^1 + x^2 dx^2 + x^3 dx^3 + x^4 dx^4 = R \frac{\partial}{\partial R}, \\
n^\flat &= \frac{1}{R}(x^1 dx^1 + x^2 dx^2 + x^3 dx^3 + x^4 dx^4) = dR
\end{aligned}$$

and its dual 3-form

$$\begin{aligned}
\omega = {}^*X^\flat &= x^i \eta_{i[jk\ell]} dx^{jk\ell} = \sum_i x^i (-1)^{i-1} dx^{\langle i \rangle} \\
&= x^1 dx^{234} + x^2 dx^{314} + x^3 dx^{124} - x^4 dx^{123} = R^4 \sin^2 \chi \sin \theta d\chi \wedge d\phi \wedge d\theta \\
&= R \hat{n}^i \eta_{i[jk\ell]} dx^{jk\ell},
\end{aligned}$$

with exterior derivative

$$d\omega = d^*X^\flat = dx^i \eta_{i|j k \ell|} dx^{j k \ell} = 4 dx^{1234} = 4R^3 \sin^2 \chi \sin \theta dR \wedge d\chi \wedge d\phi \wedge d\theta = 4\eta.$$

Since R is an arclength coordinate, $\partial/\partial R = n$ is a unit vector field and it is the field of outward unit normals to the coordinate 3-spheres. By evaluating the first argument of the volume 4-form on this unit vector, we obtain a 3-form which evaluates the 3-volume of the parallelepiped formed by 3 tangent vectors in subspace of the tangent space tangent to the 3-sphere, i.e., it acts as the unit volume 3-form on the coordinate 3-spheres, which is the differential of hypersurface area d^3S on those 3-spheres

$$d^3S = n^i \eta_{i|j k \ell|} dx^{j k \ell} = R^3 \sin^2 \chi \sin \theta d\chi \wedge d\phi \wedge d\theta.$$

Suppose we integrate ω over a spherical coordinate sphere $\partial\Sigma$ of radius R , which is the boundary of the coordinate ball Σ of radius $R = a$. Then Stokes' theorem for this configuration is

$$\begin{aligned} \int_{\partial\Sigma} *X^\flat &= \int_{\Sigma} d(*X^\flat) = \int_{\Sigma} 4\eta \\ &= 4 \int_0^a \int_0^\pi \int_0^\pi \int_0^{2\pi} R^3 \sin^2 \chi \sin \theta dR \wedge d\chi \wedge d\theta \wedge d\phi = 4V_a, \end{aligned}$$

where

$$V_a = \left(\frac{a^4}{4}\right) \left(\frac{\pi}{2}\right) (2)(2\pi) = \frac{\pi^2 a^4}{2}$$

is the volume of a 3-sphere of radius a . The left hand side (3-surface integral on the 3-sphere $R = a$) is

$$\begin{aligned} \int_{\partial\Sigma} *X^\flat &= \int_{\partial\Sigma} a^4 \sin^2 \chi \sin \theta d\chi \wedge d\theta \wedge d\phi \\ &= \int_0^\pi \int_0^\pi \int_0^{2\pi} a^4 \sin^2 \chi \sin \theta d\chi d\theta d\phi \\ &= a^4 \left(\frac{\pi}{2}\right) (2)(2\pi) = 2\pi^2 a^4 = aS_a, \end{aligned}$$

where S_a is the surface area of a 3-sphere of radius a .

For any vector field X , the hypersurface integral of the associated 3-form $*X^\flat$ on a coordinate sphere $\partial\Sigma$ can be represented as the integral of the outward normal component $n \cdot X$ of X with respect to the scalar differential of hypersurface area

$$\int_{\partial\Sigma} *X^\flat = \int_{\partial\Sigma} X^i \eta_{i|j k \ell|} dx^{j k \ell} = \int_{\partial\Sigma} X^i d^3S_i = \int_{\partial\Sigma} X^i n_i d^3S,$$

where

$$d^3S^i = n^i d^3S, \quad d^3S = n^i \eta_{i|j k \ell|} dx^{j k \ell}$$

are the vector and scalar differentials of hypersurface area. Since $** = (-1)^{p(4-p)} = (-1)^p$, for the $p = 4$ -form d^*X^b , the right hand side of Stokes' theorem can be rewritten

$$\int_{\Sigma} d^*X^b = \int_{\Sigma} **d^*X^b = \int_{\Sigma} *(\operatorname{div} X) = \int_{\Sigma} \operatorname{div} X \eta$$

leading to the more suggestive form

$$\int_{\partial\Sigma} X \cdot d^3\vec{S} = \int_{\partial\Sigma} X \cdot n d^3S = \int_{\Sigma} \operatorname{div} X d^4V.$$

This holds for any closed hypersurface in \mathbb{R}^4 , where n is the outward unit normal.

Suppose we consider the 2-form

$$\begin{aligned} \omega &= x^1 dx^2 \wedge dx^3, \\ &= R^2 \sin^2 \chi \sin \theta \cos \phi (\sin \chi \sin \phi d\theta \wedge dR + \sin \chi \sin \theta \cos \theta \cos \phi d\phi \wedge dR \\ &\quad + R \cos \chi \cos \phi d\phi \wedge d\chi + R \cos \chi \sin \theta \cos \theta \cos \phi d\phi \wedge d\chi \\ &\quad + R \sin \chi \sin \theta \cos \phi d\theta \wedge d\phi). \end{aligned}$$

whose differential is

$$\begin{aligned} d\omega &= dx^1 \wedge dx^2 \wedge dx^3 \\ &= R^3 \sin^2 \chi \cos \chi \sin \theta \cos \phi d\chi \wedge d\theta \wedge d\phi + \text{terms with factors of } dR. \end{aligned}$$

Exercise 11.12.1.

3-sphere exterior derivatives

Use a computer algebra system to evaluate ω and $d\omega$ to verify the above results and fill in the remaining terms not explicitly given. ■

For the $p = 3$ Stokes' theorem we need a 3-surface with boundary so take the ring strip on a coordinate 3-sphere $\Sigma : R = a, 0 < \chi_1 \leq \chi \leq \chi_2 < \pi$. This can be parametrized by the ordered triplet (χ, ϕ, θ) , which has the induced orientation on the 3-sphere from the standard orientation on its interior, and the boundary piece $\partial\Sigma_+ : \chi = \chi_2$ has the induced orientation (ϕ, θ) , while $\partial\Sigma_- : \chi = \chi_1$ has the induced orientation (θ, ϕ) . Then

$$\begin{aligned} \int_{\partial\Sigma} \omega &= \int_{\partial\Sigma} a^3 \sin^3 \chi \sin^2 \theta \cos^2 \phi d\theta \wedge d\phi \\ &= \int_{\partial\Sigma_+} -a^3 \sin^3 \chi_1 \sin^2 \theta \cos^2 \phi d\phi \wedge d\theta + \int_{\partial\Sigma_-} a^3 \sin^3 \chi_2 \sin^2 \theta \cos^2 \phi d\phi \wedge d\theta \\ &= \int_0^{2\pi} \int_0^\pi a^3 (-\sin^3 \chi_1 + \sin^3 \chi_2) \sin^2 \theta \cos^2 \phi d\theta d\phi = \frac{4\pi a^3}{3} (-\sin^3 \chi_2 + \sin^3 \chi_1). \end{aligned}$$

On the other hand

$$\begin{aligned} \int_{\Sigma} d\omega &= - \int_{\Sigma} a^3 \sin^2 \chi \cos \chi \sin \theta \cos \phi d\chi \wedge d\phi \wedge d\theta \\ &= - \int_{\chi_1}^{\chi_2} \int_0^{2\pi} \int_0^{\pi} a^3 \sin^2 \chi \cos \chi \sin \theta \cos \phi d\chi d\theta d\phi = -\frac{4\pi a^3}{3} (\sin^3 \chi_2 - \sin^3 \chi_1). \end{aligned}$$

Of course there was nothing to prevent us from using the opposite orientation for Σ associated with the ordered triplet (χ, θ, ϕ) , which is a right handed frame within the 3-sphere just like (r, θ, ϕ) corresponds to a right handed frame in \mathbb{R}^3 , in which case the induced orientation would correspond to the more usual inner orientation (θ, ϕ) on the “outer” 2-sphere $\chi = \chi_2$ (corresponding to the outward normal within the 3-sphere) and the opposite sign on the “inner sphere” $\chi = \chi_1$ (corresponding to the inward normal within the 3-sphere), just like in the case between outer and inner spheres $r = r_2$ and $r = r_1$ in \mathbb{R}^3 .

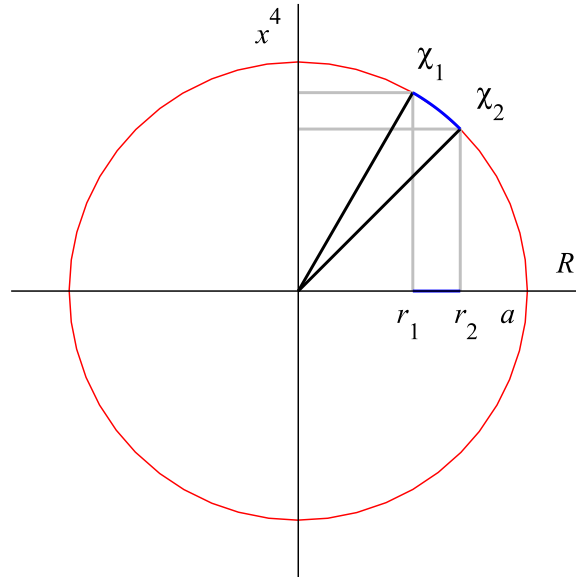


Figure 11.27: The integral of the 3-form $dx^1 \wedge dx^2 \wedge dx^3$ between two 2-spheres on a 3-sphere turns out to be the same as the volume between two 2-spheres of the same radii in \mathbb{R}^3 , but the radial separation of the 2-spheres is longer within the 3-sphere, and the pull back of this 3-form to the 3-sphere is smaller: $\cos \chi \cos \phi d^3 S$, and the two factors apparently compensate exactly to give the equivalent results.

Exercise 11.12.2. integration between 2-spheres

Consider the corresponding problem in spherical coordinates on \mathbb{R}^3 where Σ is the region between two coordinate spheres: $r_1 = a \sin \chi_1 < r_2 = a \sin \chi_2$, $r_1 \leq r \leq r_2$, with the usual

orientation in which $dx^1 \wedge dx^2 \wedge dx^3 = r^2 \sin \theta dr \wedge d\theta \wedge d\phi$ has positive orientation. Show that the Stokes' theorem statement for $\omega = x^1 dx^2 \wedge dx^3 = *(x^1 \partial/\partial x^1)^\flat$ with differential $d\omega = dx^1 \wedge dx^2 \wedge dx^3$ on this configuration has the same numerical values of the left and right hand side as the previous problem. The surface integral is the surface integral of the vector field $X = x^1 \partial/\partial x^1$, with unit divergence $\text{div } X = 1$, so the volume integral is simply the volume between the two 2-spheres.

However, the distance between the two 2-spheres within the 3-sphere is just $L_{S_3} = a(\chi_2 - \chi_1)$ while the corresponding distance in \mathbb{R}^3 is $L_{\mathbb{R}^3} = r_2 - r_1 = a(\sin \chi_2 - \sin \chi_1) < L_{S_3}$. ■

3-cylinders in \mathbb{R}^4

The 3-cylinder in \mathbb{R}^4 of constant radius r has horizontal cross-sections ($x^4 = x_0^4$) which are spheres of radius r , so coordinates adapted to it are just spherical coordinates in those subspaces

$$\begin{aligned} x^1 &= r \sin \theta \cos \phi, \\ x^2 &= r \sin \theta \sin \phi, \\ x^3 &= r \cos \theta, \\ x^4 &= x_0^4. \end{aligned}$$

2-cylinders in \mathbb{R}^4

The 2-cylinder in \mathbb{R}^4 of constant r has 2-plane cross-sections ($x^3 = x_0^3, x^4 = x_0^4$) which are circles of radius ρ , so coordinates adapted to it are just cylindrical coordinates in those subspaces

pseudospheres in \mathbb{M}^4

cylinders in \mathbb{M}^4

constant inertial time hypersurface

Chapter 12

Wrapping things up

12.1 Final remarks

1991

Okay, time for parting words.

One semester is so short a time. There are still many basic notions remaining, among the most important: group of transformations and their associated derivative operator—the Lie derivative. This is also important for the metric geometry we have explored—to describe symmetries of the geometry.

The language I have partially introduced you to is basic to the description of finite-dimensional continuous physical systems (and some infinite-dimensional ones too). It is interesting in its own right as pure mathematics, and a very powerful tool for describing many aspects of how our world works. I hope you have enjoyed seeing some of this structure a fraction as much as I have enjoyed the opportunity to rethink some of these ideas.

2013

Looks like finally bob found some time to incorporate groups into the main text, and some relativity and other junk. What next?

12.2 MATH 5600 Spring 1991 Differential Geometry: Take Home Final

Paraboloidal coordinates on \mathbb{R}^3

These result from a change of coordinates in the ρ - z plane of cylindrical coordinates from polar coordinates to coordinates based on two mutually orthogonal families of parabolas.

$$\begin{aligned} x &= \rho \cos \phi = \mu\nu \cos \phi \\ y &= \rho \sin \phi = \mu\nu \sin \phi \\ z &= \underbrace{z}_{\text{cylindrical}} = \underbrace{\frac{1}{2}(\mu^2 - \nu^2)}_{\text{paraboloidal}} \end{aligned}$$

ρ - z plane transformation:

$$\rho = \mu\nu \quad z = \frac{1}{2}(\mu^2 - \nu^2)$$

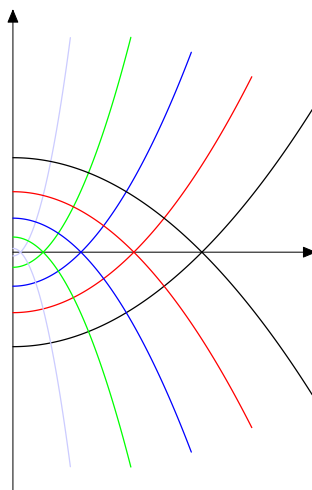


Figure 12.1: The coordinate lines for μ and ν are two mutually orthogonal families of parabolas.

Revolving this figure around the z -axis gives the 3-dimensional picture. The μ - ν coordinate surfaces are paraboloids of revolution. Compared to the original cylindrical coordinates from which these are derived, the ϕ coordinate surfaces are still the ρ - z half planes. The μ and ν coordinate lines are parabolas, while the ϕ coordinate lines are still circles about the z -axis. From the figure one can see that

$$\{e_\mu, e_\nu, e_\phi\} \equiv \left\{ \frac{\partial}{\partial \mu}, \frac{\partial}{\partial \nu}, \frac{\partial}{\partial \phi} \right\} \equiv \left\{ \frac{\partial}{\partial \bar{x}^i} \right\}$$

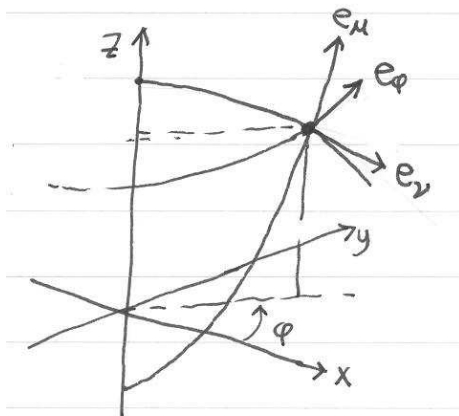


Figure 12.2: The coordinate lines and frame vectors for paraboloidal coordinates.

is a righthanded frame ($e_\mu \times e_\nu$ is along e_ϕ). The coordinate ranges are

$$\mu \geq 0, \quad \nu \geq 0, \quad 0 \leq \phi < 2\pi \quad \text{or} \quad -\pi < \phi \leq \pi.$$

1) Show that the transformation between ρ and z and μ and ν may be inverted to obtain

$$\mu = \sqrt{z + \sqrt{z^2 + \rho^2}}, \quad \nu = \sqrt{-z + \sqrt{z^2 + \rho^2}}$$

so the coordinate map is

$$\begin{aligned} \mu &= \sqrt{z + \sqrt{x^2 + y^2 + z^2}}, \\ \nu &= \sqrt{-z + \sqrt{x^2 + y^2 + z^2}}, \\ \phi &= \tan^{-1} \frac{y}{x} + \begin{cases} 0 & \text{quads: I, IV} \\ \pi & \text{quad: II} \\ -\pi & \text{quad: IV} \end{cases} \end{aligned}$$

2) Compute the transformation matrix

$$A^{-1}(\bar{x})^i_j = \frac{\partial x^i}{\partial \bar{x}^j}$$

by evaluating the differentials

$$dx^i = A^{-1}(\bar{x})^i_j d\bar{x}^j.$$

3) Since

$$\frac{\partial}{\partial \bar{x}^i} = A^{-1}(\bar{x})^j_i \frac{\partial}{\partial x^j},$$

the columns of $A^{-1}(\bar{x})$ represent the Cartesian coordinate components of the new coordinate frame vectors. Their dot products, considered as vectors in \mathbb{R}^3 give the dot products $\bar{g}_{ij} = \bar{e}_i \cdot \bar{e}_j$ of the new coordinate frame vectors $\{\bar{e}_i\} = \{\partial_\mu, \partial_\nu, \partial_\phi\}$. Show that they are orthogonal and evaluate their lengths, namely

$$(\bar{g}_{ij}) = [\underline{A}^{-1}(\bar{x})]^T \underline{A}^{-1}(\bar{x}).$$

Using these results, express the metric

$$g = \bar{g}_{ij} d\bar{x}^i \otimes d\bar{x}^j$$

in this orthogonal coordinate system.

4) Evaluate the oriented unit volume 3-form

$$\eta = dx \wedge dy \wedge dz = [\det \underline{A}^{-1}(\bar{x})] \underbrace{d\bar{x}^1 \wedge d\bar{x}^2 \wedge d\bar{x}^3}_{d\mu \wedge d\nu \wedge d\phi}$$

Since $[\det \underline{A}^{-1}(\bar{x})]$ is positive, these are oriented coordinates and $[\det \bar{g}]^{1/2} = [\det \underline{A}^{-1}(\bar{x})]$.

5) Introduce the associated orthonormal frame and its dual frame (where $(\bar{\omega}^i) \equiv (d\bar{x}^i) = (d\mu, d\nu, d\phi)$ is the orthogonal coordinate dual frame)

$$\begin{aligned} \{\bar{\mathbf{e}}_i\} &= \{e_{\hat{\mu}}, e_{\hat{\nu}}, e_{\hat{\phi}}\}, & \bar{e}_i &= (\bar{g}_{ij})^{-1/2} \bar{e}_i, \\ \{\bar{\omega}^i\} &= \{\omega^{\hat{\mu}}, \omega^{\hat{\nu}}, \omega^{\hat{\phi}}\}, & \bar{\omega}^i &= (\bar{g}_{ij})^{1/2} \bar{\omega}^i. \end{aligned}$$

Let $\underline{\mathcal{A}}(\bar{x})$ be the transformation matrix between the old and new orthonormal frames:

$$\bar{e}_i = \underline{\mathcal{A}}(\bar{x})^{-1j}{}_i \frac{\partial}{\partial x^j}, \quad \bar{\omega}_i = \underline{\mathcal{A}}(\bar{x})^i{}_j dx^j.$$

Then this orthogonal matrix is

$$\underline{\mathcal{A}}(\bar{x})^{-1i}{}_j = (\bar{g}_{jj}) A(\bar{x})^{-1i}{}_j. \quad (\text{normalize columns of } \underline{A}(\bar{x})^{-1})$$

Evaluate it explicitly.

Take its transpose to obtain $\underline{\mathcal{A}}(\bar{x})$.

Get $\underline{A}(\bar{x})$ by dividing the rows of $\underline{\mathcal{A}}(\bar{x})$ by the same normalizing factors used to multiply the columns of $\underline{A}(\bar{x})^{-1}$

$$A(\bar{x})^i{}_j = \underline{\mathcal{A}}(\bar{x})^i{}_j (\bar{g}_{ii})^{-1/2} = \frac{\partial \bar{x}^i}{\partial x^j}(x).$$

6) By differentiating the coordinate map of part 1) and re-expressing its matrix of entries in terms of the new coordinates, verify that $A(\bar{x})$ is the value obtained in 5). [Check also that $\underline{A}(\bar{x}) \underline{A}^{-1}(\bar{x}) = \underline{I}$.]

7) Compute the independent structure functions of the orthonormal frame

$$\{\bar{C}^{\hat{n}}_{\hat{j}\hat{k}}\}_{j < k}$$

defined by

$$[\bar{e}_{\hat{j}}, \bar{e}_{\hat{k}}] = \bar{C}^{\hat{i}}_{\hat{j}\hat{k}} \bar{e}_{\hat{i}}.$$

8) Compute the components of the covariant derivative in the coordinate and associated orthonormal frame using the formulas

$$\begin{aligned} \underline{A} d\underline{A}^{-1} &= \underline{\omega} = (\bar{\Gamma}^i_{kj} d\bar{x}^k) \\ \underline{\mathcal{A}} d\underline{\mathcal{A}}^{-1} &= \underline{\hat{\omega}} = (\bar{\Gamma}^{\hat{i}}_{\hat{k}\hat{j}} \bar{\omega}^{\hat{k}}) \end{aligned}$$

The entries of these matrices are called the connection 1-forms.

9) Verify these results using the formulas involving the derivatives of the metric and the structure functions.

10) Now for something new, well not new, but a putting together of things we already know. Consider the coordinate frame formula

$$\begin{aligned} R^i_{jmn} &= \partial_m \Gamma^i_{nj} - \partial_n \Gamma^i_{mj} + \Gamma^i_{m\ell} \Gamma^\ell_{nj} - \Gamma^i_{n\ell} \Gamma^\ell_{mj} \\ &= 2\partial_{[m} \Gamma^i_{n]j} + 2\Gamma^i_{[m|\ell|} \Gamma^\ell_{n]j} = R^i_{j[mn]}, \end{aligned}$$

where $|\ell|$ means to not include this index in the antisymmetrization, and define

$$\begin{aligned} \Omega^i_j &\equiv \frac{1}{2} R^i_{jmn} dx^{mn} = \frac{1}{2} \underbrace{[2\partial_{[m} \Gamma^i_{n]j} dx^{mn} + 2\Gamma^i_{[m\ell} \Gamma^\ell_{n]j} dx^{mn}]}_{[d\omega^i_j]_{mn}} \\ &= d\omega^i_j + \omega^i_\ell \wedge \omega^\ell_j. \end{aligned}$$

By introducing a curvature 2-form matrix $\underline{\Omega} = (\Omega^i_j)$ one can more efficiently compute the curvature tensor components using matrices

$$\underline{\Omega} = d\underline{\omega} + \underline{\omega} \wedge \underline{\omega}.$$

where the combined wedge and matrix product means multiply the matrices keeping the factor ordering of the 1-form entries, and wedge them in the product matrix entries. From the matrix $\underline{\Omega}$, one can read off the curvature tensor components: the matrix indices give the left pair of tensor indices, while the coefficients of dx^{mn} give the second pair.

If we use this in the new coordinate frame then

$$\bar{\underline{\Omega}} = d\bar{\underline{\omega}} + \bar{\underline{\omega}} \wedge \bar{\underline{\omega}} = \underbrace{d(\underline{A} d\underline{A}^{-1}) + \underline{A} d\underline{A}^{-1} \wedge \underline{A} d\underline{A}^{-1}}_{d\underline{A} \wedge d\underline{A}^{-1}}.$$

But

$$\begin{aligned} \underline{A} \underline{A}^{-1} &= \underline{I} \rightarrow [d\underline{A} \underline{A}^{-1} + \underline{A} d\underline{A}^{-1} = 0] \underline{A}, \\ &\rightarrow d\underline{A} + \underline{A} d\underline{A}^{-1} \underline{A} = 0 \quad \rightarrow \quad d\underline{A} = -\underline{A} d\underline{A}^{-1} \underline{A}, \end{aligned}$$

so

$$d\underline{A} \wedge d\underline{A}^{-1} = -\underline{A} d\underline{A}^{-1} \underline{A} \wedge d\underline{A}^{-1} = -\underline{A} d\underline{A}^{-1} \wedge \underline{A} d\underline{A}^{-1}$$

where the wedge can be anywhere between the differentials since the scalar matrix factors don't interfere with it. Thus $\underline{\Omega} = 0$. Of course we knew the curvature tensor to be zero, but this matrix method most efficiently achieves this result.

Note: if $\underline{\omega} \wedge \underline{\omega}$ bothers you, here is an example of wedge multiplying 2×2 matrices of 1-forms

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \wedge \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} \alpha \wedge A + \beta \wedge C & \alpha \wedge B + \beta \wedge D \\ \gamma \wedge A + \delta \wedge C & \gamma \wedge B + \delta \wedge D \end{pmatrix},$$

where all the entries are assumed to be 1-forms (or even p -forms).

11) You only had to follow 10), not do anything. Now, from your results for $\bar{\Gamma}^i_{jk}$ you can read off the components of the covariant derivative for the ν coordinate surfaces (upturned parabolas of revolution) with the metric

$$^{(2)}g = g|_{\nu=\nu_0, d\nu=0}$$

on which μ, ϕ are local coordinates. [These components can only be defined by the 2-dimensional formula in terms of the metric derivatives.]

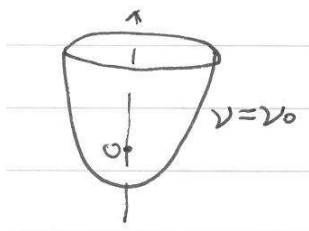


Figure 12.3: The ν coordinate surfaces are upturned parabolas of revolution.

Evaluate the 2-dimensional matrix

$$^{(2)}\underline{\omega} = (^{(2)}\underline{\omega}^\alpha_\beta) = (^{(2)}\Gamma^\alpha_{\gamma\beta} d\bar{x}^\gamma),$$

where the indices $\alpha, \beta, \dots = \mu, \phi$ correspond numerically to 1,3 in terms of the original 3 coordinates.

Next compute the corresponding 2-form matrix

$$^{(2)}\underline{\Omega} = d^{(2)}\underline{\omega} + ^{(2)}\underline{\omega} \wedge ^{(2)}\underline{\omega} = (\tfrac{1}{2} ^{(2)}R^\alpha_{\beta\gamma\delta} d\bar{x}^\gamma \wedge d\bar{x}^\delta).$$

Read off the two nonzero components

$$^{(2)}R^\mu_{\phi\mu\phi}, \quad ^{(2)}R^\phi_{\mu\mu\phi}.$$

Does

$$^{(2)}R_{\mu\phi\mu\phi} = -^{(2)}R_{\phi\mu\phi\mu}?$$

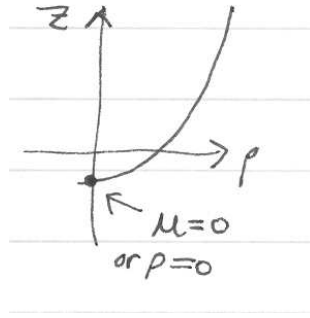


Figure 12.4: The μ coordinate lines are upturned half parabolas.

12) Evaluate ${}^{(2)}R^{\hat{\mu}}_{\hat{\phi}\hat{\mu}\hat{\phi}} = (\bar{g}_{\phi\phi})^{-1} {}^{(2)}R^{\mu}_{\phi\mu\phi}$.

What is its value at $\mu = 0$, the vertex of the parabola of revolution?

The parabola (μ coordinate line) which is revolved around the z axis, when expressed in terms of the cylindrical coordinates ρ, z , is

$$\begin{aligned}\rho &= \mu\nu_0 \quad \text{or} \quad \mu = \rho/\nu_0, \\ z &= \frac{1}{2}(\mu^2 - \nu_0^2) = \frac{1}{2}(\rho^2/\nu_0^2 - \nu_0^2).\end{aligned}$$

Since these are Cartesian coordinates in the ρ - z plane, we can use the multivariable calculus plane curve curvature formula to evaluate the curvature of this parabola at any point

$$\mathcal{K} = \frac{|d^2z/d\rho^2|}{[1 + (dz/d\rho)^2]^{3/2}}.$$

Evaluate $\mathcal{K}(\rho = 0)$ and compare it to the value of the single independent orthonormal component ${}^{(2)}R^{\hat{\mu}}_{\hat{\phi}\hat{\mu}\hat{\phi}}$ of the 2-dimensional curvature tensor. Do you notice any relationship?

13) Show that the μ coordinate lines are geodesics on these parabolas of revolution, but that the ϕ coordinate lines are not.

14) What is the single independent structure function $C^{\hat{\phi}}_{\hat{\mu}\hat{\phi}}$ for the 2-dim orthonormal frame?

Use it to compute the components of the covariant derivative in the orthonormal frame

$${}^{(2)}\Gamma^{\alpha}_{\beta\gamma}.$$

Use them to show that $e_{\hat{\mu}}$ and $e_{\hat{\phi}}$ are parallel transported along the μ coordinate lines.

15) All of these computations (with the exception of the curvature 2-form notation) have been done with either cylindrical or spherical coordinates in the notes, so you should have no problem if you understand them.

16) Let

$$\begin{cases} X = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y} + (x^2 + y^2 + z^2)\frac{\partial}{\partial z} \\ X^{\flat} = -ydx + xdy + (x^2 + y^2 + z^2)dz \end{cases}$$

Evaluate X^b in paraboloidal coordinates. Find X in these coordinates.

Evaluate $\nabla_{e_\mu} X$.

17) Let Σ be the 2-surface

$$\begin{cases} \nu = \nu_0 \\ 0 \leq \nu \leq \nu_0 \end{cases}$$

parametrized by $\{\nu, \phi\}$.

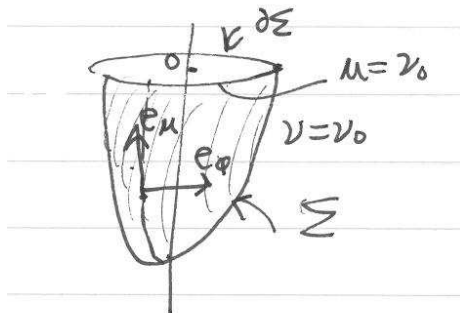


Figure 12.5: The ν coordinate surfaces $\nu = \nu_0$ are upturned parabolas of revolution: this surface Σ with boundary circle $\mu = \nu_0$ corresponds to $0 \leq \mu \leq \nu_0$.

What choice of normal does this inner orientation imply by the right hand rule, the inward/upward normal or the outward/downward normal?

Looking down from above, what is the induced orientation of $\partial\Sigma$: clockwise or counter-clockwise?

Following the example in the Stokes' theorem section, verify Stokes' theorem for this surface with boundary

$$\int_{\partial\Sigma} X^b = \int_{\Sigma} dX^b$$

for the 1-form X^b of part 16).

18) That's all folks. Have fun. Stop by if you have any difficulty. I need your work by 5pm Friday May 3, 1991 to make up grades for Monday.

Appendices

Appendix A

From trigonometry to hyperbolic functions and hyperbolic geometry

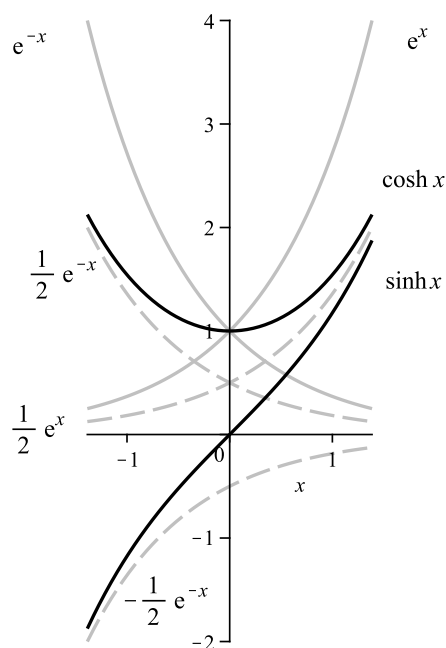


Figure A.1: The basic hyperbolic functions \cosh and \sinh are even and odd combinations of the basic increasing and decreasing exponentials, here shown together with their three asymptotic exponentials as well.

The hyperbolic cosine and sine are respectively even and odd functions, like their trigonometric counterparts

$$\cosh(-x) = \cosh x, \quad \sinh(-x) = -\sinh x,$$

and represent the even and odd combinations of the basic growing and decaying exponential

functions e^x and e^{-x}

$$\cosh x = \frac{1}{2}(e^x + e^{-x}), \quad \sinh x = \frac{1}{2}(e^x - e^{-x}).$$

This implies that their power series representation must also consist of only even and odd terms respectively, and when one combines term by term the power series for the two exponential function terms, indeed one finds that the odd and even powers respectively simply cancel out while the even and odd terms respectively are retained from the power series representation of the exponential function itself

$$\begin{aligned} e^x &= \sum_{n=0}^{\infty} \frac{x^n}{n!} = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} + \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!} = \cosh x + \sinh x, \\ e^{-x} &= \sum_{n=0}^{\infty} (-1)^n \frac{x^n}{n!} = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} - \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!} = \cosh x - \sinh x, \\ \cosh x &= \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}, \quad \sinh x = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!}. \end{aligned}$$

Like their trigonometric counterparts their derivatives interchange them but without a sign change in one of the formulas, even simpler, as easily follows from inspection differentiating their definitions

$$\frac{d}{dx} \cosh x = \sinh x, \quad \frac{d}{dx} \sinh x = \cosh x.$$

The basic hyperbolic identity, which differs only by a sign from the corresponding trigonometric identity, is a trivial consequence of the laws of exponents

$$\cosh^2 x - \sinh^2 x = \left(\frac{e^x + e^{-x}}{2} \right)^2 - \left(\frac{e^x - e^{-x}}{2} \right)^2 = \dots = 1$$

Exercise A.0.1.

Fill in the dots in the previous calculation and confirm the derivative formulas for the hyperbolic cosine and sine. ■

Every formula and identity in trigonometry is mirrored exactly with a crucial change in sign by those of hyperbolic geometry. The remaining hyperbolic functions are defined by the same ratios as in the trigonometric case

$$\begin{aligned} \tanh x &= \frac{\sinh x}{\cosh x}, \quad \coth x = \frac{\cosh x}{\sinh x} = \frac{1}{\tanh x}, \\ \operatorname{csch} x &= \frac{1}{\cosh x}, \quad \operatorname{sech} x = \frac{1}{\sinh x}. \end{aligned}$$

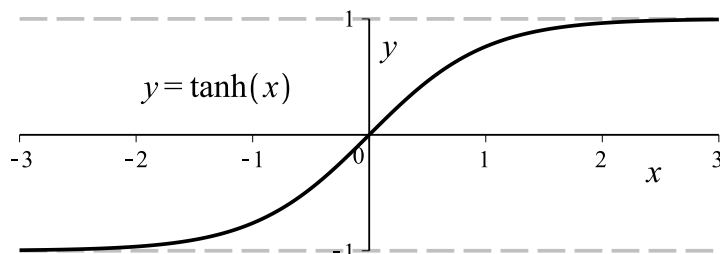


Figure A.2: The hyperbolic tangent is an odd function which interpolates between its asymptotic values of -1 and 1 .

The inverse hyperbolic functions can be re-expressed in terms of the inverse of the exponential function, namely the natural logarithm \ln . The derivatives of all of these functions are easily derived and appear in most calculus textbooks, although this section is often skipped in practice. The hyperbolic tangent is an odd function which interpolates between its asymptotic values of -1 and 1 , which it approaches pretty quickly.

Exercise A.0.2.

All the hyperbolic trigonometric functions can be inverted by solving a quadratic relationship in $(e^x)^2$ and then using the inverse of the exponential function: the natural logarithm \ln . For example, choosing $x > 0$ to get a 1-1 relationship and then the positive sign in the quadratic formula, one finds

$$\begin{aligned} 0 &= \left(\frac{e^x + e^{-x}}{2} - y \right) 2e^x = (e^x)^2 - (2y)e^x + 1 \rightarrow e^x = y + \sqrt{y^2 - 1} \\ &\rightarrow x = \ln \left(y + \sqrt{y^2 - 1} \right) = \cosh^{-1} y. \end{aligned}$$

From the definition of the hyperbolic tangent, clearing fractions and again multiplying by e^x yields a quadratic equation in e^x which can be solved similarly to show that for $|x| < 1$

$$\tanh^{-1} x = \frac{1}{2} \ln \frac{1+x}{1-x}.$$

Show this. ■

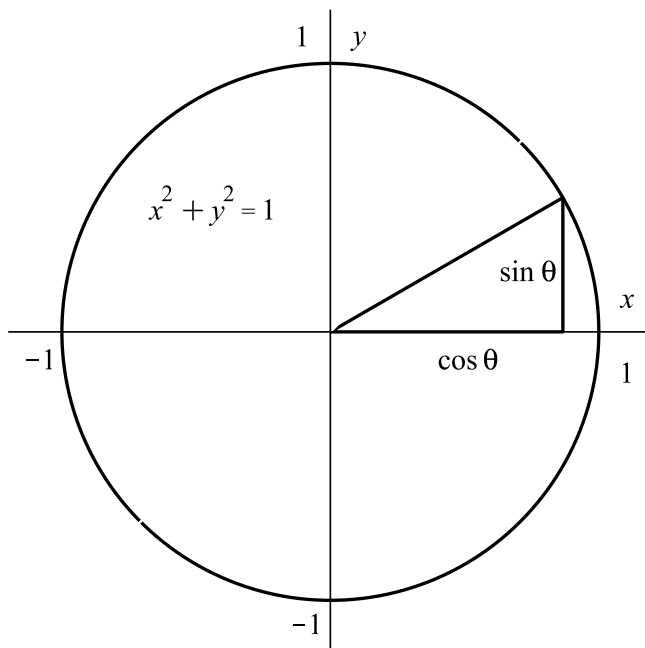


Figure A.3: Trigonometry flows from the geometry of the unit circle. The reference triangle sitting on the horizontal axis can be in one of four quadrants.

The basic geometry of trigonometry comes from the unit circle $x^2 + y^2 = 1$ in the plane (see figure A.3), and with a simple sign change, one gets hyperbolic geometry from the two analogous unit hyperbolas $x^2 - y^2 = \pm 1$ which have their symmetry axis either horizontal (+) or vertical (−). This geometry illustrated in figure A.4 covers four disjoint regions into which the plane is separated by the degenerate hyperbolas $x^2 - y^2 = 0$, the latter of which play a special role in the global hyperbolic geometry of the whole plane. Just as the trigonometric angle θ parametrizes the additive group of rotations of the unit circle, the hyperbolic angle α parametrizes the additive group of hyperbolic rotations (pseudorotations) of the two unit hyperbolas.

In the same way that polar coordinates in the plane are adapted to the trigonometric geometry of the Euclidean plane (see figure A.5), pseudo-polar coordinates can be introduced in an analogous way to adapt to the hyperbolic geometry but we need four separate coordinate patches to cover the four sectors into which the plane is separated by that geometry, two of which are illustrated in figure A.6. By using a signed pseudo-radial coordinate, only two patches are necessary: inside and outside the “cone” formed by the two oblique asymptotes to the hyperbola. The hyperbolic tangent is the ratio of the vertical to the horizontal leg of the reference triangle of a point, and hence approaches 1 in absolute value as a point approaches these asymptotes.

Exercise A.0.3. pseudo-spherical coordinates

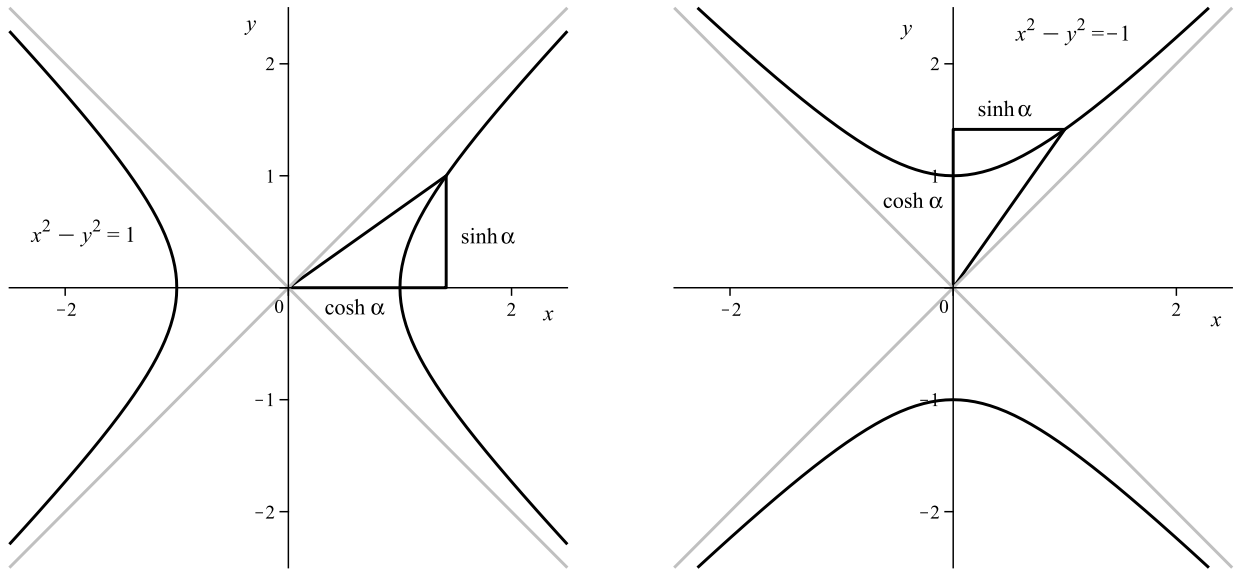


Figure A.4: Hyperbolic geometry flows from the geometry of the two unit hyperbolas. The horizontal hyperbola geometry (left) has four different sectors for the reference triangle sitting on the horizontal axis. The vertical hyperbola geometry (right) has four different sectors for the reference triangle sitting on the vertical axis.

Consider the two unit hyperbolas in the x - y plane shown in Fig. A.4.

a) By substituting the differentials of $x = r \cos \theta, y = r \sin \theta$ into the squared differential of arclength $ds^2 = dx^2 + dy^2$, the result simplifies to $ds^2 = dr^2 + r^2 d\theta^2$ using the fundamental trigonometric identity. Show this.

b) Now repeat for $x = \ell \cosh \alpha, y = \ell \sinh \alpha$ and $ds^2 = dx^2 - dy^2 = d\ell^2 - \ell^2 d\alpha^2$ using the fundamental hyperbolic identity.

c) Then repeat for $x = \tau \sinh \alpha, y = \tau \cosh \alpha$ and $ds^2 = dx^2 - dy^2 = -d\tau^2 + \tau^2 d\alpha^2$ using the fundamental hyperbolic identity.

■

The rotations of the plane are easily expressed using the addition formulas for the sine and cosine. If we start with a point $(x_0, y_0) = (r_0 \cos \theta_0, r_0 \sin \theta_0)$ in the plane and rotate it by an angle θ , we simply add θ to its polar angle θ_0

$$\begin{aligned}
 (x_0, y_0) &= (r_0 \cos \theta_0, r_0 \sin \theta_0) \rightarrow \\
 (x, y) &= (r_0 \cos(\theta_0 + \theta), r_0 \sin(\theta_0 + \theta)) \\
 &= (r_0(\cos \theta_0 \cos \theta - \sin \theta_0 \sin \theta), r_0(\sin \theta_0 \cos \theta + \cos \theta_0 \sin \theta)) \\
 &= ((r_0 \cos \theta_0) \cos \theta - (r_0 \sin \theta_0) \sin \theta, (r_0 \sin \theta_0) \cos \theta + (r_0 \cos \theta_0) \sin \theta) \\
 &= (x_0 \cos \theta - y_0 \sin \theta, x_0 \sin \theta + y_0 \cos \theta) ,
 \end{aligned}$$

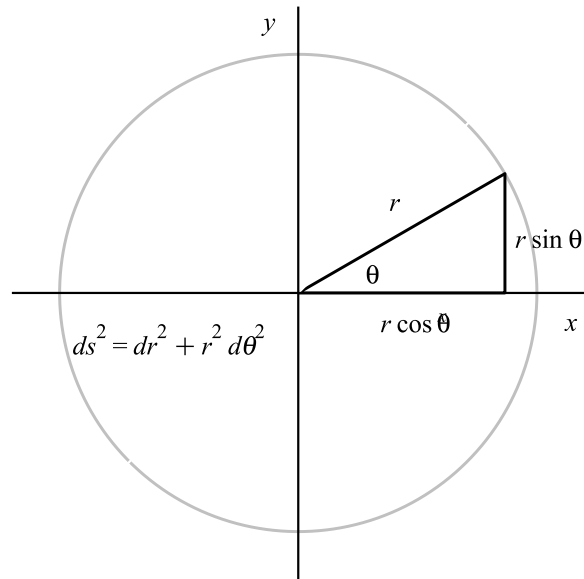


Figure A.5: Polar coordinates in the plane are adapted to its Euclidean geometry.

or in matrix form

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

Exercise A.0.4.

