General Relativity

Physics 568/821

Lectures by Prof. David Kastor

Notes compiled by Siyi Zhang

Fall 2025

Jump to lecture

Lecture 1 . . . 10
Lecture 2 . . . 12
Lecture 4 . . . 15
Lecture 5 . . . 24
Lecture 6 . . . 29
Lecture 7 . . . 34
Lectures 8–9 . . . 39
Lecture 10 . . . 49
Lecture 11 . . . 54
Lecture 12 . . . 58
Lecture 13 . . . 63
Lecture 14 . . . 68
Lecture 15 . . . 70
Lectures 16–18 . . . 73
Lecture 19 . . . 83
Lecture 20 . . . 86
Lectures 21–22 . . . 90
Lecture 23 . . . 97
Lecture 24 . . . 101
Lectures 25–26 . . . 103
Lecture 27 . . . 110
Lecture 28 . . . 114
Lectures 29–30 . . . 121

Content

Introduction 8

References and further reading 9

I Gravity and the Geometry of Space 10

1 Why a new theory of gravity was needed 10
- 1.1 Newton's inverse-square law . . . 10
- 1.2 Instantaneous action at a distance . . . 11

2 Special relativity and electromagnetism 12
- 2.1 Galilean transformations and Newtonian mechanics . . . 12
- 2.2 Lorentz transformations and spacetime . . . 12
- 2.3 Electrostatics is only the static limit . . . 13

3 Gravity as a nonlinear geometry 13
- 3.1 Why gravitational waves must interact with gravity . . . 13
- 3.2 Curvature and freely falling motion . . . 14
- 3.3 Weak gravity and large-scale consequences . . . 15

4 Lengths of curves and coordinate independence 15
- 4.1 The Euclidean line element . . . 15
- 4.2 A curve and two different parametrizations . . . 16
- 4.3 General reparametrization invariance . . . 16

5 Line elements in different coordinates 17
- 5.1 Polar coordinates and their domain . . . 17
- 5.2 Deriving the polar line element . . . 18

6 Computing curve lengths with the polar metric 20
- 6.1 A quarter-circle . . . 20
- 6.2 The same hyperbola in polar coordinates . . . 20

7 Three-dimensional Euclidean space 22
- 7.1 Cartesian coordinates and curve length . . . 22
- 7.2 Spherical coordinates . . . 22
- 7.3 Deriving the spherical line element . . . 23

8 The round two-sphere as a curved space 24
- 8.1 Restricting the line element to a surface . . . 24
- 8.2 Circumference in the Euclidean plane . . . 25
- 8.3 The intrinsic radius of a circle on the sphere . . . 25
- 8.4 Circumference of a geodesic circle . . . 26
- 8.5 Local flatness and the small-circle limit . . . 26
- 8.6 The equator and the antipodal pole . . . 27

9 Metric components and indexed coordinates 28

9.1 Coordinates labelled by indices . . . 28
9.2 The symmetric quadratic form . . . 28

10 Metric components and index notation **29**
10.1 Why the metric is symmetric . . . 29
10.2 Cartesian and polar examples . . . 30
10.3 Explicit sums and the Einstein convention . . . 30

II Special Relativity and Minkowski Spacetime **32**

11 From Newtonian space and time to spacetime **32**
11.1 Absolute time and inertial frames . . . 32
11.2 Galilean transformations . . . 32
11.3 Why electromagnetism requires a different transformation . . . 33

12 The Minkowski metric and units **34**
12.1 The spacetime interval . . . 34
12.2 Setting the speed of light to one . . . 35

13 Euclidean rotations as a model for spacetime symmetries **35**
13.1 Translations and rotated axes . . . 35
13.2 Derivation using orthonormal basis vectors . . . 36
13.3 The same transformation from projections on the axes . . . 36
13.4 Checking invariance explicitly . . . 37
13.5 Derivation using polar coordinates . . . 37
13.6 The rotation group . . . 38

14 Hyperbolic geometry in spacetime **39**
14.1 Why begin with one space and one time dimension? . . . 39
14.2 Circular orbits and the angle coordinate . . . 39
14.3 Why ordinary polar coordinates do not expose boost symmetry . . . 40
14.4 The two families of hyperbolas . . . 40
14.5 Hyperbolic functions . . . 41

15 Milne coordinates **42**
15.1 Parametrizing the future hyperbolas . . . 42
15.2 Deriving the Milne line element . . . 43
15.3 Coordinate lines and an expanding family of observers . . . 43
15.4 The past wedge . . . 44

16 Rindler coordinates in the right and left wedges **45**
16.1 The right wedge and its metric . . . 45
16.2 The left wedge . . . 46

17 Coordinate patches and coordinate singularities **46**

18 Lorentz boosts as hyperbolic rotations **47**
18.1 Useful identities and small-parameter expansions . . . 47
18.2 From a hyperbolic-angle shift to a linear transformation . . . 48

19 Lorentz boosts, rapidity, and inertial frames **49**
19.1 The boost matrix and its invariant quadratic form . . . 49
19.2 A boost as a translation of the hyperbolic angle . . . 50
19.3 The Lorentz and Poincaré groups in four dimensions . . . 50
19.4 Relating rapidity to relative velocity . . . 51
19.5 The Galilean limit . . . 52
19.6 How the axes transform on a spacetime diagram . . . 53

20 Simultaneity, intervals, and causality **54**
20.1 Surfaces of constant primed coordinates . . . 54
20.2 Newtonian simultaneity and relativistic simultaneity . . . 54
20.3 The invariant interval and its three causal types . . . 55
20.4 Which events can change their time ordering? . . . 56
20.5 Particle trajectories and the principle of causality . . . 57

21 The Minkowski metric and the time measured by a clock **58**
21.1 Metric components and the summation convention . . . 58
21.2 Proper time on a straight timelike path . . . 59
21.3 Proper time on a general timelike curve . . . 60
21.4 Proper time and proper length in four dimensions . . . 61

III Vector Spaces and Tensor Algebra **62**

22 Real vector spaces and the meaning of a vector **62**
22.1 Why vectors and tensors are needed . . . 62
22.2 Linear dependence and dimension . . . 62
22.3 Bases and components . . . 63

23 Tangent spaces and vector fields **63**
23.1 A vector at a point is a tangent to a curve . . . 63
23.2 Tangent vectors do not connect arbitrary points . . . 64
23.3 The tangent bundle and a vector field . . . 65

24 Dual vectors, cotangent spaces, and tensors **65**
24.1 The dual vector space . . . 65
24.2 Constructing the dual basis . . . 66
24.3 The natural pairing of a vector and a covector . . . 67
24.4 A vector as a linear map on covectors . . . 67
24.5 Tensors as multilinear maps . . . 68

25 Tensors as multilinear maps **68**
25.1 Vectors and covectors as maps . . . 68
25.2 The definition and the meaning of the slots . . . 69
25.3 Scalars, vectors, covectors, and tensor fields . . . 69
25.4 The vector space of tensors . . . 70

26 Tensor products, bases, and components **70**
26.1 The tensor product of two tensors . . . 70
26.2 Building a basis for the tensor space . . . 71

26.3 Evaluating a tensor using its components 72

27 The metric tensor and index lowering **73**
27.1 A symmetric bilinear form at each point 73
27.2 Minkowski inner products and causal character 73
27.3 The metric maps vectors to covectors 74

28 Mixed tensors and the inverse metric **75**
28.1 A type $(1,1)$ tensor as a linear map 75
28.2 The Kronecker tensor is the identity 76
28.3 The inverse metric 77
28.4 Raising an index and undoing index lowering 77

29 Contractions and traces **78**
29.1 Contracting an upper index with a lower index 78
29.2 Taking traces with the metric 79

30 Symmetrization and antisymmetrization **79**
30.1 Two-index symmetry and its projections 79
30.2 Products that are symmetric or antisymmetric 80
30.3 Symmetric–antisymmetric contractions vanish 81
30.4 Symmetrizing more indices or only selected slots 81
30.5 Mixed symmetry at higher rank 82

IV Differential Geometry **83**

31 Manifolds and global topology **83**
31.1 What it means to look locally like Euclidean space 83
31.2 Why coordinate patches are open 83
31.3 The sphere and stereographic coordinates 84
31.4 Examples and nonexamples 85

32 Charts, atlases, and tangent spaces **86**
32.1 Formal coordinate maps and their domains 86
32.2 Tangent vectors from curves 88
32.3 Why tangent vectors form an n-dimensional vector space 88
32.4 How coordinate bases transform 89
32.5 How vector components transform 89

33 Cartesian and polar coordinates **90**
33.1 Components, Jacobians, and inverse transformations 90
33.2 Coordinate bases are not necessarily unit bases 91

34 Vector fields and their commutator **91**
34.1 Vector fields as first-order differential operators 91
34.2 Deriving the components and canceling second derivatives 92
34.3 Why either directional component derivative fails separately 92
34.4 Simple checks on the bracket 93

35 Cotangent vectors and differential one-forms **94**
35.1 The differential of a function . . . 94
35.2 The dual coordinate basis . . . 94
35.3 Covector components and bases transform inversely . . . 95
35.4 The scalar pairing is invariant . . . 95

36 General tensors and the metric **96**
36.1 Tensor-product bases . . . 96
36.2 The full tensor transformation law . . . 96
36.3 The metric as a symmetric bilinear form . . . 96

37 Inverse metrics, identity tensors, and signature **97**
37.1 The inverse metric is a contravariant tensor . . . 97
37.2 Why only the mixed Kronecker tensor is the identity . . . 98
37.3 Canonical form and metric signature . . . 98

38 Why ordinary partial derivatives fail **98**
38.1 Ordinary derivatives are not tensor components . . . 98
38.2 A constant vector in polar coordinates . . . 99
38.3 The defining properties of a covariant derivative . . . 100

39 Connections and covariant differentiation **101**
39.1 A constant vector field in polar coordinates . . . 101
39.2 The connection coefficients . . . 101
39.3 Contractions and the derivative of a covector . . . 102
39.4 The derivative of an arbitrary tensor . . . 102

40 The Levi–Civita connection **103**
40.1 The freedom in choosing a connection . . . 103
40.2 Torsion and symmetry of the lower indices . . . 103
40.3 Metric compatibility and moving indices . . . 104
40.4 Deriving the Christoffel formula . . . 104
40.5 Christoffel symbols of the two-sphere . . . 105
40.6 All four components of the vector derivative on the sphere . . . 106
40.7 Why some partial-derivative combinations are tensors . . . 106

41 Parallel transport **107**
41.1 Constant direction along a curve . . . 107
41.2 Parallel transport as a linear initial-value problem . . . 107
41.3 A flat-space circle in polar coordinates . . . 108

42 Geodesics and the variational principle **110**
42.1 Parallel transport of the tangent vector . . . 110
42.2 Stationary length and the local nature of the problem . . . 111
42.3 The length functional and fixed-endpoint variations . . . 111
42.4 Integration by parts and the geodesic equation . . . 113
42.5 Affine parameters and constant tangent norm . . . 114

43 Causal character and geodesic parameters **114**
43.1 Stationarity, minima, and maxima . . . 114

43.2 Straight lines in flat space and spacetime . . . 115
43.3 Great circles on the sphere . . . 116
43.4 Length does not depend on the choice of parameter . . . 116
43.5 Deriving the nonaffine geodesic equation . . . 117

44 Parallel transport around closed curves **118**
44.1 Curvature as failure to return . . . 118
44.2 Parallel transport preserves lengths and inner products . . . 119
44.3 Transport around a circle of constant latitude . . . 119
44.4 The small-loop limit near the pole . . . 120

45 Curvature as a local tensor **121**
45.1 Infinitesimal loops and the meaning of the indices . . . 121
45.2 Definition through a commutator . . . 122
45.3 Deriving the connection formula . . . 122

46 The curvature commutator on tensors **123**
46.1 Scalars and the torsion-free condition . . . 123
46.2 Vectors: two ways to obtain the answer . . . 123
46.3 One curvature term for each index . . . 124

47 Riemann symmetries and identities **125**
47.1 Pair exchange and the algebraic identity . . . 125
47.2 The differential Bianchi identity . . . 126

48 Ricci curvature and scalar curvature **126**
48.1 The Ricci contraction . . . 126
48.2 The scalar contraction . . . 126

49 Curvature of a two-sphere **127**
49.1 The independent Riemann component . . . 127
49.2 Computing every Ricci component . . . 127
49.3 Scalar curvature and the radius . . . 128

50 Einstein spaces **129**

Introduction

These notes were compiled by **Siyi Zhang** for **Physics 568/821: General Relativity**, taught by **Prof. David Kastor** in **Fall 2025**. The course develops the conceptual and mathematical foundations of special and general relativity, with differential geometry providing the language in which gravitation is described.

Course content

The material in this volume follows the course's progression from geometry and special relativity to tensor algebra and curved spaces:

1. **Gravity and the geometry of space.** The limitations of Newtonian descriptions; lengths of curves, coordinate and parameter independence, polar and spherical coordinates, the geometry of a two-sphere, and metric notation.

2. **Special relativity and Minkowski spacetime.** Galilean transformations, Minkowski geometry, rotations and boosts, Milne and Rindler coordinates, rapidity, simultaneity, causality, proper time, and proper length.

3. **Vector spaces and tensor algebra.** Bases, dual spaces, tensors as multilinear maps, tensor products, components, metric and inverse metric, raising and lowering indices, contractions, and symmetry and antisymmetry.

4. **Differential geometry.** Manifolds, charts and atlases; tangent vectors as derivations; vector fields and their commutators; covectors and tensor fields; metric signature, covariant derivatives, the Levi–Civita connection, parallel transport, and geodesics.

5. **Curvature.** Holonomy, the Riemann tensor and its identities, Ricci and scalar curvature, explicit curvature calculations on the two-sphere, and Einstein spaces.

The lectures place substantial emphasis on the mathematical framework before developing its physical applications. Definitions are followed by component calculations and worked examples so that the abstract and coordinate descriptions can be used together. The main text ends at the material reached in the supplied lectures.

References and further reading

The course syllabus assigns the following main textbook. The remaining books are the supplemental references listed by the instructor.

Main textbook

- Sean M. Carroll, *Spacetime and Geometry*. Cambridge University Press, 2019.

Supplemental references

- James B. Hartle, *Gravity: An Introduction to Einstein's General Relativity*. A physics-oriented introduction requiring less of the full machinery of differential geometry.
- Bernard Schutz, *A First Course in General Relativity*, 3rd ed., 2022. An introductory treatment with astrophysics, cosmology, and gravitational-wave applications.
- Steven Weinberg, *Gravitation and Cosmology*. An approach emphasizing astrophysics and cosmology.
- Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler, *Gravitation*. A comprehensive treatment with extensive physical and geometrical explanations.
- Ray d'Inverno, *Introducing Einstein's Relativity*. A general textbook on the subject.
- Norbert Straumann, *General Relativity and Relativistic Astrophysics*. A treatment with emphasis on astrophysical applications.
- Robert M. Wald, *General Relativity*. A mathematical treatment including advanced topics such as singularity theorems.
- Bernard Schutz, *Geometrical Methods of Mathematical Physics*. Further differential geometry from a physics perspective.
- Eric Poisson, *A Relativist's Toolkit: The Mathematics of Black-Hole Mechanics*. Mathematical tools for black-hole physics.

Part I
Gravity and the Geometry of Space

Lecture 1

1 Why a new theory of gravity was needed

General relativity is Einstein's theory of gravity, completed in 1915. Its central idea is that gravitational phenomena are described by the geometry of spacetime. To develop that idea, we first need to understand how geometry is expressed in coordinates, how geometrical quantities remain independent of those coordinates, and how the same methods extend from space to spacetime.

The course therefore makes a substantial mathematical investment before turning to applications: roughly sixty percent concerns the geometrical framework and the formulation of general relativity, while the remaining forty percent develops its physical consequences. The Newtonian limit, gravitational radiation, black holes, and cosmology will provide the eventual applications of the metric, tensor calculus, and curvature.