Verify that the set of rotation matrices for all values of the angle of rotation form a group by showing that their matrix products amount to addition of the angles (closure under matrix multiplication), hence the inverse rotation corresponds to the sign-reversed angle and associativity is guaranteed by associativity of addition of real numbers

$$\begin{pmatrix} \cos \theta_3 & -\sin \theta_3 \\ \sin \theta_3 & \cos \theta_3 \end{pmatrix} = \begin{pmatrix} \cos \theta_1 & -\sin \theta_1 \\ \sin \theta_1 & \cos \theta_1 \end{pmatrix} \begin{pmatrix} \cos \theta_2 & -\sin \theta_2 \\ \sin \theta_2 & \cos \theta_2 \end{pmatrix}, \quad \text{where } \theta_3 = \theta_1 + \theta_2.$$

This group is called the special orthogonal group in 2 real dimensions, symbolized by $SO(2, R)$. ■

Exercise A.0.5.

a) Verify the addition formulas for the hyperbolic cosine and sine by simply re-expressing the left and right hand sides in exponential notation and expanding the right hand side out using rules of exponents

$$\begin{aligned} \cosh(\alpha_1 + \alpha_2) &= \cosh(\alpha_1) \cosh(\alpha_2) + \sinh(\alpha_1) \sinh(\alpha_2), \\ \sinh(\alpha_1 + \alpha_2) &= \sinh(\alpha_1) \cosh(\alpha_2) + \cosh(\alpha_1) \sinh(\alpha_2). \end{aligned}$$

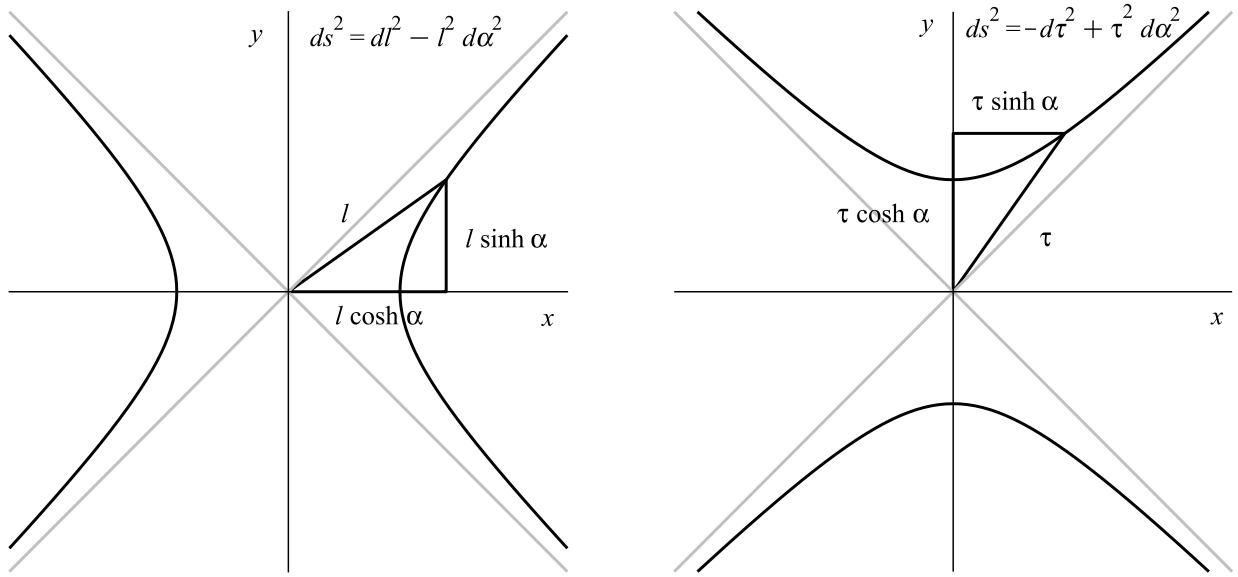


Figure A.6: Pseudopolar coordinates in the plane are adapted to its hyperbolic geometry. Four different coordinate systems are required for the four sectors of the hyperbolic plane, two of which are shown here. The same formulas with a minus sign multiplying the cosh term works in the other two disjoint sectors obtained by reflection through the origin, while allowing negative values of the hyperbolic angle α to handle the sinh term.

b) Now take the quotient of the left right hand sides (second over first) to obtain the hyperbolic tangent on the left and then divide all four terms in the right hand side quotient by $\cosh(\alpha_1) \cosh(\alpha_2)$ to then re-express each of them in terms of the hyperbolic tangent to get its addition formula

$$\tanh(\alpha_1 + \alpha_2) = \frac{\tanh \alpha_1 + \tanh \alpha_2}{1 + \tanh \alpha_1 \tanh \alpha_2}.$$

If we let $v_i = \tanh(\alpha_i)$, this becomes the formula for the relativistic addition of velocities

$$v_3 = \frac{v_1 + v_2}{1 + v_1 v_2}.$$

This is relevant to the vertical hyperbola geometry of Fig. A.4, as will be explained later.

c) For a point in the right horizontal sector of the plane for which an initial point can be represented as $(x_0, y_0) = (\ell_0 \cosh \alpha_0, \ell_0 \sinh \alpha_0)$, a hyperbolic rotation applied to the point consists of simply adding a hyperbolic angle α to its pseudoangle α_0 . Repeat the calculation done above for an ordinary rotation for this new situation. Show that the final result is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cosh \alpha & \sinh \alpha \\ \sinh \alpha & \cosh \alpha \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

Confirm that these matrices also form a group under matrix multiplication law with the same additive law for the hyperbolic angle. Notice that the determinant of such matrices is identically

1 by the fundamental hyperbolic identity. This group turns out to be a subgroup of the special linear group in 2 real dimensions, denoted by $SL(2, R)$, consisting of all matrices with unit determinant.

d) Repeat c) for a point in the top sector of the plane where an initial point can be represented as $(x_0, y_0) = (\tau_0 \sinh \alpha_0, \tau_0 \cosh \alpha_0)$, and show what happens when α_0 is changed to $\alpha_0 + \alpha$.

If we rename $y \rightarrow t$, $\cosh \alpha \rightarrow \gamma$, $\sinh \alpha \rightarrow \gamma v$ so that $\tanh \alpha \rightarrow v$ as above, and interchange the order of the two coordinates, show that the previous matrix takes the form of a so called Lorentz transformation in the 2-dimensional spacetime with time coordinate t and space coordinate x relevant for 1-dimensional motion.

$$\begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} \gamma & \gamma v \\ \gamma v & \gamma \end{pmatrix} \begin{pmatrix} t_0 \\ x_0 \end{pmatrix} = \begin{pmatrix} \gamma(t_0 + vx_0) \\ \gamma(vt_0 + x_0) \end{pmatrix}.$$

■

Under a rotation of the plane by a fixed angle, as described by the matrix multiplication above, each point of a circle centered at the origin moves around that circle by that same angle as illustrated in figure A.7, where a reference triangle with a given angle with respect to each of the four axes is shown before and after a rotation. Similarly, a hyperbolic rotation (or just “pseudo-rotation”) of the entire plane by a fixed hyperbolic angle is defined by the matrix transformation of this previous exercise derived for the right horizontal sector and the top vertical sector. Under matrix multiplication, all points along the pair of hyperbolas $x^2 = y^2 = C$ for a fixed value of C are moved along as shown in figure A.8 from the 8 standard positions located at a given hyperbolic angle from each of the four axes. For example on the upper branch of the vertical hyperbola, points move to the right, and on the lower branch to the left. Similarly on the horizontal hyperbola, points on the right branch move up, and on the left branch move down. Points on the asymptotes move along those asymptotes in the directions indicated.

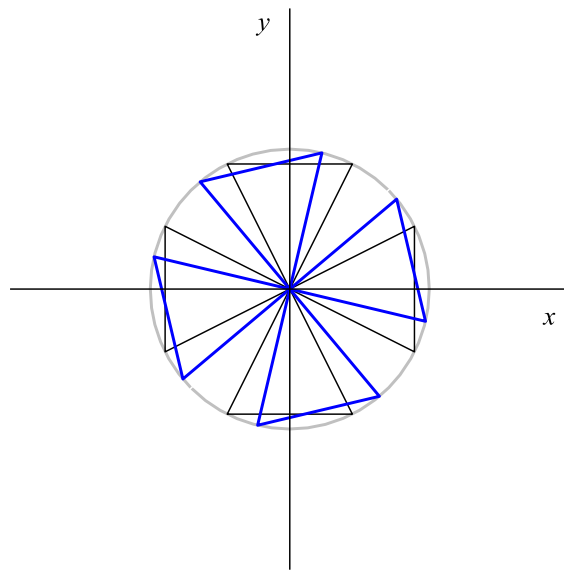


Figure A.7: A rotation of the plane by an angle θ moves each point of a circle centered at the origin along that circle by that angle.

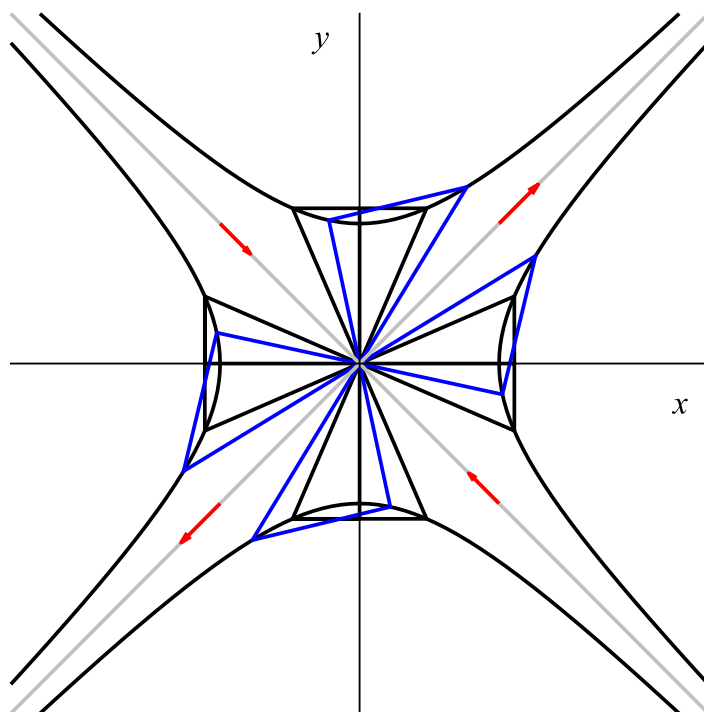


Figure A.8: A pseudo-rotation of the plane by a hyperbolic angle α moves each point of a given branch of the pair of hyperbolas $x^2 - y^2 = \pm C$ along that hyperbola as shown for the 8 different standard reference triangle positions, while stretching or shrinking respectively points located on the common asymptotes $y = \pm x$.

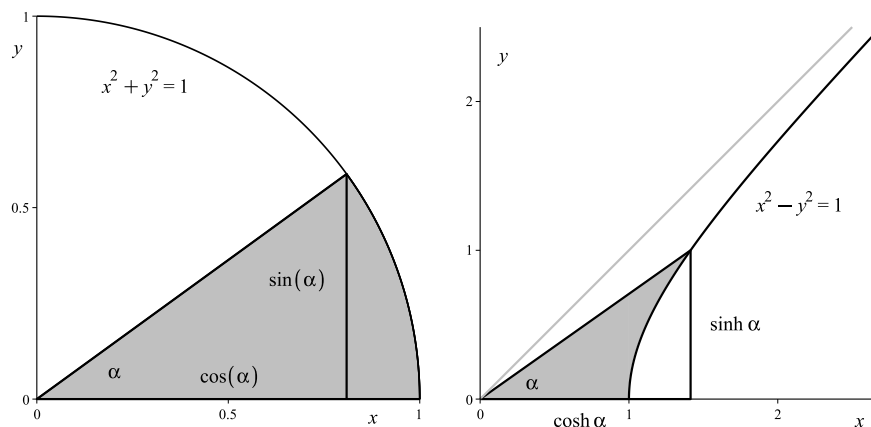


Figure A.9: The area of a sector of the unit circle of angle α equals half the angle. Similarly the area of an analogous hyperbolic sector of a unit hyperbola of hyperbolic angle α equals half that angle.

Exercise A.0.6.

Consider the geometry of the unit circle and a unit hyperbola in the x - y plane as shown in Fig. A.9.

a) As an exercise in double integration with a computer algebra system, you can set up an integral for the area of the sector of the unit circle and the corresponding sector of a unit hyperbola as a single iterated integral in Cartesian coordinates $\int_a^b \int_{f(y)}^{g(y)} 1 \, dx \, dy$ (the x integration is trivial), first integrating with respect to x from the bounding ray of the sector to the circle/hyperbola, then integrating over y from $y = 0$ to $y = \sin \alpha$ or $y = \sinh \alpha$ respectively. Note that the bounding ray has the equation $x = y \cot \alpha$ or $x = y \coth \alpha$ respectively, while the conic section has the equation $x = \sqrt{1 \pm y^2}$. Confirm these statements.

Maple won't simplify its result in the trigonometric case unless you substitute $\cos \alpha$ for the absolute value $|\cos \alpha|$, which is true for a positive acute angle: `subs(|cos(alpha)| = cos(alpha))`, although you can simplify this by eye to $\alpha/2$ easily.

Maple won't evaluate the outer integral in the hyperbolic case unless you separate it into a difference of two integrals (the two terms that result from the trivial inner integral), and then use `simplify(%,symbolic)` to get the desired result $\alpha/2$.

b) In polar coordinates $x = r \cos \theta$, $y = r \sin \theta$ we can use the determinant of the Jacobian of the coordinate transformation to get this result more easily:

$$dA = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} dr \, d\theta = r \, dr \, d\theta.$$

Setting up the iterated integral in polar coordinates leads easily to the familiar result (check

it!)

$$A = \int_0^\alpha \int_0^1 r \, dr \, d\theta = \frac{1}{2}\alpha.$$

Now repeat for the hyperbolic case with the coordinates $x = \ell \cosh \beta$, $y = \ell \sinh \beta$ in terms of which $0 \leq \ell \leq 1$, $0 \leq \beta \leq \alpha$ describes the hyperbolic sector of hyperbolic angle α . Show that by evaluating the new Jacobian determinant we get the result $dA = \ell d\ell d\beta$, and its double integral easily gives the result $\alpha/2$.

The result is simple because like polar coordinates, these are orthogonal coordinates even though there is a minus sign in the arc length formula, as shown in Fig. A.6, and as evaluated in Exercise A.0.3. The differential of area $dA = d\ell (\ell d\beta)$ is just the product of the differentials of arclength $d\ell$ in the ℓ direction and $\ell d\alpha$ in the α direction (pseudo-radius of the arc times the differential of the hyperbolic angle), directly analogous to the Euclidean case: $dA = dr (r d\theta)$. The area of a unit square $0 \leq x \leq 1$, $0 \leq y \leq 1$ in either the Euclidean geometry $ds^2 = dx^2 + dy^2$ or the Lorentzian geometry $ds^2 = dx^2 - dy^2$ is 1, the product of the lengths of the orthogonal unit length sides of the square. This means that the area of a region of the plane does not depend on this sign change. ■

Appendix B

Hyperbolic geometry and special relativity

So what does the hyperbolic geometry of the plane have to do with special relativity? Consider 1-dimensional motion along an x axis as studied first in high school physics. We can make a diagram of position along the axis versus the time t , but instead of having the time as the horizontal axis as usual, we agree that the time axis will always be the vertical axis to track the passage of time as we move up in our diagrams. This “Lorentz plane” represents the set of all events, namely all points on the spatial axis at all times.

The path of a point moving along the x -axis is then represented by a curve in the t - x plane diagram which is called its world line. World lines for motion at constant velocity are straight lines in this diagram, with constant velocity $dx/dt = v$, which is the reciprocal of the slope of the line in the spacetime diagram. Motion of real bodies must have speed less than the speed c of light, so if we introduce the new time variable $T = ct$ (which has the dimensions of length, “speed times time = distance”) then

$$\left| \frac{dx}{dt} \right| = |v| \leq c \text{ or } \left| \frac{dx}{dT} \right| = \left| \frac{1}{c} \frac{dx}{dt} \right| = \left| \frac{v}{c} \right| \leq 1$$

Using the new time variable T just corresponds to using length units for time so that space and time have the same units, and with this time, velocity becomes a dimensionless quantity and motion at the speed of light means unit speed in these units

$$|v| = c \rightarrow \left| \frac{dx}{dT} \right| = 1,$$

while motion of any real bodies must have a proper fractional speed $|dx/dT| \leq 1$, with strict inequality holding for the world lines of inertial observers: imagined observers traveling at constant velocity. Rather than having to use a new symbol T for this new time coordinate in terms of which the speed of light is 1, we can just continue using the symbol t but assume that it is measured in length units by using the speed of light as a conversion factor. This is not unusual. When we animate an arclength parametrized curve in a computer algebra system, we equate the mathematical arclength to time in seconds. If we plot a curve of total length

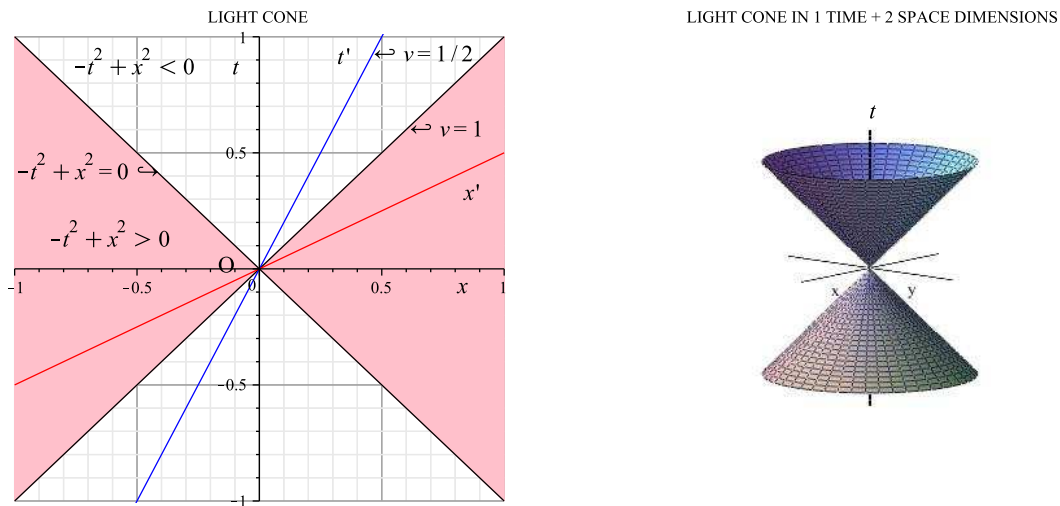


Figure B.1: **Left:** the t - x plane spacetime diagram for motion in 1 spatial dimension. The light cone $-t^2 + x^2 = 0$ divides the plane of 1 time plus 1 space dimension into regions representing events which are separated from the origin by timelike, lightlike and spacelike separations according to $-t^2 + x^2 > 0, = 0, < 0$. **Right:** the light cone in 1 time plus 2 space dimensions, a true cone, which generalizes to the 3-cone in 4-dimensional spacetime: an ordinary 2-sphere expanding at the speed of light from the event at the origin. World lines of inertial observers passing through the origin are confined to the interior of the light cone.

10, selecting 10 frames per second, and choosing the number of frames to be 100, then the animation which traces out the curve in space will last 10 seconds. Thus we have converted length into time with a conversion factor of 1 in time units of seconds. Converting time into length goes in the opposite direction, but the large value of the speed of light in centimeters per second compresses the time axis incredibly compared to ordinary time scales.

If we consider only inertial observers whose world lines pass through the origin of our coordinates, then their straight world lines through the origin are confined to the “interior” $|x| < |t|$ of the light cone: $x = \pm t$, namely the white region in Fig. B.1. For any point or “event” in the future interior of the light cone, the straight line connecting it to the origin is the world line of an inertial observer for whom the event occurs at the same spatial location as the observer in the observer’s own reference frame but at a later time compared to the event at the origin. This can be encoded into the “dot product” we use on the plane for 2-vectors $\langle t, x \rangle$ representing the position of events relative to the chosen origin by introducing a minus sign relative to the usual Euclidean dot product

$$\langle t_1, x_1 \rangle \bullet \langle t_2, x_2 \rangle = -t_1 t_2 + x_1 x_2,$$

so that instead of the distance formula $s^2 = x^2 + y^2$ for a displacement from the origin we get the “spacetime interval”

$$s^2 = \langle t, x \rangle \bullet \langle t, x \rangle = -t^2 + x^2 = \begin{cases} < 0 & \text{timelike separation,} \\ = 0 & \text{lightlike or null separation,} \\ > 0 & \text{spacelike separation.} \end{cases}$$

This Lorentz inner product governs the geometry of special relativity.

Consider the coordinate grid associated with the inertial coordinates (t, x) shown as dashed lines in Fig. B.2. The vertical lines of constant x (time coordinate lines) represent the world lines of points at rest with respect to the inertial observer of the coordinate system (t, x) , while the horizontal lines of constant t (the x coordinate lines) represent a moment of time in the reference frame of this inertial observer, or a simultaneity slice of the plane, consisting of all events which appear to be simultaneous to that inertial observer: “space” at a moment of time. The only way the speed of light can be 1 in both reference frames is if the new simultaneity slices are orthogonal (with respect to the Lorentz dot product) to the new time lines in the inertial coordinates associated with the moving observer. The new coordinate grid is simply related by a hyperbolic rotation with boost parameter $v = \tanh \alpha$, taken to be 1/2 in this spacetime diagram, called a Lorentz transformation. The new simultaneity slices have this slope. The fact that the Lorentz dot product and the spacetime separation are left invariant by a hyperbolic rotation means that

$$-t^2 + x^2 = -(t')^2 + (x')^2 = 0 \leftrightarrow \frac{dx}{dt} = \pm 1 = \frac{dx'}{dt'},$$

i.e., the speed of light remains the same in all such inertial coordinate systems. This is the heart of special relativity and the big difference with Newtonian nonrelativistic theory: in the latter the simultaneity slices never change, always remaining horizontal. Only the world lines of

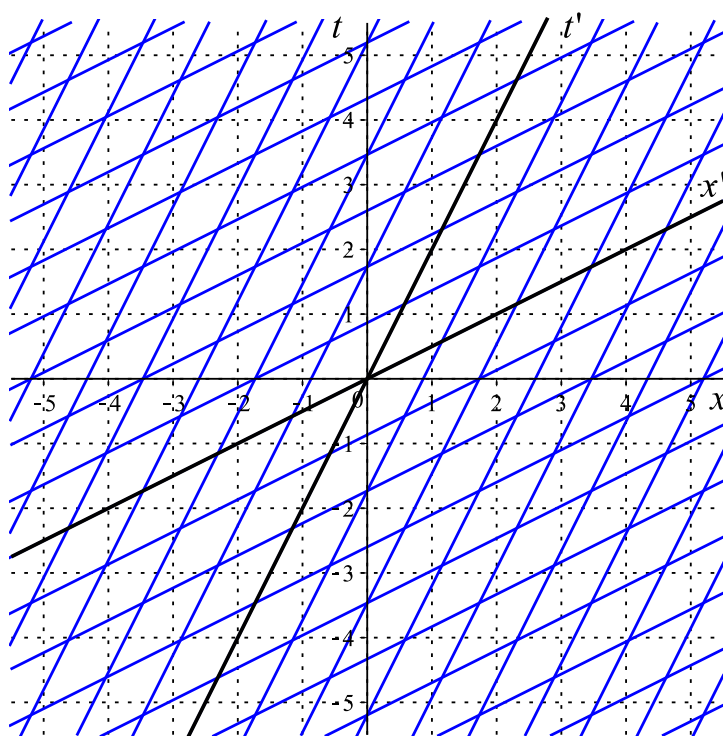


Figure B.2: The old dashed line (t, x) and new solid line (t', x') inertial coordinate grids on the Lorentz plane. The primed observer grid is moving at speed $1/2$ in the forward x direction (slope 2 in the spacetime diagram) with respect to the unprimed observer.

the points at constant relative velocity change: time was absolute in the old theory. In special relativity the notion of simultaneity depends on the observer.

The relation between the old and new coordinates is

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = \begin{pmatrix} \cosh \alpha & -\sinh \alpha \\ -\sinh \alpha & \cosh \alpha \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma v \\ -\gamma v & \gamma \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix},$$

where

$$[\cosh \alpha, \sinh \alpha, \tanh \alpha] = [\gamma, \gamma v, v]$$

and

$$\gamma = \frac{1}{\sqrt{1-v^2}} = \cosh \alpha \geq 1$$

is the associated gamma factor. This hyperbolic rotation of the grids explains length contraction and time dilation. It also gives meaning to the spacetime interval. For any point (t, x) in the future light cone of the origin at a timelike separation from the origin, there exists an inertial observer moving at relative velocity $v = x/t$. This inertial observer sees the event as occurring at the same location as the origin O, but at a time $\tau = \sqrt{t^2 - x^2}$. For any point outside the light cone at a spacelike separation from the origin, there exists an inertial observer with speed $v = t/x$ such that the two events occur simultaneously but at a distance $\ell = \sqrt{x^2 - t^2}$. In other words

$$s^2 = -t^2 + x^2 = \begin{cases} -\tau^2 & \text{timelike separation,} \\ 0 & \text{lightlike or null separation,} \\ \ell^2 & \text{spacelike separation.} \end{cases}$$

The time separation $\tau > 0$ is called the proper time between the events along the straight line segment between them, while the spatial distance separation $\ell > 0$ is called the proper distance between the events along the straight line segment between them. Events separated by 0 spacetime interval can be connected by a light signal.

Consider first length contraction. Fig. B.3 shows the world line of the two ends of a unit length 1-dimensional ruler at the tickmarks $x' = 0, 1$ of the primed coordinate grid (so that $\Delta x' = 1$) moving with velocity $v = 0.5$ with respect to the unprimed frame, and gamma factor of $\gamma = 1.155, \gamma^{-1} = 0.866$. In its rest frame it is a $L' = \Delta x' = 1$ length ruler (spacetime interval between event O and event C) but in the laboratory frame, measuring its two ends at the same time $t = 0$, for example, it appears to be

$$L = \Delta x = \gamma^{-1} L' = 0.866$$

in length (spacetime interval between event O and event B), exhibiting “Lorentz length contraction.”

The primed observer sees the unprimed observer at $x = 0$ cross from the right end of the ruler at $t = -2$ (event A) to the left end of the ruler at $t = 0$ (event O, see the right figure, it takes 2 time units to move 1 length unit at speed $1/2$), while the unprimed observer sees the right end of the ruler pass at $t = -2/\gamma = -1.732$ and the left end pass at $t = 0$ (it takes less time for the contracted ruler to pass traveling at speed $1/2$).

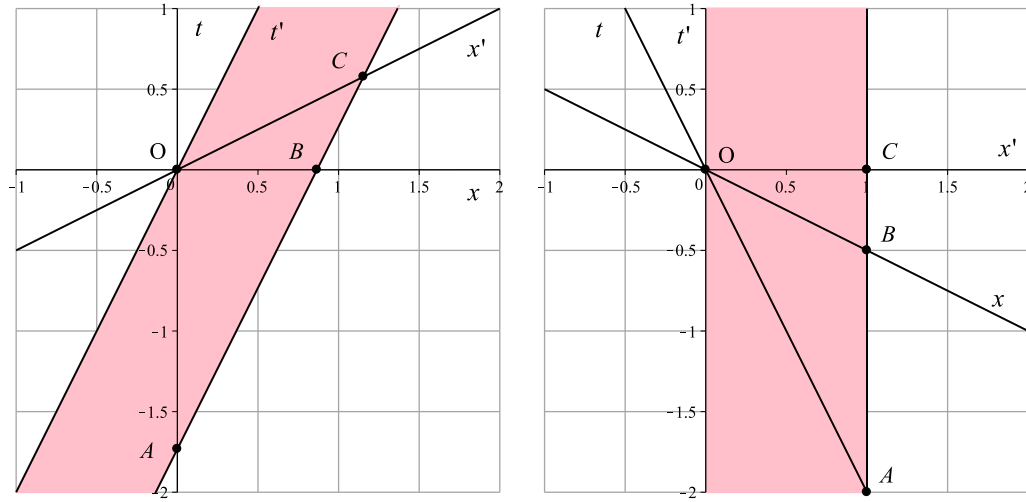


Figure B.3: World sheet of a unit length 1-dimensional ruler moving at constant velocity 0.5 with respect to the inertial frame (length of the line segment OC), with gamma factor 1.155 contracting its length to 0.866. **Left:** Passage of ruler as seen by the unprimed inertial observer at $x = 0$. **Right:** Passage of unprimed observer from the right end of the ruler (A) to the left end (O) as seen by the primed observer in its reference frame.

Fig. B.4 illustrates time dilation. The events O and A occur at position of the moving inertial observer a unit time interval apart: $\Delta t' = 1$. However, the stationary inertial observer sees these events occurring at different locations at a time interval corresponding to the events O and B, which is longer than the primed time interval

$$\Delta t = \gamma \Delta t' = 1.155 \Delta t' > \Delta t = 1.$$

The arc of a hyperbola between A and C is the set of points a unit distance from the origin. All these events are a unit time interval in the future along the world line of some inertial observer passing through the origin.

This 2-dimensional spacetime geometry is the the only new aspect of special relativity, since the dimensions orthogonal to the direction of relative motion are unchanged under Lorentz transformations between inertial frames in relative motion. For 4-dimensional “Lorentz” spacetime of 1 time and 3 space dimensions, we let $x^0 = t$ be the first coordinate with index 0 (the zero index helps remind us that time is very different from space) so we can let the indices 1,2,3 label the space coordinates as usual (and don’t have to relabel the time index if it is instead last in the cases of 1, 2 or 3 space dimensions). Then one has the Lorentz dot product

$$\langle x^0, x^1, x^2, x^3 \rangle \bullet \langle y^0, y^1, y^2, y^3 \rangle = -x^0 y^0 + x^1 y^1 + x^2 y^2 + x^3 y^3 = -x^0 y^0 + \vec{x} \cdot \vec{y}.$$

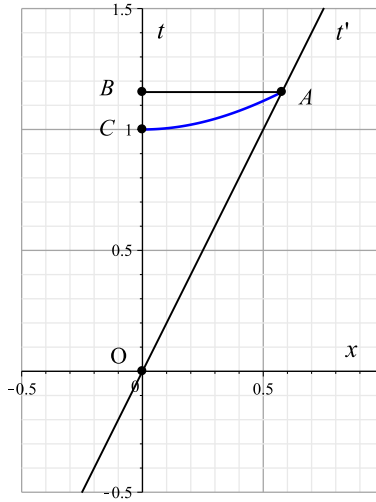


Figure B.4: World line of a point fixed in the primed coordinate grid moving at speed $v = 0.5$ with events O and A seen to occur at the same location but separated by a unit time interval in the moving frame. (All points on the hyperbola are the same spacetime interval from the origin, interpreted as a unit time interval by both observers.) The two events occur at different locations in the unprimed frame, and the time interval is seen to be longer by the gamma factor (time dilation).

As a quadratic form this is

$$(x^0 \ x^1 \ x^2 \ x^3) \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y^0 \\ y^1 \\ y^2 \\ y^3 \end{pmatrix} = x^\mu \eta_{\mu\nu} y^\nu,$$

where $\mu, \nu = 0, 1, 2, 3$. For the self dot product of two vectors, we get the spacetime separation formula, and analogous to the Euclidean case, we get the corresponding quadratic form in the coordinate differentials

$$ds^2 = -dt^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2 = \eta_{\mu\nu} dx^\mu dx^\nu$$

that defines the metric on 4-dimensional spacetime (called Minkowski spacetime) in inertial coordinates $(x^\mu) = (x^0, x^1, x^2, x^3)$, with $x^0 = t$.

Returning to the Lorentz plane with this Lorentz metric in 2-dimensions, it is natural to consider pseudo-polar coordinates directly analogous to polar coordinates in the Euclidean plane, called Rindler coordinates. One needs 4 disjoint coordinate patches separated by the two crossed light cone lines $-t^2 + x^2 = 0$

$$\begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} \rho \cosh \chi \\ \rho \sinh \chi \end{pmatrix} \text{ if } -t^2 + x^2 < 0 \text{ (timelike)} \\ \begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} \rho \sinh \chi \\ \rho \cosh \chi \end{pmatrix} \text{ if } -t^2 + x^2 > 0 \text{ (spacelike)} \end{pmatrix}$$

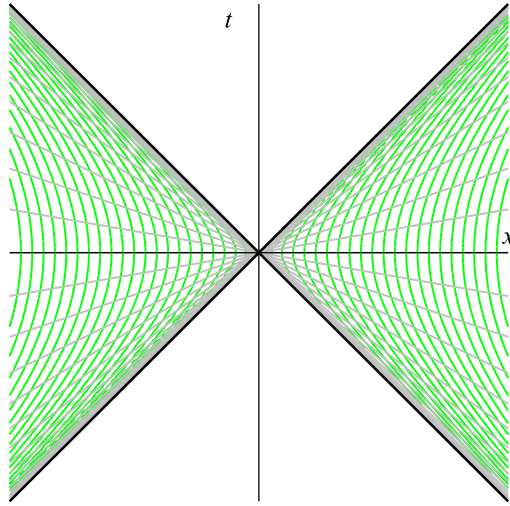


Figure B.5: The Rindler coordinate grid for two disjoint spacelike regions of the light cone.

where

$$-\infty < \rho, \chi < \infty.$$

The two disjoint coordinate grids in the spacelike region of the light cone of the origin are called Rindler wedges. The Rindler coordinate grid consists of concentric hyperbolas of constant spacetime interval from the origin (the time lines) and rays from the origin (the constant time curves), the world lines of the Rindler observers. Each such world line is of constant curvature $\kappa = 1/|\rho|$ whose radius of curvature is equal to the constant radius $|\rho|$ of each such hyperbola. The 4-acceleration equals this curvature, which goes infinite approaching the origin. The coordinates have a “horizon” at the light cone where the coordinates break down, analogous to the origin for ordinary polar coordinates where the polar angle is undetermined. Each Rindler observer thus has a constant but distinct 4-acceleration.

Particle motion: Euclidean versus Lorentzian

The whole idea of curvature starts with circles in the plane. For a circle of radius r , the curvature is defined to be the reciprocal $\kappa = 1/r$. This is then extended to more general curves using the idea of the osculating circle in the plane of the unit tangent and unit normal, which are the arclength first and second derivatives of the position vector along the curve. A parallel development can be done for a timelike curve in the Lorentz plane representing the motion of a point in spacetime, with hyperbolas taking the place of circles in the constant curvature curve starting point (pseudo-circles) and in the osculating pseudo-circle generalization. Furthermore, if one is interested in the geometry near a particular curve, one can introduce an orthogonal coordinate system adapted to this curve, called a Fermi coordinate system.

First consider the Euclidean plane and focus on a particular circle $r = r_0$ in polar coordinates, with curvature $k_0 = 1/r_0$, introducing the arclength coordinate $S = r_0\theta$ along this curve,

and the arclength difference coordinate in the orthogonal radial direction $R = r - r_0$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix} = \begin{pmatrix} (r_0 + R) \cos(S/r_0) \\ (r_0 + R) \sin(S/r_0) \end{pmatrix},$$

for which one has the differential relations $dS = r_0 d\theta$, $dR = dr$. Re-expressing the metric leads easily to

$$ds^2 = dR^2 + (r_0 + R)^2 dS^2 / r_0^2 = dR^2 + (1 + \kappa_0 R)^2 dS^2$$

where the arclength correction factor

$$N = 1 + \kappa_0 R = 1 + (\kappa_0 \hat{\mathbf{r}}) \cdot (R \hat{\mathbf{r}}) = 1 + \vec{\mathbf{a}}_0 \cdot \vec{\mathbf{R}}$$

describes the “arclength acceleration” correction factor of nearby azimuthal coordinate lines in this new orthogonal coordinate system relative to the original circle. The circle from which this coordinate system has been constructed has arclength velocity and acceleration

$$\begin{aligned} \vec{\mathbf{x}}_0 &= \langle r_0 \cos(S/r_0), r_0 \sin(S/r_0) \rangle \\ \vec{\mathbf{v}}_0 &= \frac{d\vec{\mathbf{x}}_0}{ds} = \langle -\sin(S/r_0), \cos(S/r_0) \rangle = \hat{\mathbf{T}}_0 \\ \vec{\mathbf{a}}_0 &= \frac{d^2\vec{\mathbf{x}}_0}{ds^2} = -(1/r_0) \langle \cos(S/r_0), \sin(S/r_0) \rangle = \kappa_0 \hat{\mathbf{N}}_0, \end{aligned}$$

where $\hat{\mathbf{T}}_0$ and $\hat{\mathbf{N}}_0 = -\vec{\mathbf{x}}/r_0 = -\hat{\mathbf{r}}$ are the unit tangent and unit normal to that curve.

The curvature of the circle is the magnitude of the arclength acceleration

$$\kappa_0 = \sqrt{\vec{\mathbf{a}}_0 \cdot \vec{\mathbf{a}}_0},$$

while the curvature is defined to be its reciprocal $\rho_0 = 1/\kappa_0$. The osculating circle is defined to be a circle of radius ρ_0 with center position vector which is equal to the radius of curvature r_0 times the unit normal added to the position vector

$$\vec{\mathbf{C}} = \vec{\mathbf{x}} + \rho_0 \hat{\mathbf{N}}_0,$$

and the evolute of the original curve is the set of all such centers

$$\vec{\mathbf{E}}(S) = \vec{\mathbf{x}}(S) + \rho_0 \hat{\mathbf{N}}_0(S).$$

For the circle this evolute consists of a single point where all the normal lines intersect, and which gives a coordinate singularity where the new coordinates break down. For a noncircular curve, the evolute is another curve, which marks the limiting interval of R for each S for which the new coordinates are valid.

Now consider the same analysis for a timelike pseudo-circle $\ell = \ell_0$ about the origin in the Lorentz plane, introducing the arclength coordinate $\tau = \ell_0 \chi$ along this curve, and the arclength difference coordinate in the orthogonal “radial” direction $X = \ell - \ell_0$

$$\begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} \ell \sinh \chi \\ \ell \cosh \chi \end{pmatrix} = \begin{pmatrix} (\ell_0 + X) \sinh(\tau/\ell_0) \\ (\ell_0 + X) \cosh(\tau/\ell_0) \end{pmatrix},$$

for which one has the differential relations $d\tau = \ell_0 d\chi$, $dX = d\ell$. Re-expressing the metric leads easily to

$$ds^2 = dX^2 - (\ell_0 + X)^2 d\tau^2 / \ell_0^2 = dX^2 - (1 + \kappa_0 X)^2 d\tau^2$$

where the arclength correction factor

$$N = 1 + \kappa_0 X = 1 + \left(\kappa_0 \hat{\ell} \right) \cdot \left(X \hat{\ell} \right) = 1 + \vec{\mathbf{a}}_0 \cdot \vec{\mathbf{X}}$$

describes the “arclength acceleration” correction factor of nearby time coordinate lines in this new orthogonal coordinate system relative to the original pseudo-circle. The pseudo-circle from which this coordinate system has been constructed has arclength velocity and acceleration

$$\begin{aligned} \vec{\mathbf{x}}_0 &= \langle \ell_0 \sinh(\tau/\ell_0), \ell_0 \cosh(\tau/\ell_0) \rangle \\ \vec{\mathbf{u}}_0 &= \frac{d\vec{\mathbf{x}}_0}{d\tau} = \langle \cosh(\tau/\ell_0), \sinh(\tau/\ell_0) \rangle = \hat{\mathbf{T}}_0 \\ \vec{\mathbf{a}}_0 &= \frac{d^2\vec{\mathbf{x}}_0}{d\tau^2} = (1/\ell_0) \langle \sinh(\tau/\ell_0), \cosh(\tau/\ell_0) \rangle = \kappa_0 \hat{\mathbf{N}}_0, \end{aligned}$$

where $\hat{\mathbf{T}}_0$ and $\hat{\mathbf{N}}_0 = \vec{\mathbf{x}}/\ell_0 = -\hat{\mathbf{r}}$ are the unit tangent and unit normal to that curve. However, now $\hat{\mathbf{T}}_0 \cdot \hat{\mathbf{T}}_0 = -1$ since this is a timelike curve, while its normal is spacelike: $\hat{\mathbf{N}}_0 \cdot \hat{\mathbf{N}}_0 = 1$.

For a timelike curve one instead has an osculating pseudo-circle (hyperbola) whose center is defined exactly as in the Euclidean case, except for a change in sign since the center of the hyperbola is on the opposite side from its normal vector

$$\vec{\mathbf{C}} = \vec{\mathbf{x}} - r_0 \hat{\mathbf{N}}_0,$$

For a pseudo-circle itself, the evolute is a single point (the origin in this case) but for a general timelike curve (world line), the evolute marks the interval in X away from the original world line where this new coordinate system breaks down due to the crossing of its spatial coordinate lines.

These new orthogonal coordinates are called Fermi coordinates and are orthonormal along the curve used to construct them. They are easily generalized to 4-dimensional spacetime. They were first introduced by a college student, Enrico Fermi, only a few years after the birth of general relativity, Einstein’s theory of gravity. He was motivated by the problem of electromagnetic mass. Fig. B.6 shows an osculating hyperbola on the spacetime helix in 3-dimensional Minkowski spacetime corresponding to the above circular motion, as derived in the following exercise and next section.

Exercise B.0.7.

Consider a circular orbit of a small mass around a large central mass, which is a helix in spacetime. We only need 3-dimensional spacetime to describe this problem since the orbit is confined to a plane in space. The circular orbit, when moving along the time direction, becomes

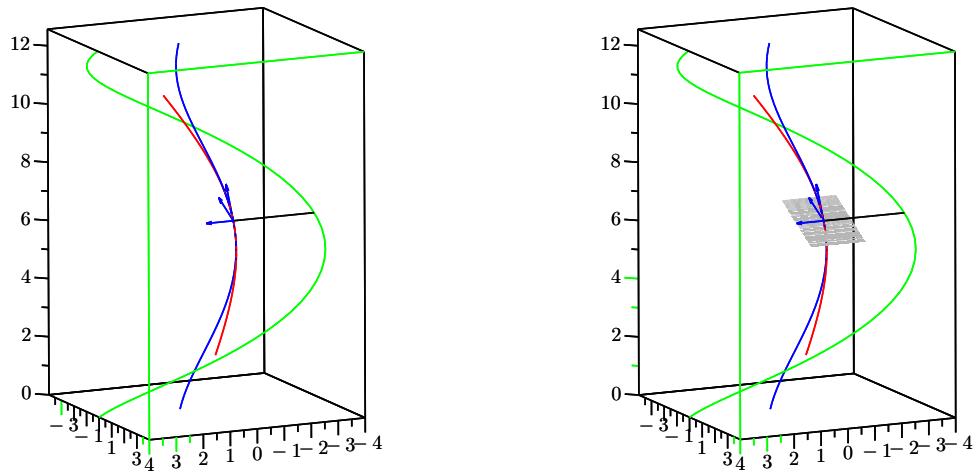


Figure B.6: The Lorentzian helix with its osculating hyperbola and evolute, including a parallelogram from the local rest space (right). The left figure more clearly shows the Serret-Frenet orthonormal frame and the osculating circle, with the line segment along the normal direction extending to the center of the osculating hyperbola. The evolute is also a helix, contained in a cylinder about the time axis which limits the validity of the Fermi coordinates based on the original helix, representing a point particle in a circular orbit about the spatial origin of coordinates.

a helix about the time axis at which the central mass is located. The proper time parametrized world line is

$$\begin{aligned}\vec{x} &= \hat{\mathbf{R}}(\tau) = \langle t, x, y \rangle = \langle \gamma\tau, a \sin(\omega_0\tau), a \cos(\omega_0\tau) \rangle = \langle t, a \sin(\omega t), a \cos(\omega t) \rangle, \\ \vec{u} &= \frac{d\vec{x}}{d\tau} = \hat{\mathbf{R}}'(\tau) = \hat{\mathbf{T}}, \\ \vec{a} &= \frac{d\vec{u}}{d\tau} = \frac{d^2\vec{x}_0}{d\tau^2} = \hat{\mathbf{R}}''(\tau) = \kappa\hat{\mathbf{N}}.\end{aligned}$$

This is spatially periodic with proper time period $P_0 = 2\pi/\omega_0$ and longer coordinate time period $P = 2\pi/\omega = \gamma P_0$ (time dilation), returning to the point with the same spatial coordinates after one period.

a) Determine the gamma factor $\gamma > 0$ such that the 4-velocity is a unit timelike vector: $\vec{u} \cdot \vec{u} = -1$.

b) Express the coordinate time frequency ω in terms of the proper time frequency ω_0 and vice versa. Express the gamma factor in terms of ω . Express the speed $v = |d\vec{x}/dt|$ in terms of the coordinate frequency and the proper time frequency.

c) Evaluate the constant curvature κ of this helix in spacetime, the reciprocal radius of curvature ρ , and the unit normal.

d) Evaluate the parametrized osculating pseudo-circle, parametrized by the hyperbolic angle α

$$\vec{x} = \vec{\mathbf{C}} + \rho \left(\cosh \alpha \hat{\mathbf{N}} + \sinh \alpha \hat{\mathbf{u}} \right).$$

e) What is the parametrized equation for the evolute? Note that it too is a helix, contained in a cylinder about the time axis of radius r_e . Evaluate r_e and then re-express it in terms of the radius a of the circular orbit and the radius of curvature ρ . Plot one revolution of the helix $\tau = 0..P_0$ together with its Lorentzian evolute for the parameter values $(a, v, \gamma, \omega_0 = \gamma v) = (1/2, 2/\sqrt{3}, 1/\sqrt{3})$. Add in the osculating hyperbola at the halfway point $\tau = P_0/2$ (say for $\alpha = -1..1$, why is this a good choice considering the dimensions of the boxed plot for one period of the helix?) together with the horizontal line segment from the point of tangency to osculating hyperbola center.

f) Repeat this for the Euclidean case, letting $(t, \tau, \gamma) \rightarrow (z, s, \Gamma)$ to determine $\Gamma > 0$ so that the tangent is a unit vector (moving z to the last position in the position vector) and then evaluate c .

$$\begin{aligned}\vec{x} &= \langle x, y, z \rangle = \langle a \sin(\omega_0 s), a \cos(\omega_0 s), \gamma s \rangle = \langle a \sin(\phi), a \cos(\phi), c\phi \rangle, \\ \vec{v} &= \frac{d\vec{x}}{ds} = \hat{\mathbf{T}}, \\ \vec{a} &= \frac{d\vec{v}}{ds} = \frac{d^2\vec{x}_0}{ds^2} = \kappa\hat{\mathbf{N}}.\end{aligned}$$

The Euclidean inclination angle of the helix is $\tan \chi = c/a$ but to compare with the Lorentzian case, consider the Euclidean angle of inclination away from the vertical axis $\tan \xi = a/c = \tanh \beta$, where the “rapidity” β is the hyperbolic angle defined by the inclination of the unit tangent away from the time axis in the Lorentzian geometry. Express β in terms of ξ . What

is the parametrized equation for the evolute in this case? What is the value of this inclination angle for the parameter values of part e)?

g) Evaluate the binormal $\hat{\mathbf{B}} = \hat{\mathbf{T}} \times \hat{\mathbf{N}}$ for the Euclidean helix. As we will learn in Chapter 1, the cross-product in the Lorentzian case is a “covector” whose value on either $\hat{\mathbf{T}}$ or $\hat{\mathbf{N}}$ is zero, so it can be “index-raised” to a vector by changing the sign of its time component to obtain a spacelike Lorentzian binormal vector orthogonal to the plane of the 4-velocity and 4-acceleration vectors, but may require normalization, i.e., division by its length to make it a normal vector $\hat{\mathbf{B}}$ in the Lorentzian geometry. Evaluate this vector. Add the three orthonormal Lorentzian Serret-Frenet vectors to your previous plot, and include the parallelogram spanned by the two unit normals and their negatives:

$$\hat{\mathbf{R}}(P/2) + t_1 \hat{\mathbf{N}}(P/2) + t_2 \hat{\mathbf{B}}(P/2), t_1 = -1..1, t_2 = -1..1$$

This lies in the “local rest space” orthogonal to the 4-velocity. Although it is a square in the Lorentz geometry, it is not square in the Euclidean geometry in which we view the plot since it is tilted with respect to the horizontal.

h) In each local rest space, we can introduce local orthonormal coordinates based on the orthonormal vectors $(\hat{\mathbf{N}}, \hat{\mathbf{B}})$ and complete them to spacetime coordinates (X, Y, τ) using the proper time along the helix. These are Fermi coordinates, valid within the cylinder of radius r_e . Can you express the original coordinates in terms of the new coordinates? Just simplify:

$$\hat{\mathbf{x}} = \hat{\mathbf{R}}(\tau) + X\hat{\mathbf{N}}(\tau) + Y\hat{\mathbf{B}}(\tau).$$

Can you invert this to obtain the new coordinates as functions of the old ones? If one plots one local rest space for the above parameter values, letting X and Y range from -3 to 3 , since $3 = r_e - a$ is the horizontal distance from the helix to its evolute, then plotting a second one slightly later shows these two planes intersecting at the evolute. Try to create such a plot. ■

Appendix C

Curves in 3-space

The Euclidean helix

Curves play a fundamental role in differential geometry, so it is important to recall their basics from multivariable calculus. Let's use an explicit example to see how things work in a best case scenario, that of a helix. Using Cartesian coordinates $(x, y, z) = (\rho \cos(\phi), \rho \sin(\phi), z)$ or cylindrical coordinates (ρ, ϕ, z) , a helix around the z -axis can be parametrized in the following natural way, using the azimuthal angle ϕ as the parameter t , with fixed radial variable $\rho = a$

$$x = a \cos(t), \quad y = a \sin(t), \quad z = ct \leftrightarrow \vec{r} = \langle x, y, z \rangle = \langle a \cos(t), a \sin(t), ct \rangle,$$

where the position vector notation \vec{r} is useful so we can use the approach of vector calculus. We can let $\vec{r}(t)$ stand for the actual parametrized vector

$$\vec{r}(t) = \langle a \cos(t), a \sin(t), ct \rangle.$$

Note that $c > 0, c < 0$ means that this is a right-handed/left-handed helix, rising/descending as one moves in the counterclockwise (positive azimuthal direction) around the axis of symmetry, with an inclination angle $\eta = \arctan(c/a)$ with respect to the horizontal, and is contained in the cylinder $\rho = a \geq 0$ around the vertical z -axis. When $c = 0$ we get the special case of a circle in the plane $z = 0$ of radius a .

The tangent is then the first derivative (the velocity $\vec{v}(t)$ if we interpret t as the time and $\vec{r}(t)$ as the position vector of a moving point)

$$\vec{r}'(t) = \frac{d}{dt} \vec{r}(t) = \langle -a \sin(t), a \cos(t), c \rangle,$$

while the second derivative (the acceleration $\vec{a}(t)$ in the particle motion language) is

$$\vec{r}''(t) = \frac{d^2}{dt^2} \vec{r}(t) = \langle -a \cos(t), -a \sin(t), 0 \rangle.$$

The plane of the first and second derivatives (the velocity-acceleration plane, also called the osculating plane for the Greek root which means “kiss”, since the osculating circle just kisses the

curve at the point of tangency) is where the turning of the curve is taking place instantaneously. Its unit normal is called the binormal

$$\hat{B}(t) = \frac{\vec{r}'(t) \times \vec{r}''(t)}{|\vec{r}'(t) \times \vec{r}''(t)|} = \frac{1}{\sqrt{a^2 + c^2}} \langle c \sin(t), -c \cos(t), a \rangle,$$

where the overhat reminds us that it is a unit vector. The Euclidean geometry enters the picture with the cross-product and when we evaluate lengths with the self-dot product, while dividing out the length leads to a unit vector

$$[\vec{X} \times \vec{Y}]_i = \epsilon_{ijk} X^j Y^k, \quad \epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1 = -\epsilon_{132} = -\epsilon_{213} = -\epsilon_{321}, \quad (\text{C.1})$$

$$|\vec{X}| = \sqrt{\vec{X} \cdot \vec{X}}, \quad \hat{X} = \frac{\vec{X}}{|\vec{X}|} \rightarrow \hat{X} \cdot \hat{X} = \delta_{ij} \hat{X}^i \hat{X}^j = 1. \quad (\text{C.2})$$

For the tangent vector this leads to its length (the speed $v = |\vec{v}|$) and the unit tangent (direction of motion)

$$|\vec{r}'(t)| = \sqrt{\vec{r}'(t) \cdot \vec{r}'(t)} = \sqrt{a^2 + c^2}, \quad \hat{T}(t) = \frac{\vec{r}'(t)}{|\vec{r}'(t)|} = \frac{1}{\sqrt{a^2 + c^2}} \langle -a \sin(t), a \cos(t), c \rangle.$$

To get the unit normal $\hat{N}(t)$, which is the direction in which the unit tangent is rotating within the velocity-acceleration plane, we just differentiate the unit tangent and divide the result by its length

$$\hat{N}(t) = \frac{\hat{T}'(t)}{|\hat{T}'(t)|} = \langle -\cos(t), -\sin(t), 0 \rangle.$$

Although we jumped the gun on defining the binormal directly from the first and second derivatives, we could have waited and defined it by

$$\hat{B}(t) = \hat{T}(t) \times \hat{N}(t),$$

since those two derivatives span the same subspace as the unit tangent and unit normal. On the other hand, it is often easier to first calculate $\hat{B}(t)$ directly as the normalized cross-product of $\vec{r}'(t)$ and $\vec{r}''(t)$, and then calculate $\hat{N}(t) = \hat{B}(t) \times \hat{T}(t)$.

By construction this ordered set of three vectors $\{\hat{E}_1(t), \hat{E}_2(t), \hat{E}_3(t)\} = \{\hat{T}(t), \hat{N}(t), \hat{B}(t)\}$ forms a right handed triad of mutually orthogonal unit vectors called the Frenet-Serret frame, defined at each point along the curve that we do not encounter a zero length tangent (momentarily at rest, zero speed, no direction of motion). By right-handed we mean that the cross-product of any two of these vectors in cyclic order is the third: $\hat{E}_3 \times \hat{E}_1 = \hat{E}_2$, etc. Even when we are only doing geometry, tracing out the curve means an animation, which means that the physics picture of a moving point particle is relevant. This is very useful for understanding the geometry of curves. The Frenet-Serret frame may be thought of as moving along the curve, and is often called a “moving frame.”

One thing is still missing, a measure of the curvature of the curve. For this we need the arclength derivatives. The differential of arclength along the curve is

$$ds^2 = dx^2 + dy^2 + dz^2 \rightarrow \frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} = |\vec{r}'(t)|,$$

and in the few cases where this differential equation can be explicitly integrated exactly

$$s = \int_{t_0}^t |\vec{r}'(u)| du$$

and also inverted to express $t = t(s)$, then we can substitute this relationship into the parametrized curve to reparametrize it in terms of the arclength from some arbitrary reference point $\vec{r}(t_0)$. For the helix this is easy since the arclength is linear in the original parameter, and choosing their zeros to coincide gives

$$s = \sqrt{a^2 + c^2} t \leftrightarrow t = s / \sqrt{a^2 + c^2},$$

so that

$$\vec{r}(s) = \left\langle a \cos \left(\frac{s}{\sqrt{a^2 + c^2}} \right), a \sin \left(\frac{s}{\sqrt{a^2 + c^2}} \right), \frac{cs}{\sqrt{a^2 + c^2}} \right\rangle,$$

where we abuse the functional notation by using the same symbol for the composed vector function $\vec{r}(t(s)) \rightarrow \vec{r}(s)$. This is equivalent to tracing out the curve at unit speed, since the first derivative comes out automatically to be a unit vector by the way the differential of arclength has been defined.

Now we can recompute the first and second derivatives.

$$\begin{aligned} \frac{d}{ds} \vec{r}(s) &= \hat{T}(s) = \frac{1}{\sqrt{a^2 + c^2}} \left\langle -a \sin \left(\frac{s}{\sqrt{a^2 + c^2}} \right), a \cos \left(\frac{s}{\sqrt{a^2 + c^2}} \right), c \right\rangle \\ \frac{d}{ds} \hat{T}(s) &\equiv \underbrace{\kappa(s)}_{\frac{a}{a^2 + c^2}} \underbrace{\hat{N}(s)}_{\left\langle -\cos \left(\frac{s}{\sqrt{a^2 + c^2}} \right), -\sin \left(\frac{s}{\sqrt{a^2 + c^2}} \right), 0 \right\rangle}. \end{aligned}$$

The magnitude of the second arclength derivative is defined to be the curvature $\kappa \geq 0$, and dividing it out gives the direction of the rate of change of the unit tangent, which is the unit normal \hat{N} . The helix has constant curvature, like its special case $c = 0$ where it degenerates to a circle and the curvature reduces to the reciprocal of the radius a of the circle, the starting point for defining curvature for all curves. In analogy, we define the reciprocal of curvature here to be the radius of curvature

$$\rho(s) = \frac{1}{\kappa(s)} = \frac{a^2 + c^2}{a} = a \left(1 + \left(\frac{c}{a} \right)^2 \right) \geq a.$$

For the helix, stretching the circle of radius a vertically to make a helix makes the curve less curved, so the radius of curvature is larger than a for the helix.

If we define the osculating circle at each point of the parametrized curve to be the circle which radius equal to the radius of curvature and center located a distance equal to that radius along the unit normal from the tip of the position vector, namely

$$\vec{C}(s) = \vec{r}(s) + \rho(s) \hat{N}(s).$$

If we zoom into the point of tangency of this circle with the original curve at $\vec{r}(s)$, then the curve and circle merge in a quadratic approximation to the curve before we zoom too far so that both straighten out to the tangent line to the curve.

It is easy to parametrize this osculating circle, recalling how we parametrize a circle about the origin in the x - y plane

$$\langle x, y \rangle = \langle a \cos \theta, a \sin \theta \rangle = a(\cos \theta e_1 + \sin \theta e_2).$$

Replacing the unit vectors (e_1, e_2) along the coordinate axes here by the unit vectors $(-\hat{N}(s), \hat{T}(s))$, we get a parametrization of a circle about the tip of the position vector starting at the point of tangency when $\theta = 0$ and moving from $\hat{N}(s)$ towards $\hat{T}(s)$. All that remains is to add this to the tip of the position vector

$$\vec{r}(t, \theta) = \vec{C}(s) + \rho(s)(-\cos \theta \hat{N}(s) + \sin \theta \hat{T}(s)).$$

Thus at $\theta = 0$, we get

$$\vec{r}(t, 0) = \vec{C}(s) + \rho(s)(-\hat{N}(s)) = (\vec{r}(t) + \rho(t)\vec{N}(t)) - \rho(s)\hat{N}(s) = \vec{r}(t).$$

Exercise C.0.8.
the helix torsion

By explicitly differentiating

$$\hat{B}(s) = \frac{1}{\sqrt{a^2 + c^2}} \left\langle c \sin \left(\frac{s}{\sqrt{a^2 + c^2}} \right), -c \cos \left(\frac{s}{\sqrt{a^2 + c^2}} \right), a \right\rangle,$$

identify the torsion defined by the relation $d\hat{B}/ds = -\tau\hat{N}$ explained below to find the formula

$$\tau(s) = \frac{c}{a^2 + c^2}.$$

The torsion somehow describes the rotation of the orientation of the normal to the velocity-acceleration plane along the curve and vanishes for plane curves, namely those curves which lie in some plane, so that the binormal is always equal to unit normal to that fixed plane containing the curve. ■

Exercise C.0.9.
the helix osculating circle

Evaluate first the osculating circle center $\vec{C}(s)$ for the helix, then the osculating circle $\vec{r}(s, \theta)$ itself.



Any unit vector function can only rotate, like the unit tangent and the two unit normals, so its derivative has to be in the plane perpendicular to the vector itself. The unit normal and unit binormal are locked to the unit tangent, so as the unit tangent rotates, they must react to stay orthogonal, so their rates of change are all interlocked as well, and must be orthogonal to themselves. One can easily compute the derivatives of the unit normal and binormal for general space curves, which is something that is usually not done in a typical multivariable calculus class, but is not any more difficult than the calculations we have already done. The derivative of the unit normal must be a linear combination of the unit tangent and the unit binormal, and the derivative of the unit binormal likewise

$$\frac{d\hat{N}}{ds} = \alpha\hat{T} + \beta\hat{B}, \quad \frac{d\hat{B}}{ds} = \gamma\hat{T} + \delta\hat{N}.$$

The orthogonality relations then force

$$0 = \frac{d}{ds} (\hat{T} \cdot \hat{N}) = \frac{d\hat{T}}{ds} \cdot \hat{N} + \hat{T} \cdot \frac{d\hat{N}}{ds} = \kappa + \alpha \rightarrow \alpha = -\kappa$$

and

$$0 = \frac{d}{ds} (\hat{B} \cdot \hat{N}) = \frac{d\hat{B}}{ds} \cdot \hat{N} + \hat{B} \cdot \frac{d\hat{N}}{ds} = \delta + \beta$$

and

$$0 = \frac{d}{ds} (\hat{T} \cdot \hat{B}) = \frac{d\hat{T}}{ds} \cdot \hat{B} + \hat{T} \cdot \frac{d\hat{B}}{ds} = 0 + \gamma.$$

Summarizing we get the Frenet-Serret relations for Euclidean 3-space (define $\tau = \beta$)

$$\begin{aligned} \frac{d\hat{T}}{ds} &= \kappa\hat{N}, \\ \frac{d\hat{N}}{ds} &= -\kappa\hat{T} + \tau\hat{B}, \\ \frac{d\hat{B}}{ds} &= -\tau\hat{N}. \end{aligned}$$

or in matrix form

$$\begin{pmatrix} \frac{d\hat{T}}{ds} & \frac{d\hat{N}}{ds} & \frac{d\hat{B}}{ds} \end{pmatrix} = (\hat{T} \quad \hat{N} \quad \hat{B}) \begin{pmatrix} 0 & -\kappa & 0 \\ \kappa & 0 & -\tau \\ 0 & \tau & 0 \end{pmatrix}.$$

The second Frenet-Serret scalar τ is called the torsion. Notice that the coefficient matrix here is antisymmetric, as it must be if it is the derivative of a rotation matrix. We first explained this in Section 1.7.

In the motion language, the curvature is the instantaneous arclength rate of change of the angle of rotation of the pair of vectors \hat{T} and \hat{N} in their plane, or the angular velocity of those

vectors about the binormal. Similarly the torsion is the instantaneous arclength rate of change of the angle of rotation of the vectors \hat{N} and \hat{B} in their plane, called the normal plane, or the angular velocity of the normal plane about the unit tangent. In fact this matrix defines an angular velocity vector

$$\vec{\omega} = \tau\hat{T} + \kappa\hat{B} \equiv \omega^1\hat{E}_1 + \omega^3\hat{E}_3$$

with magnitude

$$|\vec{\omega}| = \sqrt{\kappa^2 + \tau^2}$$

such that if we express a vector along the curve in terms of its components with respect to the frame vectors

$$\vec{X} = X^1\hat{T} + X^2\hat{N} + X^3\hat{B} = X^1\hat{E}_1 + X^2\hat{E}_2 + X^3\hat{E}_3$$

then in matrix form

$$\frac{d}{ds} \begin{pmatrix} X^1 \\ X^2 \\ X^3 \end{pmatrix} = \begin{pmatrix} 0 & -\kappa & 0 \\ \kappa & 0 & -\tau \\ 0 & \tau & 0 \end{pmatrix} \begin{pmatrix} X^1 \\ X^2 \\ X^3 \end{pmatrix}$$

or in vector form

$$\frac{d}{ds} \langle X^1, X^2, X^3 \rangle = \vec{\omega} \times \langle X^1, X^2, X^3 \rangle,$$

which should ring a bell from elementary physics discussions of angular velocity. The cross product of the angular velocity with the position vector gives the velocity vector of a moving point. The relation between the antisymmetric matrix and the angular velocity is exactly that was first explored in Exercise 1.2.4.

Note that the curvature is the angular velocity about the binormal in the velocity-acceleration plane, always nonnegative since the tangent rotates towards the unit normal by definition (the unit normal is on the side of the tangent line in which the curve is concave way from that tangent line, while the torsion is the angular velocity in the normal plane about the unit tangent, which can be either positive or negative depending on whether the normal rotates towards the binormal or away from it.

Exercise C.0.10.

Verify the above cross product relation. ■

Remark.

Several times it has been claimed that the curvature may be interpreted as the arclength rate of change of the angle of rotation of the unit tangent in the osculating plane. How can we back up this claim? Consider a curve in the x - y plane with a unit tangent $\hat{T}(s) = \langle \cos \theta(s), \sin \theta(s) \rangle$ which can be easily parametrized by its angle with respect to the positive horizontal direction. Then

$$\frac{d\hat{T}(s)}{ds} = \langle -\sin \theta(s), \cos \theta(s) \rangle \frac{d\theta(s)}{ds} = \kappa(s)\hat{N}(s),$$

where $\kappa(s) \geq 0$ so that one can identify

$$\left| \frac{d\theta(s)}{ds} \right| = \kappa(s) = \hat{N}(s) \cdot \frac{d\hat{T}(s)}{ds}.$$

The right hand side equality is the definition of curvature in the case of a general space curve, which can be used to define an equivalent rate of change of an angle (up to sign) of rotation of the unit tangent in the osculating plane there.

In a completely parallel way, the relation $\tau(s) = -\hat{N}(s) \cdot d\hat{B}(s)/ds$ describes the angular rate of change of the binormal as it rotates around the unit tangent direction away from the tip of $\hat{N}(s)$.

▲

In the arclength parametrization of a curve, the geometry is easy to evaluate. However, for most curves one cannot reparametrize the curve by an arclength function, so we instead have to evaluate these arclength derivative definitions using the chain rule

$$\frac{df}{ds} = \frac{df/dt}{ds/dt} = \frac{df/dt}{|\vec{r}'|}.$$

Thus the curvature and torsion are

$$\kappa(t) = |\hat{T}'(t)|/|\vec{r}'(t)| \geq 0, \quad \tau(t) = \vec{B}(t) \cdot \vec{N}'(t)/|\vec{r}'(t)|.$$

The torsion vanishes identically only when the binormal is a constant vector, i.e., when the orientation of the velocity-acceleration plane does not change, which is only possible for a curve confined to a single plane, called a plane curve. If the curvature is identically zero, then the unit tangent is constant, which is only possible for a straight line.

Exercise C.0.11.

a) Using the chain rule and the facts that $\hat{T} = \vec{r}'/|\vec{r}'|$ and $|\hat{T} \times \hat{N}| = |-\hat{B}| = 1$, as well as $\kappa = |d\hat{T}/ds| \geq 0$, confirm the steps in the following calculation

$$\begin{aligned} \kappa &= |\hat{T} \times (\kappa \hat{N})| = \left| \hat{T} \times \frac{d\hat{T}}{ds} \right| = \left| \frac{\vec{r}'}{|\vec{r}'|} \times \frac{1}{|\vec{r}'|} \left(\frac{\vec{r}'}{|\vec{r}'|} \right)' \right| \\ &= \frac{|\vec{r}' \times \vec{r}''|}{|\vec{r}'|^3}, \end{aligned}$$

which yields a simple direct formula for the curvature in any parametrization found in all multivariable calculus textbooks.

b) Confirm the formula found above for the curvature of a helix using this new method of evaluation.

c) For the twisted cubic curve segment

$$\vec{r}(t) = \langle t, t^2, t^3 \rangle, \quad -1 \leq t \leq 1,$$

evaluate the speed $v(t)$, the Frenet-Serret frame vectors (in the order \hat{T} , \hat{B} , then \hat{N} , and the curvature. Evaluate the parametrized curve $\vec{C}(t) = \vec{r}(t) + \rho(t)\hat{N}(t)$ representing the center of the osculating circles (called the evolute of the original curve). Plot the osculating circle along the curve using technology. Superimpose the evolute curve segment. ■

Returning to our helix, the derivative of the unit binormal is

$$\frac{d\hat{B}(t)}{ds} = - \underbrace{\frac{c}{a^2 + c^2}}_{\tau} \underbrace{\langle -\cos(t), -\sin(t), 0 \rangle}_{\hat{N}},$$

so

$$\tau = \frac{c}{a^2 + c^2}.$$

The osculating plane at the point $\vec{r}(t)$ of the curve is the velocity-acceleration plane spanned either by \vec{r}' and \vec{r}'' or by \hat{T} and \hat{N} . The osculating circle is defined to be the circle in this osculating plane of radius $\rho(t)$ whose center is a distance $\rho(t)$ along the unit normal from the tip of the position vector as derived in Exercise C.0.9

$$\vec{C}(t) = \vec{r}(t) + \rho(t)\hat{N}(t) = \frac{c}{a} \langle -c \cos(t), -c \sin(t), at \rangle,$$

which for the helix traces out another helix, called the evolute of the original helix, with inclination angle $-\arctan(a/c)$, which corresponds to the direction orthogonal to the direction of the original helix. (For an actual circle with $c = 0$, this reduces to the center of the circle at the origin.) We can parametrize the osculating circle itself by the usual parametrization of a circle at the origin of the x - y plane, choosing the angle so that we are at the point of contact at $\theta = 0$ and begin to move in the unit tangent direction as we increase that angle. The result of Exercise C.0.9 is

$$\begin{aligned} \vec{r}_C(\theta, t) &= \vec{C}(t) + \rho(t) \left(-\cos(\theta)\hat{N}(t) + \sin(\theta)\hat{T}(t) \right) \\ &= \frac{1}{a} \langle \cos(t)(-c^2 + (a^2 + c^2)\cos(\theta)) - a\sqrt{a^2 + c^2}\sin(t)\sin(\theta), \\ &\quad \sin(t)(-c^2 + (a^2 + c^2)\cos(\theta)) + a\sqrt{a^2 + c^2}\cos(t)\cos(\theta), \\ &\quad cat + c\sqrt{a^2 + c^2}\sin(\theta) \rangle. \end{aligned}$$

For any parametrized curve the center of the osculating circle is where nearby normal lines (lines through the curve along the unit normal direction) intersect as one moves along the curve. The evolute curve consisting of these centers connects up these intersection points.

Exercise C.0.12.

helix osculating circle graphics

Verify these last two results for $\vec{C}(t)$ and $\vec{r}_C(\theta, t)$ with a computer algebra system, and plot the osculating circle together with the helix at $t = 0$ and with $c = a = 1$.

Exercise C.0.13.**twisted cubic osculating circle graphics**

A simple rescaling of the twisted cubic example found in most multivariable calculus textbooks results in a perfect square for the self dot product of the tangent vector

$$\vec{r}(t) = \langle 2t, t^2, t^2/3 \rangle.$$

This makes the calculations of all the geometric quantities along the curve a little bit easier. Evaluate the equation for the osculating circle of this parametrized curve and plot it at $t = 1$ using a computer algebra system.

Circles to pseudo-circles: hyperbolas

The circles $x^2 + y^2 = r^2$ of the Euclidean plane become the pseudo-circular hyperbolas $x^2 - t^2 = \pm s^2$ of the Lorentz plane, with the pseudo-radius of the hyperbola being the distance s (called the spacetime interval) from its center to any point on the curve in the Lorentzian geometry. The radius of an approximating osculating circle is used to define the radius of curvature of a general Euclidean curve, and in the Lorentz plane, the pseudocircular hyperbolas are the curves of constant curvature which can be used to define the osculating hyperbola to a curve in that plane in order to use the Lorentz geometry consistently in the description of the local behavior of the curve. In practice one could use any quadratic curve to approximate another curve at a point, as long as that approximating curve has the same curvature. Thus for the same curve in the plane, we can use either a tangent circle or tangent hyperbola to approximate the curvature of the same curve, depending on which geometry we wish to use, for interpretational purposes.

If we contemplate using hyperbolas to approximate more general curves, we should first examine carefully the curvature of a hyperbola itself, revisited using Lorentz instead of Euclidean geometry. If (x, t) are respectively the horizontal and vertical coordinates in the Lorentz plane, then the hyperbola of all points a distance $\ell > 0$ from the origin opening up about the positive x axis is $x^2 - t^2 = \ell^2, x > 0$ and can be parametrized by the hyperbolic angle as

$$\vec{r} = \langle x, t \rangle = \langle \ell \cosh \alpha, \ell \sinh \alpha \rangle,$$

with tangent vector

$$\frac{d\vec{r}}{d\alpha} = \left\langle \frac{dx}{d\alpha}, \frac{dt}{d\alpha} \right\rangle = \langle \ell \sinh \alpha, \ell \cosh \alpha \rangle.$$

This is illustrated in Fig. C.1. In the Lorentz inner product, the tangent vector has the self-dot product

$$\frac{d\vec{r}}{d\alpha} \cdot \frac{d\vec{r}}{d\alpha} = \left(\frac{dx}{d\alpha} \right)^2 - \left(\frac{dt}{d\alpha} \right)^2 = \ell^2 (\sinh^2 \alpha - \cosh^2 \alpha) = -\ell^2 = - \left(\frac{d\tau}{d\alpha} \right)^2,$$

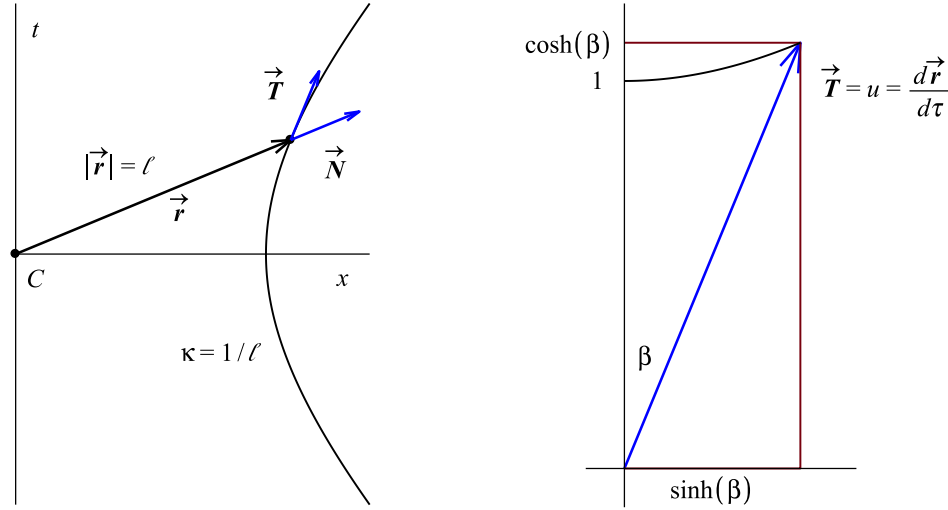


Figure C.1: **Left:** The unit tangent and unit normal for one half of a timelike pseudocircular hyperbola centered at the origin, of pseudoradius ℓ and curvature $\kappa = 1/\ell$. The center is a distance ℓ along the normal line on the opposite side of the curve from the direction along which the unit normal \hat{N} is pointing. **Right:** the unit tangent spacetime velocity $u = \hat{T}$ has components parametrized by the hyperbolic angle from the time axis, called the rapidity β . The space velocity $v = dx/dt = \tanh \beta$ is just the reciprocal slope of the tangent line.

which means that it is a timelike curve and we can reparametrize it by the proper time, choosing $\tau = \ell\alpha$, which is analogous to the Euclidean relationship that arc length of a circle is the radius times the angle—in this case the proper spacetime distance along an arc of a pseudo-circle (a proper time interval in this case because it is a timelike curve) is just the radius of the hyperbola times the hyperbolic angle. If we reparametrize the curve by the proper time, its tangent will be a unit vector just as in the Euclidean case, and its second derivative will be the nonnegative curvature times the unit normal

$$\begin{aligned}\vec{r}(\tau) &= \langle x(\tau), t(\tau) \rangle = \langle \ell \cosh(\tau/\ell), \ell \sinh(\tau/\ell) \rangle, \\ \hat{T}(\tau) = \frac{d\vec{r}(\tau)}{d\tau} &= \langle \sinh(\tau/\ell), \cosh(\tau/\ell) \rangle \equiv \hat{u}(\tau), \\ \frac{d^2\vec{r}(\tau)}{d\tau^2} &= \underbrace{\frac{1}{\ell}}_{\kappa} \underbrace{\langle \cosh(\tau/\ell), \sinh(\tau/\ell) \rangle}_{\hat{N}} \equiv \vec{a}(\tau).\end{aligned}$$

The timelike unit tangent $\hat{T} = \hat{u}$ is called the spacetime velocity of the world line, and its derivative the spacetime acceleration $\vec{a} = d\hat{u}/d\tau$. The spacelike unit normal is the direction of the acceleration and is along the spacelike pseudo-radial direction pointing away from the center of the hyperbola at the origin (it is obvious that $\hat{N} \cdot \hat{N} = 1$), on the opposite side of the curve from the direction in which the normal is pointing, in contrast with the circular case where the center is on the same side as the direction in which the normal is pointing. The curvature κ

(magnitude of the spacetime acceleration) is just the reciprocal of the pseudo-radius ℓ of the hyperbola.

However, the proper time derivative of the position is a timelike unit vector interpreted as the spacetime velocity $\hat{u} = \hat{T}$ and its derivative in turn is the spacetime acceleration $\vec{a} = d\hat{u}/d\tau = d^2\vec{r}/d\tau^2$. The constant curvature $\kappa = 1/\ell$ is then interpreted as the magnitude $|\vec{a}|$ of the spacetime acceleration. The pseudo-radius of curvature is then the reciprocal of that scalar acceleration, equal to the pseudoradius of the hyperbola. The 1-dimensional motion of a charged particle along the direction of a constant uniform electric field is characterized by constant spacetime acceleration, and its path in the Lorentz plane is such a hyperbola.

Note the space velocity of the world line is just the coordinate time derivative of the position, namely

$$v = \frac{dx}{dt} = \frac{dx/d\tau}{dt/d\tau} = \tanh(\tau/\ell) \equiv \tanh \beta,$$

which defines the hyperbolic angle parametrizing this timelike unit vector, called the rapidity β . Note that the relation $\tau = \ell\beta$ is the Lorentz analog of the arclength relation to the trigonometric angle and arc radius $s = a\theta$.

Remark.

From the successive hyperbolic identities $\cosh^2 \beta - \sinh^2 \beta = 1$, $1 - \tanh^2 \beta = 1/\cosh^2 \beta$, $\cosh \beta = 1/(1 - \tanh^2 \beta)^{1/2}$, one has the relation $\cosh \beta = (1 - v^2)^{-1/2} \equiv \gamma$. Then $\sinh \beta = \tanh \beta \cosh \beta = \gamma v$. Thus we can rewrite the three hyperbolic functions in terms of the space velocity as

$$(\gamma, \gamma v, v) = (\cosh \beta, \sinh \beta, \tanh \beta).$$

For 1-dimensional motion in time like this example, both v and β can be negative, but these same formulas can be extended to the speed $v \geq 0$ for spatial motion in more dimensions, letting $\beta \geq 0$.

▲

Exercise C.0.14.

Twin paradox family of hyperbolas

Suppose we want to find the family of pseudo-circular timelike hyperbolas centered on the x -axis which connect the two points $(x, t) = (0, -1), (0, 1)$ separated by a proper time of 2 units along the time axis as illustrated in Fig. C.2. These represent uniformly accelerated (decelerated) world lines which start out at the earlier time at the origin of spatial coordinates and then move away and then return to the same spatial origin at a later time. We would like to evaluate the elapsed proper time along each of these possible world lines connecting the two events. This will show that the maximum time interval is the unaccelerated straight line path along the time axis itself, the point at rest in this reference frame.

a) These hyperbolas are of the form: $(x - a)^2 - t^2 = b^2$. The two parameters may be determined as a function of the intercept c on the x -axis by requiring that the points $(c, 0)$ and $0, \pm 1$) satisfy the equation. Show that the following is a solution

$$a = \frac{1 + c^2}{2c}, \quad b = \frac{1 - c^2}{2c}, \quad -1 < c < 1.$$

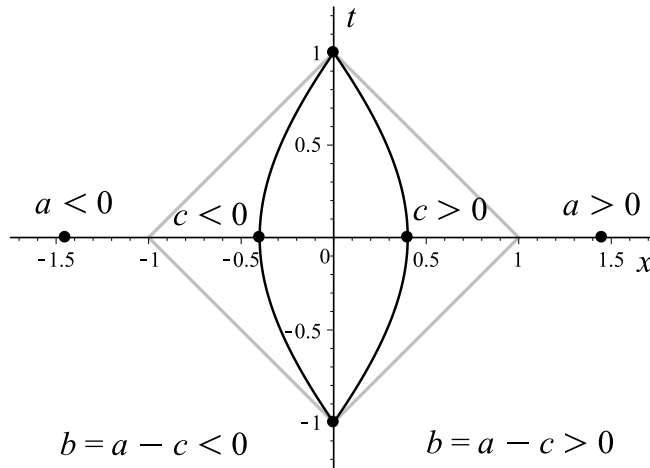


Figure C.2: The family of pseudocircular hyperbolas passing through two points on the time axis with centers at $x = a$ on the x -axis, parametrized by their intercepts $x = c$ with that axis, with $-1 \leq c \leq 1$. The difference $b = a - c$ is either > 0 (for $c > 0$) or < 0 (for $c < 0$). Its absolute value $|b|$ is the pseudoradius of the hyperbola. The path $c = 0$ along the time axis has the longest length in the Lorentzian geometry.

Note that b has the same sign as c , and $a - c = b$ is the signed distance of the center of the hyperbola from the vertex along the x -axis, as shown in Fig. C.2.

b) By setting $x - a = -b \cosh \lambda = a - b \cosh(\tau/b)$ and $t = |b| \cosh \lambda = |b| \cosh(\tau/b)$, we get a proper time parametrization of these hyperbolas for which t increases with τ . We have already shown that the τ derivative of the position vector leads to the unit tangent (future-pointing spacetime velocity in this case) and the second such derivative to the acceleration which has magnitude $\kappa = 1/|b|$. By symmetry the initial and final proper times corresponding to the initial and final events $t = \pm 1$ on the time axis are $\tau_{\pm} = \pm b \operatorname{arcsinh}(1/b)$, with a total elapsed proper time of $\Delta\tau = \tau_+ - \tau_- = 2b \operatorname{arcsinh}(1/b) = 2\kappa^{-1} \operatorname{arcsinh} \kappa$. Show that as $c \rightarrow 0$, then $\kappa \rightarrow 0$ and $\delta\tau \rightarrow 2^-$ and that as $c \rightarrow \pm 1$ and $\kappa \rightarrow \infty$ and $\Delta\tau \rightarrow 0$.

c) Plot this proper time interval as a function of c for $-1 < c < 1$. This can be interpreted as the amount by which a twin ages in a rocket which speeds away from his or her twin at $x = 0$ and returns 2 time units later in the clock time of the twin left behind. The traveling twin returns younger than the stationary twin. By increasing this constant acceleration, one can make aging of the moving twin as small as desired.

■

The Lorentz helix

So let's return to our Euclidean helix in \mathbb{R}^3 , and re-examine the same curve in a Lorentzian geometry. Since we already used the physics motion language in discussing the Euclidean helix, suppose we consider circular motion in the x - y plane but now reimagine the z -axis as the classical time t -axis so we can show the position in the plane at different times using the extra dimension. The helix becomes a timelike curve in 3-dimensional Minkowski spacetime.

In order to handle the angular velocity of this motion (now that t is the name for what used to be z), we cannot continue to use t for our parameter, so let's call it λ . Then

$$\begin{aligned}\vec{r} &= \langle x, y, t \rangle = \langle a \cos(\lambda), a \sin(\lambda), c\lambda \rangle \\ &= \left\langle a \cos\left(\frac{t}{c}\right), a \sin\left(\frac{t}{c}\right), t \right\rangle \equiv \langle a \cos(\Omega t), a \sin(\Omega t), t \rangle\end{aligned}$$

shows that the angular velocity of the circular motion about the vertical axis in 3-space (the dimension we have suppressed) is therefore $\Omega = 1/c$. The tangent is

$$\vec{r}'(t) = \langle -a\Omega \sin(\Omega t), a\Omega \cos(\Omega t), 1 \rangle,$$

and the first two components are the velocity in space of the circular motion in the x - y plane, showing that $v = |a\Omega| = |a/c|$ is the speed, which can have any value in nonrelativistic physics.

This spacetime-diagram for the circular motion has horizontal time planes and although we might consider a transformation to a moving observer in relative motion at constant velocity by defining $\langle \bar{x}, \bar{y} \rangle = \langle x, y \rangle + \vec{V}t$, the time does not change in Newtonian physics. Newtonian time is universal, the same for all unaccelerated (read “inertial”) observers in constant relative motion.

Not so in relativistic physics, as we saw in the previous sections of appendix A. In the 3-dimensional Minkowski spacetime of special relativity with one spatial dimension suppressed, like for motion in a plane as occurs in orbits around central forces, we also have the same helical motion in the 3-spacetime but assuming we use geometrical units for time, the speed for any massive particle is limited by the speed of light, which is 1 in those units: $v = |a\Omega| < 1$. This is a timelike curve in the spacetime. Of course we can also consider spacelike curves which have other interpretations, not as the motion of a particle. Indeed for spacelike curves, the Frenet-Serret relations are very similar to what we have already done in the Euclidean case including an osculating circle similarly defined, so let's study the timelike case where instead we have an osculating pseudo-circle, namely a hyperbola.

The only thing we need to change is the places where the Euclidean dot product entered our calculations and replace it by the Lorentzian dot product, and also reinterpret the cross-product to generate a Lorentzian orthogonal vector. Let's replace the index 3 by 0 but leave it at the end for comparison with the Euclidean calculations we have already done

$$\vec{X} \cdot \vec{Y} = \langle X^1, X^2, X^0 \rangle \cdot \langle Y^1, Y^2, Y^0 \rangle = X^1 Y^1 + X^2 Y^2 - X^0 Y^0 \equiv \eta_{ij} X^i Y^j,$$

where

$$(\eta_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = (\eta^{ij})$$

is the matrix of components of the flat Lorentz metric on \mathbb{R}^3 , replacing the unit matrix for the Euclidean metric, and we are using the Einstein summation convention in which repeated indices, one up and one down in a formula, are summed over their allowed range, as discussed in Chapter 0.

The cross-product is also easily handled through the triple scalar product, which for X and Y fixed is just a linear function of the third vector Z , and hence is by definition a covector. This triple scalar product is just the determinant of the matrix whose columns are the components of the 3 vectors, using Maple notation for this matrix

$$(\vec{X} \times \vec{Y}) \cdot \vec{Z} = \det\langle \vec{X} | \vec{Y} | \vec{Z} \rangle = \epsilon_{ijk} X^i Y^j Z^k = [\vec{X} \times \vec{Y}]_k Z^k \equiv [\vec{X} \times \vec{Y}]^\sharp \cdot \vec{Z}.$$

The vector $[\vec{X} \times \vec{Y}]^\sharp$ obtained from this covector is such that its dot product with \vec{Z} is the value on \vec{Z} of the linear function represented by the cross product. This requires only reversing the sign of the last component of the cross-product vector associated with the minus sign in the dot product, so that when we use the new dot product with the last component multiplied by a minus sign, we get the same result as in the Euclidean case. As we will learn in Chapter 1, this corresponds to raising the lowered “covariant” index on the cross-product to produce a vector with an upper “contravariant” index which is automatically orthogonal to either of them

$$\begin{aligned} [\vec{X} \times \vec{Y}]_k Z^k &\equiv \eta_{ik} [\vec{X} \times \vec{Y}]^\sharp{}^i Z^k, \\ [\vec{X} \times \vec{Y}]^\sharp{}^i &\equiv [\vec{X} \times \vec{Y}]^\sharp{}^i = \eta^{ik} [\vec{X} \times \vec{Y}]_k = \eta^{ik} \epsilon_{kmn} X^m Y^n. \end{aligned}$$

Since repeating any vector in a triple cross product gives zero, this new vector with its last component sign-reversed will be orthogonal to both \vec{X} and \vec{Y} in the Lorentzian geometry and hence to the plane they span. However, if one wishes to have a future-pointing vector that still obeys the right hand rule, one must reverse the sign of the whole vector to achieve this correspondence.