1.1 Newton's inverse-square law

Newton's gravitational theory, published in 1687, already explained both the motion of planets around the Sun and the motion of projectiles near the Earth's surface. For two point masses M_1 and M_2, let

$$\boldsymbol{r} = \boldsymbol{x}_2 - \boldsymbol{x}_1, \qquad r = |\boldsymbol{r}|, \qquad \hat{\boldsymbol{r}} = \frac{\boldsymbol{r}}{r}.$$

Thus $\hat{\boldsymbol{r}}$ points from M_1 toward M_2. The force exerted by M_1 on M_2 is

$$\boldsymbol{F}_{1\to 2} = -\frac{GM_1M_2}{r^2}\hat{\boldsymbol{r}}, \qquad G \simeq 6.67 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}. \tag{1}$$

The minus sign makes the interaction attractive. The inverse-square dependence determines how the force weakens with separation.

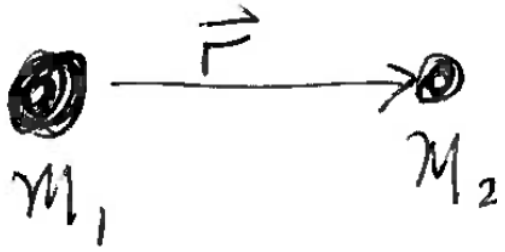

Figure 1: The separation vector points from M_1 to M_2; the gravitational force on M_2 points in the opposite direction.

Combining (1) with Newton's second law gives the test body's acceleration,

$$\frac{\mathrm{d}^2\boldsymbol{x}_2}{\mathrm{d}t^2} = -\frac{GM_1}{r^2}\hat{\boldsymbol{r}}.$$

The mass of the test body cancels. The same law therefore governs falling bodies of different masses when other forces can be neglected.

The empirical success of this theory makes the next question important: why was a different understanding of gravity required? One fundamental problem is how a change in the source is communicated to a distant body.

1.2 Instantaneous action at a distance

In the Newtonian force law, the force at time t depends on the source position at that same time:

$$\boldsymbol{F}_{1\to 2}(t) = -GM_1M_2\,\frac{\boldsymbol{x}_2(t)-\boldsymbol{x}_1(t)}{|\boldsymbol{x}_2(t)-\boldsymbol{x}_1(t)|^3}.$$

If M_1 is moved, this expression changes immediately, no matter how large the separation is. This is the **action-at-a-distance problem**.

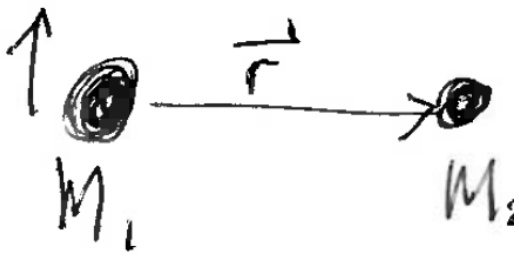

Figure 2: Moving the source mass changes the force immediately in the Newtonian description. A relativistic theory must instead describe how the change propagates.

Special relativity imposes a finite limiting speed c for causal influences. A source cannot send a physical signal to a distant object instantaneously. The conflict is more serious than merely replacing one numerical speed by another: superluminal signals would not have an observer-independent ordering of emission and reception.

To see the issue, consider two events separated by $\Delta x = v\Delta t$, with $\Delta t > 0$ and $v > c$. Under a Lorentz boost of speed u in the same direction,

$$\Delta t' = \gamma_u\left(\Delta t - \frac{u\Delta x}{c^2}\right) = \gamma_u\Delta t\left(1-\frac{uv}{c^2}\right), \qquad \gamma_u = \frac{1}{\sqrt{1-u^2/c^2}}. \tag{2}$$

Because $v > c$, it is possible to choose a physical boost satisfying $c^2/v < u < c$. Then $\Delta t' < 0$: the event called reception in the first frame occurs before the event called emission in the second. If freely controllable superluminal signaling were allowed between inertial observers, this change of time order could be used to violate the usual causal ordering of cause and effect. A relativistic gravitational theory must avoid that possibility.

Lecture 2

2 Special relativity and electromagnetism

2.1 Galilean transformations and Newtonian mechanics

Einstein's special relativity, published in 1905, addressed the relationship between mechanics and Maxwell's electromagnetic theory, developed in 1861–1862. Newtonian mechanics uses Galilean transformations between inertial observers moving with constant relative velocity.

Let the primed frame move with constant velocity $\boldsymbol{u}$ relative to the unprimed frame, and let the origins coincide at $t = 0$. With no relative rotation,

$$\boldsymbol{x}' = \boldsymbol{x} - \boldsymbol{u}t, \qquad t' = t. \tag{3}$$

Time is absolute in this description. Differentiating gives

$$\begin{aligned} \boldsymbol{v}' &= \frac{\mathrm{d}\boldsymbol{x}'}{\mathrm{d}t'} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} - \boldsymbol{u} = \boldsymbol{v} - \boldsymbol{u}, \\ \boldsymbol{a}' &= \frac{\mathrm{d}\boldsymbol{v}'}{\mathrm{d}t'} = \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \boldsymbol{a}. \end{aligned}$$

Thus inertial observers disagree about velocities but agree about accelerations. With the same inertial mass and a Galilean-invariant force law, $\boldsymbol{F} = m\boldsymbol{a}$ has the same form in both frames. In particular, separations between two particles at the same time are unchanged by a common Galilean boost, so the Newtonian inverse-square force fits this framework.

Maxwell's equations together with the Lorentz force law do not have Galilean invariance. They contain a fixed propagation speed. In vacuum, Maxwell's equations imply

$$\nabla^2 \boldsymbol{E} - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{E}}{\partial t^2} = 0, \qquad \nabla^2 \boldsymbol{B} - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{B}}{\partial t^2} = 0, \qquad c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}.$$

A light pulse moving along x with $x = ct$ would have $x' = (c - u)t'$ under a Galilean transformation, giving a different speed. Galilean kinematics therefore cannot preserve the same light-propagation law for every inertial observer.

2.2 Lorentz transformations and spacetime

Lorentz transformations mix space and time. For motion along one axis,

$$x' = \gamma_u(x - ut), \qquad t' = \gamma_u\left(t - \frac{ux}{c^2}\right), \qquad y' = y, \quad z' = z. \tag{4}$$

For the light trajectory $x = ct$,

$$x' = \gamma_u(c - u)t, \qquad ct' = \gamma_u(c - u)t,$$

so $x' = ct'$: the primed observer measures the same light speed. Space and time are therefore parts of one spacetime geometry rather than separate structures with a universal time coordinate.

For $u \ll c$, $\gamma_u \simeq 1$ and the transformation approaches the Galilean form in the nonrelativistic regime. Special relativity modifies mechanics so that its laws, like the electromagnetic laws, transform consistently under Lorentz transformations. The resulting mechanics reduces to Newtonian mechanics at small speeds. Its geometrical setting is flat **Minkowski spacetime**, which will be studied after the corresponding ideas for ordinary space.

2.3 Electrostatics is only the static limit

Electrostatics appears to have the same action-at-a-distance difficulty as Newtonian gravity. The Coulomb force on charge Q_2 due to charge Q_1 is

$$\boldsymbol{F}_{1\to 2} = \frac{1}{4\pi\epsilon_0}\frac{Q_1 Q_2}{r^2}\hat{\boldsymbol{r}}. \tag{5}$$

If this were the complete dynamical theory, moving Q_1 would seem to change the force on Q_2 immediately. But Coulomb's static law is only part of electrodynamics.

In the full electromagnetic theory, a change in the motion of a source produces a disturbance in its electric and magnetic fields. That disturbance propagates outward at speed c. A distant charge responds to the changed field only after the disturbance reaches it. If the source and receiver are separated by distance r, the first possible influence arrives after a time of order r/c.

The propagating disturbance is electromagnetic radiation: it does not merely travel at the speed of light; it is light in the general sense of an electromagnetic wave. An accelerated charge radiates electromagnetic energy. Maxwell's equations provide a local field description that replaces the misleading instantaneous interpretation of the electrostatic force law.

This analogy suggests what is needed for gravity. A relativistic theory should possess propagating gravitational disturbances, and its static or slowly varying limit should reproduce Newtonian gravity. These disturbances are **gravitational waves** or **gravitational radiation**. Constructing their complete dynamics, however, is more demanding than copying Maxwell's equations.

3 Gravity as a nonlinear geometry

3.1 Why gravitational waves must interact with gravity

Special relativity identifies rest mass as one form of energy through $E = mc^2$. A relativistic gravitational theory must account for all forms of energy, rather than coupling only to a separately distinguished Newtonian mass.

A gravitational wave carries energy. That energy must itself contribute to gravity. The resulting gravitational disturbance can in turn carry energy and contribute

further to the field. Thus the gravitational field must interact with itself. This self-coupling is the physical reason to expect nonlinear equations.

The contrast with classical vacuum electrodynamics is instructive. Electromagnetic waves carry energy but no electric charge. The source of the Maxwell field is charge and current, so electromagnetic radiation does not supply an additional electric-charge source for the same classical Maxwell equations. Consequently the vacuum Maxwell equations are linear, and sums of electromagnetic wave solutions are again solutions. Gravitational energy cannot be excluded from the gravitational interaction in the same way.

One can develop general relativity by systematically adding the self-interactions required for a gravitational wave field. That route is technically involved. Einstein's geometrical approach provides a different organizing principle: the gravitational field is part of the structure that defines spacetime distances and free motion.

3.2 Curvature and freely falling motion

In general relativity, gravity is an aspect of **spacetime curvature**. The mass, energy, momentum, and stresses of matter influence this curvature. The dynamical quantity is the **spacetime metric**, which tells us how to calculate intervals and thereby determines the geometry.

Special relativity describes the absence of gravitational curvature by the flat Minkowski metric. General relativity permits curved spacetime metrics. A freely falling test body, subject to no nongravitational force, follows a **geodesic**: the curved-spacetime generalization of an inertial straight worldline.

In Euclidean geometry, a straight segment gives the shortest distance between its endpoints. The corresponding variational interpretation in spacetime must respect the Lorentzian signature: a timelike geodesic makes proper time stationary and locally maximizes it over sufficiently short segments with fixed endpoints. It is therefore the notion of straight or freely falling motion, rather than an unqualified shortest-distance rule, that carries over to spacetime.

In this geometrical description, one does not add a separate gravitational force to the equation of an otherwise freely moving test body. The worldline is determined by the geometry. Nongravitational interactions, such as electromagnetic forces or the force exerted by a supporting surface, can cause departures from a geodesic.

The metric in four dimensions is a symmetric 4×4 array $g_{\mu\nu}$. It has

$$\frac{4(4+1)}{2} = 10$$

independent components. Einstein's field equations form ten coupled nonlinear second-order component equations for this metric, with coordinate freedom and constraint relations that must be handled consistently. Their physical content will be developed after the required geometry.

3.3 Weak gravity and large-scale consequences

The slow-motion, weak-field limit of general relativity recovers Newtonian gravity. More distinctly relativistic applications include gravitational waves, expanding-universe cosmology, and black holes. They are not equally apparent in everyday life because ordinary laboratory and solar-system gravitational fields are weak in relativistic units.

For a mass M, the characteristic Schwarzschild length is

$$r_s = \frac{2GM}{c^2}. \tag{6}$$

For the Sun, r_s is approximately 3 km, whereas the solar radius is approximately 7×10^5 km. Their ratio is

$$\frac{r_s}{R_\odot} \simeq 4 \times 10^{-6}.$$

The associated surface potential scale is $GM_\odot/(R_\odot c^2) = r_s/(2R_\odot)$, also very small. A large amount of mass concentrated into a sufficiently small region is required for strongly relativistic gravitational effects.

Nevertheless, gravity dominates much of the large-scale organization of ordinary matter. Large collections of matter tend to be electrically neutral, so their long-range electric fields largely cancel. The Newtonian gravitational attractions of positive masses instead add. Gravity therefore governs planetary systems, galaxies, and the evolution of the universe on large scales, despite its weakness in many local experiments.

Lecture 4

4 Lengths of curves and coordinate independence

4.1 The Euclidean line element

In the Euclidean plane, Cartesian coordinates (x, y) express Pythagoras' theorem in differential form:

$$\mathrm{d}\ell^2 = \mathrm{d}x^2 + \mathrm{d}y^2. \tag{7}$$

The **line element** describes the length of an infinitesimal displacement in terms of its coordinate changes. For a parametrized curve

$$x = x(s), \qquad y = y(s), \qquad s_a \le s \le s_b,$$

the coordinate differentials are $\mathrm{d}x = x'(s)\mathrm{d}s$ and $\mathrm{d}y = y'(s)\mathrm{d}s$. Hence its length is

$$\ell = \int_{s_a}^{s_b} \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}s}\right)^2}\, \mathrm{d}s. \tag{8}$$

The parameter labels points on the curve. It need not itself be distance, and in general it is not an arc-length parameter.

4.2 A curve and two different parametrizations

Consider the segment of the curve $y = 1/x$ from $(1, 1)$ to $(2, 1/2)$. A convenient parametrization is

$$x(s) = 1 + s, \qquad y(s) = \frac{1}{1+s}, \qquad 0 \le s \le 1.$$

Then

$$\frac{\mathrm{d}x}{\mathrm{d}s} = 1, \qquad \frac{\mathrm{d}y}{\mathrm{d}s} = -\frac{1}{(1+s)^2},$$

and therefore

$$\ell = \int_0^1 \sqrt{1 + \frac{1}{(1+s)^4}}\,\mathrm{d}s = \int_1^2 \sqrt{1 + \frac{1}{x^4}}\,\mathrm{d}x \simeq 1.1321. \tag{9}$$

The integral is easily evaluated numerically even though its integrand does not have a simple elementary antiderivative.

The same geometric curve can be traversed with a different parameter u:

$$x(u) = 1 + u^2, \qquad y(u) = \frac{1}{1+u^2}, \qquad 0 \le u \le 1.$$

Its derivatives are

$$\frac{\mathrm{d}x}{\mathrm{d}u} = 2u, \qquad \frac{\mathrm{d}y}{\mathrm{d}u} = -\frac{2u}{(1+u^2)^2}.$$

The length computed from this parametrization is

$$\begin{aligned}\ell &= \int_0^1 \sqrt{4u^2 + \frac{4u^2}{(1+u^2)^4}}\,\mathrm{d}u \\ &= \int_0^1 2u\sqrt{1 + \frac{1}{(1+u^2)^4}}\,\mathrm{d}u.\end{aligned}$$

With $s = u^2$ and $\mathrm{d}s = 2u\mathrm{d}u$, this becomes exactly the integral in (9). Changing the speed at which the parameter runs through the curve does not change its length.

4.3 General reparametrization invariance

Let $s = s(u)$ be a monotone change of parameter describing the same curve. The chain rule gives

$$\frac{\mathrm{d}x}{\mathrm{d}u} = \frac{\mathrm{d}x}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}u}, \qquad \frac{\mathrm{d}y}{\mathrm{d}u} = \frac{\mathrm{d}y}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}u}.$$

Thus

$$\sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}u}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}u}\right)^2} = \left|\frac{\mathrm{d}s}{\mathrm{d}u}\right| \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}s}\right)^2}.$$

For an increasing reparametrization, this is precisely the factor required by a change of integration variable. For a decreasing one, the absolute value and reversal of the integration limits again leave the positive length unchanged.

This should be distinguished from a change of *spatial coordinates*. Reparametrization changes the label along a curve; a spatial coordinate transformation changes the numbers assigned to points in the plane. Neither operation changes the geometric curve itself. A length obtained from the geometry must be invariant under both kinds of relabelling.

5 Line elements in different coordinates

A geometric quantity such as the length of a fixed curve should not depend on the coordinates used to describe it. This **coordinate invariance** will be a central organizing principle of general relativity.

A line element is the local prescription for computing length. More generally, a two-dimensional spatial geometry can have the form

$$\mathrm{d}\ell^2 = A(x,y)\mathrm{d}x^2 + B(x,y)\mathrm{d}y^2 + C(x,y)\big(\mathrm{d}x\mathrm{d}y + \mathrm{d}y\mathrm{d}x\big). \tag{10}$$

The functions A, B, C are components of the metric in the chosen coordinates. Since the differential products here are ordinary symmetric products, the cross term is $2C\mathrm{d}x\mathrm{d}y$.