The helix parametrized by the third coordinate t is then

$$\vec{r}(t) = \langle a \cos(t/c), a \sin(t/c), t \rangle = \langle a \cos(\Omega t), a \sin(\Omega t), t \rangle.$$

so its first two derivatives are

$$\begin{aligned} \vec{r}'(t) &= \left\langle -\frac{a}{c} \sin\left(\frac{t}{c}\right), \frac{a}{c} \cos\left(\frac{t}{c}\right), 1 \right\rangle = \langle -\Omega a \sin(\Omega t), \Omega a \cos(\Omega t), 1 \rangle, \\ \vec{r}''(t) &= \frac{a}{c^2} \left\langle -\cos\left(\frac{t}{c}\right), -\sin\left(\frac{t}{c}\right), 0 \right\rangle = \Omega^2 a \langle -\cos(\Omega t), -\sin(\Omega t), 0 \rangle, \end{aligned}$$

and the self-dot product of the tangent vector is

$$\vec{r}'(t) \cdot \vec{r}'(t) = \frac{a^2}{c^2} - 1 = \frac{a^2 - c^2}{c^2} = \Omega^2 a^2 - 1$$

which must be negative to interpret as the tangent to a timelike world line, and the square root of its absolute value is then the length

$$\frac{ds}{dt} = |\vec{r}'(t)| = \sqrt{\left| \frac{a^2 - c^2}{c^2} \right|} = \sqrt{1 - \frac{a^2}{c^2}} \equiv \gamma^{-1} < 1,$$

which defines the differential of spacetime arclength along the curve. It is useful to introduce the spatial speed $v = a/c = \Omega a$ and spatial velocity $\vec{v} = \langle v \cos(t/c), v \sin(t/c), 0 \rangle$, in terms of which the Lorentz gamma factor is

$$\gamma = \frac{1}{\sqrt{1-v^2}} = \frac{1}{\sqrt{1-(a/c)^2}} = \frac{|c|}{\sqrt{c^2-a^2}}.$$

The unit tangent is then

$$\hat{T}(t) = \left\langle -\gamma \frac{a}{c} \sin\left(\frac{t}{c}\right), \gamma \frac{a}{c} \cos\left(\frac{t}{c}\right), 1 \right\rangle = \langle -\gamma \Omega a \sin(\Omega t), \gamma \Omega a \cos(\Omega t), 1 \rangle.$$

We can also reparametrize the curve using the arclength interpreted as the proper time along the curve as described in the previous appendices

$$\tau = \gamma^{-1}t = \frac{\sqrt{c^2-a^2}}{|c|}t, \quad t = \gamma\tau,$$

so

$$\vec{r}(\tau) = \langle a \cos(\gamma\tau/c), a \sin(\gamma\tau/c), 1\tau \rangle = \langle a \cos(\gamma\Omega\tau), a \sin(\gamma\Omega\tau), \gamma\tau \rangle,$$

which defines the proper time angular velocity $\Omega_o \equiv \gamma\Omega$. Similarly the period $T = 2\pi/\Omega$ and proper period of the motion $T_o = 2\pi/\Omega_o = \gamma T$ are defined in the obvious way. Then

$$\begin{aligned} \vec{r}'(\tau) &= \gamma \left\langle -\frac{a}{c} \sin\left(\frac{\gamma\tau}{c}\right), \frac{a}{c} \cos\left(\frac{\gamma\tau}{c}\right), 1 \right\rangle \\ &= \gamma \langle -\Omega a \sin(\gamma\Omega\tau), \Omega a \cos(\gamma\Omega\tau), 1 \rangle = \hat{T}(\tau) \equiv \hat{U}(\tau), \\ \vec{r}''(\tau) &= \gamma^2 \frac{a}{c^2} \left\langle -\cos\left(\frac{\gamma\tau}{c}\right), -\sin\left(\frac{\gamma\tau}{c}\right), 0 \right\rangle \\ &= \underbrace{\gamma^2 \Omega^2 a}_{\kappa} \underbrace{\langle -\cos(\gamma\Omega\tau), -\sin(\gamma\Omega\tau), 0 \rangle}_{\hat{N}(\tau)} \equiv \vec{A}(\tau), \end{aligned}$$

so we can identify the magnitude of the spacelike spacetime acceleration as the curvature

$$\kappa(\tau) = \frac{a}{c^2-a^2} = \gamma^2 \Omega^2 a = \gamma^2 v^2/a$$

and radius of curvature

$$\rho(\tau) = 1/\kappa(\tau) = \frac{c^2-a^2}{a} = \gamma^{-2} \Omega^{-2} a^{-1} = a \gamma^{-2} v^{-2}.$$

This agrees with the Euclidean result with the substitution $-a^2 \rightarrow a^2$.

As before we can get a binormal from the cross-product of the first two derivatives, provided we reverse the sign of the timelike component to get a spacelike spacetime vector which will be

orthogonal in the new dot product (which multiplies that component by another sign to give the same result as in the Euclidean case for the same helical curve),

$$\begin{aligned}\hat{B} &= \frac{\vec{r}'(t) \times_{\sharp} \vec{r}''(t)}{|\vec{r}'(t) \times_{\sharp} \vec{r}''(t)|} = \frac{1}{\sqrt{c^2 - a^2}} \langle c \sin(\gamma \Omega \tau), -c \cos(\gamma \Omega \tau), -a \rangle \\ &= \gamma \operatorname{sgn}(c) \langle \sin(\gamma \Omega \tau), -\cos(\gamma \Omega \tau), -v \rangle.\end{aligned}$$

Its proper time derivative is

$$\begin{aligned}\frac{d\hat{B}}{d\tau} &= \frac{\gamma \Omega c}{\sqrt{c^2 - a^2}} \langle \cos(\gamma \Omega \tau), \sin(\gamma \Omega \tau), 0 \rangle \\ &= \underbrace{(\gamma^2/c)}_{\omega_{\text{fw}}} \underbrace{\langle \cos(\gamma \Omega \tau), \sin(\gamma \Omega \tau), 0 \rangle}_{-\hat{N}(\tau)},\end{aligned}$$

allowing us to identify the spacetime torsion,

$$\omega_{\text{fw}} = \frac{c}{c^2 - a^2},$$

which agrees with the Euclidean result with the substitution $-a^2 \rightarrow a^2$. We use the symbol ω_{fw} instead of τ since the later symbol is already being used for the proper time along the curve, and because this is called the Fermi-Walker angular velocity as will be explained below. The remaining proper time derivative is

$$\frac{d\hat{N}}{d\tau} = \kappa \hat{T} + \omega_{\text{fw}} \hat{B},$$

where the two coefficients follow from the orthogonality relations exactly as in the Euclidean case.

Thus we have the Lorentz version of the Frenet-Serret relations for a timelike curve in 3-dimensional Minkowski spacetime

$$\begin{aligned}\frac{d\hat{T}}{d\tau} &= \kappa \hat{N}, \\ \frac{d\hat{N}}{d\tau} &= \kappa \hat{T} + \omega_{\text{fw}} \hat{B}, \\ \frac{d\hat{B}}{d\tau} &= -\omega_{\text{fw}} \hat{N}.\end{aligned}$$

or in matrix form

$$\begin{pmatrix} \frac{d\hat{T}}{ds} & \frac{d\hat{N}}{ds} & \frac{d\hat{B}}{ds} \end{pmatrix} = \begin{pmatrix} \hat{T} & \hat{N} & \hat{B} \end{pmatrix} \begin{pmatrix} 0 & \kappa & 0 \\ \kappa & 0 & -\omega_{\text{fw}} \\ 0 & \omega_{\text{fw}} & 0 \end{pmatrix}.$$

such that if we express a vector along the curve in terms of its components with respect to the the frame vectors

$$\vec{X} = X^1 \hat{T} + X^2 \hat{N} + X^3 \hat{B}$$

then in matrix form

$$\frac{d}{ds} \begin{pmatrix} X^1 \\ X^2 \\ X^3 \end{pmatrix} = \begin{pmatrix} 0 & \kappa & 0 \\ \kappa & 0 & -\omega_{\text{fw}} \\ 0 & \omega_{\text{fw}} & 0 \end{pmatrix} \begin{pmatrix} X^1 \\ X^2 \\ X^3 \end{pmatrix}$$

The curvature/acceleration scalar boosts the unit tangent vector (called the 4-velocity in the 4-dimensional setting) to remain tangent to the curve with respect to a constant (parallel transported) vector along the curve, while the Fermi-Walker angular velocity is the instantaneous rotation in the normal plane (called the local rest space of the observer following the world line) compared to axes which are momentarily constant.

Notice that the upper left 2×2 block of the coefficient matrix in this vector differential equation is symmetric, generating a hyperbolic rotation. In fact if we lower the first index of the components of this linear transformation matrix, so both indices are covariant, the entire matrix becomes antisymmetric, true for all pseudo-orthogonal matrix derivatives in this context.

For a timelike curve we can introduce an osculating pseudocircle (hyperbola) in the plane of the unit tangent and unit normal, the velocity-acceleration plane, in the same way except for the fact that the center of a hyperbola is on the opposite side of the vertex along the normal line compared to the Euclidean case

$$\vec{C}(\tau) = \vec{r}(\tau) - \rho(\tau) \vec{N}(\tau) = \frac{c}{a} \langle c \cos(\Omega t), c \sin(\Omega t), at \rangle = \left\langle \frac{a}{v^2} \cos(\gamma \Omega \tau), \frac{a}{v^2} \sin(\gamma \Omega \tau), \gamma v \tau \right\rangle ,$$

Again the tangent hyperbola which starts at $\alpha = 0$ at the point of tangency and moves along $\hat{T}(\tau)$ as α increases is

$$\begin{aligned} \vec{r}_C(\alpha, \tau) &= \vec{C}(\tau) + \rho(\tau) \left(\cosh(\alpha) \hat{N}(\tau) + \sinh(\alpha) \vec{T}(\tau) \right) \\ &= \frac{1}{a} \langle \cos(\Omega t)(c^2 - (c^2 - a^2) \cosh(\alpha)) + a\sqrt{c^2 - a^2} \sin(\Omega t) \sinh(\alpha), \\ &\quad \sin(\Omega t)(c^2 - (c^2 - a^2) \cosh(\alpha)) + a\sqrt{c^2 - a^2} \cos(\Omega t) \cosh(\alpha), \\ &\quad cat + c\sqrt{c^2 - a^2} \sinh(\alpha) \rangle \\ &= \frac{1}{a} \langle \cos(\gamma \Omega \tau)(c^2 - (c^2 - a^2) \cosh(\alpha)) + a\sqrt{c^2 - a^2} \sin(\gamma \Omega \tau) \sinh(\alpha), \\ &\quad \sin(\gamma \Omega \tau)(c^2 - (c^2 - a^2) \cosh(\alpha)) + a\sqrt{c^2 - a^2} \cos(\gamma \Omega \tau) \cosh(\alpha), \\ &\quad \gamma ca\tau + c\sqrt{c^2 - a^2} \sinh(\alpha) \rangle . \end{aligned}$$

Fig. C.3 shows this at the half period point on one cycle of the helix. The osculating plane is identical to the Euclidean case, since the plane of the first and second derivatives of the parametrized helix is independent of the geometry.

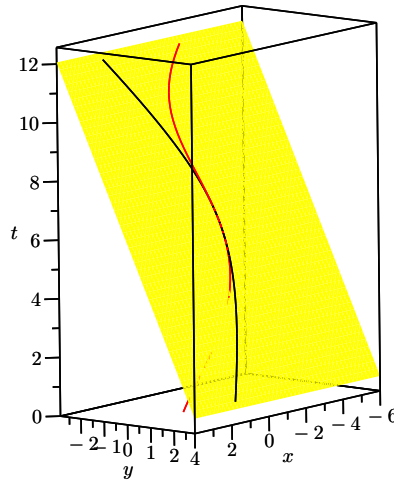


Figure C.3: The Lorentzian helix (red) with its osculating hyperbola (black), including a parallelogram from the velocity-acceleration plane (osculating plane) which contains that hyperbola.

Exercise C.0.15.
magnetic helix

This problem requires first reading Appendix A.3 which reviews how a unit tangent and unit normal behave along an arclength parametrized curve or in any parametrization of a curve in ordinary space, extending those notions to the Minkowski spacetime case.

In a uniform constant magnetic field, a charged particle moves in a helix, spiraling around an axis parallel to the magnetic field lines. Suppose the magnetic field is aligned with the z -axis $\langle B_1, B_2, B_3 \rangle = \langle 0, 0, B \rangle$, and the particle of charge q has constant transverse speed v_\perp and constant velocity v_3 along the z -axis, for a total speed of $v = (v_\perp^2 + v_3^2)^{1/2}$. Define the gyration radius $R = mv_\perp/(qB)$ and Lorentz gamma factor $\gamma = (1 - v_\perp^2 - v_3^2)^{-1/2} \equiv dt/d\tau$. The unit 4-velocity is just

$$u = \frac{dx}{d\tau}, \quad u^0 = \frac{dt}{d\tau} = \gamma.$$

By appropriate choice of the initial conditions

$$\langle x^0(0), x^1(0), x^2(0), x^3(0) \rangle = \langle 0, R, 0, 0 \rangle, \quad \langle u^0(0), u^1(0), u^2(0), u^3(0) \rangle = \gamma \langle 0, 0, v_\perp, v_3 \rangle,$$

the particle trajectory is the following world line in Minkowski spacetime specified by giving the inertial coordinates as the following functions of the inertial time t

$$x = \langle x^0, x^1, x^2, x^3 \rangle = \langle t, R \cos \left(\frac{v_\perp t}{R} \right), R \sin \left(\frac{v_\perp t}{R} \right), v_3 t \rangle.$$

The azimuthal angle is $\phi = v_{\perp}t/R = \gamma v_{\perp}\tau/R$ so the inertial time angular velocity is $d\phi/dt = \omega = v_{\perp}/R = qB/m$ (called the Larmour frequency), while the proper time angular velocity is $d\phi/d\tau = \omega_o = \gamma\omega$. Since ω is independent of the velocity, for a uniform magnetic field, charged particles at different radii from the axis aligned with that magnetic field all spiral around the field lines at this same frequency.

a) This is both a helix around the time axis as well as around the z axis—in fact it is an ordinary helix around the new time axis in which the new z coordinate moves with the same velocity v_3 as the particle

$$\text{motion along } z\text{-axis:} \quad \langle x^0, x^1, x^2, x^3 \rangle = \langle t, 0, 0, v_3 t \rangle,$$

which has unit tangent $U = \langle 1, 0, 0, v_3 \rangle / (1 - v_3^2)^{1/2}$. In fact the helical particle trajectory lies in the hyperplane $n_i x^i = -v_3 t + z = 0$, which has a normal $\langle n^i \rangle = \langle v_3, 0, 0, 1 \rangle$ which can be normalized to a unit normal $E_3 = \langle v_3, 0, 0, 1 \rangle / (1 - v_3^2)^{1/2}$. Show that $E_3 \cdot E_3 = 1$ and that $E_3 \cdot U = 0$.

b) Reparametrize this helix by the proper time $\tau = \gamma t$ so that the curve is parametrized by the spacetime arclength.

c) Evaluate the 4-velocity

$$e_0 \equiv u = \frac{dx}{d\tau}$$

and show that $u \cdot u = -1$. Verify that $u = \gamma(\hat{t} + v_{\perp}\hat{\phi} + v_3\hat{z})$, where $\hat{\phi} = \langle 0, -\sin\phi, \cos\phi, 0 \rangle$ and $\phi = v_{\perp}t/R = \gamma v_{\perp}\tau/R$. Thus confirm that the spatial speed is $v = \|v_{\perp}\hat{\phi} + v_3\hat{z}\| = (v_{\perp}^2 + v_3^2)^{1/2}$.

d) Evaluate the 4-acceleration

$$a = \frac{du}{d\tau}$$

and its magnitude $\kappa \equiv \|a\|$ and show that $a \cdot u = 0$. Evaluate the direction unit vector of the acceleration $e_1 = a/\kappa$.

e) Let F be the mixed electromagnetic field tensor associated with zero electric field and this magnetic field. Show that this world line satisfies the Lorentz force law

$$m \frac{du}{d\tau} = q \underline{F} \cdot u.$$

f) The Serret-Frenet relations for a timeline world line in Minkowski spacetime with unit tangent $u \equiv e_0$ has a unit normal $e_1 = a/\|a\|$ and two binormals e_2 and e_3 which together make an orthonormal set of vectors which satisfy the relations

$$\frac{d}{d\tau} \begin{pmatrix} e_0 \\ e_1 \\ e_2 \\ e_3 \end{pmatrix} = \begin{pmatrix} 0 & \kappa & 0 & 0 \\ \kappa & 0 & \tau_1 & 0 \\ 0 & -\tau_1 & 0 & \tau_2 \\ 0 & 0 & -\tau_2 & 0 \end{pmatrix} \begin{pmatrix} e_0 \\ e_1 \\ e_2 \\ e_3 \end{pmatrix},$$

where $\kappa \geq 0$ is the curvature and τ_1, τ_2 are the first and second torsions. This generalizes the 3-dimensional discussion in Appendix C and as a matrix, itself corresponds to a mixed electromagnetic field component matrix with $E_1 \rightarrow \kappa, B_3 \rightarrow \tau_1, B_1 \rightarrow \tau_2$ and all other components

zero. In order for this basis to remain orthonormal along the world line, its rate of change must be the result of matrix multiplication by a matrix of this form as explained in section 1.7. Similarly for the unit vector u to remain a unit vector along the world line, it can only undergo a pseudo-rotation (Lorentz transformation) which preserves its length, so its rate of change along the world line can only allow its tip to pseudo-rotate, which is what this matrix does.

Use a computer algebra system to evaluate $de_1/d\tau$. Use the third Frenet-Serret equation to define $\tau_1 e_2$, knowing κ . Let $\tau_1 = ||\tau_1 e_2|| \geq 0$ and evaluate the corresponding unit vector e_2 .

g) Evaluate $de_2/d\tau$ and show that by setting $\tau_2 = 0$ we satisfy the third Frenet-Serret relation. If we set $e_3 = E_3$ (a constant vector so that the final relation is satisfied), show that the we complete these first 3 unit vectors to an orthonormal basis which satisfies the full set of Frenet-Serret relations, i.e., show that E_3 is orthogonal to e_0, e_1, e_2 . The last torsion measures the rotation of the normal to the 3-plane spanned by $u, a, da/d\tau$, but since the motion is confined to a hyperplane, this is zero.

h) Show that the first torsion equals

$$\tau_1 = \omega_o(1 + \gamma^2 v_\perp^2)^{1/2} = \omega_o \frac{\gamma}{\gamma_3}, \quad \gamma_3 = (1 - v_3^2)^{-1/2}$$

by combining both terms inside the square root expression, which multiplies the proper time angular velocity by the ratio of the gamma factor of the particle rest frame and the rest frame of the axis of the helix.

■

Appendix D

Surfaces in 3-space

Surfaces play a fundamental role in differential geometry, so it is important to recall their basics from multivariable calculus. In my university, parametrized surfaces are not even covered, so it is important to extend what we do teach about graphs of functions of two variables to that case. The first encounter with multivariable functions after calculus of a single variable is with functions of two independent variables, which we visualize through their graphs in space, adding one extra “dependent” variable: $z = f(x, y)$. For each point in the x - y plane, we plot the value of the function in the z -direction to create a surface which we can think of as “parametrized” by the coordinates x and y in the same sense that a curve in space is parametrized by a single variable: $\langle x, y, z \rangle = \langle x(t), y(t), z(t) \rangle$. This analogy requires us to write the position vector of points on this graph in the form

$$\langle x, y, z \rangle = \langle x, y, f(x, y) \rangle,$$

but we can more easily understand this as a “parametrization” by naming the parameters with different variable names than the coordinates with which they agree

$$\langle x, y, z \rangle = \langle t_1, t_2, f(t_1, t_2) \rangle,$$

This is also exactly how we can view single variable function graphs as parametrized curves: $y = f(x)$ becomes $\langle x, y \rangle = \langle t, f(t) \rangle$ if we simply rename $x = t$. However, for surfaces it is more usual to use the letters (u, v) for the parameters, or (u^1, u^2) if we want to use numbered variable names, like $(x, y, z) = (x^1, x^2, x^3)$. Let’s adopt this notation to get used to how we will treat multiple objects in this book. Our function graph then looks like

$$\vec{r} = \langle x^1, x^2, x^3 \rangle = \langle u^1, u^2, f(u^1, u^2) \rangle.$$

Partial derivatives of the function f are then introduced in this context and visualized in terms of the slopes of the vertical plane cross-sections that result from holding one of the two Cartesian coordinates fixed. In fact, these are just space curves, which we already discussed in the previous appendix. Namely we can think of the graph as a 1-parameter family of curves in two ways, and for each one we can introduce their tangent vectors. Holding u^2 fixed, for each value we get a curve parametrized by u^1 , and vice versa. Graphing software for 3d plotting of

graphs like this uses equally spaced values of each coordinate to imprint the Cartesian coordinate grid in the x - y plane onto the graph above it to give it more 3-dimensional perspective.

One then figures out a way to describe the tangent plane to the graph at each point, obtaining a normal vector to the plane. Let's try in this new notation. Here are the tangent vectors to the two families of parameter curves

$$\begin{aligned}\vec{r}_1(u^1, u^2) &\equiv \frac{\partial \vec{x}}{\partial u^1}(u^1, u^2) = \left\langle 1, 0, \frac{\partial f}{\partial u^1}(u^1, u^2) \right\rangle, \\ \vec{r}_2(u^1, u^2) &\equiv \frac{\partial \vec{x}}{\partial u^2}(u^1, u^2) = \left\langle 0, 1, \frac{\partial f}{\partial u^2}(u^1, u^2) \right\rangle.\end{aligned}$$

Extending these vectors from the point on the graph where we have evaluated them where we imagine their initial points are located, we get the two tangent lines to those curves. These two intersecting lines determine the tangent plane, whose normal vector is just the cross-product of the two tangent vectors, in this order yielding an upward normal since its third component is positive

$$\vec{N}(u^1, u^2) = \vec{r}_1(u^1, u^2) \times \vec{r}_2(u^1, u^2) = \left\langle -\frac{\partial f}{\partial u^1}(u^1, u^2), -\frac{\partial f}{\partial u^2}(u^1, u^2), 1 \right\rangle.$$

The first two components are just the sign-reversal of the gradient of the function in the x - y plane,

$$\vec{\nabla} f(x^1, x^2) = \left\langle \frac{\partial f}{\partial x^1}(x^1, x^2), \frac{\partial f}{\partial x^2}(x^1, x^2) \right\rangle,$$

pointing in the direction in which the function decreases, exactly right since the upward normal must tilt backwards to the direction in which the function graph is increasing.

A more direct route to the normal vector bypassing the cross product is by introducing the function $F(x^1, x^2, x^3) = x^3 - f(x^1, x^2)$ whose level surface $F(x^1, x^2, x^3) = 0$ is the graph of f , and the gradient of this new function is orthogonal to the surface

$$\begin{aligned}\vec{N}(x^1, x^2) &= \vec{\nabla} F(x^1, x^2, x^3) = \left\langle \frac{\partial F}{\partial x^1}(x^1, x^2, x^3), \frac{\partial F}{\partial x^2}(x^1, x^2, x^3), \frac{\partial F}{\partial x^3}(x^1, x^2, x^3) \right\rangle \\ &= \left\langle -\frac{\partial f}{\partial x^1}(x^1, x^2), -\frac{\partial f}{\partial x^2}(x^1, x^2), 1 \right\rangle.\end{aligned}$$

In fact the real meaning of the gradient is connected to the chain rule for the derivative of a function along a curve. If $\vec{r} = \vec{r}(t)$ is a curve through the point (x^1, x^2, x^3) such that $\vec{r}(t_0) = \langle x^1, x^2, x^3 \rangle$, then the derivative of any function $F(x^1, x^2, x^3)$ along this curve there is just

$$\begin{aligned}\left. \frac{dF}{dt}(\vec{r}(t)) \right|_{t=t_0} &= \left(\frac{dF}{dx^1} \frac{dx^1}{dt} + \frac{dF}{dx^2} \frac{dx^2}{dt} + \frac{dF}{dx^3} \frac{dx^3}{dt} \right) \Big|_{t=t_0} \\ &= \vec{\nabla} F(\vec{r}(t_0)) \cdot \vec{r}'(t_0),\end{aligned}$$

where we use the notation $F(\vec{r}) = G(x^1, x^2, x^3)$ to reduce the length of our formulas. If $\vec{X} = \vec{r}'(t_0)$ is such a tangent vector, this becomes the directional derivative of G along \vec{X}

$$\vec{\nabla} F(x^1, x^2, x^3) \cdot \vec{X} \equiv \left(\nabla_{\vec{X}} F \right) (x^1, x^2, x^3).$$

Although usually this formula is restricted to a unit vector \vec{X} , there is no reason we cannot use the same formula for any vector \vec{X} , and this shows that the gradient of G at the point (x^1, x^2, x^3) is really just a linear function of tangent vectors \vec{X} there, namely the components of the gradient are just the coefficients of the vector in this directional derivative which defines a linear function like any set of coefficients

$$a_1 X^1 + a_2 X^2 + a_3 X^3.$$

In Chapter 1 we learn to call this a covector, or covariant vector, to distinguish the linear function role of the vector in producing a real number from another vector through what we normally denote as the dot product of the two vectors.

We can jump to a general parametrized surface by simply letting all three coordinates be functions of two parameters which are no longer associated with the first two coordinates

$$\vec{r}(u^1, u^2) = \langle x^1(u^1, u^2), x^2(u^1, u^2), x^3(u^1, u^2) \rangle.$$

We still have a 1-parameter family of curves in two senses to form a parameter grid on the surface, and as long as the relationship between points in the parameter space (the u^1 - u^2 plane) and the image points in the 3-space of the Cartesian coordinates x^1, x^2, x^3 is one-to-one, we can think of the two parameters as coordinates on the surface. We can in principle invert the relationship to identify unique values of (u^1, u^2) with each point in the parametrized surface, although in practice it may not actually be possible to solve the relationship in closed form.

Again we can introduce the two tangent vectors to the grid lines

$$\begin{aligned}\vec{r}_1(u^1, u^2) &\equiv \frac{\partial \vec{x}}{\partial u^1}(u^1, u^2), \\ \vec{r}_2(u^1, u^2) &\equiv \frac{\partial \vec{x}}{\partial u^2}(u^1, u^2),\end{aligned}$$

and the tangent plane is again the span of this set of two vectors, whose normal can be found by the cross product

$$\vec{N}(u^1, u^2) = \vec{r}_1(u^1, u^2) \times \vec{r}_2(u^1, u^2).$$

Recall that the cross product of two vectors \vec{a} and \vec{b} has a magnitude which is the area of the natural parallelogram formed by the two vectors as the two sides adjacent to one vertex of a parallelogram, while the direction of their cross product gives a normal to the plane they form. Thus

$$dS = |\vec{N}(u^1, u^2)| du^1 du^2$$

is the differential area of the parallelogram formed in the space of vectors at the point $R(u^1, u^2)$ by the vectors $\vec{r}_1(u^1, u^2) du^1$ and $\vec{r}_2(u^1, u^2) du^2$ tangent to the surface grid lines. In the limit of very small differentials of the parameters, the actual grid lines corresponding to these differentials form a figure which gets closer and closer to the parallelogram in the tangent space. Integrating this over a region of the parameter space gives the surface area of the corresponding

part of the surface. This is very similar to finding the arclength of a parametrized curve by integrating the length of the tangent vector, where the differential of arclength is

$$ds = |\vec{r}'(t)|.$$

Take the sphere of radius a for example, easily parametrized in spherical coordinates

$$\begin{aligned}\vec{r}(\theta, \phi) &= \langle a \sin \theta \cos \phi, a \sin \theta \sin \phi, a \cos \theta \rangle, \\ \vec{r}_1(\theta, \phi) &= \langle a \cos \theta \cos \phi, a \cos \theta \sin \phi, -a \sin \theta \rangle, \\ \vec{r}_2(\theta, \phi) &= \langle -a \sin \theta \sin \phi, a \sin \theta \cos \phi, 0 \rangle,\end{aligned}$$

and

$$\vec{N}(\theta, \phi) = \vec{r}_1(\theta, \phi) \times \vec{r}_2(\theta, \phi) = \langle a \sin \theta \cos \phi, a \sin \theta \sin \phi, a \cos \theta \rangle$$

with

$$dS = |\vec{N}(\theta, \phi)| d\theta d\phi = a^2 \sin^2 \theta d\theta d\phi.$$

The integral over the whole sphere gives the area of the sphere

$$\int_0^{2\pi} \int_0^\pi a^2 \sin^2 \theta d\theta d\phi = 4\pi a^2.$$

We will study integration in Chapter 11, so no need to worry about this aspect of surfaces yet. The key thing is that the parametrization of a surface contains the information about the geometry of the surface through its grid.

From surface area integrals to the integral of the flux of a vector field through a surface is a small step. We need to pick a direction to measure this flux in, so we have to pick a unit normal on one side or the other of the surface, which if it can be done consistently makes the surface an “orientable surface.” The most well known counterexample is a Mobius strip, but we are only interested in local considerations so we will not worry about this complication. Take the radially outward pointing vector field $\vec{F} = (x^2 + y^2 + z^2)^{-3/2} \langle x, y, z \rangle = (x^2 + y^2 + z^2)^{-1} \hat{r}$ in \mathbb{R}^3 whose magnitude is $|\vec{F}| = 1/r^2$, where $r = (x^2 + y^2 + z^2)^{1/2}$ is the distance from the origin. The flux of this vector field outward through a sphere of radius a about the origin is defined to be the product of the area $4\pi a^2$ of the surface times the constant outward normal component of the vector field $1/a$ on the sphere: $Flux = 4\pi a^2 / a = 4\pi$, which also turns out to be a constant. For a vector field whose normal component is not constant we simply integrate that component with respect to the differential of surface area to define the flux, which is called the surface integral of the vector field over the oriented surface.

Given a parametrized surface with an ordering of the parameters (u^1, u^2) , the normal defined above orients the surface, so we define the surface integral of the vector field over the parametrized surface by

$$\begin{aligned}\int_{\Sigma} \vec{F} \cdot d\vec{S} &= \int_{\Sigma} \vec{F}(\vec{r}(u^1, u^2)) \cdot \underbrace{\hat{N}(u^1, u^2)}_{\hat{N}(u^1, u^2) \|\vec{N}(u^1, u^2)\|} dS \\ &= \int_{\Sigma} \vec{F}(\vec{r}(u^1, u^2)) \cdot \vec{N}(u^1, u^2) du^1 du^2,\end{aligned}$$

where the length of the unnormalized normal $\vec{N}(u^1, u^2)$ naturally combines with its unit direction to produce the dot product with the unnormalized normal that is generated by the parametrization—easy! The same combination of length and direction occurs if we define the integral of a vector field along a parametrized curve C (oriented by the increasing parameter t) as the integral of the tangential component with respect to the differential of arclength $ds/dt = \|\vec{r}'(t)\|$

$$\begin{aligned}\int_C \vec{F} \cdot d\vec{s} &= \int_C \vec{F}(\vec{r}(t)) \cdot \underbrace{\hat{T}(t)}_{\hat{T}(t)\|\vec{r}'(t)\|dt = \vec{r}'(t)} ds \\ &= \int_C \vec{F}(\vec{r}(t)) \cdot \vec{r}'(t) dt,\end{aligned}$$

since $\hat{T}(t) = \vec{r}'(t)/\|\vec{r}'(t)\|$ is the unit direction of the tangent vector. In each case the actual definite integrals are taken over the ranges of the parameters involved.

Exercise D.0.16.

surface area, conics of revolution

a) Find the surface area $S = 17(\sqrt{17} - 1)\pi/6$ of the parabola of revolution $z = 4 - x^2 - y^2$ above the plane $z = 0$. It is helpful to parametrize this by polar coordinates in the x - y plane:

$$\vec{r}(u, v) = \langle u \cos v, u \sin v, 4 - u^2 \rangle,$$

which orients the surface by the upward normal.

b) Evaluate the surface integral of the vector field $\langle 0, 0, z \rangle$ over this surface oriented by the upward normal, showing that its value is 8π .

c) Find the surface area of the upper half of the hyperbola of revolution $x^2 + y^2 - z^2 = -1, z > 0$ below the plane $z = 2$ using the parametrization

$$\vec{r}(u, v) = \langle \cosh u \cos v, \cosh u \sin v, \sinh u \rangle.$$

d) Evaluate the surface integral of the vector field $\langle 0, 0, z \rangle$ over this surface oriented by the upward normal. ■

Remark.

Why is the normal component of a vector field with respect to a surface interesting? Why do we have the seasons at our latitude on the Earth? For exactly this reason. Think of the solar energy per unit area hitting the Earth's surface as a vector field on the surface of the Earth with length equal to that energy per unit area and direction away from the sun. The amount of energy hitting the surface is proportional to the amount of area being hit by the solar energy, but if we fix a small patch on the surface, then the area of the cross-section of the impinging solar energy vector field is proportional to the cosine of the angle between that

vector field and inward normal to the Earth's surface, zero if the surface is parallel to the rays, 1 if the surface is at right angles. The dot product of the vector field with the unit normal to the surface is thus the local measure of the surface density of energy deposited per unit time on Earth's surface. It is maximized in the summer, minimized in the winter.

▲

Exercise D.0.17.

surface area on unit pseudosphere

a) Consider the hyperboloid $z = \sqrt{1 + x^2 + y^2}$ in \mathbb{R}^3 and use polar coordinates in the plane to evaluate its surface area below the plane $z = 2$, or use the parametrization in the previous Exercise.

b) Now consider the same surface $t = \sqrt{1 + x^2 + y^2}$ in 3-dimensional Minkowski spacetime with the same parametrization and evaluate its surface area below the plane $t = 2$, which was already done in the previous Exercise in its natural hyperbolic parametrization.

The radial differential of arclength is contracted as the hyperboloid tilts up towards the null radial direction which has zero length, thus decreasing the surface area of the Minkowski spacetime surface relative to the Euclidean surface.

■

While we are looking at parametrized surfaces, we cannot ignore the geometry of the grid that the parametrization imposes on the surface, in terms of the two 1-parameter family of curves we get on that surface by holding each of the parameters fixed in turn and choosing equally spaced intervals of that fixed parameter to generate those curves along which the other parameter varies. Computer software uses this grid to give perspective to the graphs of functions of two variables in space, and for the more general parametrized surfaces.

The first question we can ask about this grid is whether or not the gridlines are orthogonal to each other when they meet, i.e., as determined by the angles between their tangent vectors $\vec{r}_1(u^1, u^2)$ and $\vec{r}_2(u^1, u^2)$. We can simply define their matrix of inner products

$$G_{ij}(u^1, u^2) = \vec{r}_i(u^1, u^2) \cdot \vec{r}_j(u^1, u^2),$$

Orthogonality of the grid requires that this matrix be diagonal everywhere: $G_{12}(u^1, u^2) = \vec{r}_1(u^1, u^2) \cdot \vec{r}_2(u^1, u^2) = 0$. Orthogonal grids are very useful on a surface since it helps us view its geometry in a way that is as close as possible to our rectangular grids in Cartesian coordinates in the plane.

This matrix of inner products is useful to re-express the differential of arclength for any curve in the surface

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 = d\vec{x} \cdot d\vec{x} \\ &= \frac{\partial \vec{x}}{\partial u^i} du^i \cdot \frac{\partial \vec{x}}{\partial u^j} du^j = G_{ij} du^i du^j \\ &= G_{11}(du^1)^2 + 2G_{12} du^1 du^2 + G_{22}(du^2)^2. \end{aligned}$$

For an orthogonal grid this the middle term is zero.

Exercise D.0.18.**arclength on a parametrized surface**

- a) Evaluate this differential of arclength for the sphere already discussed above in the text.
- b) Repeat for the unit hyperboloid using the hyperbolic function parametrization of Exercise [D.0.16](#) using both the Euclidean inner product and the Minkowski inner product and compare the radial arclength from the symmetry axis at $z = 1$ to the circle at $z = 2$. Which one is smaller?



Appendix E

Visualizing vector space duality in the vector space R^3

All of tensor analysis is based on “multilinear” functions of a certain number of vector arguments (simultaneously linear in each vector argument separately), just like the familiar dot product for ordinary 3-vectors. It is therefore worthwhile understanding well a simple real-valued linear function of a single vector argument and its simple geometry that enables it to be visualized and distinguished from the vector with which it is usually identified in calculus. This is the foundation from which contravariant and covariant tensors spring forth.

Let us identify a vector in R^3 with a column matrix when it appears in formulas involving matrices, as the computer algebra system Maple does

$$\vec{X} = \langle X^1, X^2, X^3 \rangle \leftrightarrow \begin{pmatrix} X^1 \\ X^2 \\ X^3 \end{pmatrix}. \quad (\text{E.1})$$

The corresponding transposed row matrix will then be denoted by

$$\left(\vec{X}\right)^T \leftrightarrow (X^1 \quad X^2 \quad X^3). \quad (\text{E.2})$$

The dot product is a bilinear function “dot” of a pair of vectors in R^3 ,

$$\vec{X} \cdot \vec{Y} = \left(\vec{X}\right)^T \vec{Y} = \sum_{i=1}^3 X^i Y^i \quad (\text{E.3})$$

which happens to be symmetric in the sense that the order of the two vector arguments does not matter:

$$\text{dot}(\vec{X}, \vec{Y}) = \vec{X} \cdot \vec{Y} = \vec{Y} \cdot \vec{X} = \text{dot}(\vec{Y}, \vec{X}).$$

If you double one of the vectors, you double the value of their dot product:

$$\text{dot}(2\vec{X}, \vec{Y}) = (2\vec{X}) \cdot \vec{Y} = 2(\vec{X} \cdot \vec{Y}) = 2\text{dot}(\vec{X}, \vec{Y}),$$

or if you dot the sum of two vectors into a third vector, the result is the same as the sum of the individual dot products

$$\text{dot}(\vec{X} + \vec{Y}, \vec{Z}) = (\vec{X} + \vec{Y}) \cdot \vec{Z} = \vec{X} \cdot \vec{Z} + \vec{Y} \cdot \vec{Z}.$$

These are the basic properties which define any linear function of a vector, and in this case they apply to each of the two input vectors separately, making this a multilinear function of two vector arguments, i.e., a bilinear function. The immediate consequence is that the value on a linear combination is the linear combination of the values

$$\text{dot}(a\vec{X} + b\vec{Y}, \vec{Z}) = a \text{dot}(\vec{X}, \vec{Z}) + b \text{dot}(\vec{Y}, \vec{Z}) \quad (\text{E.4})$$

We should start at the beginning, with a single real-valued linear function A of one vector on R^3

$$A(\vec{x}) = a_1x^1 + a_2x^2 + a_3x^3 = \vec{a} \cdot \vec{x}, \quad \vec{x} = \langle x^1, x^2, x^3 \rangle \in R^3.$$

The vector variable \vec{x} is the usual variable representing vectors in R^3 , while the vector of coefficients of the linear function is identified with a constant vector in R^3 so that its linear combination with the components of the vector variable is realized through the dot product.

Every function on R^3 determines a family of “level surfaces,” each of which consists of all points (vectors in the space) which share the same value of the function. For a linear function A these surfaces are planes

$$A(\vec{x}) = a_1x^1 + a_2x^2 + a_3x^3 = c.$$

One can imagine representing this function visually by those level planes associated with the integer values of the function to get a family of equally spaced parallel planes dividing up the space, any two of which can be used to reconstruct the entire family. In particular the values 0 and 1 give the plane through the origin (zero vector) and the next plane in this sampled family with a greater value, and these two planes can be used as a visualization of the linear function in the same way a vector is visualized by an arrow (directed line segment from the origin to the point corresponding to a vector). The latter is an extremely useful visualization of a vector since it permits a geometric interpretation of vector addition as a tip to tail path or as the main diagonal of the parallelogram formed by two vectors emanating from the origin. For the linear function one also needs to know of this basic pair of parallel planes, which one corresponds to 0 and 1 in order to know in which direction the function is increasing, so it is a directed pair of planes that we will take as our visualization of the linear function

$$A(\vec{x}) = a_1x^1 + a_2x^2 + a_3x^3 = 0, 1.$$

This plane pair helps us visualize that separation as well as the orientation of the family in space.

Notice that scaling a linear function by a factor of 2 decreases the separation of the basic pair of planes representing that function compared to the original function by a factor of 2

$$2A(\vec{x}) = 2a_1x^1 + 2a_2x^2 + 2a_3x^3 = 1 \rightarrow A(\vec{x}) = a_1x^1 + a_2x^2 + a_3x^3 = 1/2.$$

This is exactly right since level surfaces of a function are closer together when it is increasing more rapidly. Thus the idea of a larger vector being associated with a longer arrow, while a larger linear function is associated with a smaller spacing of its basic plane pair go in opposite directions in their corresponding geometrical representations, but if we adjust our point of view, the bigger linear form has more representative planes per interval oblique to those parallel planes, which goes in the right direction.

Of course we also have another geometrical representation of the linear function by interpreting its ordered triplet of coefficients (a_1, a_2, a_3) as a vector

$$A \mapsto \vec{a} = \langle a_1, a_2, a_3 \rangle,$$

so that

$$A(\vec{x}) = \vec{a} \cdot \vec{x}, \quad (\text{E.5})$$

but we do this without realizing that it depends on the dot product that we take for granted on R^3 , and untangling this association is important for understanding metric geometry. This vector visualization behaves in the usual way, a larger linear function A leads to a larger vector \vec{a} of coefficients, so why to we need this complementary realization of a linear function as a pair of planes? The short answer is we don't and some people prefer to avoid it if they can, but doing so limits our understanding of the geometry involved and removes the real distinction which separates the roles played by vectors and linear functions of vectors. The dot product so far only represents the natural pairing between coefficients and variables of a linear function, but we then go on to use it to introduce length and angle geometry on R^3 , and reinterpret linear relations in terms of this metric geometry. The problem is that to get serious about metric geometry, we have to give up the dot product for a more general structure, so by keeping the distinction, it is much easier to handle that structure and its consequences.

For the moment let's use the dot product geometry, introducing the length of the coefficient covector (called a "normal vector" to the level planes of the linear function)

$$|\vec{a}| = ((a_1)^2 + (a_2)^2 + (a_3)^2)^{1/2}$$

and associated unit vector

$$\hat{a} = \frac{\vec{a}}{|\vec{a}|}.$$

Then the identity

$$\vec{a} \cdot \left(\frac{\vec{a}}{|\vec{a}|^2} \right) = 1$$

shows that it is the vector $\hat{a}/|\vec{a}| = \vec{a}/|\vec{a}|^2$ whose tip lies in the representation plane $A(\vec{x}) = 1$. This is an "orthogonal connecting vector" whose initial point at the origin connects up the plane $A(\vec{x}) = 0$ to the plane $A(\vec{x}) = 1$, and is at right angles to that plane in the dot product geometry in which angles are defined by the arccosine of the dot product of unit vectors (i.e., $\hat{a} \cdot \hat{b} = \cos \theta$). Since the length of the connecting vector is the reciprocal of the length of the original coefficient vector, increasing the linear function by a scale factor decreases the length of the separation vector by that same amount.

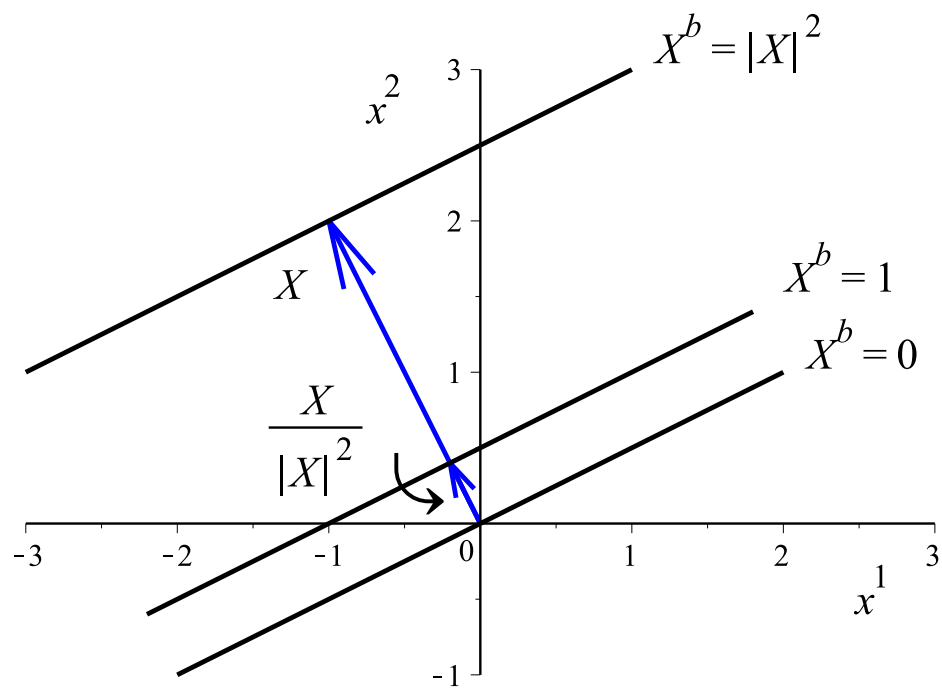


Figure E.1: Visualizing the relation between a vector X and its corresponding 1-form X^b with the same components in the standard basis of R^2 .