In a spatial geometry, the quadratic form is positive for a nonzero displacement. The Cartesian Euclidean choice is $A = B = 1$, $C = 0$. A nonconstant set of coefficients need not indicate intrinsic curvature: even flat Euclidean space acquires position-dependent metric components in polar coordinates. Conversely, in a genuinely curved geometry no coordinate transformation can make the metric globally Euclidean throughout a region with nonzero curvature.

The metric and its associated line element are the central dynamical quantities in general relativity. Giving a line element is often the most useful way to specify a spacetime solution, such as a black-hole geometry. Before studying those cases, we can learn the coordinate bookkeeping in flat space.

5.1 Polar coordinates and their domain

Define polar coordinates by

$$x = r\cos\theta, \qquad y = r\sin\theta, \qquad r > 0, \quad 0 \le \theta < 2\pi. \tag{11}$$

The radial coordinate measures the distance from the origin and θ specifies the direction.

At the origin, every value of θ would describe the same point. The angle is not defined there, so the polar coordinate system fails at $r = 0$. This is a **coordinate singularity**; the plane itself is perfectly regular. An angular branch also has to be chosen to obtain a single-valued coordinate chart, since angles differing by 2π represent the same direction.

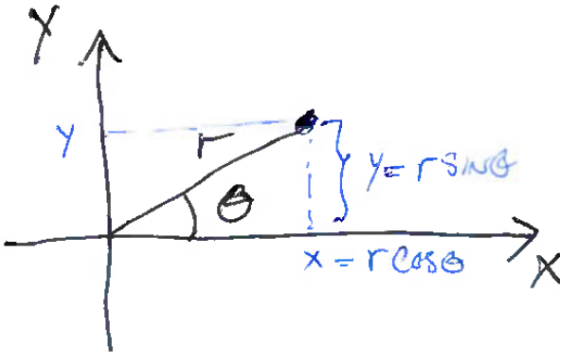

Figure 3: Cartesian and polar coordinates of the same point in the Euclidean plane.

Distinguishing a coordinate failure from a geometric singularity will be important for black holes. The Schwarzschild coordinates initially used for that solution become singular at the event horizon, although the horizon is not an intrinsic curvature singularity. The simple polar-coordinate example already shows why a problematic coordinate formula is not by itself evidence that the geometry is singular.

5.2 Deriving the polar line element

Consider a small displacement from (r, θ) to $(r + \mathrm{d}r, \theta + \mathrm{d}\theta)$. We want to rewrite (7) in terms of $\mathrm{d}r$ and $\mathrm{d}\theta$.

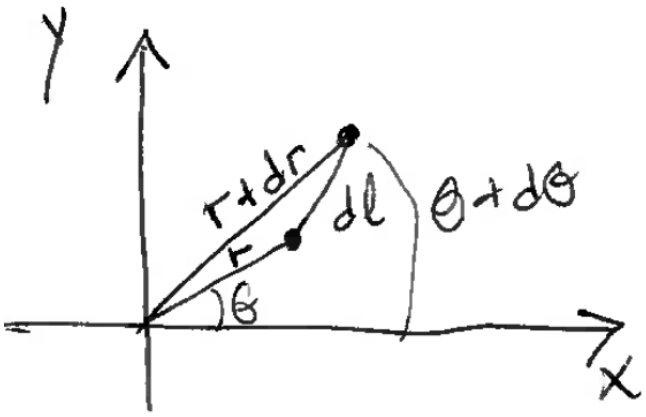

Figure 4: An infinitesimal displacement described by radial and angular coordinate changes. The displacement is exaggerated in the drawing.

The total differentials of (11) are

$$\mathrm{d}x = \frac{\partial x}{\partial r}\mathrm{d}r + \frac{\partial x}{\partial \theta}\mathrm{d}\theta = \cos\theta\,\mathrm{d}r - r\sin\theta\,\mathrm{d}\theta,$$
$$\mathrm{d}y = \frac{\partial y}{\partial r}\mathrm{d}r + \frac{\partial y}{\partial \theta}\mathrm{d}\theta = \sin\theta\,\mathrm{d}r + r\cos\theta\,\mathrm{d}\theta.$$

Substitution gives

$$\begin{aligned}\mathrm{d}\ell^2 &= (\cos\theta\,\mathrm{d}r - r\sin\theta\,\mathrm{d}\theta)^2 + (\sin\theta\,\mathrm{d}r + r\cos\theta\,\mathrm{d}\theta)^2 \\ &= (\cos^2\theta + \sin^2\theta)\mathrm{d}r^2 + r^2(\sin^2\theta + \cos^2\theta)\mathrm{d}\theta^2 \\ &\quad + (-2r\sin\theta\cos\theta + 2r\sin\theta\cos\theta)\mathrm{d}r\mathrm{d}\theta.\end{aligned}$$

The cross terms cancel, leaving

$$\boxed{\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2.} \tag{12}$$

The metric components now depend on position through r^2, but the underlying space remains the same Euclidean plane.

Two simple motions explain the two terms. Along a radial line, $\mathrm{d}\theta = 0$ and $\mathrm{d}\ell = |\mathrm{d}r|$. Along a circle of fixed radius, $\mathrm{d}r = 0$ and

$$\mathrm{d}\ell = r|\mathrm{d}\theta|.$$

Thus the physical arc length associated with a fixed angular change grows in proportion to the radius.

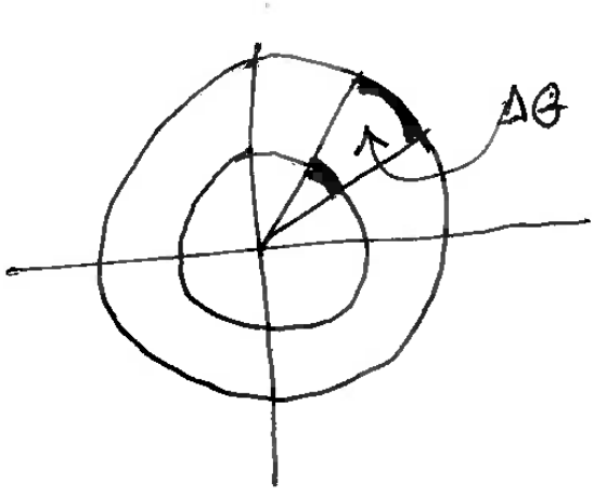

Figure 5: The same angular interval spans a larger physical arc on a circle of larger radius.

Angles measured in radians are dimensionless. Consequently the coefficient multiplying $\mathrm{d}\theta^2$ must carry dimensions of length squared in order for $\mathrm{d}\ell^2$ to have the correct units. The factor r^2 passes this dimensional check.

The coefficients are independent of θ. A rotation $\theta \mapsto \theta + \theta_0$ leaves both the differentials and the line element unchanged, making rotational invariance manifest in these coordinates.

6 Computing curve lengths with the polar metric

For a curve described by $r = r(s)$ and $\theta = \theta(s)$, the polar line element gives

$$\ell = \int_{s_a}^{s_b} \sqrt{\left(\frac{\mathrm{d}r}{\mathrm{d}s}\right)^2 + r(s)^2 \left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^2}\, \mathrm{d}s. \tag{13}$$

The factor $r(s)^2$ must be evaluated along the curve; it is not an independent constant unless the curve lies on a circle centred at the origin.

6.1 A quarter-circle

Consider the quarter-circle $r = R$ with $0 \le \theta \le \pi/2$. We expect its length to be one quarter of $2\pi R$.

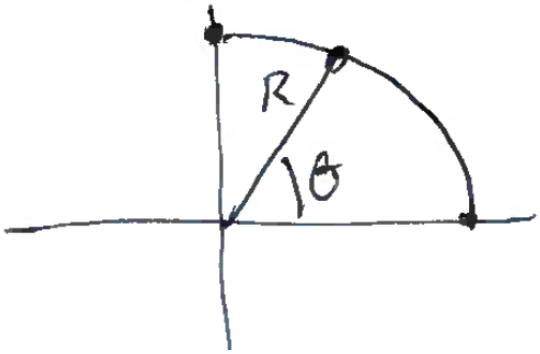

Figure 6: A quarter-circle of radius R is particularly simple in polar coordinates.

Choose the parameter range $0 \le s \le 1$ and set

$$r(s) = R, \qquad \theta(s) = \frac{\pi}{2} s.$$

Then $\mathrm{d}r/\mathrm{d}s = 0$ and $\mathrm{d}\theta/\mathrm{d}s = \pi/2$. Therefore

$$\begin{aligned} \ell &= \int_0^1 \sqrt{0 + R^2 \left(\frac{\pi}{2}\right)^2}\, \mathrm{d}s \\ &= \frac{\pi R}{2} \int_0^1 \mathrm{d}s = \frac{\pi R}{2}. \end{aligned}$$

The angular term in the metric correctly converts an angular interval into its physical arc length.

6.2 The same hyperbola in polar coordinates

Return to $y = 1/x$ between $x = 1$ and $x = 2$. The curve is not especially simple in polar coordinates, but its length must still agree with (9).

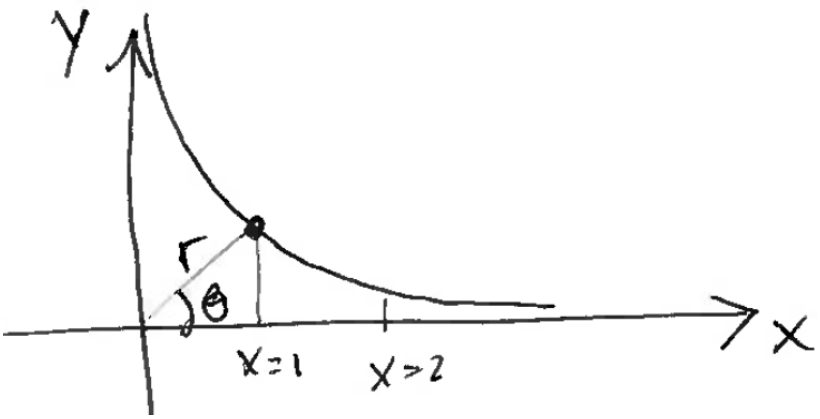

Figure 7: The curve $y = 1/x$, together with the polar coordinates of a point on it.

Keep the original parameter s and write $q = 1 + s$, so $x = q$ and $y = q^{-1}$ with $1 \le q \le 2$. The polar coordinates are

$$r(s) = \sqrt{x(s)^2 + y(s)^2} = \sqrt{q^2 + q^{-2}} = q\sqrt{1 + q^{-4}}, \tag{14}$$

$$\theta(s) = \arctan \frac{y(s)}{x(s)} = \arctan(q^{-2}). \tag{15}$$

The angle formula uses the first quadrant, where this branch of arctan is appropriate. Differentiating,

$$\frac{\mathrm{d}r}{\mathrm{d}s} = \frac{q - q^{-3}}{\sqrt{q^2 + q^{-2}}},$$
$$\frac{\mathrm{d}\theta}{\mathrm{d}s} = \frac{-2q^{-3}}{1 + q^{-4}} = -\frac{2q}{q^4 + 1}.$$

Their contributions to the squared speed are

$$\left(\frac{\mathrm{d}r}{\mathrm{d}s}\right)^2 = \frac{(q^4 - 1)^2}{q^4(q^4 + 1)}, \qquad r^2 \left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^2 = \frac{4}{q^4 + 1}.$$

Adding them gives

$$\left(\frac{\mathrm{d}r}{\mathrm{d}s}\right)^2 + r^2 \left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^2 = \frac{(q^4 - 1)^2 + 4q^4}{q^4(q^4 + 1)}$$
$$= \frac{(q^4 + 1)^2}{q^4(q^4 + 1)} = 1 + q^{-4}.$$

Hence

$$\ell = \int_0^1 \sqrt{1 + \frac{1}{(1 + s)^4}}\, \mathrm{d}s,$$

exactly as in Cartesian coordinates. This example distinguishes *invariance* from *convenience*: all valid coordinates give the same answer, but coordinates adapted to a curve can make the calculation much shorter.

7 Three-dimensional Euclidean space

7.1 Cartesian coordinates and curve length

The extension from two to three dimensions follows the same pattern. In Cartesian coordinates (x, y, z),

$$\mathrm{d}\ell^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{16}$$

For a curve $(x(s), y(s), z(s))$, its length is

$$\ell = \int_{s_a}^{s_b} \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}z}{\mathrm{d}s}\right)^2}\,\mathrm{d}s. \tag{17}$$

Again, the coordinates and the curve parameter are descriptive choices; the length is the geometric result.

7.2 Spherical coordinates

Introduce a radius r, a polar angle θ measured from the positive z axis, and an azimuthal angle ϕ in the xy plane:

$$x = r\sin\theta\cos\phi, \qquad y = r\sin\theta\sin\phi, \qquad z = r\cos\theta. \tag{18}$$

Their usual ranges are

$$r > 0, \qquad 0 \le \theta \le \pi, \qquad 0 \le \phi < 2\pi.$$

As with polar coordinates, this notation includes places where angular coordinates cease to give a regular chart. The azimuth is undefined on the polar axis, and both angles are undefined at the origin.

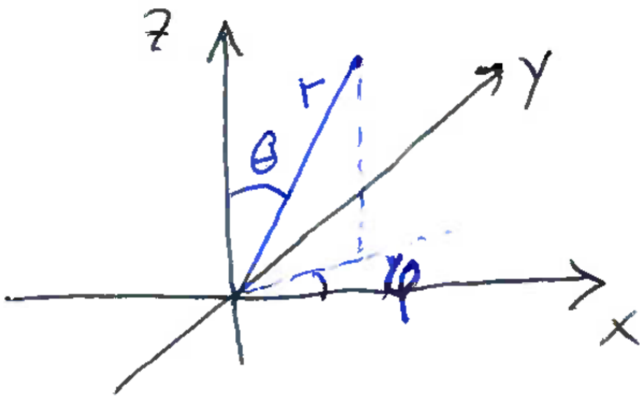

Figure 8: The radius r, polar angle θ, and azimuthal angle ϕ used for spherical coordinates.

The coordinate definitions satisfy

$$\begin{aligned} x^2 + y^2 + z^2 &= r^2\sin^2\theta(\cos^2\phi + \sin^2\phi) + r^2\cos^2\theta \\ &= r^2. \end{aligned}$$

Thus a surface of fixed r is a sphere, and (θ, ϕ) label its points away from coordinate singularities. This observation will be useful both for curved surfaces and for spacetimes with spherical symmetry.

7.3 Deriving the spherical line element

Exercise 7.1 (The Euclidean metric in spherical coordinates). Starting from (16) and the coordinate definitions (18), derive the line element in spherical coordinates.

Solution. The total differentials are

$$\begin{aligned} \mathrm{d}x &= \sin\theta\cos\phi\,\mathrm{d}r + r\cos\theta\cos\phi\,\mathrm{d}\theta - r\sin\theta\sin\phi\,\mathrm{d}\phi, \\ \mathrm{d}y &= \sin\theta\sin\phi\,\mathrm{d}r + r\cos\theta\sin\phi\,\mathrm{d}\theta + r\sin\theta\cos\phi\,\mathrm{d}\phi, \\ \mathrm{d}z &= \cos\theta\,\mathrm{d}r - r\sin\theta\,\mathrm{d}\theta. \end{aligned}$$

We can organize the expansion by collecting the coefficients of each quadratic differential. The coefficient of $\mathrm{d}r^2$ is

$$\sin^2\theta\cos^2\phi + \sin^2\theta\sin^2\phi + \cos^2\theta = 1.$$

The coefficient of $\mathrm{d}\theta^2$ is

$$r^2\cos^2\theta\cos^2\phi + r^2\cos^2\theta\sin^2\phi + r^2\sin^2\theta = r^2.$$

The coefficient of $\mathrm{d}\phi^2$ is

$$r^2\sin^2\theta\sin^2\phi + r^2\sin^2\theta\cos^2\phi = r^2\sin^2\theta.$$

All three kinds of mixed terms cancel. For $\mathrm{d}r\mathrm{d}\theta$ the coefficient is

$$2r\sin\theta\cos\theta(\cos^2\phi + \sin^2\phi) - 2r\sin\theta\cos\theta = 0.$$

For $\mathrm{d}r\mathrm{d}\phi$ it is

$$-2r\sin^2\theta\cos\phi\sin\phi + 2r\sin^2\theta\sin\phi\cos\phi = 0,$$

and for $\mathrm{d}\theta\mathrm{d}\phi$ it is

$$-2r^2\sin\theta\cos\theta\cos\phi\sin\phi + 2r^2\sin\theta\cos\theta\sin\phi\cos\phi = 0.$$

Therefore

$$\boxed{\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2.} \tag{19}$$

□

It is convenient to introduce the angular line element

$$\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2, \qquad \mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2\mathrm{d}\Omega^2. \tag{20}$$

The notation $\mathrm{d}\Omega^2$ here denotes a quadratic line element, not the square of the solid-angle area element $\sin\theta\,\mathrm{d}\theta\mathrm{d}\phi$.

The factors have direct geometrical interpretations. At fixed r and ϕ, an angular displacement $\mathrm{d}\theta$ has length $r|\mathrm{d}\theta|$. At fixed r and θ, a change $\mathrm{d}\phi$ moves around a latitude circle with length

$$\mathrm{d}\ell = r \sin\theta \, |\mathrm{d}\phi|.$$

The latitude circle has radius $r \sin\theta$ in the surrounding Euclidean space. It is largest at the equator, $\theta = \pi/2$, and shrinks to zero at the poles, $\theta = 0, \pi$. The overall factor r^2 in the angular terms records how the size of spheres grows with radius.

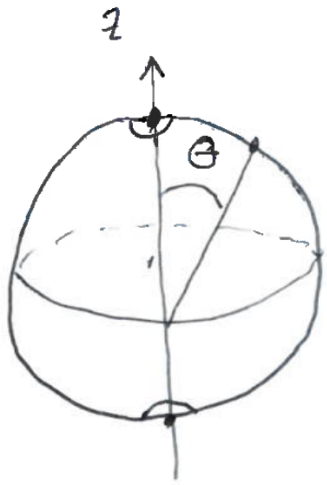

Figure 9: Latitude circles on a sphere. Their radius is largest at the equator and vanishes at the two poles.

Lecture 5

8 The round two-sphere as a curved space

8.1 Restricting the line element to a surface

Fix the radial coordinate in (19) to be $r = R$. A displacement tangent to this surface has $\mathrm{d}r = 0$, so its line element is

$$\boxed{\mathrm{d}\ell^2 = R^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi^2\right).} \tag{21}$$

This surface is the **two-sphere**, denoted S^2. Its dimension is two because only two independent coordinates, θ and ϕ, are needed locally. The ambient Euclidean space has dimension three, but it is not part of the intrinsic coordinate count of the surface.

The word *sphere* refers to the surface alone. Including the interior gives a *ball*. For a unit two-sphere, $R = 1$ in the chosen length units, and

$$\mathrm{d}\ell^2 = \mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi^2 = \mathrm{d}\Omega^2. \tag{22}$$

Although we obtained this metric by restricting flat three-dimensional space, the resulting two-dimensional geometry is intrinsically curved. A curved subspace need not inherit the flatness of its ambient space.

Later, the Riemann curvature tensor will provide a systematic local test of curvature using first and second derivatives of the metric. Before introducing that machinery, we can detect a difference from the Euclidean plane by comparing the circumference of a circle with its radius measured *within* the geometry.

8.2 Circumference in the Euclidean plane

A circle of radius ρ in the plane can be parametrized in polar coordinates as

$$r(s) = \rho, \qquad \phi(s) = 2\pi s, \qquad 0 \le s \le 1.$$

Here ϕ is the planar angular coordinate. Using $\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2\mathrm{d}\phi^2$,

$$\begin{aligned} C_{\text{flat}}(\rho) &= \int_0^1 \sqrt{\left(\frac{\mathrm{d}r}{\mathrm{d}s}\right)^2 + r^2\left(\frac{\mathrm{d}\phi}{\mathrm{d}s}\right)^2}\,\mathrm{d}s \\ &= \int_0^1 \sqrt{0 + \rho^2(2\pi)^2}\,\mathrm{d}s = 2\pi\rho. \end{aligned}$$

The familiar formula follows directly from the metric. The question is whether the same relation holds for an intrinsically defined circle on the sphere.

8.3 The intrinsic radius of a circle on the sphere

Choose the north pole as the centre. A circle on the sphere consists of the points whose distance from that pole, measured along the sphere, is a fixed value ρ.

Along a meridian, $\mathrm{d}\phi = 0$. The sphere metric reduces to

$$\mathrm{d}\ell = R|\mathrm{d}\theta|.$$

The meridian distance from $\theta = 0$ to a point with polar angle θ is therefore

$$\rho = \int_0^\theta R\,\mathrm{d}\theta' = R\theta, \qquad \theta = \frac{\rho}{R}. \tag{23}$$

The ratio ρ/R is dimensionless, as an angle must be.

A meridian does give the shortest distance from the pole to that latitude. For any path joining the pole to polar angle θ, its length satisfies

$$\begin{aligned} \ell &= R\int \sqrt{\left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^2 + \sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}s}\right)^2}\,\mathrm{d}s \\ &\ge R\int \left|\frac{\mathrm{d}\theta}{\mathrm{d}s}\right|\mathrm{d}s \ge R\theta. \end{aligned}$$

A path with constant ϕ and monotone θ attains the bound. Thus the intrinsic radius is indeed $R\theta$, not the radius of the latitude circle measured in the surrounding three-dimensional space.

That latter radius is $R\sin\theta$. Confusing $R\theta$ with $R\sin\theta$ would obscure the curvature effect we want to measure. Both are meaningful lengths, but they answer different geometrical questions.

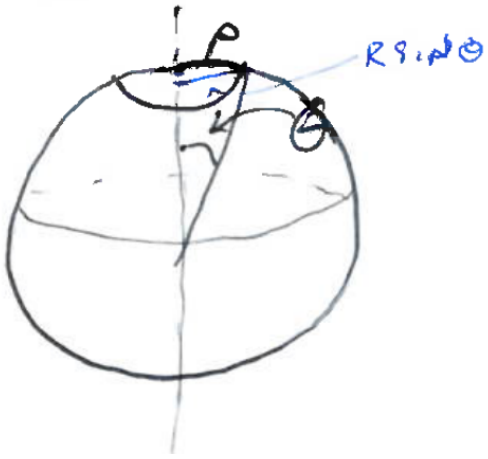

Figure 10: A circle centred at the north pole: its intrinsic radius $\rho = R\theta$ follows the sphere, whereas its radius in the ambient latitude plane is $R\sin\theta$.

8.4 Circumference of a geodesic circle

Parametrize the latitude circle at fixed intrinsic radius ρ by

$$\theta(s) = \frac{\rho}{R}, \qquad \phi(s) = 2\pi s, \qquad 0 \le s \le 1.$$

The parameter range for the radius is $0 \le \rho \le \pi R$, from the north pole to the south pole. Since θ is constant along the circle,

$$\begin{aligned} C_{S^2}(\rho) &= R\int_0^1 \sqrt{0 + \sin^2\!\left(\frac{\rho}{R}\right)(2\pi)^2}\,\mathrm{d}s \\ &= 2\pi R\sin\!\left(\frac{\rho}{R}\right). \end{aligned}$$

The sine is nonnegative over the stated range. We have found

$$\boxed{C_{S^2}(\rho) = 2\pi R\sin\!\left(\frac{\rho}{R}\right), \qquad C_{\text{flat}}(\rho) = 2\pi\rho.} \tag{24}$$

These are different relations between quantities that can be measured entirely within each two-dimensional geometry. No change of coordinates can remove the difference.

8.5 Local flatness and the small-circle limit

For $\rho \ll R$, use

$$\sin z = z - \frac{z^3}{6} + \frac{z^5}{120} + O(z^7).$$

Substituting $z = \rho/R$ into (24) gives

$$C_{S^2}(\rho) = 2\pi\rho\left[1 - \frac{\rho^2}{6R^2} + \frac{\rho^4}{120R^4} + O\!\left(\frac{\rho^6}{R^6}\right)\right]. \tag{25}$$

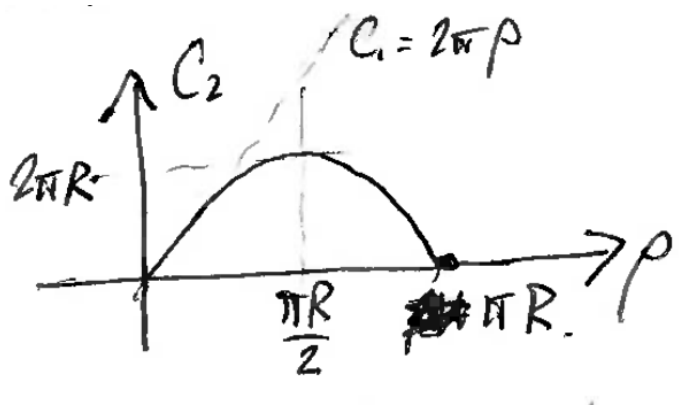

Figure 11: Circumference versus intrinsic radius: the sphere curve reaches a maximum at the equator and returns to zero at the antipodal pole; the planar result continues to grow linearly.

To leading order, $C_{S^2}(\rho) \simeq 2\pi\rho$: a sufficiently small region of the sphere is approximately flat. This is why an everyday patch of the Earth's surface can be treated as a plane, while measurements over large distances reveal its curvature.

The leading circumference deficit is

$$2\pi\rho - C_{S^2}(\rho) = \frac{\pi\rho^3}{3R^2} + O\left(\frac{\rho^5}{R^4}\right).$$

Thus the departure from flat geometry is suppressed by the squared ratio ρ^2/R^2 relative to the leading circumference. The sphere can be very accurately approximated by a plane on a small scale without being globally flat.

The same conclusion is visible directly in the metric. Write $\rho = R\theta$ near the north pole. Then

$$\mathrm{d}\ell^2 = \mathrm{d}\rho^2 + R^2 \sin^2\left(\frac{\rho}{R}\right) \mathrm{d}\phi^2 \simeq \mathrm{d}\rho^2 + \rho^2 \mathrm{d}\phi^2.$$

The leading expression is the Euclidean polar line element. Curvature appears in the higher-order corrections, not in a failure of the local distance formula to have a flat leading approximation.

8.6 The equator and the antipodal pole

The full circumference function contains more information than its small-radius expansion. Differentiating,

$$\frac{\mathrm{d}C_{S^2}}{\mathrm{d}\rho} = 2\pi \cos\left(\frac{\rho}{R}\right).$$

It increases from zero at $\rho = 0$ to a maximum at

$$\rho = \frac{\pi R}{2}, \qquad \theta = \frac{\pi}{2}, \qquad C_{S^2} = 2\pi R.$$

This is the equator. Beyond it, the circles of fixed distance from the north pole shrink as their intrinsic radius continues to increase. At the south pole,

$$\rho = \pi R, \qquad \theta = \pi, \qquad C_{S^2} = 0.$$

All azimuthal labels describe the same antipodal point there. The vanishing of the angular coefficient at a pole is a failure of the angular chart, not a physical singularity of the round sphere.

9 Metric components and indexed coordinates

9.1 Coordinates labelled by indices

To write general line elements compactly, label the coordinates by an index:

$$(x^1, x^2)$$

in two dimensions, and more generally $(x^1, \dots, x^n)$ in n spatial dimensions. The superscripts in these expressions are *coordinate labels*, not powers. For example, one may choose $x^1 = x$, $x^2 = y$, or instead $x^1 = r$, $x^2 = \theta$.

The tensor definition of the metric will be developed later. For now, its components are the coefficient functions that enter the quadratic line element. A general two-dimensional line element is

$$\begin{aligned} \mathrm{d}\ell^2 = {} & g_{11}(x^1, x^2)(\mathrm{d}x^1)^2 + g_{22}(x^1, x^2)(\mathrm{d}x^2)^2 \\ & + g_{12}(x^1, x^2)\mathrm{d}x^1\mathrm{d}x^2 + g_{21}(x^1, x^2)\mathrm{d}x^2\mathrm{d}x^1. \end{aligned} \tag{26}$$

It gives the squared infinitesimal distance between nearby coordinate values (x^1, x^2) and $(x^1 + \mathrm{d}x^1, x^2 + \mathrm{d}x^2)$.

The displayed arguments emphasize that each coefficient can depend on position. They are often suppressed to reduce clutter:

$$\mathrm{d}\ell^2 = g_{11}(\mathrm{d}x^1)^2 + g_{22}(\mathrm{d}x^2)^2 + g_{12}\mathrm{d}x^1\mathrm{d}x^2 + g_{21}\mathrm{d}x^2\mathrm{d}x^1.$$

Suppressing the arguments does not make the coefficients constant.

9.2 The symmetric quadratic form

The products in a line element are symmetric:

$$\mathrm{d}x^1\mathrm{d}x^2 = \mathrm{d}x^2\mathrm{d}x^1.$$

Therefore the two mixed terms combine as

$$g_{12}\mathrm{d}x^1\mathrm{d}x^2 + g_{21}\mathrm{d}x^2\mathrm{d}x^1 = (g_{12} + g_{21})\mathrm{d}x^1\mathrm{d}x^2.$$

Only the symmetric part of the coefficient array contributes to lengths. We consequently take the metric components to be symmetric, $g_{12} = g_{21}$, and write

$$\mathrm{d}\ell^2 = g_{11}(\mathrm{d}x^1)^2 + 2g_{12}\mathrm{d}x^1\mathrm{d}x^2 + g_{22}(\mathrm{d}x^2)^2. \tag{27}$$

The factor of two in the mixed term is important when reading metric components from a line element.

In matrix notation,

$$\mathrm{d}\ell^2 = \begin{pmatrix} \mathrm{d}x^1 & \mathrm{d}x^2 \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{pmatrix} \begin{pmatrix} \mathrm{d}x^1 \\ \mathrm{d}x^2 \end{pmatrix}.$$

For the examples already studied,

Geometry	Coordinates	Metric components
Euclidean plane	(x, y)	$\mathrm{diag}(1, 1)$
Euclidean plane	(r, θ)	$\mathrm{diag}(1, r^2)$
Round sphere	(θ, ϕ)	$\mathrm{diag}(R^2, R^2 \sin^2\theta)$

The first two rows describe the same flat geometry in different coordinates. The third describes a different, intrinsically curved two-dimensional space. Metric coefficients encode both the coordinate description and the underlying geometry; distinguishing the two is the reason to develop coordinate-invariant methods.

10 Metric components and index notation

Lecture 6

10.1 Why the metric is symmetric

A line element in two coordinates can be written

$$\mathrm{d}\ell^2 = g_{11}(\mathrm{d}x^1)^2 + (g_{12} + g_{21})\mathrm{d}x^1\mathrm{d}x^2 + g_{22}(\mathrm{d}x^2)^2. \tag{28}$$

The products of coordinate differentials in a quadratic form commute: $\mathrm{d}x^1\mathrm{d}x^2 = \mathrm{d}x^2\mathrm{d}x^1$. Therefore the cross term depends only on the sum of the two off-diagonal entries. To make this explicit, write

$$g_{12} = S + A, \qquad g_{21} = S - A.$$

Then

$$g_{12} + g_{21} = 2S, \qquad g_{12} - g_{21} = 2A,$$

and (28) contains S but no A. An antisymmetric contribution cannot change the squared length.

The same argument works in any dimension. Decompose an array into its symmetric and antisymmetric parts,

$$g_{(ab)} = \frac{1}{2}(g_{ab} + g_{ba}), \qquad g_{[ab]} = \frac{1}{2}(g_{ab} - g_{ba}).$$

Relabeling the two summed indices gives

$$g_{[ab]}\mathrm{d}x^a\mathrm{d}x^b = g_{[ba]}\mathrm{d}x^b\mathrm{d}x^a = -g_{[ab]}\mathrm{d}x^a\mathrm{d}x^b.$$

The expression must therefore be zero. In general relativity, the metric is taken to be a symmetric, nondegenerate tensor:

$$g_{ab} = g_{ba}. \tag{29}$$

For a two-dimensional metric, this means $g_{12} = g_{21}$, and the cross term in the line element is $2g_{12}\mathrm{d}x^1\mathrm{d}x^2$. In particular, the coefficient of a displayed cross term is twice the corresponding off-diagonal matrix entry.