We are now in a position to extend the duality between the vectors and linear functions realized visually by vectors and planes to the idea of bases of the vector space R^3 and corresponding “dual bases” of the space of linear functions on R^3 : let’s designate this latter space by $(R^3)^*$.

Let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ be any ordered set of three linearly independent vectors in R^3 : a so called “basis” of the vector space, like the familiar natural basis $\{\langle 1, 0, 0 \rangle, \langle 0, 1, 0 \rangle, \langle 0, 0, 1 \rangle\}$. Let

$$B = (B^i_j) = (e^i_j) = \langle \vec{e}_1 | \vec{e}_2 | \vec{e}_3 \rangle \quad (\text{E.6})$$

be the matrix whose columns are the components of these vectors in the natural basis of R^3 . These three vectors visualized as arrows emanating from the vertex at the origin can be extended to form a parallelopiped with 8 vertices, 12 edges and 6 parallelogram faces. For example the tip of $\vec{e}_1 + \vec{e}_2 + \vec{e}_3$ is at the end of the “main diagonal” of this parallelopiped. The vertices and edges of the parallelopiped define its “skeleton” vector representation, a visual representation by vectors, the original 3 and their various tip to tail translations that make up the remaining edges. Filling in this skeleton with the 6 faces makes up the surface of the parallelopiped, while the solid parallelopiped corresponds to all possible linear combinations of the basic 3 vectors with coefficients whose values are confined to the closed interval from 0 to 1. The 3 pairs of parallel face planes define 3 corresponding linear functions whose coefficient vectors are related to the pair as normal vectors as described above. Let ω^3 be the linear function associated with the faces spanned by the first two vectors \vec{e}_1, \vec{e}_2 , and similarly ω^2 associated with \vec{e}_2, \vec{e}_3 and finally ω^1 . By definition $\omega^3(\vec{e}_1) = 0 = \omega^3(\vec{e}_2)$ since these two vectors lie in the plane $\omega^3(\vec{x}) = 0$, while $\omega^3(\vec{e}_3) = 1$ since the tip of \vec{e}_3 lies in the plane $\omega^3(\vec{x}) = 1$. Similar relations hold for the remaining linear functions, which together as an ordered set are called the dual basis, and they form a basis of the dual space of linear functions, or “1-forms.” These various evaluation relations altogether make the following array.

$$\omega^1(\vec{e}_1) = 1, \quad \omega^1(\vec{e}_2) = 0, \quad \omega^1(\vec{e}_3) = 0, \quad (\text{E.7})$$

$$\omega^2(\vec{e}_1) = 0, \quad \omega^2(\vec{e}_2) = 1, \quad \omega^2(\vec{e}_3) = 0,$$

$$\omega^3(\vec{e}_1) = 0, \quad \omega^3(\vec{e}_2) = 0, \quad \omega^3(\vec{e}_3) = 1.$$

$$(\text{E.8})$$

In terms of their components these so called duality relations can be written

$$\omega^i_m e^m_j = \delta^i_j \quad \leftrightarrow \quad (\omega^i_j) \underline{B} = \underline{I}, \quad (\text{E.9})$$

where $I = (\delta^i_j)$ is the identity matrix. Comparison of this matrix product relation with the definition of the inverse matrix $B^{-1}B = I$ shows that the matrix of components of the dual basis with respect to the natural basis of R^3 is just the inverse matrix

$$(\omega^i_j) = B^{-1} = \begin{pmatrix} (\omega^1_i) \\ (\omega^2_i) \\ (\omega^3_i) \end{pmatrix} = \begin{pmatrix} (\vec{w}_1)^T \\ (\vec{w}_2)^T \\ (\vec{w}_3)^T \end{pmatrix}. \quad (\text{E.10})$$

The rows of this inverse matrix correspond to the components of the dual basis 1-forms. However, they can also be interpreted directly as the components of vectors with respect to the

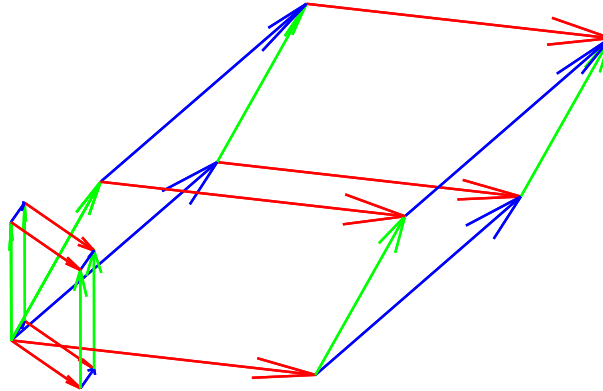


Figure E.2: Visualizing the dual basis of 1-forms and the reciprocal basis vectors. The three pairs of parallel faces of the (larger) parallelepiped whose edges originating at the origin coincide with the basis vectors lie in the planes $\omega^i(X) = 0, 1$ characterizing each of the dual 1-forms. The reciprocal basis vectors are respectively orthogonal to these 3 pairs of planes and form a corresponding (smaller) parallelepiped in this example. Scaling down the original set of basis vectors by an overall constant scales up the reciprocal basis by that same constant.

natural basis, which is a nontrivial identification that allows some presentations of tensor algebra to avoid the subject of duality, and the abstraction of linear functions. The ordered set $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ is called the reciprocal basis.

The duality relations state that the dot products of \vec{w}_1 with \vec{e}_2 and \vec{e}_3 vanish, so it is orthogonal to their plane in the dot product geometry, and so on, while the duality relation $\vec{w}_1 \cdot \vec{e}_1 = 1$ then states that these two vectors have lengths which vary inversely: if \vec{e}_1 is stretched, \vec{w}_1 is compressed. In fact this reciprocal relationship of their lengths is clear from the more explicit relationship $\vec{w}_1 \cdot \vec{e}_1 = |\vec{w}_1||\vec{e}_1| \cos \theta = 1$, where θ is the fixed angle between the two vectors.

It is obvious that

$$\vec{w}_1 = \frac{\vec{e}_2 \times \vec{e}_3}{\vec{e}_1 \cdot (\vec{e}_2 \times \vec{e}_3)} \quad (\text{E.11})$$

easily solves the 3 duality conditions on \vec{w}_1 , with cyclic permutations giving the other two reciprocal vectors. Thus the three reciprocal vectors are orthogonal to the three faces of the parallelepiped formed by the original basis $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$, on the same side of these faces as the corresponding basis vectors but with appropriately altered lengths. Of course this vector formula for the reciprocal basis vectors is just an equivalent expression for the rows of the inverse matrix B^{-1} , which is the transpose of the matrix of minors of B divided by its determinant $\det B = \vec{e}_1 \cdot (\vec{e}_2 \times \vec{e}_3)$.

Okay, what is all this good for? One introduces new bases in order to adapt the coordinates

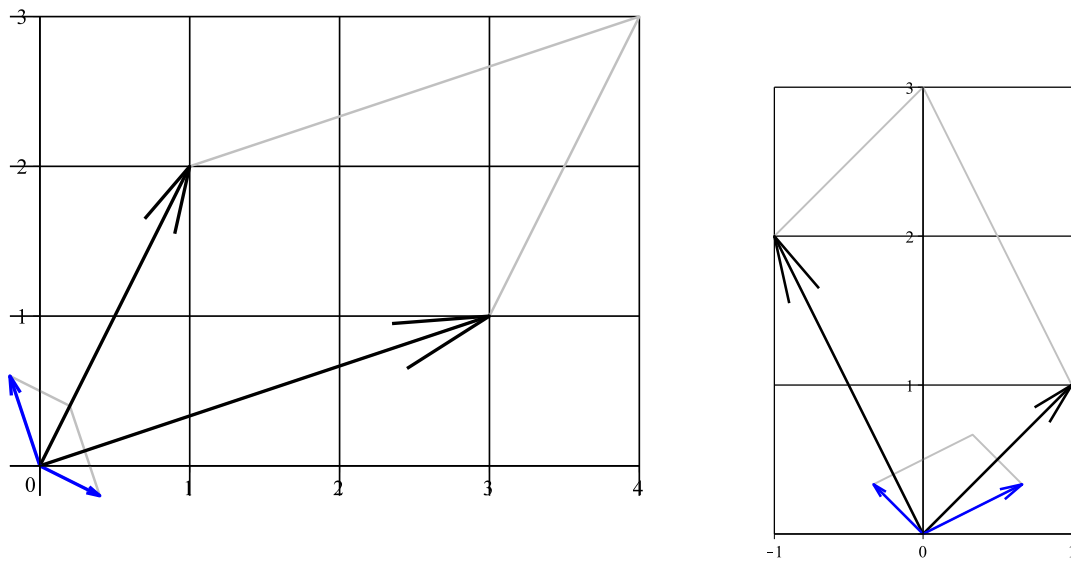


Figure E.3: Visualizing the dual basis of 1-forms and the reciprocal basis vectors in 2 dimensions is a bit easier than in 3 dimensions. Here we have the bases $\{\langle 3, 1 \rangle, \langle 1, 2 \rangle\}$ (left) and $\{\langle 1, 1 \rangle, \langle -1, 2 \rangle\}$ (right) shown in black. The reciprocal bases are much smaller, shown in blue. The corresponding unit parallelograms consisting of the two pairs of sides represent the dual bases (gray).

to directions that are special for the linear problem under study.

$$\vec{x} = \langle x^i \rangle = y^j \vec{e}_j \quad \text{or} \quad x^i = e^i_j y^j = B^i_j y^j \leftrightarrow \vec{x} = B \vec{y}, \quad \vec{y} = B^{-1} \vec{x}, \quad (\text{E.12})$$

where in the matrix equations, the symbols \vec{x} and \vec{y} are the corresponding column matrices. But the new coordinate is just the value of the corresponding dual vector on \vec{x}

$$y^i = B^{-1i}_j x^j = \omega^i_j x^j = \omega^i(\vec{x}) = \vec{w}_i \cdot \vec{x}. \quad (\text{E.13})$$

The reciprocal basis allows one to avoid talking about linear functions and the dual space, while providing a new basis of the space adapted to the normals to the faces of the parallelepiped formed by the original basis vectors rather than its edges. Here we have used subscripted reciprocal basis vectors as in the original basis simply to label the ordering of the vectors, but one must use superscripted variables like the dual 1-forms to play tensor algebra with them. However, this obscures the simpler mathematics of linearity by the more complicated geometry of an inner product space. There is much of modern differential geometry that is not related to inner products or metrics so if one wants to understand the bigger picture, one should clearly distinguish this structure.

Part III

Supplementary materials

Solutions to Exercises

In Progress: A lot of work still to do here...

Chapter 0

0.0.1: arclength in the plane

a) Filling in the blanks

$$\begin{aligned}
 ds^2 &= (\cos \theta dr - r \sin \theta d\theta)^2 + (\sin \theta dr + r \cos \theta d\theta)^2 \\
 &= (\cos^2 \theta dr^2 - 2r \cos \theta \sin \theta dr d\theta + r^2 \sin^2 \theta d\theta^2) \\
 &\quad + (\sin^2 \theta dr^2 + 2r \cos \theta \sin \theta dr d\theta + r^2 \cos^2 \theta d\theta^2) \\
 &= (\cos^2 \theta + \sin^2 \theta) (dr^2 + r^2 d\theta^2) \\
 &= dr^2 + r^2 d\theta^2.
 \end{aligned}$$

b) The Jacobian matrix is

$$\underline{J} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix},$$

so

$$\begin{aligned}
 \underline{J}^T \underline{J} &= \begin{pmatrix} \cos \theta & r \sin \theta \\ \sin \theta & -r \cos \theta \end{pmatrix} \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} = \begin{pmatrix} \cos^2 \theta + \sin^2 \theta & 0 \\ 0 & r^2(\sin^2 \theta + \cos^2 \theta) \end{pmatrix} \\
 &= \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}.
 \end{aligned}$$

c) Starting from $x = uv, y = \frac{1}{2}(u^2 - v^2)$ we get

$$dx = v du + u dv, \quad dy = u du - v dv$$

or

$$\underline{J} = \begin{pmatrix} v & u \\ u & -v \end{pmatrix} = \underline{J}^T,$$

so for example

$$\underline{J}^T \underline{J} = \begin{pmatrix} v & u \\ u & -v \end{pmatrix} \begin{pmatrix} v & u \\ u & -v \end{pmatrix} = \begin{pmatrix} u^2 + v^2 & 0 \\ 0 & u^2 + v^2 \end{pmatrix} = (u^2 + v^2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which means

$$ds^2 = (u^2 + v^2)(du^2 + dv^2).$$

This is an example of what are called isotropic coordinates since the line element is a multiple of the flat Euclidean line element $du^2 + dv^2$.

0.0.2: matrix multiplication and the trace

First, cyclicly permuting the scalar component factors easily does it

$$\text{Tr}(\underline{A}^{-1} \underline{B} \underline{A}) = A^{-1i}{}_j B^j{}_k A^k{}_i = A^k{}_i A^{-1i}{}_j B^j{}_k \delta^k{}_i = B^i{}_i = \text{Tr} \underline{B}.$$

Second, doing the same as before

$$\text{Tr}(\underline{A} \underline{B} \underline{C}) = A^i{}_j B^j{}_k C^k{}_i = C^k{}_i A^i{}_j B^j{}_k = \text{Tr}(\underline{C} \underline{A} \underline{B}).$$

Chapter 1

?: 2×2 matrices as a vector space

See Maple worksheet: gl2R-traceinnerproduct.mw.

1.2.2: 2×2 complex matrices as a real vector space

See Maple worksheet: gl2R-traceinnerproduct.mw.

1.2.3: up to quadratic functions

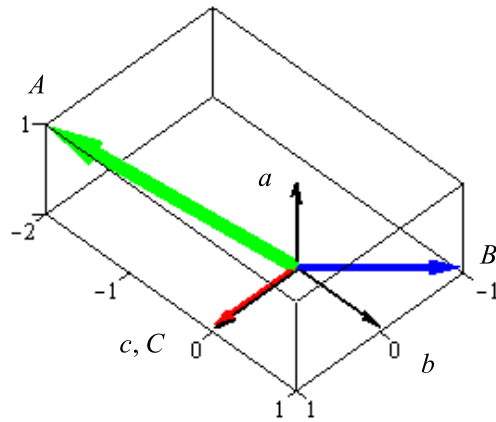


Figure E.4: Visualizing two bases for the space of at most quadratic polynomials: $ax^2 + bx + c = A(x - 1)^2 + B(x - 1) + C$. The letters in the figure indicate the old and new coordinate bases.

We just expand the expression and compare with the previous one

$$\begin{aligned} A(x - 1)^2 + B(x - 1) + C(1) &= (Ax^2 - 2Ax + A) + (Bx - B) + (C) \\ &= Ax^2 + (B - 2A)x + (A - B + C) = ax^2 + bx + c \end{aligned} \quad (\text{E.14})$$

leading to the identification

$$c = C - B + A, \quad b = B - 2A, \quad a = A. \quad (\text{E.15})$$

This is easily inverted, which leads to the Taylor coefficients in the expansion about $x = 1$

$$C = c + b + a, \quad B = b + 2a, \quad A = a. \quad (\text{E.16})$$

This can be put into matrix form

$$\begin{pmatrix} c \\ b \\ a \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} C \\ B \\ A \end{pmatrix}, \quad \begin{pmatrix} C \\ B \\ A \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c \\ b \\ a \end{pmatrix}. \quad (\text{E.17})$$

The columns of the first coefficient matrix are the old components of the new basis vectors, both sets of which are shown in Fig. E.4.

See also the Maple worksheet: quadratics.mw.

1.2.4: 3×3 antisymmetric matrices and the cross product

a) Matrix multiplication immediately gives the cross product formula

$$\underline{A}\underline{b} = \begin{pmatrix} 0 & -a^3 & a^2 \\ a^3 & 0 & -a^1 \\ -a^2 & a^1 & 0 \end{pmatrix} \begin{pmatrix} b^1 \\ b^2 \\ b^3 \end{pmatrix} = \begin{pmatrix} -a^3b^2 + a^2b^3 \\ a^3b^1 - a^1b^3 \\ -a^2b^1 + a^1b^2 \end{pmatrix} = \underline{\vec{a}} \times \underline{\vec{b}}. \quad (\text{E.18})$$

b) Doing the matrix product and difference leads to

$$\begin{aligned} & \underline{A}\underline{B} - \underline{B}\underline{A} \\ &= \begin{pmatrix} 0 & -a^3 & a^2 \\ a^3 & 0 & -a^1 \\ -a^2 & a^1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -b^3 & b^2 \\ b^3 & 0 & -b^1 \\ -b^2 & b^1 & 0 \end{pmatrix} - \begin{pmatrix} 0 & -a^3 & a^2 \\ a^3 & 0 & -a^1 \\ -a^2 & a^1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -b^3 & b^2 \\ b^3 & 0 & -b^1 \\ -b^2 & b^1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -a^3b^3 - a^2b^2 & a^2b^1 & a^3b^1 \\ a^1b^2 & -a^3b^3 - a^1b^1 & a^3b^2 \\ a^1b^3 & a^2b^3 & -a^2b^2 - a^1b^1 \end{pmatrix} \\ &\quad - \begin{pmatrix} -a^3b^3 - a^2b^2 & a^2b^1 & a^3b^1 \\ a^1b^2 & -a^3b^3 - a^1b^1 & a^3b^2 \\ a^1b^3 & a^2b^3 & -a^2b^2 - a^1b^1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -(a^1b^2 - a^2b^1) & a^3b^1 - a^1b^3 \\ a^1b^2 - a^2b^1 & 0 & -(a^2b^3 - a^3b^2) \\ -(a^3b^1 - a^1b^3) & a^2b^3 - a^3b^2 & 0 \end{pmatrix} \end{aligned}$$

from which one can identify the three independent components of this antisymmetric matrix as those of $\underline{\vec{a}} \times \underline{\vec{b}}$. Then using associativity of matrix multiplication and the previous correspondence

$$\underline{a} \times (\underline{b} \times \underline{u}) - \underline{b} \times (\underline{a} \times \underline{u}) = \underline{A}(\underline{B}\underline{u}) - \underline{B}(\underline{A}\underline{u}) = (\underline{A}\underline{B} - \underline{B}\underline{A})\underline{u} = (\underline{a} \times \underline{b}) \times \underline{u}.$$

See the Maple worksheet: asymmatrices-crossprod.mw.

1.2.5: complex numbers as 2-dimensional real vector space

See the Maple worksheet c2algebra.mw.

1.3.1: dual space closure

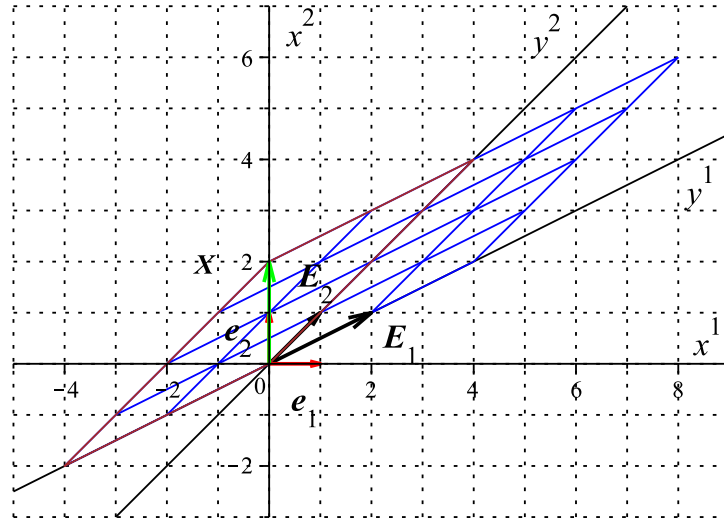


Figure E.5: Visualizing the new components of a vector using the basis and dual basis. The vector $X = \langle 0, 2 \rangle$ is the vector sum of $-2E_1$ and $4E_2$, namely minus 2 tickmarks on the y^1 axis and plus 4 tickmarks on the y^2 axis.

One has the following sequence of equalities starting and ending with the definition of a linear combination of linear functions as a new linear function

$$\begin{aligned}
 (c_1 f + c_2 g)(au + bv) &= c_1 f(au + bv) + c_2 g(au + bv) && \text{(definition)} \\
 &= c_1 (af(u) + bf(v)) + c_2 (ag(u) + bg(v)) && \text{(linearity of } f, g) \\
 &= a(c_1 f(u) + c_2 g(u)) + b(c_1 f(v) + c_2 g(v)) && \text{(recombination)} \\
 &= a(c_1 f + c_2 g)(u) + b(c_1 f + c_2 g)(v), && \text{(definition in reverse)}
 \end{aligned}$$

showing that a linear combination of linear functions is itself a linear function.

1.3.2: change of basis in the plane

We have the system of 4 equations to solve, amounting to two decoupled systems of 2 equations for 2 unknowns

$$\begin{aligned}
 W^1(E_1) &= a\omega^1(E_1) + b\omega^2(E_1) = 2a + b = \delta^1_1 = 1 \\
 W^1(E_2) &= a\omega^1(E_2) + b\omega^2(E_2) = a + b = \delta^1_2 = 0 \\
 W^2(E_1) &= c\omega^1(E_1) + d\omega^2(E_1) = 2c + d = \delta^2_1 = 0 \\
 W^2(E_2) &= c\omega^1(E_2) + d\omega^2(E_2) = c + d = \delta^2_2 = 1.
 \end{aligned}$$

Solving these two simple systems gives $(a, b) = (1, -1), (c, d) = (-1, 2)$. Let these define vectors $\vec{W}^1 \equiv \langle 1, -1 \rangle$ and $\vec{W}^2 \equiv \langle -1, 2 \rangle$ from the components of these two linear functions. Then $W^1(u) = \vec{W}^1 \cdot u$ so if $W^1(u) = 0$, then u is orthogonal to the vector \vec{W}^1 of components of W^1 with respect to $\{e_i\}$. Thus \vec{W}^1 is orthogonal to E_2 and \vec{W}^2 is orthogonal to E_1 while $W^2(\langle 5, -2 \rangle) = \langle -1, 2 \rangle \cdot \langle 5, -2 \rangle = -5 - 4 = -9$ or $= -\omega^1(\langle 5, -2 \rangle) + 2\omega^2(\langle 5, -2 \rangle) = -5 + 2(-2) = -9$. We could have written W^i (i -th covector) $= W^i_j \omega^j$ (j -th component with $\{e_i\}$ of W^i), leading to a matrix $(W^i_j) = (W^i(e_j))$ which “changes the basis.” More later.

Fig. E.5 is a graphical representation of the integer level surfaces of W^1 and W^2 and an example of decomposing a vector into components with respect to $\{E_i\}$ using the dual basis.

Note: $W^1 = x - y$, $W^2 = -x + 2y$ in Cartesian coordinates $\{x, y\}$ on \mathbb{R}^2 .

1.4.1: rotations of the plane, pseudorotations of the Lorentz plane

See the maple worksheet `matrixexponential2by2.mw`. Apart from multiplying 2×2 matrices, easily done by hand, one only needs the two trig addition formulas:

$$\begin{aligned}\cos(\theta_1 + \theta_2) &= \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2, \\ \sin(\theta_1 + \theta_2) &= \sin \theta_1 \cos \theta_2 + \cos \theta_1 \sin \theta_2,\end{aligned}$$

and their hyperbolic analogs

$$\begin{aligned}\cosh(\theta_1 + \theta_2) &= \cosh \theta_1 \cosh \theta_2 + \sinh \theta_1 \sinh \theta_2, \\ \sinh(\theta_1 + \theta_2) &= \sinh \theta_1 \cosh \theta_2 + \cosh \theta_1 \sinh \theta_2.\end{aligned}$$

1.4.2: determinants and the cross product

This is almost obvious, once you look at concrete values of the indices.

1.4.3: quadruple scalar product

This is almost obvious, once you look at concrete values of the indices.

1.4.4: transforming a tensor on \mathbb{R}^2

1.4.4: 2 index tensor in a frame

If we let $A^i_j = \mathbb{A}(W^i, E_j)$ be the components of \mathbb{A} with respect to $\{E_i\}$, then

$$\mathbb{A} = A^i_j E_i \otimes W^j = E_1 \otimes W^1 + 2E_1 \otimes W^2 - E_2 \otimes W^1.$$

But both $\{E_i\}$ and $\{W^j\}$ are linear combinations of the standard basis and dual basis, so we can just substitute and expand

$$\begin{aligned}\mathbb{A} &= (2e_1 + e_2) \otimes (\omega^1 - \omega^2) - 2(2e_1 + e_2) \otimes (-\omega^1 - 2\omega^2) - (e_1 + e_2) \otimes (\omega^1 - \omega^2) \\ &= (2e_1 + e_2) \otimes ((\omega^1 - \omega^2) + 2(-\omega^1 - 2\omega^2)) - (e_1 + e_2) \otimes (\omega^1 - \omega^2) \\ &= -2e_1 \otimes \omega^1 - e_2 \otimes \omega^1 + 6e_1 \otimes \omega^2 + 3e_2 \otimes \omega^2 - e_1 \otimes \omega^1 - e_2 \otimes \omega^1 + e_1 \otimes \omega^2 + e_2 \otimes \omega^2 \\ &= -3e_1 \otimes \omega^1 + 7e_1 \otimes \omega^2 - 2e_2 \otimes \omega^1 + 4e_2 \otimes \omega^2 = A(\omega^i, e_j) e_i \otimes \omega^j,\end{aligned}$$

1.5.1: eigenvectors of a matrix of eigenvectors

See the Maple worksheet [Remove this problem??]

1.5.2: change of coordinates in the plane

See the Maple worksheet gridsinplane.mw.

1.5.3: change of coordinates in \mathbb{R}^3

See the Maple worksheet eigenvectors_uppertriangular.mw.

1.6.1: Euclidean inner product on $h(2)$

See the Maple worksheet su2matrices.mw.

1.6.2: two inner products on $gl(2, R)$

See the Maple worksheet sl2matrices.mw and gl2R-traceinnerproduct.mw.

1.6.3: pseudo-orthogonality in the Lorentz plane

a) Since \underline{A} and \underline{A}^{-1} are symmetric matrices, the condition is $\underline{A}^{-1} \underline{G} \underline{A}^{-1} = \underline{G}$, easily verified by explicit multiplication.

b) Note $\langle \pm 1, \pm 1 \rangle \cdot \langle \pm 1, \pm 1 \rangle = -(\pm 1)^2 + (\pm 1)^2 = -1 + 1 = 0$.

1.6.4: Euclidean and Lorentzian dot products

a) $v \cdot v > 0$ for all $v \neq 0$, so $\text{sgn}(v) = 1$.

b) Only the zero vector has zero length with the usual dot product.

c) $\langle 0, 1, -1, 1 \rangle \cdot \langle 0, 1, -1, 1 \rangle = -(0)^2 + 1^2 + (-1)^2 + 1^2 = 3 > 0$

$\langle 2, 1, 0, 0 \rangle \cdot \langle 2, 1, 0, 0 \rangle = -2^2 + 1^2 + 0^2 + 0^2 = -1 < 0$

$\langle 1, 0, 0, 1 \rangle \cdot \langle 1, 0, 0, 1 \rangle = -1^2 + 0^2 + 0^2 + 1^2 = 0$

The first can be normalized by dividing by $\sqrt{3}$. The second is already a unit vector.

d) It only must satisfy $-(v^0)^2 + (v^1)^2 + (v^2)^2 + (v^3)^2 = 0$, i.e., must lie on the light cone.

1.6.5: trace inner product for 3×3 antisymmetric matrices

One easily calculates the square which has diagonal entries $-(\omega^2) - (\omega^3)^2$, $-(\omega^3) - (\omega^1)^2$, $-(\omega^1) - (\omega^2)^2$, so the trace becomes minus twice the self-dot product of this vector.

See the Maple worksheet asymmatrices-crossprod.mw.

1.6.6: electromagnetic field matrices

See the Maple worksheet emfieldmatrix.mw.

1.6.7: visualizing a covector in the plane**1.6.8: transformation of dot products****1.6.9: inner products on spaces of square matrices and symmetry**

(i) If $\underline{A} = A^m_n \underline{e}^n_m$ then

$$\omega^i_j(\underline{A}) = \omega^i_j(A^m_n \underline{e}^j_m) = A^m_n \underbrace{\omega^i_j(\underline{e}^j_m)}_{\delta^i_m \delta^n_j} = A^i_j$$

(ii)

$$\underline{A} \underline{B} = (A^j_i \underline{e}^i_j)(B^n_m \underline{e}^m_n) = A^j_i B^n_m \underbrace{\underline{e}^i_j \underline{e}^m_n}_{\underline{e}^i_j \underline{e}^m_n = \delta^i_n \underline{e}^m_j} = A^j_n B^n_m \underline{e}^m_j$$

so

$$[\underline{A} \underline{B}]^j_m = A^j_n B^n_m.$$

(iii)–(iv) Using $\underline{A}^T = \underline{A}$ and $\underline{B}^T = -\underline{B}$ and the trace transpose and product identities, and the cyclic permutation trace identity

$$\mathcal{G}(\underline{A}, \underline{B}) = \text{Tr } \underline{A} \underline{B} = -\text{Tr } \underline{A}^T \underline{B}^T = -\text{Tr}(\underline{B} \underline{A})^T = -\text{Tr}(\underline{B} \underline{A}) = -\text{Tr}(\underline{A} \underline{B}),$$

so this must vanish, but

$$G(\underline{A}, \underline{B}) = \text{Tr } \underline{A}^T \underline{B} = \text{Tr } \underline{A} \underline{B} = \mathcal{G}(\underline{A}, \underline{B})$$

so the antisymmetric and symmetric matrices are orthogonal with respect to both metrics.

(v)

$$\mathcal{G}(\underline{e}^i_j, \underline{e}^m_n) = \text{Tr } \underline{e}^i_j \underline{e}^m_n = \text{Tr } \delta^i_n \underline{e}^m_j = \delta^i_n \underbrace{\text{Tr } \underline{e}^m_j}_{\delta^m_j \text{ (think why)}} = \delta^i_n \delta^m_j \neq \delta^{im} \delta_{jn},$$

while

$$G(\underline{e}^i_j, \underline{e}^m_n) = \text{Tr}(\underline{e}^i_j)^T \underline{e}^m_n = \text{Tr } \underline{e}^j_i \underline{e}^m_n = \delta^j_n \text{Tr } \underline{e}^m_i = \delta^j_n \delta^m_i = \delta^{im} \delta_{jn} = \begin{cases} 1 & \text{if } (i, j) = (m, n), \\ 0 & \text{otherwise,} \end{cases}$$

so the basis is orthonormal with respect to G but not \mathcal{G} .

(vi) Because $A - A^T = 0$ or $B + B^T = 0$ are linear conditions on the entries of the matrix, they are preserved under linear combinations and hence define linear subspaces. The dimension of the symmetric subspace equals the number of entries in an upper triangular matrix, which is, working up from the bottom corner by row, equal to $1 + 2 + \dots + n = n(n+1)/2$. This must be reduced by the n diagonal entries for the antisymmetric subspace, namely $n(n+1)/2 - n = n(n-1)/2$.

(vii) First we must verify the expansion of any matrix in terms of this basis. The diagonal contributions to the sum are clear, so consider only the offdiagonal contributions to the following sum

$$\begin{aligned} \check{A}^i_j \check{\underline{E}}^j_i + \check{A}^j_i \check{\underline{E}}^i_j &= \frac{1}{2}[(A^i_j + A^j_i)(\underline{e}^j_i + \underline{e}^i_j) + (A^i_j - A^j_i)(\underline{e}^j_i - \underline{e}^i_j)] \\ &= \frac{1}{2}[2A^i_j \underline{e}^j_i + 0A^j_i \underline{e}^j_i + 0A^i_j \underline{e}^i_j + 2A^j_i \underline{e}^i_j] = 2A^i_j \underline{e}^j_i, \end{aligned}$$

which is correct for those offdiagonal contributions if one instead sums only over $i < j$.

First we evaluate the upper offdiagonal inner products for the symmetric basis matrices for which $i < j$ and $m < n$

$$\begin{aligned}
 \mathcal{G}(\check{\underline{E}}_j^i, \check{\underline{E}}_n^m) &= \text{Tr } \check{\underline{E}}_j^i \check{\underline{E}}_n^m = \frac{1}{2} \text{Tr}(\underline{e}_j^i + \underline{e}_i^j)(\underline{e}_n^m + \underline{e}_m^n) \\
 &= \frac{1}{2}(\text{Tr } \underline{e}_j^i \underline{e}_n^m + \text{Tr } \underline{e}_i^j \underline{e}_n^m + \text{Tr } \underline{e}_j^i \underline{e}_m^n + \text{Tr } \underline{e}_i^j \underline{e}_m^n) \\
 &= \frac{1}{2}(\delta_n^i \text{Tr } \underline{e}_j^m + \delta_n^j \text{Tr } \underline{e}_i^m + \delta_m^i \text{Tr } \underline{e}_j^n + \delta_m^j \text{Tr } \underline{e}_i^n) = \delta^{im} \delta_{jn} + \delta_n^i \delta_j^m \\
 &= \delta^{im} \delta_{jn},
 \end{aligned}$$

since the last term $\delta_n^i \delta_j^m$ is nonzero only if $i = n$ and $m = j$, but from our inequalities, $i < j$ then implies $n < m$, which is a contradiction. For the diagonal case $(\underline{e}_i^i)^2 = \underline{e}_i^i$ which has unit trace so these are orthonormal among themselves, and the trace of a diagonal and offdiagonal matrix is zero since their product remains offdiagonal, remarks which apply also to the inner product G . For the antisymmetric upper offdiagonal inner products a similar calculation holds

$$\begin{aligned}
 \mathcal{G}(\check{\underline{E}}_j^i, \check{\underline{E}}_n^m) &= \text{Tr } \check{\underline{E}}_j^i \check{\underline{E}}_n^m = \frac{1}{2} \text{Tr}(\underline{e}_j^i - \underline{e}_i^j)(\underline{e}_n^m - \underline{e}_m^n) \\
 &= \frac{1}{2}(\text{Tr } \underline{e}_j^i \underline{e}_n^m - \text{Tr } \underline{e}_i^j \underline{e}_n^m - \text{Tr } \underline{e}_j^i \underline{e}_m^n + \text{Tr } \underline{e}_i^j \underline{e}_m^n) \\
 &= \frac{1}{2}(\delta_n^i \text{Tr } \underline{e}_j^m - \delta_n^j \text{Tr } \underline{e}_i^m - \delta_m^i \text{Tr } \underline{e}_j^n + \delta_m^j \text{Tr } \underline{e}_i^n) = -\delta^{im} \delta_{jn} + \delta_n^i \delta_j^m \\
 &= -\delta^{im} \delta_{jn}.
 \end{aligned}$$

Because of the transpose on one factor, the metric G leads to a sign reversal of these self inner products (it is positive definite). Finally the inner products with either metric of symmetric with antisymmetric matrices is zero since the trace of such a product is always zero. Thus this basis is orthonormal with respect to both inner products.

(viii)

$$f(\underline{A}) = f^i_j \omega^j_i(\underline{A}) = f^i_j A^j_i \equiv \text{Tr } \underline{F} \underline{A}$$

(ix)

$$\begin{aligned}
 \mathcal{G}^{i\ m}_j &= \mathcal{G}(\underline{e}_j^i, \underline{e}_n^m) = \text{Tr } \underline{e}_j^i \underline{e}_n^m = \text{Tr } \delta_n^i \underline{e}_j^m = \delta_n^i \delta_j^m \\
 G^{i\ m}_j &= G(\underline{e}_j^i, \underline{e}_n^m) = \text{Tr } \underline{e}_i^j \underline{e}_n^m = \text{Tr } \delta_n^j \underline{e}_i^m = \delta_n^j \delta_i^m = \delta^{im} \delta_{jn} \\
 \mathcal{G} &= \delta_n^i \delta_j^m \omega^j_i \otimes \omega^n_m = \omega^j_i \otimes \omega^i_j = \text{Tr } \underline{\omega} \otimes \underline{\omega} \\
 G &= \delta^{im} \delta_{jn} \omega^j_i \otimes \omega^n_m = \omega^j_i \otimes \omega^j_i = \text{Tr } \underline{\omega}^T \otimes \underline{\omega}
 \end{aligned}$$

(x) Note also that the trace is a real valued linear function on V , i.e., a covector

$$\text{Tr } \underline{e}_j^i = \delta^i_j \longrightarrow \text{Tr} = \delta^i_j \omega^j_i = \omega^i_i = \text{Tr } \underline{\omega}.$$

(xi) This is true only in the uninteresting case of $p = 1$ since it is not linear with more than 1 factor

$$\det((\underline{A} + \underline{B})\underline{C}) \neq \det \underline{A} \underline{C} + \det \underline{B} \underline{C}.$$

If you have trouble with indices in any of the above calculations, look at the $n = 2$ or $n = 3$ cases. For example, if $n = 2$ which is all you need to understand what happens in general, then

$$\{\underline{e}^1_1, \underline{e}^2_1, \underline{e}^1_2, \underline{e}^2_2\} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

Then

$$\text{Tr } \underline{e}^1_1 = \text{Tr } \underline{e}^2_2 = 1, \text{Tr } \underline{e}^2_1 = \text{Tr } \underline{e}^1_2 = 0$$

and

$$\check{\underline{E}}^1_2 = 2^{-1/2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \check{\underline{E}}^1_2 = 2^{-1/2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

so

$$\text{Tr}(\check{\underline{E}}^1_2)^2 = \frac{1}{2} \text{Tr} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{1}{2} \text{Tr} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1$$

and

$$\text{Tr}(\check{\underline{E}}^1_2)^2 = \frac{1}{2} \text{Tr} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \frac{1}{2} \text{Tr} \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -1,$$

etc.

?: Lorentz inner product on $gl(2, \mathbb{R})$

1.6.10: projections in \mathbb{R}^3

1.6.11: Gram-Schmidt orthonormalization

1.6.12: second derivative test

1.6.13: visualizing positive-definite inner products for the plane

1.7.1: 2×2 matrix exponentials

See the Maple worksheet `matrixexponential2by2.mw`.

1.7.2: differential of a family of matrices preserving an inner product

1.7.3: linear transformations plus translations: the inhomogeneous linear group

See the Maple worksheet `GLplusT.mw`.

1.7.4: $U(1)$, unit complex numbers

See the Maple worksheet `u1.mw`.

1.7.5: left and right translations and the adjoint action of a group

1.7.6: commutators of antisymmetric 3×3 matrices

1.7.7: commutator of small rotations

1.7.8: matrix Lie algebra commutators

a)

$$\begin{aligned}
 0 &= [\underline{E}_a, [\underline{E}_b, \underline{E}_c]] + [\underline{E}_b, [\underline{E}_c, \underline{E}_a]][\underline{E}_c, [\underline{E}_a, \underline{E}_b]] \\
 &= C^e_{bc}[\underline{E}_a, \underline{E}_e] + C^e_{ca}[\underline{E}_b, \underline{E}_e] + C^e_{ab}[\underline{E}_c, \underline{E}_e] \\
 &= C^e_{bc}C^d_{ae}\underline{E}_d + C^e_{ca}C^d_{be}\underline{E}_d + C^e_{ab}C^d_{ce}\underline{E}_d \\
 &= (C^e_{bc}C^d_{ae} + C^e_{ca}C^d_{be} + C^e_{ab}C^d_{ce})\underline{E}_d.
 \end{aligned}$$

The coefficient is therefore 0, which is the target identity.

b) We are aiming for

$$[\underline{k}_a, \underline{k}_b] = C^c_{ab} \underline{k}_c$$

starting from

$$C^d_{ea}C^e_{bc} + C^d_{eb}C^e_{ca} + C^d_{ec}C^e_{ab} = 0.$$

Notice that the right hand side of the former is $(C^c_{ab} \underline{k}_c)^d_c$, so we need to see the first two terms as the same components $(\dots)^d_c$, but also we need the a and b indices to be the labels of \underline{k}_a and \underline{k}_b , so we need to use the antisymmetry on the lower indices to change the order in the first two terms (one sign change in the first term, two in the second term for no net change in sign)

$$-C^d_{ae}C^e_{bc} + C^d_{be}C^e_{ac} + C^d_{ec}C^e_{ab} = 0$$

which are the $(\dots)^d_c$ components of

$$-\underline{k}_a \underline{k}_b + \underline{k}_b \underline{k}_a + C^e_{ab} \underline{k}_e = 0$$

which yields the desired result when the first two terms are moved to the opposite side of the equation (dummy index e becomes c).

c) Just express in components

$$\text{ad}(\underline{X})\underline{Y} = [X^a \underline{E}_a, Y^b \underline{E}_b] = X^a Y^b C^e_{ab} \underline{E}_e = (X^a C^e_{ab}) Y^b \underline{E}_e = (\text{ad}(\underline{X}))^e_b Y^b \underline{E}_e$$

so the matrix of this linear transformation is

$$Y^e \rightarrow (\text{ad}(\underline{X}))^e_b Y^b = (X^a C^e_{ab}) Y^b = (X^a \underline{k}_a)^e_b Y^b$$

so that

$$\underline{\text{ad}}(\underline{X}) = X^a \underline{k}_a.$$

Chapter 2

2.2.1: counting independent components

A symmetric matrix has the same number of independent components as an upper triangular matrix since the entries below the main diagonal are equal to those above it. There are n upper triangular entries in the first row, $n - 1$ in the second, etc., until there is only 1 upper diagonal entry in the last row, so the total is the sum of the first n integers or $n(n + 1)/2$. An antisymmetric matrix has n fewer independent entries since the diagonal entries are zero, so it has $n(n + 1)/2 - n = n(n - 1)/2$ independent components.

2.2.2: trace inner products and symmetry

Using a computer algebra system makes this painless. The two projections of the matrix are

$$\underline{S} = \text{SYM}(\underline{A}) = \begin{pmatrix} 1 & 3 & 5 \\ 3 & 5 & 7 \\ 5 & 7 & 9 \end{pmatrix}, \quad \underline{A} = \text{ALT}(\underline{A}) = \begin{pmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{pmatrix},$$

and their inner products are

$$\text{Tr}(\underline{S} \underline{S}) = 273, \quad \text{Tr}(\underline{A} \underline{A}) = -12, \quad \text{Tr}(\underline{S} \underline{A}) = 0,$$

or

$$\text{Tr}(\underline{S}^T \underline{S}) = 273, \quad \text{Tr}(\underline{A}^T \underline{A}) = 12, \quad \text{Tr}(\underline{S}^T \underline{A}) = 0.$$

Finally

$$\text{Tr}(\underline{A} \underline{A}) = 273 - 12 \quad \text{Tr}(\underline{A}^T \underline{A}) = 273 + 12.$$

2.2.3: counting transpositions

Each time you uncross one pair of strings you do a transposition, so it is just a matter of looking at a few examples to convince yourself how it works. Be my guest to attempt a proof. Look at this example to convince yourself how it works. Draw in the 3 connecting lines between like integers on the first and second rows.

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$$

Then $T(4, 1)$ uncrosses one pair and then $T(1, 2)$, $T(3, 4)$ uncross the other two, for a total of 3 transpositions. But that was too simple. Consider instead the case of 6 crossings (make sure only 2 connecting lines cross at one time)

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$$

Then $T(4, 1)$ uncrosses all but one pair, so that $T(2, 3)$ finishes the job, a total of 2 transpositions, an even number like 6.