10.2 Cartesian and polar examples

Superscripts on coordinate labels are indices, not powers. Thus x^2 means the second coordinate, whereas $(x^2)^2$ means its square. For the Euclidean plane in Cartesian coordinates, choose $x^1 = x$, $x^2 = y$. Then

$$\mathrm{d}\ell^2 = \mathrm{d}x^2 + \mathrm{d}y^2 = (\mathrm{d}x^1)^2 + (\mathrm{d}x^2)^2, \qquad (g_{ab}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Here $g_{11} = g_{22} = 1$ and $g_{12} = g_{21} = 0$.

For the same flat plane in polar coordinates, choose $x^1 = r$ and $x^2 = \theta$. The line element is

$$\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2, \qquad (g_{ab}) = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}. \tag{30}$$

In indexed notation this says $g_{11} = 1$, $g_{22} = (x^1)^2$, and $g_{12} = g_{21} = 0$. The coordinate dependence of g_{22} does not make the plane curved. It records the fact that the same angular increment corresponds to a larger distance at larger radius.

In general formulas it is convenient to write $x^1, \ldots, x^D$. In a concrete example, names such as r, θ often communicate more clearly. Accordingly, $g_{rr} = 1$ and $g_{\theta\theta} = r^2$ are the same statements as the indexed components above.

A metric matrix is **diagonal** in a given coordinate system when its off-diagonal components vanish. A nonzero $g_{r\theta}$ would produce the cross term $2g_{r\theta}\mathrm{d}r\mathrm{d}\theta$. Many examples will use diagonal line elements, but this is not a general requirement. Off-diagonal terms have important applications, including the spacetime geometry of a rotating black hole. Whether the matrix is diagonal must always be stated relative to coordinates.

10.3 Explicit sums and the Einstein convention

In two dimensions, let a, b each take the values $1, 2$. Then x^a means either coordinate, and g_{ab} selects an entry in the metric matrix. The quadratic form is

$$\begin{aligned} \mathrm{d}\ell^2 &= \sum_{a=1}^{2}\sum_{b=1}^{2} g_{ab}\mathrm{d}x^a\mathrm{d}x^b \\ &= \sum_{a=1}^{2}\left(g_{a1}\mathrm{d}x^a\mathrm{d}x^1 + g_{a2}\mathrm{d}x^a\mathrm{d}x^2\right) \\ &= g_{11}(\mathrm{d}x^1)^2 + g_{21}\mathrm{d}x^2\mathrm{d}x^1 + g_{12}\mathrm{d}x^1\mathrm{d}x^2 + g_{22}(\mathrm{d}x^2)^2. \end{aligned}$$

Repeated indices are a convenient way to express these sums without writing the summation symbols. In the **Einstein summation convention**, an index repeated once upstairs and once downstairs is summed over its range. Thus

$$\boxed{\mathrm{d}\ell^2 = g_{ab}\mathrm{d}x^a\mathrm{d}x^b.} \tag{31}$$

Both a and b are summed; neither is a free index. The expression is a scalar, as a squared length should be. The particular letters used for summed indices do not

matter, provided the replacements are consistent and do not collide with another index.

Equation (31) has exactly the same form in D dimensions. For the spatial examples in this section, $a, b = 1, \ldots, D$. In flat Cartesian coordinates,

$$g_{ab} = \delta_{ab}, \qquad \mathrm{d}\ell^2 = \delta_{ab}\mathrm{d}x^a\mathrm{d}x^b = \sum_{a=1}^{D}(\mathrm{d}x^a)^2.$$

The notation separates the geometry, contained in the metric, from the number and names of the coordinates.

Part II
Special Relativity and Minkowski Spacetime

11 From Newtonian space and time to spacetime

11.1 Absolute time and inertial frames

In Newtonian mechanics, time is separate from space. One can picture space as a sequence of spatial slices, one for each value of an absolute time t. An event has both a spatial location and a time. All observers using this Newtonian time agree on which events are simultaneous, and agree on the temporal order of events occurring at different times.

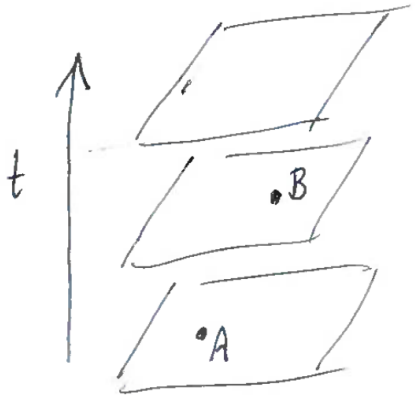

Figure 12: Newtonian space at successive values of absolute time. Events A and B lie on different time slices.

An **inertial frame** is a frame in which a force-free particle has zero acceleration. More generally, Newton's equation $\boldsymbol{F} = m\boldsymbol{a}$ holds there without adding fictitious forces. The rest frame of an accelerating observer is not inertial. For example, if the origin of a new frame follows $\boldsymbol{X}(t)$, then

$$\boldsymbol{x}' = \boldsymbol{x} - \boldsymbol{X}(t) \quad \Longrightarrow \quad \boldsymbol{a}' = \boldsymbol{a} - \ddot{\boldsymbol{X}}(t).$$

A free particle can accelerate in this new frame. To retain Newton's form there, one would have to include the inertial force $-m\ddot{\boldsymbol{X}}$.

11.2 Galilean transformations

Consider two inertial frames with parallel axes. Let the primed origin move with constant velocity $\boldsymbol{v}$ relative to the unprimed frame, and assume their origins coincide

at $t = 0$. An event at $\boldsymbol{x}(t)$ has primed coordinates

$$\boldsymbol{x}'(t') = \boldsymbol{x}(t) - \boldsymbol{v}t, \qquad t' = t. \tag{32}$$

The primed observer sees the unprimed origin moving at $-\boldsymbol{v}$. The equality $t' = t$ expresses absolute time: the transformation changes spatial descriptions while preserving the common time label.

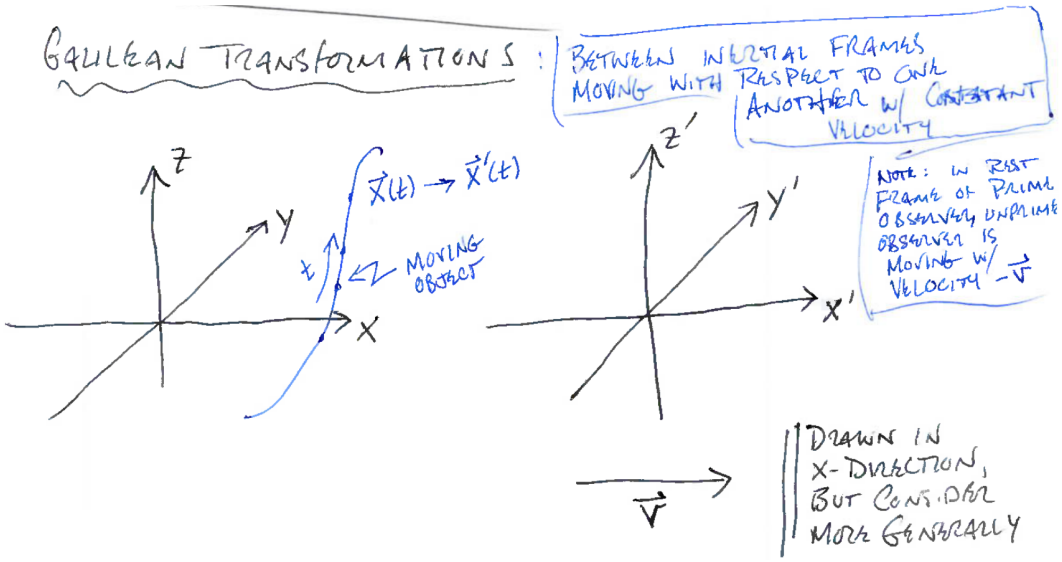

Figure 13: Two inertial coordinate systems related by constant relative velocity. The particle trajectory has coordinate descriptions $\boldsymbol{x}(t)$ and $\boldsymbol{x}'(t)$.

Because $\boldsymbol{v}$ is constant and $\mathrm{d}t' = \mathrm{d}t$,

$$\boldsymbol{u}' = \frac{\mathrm{d}\boldsymbol{x}'}{\mathrm{d}t'} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} - \boldsymbol{v} = \boldsymbol{u} - \boldsymbol{v},$$
$$\boldsymbol{a}' = \frac{\mathrm{d}^2}{\mathrm{d}t^2}(\boldsymbol{x} - \boldsymbol{v}t) = \frac{\mathrm{d}^2\boldsymbol{x}}{\mathrm{d}t^2} = \boldsymbol{a}.$$

Thus a free particle remains unaccelerated in either frame. With the usual Newtonian transformation of forces, the equations of motion retain their form. This is why Galilean transformations relate Newtonian inertial frames.

The absolute Newtonian time can also be used by noninertial observers; it is the simple form of the force law that distinguishes inertial frames. These are separate issues: changing the spatial motion of an observer does not, within Newtonian theory, change the universal time.

11.3 Why electromagnetism requires a different transformation

The Newtonian framework works extremely well in its domain, but Maxwell's equations are not invariant under Galilean boosts. They contain a fixed vacuum propagation speed c. If a light ray has velocity c along the x axis in one frame, the Galilean

velocity rule would assign it velocity $c - v$ in another. This conflicts with having the same vacuum Maxwell equations, with the same c, in both inertial frames.

The incompatibility can also be seen in a wave equation. For $x' = x - vt$ and $t' = t$, the chain rule gives

$$\partial_x = \partial_{x'}, \qquad \partial_t = \partial_{t'} - v\partial_{x'}.$$

Consequently,

$$\partial_x^2 - \frac{1}{c^2}\partial_t^2 = \left(1 - \frac{v^2}{c^2}\right)\partial_{x'}^2 + \frac{2v}{c^2}\partial_{t'}\partial_{x'} - \frac{1}{c^2}\partial_{t'}^2.$$

The transformed differential operator does not have the original form. The mixed derivative is one visible sign of the failure.

Maxwell's equations instead respect **Lorentz transformations**, which mix time and space coordinates. This led to the spacetime framework of special relativity: the speed of light is the same in all inertial frames, while a common, observer-independent division into space and time is no longer assumed. The dynamics of material objects moving close to c must also be described relativistically. Lorentz boosts reduce to the Galilean description in the appropriate low-velocity limit, explaining the success of Newtonian mechanics for ordinary slow motion.

12 The Minkowski metric and units

Lecture 7

12.1 The spacetime interval

The geometry of flat spacetime is specified by the **Minkowski line element**

$$\boxed{\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2.} \tag{33}$$

The factor of c^2 gives the temporal term the same units as the spatial terms. Unlike a Euclidean spatial line element, this quadratic form is not positive definite: a purely temporal separation has $\mathrm{d}s^2 < 0$, a purely spatial separation has $\mathrm{d}s^2 > 0$, and some nonzero separations have $\mathrm{d}s^2 = 0$.

This is why the notation $\mathrm{d}s^2$, rather than a positive length $\mathrm{d}\ell^2$, is useful. The quantity is the spacetime interval, not an ordinary Euclidean squared distance. Lorentz transformations preserve it. For two events in Cartesian inertial coordinates, their finite coordinate differences have the corresponding invariant

$$\Delta s^2 = -c^2(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2.$$

Defining $x^0 = ct$, $x^1 = x$, $x^2 = y$, $x^3 = z$, the metric components are

$$\eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1), \qquad \mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu.$$

Here Greek indices range over $0, 1, 2, 3$; spatial indices range over $1, 2, 3$. The signature throughout is $(-, +, +, +)$.

12.2 Setting the speed of light to one

In SI units, $c \simeq 3 \times 10^8\,\mathrm{m\,s^{-1}}$. A **light-second** is the distance traveled by light in one second:

$$1 \text{ light-second} = c(1\text{ s}) \simeq 3 \times 10^8 \text{ m}.$$

If lengths are measured in light-seconds and times in seconds, the numerical value of the speed of light is one. Alternatively, keep the metre as a unit of length and measure time in the time light takes to travel one metre. Both choices express time and length in compatible units.

With this convention, we write $c = 1$ and

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{34}$$

The temporal coordinate is now understood in length units; an implicit factor of c must be restored when returning to standard numerical units. Velocities written in these units are measured as fractions of c. Choosing $c = 1$ is a unit convention, not a change in the physical speed of light.

Later it is also useful to set Newton's constant $G = 1$. The combination GM/c^2 has dimensions of length, so this convention permits mass to be expressed geometrically as a length. With $c = G = 1$, there is still one overall scale to choose. Setting $\hbar = 1$ as well fixes it, leading to Planck units. Restoring the constants, the corresponding basic scales are

$$\ell_{\mathrm{P}} = \sqrt{\frac{\hbar G}{c^3}}, \qquad t_{\mathrm{P}} = \sqrt{\frac{\hbar G}{c^5}}, \qquad m_{\mathrm{P}} = \sqrt{\frac{\hbar c}{G}}.$$

These satisfy $\ell_{\mathrm{P}}/t_{\mathrm{P}} = c$, $Gm_{\mathrm{P}}/\ell_{\mathrm{P}} = c^2$, and $m_{\mathrm{P}} c \ell_{\mathrm{P}} = \hbar$. They are often called natural or Planck units in quantum-gravity discussions. For the following spacetime calculations, only the convention $c = 1$ is needed.

13 Euclidean rotations as a model for spacetime symmetries

13.1 Translations and rotated axes

Before studying Lorentz boosts, consider the simpler positive-definite line element

$$\mathrm{d}\ell^2 = \mathrm{d}x^2 + \mathrm{d}y^2.$$

Translations $x' = x + a$, $y' = y + b$ immediately preserve it because $\mathrm{d}x' = \mathrm{d}x$ and $\mathrm{d}y' = \mathrm{d}y$. Rotations preserve it as well, but mix the two coordinate differentials.

Rotate the coordinate axes counterclockwise by a fixed angle α, while holding the geometric point P fixed. Its old and new coordinates are (x, y) and (x', y'). This is a passive coordinate rotation. It must be distinguished from rotating the point itself while holding the original axes fixed; that active operation uses the inverse coordinate matrix.

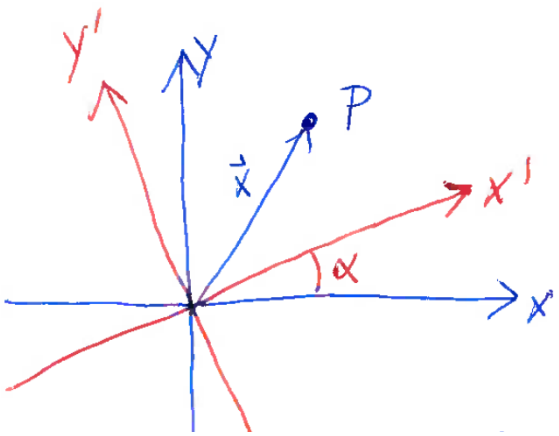

Figure 14: A fixed point P described using original and rotated Cartesian axes. The primed axes are rotated through α.

13.2 Derivation using orthonormal basis vectors

Let $\hat{\boldsymbol{e}}_x, \hat{\boldsymbol{e}}_y$ be the original unit vectors and $\hat{\boldsymbol{e}}_{x'}, \hat{\boldsymbol{e}}_{y'}$ the rotated ones. The geometry of the rotation gives

$$\begin{aligned} \hat{\boldsymbol{e}}_{x'} &= \cos\alpha\, \hat{\boldsymbol{e}}_x + \sin\alpha\, \hat{\boldsymbol{e}}_y, \\ \hat{\boldsymbol{e}}_{y'} &= -\sin\alpha\, \hat{\boldsymbol{e}}_x + \cos\alpha\, \hat{\boldsymbol{e}}_y. \end{aligned}$$

Relative to a chosen origin, the position vector is one geometric vector with two coordinate expansions:

$$\boldsymbol{r} = x\hat{\boldsymbol{e}}_x + y\hat{\boldsymbol{e}}_y = x'\hat{\boldsymbol{e}}_{x'} + y'\hat{\boldsymbol{e}}_{y'}.$$

Taking its Euclidean dot product with the primed unit vectors yields

$$\begin{aligned} x' &= \boldsymbol{r}\cdot\hat{\boldsymbol{e}}_{x'} = x\cos\alpha + y\sin\alpha, \\ y' &= \boldsymbol{r}\cdot\hat{\boldsymbol{e}}_{y'} = -x\sin\alpha + y\cos\alpha. \end{aligned}$$

Hence

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = R(\alpha)\begin{pmatrix} x \\ y \end{pmatrix}, \qquad R(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}. \tag{35}$$

The position-vector construction uses the affine structure of flat space and a selected origin. On a general curved manifold, coordinates of a point do not by themselves form an intrinsic position vector. The flat-space example should not be interpreted as asserting one.

13.3 The same transformation from projections on the axes

One can also determine the two columns of $R(\alpha)$ separately. First take P on the original x axis, with coordinates $(x, 0)$. Its projection on x' is $x\cos\alpha$, while its projection on y' is

$$-x\cos\left(\frac{\pi}{2} - \alpha\right) = -x\sin\alpha.$$

Next take P on the original y axis, with coordinates $(0, y)$. Its projections are

$$x' = y\cos\left(\frac{\pi}{2} - \alpha\right) = y\sin\alpha, \qquad y' = y\cos\alpha.$$

For a general point, add the contributions of $(x, 0)$ and $(0, y)$. Linearity reproduces both rows of (35). This axis-by-axis construction is a direct geometric way to read off a linear transformation matrix.

13.4 Checking invariance explicitly

Since α is constant,

$$\mathrm{d}x' = \cos\alpha\,\mathrm{d}x + \sin\alpha\,\mathrm{d}y, \qquad \mathrm{d}y' = -\sin\alpha\,\mathrm{d}x + \cos\alpha\,\mathrm{d}y.$$

Substitution gives

$$\begin{aligned}\mathrm{d}x'^2 + \mathrm{d}y'^2 &= (\cos\alpha\,\mathrm{d}x + \sin\alpha\,\mathrm{d}y)^2 + (-\sin\alpha\,\mathrm{d}x + \cos\alpha\,\mathrm{d}y)^2 \\ &= (\cos^2\alpha + \sin^2\alpha)\mathrm{d}x^2 + (\sin^2\alpha + \cos^2\alpha)\mathrm{d}y^2 \\ &\quad + (2\sin\alpha\cos\alpha - 2\sin\alpha\cos\alpha)\mathrm{d}x\mathrm{d}y \\ &= \mathrm{d}x^2 + \mathrm{d}y^2.\end{aligned}$$

The mixed terms cancel, and the coefficients of the squared differentials become one. Thus rotations are symmetries of the Euclidean line element. In matrix notation the same calculation is $R(\alpha)^{\mathsf{T}}R(\alpha) = I$.

13.5 Derivation using polar coordinates

A rotation becomes especially simple in polar coordinates:

$$x = r\cos\theta, \qquad y = r\sin\theta.$$

Rotating the axes through α leaves the distance from the origin unchanged and subtracts α from the coordinate angle:

$$r' = r, \qquad \theta' = \theta - \alpha. \tag{36}$$

Because $\mathrm{d}r' = \mathrm{d}r$ and $\mathrm{d}\theta' = \mathrm{d}\theta$,

$$\mathrm{d}r'^2 + r'^2\mathrm{d}\theta'^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2.$$

This proves invariance without expanding any cross terms. It also recovers the Cartesian transformation through the angle-subtraction identities:

$$\begin{aligned}x' &= r'\cos\theta' = r\cos(\theta - \alpha) \\ &= r(\cos\theta\cos\alpha + \sin\theta\sin\alpha) = x\cos\alpha + y\sin\alpha, \\ y' &= r'\sin\theta' = r\sin(\theta - \alpha) \\ &= r(\sin\theta\cos\alpha - \cos\theta\sin\alpha) = -x\sin\alpha + y\cos\alpha.\end{aligned}$$

The basis-vector, axis-projection, and polar-coordinate approaches therefore give the same transformation, with consistent signs.

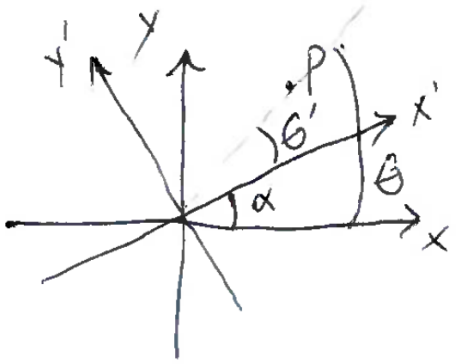

Figure 15: The angular coordinates of one point satisfy $\theta' = \theta - \alpha$ when the axes are rotated.

13.6 The rotation group

Successive rotations compose according to

$$R(\alpha)R(\beta) = R(\alpha + \beta). \tag{37}$$

For example, the first row of the product is

$$\big(\cos\alpha\cos\beta - \sin\alpha\sin\beta,\ \cos\alpha\sin\beta + \sin\alpha\cos\beta\big),$$

which is $(\cos(\alpha+\beta), \sin(\alpha+\beta))$. The second row follows similarly. Also,

$$R(0) = I, \qquad R(-\alpha) = R(\alpha)^{-1} = R(\alpha)^{\mathsf{T}}.$$

Matrix multiplication is associative, so these transformations form a group. Its name is $SO(2)$:

$$SO(2) = \{R \in \mathbb{R}^{2\times 2} : R^{\mathsf{T}}R = I,\ \det R = 1\}.$$

The word "orthogonal" refers to $R^{-1} = R^{\mathsf{T}}$; "special" refers to determinant $+1$. Directly, $\det R(\alpha) = \cos^2\alpha + \sin^2\alpha = 1$. The integer two is the size of the matrices and the dimension of space on which they act. The angle is periodic modulo 2π.

Thus $SO(2)$ is the orientation-preserving rotation subgroup of the symmetries of the plane. The full Euclidean line element also admits translations, as noted above, and reflections if orientation reversal is allowed. It is important not to identify its entire isometry group with the rotation subgroup alone.

For three-dimensional flat space,

$$\mathrm{d}\ell^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2,$$

rotations are represented by 3×3 real matrices satisfying $R^{\mathsf{T}}R = I$ and $\det R = 1$. They form $SO(3)$. A general spatial rotation requires three continuous parameters. Rotations about the coordinate axes give useful elementary examples; a rotation about z uses the 2×2 block in (35) and leaves z unchanged. The same orthogonality calculation proves invariance of the three-dimensional line element. The essential comparison for spacetime is that ordinary rotations preserve a sum of squares, while boosts will preserve a difference of squares.

14 Hyperbolic geometry in spacetime

Lectures 8–9

14.1 Why begin with one space and one time dimension?

The notation $1+1$ **dimensions** means one time dimension and one space dimension. Our physical spacetime has 3+1 dimensions, but the main difference between a spatial rotation and a spacetime boost is already visible in the two-dimensional line element

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2. \tag{38}$$

We continue to use units in which $c = 1$. A transformation is an isometry when it preserves this line element. Constant translations provide immediate examples:

$$t' = t + a, \qquad x' = x + b \quad \Longrightarrow \quad \mathrm{d}t' = \mathrm{d}t, \qquad \mathrm{d}x' = \mathrm{d}x.$$

Neither metric component depends on t or x, so these translations leave the metric unchanged. The less obvious isometries mix space and time. To find them, it is helpful to revisit the geometry behind ordinary rotations.

14.2 Circular orbits and the angle coordinate

A rotation about the origin of the Euclidean plane preserves $x^2 + y^2$. Thus the orbit of a point at distance $r > 0$ is the circle

$$x^2 + y^2 = r^2.$$

Writing $X = x/r$ and $Y = y/r$ reduces this to the unit circle $X^2 + Y^2 = 1$. Its usual parametrization is

$$X = \cos\theta, \qquad Y = \sin\theta,$$

and therefore

$$x = r\cos\theta, \qquad y = r\sin\theta.$$

Angles differing by 2π label the same point. A rotation changes θ by a constant and leaves r fixed.

There is also a metric interpretation of the angle. Along a circle of fixed radius, the Euclidean polar line element becomes

$$\mathrm{d}\ell^2 = r^2\mathrm{d}\theta^2, \qquad \mathrm{d}\ell = r|\mathrm{d}\theta|.$$

Consequently, an arc through an angle $|\Delta\theta|$ has length $r|\Delta\theta|$. The angle parametrizes the orbit at a constant rate with respect to arc length. More generally,

$$\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2$$

is independent of θ, making the symmetry under constant shifts of θ explicit. We seek the corresponding orbit and parameter for a metric with a minus sign.

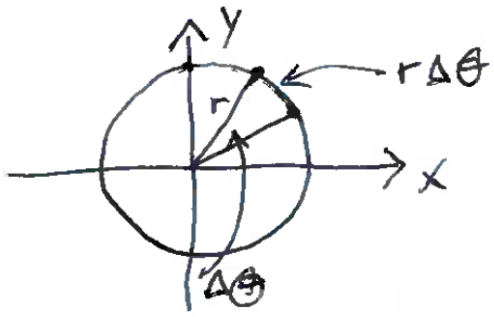

Figure 16: A circular arc in the original lecture sketch. At fixed radius, its length is proportional to the change in angle.

14.3 Why ordinary polar coordinates do not expose boost symmetry

One can introduce ordinary polar coordinates in the (t, x) coordinate plane, but they are not adapted to the Minkowski metric. To see the problem explicitly, set

$$t = R\cos\theta, \qquad x = R\sin\theta.$$

Their differentials are

$$\begin{aligned} \mathrm{d}t &= \cos\theta\,\mathrm{d}R - R\sin\theta\,\mathrm{d}\theta, \\ \mathrm{d}x &= \sin\theta\,\mathrm{d}R + R\cos\theta\,\mathrm{d}\theta. \end{aligned}$$

Substituting into (38) gives

$$\begin{aligned} \mathrm{d}s^2 &= -\big(\cos\theta\,\mathrm{d}R - R\sin\theta\,\mathrm{d}\theta\big)^2 + \big(\sin\theta\,\mathrm{d}R + R\cos\theta\,\mathrm{d}\theta\big)^2 \\ &= (\sin^2\theta - \cos^2\theta)\,\mathrm{d}R^2 + R^2(\cos^2\theta - \sin^2\theta)\,\mathrm{d}\theta^2 \\ &\quad + 4R\sin\theta\cos\theta\,\mathrm{d}R\,\mathrm{d}\theta \\ &= -\cos(2\theta)\,\mathrm{d}R^2 + R^2\cos(2\theta)\,\mathrm{d}\theta^2 + 2R\sin(2\theta)\,\mathrm{d}R\,\mathrm{d}\theta. \end{aligned}$$

The cross terms now add instead of cancelling. Moreover, all the metric components depend on θ, so an angular shift at fixed R is not an isometry. These are legitimate local coordinates away from their usual polar-coordinate limitations; they simply obscure the symmetry we want. The invariant curves for a difference of squares are hyperbolas, not circles.

14.4 The two families of hyperbolas

There are two signs to consider. For a positive constant T, the curves

$$t^2 - x^2 = T^2 \tag{39}$$

have two branches,

$$t = +\sqrt{x^2 + T^2} \qquad \text{and} \qquad t = -\sqrt{x^2 + T^2}.$$

They meet the time axis at $t = +T$ and $t = -T$, respectively. Since $t^2 > x^2$, the upper branch lies in the region $t > |x|$ and the lower branch in $t < -|x|$. For large $|x|$, each branch approaches one of the straight lines $t = \pm x$.

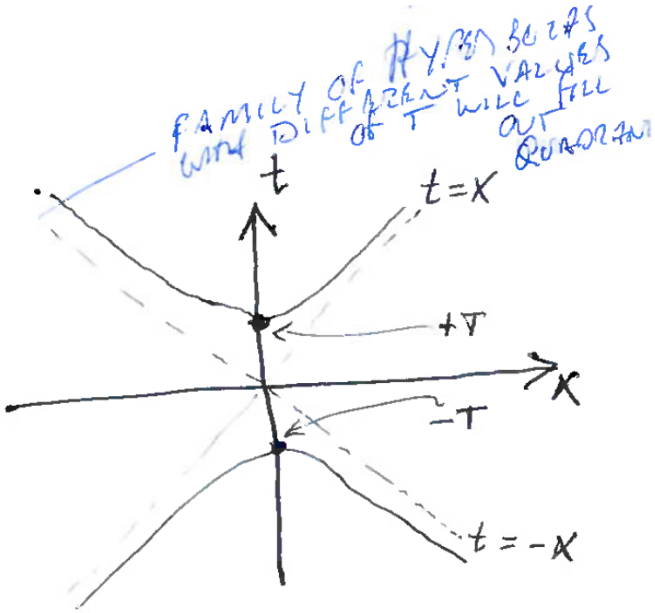

Figure 17: The future and past branches of $t^2 - x^2 = T^2$. Varying T fills the corresponding wedges with hyperbolas.

For a positive constant R, the other family is

$$x^2 - t^2 = R^2. \tag{40}$$

Its branches are

$$x = +\sqrt{t^2 + R^2} \qquad \text{and} \qquad x = -\sqrt{t^2 + R^2}.$$

They meet the space axis at $x = +R$ and $x = -R$. The right branch lies in $x > |t|$, the left branch in $x < -|t|$, and they have the same asymptotes $t = \pm x$.

The two null lines $t = \pm x$ thus divide the plane into four wedges. The future and past wedges contain the first family of hyperbolas; the right and left wedges contain the second. A boost preserves the value of $-t^2 + x^2$, so a continuously varying boost moves a point along one of these hyperbola branches.

14.5 Hyperbolic functions

The appropriate analogues of sine and cosine are

$$\cosh\theta = \frac{e^{\theta} + e^{-\theta}}{2}, \qquad \sinh\theta = \frac{e^{\theta} - e^{-\theta}}{2}. \tag{41}$$

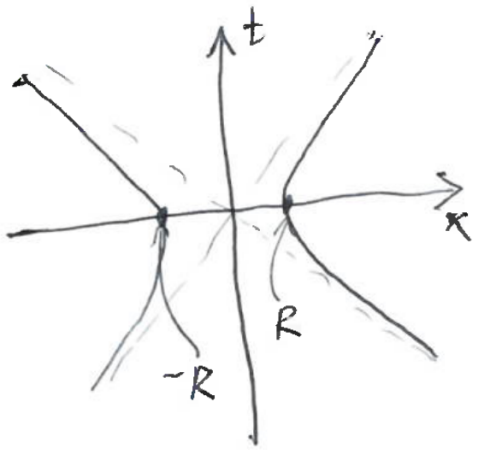

Figure 18: The two branches of $x^2 - t^2 = R^2$, with asymptotes on the null lines.

They obey

$$\cosh^2\theta - \sinh^2\theta = \frac{(\mathrm{e}^{\theta} + \mathrm{e}^{-\theta})^2 - (\mathrm{e}^{\theta} - \mathrm{e}^{-\theta})^2}{4} = \frac{4\mathrm{e}^{\theta}\mathrm{e}^{-\theta}}{4} = 1.$$

This is the difference-of-squares identity that replaces $\cos^2\theta + \sin^2\theta = 1$. The derivatives follow directly from the exponential definitions:

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\cosh\theta = \frac{\mathrm{e}^{\theta} - \mathrm{e}^{-\theta}}{2} = \sinh\theta, \qquad \frac{\mathrm{d}}{\mathrm{d}\theta}\sinh\theta = \frac{\mathrm{e}^{\theta} + \mathrm{e}^{-\theta}}{2} = \cosh\theta.$$

Both signs in these derivative relations are positive. This sign pattern is responsible for the cancellation of cross terms in the adapted spacetime coordinates.

15 Milne coordinates

15.1 Parametrizing the future hyperbolas

On the upper branch of (39), use

$$t = T\cosh\theta, \qquad x = T\sinh\theta, \qquad T > 0, \quad -\infty < \theta < \infty. \tag{42}$$

The identity above immediately gives

$$t^2 - x^2 = T^2(\cosh^2\theta - \sinh^2\theta) = T^2.$$

Since $\cosh\theta > 0$, this choice has $t > 0$ and covers precisely the future wedge $t > |x|$. The parameter θ is a **hyperbolic angle**; unlike an ordinary angle, it is not periodic.