2.3.1: quadruple scalar product

a) This has 3 different symmetries. It is antisymmetric in the first and second pairs individually, and symmetric under interchange of the two pairs

$$\begin{aligned} T(X, Y, Z, W) &= (X \cdot Z)(Y \cdot W) - (X \cdot W)(Y \cdot Z), \\ T(X, Y, W, Z) &= (X \cdot Z)(Y \cdot Z) - (X \cdot Z)(Y \cdot W) = -T(X, Y, Z, W), \\ T(Y, X, Z, W) &= (Y \cdot Z)(X \cdot W) - (Y \cdot W)(X \cdot Z) = -T(X, Y, Z, W), \\ T(Z, W, X, Y) &= (Z \cdot X)(W \cdot Y) - (Z \cdot Y)(W \cdot X) = T(X, Y, Z, W). \end{aligned}$$

b) This is a simple expansion and cancellation of the six terms in three pairs due to the commutivity of the dot product

$$\begin{aligned} &T(W, X, Y, Z) + T(W, Y, Z, X) + T(W, Z, X, Y) \\ &= (W \cdot Y)(X \cdot Z) - (W \cdot Z)(X \cdot Y) \\ &+ (W \cdot Z)(Y \cdot X) - (W \cdot X)(Y \cdot Z) \\ &+ (W \cdot X)(Z \cdot Y) - (W \cdot Y)(Z \cdot X) \\ &= 0. \end{aligned}$$

c) Indices cannot be repeated in the first pair or in the second pair without making the component zero.

$$\begin{aligned} T_{\alpha\beta\delta\gamma} &= T(e_\alpha, e_\beta, e_\gamma, e_\delta), \\ T_{1122} &= (e_1 \cdot e_2)(e_1 \cdot e_2) - (e_1 \cdot e_2)(e_1 \cdot e_2) = 0, \\ T_{1212} &= (e_1 \cdot e_1)(e_2 \cdot e_2) - (e_1 \cdot e_2)(e_2 \cdot e_1) = 1, \\ T_{2121} &= (e_2 \cdot e_2)(e_1 \cdot e_1) - (e_2 \cdot e_1)(e_1 \cdot e_2) = 1. \end{aligned}$$

d) Suppose we start from the simpler vector identities

$$\begin{aligned} \vec{A} \times (\vec{B} \times \vec{C}) &= (\vec{A} \cdot \vec{C})\vec{B} - (\vec{A} \cdot \vec{B})\vec{C}, & (\text{triple cross product}) \\ (\vec{A} \times \vec{B}) \cdot \vec{C} &= (\vec{B} \times \vec{C}) \cdot \vec{A}. & (\text{triple scalar product, cyclic symmetry}) \end{aligned}$$

Then a simple calculation shows

$$\begin{aligned} S(X, Y, Z, W) &= (\vec{X} \times \vec{Y}) \cdot (\vec{Z} \times \vec{W}) \\ &= [\vec{Y} \times (\vec{Z} \times \vec{W})] \cdot \vec{X} \\ &= [(\vec{Y} \cdot \vec{W})\vec{Z} - (\vec{Y} \cdot \vec{Z})\vec{W}] \cdot \vec{X} \\ &= (Y \cdot W)(Z \cdot X) - (Y \cdot Z)(W \cdot X) \\ &= T(X, Y, Z, W) \end{aligned}$$

This tensor is just the totally covariant form of the $\binom{2}{2}$ generalized Kronecker delta $\delta_{\gamma\delta}^{\alpha\beta}$, namely

$$T_{\alpha\beta\delta\gamma} = \delta_{\alpha\mu}\delta_{\beta\nu}\delta_{\gamma\delta}^{\mu\nu}.$$

This same tensor with the first argument left unevaluated (and index raised) is just the triple cross product

$$T^\alpha{}_{\beta\mu\nu}X^\beta Y^\gamma Z^\delta = \epsilon^\alpha{}_{\beta\gamma}X^\beta \epsilon^\gamma{}_{\mu\nu}Y^\mu Z^\nu .$$

There are two kinds of quadruple cross products which involve three cross products and produce a vector, defining a 5 index tensor constructed from three epsilons and some deltas for index raising. Identities involving the epsilons then lead to classical vector identities. Google some of these and analyze them as we have done above.

2.3.2: higher dimension contractions of the $p = 2$ generalized Kronecker delta

2.3.3: Jacobian matrix

2.3.4: quadruple scalar product again

2.3.5: differential of the determinant

2.3.6: inverse matrix differential

2.3.7: relative differential rotations and boosts

2.3.8: antisymmetry of the electromagnetic field tensor

2.4.1: linear independence of basis p -vectors

2.4.2: p -vectors in \mathbb{R}^4

For the case $n = 4$, write out explicitly the following sums

$$\begin{aligned} S &= S^{ij}e_{|ij|} = S^{14}e_{14} + S^{24}e_{24} + S^{34}e_{34} + S^{23}e_{23} + S^{13}e_{13} + S^{12}e_{12} , \\ T &= T^{ijk}e_{|ijk|} = T^{234}e_{234} + T^{134}e_{134} + T^{124}e_{124} + T^{123}e_{123} . \end{aligned}$$

2.5.1: multivariable Taylor series example

See Maple worksheet multitaylor.mw.

2.5.2: Quadratic function graph approximation to sphere, ellipsoid at a pole

See Maple worksheet multitaylor.mw.

2.5.3: moments of inertia of hemisphere

See Maple worksheet momentsofinertia.mw.

2.5.4: moment of inertia for snow cone

See Maple worksheet momentsofinertia.mw.

Chapter 3

3.1.1: covector addition

3.1.2: deWitt inner product for symmetric tensors

Chapter 4

4.2.1: successive antisymmetrization and the wedge product

4.2.2: wedges in \mathbb{R}^3

4.2.3: wedges in \mathbb{R}^4

4.2.4: wedges in \mathbb{R}^5

4.3.1: double natural dual sign

Replace S by $(*)T$ in the second line, matching up the $n-p$ dummy indices correctly, leading to the double Levi-Civita symbol, in which the indices must be reshuffled as explained in the problem. Then use the summation identity for the Kronecker delta. Finally the antisymmetric part is just the original tensor.

$$\begin{aligned}
 [^{(*)}T]_{i_{p+1}\dots i_n} &= \frac{1}{p!} T^{i_1\dots i_p} \epsilon_{i_1\dots i_p i_{p+1}\dots i_n} \\
 [^{(*)}S]^{i_1\dots i_p} &= \frac{1}{p!} S_{j_1\dots j_{n-p}} \epsilon^{j_1\dots j_{n-p} i_1\dots i_p} \\
 [^{(*)}({}^{(*)}T)^{i_1\dots i_p} &= \frac{1}{(n-p)!} ({}^{(*)}T_{i_{p+1}\dots i_n} \epsilon^{i_{p+1}\dots i_n i_1\dots i_p} \\
 &= \frac{1}{(n-p)!} \frac{1}{p!} T^{j_1\dots j_p} \epsilon_{j_1\dots j_p i_{p+1}\dots i_n} \epsilon^{i_{p+1}\dots i_n i_1\dots i_p} \\
 &= (-1)^{p(n-p)} \frac{1}{(n-p)!} \frac{1}{p!} T^{j_1\dots j_p} \epsilon_{i_1\dots i_{n-p} j_1\dots j_p} \epsilon^{i_1\dots i_{n-p} i_1\dots i_p} \\
 &= (-1)^{p(n-p)} \frac{1}{p!} T^{j_1\dots j_p} \delta_{j_1\dots j_p}^{i_1\dots i_p} = (-1)^{p(n-p)} T^{i_1\dots i_p},
 \end{aligned}$$

For $n = 3$, one has $(-1)^{p(3-p)} = 1$ for all $p = 0, 1, 2, 3$.

For $n = 4$, one has $(p, (-1)^{p(4-p)}) = (0, 1), (1, -1), (2, 1), (3, -1), (4, 1)$, i.e., the sign alternates with p .

4.3.2: natural dual index approach

Using these two identities

$$[e_{i_1\dots i_p}]^{j_1\dots j_p} = \delta_{i_1\dots i_p}^{j_1\dots j_p} = [\omega^{j_1\dots j_p}]_{i_1\dots i_p},$$

first the p -vector case

$$\begin{aligned}
 [^{(*)}T]_{i_{p+1}\dots i_n} &= \frac{1}{p!} T^{j_1\dots j_p} \epsilon_{j_1\dots j_p i_{p+1}\dots i_n} \\
 [^{(*)}e_{i_1\dots i_p}]_{i_{p+1}\dots i_n} &= \frac{1}{p!} \delta_{i_1\dots i_p}^{j_1\dots j_p} \epsilon_{j_1\dots j_p i_{p+1}\dots i_n} \\
 &= \epsilon_{i_1\dots i_p i_{p+1}\dots i_n},
 \end{aligned}$$

and then repeat for the p -covector case

$$\begin{aligned} [^{(*)}S]^{i_{p+1}\cdots i_n} &= \frac{1}{p!} S_{i_1\cdots i_p} \epsilon^{i_1\cdots i_p i_{p+1}\cdots i_n} \\ [^{(*)}\omega^{j_1\cdots j_p}]^{i_{p+1}\cdots i_n} &= \frac{1}{p!} \delta_{i_1\cdots i_p}^{j_1\cdots j_p} \epsilon^{i_1\cdots i_p i_{p+1}\cdots i_n} \\ &= \epsilon^{j_1\cdots j_p i_{p+1}\cdots i_n} . \end{aligned}$$

4.3.3: natural duals

4.3.4: dual of decomposable p -vector

4.3.5: self-inner products of p -vectors

4.3.6: quadruple scalar product and area

4.3.7: dual of the unit n -form

4.3.8: duals in \mathbb{R}^3 : self wedge with dual in \mathbb{R}^3

4.3.9: duals for M^3

4.3.10: cross product on \mathbb{R}^3 and M^3

4.3.11: double dual sign

The double dual $^{**}T$ for a p -covector is just T modulo a sign. To evaluate this sign, consider two convenient forms of the $(n-p)$ -covector dual a p -covector T

$$[^{*}T]_{i_{p+1}\cdots i_n} = \frac{1}{p!} T_{i_1\cdots i_p} \eta^{i_1\cdots i_p}_{i_{p+1}\cdots i_n} \quad \text{or} \quad [^{*}T]_{j_1\cdots j_{n-p}} = \frac{1}{p!} T_{i_1\cdots i_p} \eta^{i_1\cdots i_p}_{j_1\cdots j_{n-p}}$$

which when iterated give

$$\begin{aligned} [^{*}({}^{*}T)]_{j_{n-p+1}\cdots j_n} &= \frac{1}{(n-p)!} ({}^{*}T)_{j_1\cdots j_{n-p}} \eta^{j_1\cdots j_{n-p}}_{j_{n-p+1}\cdots j_n} \\ &= \frac{1}{(n-p)!} \frac{1}{p!} T_{i_1\cdots i_p} \underbrace{\eta^{i_1\cdots i_p}_{j_1\cdots j_{n-p}} \eta^{j_1\cdots j_{n-p}}_{j_{n-p+1}\cdots j_n}}_{\eta^{i_1\cdots i_p j_1\cdots j_{n-p}} \underbrace{\eta_{j_1\cdots j_{n-p} j_{n-p+1}\cdots j_n}}_{(-1)^{p(n-p)} \eta_{j_{n-p+1}\cdots j_n j_1\cdots j_{n-p}}}} , \end{aligned}$$

but then using the contraction identity

$$\eta^{i_1\cdots i_p j_1\cdots j_{n-p}} \eta_{j_{n-p+1}\cdots j_n j_1\cdots j_{n-p}} = (-1)^M (n-p)! \delta^{i_1\cdots i_p}_{j_{n-p+1}\cdots j_n} ,$$

we get

$$\begin{aligned} [*(T)]_{j_{n-p+1}\dots j_n} &= \frac{1}{(n-p)!} \left(\frac{1}{p!} T_{i_1\dots i_p} \delta_{j_{n-p+1}\dots j_n}^{i_1\dots i_p} \right) (-1)^M (n-p)! (-1)^{p(n-p)} \\ &= T_{j_{n-p+1}\dots j_n} (-1)^{M+p(n-p)}. \end{aligned}$$

Thus we have found

$$**T = (-1)^{M+p(n-p)} T.$$

4.3.12: inverse of dual

Recall that $*^{-1}R = (-1)^{M+P(n-P)}R$ for a P -form. The result of these operations is a $(p-1)$ -form

$$\underbrace{\underbrace{\underbrace{(S \wedge *T)}_{\substack{1 \quad p \\ n-p}}}_{P=n-p+1}}_{n-(n-p+1)=p-1}$$

and the sign exponent $P(n-P)$ needed for the inverse dual has $P = n - p + 1$, so

$$(n-p+1)(n-(n-p+1)) = (n-p+1)(p-1) = (n+1)(p-1) - p(p-1)$$

but since $p(p-1)$ is always even, it does not change the sign and may be dropped

$$(-1)^{(n+1)(p-1)-p(p-1)} = (-1)^{(n+1)(p-1)}.$$

Thus

$$*^{-1}(S \wedge *T) = (-1)^{M+(n+1)(p-1)}(S \wedge *T).$$

4.3.13: double dual sign in \mathbb{R}^4

4.3.14: index shifting

4.3.15: 2-vector duals in \mathbb{R}^4

4.3.16: inner product of two duals for a general inner product EDIT THIS

4.3.17: M^4 duals with indices 0,1,2,3

4.3.18: M^4 duals with indices 1,2,3,4

4.3.19: Euclidean \mathbb{R}^4 duals

4.3.20: complex plane and real wedge products

4.3.21: 2-planes in \mathbb{R}^4 and wedge products

4.4.1: transforming wedge products and star duals in the plane

4.5.1: exponentiating boost matrices

4.5.2: null rotations

4.5.3: antisymmetric 3×3 matrices and the negative dual vector

4.5.4: commutators of the Lorentz group Lie algebra

4.5.5: commutators of the (pseudo-)orthogonal group Lie algebras

4.5.6: rotations in \mathbb{R}^4

4.5.7: differentials of rotation matrices

4.5.8: unitary groups

If $\underline{K} = i\underline{\mathcal{K}}$ is anti-Hermitian, i.e., $\underline{\mathcal{K}} = -i\underline{K}$, then multiplying it by i makes it Hermitian

$$\underline{\mathcal{K}}^\dagger = (-i\underline{K})^\dagger = i(-\underline{K}) = \underline{\mathcal{K}}.$$

?: $h(2)$ matrices are Hermitian

4.5.9: the special unitary group $SU(2)$ and $SO(3, \mathbb{R})$

See the Maple worksheet `matrixproducts-nullrotations.mw`.

4.5.10: $SL(2, \mathbb{R})$ and the Lorentz group in 3 dimensions

4.5.11: quaternions?

4.5.12: squared angular momentum L^2

4.5.13: unitary groups again

Chapter 5

5.1.1: Some problems from 3-d calculus

5.1.2: tangent to level surfaces

We must show that $Xf = 0 = Xg$.

$$\begin{aligned} Xf &= \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z} \right) \left(\frac{y}{x} \right) \\ &= x \left(-\frac{y}{x^2} \right) + y \left(\frac{1}{x} \right) = 0. \end{aligned}$$

$$\begin{aligned} Xg &= \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z} \right) \left(\frac{z^2}{x^2 + y^2} \right) \\ &= x \left(-\frac{2xz^2}{(x^2 + y^2)^2} \right) + y \left(-\frac{2yz^2}{(x^2 + y^2)^2} \right) + \frac{2z^2}{x^2 + y^2} \\ &= -2z^2 \frac{x^2 + y^2}{(x^2 + y^2)^2} + \frac{z^2}{x^2 + y^2} = -\frac{2z^2}{x^2 + y^2} + \frac{2z^2}{x^2 + y^2} = 0. \end{aligned}$$

$$\begin{aligned} df &= d \left(\frac{y}{x} \right) \\ &= \left(-\frac{y}{x^2} \right) dx + \left(\frac{1}{x} \right) dy, \\ df(X) &= \left(\left(-\frac{y}{x^2} \right) dx + \left(\frac{1}{x} \right) dy \right) \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z} \right) \\ &= \left(-\frac{y}{x^2} \right) x + \frac{1}{x} y = 0. \end{aligned}$$

$$\begin{aligned} dg &= d \left(\frac{z^2}{x^2 + y^2} \right) \\ &= \left(-\frac{2xz^2}{(x^2 + y^2)^2} \right) dx + \left(-\frac{2yz^2}{(x^2 + y^2)^2} \right) dy + \frac{2z}{x^2 + y^2} dz, \\ dg(X) &= \left(\left(-\frac{2xz^2}{(x^2 + y^2)^2} \right) dx + \left(-\frac{2yz^2}{(x^2 + y^2)^2} \right) dy + \frac{2z}{x^2 + y^2} dz \right) \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z} \right) \\ &= \left(-\frac{2xz^2}{(x^2 + y^2)^2} \right) x + \left(-\frac{2yz^2}{(x^2 + y^2)^2} \right) y + \frac{2z}{x^2 + y^2} z = 0. \end{aligned}$$

5.1.3: elliptical level curves

5.2.1: polar coordinate calculations**5.3.1: matrix exponential chain rule**

$$\begin{aligned}
\frac{d}{dt}e^{t\underline{A}} &= \frac{d}{dt} \left(\sum_{k=0}^{\infty} \frac{t^k}{k!} \underline{A}^k \right) \\
&= \left(\sum_{k=0}^{\infty} \frac{k t^{k-1}}{k!} \underline{A}^k \right) \\
&= \left(\sum_{k=1}^{\infty} \frac{t^{k-1}}{(k-1)!} \underline{A}^k \right) \\
&= \underline{A} \left(\sum_{k=1}^{\infty} \frac{t^{k-1}}{(k-1)!} \underline{A}^{k-1} \right) \\
&= \underline{A} \left(\sum_{j=0}^{\infty} \frac{t^j}{j!} \underline{A}^j \right) \\
&= \underline{A} e^{t\underline{A}}.
\end{aligned}$$

5.3.2: hyperbolic rotations via matrix exponential**5.3.3: space rotations via the matrix exponential****5.3.4: Cayley-Hamilton theorem for $n = 3$**

See the Maple worksheet `cayleyhamilton.mw`. The case $n = 4$ is an interesting challenge. There is probably a general theory for evaluating the coefficients of the characteristic equation in terms of the determinant and trace of the powers, but it does not seem to show up in web searches. Try an old fashioned library?

5.3.5: rotations as solutions of a system of differential equations

See Maple worksheet `rotationmatrixeigenvectors.mw`.

5.3.6: local rest space decomposition in \mathbb{M}^4 **5.3.7: logarithmic spiral group****5.4.1: Lie bracket evaluation**

$$\begin{aligned}
a) \quad & X_1 = \partial_1, \quad X_2 = \partial_2, \quad X_3 = x^1 \partial_2 - x^2 \partial_1, \\
& [X_2, X_3] = -X_1, \quad [X_3, X_1] = -X_2, \quad [X_1, X_2] = 0. \\
b) \quad & X_1 = \partial_1, \quad X_2 = \partial_2, \quad X_3 = x^1 \partial_2 + x^2 \partial_1 \\
& [X_2, X_3] = X_1, \quad [X_3, X_1] = -X_2, \quad [X_1, X_2] = 0.
\end{aligned}$$

5.4.2: Lie bracket evaluation

$$\begin{aligned}
a) \quad [u, v] &= [x\partial_x + y\partial_y + z\partial_z, y\partial_x - x\partial_y] \\
&= x(-\partial_y) + y(\partial_x) - y(\partial_x) + x(\partial_y) = 0 \\
[u, w] &= [x\partial_x + y\partial_y + z\partial_z, (x^2 + y^2)(\partial_x + \partial_y) + \partial_z] \\
&= (x(2x) + y(2y))(\partial_x + \partial_y) - (x^2 + y^2)(\partial_x + \partial_y) - \partial_z \\
&= (x^2 + y^2)(\partial_x + \partial_y) - \partial_z \\
[v, w] &= [y\partial_x - x\partial_y, (x^2 + y^2)(\partial_x + \partial_y) + \partial_z] \\
&= (y(2x) - x(2y))(\partial_x + \partial_y) - (x^2 + y^2)(-\partial_y + \partial_x) \\
&= (x^2 + y^2)(-\partial_x + \partial_y) \\
b) \quad [u, v] &= [(x^2 + y^2)^{-1/2}(x\partial_x + y\partial_y), y\partial_x - x\partial_y] \\
&= (x^2 + y^2)^{-1/2}(x(-\partial_y) + y\partial_x) \\
&\quad - y(-1/2(x^2 + y^2)^{-3/2}(2x) + x(-1/2(x^2 + y^2)^{-3/2}(2y) \\
&\quad - (x^2 + y^2)^{-1/2}(y\partial_x + x\partial_y) = 0
\end{aligned}$$

When simplified these 4 terms cancel in pairs.

$|\underline{A}^{-1}| = (x^2 + y^2)^{1/2}$ vanishes at the origin where these vectors are no longer linearly independent. In fact v vanishes there, while u has direction dependent limits there.

5.4.3: linear vector field Lie brackets

This is a straightforward calculation.

$$\begin{aligned}
[X, Y] &= [A^i_j x^j \frac{\partial}{\partial x^i}, B^m_n x^n \frac{\partial}{\partial x^m}] \\
&= A^i_j x^j \frac{\partial}{\partial x^i} (B^m_n x^n \frac{\partial}{\partial x^m}) - B^m_n x^n \frac{\partial}{\partial x^m} (A^i_j x^j \frac{\partial}{\partial x^i}) \\
&= A^i_j x^j B^m_n (\frac{\partial x^n}{\partial x^i} \frac{\partial}{\partial x^m} + x^n \frac{\partial^2}{\partial x^i \partial x^m}) \\
&\quad - B^m_n x^n A^i_j (\frac{\partial x^n}{\partial x^i} \frac{\partial}{\partial x^m} + x^j \frac{\partial^2}{\partial x^i \partial x^m}) \\
&= A^i_j x^j B^m_n \underbrace{\frac{\partial x^n}{\partial x^i}}_{\delta^n_i} \frac{\partial}{\partial x^m} - B^m_n x^n A^i_j \underbrace{\frac{\partial x^j}{\partial x^m}}_{\delta^j_m} \frac{\partial}{\partial x^i} \\
&= B^m_i A^i_j x^j \frac{\partial}{\partial x^m} - A^i_j B^j_n x^n \frac{\partial}{\partial x^i} \\
&= B^m_i A^i_j x^j \frac{\partial}{\partial x^m} - A^m_i B^i_j x^j \frac{\partial}{\partial x^m} \\
&= [\underline{B}, \underline{A}]^m_j x^j \frac{\partial}{\partial x^m} = -[\underline{A}, \underline{B}]^m_j x^j \frac{\partial}{\partial x^m}
\end{aligned}$$

$$\begin{aligned}
[X, Z] &= [A^i_j x^j \frac{\partial}{\partial x^j}, b^\ell \frac{\partial}{\partial x^\ell}] \\
&= -b^\ell A^i_j \frac{\partial x^j}{\partial x^\ell} \frac{\partial}{\partial x^j} = -A^i_j b^j \frac{\partial}{\partial x^i} \\
&= -[\underline{A} \underline{b}]^i \frac{\partial}{\partial x^i}
\end{aligned}$$

$$[Z, W] = [b^\ell \frac{\partial}{\partial x^\ell}, C^b \frac{\partial}{\partial x^k}] = 0 \quad (\text{constant components})$$

We can make these results look mathematically pretty by defining

$$\sigma(\underline{A}) = A^i_j x^j \frac{\partial}{\partial x^i}, \quad \zeta(\underline{b}) = b^i \frac{\partial}{\partial x^i}.$$

Then

$$\begin{aligned}
(1) \quad [\sigma(\underline{A}), \sigma(\underline{B})] &= -\sigma([\underline{A}, \underline{B}]) \\
(2) \quad [\sigma(\underline{A}), \zeta(\underline{b})] &= -\zeta(\underline{A}, \underline{b}) \\
(3) \quad [\zeta(\underline{b}), \zeta(\underline{c})] &= 0
\end{aligned}$$

$\sigma : \{n \times n \text{ matrices}\} \longrightarrow \{\text{vector fields on } \mathbb{R}^n\}$ is a linear map from a vector space with a commutator (the matrix commutator) into a vector space with a commutator (the Lie bracket). To verify this linearity property check the following as an exercise

$$\sigma(a\underline{A} + b\underline{B}) = a\sigma(\underline{A}) + b\sigma(\underline{B}).$$

The relation (1) says you can do the commutator before or after the map and still get the same result apart from the minus sign, which can be included as a reflection in the map

$$[-\sigma(\underline{A}), -\sigma(\underline{B})] = -\sigma([\underline{A}, \underline{B}]).$$

Thus $-\sigma$ has the desired property that the order of evaluating the map and the commutator does not matter. [A vector space with a commutator is called a Lie algebra and such a map between Lie algebra mapping one commutator into the next is called a Lie algebra homomorphism.]

5.4.4: rotation generator Lie brackets

5.4.5: Laplacian

5.4.6: total angular momentum operator and the Laplacian

5.4.7: spherical basis?

5.5.1: polar coordinates and circles not centered at the origin

5.5.2: polar coordinates and multipetal curves**5.6.1: mathematical wedding band surface boundaries****5.7.1: transforming a vector field and 1-form****5.7.2: Laplacian in cylindrical coordinates****5.7.3: paracylindrical coordinates****5.8.1: Jacobian matrices for spherical coordinates****5.8.2: spherical coordinate frame rotation****5.8.3: differential, gradient in cylindrical, spherical coordinates**

Compute df and $\text{grad } f = \vec{\nabla} f$ in cylindrical coordinate and verify that you get our previous results quoted in the text.??

Consider the function

$$f = xy = \rho^2 \sin \phi \cos \phi = \frac{1}{2} \rho^2 \sin 2\phi = \frac{1}{2} r^2 \sin^2 \theta \sin 2\phi.$$

Then

$$\begin{aligned} df &= ydx + xdy = X^\flat \\ \vec{\nabla} f &= [df]^\sharp = y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} = X \end{aligned}$$

yields our friend X from previous exercises where we saw that

$$\begin{aligned} X &= \rho \sin 2\phi \frac{\partial}{\partial \rho} + \cos 2\phi \frac{\partial}{\partial \phi} = \sin \theta \sin 2\phi (r \sin \theta \frac{\partial}{\partial r} + \cos \theta \frac{\partial}{\partial \theta}) + \cos 2\phi \frac{\partial}{\partial \phi} \\ X^\flat &= \rho \sin 2\phi d\rho + \rho^2 \sin 2\phi d\phi = \sin \theta \sin 2\phi (r \sin \theta dr + r^2 \cos \theta d\theta) + r^2 \sin^2 \theta \cos 2\phi d\phi \end{aligned}$$

$$[\frac{\partial}{\partial r}]_i = g_{ij} [\frac{\partial}{\partial r}]^j = g_{ir} \longrightarrow [\frac{\partial}{\partial r}]^\flat = g_{ir} d\bar{x}^i = g_{rr} dr = dr$$

and similarly

$$\begin{aligned} [\frac{\partial}{\partial \phi}]^\flat &= g_{\phi i} d\bar{x}^i = g_{\phi\phi} d\phi = r^2 \sin^2 \theta d\phi \\ [\frac{\partial}{\partial \theta}]^\flat &= g_{\theta i} d\bar{x}^i = g_{\theta\theta} d\theta = r^2 d\theta \end{aligned}$$

In general

$$e_i^\flat = g_{kj} e_i^j \omega^k = g_{ik} \omega^k$$

so that

$$X^\flat = (X^i e_i)^\flat = X^i e_i^\flat = X^i g_{ik} \omega^k = X_k \omega^k.$$

Similarly

$$[\omega^i]^\sharp = g^{ij} e_j$$

holds for an orthogonal frame, index shifting the frame vectors and dual frame covectors yields the corresponding basis covector or vector multiplied by the diagonal metric component or its reciprocal.

5.8.4: spherical coordinates back to Cartesian coordinates

$$\begin{aligned}\frac{\partial}{\partial r} &= \frac{\partial x}{\partial r} \frac{\partial}{\partial x} + \frac{\partial y}{\partial r} \frac{\partial}{\partial y} + \frac{\partial z}{\partial r} \frac{\partial}{\partial z} \\ &= \frac{x}{r} \frac{\partial}{\partial x} + \frac{y}{r} \frac{\partial}{\partial y} + \frac{z}{r} \frac{\partial}{\partial z} \\ \frac{\partial}{\partial \phi} &= \frac{\partial x}{\partial \phi} \frac{\partial}{\partial x} + \frac{\partial y}{\partial \phi} \frac{\partial}{\partial y} + \frac{\partial z}{\partial \phi} \frac{\partial}{\partial z} \\ &= -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}\end{aligned}$$

$$\begin{aligned}Y &= r \frac{\partial}{\partial r} + \frac{\partial}{\partial \phi} \\ &= (x^2 + y^2 + z^2)^{1/2} \left[x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z} \right] / (x^2 + y^2 + z^2)^{1/2} + \left[-y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} \right] \\ &= x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z} - y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} = (x - y) \frac{\partial}{\partial x} + (x + y) \frac{\partial}{\partial y} + z \frac{\partial}{\partial z} .\end{aligned}$$

5.8.5: spherical coordinate Laplacian

5.9.1: spherical coordinate commutator using Cartesian coordinates

Calculating the Lie brackets of the spherical coordinates in terms of Cartesian coordinates is cumbersome. For example

$$\begin{aligned}[e_r, e_\phi] &= \left[(x^2 + y^2 + z^2)^{-1/2} \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z} \right), -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} \right] \\ &= (x^2 + y^2 + z^2)^{-1/2} \left[x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}, -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} \right] \\ &\quad - \underbrace{\left(\left(-y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} \right) (x^2 + y^2 + z^2)^{-1/2} \right)}_{=0} + \underbrace{\left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z} \right)}_{\equiv Q = 0 \text{ (next line)}} = 0 ,\end{aligned}$$

since

$$\begin{aligned}Q &= x \left(\frac{\partial}{\partial x} x \right) \frac{\partial}{\partial y} + y \frac{\partial}{\partial y} (-y) \frac{\partial}{\partial x} - \left(-y \frac{\partial}{\partial x} (x) \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} (y) \frac{\partial}{\partial y} \right) \\ &= x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} + y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} = 0 .\end{aligned}$$

5.9.2: structure functions of cylindrical and spherical coordinates

The orthonormal spherical coordinate frame

$$e_{\hat{r}} = \frac{\partial}{\partial r}, \quad e_{\hat{\theta}} = \frac{1}{r} \frac{\partial}{\partial \theta}, \quad e_{\hat{\phi}} = \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi}$$

has nonzero Lie brackets, easily calculated in spherical coordinates

$$\begin{aligned} [e_{\hat{r}}, e_{\hat{\theta}}] &= \left[\frac{\partial}{\partial r}, \frac{1}{r} \frac{\partial}{\partial \theta} \right] = -\frac{1}{r^2} \frac{\partial}{\partial \theta} = \underbrace{-\frac{1}{r}}_{C^{\hat{\theta}}_{\hat{r}\hat{\theta}}} e_{\hat{\theta}}, \\ [e_{\hat{\theta}}, e_{\hat{\phi}}] &= \left[\frac{1}{r} \frac{\partial}{\partial \theta}, \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi} \right] = \frac{1}{r^2} \frac{\partial}{\partial \theta} \left(\frac{1}{\sin \theta} \right) \frac{\partial}{\partial \phi} \\ &= -\frac{\cos \theta}{r^2 \sin^2 \theta} \frac{\partial}{\partial \phi} = \underbrace{-\frac{1}{r} \cot \theta}_{C^{\hat{\phi}}_{\hat{\theta}\hat{\phi}}} e_{\hat{\phi}} \\ [e_{\hat{\phi}}, e_{\hat{r}}] &= \left[\frac{1}{r \sin \theta} \frac{\partial}{\partial \phi}, \frac{\partial}{\partial r} \right] = \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \phi} = \underbrace{\frac{1}{r}}_{C^{\hat{\phi}}_{\hat{\phi}\hat{r}}} e_{\hat{\phi}}. \end{aligned}$$

5.9.3: Lie brackets in cylindrical coordinates

5.9.4: Lie brackets in spherical coordinate orthonormal frame

For the spherical coordinate orthonormal frame the nonzero structure functions are

$$C^{\hat{\theta}}_{\hat{r}\hat{\theta}} = -\frac{1}{r}, \quad C^{\hat{\phi}}_{\hat{\theta}\hat{\phi}} = -\frac{1}{r} \cot \theta, \quad C^{\hat{\phi}}_{\hat{\phi}\hat{r}} = \frac{1}{r}.$$

I used the cyclic order $23, 31, 12 \mapsto \hat{r}\hat{\theta}, \hat{\theta}\hat{\phi}, \hat{\phi}\hat{r}$ in computing the Lie brackets, rather than $i < j$: 23, 13, 12.

5.9.5: Lie brackets in cylindrical coordinate orthonormal frame

$$\begin{aligned} e_{\hat{\rho}} &= \frac{\partial}{\partial \rho}, \quad e_{\hat{\phi}} = \frac{1}{\rho} \frac{\partial}{\partial \phi}, \quad e_{\hat{z}} = \frac{\partial}{\partial z}, \\ [e_{\hat{\rho}}, e_{\hat{\phi}}] &= \left[\frac{\partial}{\partial \rho}, \frac{1}{\rho} \frac{\partial}{\partial \phi} \right] = -\frac{1}{\rho^2} \frac{\partial}{\partial \phi} = -\frac{1}{\rho} e_{\hat{\phi}}, \\ [e_{\hat{\phi}}, e_{\hat{z}}] &= \left[\frac{1}{\rho} \frac{\partial}{\partial \phi}, \frac{\partial}{\partial z} \right] = 0, \\ [e_{\hat{z}}, e_{\hat{\rho}}] &= \left[\frac{\partial}{\partial z}, \frac{\partial}{\partial \rho} \right] = 0, \end{aligned}$$

so

$$C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} = -\frac{1}{\rho} (= -C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}})$$

is the single independent structure function.

5.9.6: duality practice

Substitute e_1, e_2, e_3 by dx, dy, dz (i.e., p -vector $\rightarrow p$ -covector)

$$*1 = dx \wedge dy \wedge dz$$

$$*(x_1 dx + x_2 dy + x_3 dz) = x_1 dy \wedge dz + x_2 dz \wedge dx + x_3 dx \wedge dy$$

$$*(x_{23} dy \wedge dz + x_{31} dz \wedge dx + x_{12} dx \wedge dy) = x_{23} dx + x_{31} dy + x_{12} dz.$$

$$*(x_{123} dx \wedge dy \wedge dz) = x_{123}$$

$$*1 = \omega^{\hat{\rho}\hat{\phi}\hat{z}} = \omega^{\hat{\rho}} \wedge \omega^{\hat{\phi}} \wedge \omega^{\hat{z}} = d\rho \wedge (pd\phi) \wedge (dz) = \rho d\rho \wedge d\phi \wedge dz.$$

$$*(X_{\hat{\rho}}\omega^{\hat{\rho}} + X_{\hat{\phi}}\omega^{\hat{\phi}} + X_{\hat{z}}\omega^{\hat{z}}) = X_{\hat{\rho}}\omega^{\hat{\rho}\hat{z}} + X_{\hat{\phi}}\omega^{\hat{z}\hat{\phi}} + X_{\hat{z}}\omega^{\hat{\rho}\hat{\phi}}$$

$$*(X_{\hat{\phi}\hat{z}}\omega^{\hat{\phi}\hat{z}} + X_{\hat{z}\hat{\rho}}\omega^{\hat{z}\hat{\rho}} + X_{\hat{\rho}\hat{\phi}}\omega^{\hat{\rho}\hat{\phi}}) = X_{\hat{\rho}\hat{z}}\omega^{\hat{\rho}} + X_{\hat{z}\hat{\rho}}\omega^{\hat{\phi}} + X_{\hat{\rho}\hat{\phi}}\omega^{\hat{z}}$$

$$*(X_{\hat{\rho}\hat{\phi}\hat{z}}\omega^{\hat{\rho}\hat{\phi}\hat{z}}) = X_{\hat{\rho}\hat{\phi}\hat{z}}$$

$$*1 = \omega^{\hat{r}\hat{\theta}\hat{\phi}} = \omega^{\hat{r}} \wedge \omega^{\hat{\theta}} \wedge \omega^{\hat{\phi}} = dr \wedge (rd\theta) \wedge (r \sin \theta d\phi) = r^2 \sin \theta dr \wedge d\theta \wedge d\phi$$

$$\begin{aligned} X &= \overbrace{\frac{1}{r^2 \sin \theta} X_{\theta\phi}}^{X_{\hat{\theta}\hat{\phi}}} \omega^{\hat{\theta}\hat{\phi}} + \frac{1}{r \sin \theta} X_{\phi r} \omega^{\hat{\phi}\hat{r}} + \frac{1}{r} X_{r\theta} \omega^{\hat{r}\hat{\theta}} \\ *X &= \frac{1}{r^2 \sin \theta} X_{\theta\phi} \omega^{\hat{r}} + \frac{1}{r \sin \theta} X_{\phi r} \omega^{\hat{\theta}} + \frac{1}{r} X_{r\theta} \omega^{\hat{\phi}} \\ *X^{\sharp} &= \frac{1}{r^2 \sin \theta} X_{\theta\phi} e_{\hat{r}} + \frac{1}{r \sin \theta} X_{\phi r} e_{\hat{\theta}} + \frac{1}{r} X_{r\theta} e_{\hat{\phi}} \\ &= \frac{1}{r^2 \sin \theta} X_{\theta\phi} \frac{\partial}{\partial r} + \frac{1}{r^2 \sin \theta} X_{\phi r} \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin \theta} X_{r\theta} \frac{\partial}{\partial \phi} \\ (*X^{\sharp})^{\hat{r}} &= \frac{1}{r^2 \sin \theta} X_{\theta\phi} \quad (*X^{\sharp})^{\hat{\theta}} = \frac{1}{r \sin \theta} X_{\phi r} \quad (*X^{\sharp})^{\hat{\phi}} = \frac{1}{r} X_{r\theta} \\ (*X^{\sharp})^r &= \frac{1}{r^2 \sin \theta} X_{\theta\phi} \quad etc. \end{aligned}$$

5.9.7: dual of 2-form in \mathbb{R}^3

5.9.8: Lie brackets and the derivatives of the frame transformation matrix

5.9.9: rotation generator Lie brackets in spherical coordinates

5.??:

on page (40c) worked [Lie brackets in cylindrical coordinates]

$$\begin{aligned}
 [X, Y]^{\hat{\rho}} &= Y^{\hat{\rho}},_{\hat{\rho}} X^{\hat{\rho}} + Y^{\hat{\rho}},_{\hat{\phi}} X^{\hat{\phi}} - X^{\hat{\rho}},_{\hat{\rho}} Y^{\hat{\rho}} + C^{\hat{\rho}}_{\hat{j}\hat{k}} X^{\hat{j}} Y^{\hat{k}} \\
 &= \frac{\partial}{\partial \rho}(\rho) \cdot (\rho \sin 2\phi) + \frac{1}{\rho} \underbrace{\frac{\partial}{\partial \phi}(\rho)}_{=0} \cdot (\rho \cos 2\phi) - \frac{\partial}{\partial \rho}(\rho \sin 2\phi) \cdot \rho \\
 &= \rho \sin 2\phi - \rho \sin 2\phi = 0 \\
 [X, Y]^{\hat{\phi}} &= Y^{\hat{\phi}},_{\hat{\rho}} X^{\hat{\rho}} + Y^{\hat{\phi}},_{\hat{\phi}} X^{\hat{\phi}} - X^{\hat{\phi}},_{\hat{\rho}} Y^{\hat{\rho}} + C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} X^{\hat{\rho}} Y^{\hat{\phi}} + C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} X^{\hat{\phi}} Y^{\hat{\rho}} \\
 &= -\frac{\partial}{\partial \rho}(\rho \cos 2\phi) \cdot \rho + \rho \cos 2\phi = 0 \\
 [X, Y]^{\hat{z}} &= Y^{\hat{z}},_{\hat{i}} X^{\hat{i}} - X^{\hat{z}},_{\hat{i}} X^{\hat{i}} + C^{\hat{z}}_{\hat{j}\hat{k}} X^{\hat{j}} Y^{\hat{k}} = 0
 \end{aligned}$$

so $[X, Y] = 0$. Compare

$$\begin{aligned}
 [X, Y] &= [y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}, x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y}] \\
 &= y \frac{\partial}{\partial x}(x) \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}(y) \frac{\partial}{\partial y} - x \frac{\partial}{\partial x}(x) \frac{\partial}{\partial y} - y \frac{\partial}{\partial y}(y) \frac{\partial}{\partial x} \\
 &= y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} - x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} = 0 .
 \end{aligned}$$

Chapter 6

6.2.1: cylindrical frame connection components

The Maple package tensor is easily used to evaluate the connection components in a coordinate system.

$(\rho, \phi, z) \sim (1, 2, 3)$. Inverting the 2×2 block of \underline{A} gives

$$\underline{A}^{-1} = \begin{pmatrix} \cos \phi & -\rho \sin \phi & 0 \\ \sin \phi & \rho \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Check that $\underline{A}^{-1} \underline{A} = \underline{I}$. [also given at bottom of page (31).]

$$\begin{aligned} \underline{\bar{\omega}} = \underline{A} d \underline{A}^{-1} &= \begin{pmatrix} C & S & 0 \\ -\rho^{-1} S & \rho^{-1} C & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -\sin \phi d\phi & -\rho \cos \phi d\phi - \sin \phi d\rho & 0 \\ \cos \phi d\phi & -\rho \sin \phi d\phi + \cos \phi d\rho & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -\rho & 0 \\ \rho^{-1} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} d\phi + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \rho^{-1} d\rho \end{aligned}$$

so

$$\bar{\Gamma}^{\rho}_{\phi\phi} = -\rho \quad , \quad \bar{\Gamma}^{\phi}_{\phi\rho} = \rho^{-1} \quad , \quad \bar{\Gamma}^{\phi}_{\rho\phi} = \rho^{-1}$$

are the only nonzero connection components, from the three nonzero entries of $\bar{\omega}$ recalling the matrix indices $(1, 2, 3) \sim (\rho, \phi, z)$

The matrix \underline{a} for the associated orthonormal frame is obtained from \underline{A} by setting $\rho = 1$, so one can derive the corresponding result by putting $\rho = 1$ into the above calculation:

$$\hat{\bar{\omega}} \equiv (\bar{\omega}^{\hat{i}}_{\hat{k}}) \equiv (\bar{\Gamma}^{\hat{i}}_{\hat{j}\hat{k}} \bar{\omega}^{\hat{j}}) = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} d\phi$$

$\bar{\Gamma}^{\hat{\rho}}_{\hat{\phi}\hat{\phi}} = -1$, $\bar{\Gamma}^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} = 1$ are the only nonzero connection components, due to the rotation by the angle ϕ of the orthonormal vector fields $e_{\hat{\rho}}$ nad $e_{\hat{\phi}}$ relative to e_x and e_y .

6.3.1: matrix Lie algebra representation map

6.3.2: rotation generator

6.3.3: pseudoorthogonal generators

6.3.4: efficient use of connection 1-forms

6.3.5: properties of the vector covariant derivative component formula

6.4.1: covariant constancy of generalized Kronecker delta

Preliminary remark. In any frame we have the definition

$$\delta^{ij}_{mn} = \delta^i_m \delta^j_n - \delta^i_n \delta^j_m .$$

If we differentiate this equation

$$\delta^{ij}_{mn;k} = \delta^i_{m;k} \delta^j_n + \delta^i_m \delta^j_{n;k} - \delta^i_{n;k} \delta^j_m - \delta^i_n \delta^j_{m;k} = 0$$

then $\nabla \delta^{(2)} = 0$ follows from $\nabla \delta = 0$ and the product rule.

However, just using the formula in the barred frame

$$\begin{aligned} \bar{\delta}^{ij}_{mn;k} &= \bar{\delta}^{ij}_{mn,k} + \bar{\Gamma}^i_{k\ell} \bar{\delta}^{\ell j}_{mn} + \bar{\Gamma}^j_{k\ell} \bar{\delta}^{i\ell}_{mn} - \bar{\Gamma}^\ell_{km} \bar{\delta}^{ij}_{\ell n} - \bar{\Gamma}^\ell_{kn} \bar{\delta}^{ij}_{m\ell} \\ &= (\bar{\Gamma}^i_{km} \bar{\delta}^j_n - \bar{\Gamma}^i_{kn} \bar{\delta}^j_m) + (\bar{\Gamma}^j_{kn} \bar{\delta}^i_m - \bar{\Gamma}^j_{km} \bar{\delta}^i_n) \\ &\quad - (\bar{\Gamma}^i_{km} \bar{\delta}^j_n - \bar{\Gamma}^j_{km} \bar{\delta}^i_n) - (\bar{\Gamma}^j_{kn} \bar{\delta}^i_m - \bar{\Gamma}^i_{kn} \bar{\delta}^j_m) \\ &= 0 . \end{aligned}$$

6.4.2: covariant constant fields in cylindrical coordinates

Remember only

$$\Gamma^\rho_{\phi\phi} = -\rho , \quad \Gamma^\phi_{\phi\rho} = \rho^{-1} , \quad \Gamma^\phi_{\rho\phi} = \rho^{-1}$$

are nonzero, and these vectors have no z components and no components depend on z , so this is basically a 2-dimensional problem (anything with a z index vanishes). So the components of $X = \partial/\partial x$ are

$$X^\rho = \cos \phi \quad X^\phi = -\frac{\sin \phi}{\rho}$$

Writing down only the nonzero terms:

$$\begin{aligned} X^\rho_{;\rho} &= X^\rho_{,\rho} + \Gamma^\rho_{\rho i} X^i = 0 \\ X^\rho_{;\phi} &= X^\rho_{,\phi} + \Gamma^\rho_{\phi\phi} X^\phi = -\sin \phi + \sin \phi = 0 \\ X^\rho_{;z} &= X^\rho_{,z} + \Gamma^\rho_{zi} X^i = 0 \\ X^\phi_{;\rho} &= X^\phi_{,\rho} + \Gamma^\phi_{\rho\phi} X^\phi = \frac{1}{\rho^2} \sin \phi - \frac{1}{\rho^2} \sin \phi = 0 \\ X^\phi_{;\phi} &= X^\phi_{,\phi} + \Gamma^\phi_{\phi\rho} X^\rho = 0 \\ X^\phi_{;z} &= X^\phi_{,z} + \Gamma^\phi_{zi} X^i = 0 \\ X^z_{;i} &= X^z_{,i} + \Gamma^z_{ij} X^j = 0 \end{aligned}$$

so $X^i_{;j} = 0$, i.e., $\nabla X = 0$. There is nothing new to be gained by verifying $\nabla[\frac{\partial}{\partial y}] = 0$ so let's

move on to the covariant constant 1-form dx

$$\begin{aligned}
[dx]_\rho &= \cos \phi \quad [dx]_\phi = -\rho \sin \phi, \quad [dx]_{i,j} = [dx]_{i,j} - \Gamma_{ji}^\ell [dx]_\ell \\
[dx]_{i,z} &= [dx]_{i,z} - \Gamma_{zi}^\ell [dx]_\ell = 0, \quad [dx]_{z,i} = [dx]_{z,i} - \Gamma_{iz}^\ell [dx]_\ell = 0 \\
[dx]_{\rho;\rho} &= [dx]_{\rho,\rho} - \Gamma_{\rho\rho}^i [dx]_i = 0 \\
[dx]_{\rho;\phi} &= [dx]_{\rho,\phi} - \Gamma_{\phi\rho}^\phi [dx]_\phi = 0 \\
[dx]_{\phi;\rho} &= [dx]_{\phi,\rho} - \Gamma_{\rho\phi}^\phi [dx]_\phi = 0 \\
[dx]_{\phi;\phi} &= [dx]_{\phi,\phi} - \Gamma_{\phi\phi}^\rho [dx]_\rho = 0
\end{aligned}$$

Similarly there is nothing new to learn from $\nabla dy = 0$. Finally

$$[dz]_\rho = 0 = [dz]_\phi, \quad [dz]_z = 1$$

so

$$[dz]_{i,j} = [dz]_{i,j} - \Gamma_{ji}^z [dz]_z = -\Gamma_{ji}^z = 0.$$

6.5.1: orthogonal coordinate connection components

6.5.2: symmetry of connection components

6.5.3: differential log metric determinant

6.5.4: trace of the connection components

6.5.5: cylindrical coordinate connection components

We start from the formulas

$$\begin{aligned}
\bullet \quad \bar{\Gamma}_{jk}^i &= \frac{1}{2} \bar{g}^{i\ell} (\bar{g}_{\ell j, k} - \bar{g}_{jk, \ell} + \bar{g}_{k\ell, j}) \quad \text{but} \quad \bar{g}_{ij} = \bar{g}_{ji} \quad \text{so} \\
\bar{\Gamma}_{kj}^i &= \frac{1}{2} \bar{g}^{i\ell} (\bar{g}_{\ell k, j} - \bar{g}_{kj, \ell} + \bar{g}_{j\ell, k}) = \bar{\Gamma}_{jk}^i \\
\bullet \quad \bar{\Gamma}_{ijk} &= \frac{1}{2} (\bar{g}_{ij, k} - \bar{g}_{jk, i} + \bar{g}_{ki, j})
\end{aligned}$$

[Note this is symmetric in (jk) for the same reason]

$$\bar{g}_{\rho\rho} = 1 = \bar{g}_{zz}, \quad \bar{g}_{\phi\rho} = \rho^2.$$

At least two indices have to be the same to get a diagonal metric component to differentiate, otherwise you differentiate an off diagonal metric component which is zero. Finally the only diagonal component with a nonzero derivative is $g_{\phi\phi} = \rho^2$ so the indices have to be some permutation of $(\phi\phi\rho)$ to get a nonzero result.

$$\begin{aligned}
\Gamma_{\rho\phi\phi} &= \frac{1}{2} (g_{\rho\phi, \phi} - g_{\phi\phi, \rho} + g_{\phi\rho, \phi}) = \frac{1}{2} (2\rho) = -\rho \\
\Gamma_{\phi\rho\phi} &= \frac{1}{2} (g_{\phi\rho, \phi} - g_{\rho\phi, \phi} + g_{\phi\phi, \rho}) = \frac{1}{2} (2\rho) = \rho \\
\Gamma_{\phi\phi\rho} &= \frac{1}{2} (g_{\phi\phi, \rho} - g_{\phi\rho, \phi} + g_{\rho\phi, \phi}) = \Gamma_{\phi\rho\phi} = \rho
\end{aligned}$$

since symmetric in last two indices in coordinate frame. Now raise first index:

$$\begin{aligned}\Gamma^\rho_{\phi\phi} &= g^{\rho\rho}\Gamma_{\rho\phi\phi} = -\rho \\ \Gamma^\phi_{\rho\phi} &= g^{\phi\phi}\Gamma_{\phi\rho\phi} = \rho^{-2}(\rho) = \frac{1}{\rho} = \Gamma^\phi_{\phi\rho}. \quad \text{Done.}\end{aligned}$$

Interpretation:

$$\nabla_{e_\phi} e_\phi = -\rho e_\rho, \quad \nabla_{e_\phi} e_\rho = \frac{1}{\rho} e_\phi = \nabla_{e_\rho} e_\phi$$

e_ϕ has length ρ . Translate its value at $(\rho, \phi\Delta\phi)$ back to (ρ, ϕ) so has same initial point as e_ϕ at (ρ, ϕ) . Difference is $\approx -\rho\Delta\phi$ in radial direction. Try interpreting another.

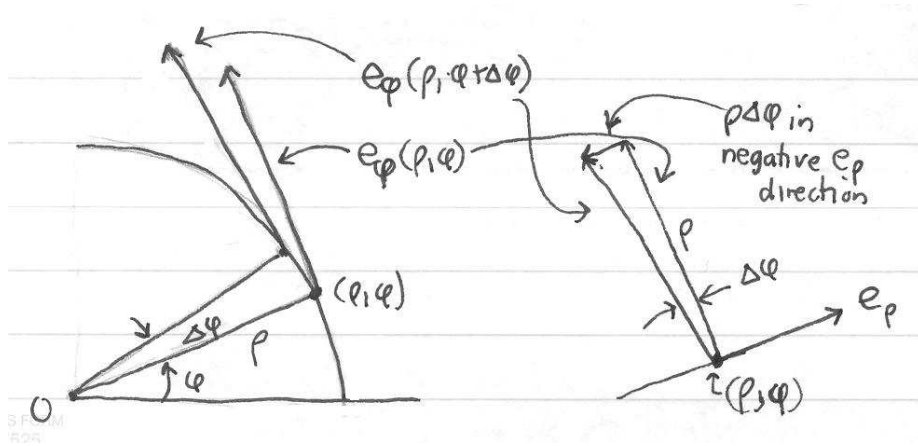


Figure E.6: Geometrically determining how e_ϕ rotates as one increases ϕ .

6.5.6: covariant constant tensor

Preliminary remark. Whatever symmetries a tensor has, its covariant derivative has the same symmetries. $T_{ij} = T_{ji}$ is symmetric so

$$\begin{aligned}T_{ij;k} &= T_{ij,k}^{(3)} - \Gamma_{ki}^\ell T_{\ell j}^{(1)} - \Gamma_{kj}^\ell T_{i\ell}^{(2)} \\ T_{ji;k} &= T_{ji,k}^{(3)} - \Gamma_{ki}^\ell T_{\ell j}^{(2)} - \Gamma_{ki}^\ell T_{j\ell}^{(1)} = T_{ij,k}^{(3)} - \Gamma_{ki}^\ell T_{\ell j}^{(1)} - \Gamma_{kj}^\ell T_{i\ell}^{(2)} \\ &= T_{ij;k}\end{aligned}$$

A symmetric 2-index object in 3-dimensions has 6 independent components. Its covariant derivative has $6 \times 3 = 18$ in general (still a lot, no?). But really this is a 2-dimensional problem because nothing depends on z and no z components are nonzero so no z -component of $T_{ij;k}$ is nonzero. So we have 3 independent components of T_{ij} times 2 for its covariant derivative for a

grand total 6. Not too bad. Thus we calculate

$$\begin{aligned}
(T_{\rho\rho}, T_{\phi\phi}, T_{\rho\phi}) &= (\cos^2 \phi, \rho^2 \sin^2 \phi, -\rho \sin \phi \cos \phi) \\
\Gamma^\rho_{\phi\phi} &= -\rho, \quad \Gamma^\phi_{\rho\phi} = \Gamma^\phi_{\phi\rho} = \rho^{-1} \\
T_{\rho\rho;\rho} &= T_{\rho\rho,\rho} - \Gamma^i_{\rho\rho} T_{i\rho} - \Gamma^i_{\rho\rho} T_{\rho i} = 0 \\
T_{\rho\rho;\phi} &= T_{\rho\rho,\phi} - \Gamma^i_{\phi\rho} T_{i\rho} - \Gamma^i_{\rho\rho} T_{\rho i} = -2 \cos \phi \sin \phi + 2 \sin \phi \cos \phi = 0 \\
T_{\phi\phi;\rho} &= T_{\phi\phi,\rho} - \Gamma^i_{\rho\phi} T_{i\phi} - \Gamma^i_{\rho\phi} T_{\phi i} = 2\rho \sin^2 \phi - 2\rho \sin^2 \phi = 0 \\
T_{\phi\phi;\phi} &= T_{\phi\phi,\phi} - \Gamma^i_{\phi\phi} T_{i\phi} - \Gamma^i_{\phi\phi} T_{\phi i} = 2\rho^2 \sin \phi \cos \phi - 2\rho^2 \sin \phi \cos \phi = 0 \\
T_{\rho\phi;\rho} &= T_{\rho\phi,\rho} - \Gamma^i_{\rho\rho} T_{i\phi} - \Gamma^i_{\rho\phi} T_{\rho i} = -\sin \phi \cos \phi + \sin \phi \cos \phi = 0 \\
T_{\rho\phi;\phi} &= T_{\rho\phi,\phi} - \Gamma^i_{\phi\rho} T_{i\phi} - \Gamma^i_{\phi\phi} T_{\rho i} = -\rho(\cos^2 \phi - \sin^2 \phi) - \rho \sin^2 \phi + \rho \cos^2 \phi = 0
\end{aligned}$$

Thus all components are zero as expected.

6.6.1: antisymmetric part of connection components

$$\begin{aligned}
\Gamma^i_{[jk]} &= \{[jk]^i\} + \frac{1}{2}(C^i_{[jk]} - C_{[jk]}^i + C_{[k\ j]}^i) \\
&= \frac{1}{2}C^i_{jk}
\end{aligned}$$

if you don't believe it :

$$C_k^i{}_{\ j} = g_{km}g^{in}C^m_{\ nj} = -g_{km}g^{in}C^m_{\ jn} = -C_{kj}^i$$

6.6.2: cylindrical coordinate orthonormal frame connection components

In Exercise 5.9.5 we found

$$C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} = -\frac{1}{\rho} = -C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}}$$

(only nonzero structure function)

So to get a nonzero component of $\Gamma^i_{\hat{j}\hat{k}}$ the indices must be a permutation of $(\hat{\phi}, \hat{\phi}, \hat{\rho})$.

$$\begin{aligned}
\Gamma^{\hat{\rho}}_{\hat{\phi}\hat{\phi}} &= \frac{1}{2}(C^{\hat{\rho}}_{\hat{\phi}\hat{\phi}} - C^{\hat{\phi}}_{\hat{\phi}\hat{\phi}} + C^{\hat{\phi}}_{\hat{\phi}\hat{\phi}}) = \frac{1}{2}(C^{\hat{\rho}}_{\hat{\phi}\hat{\phi}} - C^{\hat{\phi}}_{\hat{\phi}\hat{\phi}} + C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}}) \\
&= C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} = -\frac{1}{\rho} \\
\Gamma^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} &= \frac{1}{2}(C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} - C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} + C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}}) = \frac{1}{2}(C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} - C^{\hat{\rho}}_{\hat{\phi}\hat{\phi}} + C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}}) = 0 \\
\Gamma^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} &= \frac{1}{2}(C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} - C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} + C^{\hat{\rho}}_{\hat{\phi}\hat{\phi}}) = \frac{1}{2}(C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} - C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} + C^{\hat{\rho}}_{\hat{\phi}\hat{\phi}}) \\
&= C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} = \frac{1}{\rho}
\end{aligned}$$

Compare with page (58)?? and oops! I forgot to normalize!

$$\underline{\hat{\omega}} = (\bar{\Gamma}^i_{\hat{j}\hat{k}} \bar{\omega}^{\hat{j}}) = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} d\phi = \begin{pmatrix} 0 & -\rho^{-1} & 0 \\ \rho^{-1} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \omega^{\hat{\phi}}$$

so

$$\bar{\Gamma}^{\hat{\rho}}_{\hat{\phi}\hat{\phi}} = -\rho^{-1} \quad , \quad \bar{\Gamma}^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} = \rho^{-1}$$

agreement

6.6.3: constant metric component connection

6.6.4: the torsion tensor

6.7.1: Jacobi identity components

6.7.2: commutators of rotations and translations

6.7.3: Lie brackets of linear trasformation generating vector fields

6.7.4: polar coordinate vector fields

6.7.5: 3-sphere vector fields

6.8.1: (pseudo-) orthogonal group generators are Killing vector fields

6.8.2: comma to semicolon rule for Lie derivative of a metric

6.8.3: 1-form Lie derivative

6.8.4: Lie derivative and the Jacobi identity

6.8.5: complex numbers and rotations

6.8.6: gauge invariant derivative

6.8.7: vector potential for electromagnetic field

6.8.8: non-Abelian gauge theories

6.8.9: angular momentum ladder operators and representation theory for $SU(2) \sim SO(3, \mathbb{R})$

6.8.10: Gell-Mann matrices

6.9.1: angular momentum commutation relations

?: tracefree Lie algebra

6.??:

on page (57) worked

Preliminary remark. In any frame we have the definition

$$\delta^{ij}_{mn} = \delta^i_m \delta^j_n - \delta^i_n \delta^j_m .$$

If we differentiate this equation

$$\delta^{ij}_{mn;k} = \delta^i_{m;k} \delta^j_n + \delta^i_m \delta^j_{n;k} - \delta^i_{n;k} \delta^j_m - \delta^i_n \delta^j_{m;k} = 0$$

then $\nabla \delta^{(2)} = 0$ follows from $\nabla \delta = 0$ and the product rule.

However, just using the formula in the barred frame

$$\begin{aligned} \bar{\delta}^{ij}_{mn;k} &= \bar{\delta}^{ij}_{mn,k} + \bar{\Gamma}^i_{kl} \bar{\delta}^{\ell j}_{mn} + \bar{\Gamma}^j_{kl} \bar{\delta}^{i\ell}_{mn} - \bar{\Gamma}^\ell_{km} \bar{\delta}^{ij}_{\ell n} - \bar{\Gamma}^\ell_{kn} \bar{\delta}^{ij}_{m\ell} \\ &= (\bar{\Gamma}^i_{km} \bar{\delta}^j_n - \bar{\Gamma}^i_{kn} \bar{\delta}^j_m) + (\bar{\Gamma}^j_{kn} \bar{\delta}^i_m - \bar{\Gamma}^j_{km} \bar{\delta}^i_n) - (\bar{\Gamma}^i_{km} \bar{\delta}^j_n - \bar{\Gamma}^j_{km} \bar{\delta}^i_n) - (\bar{\Gamma}^j_{kn} \bar{\delta}^i_m - \bar{\Gamma}^i_{kn} \bar{\delta}^j_m) \\ &= 0 . \end{aligned}$$

??:

on page (59) worked.

Remember only

$$\Gamma^\rho_{\phi\phi} = -\rho , \quad \Gamma^\phi_{\phi\rho} = \rho^{-1} , \quad \Gamma^\phi_{\rho\phi} = \rho^{-1}$$

are nonzero X, Y and no Z components, no components depend on Z , so this is basically a 2-dimensional problem. (anything with a Z index vanishes).

so:

$$X^\rho = \cos \phi \quad X^\phi = -\frac{\sin \phi}{\rho} \quad \text{write down only nonzero terms:}$$

$$X^\rho_{;\rho} = X^\rho_{,\rho} + \Gamma^\rho_{\rho i} X^i = 0$$

$$X^\rho_{;\phi} = X^\rho_{,\phi} + \Gamma^\rho_{\phi\phi} X^\phi = -\sin \phi + \sin \phi = 0$$

$$X^\rho_{;z} = X^\rho_{,z} + \Gamma^\rho_{zi} X^i = 0$$

$$X^\phi_{;\rho} = X^\phi_{,\rho} + \Gamma^\phi_{\rho\phi} X^\phi = \frac{1}{\rho^2} \sin \phi - \frac{1}{\rho^2} \sin \phi = 0$$

$$X^\phi_{;\phi} = X^\phi_{,\phi} + \Gamma^\phi_{\phi\rho} X^\rho = 0$$

$$X^\phi_{;z} = X^\phi_{,z} + \Gamma^\phi_{zi} X^i = 0$$

$$X^z_{;i} = X^z_{,i} + \Gamma^z_{ij} X^j = 0$$

so $X^i_{;j} = 0$, i.e., $\nabla X = 0$. nothing new to be gained by doing $\nabla[\frac{\partial}{\partial y}] = 0$ so let's move on :

$$\begin{aligned} [dx]_\rho &= \cos \phi \quad [dx]_\phi = -\rho \sin \phi , \quad [dx]_{i;j} = [dx]_{i,j} - \Gamma^\ell_{ji} [dx]_\ell \\ [dx]_{i;z} &= [dx]_{i,z} - \Gamma^\ell_{zi} [dx]_\ell = 0 , \quad [dx]_{z;i} = [dx]_{z,i} - \Gamma^\ell_{iz} [dx]_\ell = 0 \end{aligned}$$

$$[dx]_{\rho;\rho} = [dx]_{\rho,\rho} - \Gamma^i_{\rho\rho} [dx]_i = 0$$

$$[dx]_{\rho;\phi} = [dx]_{\rho,\phi} - \Gamma^\phi_{\phi\rho} [dx]_\phi = 0$$

$$[dx]_{\phi;\rho} = [dx]_{\phi,\rho} - \Gamma^\phi_{\rho\phi} [dx]_\phi = 0$$

$$[dx]_{\phi;\phi} = [dx]_{\phi,\phi} - \Gamma^\rho_{\phi\phi} [dx]_\rho = 0$$

nothing new from $\nabla dy = 0$. Finally

$$[dz]_\rho = 0 = [dz]_\phi \quad , \quad [dz]_z = 1$$

so

$$[dz]_{i,j} = [dz]_{i,j} - \Gamma^z_{ji}[dz]_z = -\Gamma^z_{ji} = 0 \quad .$$

6.??:

on page 62 worked

$$\begin{aligned} \bullet \quad \bar{\Gamma}^i_{jk} &= \frac{1}{2} \bar{g}^{i\ell} (\bar{g}_{\ell j, k} - \bar{g}_{jk, \ell} + \bar{g}_{k\ell, j}) \quad \text{but} \quad \bar{g}_{ij} = \bar{g}_{ji} \quad \text{so} \\ \bar{\Gamma}^i_{kj} &= \frac{1}{2} \bar{g}^{i\ell} (\bar{g}_{\ell k, j} - \bar{g}_{kj, \ell} + \bar{g}_{j\ell, k}) = \bar{\Gamma}^i_{jk} \\ \bullet \quad \bar{\Gamma}_{ijk} &= \frac{1}{2} (\bar{g}_{ij, k} - \bar{g}_{jk, i} + \bar{g}_{ki, j}) \end{aligned}$$

[Note this is symmetric in (jk) for the same reason]

$$\bar{g}_{\rho\rho} = 1 = \bar{g}_{zz} \quad , \quad \bar{g}_{\phi\rho} = \rho^2 \quad .$$

At least two indices have to be the same to get a diagonal metric component to differentiate, otherwise you differentiate an off diagonal metric component which is zero. Finally the only diagonal component with a nonzero derivative is $g_{\phi\phi} = \rho^2$ so the indices have to be some permutation of $(\phi\phi\rho)$ to get a nonzero result.

$$\begin{aligned} \Gamma_{\rho\phi\phi} &= \frac{1}{2} (g_{\rho\phi, \phi} - g_{\phi\phi, \rho} + g_{\phi\rho, \phi}) = \frac{1}{2} (2\rho) = -\rho \\ \Gamma_{\phi\rho\phi} &= \frac{1}{2} (g_{\phi\rho, \phi} - g_{\rho\phi, \phi} + g_{\phi\phi, \rho}) = \frac{1}{2} (2\rho) = \rho \\ \Gamma_{\phi\phi\rho} &= \frac{1}{2} (g_{\phi\phi, \rho} - g_{\phi\rho, \phi} + g_{\rho\phi, \phi}) = \Gamma_{\phi\rho\phi} = \rho \end{aligned}$$

since symmetric in last two indices in coordinate frame. Now raise first index:

$$\begin{aligned} \Gamma^\rho_{\phi\phi} &= g^{\rho\rho} \Gamma_{\rho\phi\phi} = -\rho \\ \Gamma^\phi_{\rho\phi} &= g^{\phi\phi} \Gamma_{\phi\rho\phi} = \rho^{-2}(\rho) = \frac{1}{\rho} = \Gamma^\phi_{\phi\rho}. \quad \text{Done.} \end{aligned}$$

Interpretation:

$$\nabla_{e_\phi} e_\phi = -\rho e_\rho \quad , \quad \nabla_{e_\phi} e_\rho = \frac{1}{\rho} e_\phi = \nabla_{e_\rho} e_\phi$$

e_ϕ has length ρ . Translate its value at $(\rho, \phi\Delta\phi)$ back to (ρ, ϕ) so has same initial point as e_ϕ at (ρ, ϕ) . Difference is $\approx -\rho\Delta\phi$ in radial direction. Try interpreting another.

6.??:

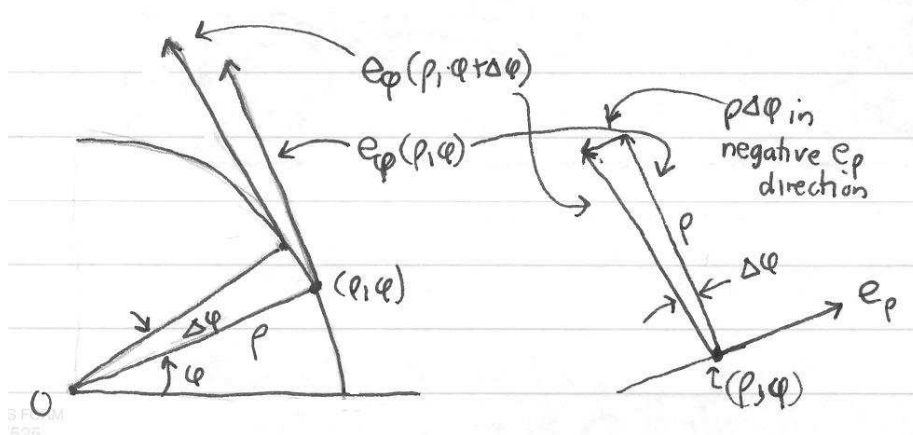


Figure E.7: Geometrically determining how $e_{\hat{\phi}}$ rotates as one increases ϕ .

last exercise on page (62) worked

Preliminary remark. Whatever symmetries a tensor has, its covariant derivative has the same symmetries.

6.??:

on page (64) worked

$$\begin{aligned}\Gamma^i_{[jk]} &= \{[jk]^i\} + \frac{1}{2}(C^i_{[jk]} - C_{[jk]}^i + C_{[k\ j]}^i) \\ &= \frac{1}{2}C^i_{jk}\end{aligned}$$

if you don't believe it :

$$C_k^i{}^j = g_{km}g^{in}C^m{}_{nj} = -g_{km}g^{in}C^m{}_{jn} = -C_{kj}^i$$

On page (53) we found

$$C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} = -\frac{1}{\rho} = -C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}}$$

(only nonzero structure function)

So to get a nonzero component of $\Gamma^{\hat{i}}_{\hat{j}\hat{k}}$ the indices must be a permutation of $(\hat{\phi}, \hat{\phi}, \hat{\rho})$.

$$\begin{aligned}
 \Gamma^{\hat{\rho}}_{\hat{\phi}\hat{\phi}} &= \frac{1}{2}(C^{\hat{\rho}}_{\hat{\phi}\hat{\phi}} - C_{\hat{\phi}\hat{\phi}}^{\hat{\rho}} + C_{\hat{\phi}}^{\hat{\rho}}_{\hat{\phi}}) = \frac{1}{2}(C^{\hat{\rho}}_{\hat{\phi}\hat{\phi}} - C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} + C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}}) \\
 &= C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} = -\frac{1}{\rho} \\
 \Gamma^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} &= \frac{1}{2}(C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} - C_{\hat{\rho}\hat{\phi}}^{\hat{\phi}} + C_{\hat{\phi}}^{\hat{\phi}}_{\hat{\rho}}) = \frac{1}{2}(C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} - C^{\hat{\rho}}_{\hat{\phi}\hat{\phi}} + C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}}) = 0 \\
 \Gamma^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} &= \frac{1}{2}(C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} - C_{\hat{\phi}\hat{\rho}}^{\hat{\phi}} + C_{\hat{\rho}}^{\hat{\phi}}_{\hat{\phi}}) = \frac{1}{2}(C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} - C^{\hat{\phi}}_{\hat{\rho}\hat{\phi}} + C^{\hat{\rho}}_{\hat{\phi}\hat{\phi}}) \\
 &= C^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} = \frac{1}{\rho}
 \end{aligned}$$

Compare with page (58??) and oops! I forgot to normalize!

$$\underline{\hat{\omega}} = (\bar{\Gamma}^{\hat{i}}_{\hat{j}\hat{k}} \bar{\omega}^{\hat{j}}) = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} d\phi = \begin{pmatrix} 0 & -\rho^{-1} & 0 \\ \rho^{-1} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \omega^{\hat{\phi}}$$

so

$$\bar{\Gamma}^{\hat{\rho}}_{\hat{\phi}\hat{\phi}} = -\rho^{-1} \quad , \quad \bar{\Gamma}^{\hat{\phi}}_{\hat{\phi}\hat{\rho}} = \rho^{-1} \quad \text{agreement}$$

Chapter 7

7.1.1: gradient in cylindrical and spherical coordinates

7.1.2: curl and div in cylindrical coordinates

7.1.3: more curl and div in cylindrical coordinates

7.1.5: still more curl and div in cylindrical and spherical coordinates

7.2.1: harmonic coordinates

7.2.2: harmonic function

11.8.2: grad, curl and div in cylindrical and spherical coordinates

7.2.4: Laplacian and angular momentum

$$\begin{aligned}
 \frac{\partial^2}{\partial x^2} &= \left(\frac{\partial}{\partial x} \right)^2 = \left(\cos \phi \frac{\partial}{\partial \rho} - \frac{\sin \phi}{\rho} \frac{\partial}{\partial \phi} \right) \left(\cos \phi \frac{\partial}{\partial \rho} - \frac{\sin \phi}{\rho} \frac{\partial}{\partial \phi} \right) \\
 &= \cos^2 \phi \frac{\partial^2}{\partial \rho^2} - \frac{2 \cos \phi \sin \phi}{\rho} \frac{\partial^2}{\partial \rho \partial \phi} - \frac{\sin \phi}{\rho} (-\sin \phi) \frac{\partial}{\partial \rho} + \frac{\sin^2 \phi}{\rho^2} \frac{\partial^2}{\partial \phi^2} \\
 &\quad + \left[\cos \phi \left(\frac{\sin \phi}{\rho^2} \right) + \frac{2 \sin \phi \cos \phi}{\rho^2} \right] \frac{\partial}{\partial \phi} \\
 &= \cos^2 \phi \frac{\partial^2}{\partial \rho^2} + \frac{\sin^2 \phi}{\rho} \frac{\partial}{\partial \rho} + \frac{\sin^2 \phi}{\rho^2} \frac{\partial^2}{\partial \phi^2} - \frac{2 \cos \phi \sin \phi}{\rho} \frac{\partial^2}{\partial \rho \partial \phi} + \frac{2 \cos \phi \sin \phi}{\rho^2} \frac{\partial}{\partial \phi}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial^2}{\partial y^2} &= \left(\frac{\partial}{\partial y} \right)^2 = \left(\sin \phi \frac{\partial}{\partial \rho} + \frac{\cos \phi}{\rho} \frac{\partial}{\partial \phi} \right) \left(\sin \phi \frac{\partial}{\partial \rho} + \frac{\cos \phi}{\rho} \frac{\partial}{\partial \phi} \right) \\
 &= \sin^2 \phi \frac{\partial^2}{\partial \rho^2} + \frac{\cos \phi (\cos \phi)}{\rho} \frac{\partial}{\partial \rho} + \frac{2 \sin \phi \cos \phi}{\rho} \frac{\partial^2}{\partial \rho \partial \phi} + \frac{\cos^2 \phi}{\rho^2} \frac{\partial^2}{\partial \phi^2} \\
 &\quad + \left[\frac{\sin \phi \cos \phi}{-\rho^2} + \frac{\cos \phi}{\rho} \left(\frac{-\sin \phi}{\rho} \right) \right] \frac{\partial}{\partial \phi} \\
 &= \sin^2 \phi \frac{\partial^2}{\partial \rho^2} + \frac{\cos^2 \phi}{\rho} \frac{\partial}{\partial \rho} + \frac{\cos^2 \phi}{\rho^2} \frac{\partial^2}{\partial \phi^2} + \frac{2 \sin \phi \cos \phi}{\rho} \frac{\partial^2}{\partial \rho \partial \phi} - \frac{2 \cos \phi \sin \phi}{\rho^2} \frac{\partial}{\partial \phi}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} &= (\cos^2 \phi + \sin^2 \phi) \frac{\partial^2}{\partial \rho^2} + \frac{\sin^2 \phi + \cos^2 \phi}{\rho} \frac{\partial}{\partial \rho} + \frac{\sin^2 \phi + \cos^2 \phi}{\rho^2} \frac{\partial^2}{\partial \phi^2} \\
 &\quad + \left[\frac{-\cos \phi \sin \phi}{\rho} + \frac{\cos \phi \sin \phi}{\rho} \right] \frac{\partial^2}{\partial \rho \partial \phi} + \left[\frac{2 \cos \phi \sin \phi}{\rho^2} - \frac{2 \cos \phi \sin \phi}{\rho^2} \right] \frac{\partial}{\partial \phi} \\
 &= \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2}
 \end{aligned}$$

Thus the Laplacian expressed in cylindrical coordinates is

$$\nabla^2 = \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}.$$

Replacing (ρ, ϕ) by (r, θ) in the preliminary result before this yields the Laplacian ∇^2 on \mathbb{R}^2 in polar coordinates

$$\nabla^2 = \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2}.$$

Note that

$$\frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial f}{\partial \rho} \right) = \frac{\partial^2 f}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial f}{\partial \rho},$$

so the first two terms can be rewritten as

$$\frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial}{\partial \rho} \right).$$

7.2.5: angular momentum and Cartesian coordinate functions

7.3.1: matrix product representation of orthonormal frame

7.3.2: spherical coordinate orthonormal frame connection vector

7.3.3: spherical coordinate connection 1-forms

??: gradient and differential in cylindrical coordinates

Chapter 8

- 8.1.1: directional derivative along a curve
- 8.1.2: directional derivative along curve in cylindrical coordinates
- 8.1.3: covariant derivative in spherical coordinates
- 8.2.1: parallel transport along lines of latitude
- 8.2.2: parallel combed hair on the sphere
- 8.2.3: Tangent cone to a sphere
- 8.3.1: covariant geodesic equation
- 8.3.2: geodesic coordinate lines
- 8.3.3: Lorentz force
- 8.4.1: metric for a surface of revolution
- 8.4.2: tangent cone to surface of revolution
- 8.4.3: planes, cylinders and cones
- 8.4.4: black hole embedding surfaces
- 8.4.5: parallel transport along circles
- 8.5.1: geodesics on a surface of revolution
- 8.5.2: surface of revolution meridian arclength
- 8.6.1: plane geodesics in polar coordinates
- 8.6.2: orbit equation for plane geodesics in polar coordinates
- 8.6.3: quadratic potential motion
- 8.6.4: radius versus time
- 8.12.1: black hole orbits
- 8.7.1: intrinsic osculating circle
- 8.7.2: geodesics on the cylinder

- 8.7.3: geodesics on the unit sphere
- 8.7.4: geodesics on the unit sphere: orbit equation
- 8.7.5: orthonormal coordinate frame connection 1-form matrix by transformation
- 8.7.6: connection in spacelike pseudospherical coordinates
- 8.7.7: connection on the spacelike pseudosphere
- 8.7.8: connection in timelike pseudospherical coordinates
- 8.7.9: geodesics on the unit timelike pseudosphere
- 8.7.10: geodesics on the unit hyperboloid of one sheet
- 8.7.11: parabola of revolution geodesics
- 8.7.12: geodesics on ellipse of revolution
- 8.8.1: toroidal coordinates
- 8.8.2: toroidal coordinates for the torus
- 8.8.3: toroidal coordinate metric
- 8.8.4: surface of revolution connection
- 8.8.5: geodesics on the unit torus
- 8.9.1: Lagrangian equations for geodesics
- 8.9.2: simple harmonic oscillator
- 8.9.3: principle of least action for a charged particle
- 8.9.4: spherical pendulum: gravity as geometry
- 8.11.1: cavatappo 2.0
- 8.11.2: the Lorentz cavatappo 2.0 surface
- 8.11.3: tilted helical surfaces
- 8.11.5: helicoids

Chapter 9

9.1.1: Riemann, Ricci and Einstein in 3 dimensions

9.1.2: Riemann, Ricci and Einstein in 2 dimensions

9.1.3: covariant components of Riemann

a) The antisymmetric part over 3 indices collapses from 6 to 3 terms because of the anti-symmetry on the last pair of indices

$$\begin{aligned} R^i_{[jkl]} &= \frac{1}{3!}(R^i_{jkl} + R^i_{k\ell j} + R^i_{\ell jk} - R^i_{j\ell k} - R^i_{kjl} - R^i_{\ell kj}) \\ &= \frac{1}{3}(R^i_{jkl} + R^i_{k\ell j} + R^i_{\ell jk}) = 0. \end{aligned}$$

d) The number of independent conditions represented by the Bianchi identities of the first kind is the number of combinations of n things taken 4 at a time:

$$\binom{n}{4} = \frac{n!}{4!(n-4)!} = \frac{1}{4!}n(n-1)(n-2)(n-3).$$

The four index values must be distinct since if one value is repeated, we saw that it reduces to some combination of the pair interchange symmetries. Given one value for the first position in the Bianchi identities, one can then use the pair interchange symmetries to move any other of three remaining values into the first slot in all three terms, so every combination of the 4 values is equivalent to any other. Their order does not matter so the number of combinations of n things taken 4 at a time gives the number of independent conditions represented by the Bianchi identities. This took me quite some time to reason out. I never would have gotten to this reasoning without deducing the count from the other two formulas using a computer algebra system.

9.1.4: symmetries of Riemann

9.1.5: curvature of planes, cylinders, spheres

9.1.6: curvature of an ellipsoid of revolution

9.1.7: curvature of an elliptical paraboloid

9.1.8: curvature of surface of revolution

9.1.9: curvature in cylindrical coordinates

9.1.10: curvature from integrability conditions

9.1.11: Jacobi and Bianchi identities

9.3.1: frame components of Riemann

9.4.1: meridians on ellipsoid of revolution

9.4.2: minimum convergence length on the ellipsoid of revolution

9.4.3: minimum convergence length on the torus

9.4.4: minimum convergence length on the catenoid surfaces

9.4.5: curvature of elliptic paraboloid

9.4.6: curvature of pseudospheres compared to corresponding hyperboloids in \mathbb{R}^3

Chapter 10

10.3.1: Taylor approximation to the sphere

10.3.2: monkeysaddle degeneracy

10.3.3: Gaussian curvature of a surface of revolution

10.3.4: Gaussian curvature of a helicoid

10.4.1: extrinsic curvature as a connection component

10.4.2: decomposition of curvature on a family of surfaces

10.4.3: spaces of constant curvature

10.4.4: spherical coordinates with a signature change

10.4.5: extrinsic curvature of pseudospheres

10.4.6: curvature of hyperbolic paraboloid

10.5.1: tilted cavatappo surface curvature

10.6.1: shape operator insensitive to length of normal

10.6.2: geodesics on the sphere and ellipsoid

10.6.3: geodesics on the approximate gyroid

10.6.4: rotation of the surface normal compared to the Frenet-Serret frame

10.6.5: relative rotation of Frenet-Serret frame and surface adapted frame

10.6.6: Gaussian curvature of implicitly defined surface

Chapter 11

11.2.1: integration in the plane over a parallelogram

11.2.2: snow cone centroid integration

11.3.1: differential of surface area

11.4.1: surface integral on a sphere

11.4.2: contraction of 2-form with 2-vector

11.4.3: integration over a triangular surface in space

11.6.1: integration of a 2-form in 3-spacetime

11.7.1: exterior derivatives in cylindrical coordinates

11.7.2: exterior derivative in a frame, curvature 2-form

11.7.3: curvature of the 3-sphere

11.7.4: $SU(2)$ gauge derivative

11.8.1: tensor-valued differential forms

11.8.2: grad curl div

11.8.3: Maxwell's equations in differential form

11.8.4: vector potential for electromagnetic field

11.8.5: codifferential versus divergence sign

b) Recall $*^{-1}S = (-1)^{M+p(n-p)}*S$ for a p -form S . Then for a p -form β we have

$$\begin{aligned}
 \delta\beta &= (-1)^{p+M} *^{-1} * d \underbrace{* \beta}_{\substack{n-p \\ n-p+1}} \\
 &\quad \underbrace{\hspace{1.5cm}}_{n-(n-p+1)=p-1} \\
 &= (-1)^Q * d^* \beta \\
 &= (-1)^{n(p-1)+1} * d^* \beta,
 \end{aligned}$$

where

$$\begin{aligned}
 Q &= p + M + M + (p-1)(n - (p-1)) = 2M + p + n(p-1) - (p-1)^2 \\
 &= n(p-1) + 2M + p - p^2 + 2p - 1 = n(p-1) - 1 + [2M + 2p - p(p-1)] \\
 &= n(p-1) + 1,
 \end{aligned}$$

since the terms in square brackets are even and $(-1)^{-1} = (-1)^1$.

11.8.6: Maxwell's equations and the codifferential

11.8.7: spacetime deRham cohomology

11.9.1: snow cone surface integral

11.11.1: paraboloidal solid integration

11.11.2: wedge of cylinder integration

11.11.3: unit ball integration

11.12.1: 3-sphere exterior derivatives

11.12.2: integration between 2-spheres

Chapter 12: final exam worked

1) Eliminating ν from the first equation $\rho = \mu\nu \rightarrow \nu = \rho/\mu$ leads to a quadratic equation for μ when substituted into the second equation

$$\begin{aligned} z &= \frac{1}{2}(\mu^2 - \nu^2) = \frac{1}{2}(\mu^2 - \rho^2/\mu^2) \\ \mu^4 - 2z\mu^2 - \rho^2 &= 0 \\ \mu^2 &= \frac{1}{2}(2z \pm \sqrt{4z^2 + 4\rho^2}) = z + \sqrt{z^2 + \rho^2} \geq 0 \\ \mu &= \sqrt{z + \sqrt{z^2 + \rho^2}} \quad (\mu \geq 0) \end{aligned}$$

or vice versa $\mu = \rho/\nu$ leads to

$$\begin{aligned} z &= \frac{1}{2}(\rho^2/\nu^2 - \nu^2) \\ 2z + \nu^2 &= \rho^2/\nu^2 \\ \nu^4 + 2z\nu^2 - \rho^2 &= 0 \\ \nu^2 &= \frac{1}{2}(2z \pm \sqrt{4z^2 + 4\rho^2}) = -z + \sqrt{z^2 + \rho^2} \\ \nu &= \sqrt{-z + \sqrt{z^2 + \rho^2}} \quad (\nu \geq 0) \end{aligned}$$

Note

$$\underbrace{z^2 + \rho^2}_{z^2 + x^2 + y^2 = r^2} = \frac{1}{4}(4\mu^2\nu^2 + \mu^4 - 2\mu^2\nu^2 + \nu^4) = \left(\frac{\mu^2 + \nu^2}{2}\right)^2$$

so

$$r = \sqrt{z^2 + \rho^2} = \frac{\mu^2 + \nu^2}{2} \quad \text{or} \quad \mu^2 + \nu^2 = \frac{r}{2}.$$

2) Evaluate the differentials and read off the partial derivative coefficients to arrange in the Jacobian matrix

$$\begin{aligned} x &= \mu\nu \cos \phi & dx &= \nu \cos \phi d\mu + \mu \cos \phi d\nu - \mu\nu \sin \phi d\phi \\ y &= \mu\nu \sin \phi & dy &= \nu \sin \phi d\mu + \mu \sin \phi d\nu - \mu\nu \cos \phi d\phi \\ z &= \frac{1}{2}(\mu^2 - \nu^2) & dz &= \mu d\mu - \nu d\nu \end{aligned}$$

$$\underline{A}^{-1}(\bar{x}) = \begin{pmatrix} \nu \cos \phi & \mu \cos \phi & -\mu\nu \sin \phi \\ \nu \sin \phi & \mu \sin \phi & \mu\nu \cos \phi \\ \mu & -\nu & 0 \end{pmatrix}$$

3) The self-dot products of the orthogonal columns give the diagonal metric components

$$g = \delta_{ij} dx^i \otimes dx^j = (\mu^2 + \nu^2)[d\mu \otimes d\mu + d\nu \otimes d\nu] + \mu^2\nu^2 d\phi d\phi$$

4) Row expansion on last row of the determinant of the Jacobian matrix

$$\begin{aligned}\det \underline{A}^{-1}(\bar{x}) &= \mu \begin{vmatrix} \mu \cos \phi & -\mu\nu \sin \phi \\ \mu \sin \phi & \mu\nu \cos \phi \end{vmatrix} - (-\nu) \begin{vmatrix} \nu \cos \phi & -\mu\nu \sin \phi \\ \nu \sin \phi & \mu\nu \cos \phi \end{vmatrix} \\ &= \mu(\mu^2\nu) + \nu(\mu\nu^2) = \mu\nu(\mu^2 + \nu^2) \geq 0\end{aligned}$$

shows that the new coordinate frame is positively oriented and the unit volume 3-form is

$$\eta = \underbrace{\mu\nu(\mu^2 + \nu^2)}_{(g_{\mu\mu}g_{\nu\nu}g_{\phi\phi})^{1/2}} d\mu \wedge d\nu \wedge d\phi$$

5) The normalized frame and dual frame are

$$\begin{aligned}e_{\hat{\mu}} &= \frac{1}{(\mu^2 + \nu^2)^{1/2}} \frac{\partial}{\partial \mu} \quad , \quad e_{\hat{\nu}} = \frac{1}{(\mu^2 + \nu^2)^{1/2}} \frac{\partial}{\partial \nu} \quad , \quad e_{\hat{\phi}} = \frac{1}{(\mu\nu)} \frac{\partial}{\partial \phi} \\ \omega^{\hat{\mu}} &= (\mu^2 + \nu^2)^{1/2} d\mu \quad , \quad \omega^{\hat{\nu}} = (\mu^2 + \nu^2)^{1/2} d\nu \quad , \quad \omega^{\hat{\phi}} = \mu\nu d\phi\end{aligned}$$

(note $\eta = \omega^{\hat{\mu}} \wedge \omega^{\hat{\nu}} \wedge \omega^{\hat{\phi}} = \omega^{\hat{\mu}\hat{\nu}\hat{\phi}}$) so normalizing the columns of the Jacobian matrix yields the orthogonal matrix

$$\underline{A}^{-1}(\bar{x}) = \begin{pmatrix} \frac{\nu}{(\mu^2 + \nu^2)^{1/2}} \cos \phi & \frac{\mu}{(\mu^2 + \nu^2)^{1/2}} \cos \phi & -\sin \phi \\ \frac{\nu}{(\mu^2 + \nu^2)^{1/2}} \sin \phi & \frac{\mu}{(\mu^2 + \nu^2)^{1/2}} \sin \phi & \cos \phi \\ \frac{\mu}{(\mu^2 + \nu^2)^{1/2}} & \frac{-\nu}{(\mu^2 + \nu^2)^{1/2}} & 0 \end{pmatrix}$$

and its inverse is its transpose

$$\underline{A}(\bar{x}) = \begin{pmatrix} \frac{\nu}{(\mu^2 + \nu^2)^{1/2}} \cos \phi & \frac{\nu}{(\mu^2 + \nu^2)^{1/2}} \sin \phi & \frac{\mu}{(\mu^2 + \nu^2)^{1/2}} \\ \frac{\mu}{(\mu^2 + \nu^2)^{1/2}} \cos \phi & \frac{\mu}{(\mu^2 + \nu^2)^{1/2}} \sin \phi & \frac{-\nu}{(\mu^2 + \nu^2)^{1/2}} \\ -\sin \phi & \cos \phi & 0 \end{pmatrix}$$

Multiplying these rows by the corresponding normalization factors used above for the columns gives the inverse Jacobian matrix expressed in terms of the new coordinates

$$\underline{A}(\bar{x}) = \begin{pmatrix} \frac{\nu}{\mu^2 + \nu^2} \cos \phi & \frac{\nu}{\mu^2 + \nu^2} \sin \phi & \frac{\mu}{\mu^2 + \nu^2} \\ \frac{\mu}{\mu^2 + \nu^2} \cos \phi & \frac{\mu}{\mu^2 + \nu^2} \sin \phi & \frac{-\nu}{\mu^2 + \nu^2} \\ \frac{-1}{\mu\nu} \sin \phi & \frac{1}{\mu\nu} \cos \phi & 0 \end{pmatrix} = \left(\frac{\partial \bar{X}^i}{\partial x^j}(x(\bar{x})) \right)$$

6) Now recall that $2\sqrt{z^2 + \rho^2} = \mu^2 + \nu^2$ which can be used to simplify the differentials of the new coordinates, so for example,

$$\begin{aligned}
 \mu &= (z + (z^2 + \rho^2)^{1/2})^{1/2} \\
 d\mu &= \frac{1}{2(z^2 + \rho^2)^{1/2}} \left[dz + \frac{1}{2} \frac{[2zdz + 2xdx + 2ydy]}{(z^2 + \rho^2)^{1/2}} \right] \\
 &= \frac{1}{2\mu} \left[\frac{(\mu^2 + \nu^2)dz + [2zdz + 2xdx + 2ydy]}{\mu^2 + \nu^2} \right] \\
 &= \frac{1}{2\mu} \left[\frac{2xdx + 2ydy + 2\mu^2 dz}{\mu^2 + \nu^2} \right] = \frac{xdx + ydy + \mu^2 dz}{\mu(\mu^2 + \nu^2)} \\
 &= \frac{\mu\nu \cos \phi dx + \mu\nu \sin \phi dy + \mu^2 dz}{\mu(\mu^2 + \nu^2)} \\
 &= \frac{\nu \cos \phi dx + \nu \sin \phi dy + \mu^2 dz}{\mu^2 + \nu^2}
 \end{aligned}$$

and these components of $d\mu$ are exactly the first rows of $\underline{A}(\bar{X})$. The second row is handled similarly. The last row comes from

$$d\phi = \frac{-\sin \phi}{\rho} dx + \frac{\cos \phi}{\rho} dy$$

which is the cylindrical coordinate result with ρ then replaced by $\mu\nu$.