The inverse transformation is also useful:

$$T = \sqrt{t^2 - x^2}, \qquad \theta = \operatorname{arctanh}\left(\frac{x}{t}\right). \tag{43}$$

Indeed, $x/t = \tanh\theta$, and $|x/t| < 1$ in this wedge, so the real inverse hyperbolic tangent is well defined. The Jacobian is

$$\det \frac{\partial(t,x)}{\partial(T,\theta)} = \det \begin{pmatrix} \cosh\theta & T\sinh\theta \\ \sinh\theta & T\cosh\theta \end{pmatrix} = T.$$

It is nonzero throughout $T > 0$, establishing that these are valid local coordinates there.

15.2 Deriving the Milne line element

Differentiating (42), with both coordinates allowed to vary, gives

$$\begin{aligned} \mathrm{d}t &= \cosh\theta\,\mathrm{d}T + T\sinh\theta\,\mathrm{d}\theta, \\ \mathrm{d}x &= \sinh\theta\,\mathrm{d}T + T\cosh\theta\,\mathrm{d}\theta. \end{aligned}$$

Therefore

$$\begin{aligned} \mathrm{d}s^2 &= -\big(\cosh\theta\,\mathrm{d}T + T\sinh\theta\,\mathrm{d}\theta\big)^2 + \big(\sinh\theta\,\mathrm{d}T + T\cosh\theta\,\mathrm{d}\theta\big)^2 \\ &= \big(-\cosh^2\theta + \sinh^2\theta\big)\mathrm{d}T^2 + T^2\big(\cosh^2\theta - \sinh^2\theta\big)\mathrm{d}\theta^2 \\ &\quad + \big(-2T\cosh\theta\sinh\theta + 2T\sinh\theta\cosh\theta\big)\mathrm{d}T\,\mathrm{d}\theta. \end{aligned}$$

The mixed coefficient vanishes, and the two remaining coefficients simplify to

$$\boxed{\mathrm{d}s^2 = -\mathrm{d}T^2 + T^2\mathrm{d}\theta^2.} \tag{44}$$

These are **Milne coordinates** on flat spacetime. The factor T^2 does not imply that spacetime is curved: the coordinate change was performed on the Minkowski line element, without altering its geometry.

The metric has no θ dependence. Consequently,

$$T' = T, \qquad \theta' = \theta - \alpha,$$

with constant α, preserves (44). This is the symmetry that ordinary polar coordinates failed to reveal. In the original (t,x) coordinates it will become a Lorentz boost.

15.3 Coordinate lines and an expanding family of observers

A constant-T coordinate line is one of the future hyperbolas. A constant-θ coordinate line obeys

$$\frac{x}{t} = \tanh\theta,$$

and hence is a straight ray through the origin lying inside the future wedge. The ray $\theta = 0$ is the positive t axis. As θ increases to $+\infty$, the ray approaches $x = t$; as θ decreases to $-\infty$, it approaches $x = -t$. Equivalently, for a nonzero value of θ, one can write $t/x = \coth\theta$.

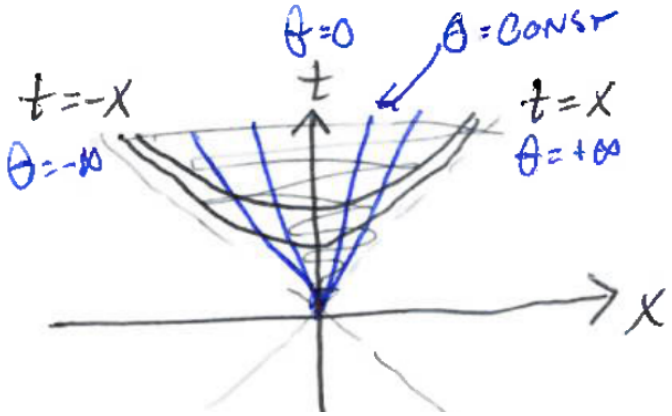

Figure 19: Milne coordinate lines in the future wedge: constant-T hyperbolas and constant-θ straight rays.

The metric also identifies the roles of the new coordinates. Along a constant-θ worldline,

$$\mathrm{d}s^2 = -\mathrm{d}T^2.$$

Thus T measures the proper time of an observer following such a ray. The coefficient of $\mathrm{d}\theta^2$ is positive, so separation in θ at fixed T is spatial. Between two fixed coordinate values θ_1 and θ_2, its length on the constant-T slice is

$$\ell(T) = \int_{\theta_1}^{\theta_2} T\,|\mathrm{d}\theta| = T|\theta_2 - \theta_1|.$$

This distance grows linearly with T. Milne coordinates therefore provide a simple cosmological description of an expanding collection of inertial observers. The expansion refers to the chosen congruence of worldlines and time slices; the underlying spacetime remains flat. Time-dependent spatial distances will also appear when we study the interior geometry of a Schwarzschild black hole.

15.4 The past wedge

The future chart does not cover the lower hyperbolas. A second chart for the past wedge is

$$t = -T\cosh\theta, \qquad x = T\sinh\theta, \qquad T > 0, \quad \theta \in \mathbb{R}. \tag{45}$$

Here $t < -|x|$. The minus sign on t does not affect the squared contribution to the line element, so again

$$\mathrm{d}s^2 = -\mathrm{d}T^2 + T^2\mathrm{d}\theta^2.$$

The constant-θ lines obey $x/t = -\tanh\theta$. In this chart, increasing T moves farther into the past, so the coordinate's chosen orientation must be distinguished from future-directed proper-time orientation. Its magnitude still measures proper-time separation along each of these straight worldlines.

16 Rindler coordinates in the right and left wedges

16.1 The right wedge and its metric

The second family of hyperbolas suggests the parametrization

$$x = R\cosh\theta, \qquad t = R\sinh\theta, \qquad R > 0, \quad \theta \in \mathbb{R}. \tag{46}$$

Now

$$x^2 - t^2 = R^2,$$

and $x > 0$, so this chart covers the right wedge $x > |t|$. Its inverse is

$$R = \sqrt{x^2 - t^2}, \qquad \theta = \operatorname{arctanh}\left(\frac{t}{x}\right).$$

The differentials are

$$\begin{aligned} \mathrm{d}x &= \cosh\theta\,\mathrm{d}R + R\sinh\theta\,\mathrm{d}\theta, \\ \mathrm{d}t &= \sinh\theta\,\mathrm{d}R + R\cosh\theta\,\mathrm{d}\theta. \end{aligned}$$

Substituting and collecting all three types of terms yields

$$\begin{aligned} \mathrm{d}s^2 &= \big(\cosh^2\theta - \sinh^2\theta\big)\mathrm{d}R^2 + R^2\big(\sinh^2\theta - \cosh^2\theta\big)\mathrm{d}\theta^2 \\ &\quad + \big(2R\sinh\theta\cosh\theta - 2R\cosh\theta\sinh\theta\big)\mathrm{d}R\,\mathrm{d}\theta \\ &= \mathrm{d}R^2 - R^2\mathrm{d}\theta^2. \end{aligned}$$

Thus the **Rindler line element** is

$$\boxed{\mathrm{d}s^2 = -R^2\mathrm{d}\theta^2 + \mathrm{d}R^2.} \tag{47}$$

It is another expression for the same flat Minkowski metric.

The coordinate roles have changed compared with the Milne chart. The negative coefficient now multiplies $\mathrm{d}\theta^2$, making θ a time coordinate, while R is spatial. A constant-R worldline traces a hyperbola; a constant-θ line satisfies

$$\frac{t}{x} = \tanh\theta$$

and is a straight ray in the right wedge. At $\theta = 0$ the ray lies on the positive x axis. Along a constant-R worldline,

$$\mathrm{d}\tau = R\,\mathrm{d}\theta$$

for increasing future-directed θ. Therefore equal changes of coordinate time θ correspond to different proper-time changes on different constant-R worldlines. As in the Milne chart, the absence of θ from the metric makes shifts in θ an isometry.

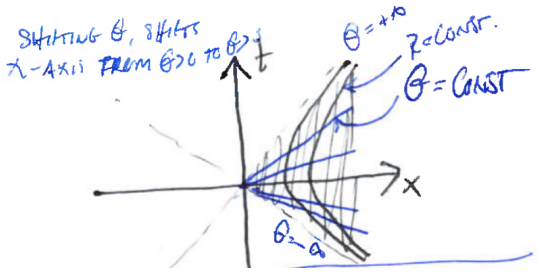

Figure 20: Rindler coordinate lines in the right wedge. Constant-R curves are hyperbolas; constant-θ curves are straight rays.

16.2 The left wedge

To cover $x < -|t|$, choose

$$x = -R\cosh\theta, \qquad t = R\sinh\theta, \qquad R > 0, \quad \theta \in \mathbb{R}. \tag{48}$$

The same calculation gives

$$\mathrm{d}s^2 = -R^2\mathrm{d}\theta^2 + \mathrm{d}R^2.$$

The hyperbolas are now the left branches of $x^2 - t^2 = R^2$. The corresponding straight rays obey $t/x = -\tanh\theta$.

17 Coordinate patches and coordinate singularities

The four charts can be summarized by their domains and the signs in the transformations. In every row, the radial coordinate is positive and θ ranges over all real numbers.

Region	Transformation	Line element
$t > \lvert x\rvert$	$t = T\cosh\theta,\ x = T\sinh\theta$	$-\mathrm{d}T^2 + T^2\mathrm{d}\theta^2$
$t < -\lvert x\rvert$	$t = -T\cosh\theta,\ x = T\sinh\theta$	$-\mathrm{d}T^2 + T^2\mathrm{d}\theta^2$
$x > \lvert t\rvert$	$t = R\sinh\theta,\ x = R\cosh\theta$	$-R^2\mathrm{d}\theta^2 + \mathrm{d}R^2$
$x < -\lvert t\rvert$	$t = R\sinh\theta,\ x = -R\cosh\theta$	$-R^2\mathrm{d}\theta^2 + \mathrm{d}R^2$

Together they cover the Minkowski plane except for its two null lines, $|t| = |x|$. The need for several charts is a limitation of these particular coordinates, not a division of spacetime into disconnected pieces. The original Cartesian coordinates cover all four wedges and the boundaries between them.

Why do the hyperbolic charts fail on the boundaries? In the future Milne wedge,

$$T^2 = t^2 - x^2,$$

so approaching a null line sends T to zero. However, setting $T = 0$ while keeping θ finite in (42) gives only the origin. Reaching a nonzero point on a null line requires a

simultaneous limit $T \to 0$ and $|\theta| \to \infty$. Likewise, a Rindler chart reaches a nonzero point on a null boundary only through a limit with $R \to 0$ and $|\theta| \to \infty$.

The failure is visible in the metric determinants:

$$\det(g_{ab}) = -T^2 \quad \text{in Milne coordinates,} \qquad \det(g_{ab}) = -R^2 \quad \text{in Rindler coordinates.}$$

At zero radial coordinate the formal expressions become degenerate. They do not define an invertible metric there because the coordinate transformation itself has ceased to be invertible. In Cartesian Minkowski coordinates the determinant remains -1 everywhere, so there is no geometric singularity on either null line.

The Euclidean polar metric supplies a familiar comparison:

$$\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2, \qquad \det(g_{ab}) = r^2.$$

At $r = 0$, every value of θ describes the same point, and the angular coordinate becomes meaningless. The plane is nevertheless perfectly regular at the origin. A vanishing or divergent metric component in a particular chart therefore cannot by itself establish a physical singularity.

This distinction becomes essential in black-hole geometry. A coordinate system can break down at an event horizon even when the spacetime geometry there is regular. Suitable alternative coordinates can remove that coordinate singularity. This is a different issue from a genuine curvature singularity, which cannot be eliminated by changing coordinates.

18 Lorentz boosts as hyperbolic rotations

18.1 Useful identities and small-parameter expansions

Some relativity texts abbreviate $\cosh\alpha$ and $\sinh\alpha$ as $\mathrm{ch}\,\alpha$ and $\mathrm{sh}\,\alpha$. We retain the full notation. The exponential definitions imply

$$\mathrm{e}^{\alpha} = \cosh\alpha + \sinh\alpha, \qquad \mathrm{e}^{-\alpha} = \cosh\alpha - \sinh\alpha.$$

Expanding the exponentials separates their even and odd terms:

$$\cosh\alpha = 1 + \frac{\alpha^2}{2!} + \frac{\alpha^4}{4!} + \cdots,$$
$$\sinh\alpha = \alpha + \frac{\alpha^3}{3!} + \frac{\alpha^5}{5!} + \cdots.$$

In particular,

$$\cosh(-\alpha) = \cosh\alpha, \qquad \sinh(-\alpha) = -\sinh\alpha.$$

The addition formulas can be obtained by multiplying exponentials. For example,

$$\begin{aligned}\cosh(\theta-\alpha) &= \frac{\mathrm{e}^{\theta}\mathrm{e}^{-\alpha} + \mathrm{e}^{-\theta}\mathrm{e}^{\alpha}}{2} \\ &= \cosh\theta\cosh\alpha - \sinh\theta\sinh\alpha,\end{aligned}$$

and the difference of the same exponential products gives

$$\sinh(\theta-\alpha) = \sinh\theta\cosh\alpha - \cosh\theta\sinh\alpha.$$

18.2 From a hyperbolic-angle shift to a linear transformation

Apply the Milne-coordinate isometry

$$T' = T, \qquad \theta' = \theta - \alpha.$$

Using the addition formulas,

$$\begin{aligned} t' &= T\cosh(\theta-\alpha) \\ &= T\cosh\theta\cosh\alpha - T\sinh\theta\sinh\alpha \\ &= t\cosh\alpha - x\sinh\alpha, \\ x' &= T\sinh(\theta-\alpha) \\ &= T\sinh\theta\cosh\alpha - T\cosh\theta\sinh\alpha \\ &= x\cosh\alpha - t\sinh\alpha. \end{aligned}$$

Thus the transformation in Cartesian spacetime coordinates is

$$\boxed{\begin{aligned} t' &= \cosh\alpha\, t - \sinh\alpha\, x, \\ x' &= -\sinh\alpha\, t + \cosh\alpha\, x. \end{aligned}} \tag{49}$$

This is a **Lorentz boost**, and α is its **rapidity**. Although the derivation used the future wedge, the linear transformation itself is defined on the entire Minkowski plane. Its metric-preserving property can be checked without using any hyperbolic coordinate chart.

Exercise 18.1 (Invariance under a boost). Verify directly that (49) preserves $\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2$.

Solution. Because α is constant, differentiation gives

$$\mathrm{d}t' = \cosh\alpha\,\mathrm{d}t - \sinh\alpha\,\mathrm{d}x, \qquad \mathrm{d}x' = -\sinh\alpha\,\mathrm{d}t + \cosh\alpha\,\mathrm{d}x.$$

Hence

$$\begin{aligned} -\mathrm{d}t'^2 + \mathrm{d}x'^2 &= -\big(\cosh\alpha\,\mathrm{d}t - \sinh\alpha\,\mathrm{d}x\big)^2 + \big(-\sinh\alpha\,\mathrm{d}t + \cosh\alpha\,\mathrm{d}x\big)^2 \\ &= \big(-\cosh^2\alpha + \sinh^2\alpha\big)\mathrm{d}t^2 + \big(\cosh^2\alpha - \sinh^2\alpha\big)\mathrm{d}x^2 \\ &\quad + \big(2\cosh\alpha\sinh\alpha - 2\sinh\alpha\cosh\alpha\big)\mathrm{d}t\,\mathrm{d}x \\ &= -\mathrm{d}t^2 + \mathrm{d}x^2. \end{aligned}$$

The mixed terms cancel, and the remaining terms use $\cosh^2\alpha - \sinh^2\alpha = 1$. □

In matrix notation, the same statement is

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = B(\alpha)\begin{pmatrix} t \\ x \end{pmatrix}, \qquad B(\alpha) = \begin{pmatrix} \cosh\alpha & -\sinh\alpha \\ -\sinh\alpha & \cosh\alpha \end{pmatrix},$$

with

$$B(\alpha)^{\mathsf{T}}\eta B(\alpha) = \eta, \qquad \eta = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Its determinant is $\cosh^2\alpha - \sinh^2\alpha = 1$. The inverse is obtained by reversing the rapidity, $B(\alpha)^{-1} = B(-\alpha)$. Successive boosts along this same spatial axis add their rapidities:

$$B(\alpha_1)B(\alpha_2) = B(\alpha_1 + \alpha_2).$$

This follows either by multiplying the two matrices and applying the hyperbolic addition formulas, or directly by composing the shifts $\theta \mapsto \theta - \alpha_2 \mapsto \theta - \alpha_2 - \alpha_1$. Ordinary rotations preserve the Euclidean identity matrix, whereas Lorentz boosts preserve the Minkowski matrix η.