7) We compute the Lie brackets of the orthonormal frame vectors, recalling $[X, Y] = XY - YX$

$$\begin{aligned}
 [e_{\hat{\nu}}, e_{\hat{\phi}}] &= \left[(\mu^2 + \nu^2)^{-1/2} \frac{\partial}{\partial \nu}, (\mu\nu)^{-1} \frac{\partial}{\partial \phi} \right] = (\mu^2 + \nu^2)^{-1/2} \frac{1}{\mu} \left(\frac{-1}{\nu^2} \right) \frac{\partial}{\partial \phi} \\
 &= \frac{-1}{(\mu^2 + \nu^2)^{1/2} \nu} e_{\hat{\phi}} \quad \rightarrow \quad C^{\hat{\phi}}_{\hat{\nu}\hat{\phi}} = \frac{-1}{(\mu^2 + \nu^2)^{1/2} \nu}
 \end{aligned}$$

$$\begin{aligned}
 [e_{\hat{\mu}}, e_{\hat{\phi}}] &= \left[(\mu^2 + \nu^2)^{-1/2} \frac{\partial}{\partial \mu}, (\mu\nu)^{-1} \frac{\partial}{\partial \phi} \right] = (\mu^2 + \nu^2)^{-1/2} \frac{1}{\nu} \left(\frac{-1}{\mu^2} \right) \frac{\partial}{\partial \phi} \\
 &= \frac{-1}{(\mu^2 + \nu^2)^{1/2} \mu} e_{\hat{\phi}} \quad \rightarrow \quad C^{\hat{\phi}}_{\hat{\mu}\hat{\phi}} = \frac{-1}{(\mu^2 + \nu^2)^{1/2} \mu}
 \end{aligned}$$

$$\begin{aligned}
[e_{\hat{\mu}}, e_{\hat{\nu}}] &= \left[(\mu^2 + \nu^2)^{-1/2} \frac{\partial}{\partial \mu}, (\mu^2 + \nu^2)^{-1/2} \frac{\partial}{\partial \nu} \right] \\
&= (\mu^2 + \nu^2)^{-1/2} \frac{\partial}{\partial \mu} (\ln(\mu^2 + \nu^2)^{-1/2}) \frac{\partial}{\partial \nu} - (\mu^2 + \nu^2)^{-1/2} \frac{\partial}{\partial \nu} (\ln(\mu^2 + \nu^2)^{-1/2}) \frac{\partial}{\partial \mu} \\
&= \frac{-1}{2} (\mu^2 + \nu^2)^{-1} \frac{2\mu}{(\mu^2 + \nu^2)} \frac{\partial}{\partial \nu} + \frac{1}{2} (\mu^2 + \nu^2)^{-1} \frac{2\nu}{(\mu^2 + \nu^2)} \frac{\partial}{\partial \mu} \\
&= \frac{1}{(\mu^2 + \nu^2)^2} [-\mu e_{\hat{\nu}} + \nu e_{\hat{\mu}}] \rightarrow \\
C^{\hat{\mu}}_{\hat{\mu}\hat{\nu}} &= \frac{\nu}{(\mu^2 + \nu^2)^2}, \quad C^{\hat{\nu}}_{\hat{\mu}\hat{\nu}} = \frac{-\mu}{(\mu^2 + \nu^2)^2}
\end{aligned}$$

8) For more manageable matrix notation, introduce the abbreviations $2r = \mu^2 + \nu^2$, $\rho = \mu\nu$, $S = \sin \phi$ and $C = \cos \phi$ a)

$$\begin{aligned}
\bar{\omega} = \underline{A} d\underline{A}^{-1} &= \begin{pmatrix} \frac{\nu}{(2r)^2} C & \frac{\nu}{(2r)^2} S & \frac{\mu}{(2r)^2} \\ \frac{\mu}{(2r)^2} C & \frac{\mu}{(2r)^2} S & \frac{-\nu}{(2r)^2} \\ \frac{-1}{\rho} S & \frac{1}{\rho} C & 0 \end{pmatrix} d \begin{pmatrix} \nu C & \mu C & -\mu\nu S \\ \nu S & \mu S & \mu\nu C \\ \mu & -\nu & 0 \end{pmatrix} \\
&\quad \begin{pmatrix} 0 & C & -\nu S \\ 0 & S & \nu C \\ 1 & 0 & 0 \end{pmatrix} d\mu + \begin{pmatrix} C & 0 & -\mu S \\ S & 0 & \mu C \\ 0 & -1 & 0 \end{pmatrix} d\nu + \begin{pmatrix} -\nu S & -\nu S & -\mu\nu C \\ \nu C & \mu C & -\mu\nu S \\ 0 & 0 & 0 \end{pmatrix} d\phi \\
&= \begin{pmatrix} \frac{\mu}{(2r)^2} & \frac{\nu}{(2r)^2} & 0 \\ \frac{-\nu}{(2r)^2} & \frac{\mu}{(2r)^2} & 0 \\ 0 & 0 & \frac{1}{\mu} \end{pmatrix} d\mu + \begin{pmatrix} \frac{\nu}{(2r)^2} & \frac{-\mu}{(2r)^2} & 0 \\ \frac{\mu}{(2r)^2} & \frac{\nu}{(2r)^2} & 0 \\ 0 & 0 & \frac{1}{\nu} \end{pmatrix} d\nu + \begin{pmatrix} 0 & 0 & \frac{-\mu\nu^2}{(2r)^2} \\ 0 & 0 & \frac{-\mu^2\nu}{(2r)^2} \\ \frac{\nu}{\rho} & \frac{\mu}{\rho} & 0 \end{pmatrix} d\phi \\
&= \begin{pmatrix} \bar{\Gamma}^{\mu}_{\mu\mu} & \bar{\Gamma}^{\mu}_{\mu\nu} & 0 \\ \bar{\Gamma}^{\nu}_{\mu\mu} & \bar{\Gamma}^{\nu}_{\mu\nu} & 0 \\ 0 & 0 & \bar{\Gamma}^{\phi}_{\mu\phi} \end{pmatrix} d\mu + \begin{pmatrix} \bar{\Gamma}^{\mu}_{\nu\mu} & \bar{\Gamma}^{\mu}_{\nu\nu} & 0 \\ \bar{\Gamma}^{\nu}_{\nu\mu} & \bar{\Gamma}^{\nu}_{\nu\nu} & 0 \\ 0 & 0 & \bar{\Gamma}^{\phi}_{\nu\phi} \end{pmatrix} d\nu + \begin{pmatrix} 0 & 0 & \bar{\Gamma}^{\mu}_{\phi\phi} \\ 0 & 0 & \bar{\Gamma}^{\nu}_{\phi\phi} \\ \bar{\Gamma}^{\mu}_{\phi\mu} & \bar{\Gamma}^{\nu}_{\phi\nu} & 0 \end{pmatrix} d\phi
\end{aligned}$$

Identifying the nonzero entries gives the nonzero coordinate components of the connection.

9a) Alternatively by explicit differentiation

$$\Gamma_{ijk} = \frac{1}{2} (g_{ij,k} - g_{jk,i} + g_{ki,j}) \quad \Gamma^i_{jk} = g^{ii} \Gamma_{ijk} = (g_{ii})^{-1} \Gamma_{ijk} \quad (\text{orthogonal coordinate})$$

$$g_{\mu\mu} = g_{\nu\nu} = \mu^2 + \nu^2 \quad , \quad g_{\phi\phi} = \mu\nu$$

$$\begin{aligned} \Gamma_{\mu\mu\mu} &= \frac{1}{2}(g_{\mu\mu,\mu} - g_{\mu\mu,\mu} + g_{\mu\mu,\mu}) = \mu & \Gamma^\mu_{\mu\mu} &= \frac{\mu}{\mu^2 + \nu^2} = [\omega^\mu_\mu]_\mu \\ \Gamma_{\mu\mu\nu} &= \frac{1}{2}(g_{\mu\mu,\nu} - g_{\mu\nu,\mu} + g_{\nu\mu,\mu}) = \nu & \Gamma^\mu_{\mu\nu} &= \frac{\nu}{\mu^2 + \nu^2} = [\omega^\mu_\nu]_\mu \\ \Gamma_{\nu\mu\mu} &= \frac{1}{2}(g_{\nu\mu,\mu} - g_{\mu\mu,\nu} + g_{\mu\nu,\mu}) = -\nu & \Gamma^\nu_{\mu\mu} &= \frac{-\nu}{\mu^2 + \nu^2} = [\omega^\nu_\mu]_\mu \\ \Gamma_{\nu\mu\nu} &= \frac{1}{2}(g_{\nu\mu,\nu} - g_{\mu\nu,\nu} + g_{\nu\nu,\mu}) = \mu & \Gamma^\nu_{\mu\nu} &= \frac{\mu}{\mu^2 + \nu^2} = [\omega^\nu_\nu]_\mu \end{aligned}$$

$$\begin{aligned} \Gamma_{\phi\phi\mu} &= \frac{1}{2}(g_{\phi\phi,\mu} - g_{\phi\mu,\phi} + g_{\phi\mu,\phi}) = \mu\nu^2 & \Gamma^\phi_{\phi\mu} &= \frac{1}{\mu} = \frac{\nu}{\rho} \\ \Gamma_{\phi\phi\nu} &= \frac{1}{2}(g_{\phi\phi,\nu} - g_{\phi\nu,\phi} + g_{\phi\nu,\phi}) = \nu\mu^2 & \Gamma^\phi_{\phi\nu} &= \frac{1}{\nu} = \frac{\mu}{\rho} \end{aligned}$$

$$\begin{aligned} \Gamma_{\phi\mu\phi} &= \frac{1}{2}(g_{\phi\mu,\phi} - g_{\mu\phi,\phi} + g_{\phi\phi,\mu}) = \mu\nu^2 & \Gamma^\phi_{\mu\phi} &= \frac{1}{\mu} \\ \Gamma_{\phi\nu\phi} &= \frac{1}{2}(g_{\phi\nu,\phi} - g_{\nu\phi,\phi} + g_{\phi\phi,\nu}) = \mu^2\nu & \Gamma^\phi_{\nu\phi} &= \frac{1}{\nu} \end{aligned}$$

8b)

$$P \equiv \frac{\mu}{(\mu^2 + \nu^2)^{1/2}} \quad , \quad Q \equiv \frac{\nu}{(\mu^2 + \nu^2)^{1/2}} \quad , \quad P^2 + Q^2 = 1$$

$$\begin{aligned} \frac{\partial P}{\partial \mu} &= \frac{(\mu^2 + \nu^2)^{1/2} \cdot 1 - \mu \cdot \frac{1}{2} \frac{2\mu}{(\mu^2 + \nu^2)^{1/2}}}{(\mu^2 + \nu^2)} = \frac{(\mu^2 + \nu^2) - \mu^2}{(\mu^2 + \nu^2)^{3/2}} = \frac{\nu^2}{(\mu^2 + \nu^2)^{3/2}} \\ \frac{\partial P}{\partial \nu} &= \frac{(\mu^2 + \nu^2)^{1/2} \cdot 0 - \mu \cdot \frac{1}{2} \frac{2\nu}{(\mu^2 + \nu^2)^{1/2}}}{(\mu^2 + \nu^2)} = \frac{-\mu\nu}{(\mu^2 + \nu^2)^{3/2}} \\ \frac{\partial Q}{\partial \mu} &= \frac{(\mu^2 + \nu^2)^{1/2} \cdot 0 - \nu \cdot \frac{1}{2} \frac{2\mu}{(\mu^2 + \nu^2)^{1/2}}}{(\mu^2 + \nu^2)} = \frac{-\mu\nu}{(\mu^2 + \nu^2)^{3/2}} \\ \frac{\partial Q}{\partial \nu} &= \frac{(\mu^2 + \nu^2)^{1/2} \cdot 1 - \nu \cdot \frac{1}{2} \frac{2\nu}{(\mu^2 + \nu^2)^{1/2}}}{(\mu^2 + \nu^2)} = \frac{(\mu^2 + \nu^2) - \nu^2}{(\mu^2 + \nu^2)^{3/2}} = \frac{\mu^2}{(\mu^2 + \nu^2)^{3/2}} \end{aligned}$$

$$\hat{\omega} = \underline{\mathcal{A}} \, d\underline{\mathcal{A}}^{-1}$$

$$\begin{aligned} &= \begin{pmatrix} QC & QS & P \\ PC & PS & -Q \\ -S & C & 0 \end{pmatrix} \underbrace{d \begin{pmatrix} QC & PC & -S \\ QS & PS & C \\ P & -Q & 0 \end{pmatrix}} \\ &\begin{pmatrix} -\mu\nu C & \nu^2 C & 0 \\ -\mu\nu S & \nu^2 S & 0 \\ \nu^2 & \mu\nu & 0 \end{pmatrix} \frac{d\mu}{(\mu^2 + \nu^2)^{3/2}} + \begin{pmatrix} \mu^2 C & -\mu\nu C & 0 \\ \mu^2 S & -\mu\nu S & 0 \\ -\mu\nu & -\mu^2 & 0 \end{pmatrix} \frac{d\nu}{(\mu^2 + \nu^2)^{3/2}} + \begin{pmatrix} -QS & -PS & -C \\ QC & PC & -S \\ 0 & 0 & 0 \end{pmatrix} d\phi \end{aligned}$$

$$= \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{Q}{(\mu^2 + \nu^2)} [(\mu^2 + \nu^2)^{1/2} d\mu] + \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{P}{(\mu^2 + \nu^2)} [(\mu^2 + \nu^2)^{1/2} d\nu] \\ + \begin{pmatrix} 0 & 0 & -Q \\ 0 & 0 & -P \\ Q & P & 0 \end{pmatrix} \frac{1}{\mu\nu} [\mu\nu d\phi] = \left(\Gamma^{\hat{i}}_{\hat{k}\hat{j}} \right) \hat{\omega}^{\hat{k}}$$

$$\Gamma^{\hat{\mu}}_{\hat{\mu}\hat{\nu}} = \frac{\nu}{(\mu^2 + \nu^2)^{3/2}} = -\Gamma^{\hat{\nu}}_{\hat{\mu}\hat{\mu}} \quad , \quad \Gamma^{\hat{\nu}}_{\hat{\nu}\hat{\nu}} = -\frac{\mu}{(\mu^2 + \nu^2)^{3/2}} = -\Gamma^{\hat{\nu}}_{\hat{\nu}\hat{\mu}} \\ \Gamma^{\hat{\mu}}_{\hat{\phi}\hat{\phi}} = \frac{-1}{\mu(\mu^2 + \nu^2)^{1/2}} = -\Gamma^{\hat{\phi}}_{\hat{\mu}\hat{\phi}} \quad , \quad \Gamma^{\hat{\nu}}_{\hat{\phi}\hat{\phi}} = -\frac{1}{\nu(\mu^2 + \nu^2)^{1/2}} = -\Gamma^{\hat{\phi}}_{\hat{\phi}\hat{\nu}}$$

9b) or by direct evaluation of the component formulas in an orthonormal frame

$$\Gamma^{\hat{i}}_{\hat{j}\hat{k}} = \Gamma_{\hat{i}\hat{j}\hat{k}} = \frac{1}{2}(C_{\hat{i}\hat{j}\hat{k}} - C_{\hat{j}\hat{k}\hat{i}} + C_{\hat{k}\hat{i}\hat{j}}) \\ C_{\hat{\phi}\hat{\mu}\hat{\phi}} = \frac{-1}{\mu(\mu^2 + \nu^2)^{1/2}} \quad , \quad C_{\hat{\phi}\hat{\nu}\hat{\phi}} = \frac{-1}{\nu(\mu^2 + \nu^2)^{1/2}} \\ C_{\hat{\mu}\hat{\nu}\hat{\mu}} = \frac{-\nu}{(\mu^2 + \nu^2)^{3/2}} \quad , \quad C_{\hat{\nu}\hat{\mu}\hat{\nu}} = \frac{-\mu}{(\mu^2 + \nu^2)^{3/2}} \\ \Gamma^{\hat{\mu}}_{\hat{\mu}\hat{\nu}} = \frac{1}{2}(C_{\hat{\mu}\hat{\mu}\hat{\nu}} - C_{\hat{\mu}\hat{\nu}\hat{\mu}} + C_{\hat{\nu}\hat{\mu}\hat{\mu}}) = C_{\hat{\mu}\hat{\mu}\hat{\nu}} = \frac{\nu}{(\mu^2 + \nu^2)^{3/2}} \\ \Gamma^{\hat{\mu}}_{\hat{\nu}\hat{\nu}} = \frac{1}{2}(C_{\hat{\mu}\hat{\nu}\hat{\nu}} - C_{\hat{\nu}\hat{\nu}\hat{\mu}} + C_{\hat{\nu}\hat{\mu}\hat{\nu}}) = C_{\hat{\nu}\hat{\mu}\hat{\nu}} = \frac{-\mu}{(\mu^2 + \nu^2)^{3/2}} \\ \Gamma^{\hat{\mu}}_{\hat{\phi}\hat{\phi}} = \frac{1}{2}(C_{\hat{\mu}\hat{\phi}\hat{\phi}} - C_{\hat{\phi}\hat{\phi}\hat{\mu}} + C_{\hat{\phi}\hat{\mu}\hat{\phi}}) = C_{\hat{\phi}\hat{\mu}\hat{\phi}} = \frac{-1}{\mu(\mu^2 + \nu^2)^{1/2}} \\ \Gamma^{\hat{\nu}}_{\hat{\phi}\hat{\phi}} = \frac{1}{2}(C_{\hat{\nu}\hat{\phi}\hat{\phi}} - C_{\hat{\phi}\hat{\phi}\hat{\nu}} + C_{\hat{\phi}\hat{\nu}\hat{\phi}}) = C_{\hat{\phi}\hat{\nu}\hat{\phi}} = \frac{-1}{\nu(\mu^2 + \nu^2)^{1/2}}$$

11) For the 2-surface $\nu = \nu_0$

$$^{(2)}g = (\mu^2 + \nu_0^2)d\mu \otimes d\mu + \mu^2\nu_0 d\phi \otimes d\phi, \quad , \quad ^{(2)}\eta = \mu\nu_0(\mu^2 + \nu_0^2)^{1/2}d\mu \wedge d\phi$$

examining the components of the connection for the 3-metric, only the components $\bar{\Gamma}^i_{jk}$ evaluated above with no ν indices are relevant here:

$$^{(2)}\Gamma^{\mu}_{\mu\mu} = \frac{\mu}{\mu^2 + \nu_0^2} \quad , \quad ^{(2)}\Gamma^{\phi}_{\mu\phi} = \frac{1}{\mu} = ^{(2)}\Gamma^{\phi}_{\phi\mu} \quad , \quad ^{(2)}\Gamma^{\mu}_{\phi\phi} = \frac{-\mu\nu_0^2}{\mu^2 + \nu_0^2}.$$

To get $^{(2)}\underline{\omega}$, just delete the second row and column of $\hat{\omega}_{\mu}d\mu + \hat{\omega}_{\phi}d\phi$, while setting $\nu = \nu_0$:

$$^{(2)}\underline{\omega} = \begin{pmatrix} \frac{\mu}{\mu^2 + \nu_0^2} & 0 \\ 0 & \mu^{-1} \end{pmatrix} d\mu + \begin{pmatrix} 0 & \frac{-\mu\nu_0^2}{\mu^2 + \nu_0^2} \\ \mu^{-1} & 0 \end{pmatrix} d\phi = \begin{pmatrix} \frac{\mu}{\mu^2 + \nu_0^2}d\mu & \frac{-\mu\nu_0^2}{\mu^2 + \nu_0^2}d\phi \\ \mu^{-1}d\phi & \mu^{-1}d\mu \end{pmatrix}$$

$$d\left(\frac{\mu}{\mu^2 + \nu_0^2}\right) = \frac{(\mu^2 + \nu_0^2) - \mu(2\mu)}{(\mu^2 + \nu_0^2)^2} = \frac{\nu_0^2 - \mu^2}{(\mu^2 + \nu_0^2)^2} \quad d(\mu^{-1}) = -\mu^{-2}d\mu.$$

$$d^{(2)}\underline{\omega} = \begin{pmatrix} 0 & \frac{\nu_0^2(\mu^2 - \nu_0^2)}{\mu^2 + \nu_0^2} d\mu d\phi \\ -\mu^{-2}d\mu d\phi & 0 \end{pmatrix} \quad (d[f(\mu)d\mu] = f'(\mu)d\mu \wedge d\mu = 0)$$

$$\begin{aligned} {}^{(2)}\underline{\omega} \wedge {}^{(2)}\underline{\omega} &= \begin{pmatrix} \frac{\mu}{\mu^2 + \nu_0^2} d\mu & \frac{-\mu\nu_0^2}{\mu^2 + \nu_0^2} d\phi \\ \mu^{-1}d\phi & \mu^{-1}d\mu \end{pmatrix} \wedge \begin{pmatrix} \frac{\mu}{\mu^2 + \nu_0^2} d\mu & \frac{-\mu\nu_0^2}{\mu^2 + \nu_0^2} d\phi \\ \mu^{-1}d\phi & \mu^{-1}d\mu \end{pmatrix} \\ &= \begin{pmatrix} 0 & [\frac{-\mu^2\nu_0^2}{(\mu^2 + \nu_0^2)^2} + \frac{\nu_0^2}{\mu^2 + \nu_0^2}]d\mu \wedge d\phi \\ [\frac{-1}{\mu^2 + \nu_0^2} + \frac{1}{\mu^2}]d\mu \wedge d\phi & 0 \end{pmatrix} \end{aligned}$$

$${}^{(2)}\underline{\Omega} = d {}^{(2)}\underline{\omega} + {}^{(2)}\underline{\omega} \wedge {}^{(2)}\underline{\omega} = \begin{pmatrix} 0 & \frac{\mu^2\nu_0^2}{(\mu^2 + \nu_0^2)^2} d\mu \wedge d\phi \\ \frac{-1}{\mu^2 + \nu_0^2} d\mu \wedge d\phi & 0 \end{pmatrix}$$

$${}^{(2)}R^\phi_{\mu\mu\phi} = \frac{-1}{\mu^2 + \nu_0^2} \quad {}^{(2)}R^\mu_{\phi\mu\phi} = \frac{\mu^2\nu_0^2}{(\mu^2 + \nu_0^2)^2}$$

$${}^{(2)}R_{\phi\mu\mu\phi} = \frac{-\mu^2\nu_0^2}{\mu^2 + \nu_0^2} \quad {}^{(2)}R_{\mu\phi\mu\phi} = \frac{\mu^2\nu_0^2}{(\mu^2 + \nu_0^2)} = -{}^{(2)}R_{\phi\mu\mu\phi}.$$

12)

$${}^{(2)}R^{\hat{\mu}}_{\hat{\phi}\hat{\mu}\hat{\phi}} = (g_{\phi\phi})^{-1} {}^{(2)}R^\mu_{\phi\mu\phi} = \frac{1}{(\mu^2 + \nu_0^2)^2} \xrightarrow{at \ \mu=0} \frac{1}{\nu_0^4}$$

12b)

$$\begin{aligned} z &= \frac{1}{2\nu_0^2}\rho^2 - \frac{1}{2}\nu_0^2 & \frac{dz}{\nu_0^2} &= \frac{\rho}{\nu_0^2} & \frac{d^2z}{d\rho^2} &= \frac{1}{\nu_0^2} \\ \mathcal{K} &= \frac{1/\nu_0^2}{[1 + \rho^2/\nu_0^4]^{3/2}} & \mathcal{K}|_{\rho=0} &= \frac{1}{\nu_0^2}. \end{aligned}$$

The relationship is

$${}^{(2)}R^{\hat{\mu}}_{\hat{\phi}\hat{\mu}\hat{\phi}}(\mu = 0) = [\mathcal{K}(\rho = 0)]^2$$

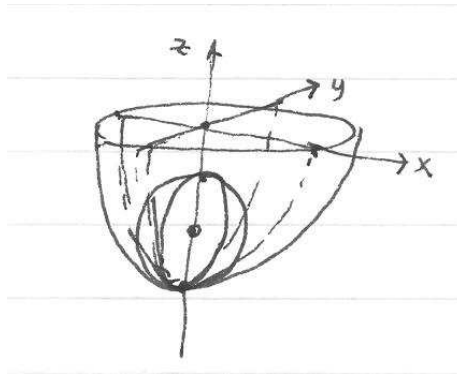


Figure E.8: The circles of best fit at the origin.

so

$${}^{(2)}R^{\alpha\beta}{}_{\gamma\delta} = {}^{(2)}\mathcal{K} \delta^{\alpha\beta}{}_{\gamma\delta} \quad , \quad {}^{(2)}\mathcal{K} = \frac{1}{(\mu^2 + \nu_0^2)^2}$$

At the vertex any orthogonal pair of vertical planes through the z -axis are the same and lead to 2 orthogonal osculating circles of best fit to those parabolas with the same radius and center.

The “2-curvature” ${}^{(2)}\mathcal{K} = R^{\hat{\mu}}{}_{\hat{\phi}\hat{\mu}\hat{\phi}}$ evaluated there there is just the product of these two “1-curvatures” $\mathcal{K} = \frac{1}{\nu_0^2}$. (See page 111. [fix])

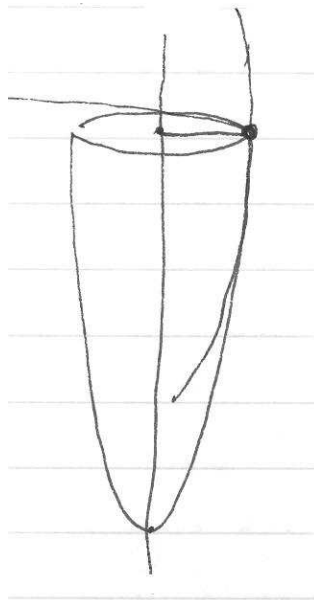


Figure E.9: The circles of best fit at away from the origin.

Suppose we consider points on this surface far from the vertex where the parallels nearly

coincide with the normal intersection of the surface by the normal plane of the meridians, namely for very large μ ($\mu \gg \nu_0$), then

$$\frac{\rho^2}{\nu_0^4} = \frac{\mu^2 \nu_0^2}{\nu_0^4} = \frac{\mu^2}{\nu_0^4} \gg 1$$

the horizontal cross-section (ϕ -coordinate circle of radius $\rho = \mu\nu_0$) is a circle whose connecting vector from its center to the point of tangency is almost along the normal direction. Together with the osculating circle of the parabola vertical cross-section, one obtains two nearly orthogonal circles of best fit.

$$\begin{aligned} \mathcal{K}_{\text{circle}} &= \frac{1}{\rho} = \frac{1}{\mu\nu_0} \\ \mathcal{K}_{\text{parabola}} &= \frac{1}{\nu_0^2 [1 + \frac{\rho^2}{\nu_0^4}]^{3/2}} \approx \frac{1}{\nu_0^2 (\frac{\rho^2}{\nu_0^4})^{3/2}} = \frac{\nu_0^4}{\rho^3} = \frac{\nu_0^4}{(\mu\nu_0)^3} = \frac{\nu_0}{\mu^3} \\ \left. \begin{aligned} \mathcal{K}_{\text{circle}} \mathcal{K}_{\text{parabola}} &\approx \left(\frac{1}{\mu\nu_0}\right) \left(\frac{\nu_0}{\mu^3}\right) = \frac{1}{\mu^4} \\ {}^{(2)}\mathcal{K} &= \frac{1}{(\mu^2 + \nu_0^2)^2} \approx \frac{1}{\mu^4} \end{aligned} \right\} \text{approximately equal.} \end{aligned}$$

13) Examine the geodesic equations for the coordinate curves in turn. One sees that the second covariant derivative of the μ coordinate lines is proportional to their tangent vector (geodesic condition), while the nonzero value for the ϕ coordinate lines is not, hence the latter are not geodesics.

μ lines:

$$\begin{aligned} \mu &= \lambda & \mu' &= 1 & \mu'' &= 0 \\ \phi &= \phi_0 & \phi' &= 0 & \phi'' &= 0 \end{aligned}$$

$$\left. \begin{aligned} \frac{D^2 \mu}{d^2 \lambda} &= \mu'' + \Gamma^\mu_{\alpha\beta} \frac{d\bar{x}^\alpha}{d\lambda} \frac{d\bar{x}^\beta}{d\lambda} \\ \frac{D^2 \phi}{d^2 \lambda} &= \phi'' + \Gamma^\phi_{\mu\mu} = 0 \end{aligned} \right\} \frac{D^2 \bar{x}^\alpha}{d^2 \lambda} = \left(\frac{\mu}{\mu^2 + \nu_0^2} \right) \frac{d\bar{x}^\alpha}{d\lambda}$$

ϕ lines:

$$\begin{aligned} \mu &= \mu_0 & \mu' &= 0 & \mu'' &= 0 \\ \phi &= \lambda & \phi' &= 1 & \phi'' &= 0 \end{aligned}$$

$$\left. \begin{aligned} \frac{D^2 \mu}{d^2 \lambda} &= \mu'' + \Gamma^\mu_{\phi\phi} \frac{d\phi}{d\lambda} \frac{d\phi}{d\lambda} = \frac{-\mu\nu_0^2}{\mu^2 + \nu^2} \\ \frac{D^2 \phi}{d^2 \lambda} &= \phi'' + \Gamma^\mu_{\phi\phi} \frac{d\phi}{d\lambda} \frac{d\phi}{d\lambda} = 0 \end{aligned} \right\} \frac{D^2 \bar{x}^\alpha}{d^2 \lambda} \text{ not proportional to } \frac{d\bar{x}^\alpha}{d\lambda}$$

14) Evaluate $C^{\hat{\phi}}_{\hat{\mu}\hat{\phi}}$ of part 7) at $\nu = \nu_0$:

$$C^{\hat{\phi}}_{\hat{\mu}\hat{\phi}} = \frac{-1}{(\mu^2 + \nu_0^2)^{1/2}\mu}$$

$$\begin{aligned} {}^{(2)}\Gamma^{\hat{\phi}}_{\hat{\phi}\hat{\mu}} &= \frac{1}{2}(C_{\hat{\phi}\hat{\phi}\hat{\mu}} - C_{\hat{\phi}\hat{\mu}\hat{\phi}} + C_{\hat{\mu}\hat{\phi}\hat{\phi}}) = -C_{\hat{\phi}\hat{\mu}\hat{\phi}} = \frac{1}{\mu(\mu^2 + \nu_0^2)^{1/2}} = -{}^{(2)}\Gamma^{\hat{\mu}}_{\hat{\phi}\hat{\phi}} \\ {}^{(2)}\Gamma^{\hat{\phi}}_{\hat{\mu}\hat{\phi}} &= \frac{1}{2}(C_{\hat{\phi}\hat{\mu}\hat{\phi}} - C_{\hat{\mu}\hat{\phi}\hat{\phi}} + C_{\hat{\phi}\hat{\phi}\hat{\mu}}) = 0 \end{aligned}$$

where of course this last line must vanish because of the antisymmetry in the outer indices.

$${}^{(2)}\underline{\hat{\omega}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \frac{1}{\mu(\mu^2 + \nu_0^2)^{1/2}} (\mu\nu_0 d\phi) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \frac{\nu_0}{(\mu^2 + \nu_0^2)^{1/2}} d\phi$$

$${}^{(2)}\underline{\hat{\omega}} \wedge {}^{(2)}\underline{\hat{\omega}} = 0$$

$$d {}^{(2)}\underline{\hat{\omega}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \left(\frac{-1}{2} \frac{2\nu_0\mu}{(\mu^2 + \nu_0^2)^{3/2}} \right) d\mu \wedge d\phi = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \frac{1}{(\mu^2 + \nu_0^2)^2} \omega^{\hat{\mu}\hat{\phi}}$$

$${}^{(2)}R^{\hat{\mu}}_{\hat{\phi}\hat{\mu}\hat{\phi}} = \frac{1}{(\mu^2 + \nu_0^2)^2}.$$

$$\left[{}^{(2)}\underline{\hat{\Omega}} = d {}^{(2)}\underline{\hat{\omega}} + {}^{(2)}\underline{\hat{\omega}} \wedge {}^{(2)}\underline{\hat{\omega}} \right]$$

since this formula is also valid for an orthonormal frame — or in general — in any frame.

14) b)

$$\begin{aligned} {}^{(2)}\nabla_{e_{\hat{\mu}}} e_{\hat{\mu}} &= \Gamma^{\hat{\mu}}_{\hat{\mu}\hat{\mu}} e_{\hat{\mu}} + \Gamma^{\hat{\phi}}_{\hat{\mu}\hat{\mu}} e_{\hat{\phi}} = 0 \\ {}^{(2)}\nabla_{e_{\hat{\mu}}} e_{\hat{\phi}} &= \Gamma^{\hat{\mu}}_{\hat{\mu}\hat{\phi}} e_{\hat{\mu}} + \Gamma^{\hat{\phi}}_{\hat{\mu}\hat{\phi}} e_{\hat{\phi}} = 0 \end{aligned}$$

so they are parallel transported along $e_{\hat{\phi}}$ which is the unit tangent to the μ coordiante lines. The first equality says $e_{\hat{\mu}}$ is auto parallel along μ and hence the curve must be a geodesic.

15) Yeah.

16) Using the matrix coordinate transformation

$$\bar{X}^i = A^i_j(\bar{x}) X^j(x(\bar{x}))$$

and recalling that $x^2 + y^2 + z^2 = r^2 = (\mu^2 + \nu^2)/2$, one finds

$$\begin{aligned} \begin{bmatrix} X^\mu \\ X^\nu \\ X^\phi \end{bmatrix} &= \begin{bmatrix} \frac{\nu C}{2r} & \frac{\nu S}{2r} & \frac{\mu}{2r} \\ \frac{\mu C}{2r} & \frac{\mu S}{2r} & \frac{-\nu}{2r} \\ \frac{-1}{\mu\nu} S & \frac{1}{\mu\nu} C & 0 \end{bmatrix} \begin{bmatrix} -\mu\nu S \\ \mu\nu C \\ \left(\frac{\mu^2 + \nu^2}{2}\right)^2 \end{bmatrix} = \begin{bmatrix} \frac{-\mu\nu^2 CS + \mu\nu^2 CS}{2r} + \frac{\mu}{\mu^2 + \nu^2} \left(\frac{\mu^2 + \nu^2}{2}\right)^2 \\ \frac{-\mu^2 \nu CS + \mu^2 \nu CS}{2r} - \frac{-\nu}{\mu^2 + \nu^2} \left(\frac{\mu^2 + \nu^2}{2}\right)^2 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{\mu}{4}(\mu^2 + \nu^2) \\ \frac{-\nu}{4}(\mu^2 + \nu^2) \\ 1 \end{bmatrix} \end{aligned}$$

$$\bar{X}_i = X_j(\mu) A^{-1j}_i(\bar{x})$$

$$\begin{aligned} [X_\mu \quad X_\nu \quad X_\phi] &= [-\mu\nu S \quad \mu\nu C \quad \left(\frac{\mu^2 + \nu^2}{2}\right)^2] \begin{pmatrix} \nu C & \mu C & -\mu\nu S \\ \nu S & \mu S & -\mu\nu C \\ \mu & -\nu & 0 \end{pmatrix} \\ &= [-\mu\nu^2 CS + \mu\nu^2 CS + \frac{\mu}{4}(\mu^2 + \nu^2)^2 \\ &\quad - \mu^2 \nu CS + \mu^2 \nu CS - \frac{\mu}{4}(\mu^2 + \nu^2)^2 \quad \mu^2 \nu^2] \\ &= \left[\frac{\mu}{4}(\mu^2 + \nu^2)^2 \quad -\frac{\mu}{4}(\mu^2 + \nu^2)^2 \quad \mu^2 \nu^2 \right] = [g_{\mu\mu} X^\mu \quad g_{\nu\nu} X^\nu \quad g_{\phi\phi} X^\phi] \end{aligned}$$

leading to

$$X^\flat = \frac{(\mu^2 + \nu^2)^2}{4}(\mu d\mu - \nu d\nu) + \mu^2 \nu^2 d\phi$$

which is consistent with the previous evaluation using the final equality above relating covariant and contravariant components.

Calculating the covariant derivative along μ

$$\begin{aligned} [\overline{\nabla_{e_\mu} X}]^i &= \bar{X}^i_{;\mu} = \bar{X}^i_{,\mu} + \bar{\Gamma}^i_{\mu j} \bar{X}^j \\ X^\mu_{;\mu} &= X^\mu_{,\mu} + \Gamma^\mu_{\mu\mu} X^\mu + \Gamma^\mu_{\mu\nu} X^\nu = \frac{1}{4}(3\mu^2 + \nu^2 + \mu^2 - \nu^2) = \mu^2 \\ X^\nu_{;\mu} &= X^\nu_{,\mu} + \Gamma^\nu_{\mu\mu} X^\mu + \Gamma^\nu_{\mu\nu} X^\nu = \frac{1}{4}(-2\mu\nu - \mu\nu - \mu\nu) = -\mu\nu \\ X^\phi_{;\mu} &= X^\phi_{,\mu} + \Gamma^\phi_{\mu\mu} X^\mu + \Gamma^\phi_{\mu\nu} X^\nu + \Gamma^\phi_{\mu\phi} X^\phi = \frac{1}{\mu} \\ \nabla_{e_\mu} X &= \mu^2 e_\mu - \mu\nu e_\nu + \mu^{-1} e_\phi. \end{aligned}$$

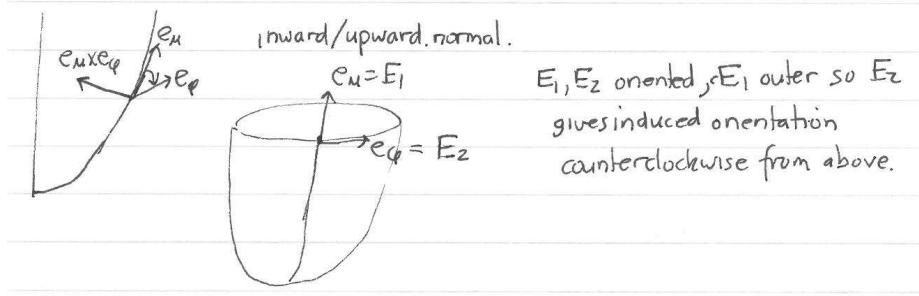


Figure E.10: The orientation of the boundary.

17) $\{E_1, E_2\}$ is oriented, E_1 points out so E_2 gives the induced orientation counterclockwise from above.

$$\partial\Sigma : \begin{cases} \mu = \nu_0 & \mu' = 0 \\ \nu = \nu_0 & \nu' = 0 \\ \phi = \lambda & \phi' = 1 \end{cases} \quad 0 \leq \lambda \leq 2\pi \quad \text{is an oriented parametrization of } \partial\Sigma.$$

$$\begin{aligned} \int_{\partial\Sigma} X^b &= \int_{\partial\Sigma} (\mu^2 + \nu^2/2)^2 (\mu d\mu - \nu d\nu) + \mu^2 \nu^2 d\phi \\ &= \int_0^{2\pi} \nu_0^4 [0 - 0 + d\lambda] = 2\pi \nu_0^4 \end{aligned}$$

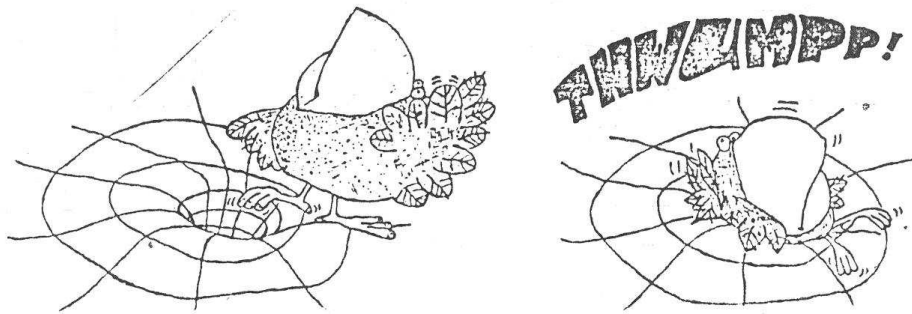
$$dX^b = \underbrace{\frac{2}{4}(\mu^2 + \nu^2)(2\nu d\nu \wedge \mu d\mu) - \frac{2}{4}(\mu^2 + \nu^2)(2\mu d\mu \wedge \nu d\nu)}_{-2\mu\nu(\mu^2 + \nu^2)d\mu \wedge d\nu} + 2\mu\nu^2 d\mu \wedge d\phi + 2\mu^2 \nu d\nu \wedge d\phi$$

$$\left. \begin{array}{ll} \mu = u^1 & 0 \leq u^1 \leq \nu_0 \\ \nu = \nu_0 & 0 \leq u^2 \leq 2\pi \\ \phi = u^2 & \end{array} \right\} \quad \text{oriented parametrization of } \Sigma$$

$$\int_{\Sigma} dX^b = \int_0^{2\pi} \int_0^{\nu_0} [0 + 2u^1 \nu_0^2 du^1 du^2 + 0] = 2\pi (u^1)^2 \nu_0^2 \Big|_0^{\nu_0} = 2\pi \nu_0^4$$

The End (for real).

A takehome final or a blackhole...?



Didn't mean to suck you into this group project
at the end — there just wasn't enough time.

I know you are all anxious to take off.



Please read over these notes.
Hope they made some impression.

— bob

Figure E.11:

Maple worksheets

Index of related Maple worksheets, correlating them with the Exercises.
TO DO ...