19 Lorentz boosts, rapidity, and inertial frames

Lecture 10

19.1 The boost matrix and its invariant quadratic form

In one spatial dimension, with $c = 1$, the boost transformation takes the form

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = B(\alpha) \begin{pmatrix} t \\ x \end{pmatrix}, \qquad B(\alpha) = \begin{pmatrix} \cosh\alpha & -\sinh\alpha \\ -\sinh\alpha & \cosh\alpha \end{pmatrix}. \tag{50}$$

The parameter α is a hyperbolic angle. The defining property of the transformation is preservation of the Minkowski line element $\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2$. In matrix notation, write

$$\eta = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathrm{d}s^2 = \begin{pmatrix} \mathrm{d}t & \mathrm{d}x \end{pmatrix} \eta \begin{pmatrix} \mathrm{d}t \\ \mathrm{d}x \end{pmatrix}.$$

The previously established invariance is equivalently

$$B(\alpha)^{\mathsf{T}} \eta B(\alpha) = \eta. \tag{51}$$

This is the indefinite-metric counterpart of the Euclidean orthogonality condition $R^{\mathsf{T}}R = I$. The minus sign in η is responsible for the appearance of hyperbolic functions instead of circular functions.

To see the cancellation directly, put $C = \cosh\alpha$ and $S = \sinh\alpha$. Then

$$\begin{aligned} -(\mathrm{d}t')^2 + (\mathrm{d}x')^2 &= -(C\mathrm{d}t - S\mathrm{d}x)^2 + (C\mathrm{d}x - S\mathrm{d}t)^2 \\ &= -(C^2 - S^2)\mathrm{d}t^2 + (C^2 - S^2)\mathrm{d}x^2 \\ &\quad + (2CS - 2SC)\mathrm{d}t\,\mathrm{d}x \\ &= -\mathrm{d}t^2 + \mathrm{d}x^2, \end{aligned}$$

using $\cosh^2\alpha - \sinh^2\alpha = 1$.

The boost matrices form a one-parameter group. First, $B(0) = I$. Next, matrix multiplication gives

$$B(\alpha)B(\beta) = \begin{pmatrix} C_\alpha C_\beta + S_\alpha S_\beta & -(C_\alpha S_\beta + S_\alpha C_\beta) \\ -(S_\alpha C_\beta + C_\alpha S_\beta) & S_\alpha S_\beta + C_\alpha C_\beta \end{pmatrix}.$$

The hyperbolic addition identities

$$\begin{aligned} \cosh(\alpha + \beta) &= \cosh\alpha\cosh\beta + \sinh\alpha\sinh\beta, \\ \sinh(\alpha + \beta) &= \sinh\alpha\cosh\beta + \cosh\alpha\sinh\beta \end{aligned}$$

therefore imply

$$B(\alpha)B(\beta) = B(\alpha+\beta), \qquad B(\alpha)^{-1} = B(-\alpha). \tag{52}$$

Associativity is inherited from matrix multiplication. Since $\det B(\alpha) = 1$ and $\cosh\alpha > 0$, these boosts form the component of $\mathrm{SO}(1,1)$ connected to the identity, often written $\mathrm{SO}^+(1,1)$. Reflections or time reversal are not generated by varying the real rapidity continuously.

19.2 A boost as a translation of the hyperbolic angle

In the future timelike wedge, introduce Milne coordinates

$$t = T\cosh\theta, \qquad x = T\sinh\theta, \qquad T > 0. \tag{53}$$

The line element is

$$\mathrm{d}s^2 = -\mathrm{d}T^2 + T^2\mathrm{d}\theta^2. \tag{54}$$

It does not depend explicitly on θ. Consequently the coordinate change

$$T' = T, \qquad \theta' = \theta - \alpha \tag{55}$$

leaves the line element unchanged: $\mathrm{d}T' = \mathrm{d}T$ and $\mathrm{d}\theta' = \mathrm{d}\theta$. Transforming back to Cartesian coordinates gives

$$\begin{aligned} t' &= T'\cosh\theta' = T\cosh(\theta-\alpha) \\ &= T\cosh\theta\cosh\alpha - T\sinh\theta\sinh\alpha \\ &= t\cosh\alpha - x\sinh\alpha, \\ x' &= T'\sinh\theta' = T\sinh(\theta-\alpha) \\ &= T\sinh\theta\cosh\alpha - T\cosh\theta\sinh\alpha \\ &= x\cosh\alpha - t\sinh\alpha. \end{aligned}$$

Thus a boost is a shift of hyperbolic angle, just as a spatial rotation is a shift of polar angle. The Milne chart covers only its timelike wedge, but the linear boost formula extends to all of Minkowski spacetime.

19.3 The Lorentz and Poincaré groups in four dimensions

In $3+1$ dimensions the Cartesian Minkowski line element is

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{56}$$

There are three independent infinitesimal boost directions, along x, y, z, and three independent spatial rotation planes, xy, yz, zx. Equivalently, the six generators are associated with the six coordinate planes tx, ty, tz, xy, yz, zx. Transformations in a plane containing time are hyperbolic rotations; transformations in purely spatial planes are ordinary rotations.

Writing $x^\mu = (t, x, y, z)$, a Lorentz transformation is linear:

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu, \qquad \Lambda^{\mathsf{T}} \eta \Lambda = \eta, \qquad \eta = \mathrm{diag}(-1, 1, 1, 1). \tag{57}$$

The full group satisfying this condition is $\mathrm{O}(1,3)$. The determinant-one subgroup is $\mathrm{SO}(1,3)$, also denoted $\mathrm{SO}(3,1)$. Boosts and spatial rotations connected to the identity generate its proper orthochronous component. The six-parameter count refers to continuous transformations; discrete reflections do not add continuous parameters.

There are also four translation symmetries. For example, $x' = x + a$ with t, y, z unchanged preserves the line element because $\mathrm{d}x' = \mathrm{d}x$. Combining translations with Lorentz transformations gives

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu. \tag{58}$$

These form the ten-dimensional Poincaré group, denoted $\mathrm{ISO}(1,3)$ or $\mathrm{ISO}(3,1)$, with the desired discrete components specified when necessary. The word *inhomogeneous* refers to the translation term: the transformation is affine rather than linear when $a^\mu \neq 0$.

Its composition rule follows by substitution:

$$(\Lambda_2, a_2) \circ (\Lambda_1, a_1) = (\Lambda_2 \Lambda_1, \Lambda_2 a_1 + a_2).$$

This is a semidirect product of translations and Lorentz transformations. In four dimensions, a metric can have at most $4(4+1)/2 = 10$ independent continuous isometries. Minkowski spacetime attains this maximum.

19.4 Relating rapidity to relative velocity

Consider two inertial frames with relative motion in the positive x direction. The spatial origin of the primed frame has $x' = 0$. Equation (50) then gives

$$0 = x \cosh\alpha - t \sinh\alpha \quad \Longrightarrow \quad \frac{x}{t} = \tanh\alpha.$$

Thus, in units $c = 1$, the relative velocity is

$$v = \tanh\alpha, \qquad \cosh\alpha = \frac{1}{\sqrt{1 - v^2}}, \qquad \sinh\alpha = \frac{v}{\sqrt{1 - v^2}}. \tag{59}$$

The last two expressions obey $\cosh^2\alpha - \sinh^2\alpha = 1$ and have the signs required for either positive or negative v.

Substitution into the hyperbolic formulas gives

$$t' = \frac{t - vx}{\sqrt{1 - v^2}}, \qquad x' = \frac{x - vt}{\sqrt{1 - v^2}}. \tag{60}$$

To restore units, use $x^0 = ct$ and replace the dimensionless velocity by v/c. Defining

$$\beta = \frac{v}{c}, \qquad \gamma = \frac{1}{\sqrt{1 - \beta^2}},$$

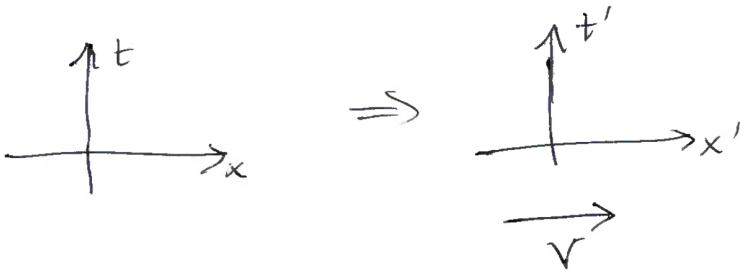

Figure 21: Two inertial coordinate systems with relative velocity in the positive x direction.

we obtain

$$ct' = \gamma(ct - \beta x), \qquad x' = \gamma(x - \beta ct), \tag{61}$$

$$t' = \gamma\left(t - \frac{vx}{c^2}\right), \qquad x' = \gamma(x - vt). \tag{62}$$

The parameter α is called the **rapidity**, with

$$\frac{v}{c} = \tanh\alpha. \tag{63}$$

A finite real rapidity always gives $|v| < c$. As $\alpha \to \pm\infty$, $v \to \pm c$ and the boost matrix entries diverge. There is no finite Lorentz transformation to the rest frame of a light ray.

Rapidity is additive for collinear boosts. If $v_i/c = \tanh\alpha_i$, then

$$\frac{v_{\text{combined}}}{c} = \tanh(\alpha_1 + \alpha_2) = \frac{v_1/c + v_2/c}{1 + v_1 v_2/c^2}.$$

Thus $v_{\text{combined}} = (v_1 + v_2)/(1 + v_1 v_2/c^2)$; the group law is simplest when expressed directly in rapidities.

19.5 The Galilean limit

For $|v|/c \ll 1$,

$$\gamma = 1 + \frac{1}{2}\frac{v^2}{c^2} + O\left(\frac{v^4}{c^4}\right).$$

The spatial transformation becomes $x' = x - vt + O(c^{-2})$ for fixed x, t, v, and

$$t' = t - \frac{vx}{c^2} + \frac{v^2 t}{2c^2} + O(c^{-4}).$$

In the Newtonian limit $c \to \infty$ with ordinary distances, velocities, and times fixed,

$$x' \longrightarrow x - vt, \qquad t' \longrightarrow t. \tag{64}$$

Absolute Newtonian time emerges as the leading approximation. At finite c, the term $-vx/c^2$ retains the relativity of simultaneity, even though it is negligible in many slow-motion experiments.

19.6 How the axes transform on a spacetime diagram

For an ordinary rotation of the Euclidean plane, the coordinate axes turn together through an angle and remain perpendicular in the Euclidean sense.

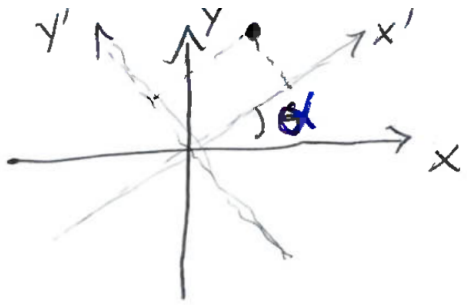

Figure 22: An ordinary spatial rotation of the coordinate axes.

A spacetime boost has a different appearance on a drawing whose vertical axis is t and horizontal axis is x. Set $c = 1$ again. The primed time axis is the locus $x' = 0$, while the primed spatial axis is $t' = 0$. From (60),

$$t' \text{ axis: } x = vt, \qquad x' \text{ axis: } t = vx. \tag{65}$$

For $0 < v < 1$, the t' axis leans right and the x' axis leans upward: the axes scissor toward the null line $t = x$. They remain orthogonal with respect to the Minkowski metric; Euclidean angles on the paper do not measure Lorentzian orthogonality.

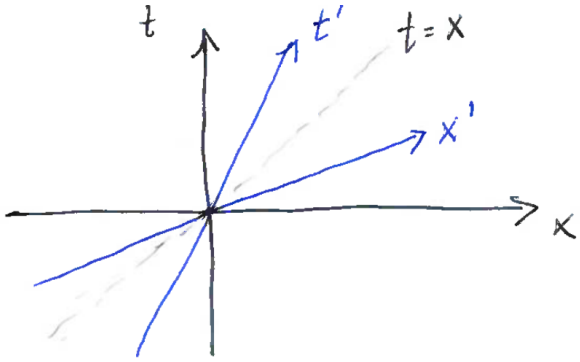

Figure 23: The primed axes scissor toward the null line under a positive boost.

The invariance of the speed of light is explicit. On $x = t$,

$$x' = \gamma(1 - v)t, \qquad t' = \gamma(1 - v)t,$$

so $x' = t'$. Likewise $x = -t$ transforms to $x' = -t'$. Both families of null rays remain null, and every inertial observer assigns them the same speed.

The hyperbolic-angle description makes the scissoring geometric. In the future timelike wedge, a constant-θ line has $x/t = \tanh\theta$. The unprimed time axis has

$\theta = 0$, while the primed time axis has $\theta' = 0$, hence $\theta = \alpha$. In the right spacelike wedge use

$$x = R\cosh\theta, \qquad t = R\sinh\theta.$$

The boost again changes θ to $\theta - \alpha$. Now $\theta = 0$ is the x axis, and $\theta = \alpha$ is the x' axis. The same hyperbolic-angle shift determines both tilted axes.

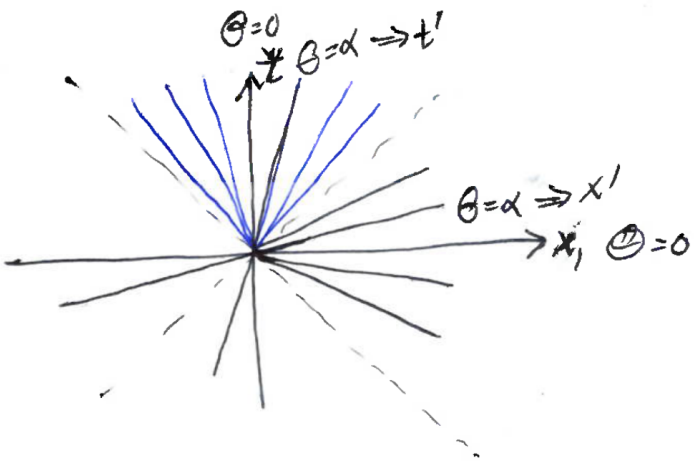

Figure 24: Families of constant hyperbolic angle show the old and boosted time and space axes.

20 Simultaneity, intervals, and causality

Lecture 11

20.1 Surfaces of constant primed coordinates

The tilted axes determine the full coordinate grid. For a fixed value $t' = T'$, the boost equation gives

$$t - vx = \frac{T'}{\gamma}, \qquad t = vx + \frac{T'}{\gamma}. \tag{66}$$

All constant-t' lines have slope v and are parallel to the x' axis. They are the events that the primed observer calls simultaneous. Similarly, fixing $x' = X'$ gives

$$x - vt = \frac{X'}{\gamma}. \tag{67}$$

For $v \neq 0$, these lines have t-versus-x slope $1/v$ and are parallel to the t' axis. For $v = 0$, they are vertical. A fixed primed spatial position is a worldline of constant velocity v in the unprimed frame.

20.2 Newtonian simultaneity and relativistic simultaneity

Under a Galilean transformation, $t' = t$ up to a common shift of the time origin. Constant-time slices are the same for all inertial observers. They agree on whether

two events are simultaneous and, if they are not, on which occurred first. Newtonian spacetime has a preferred division into spatial slices labeled by absolute time.

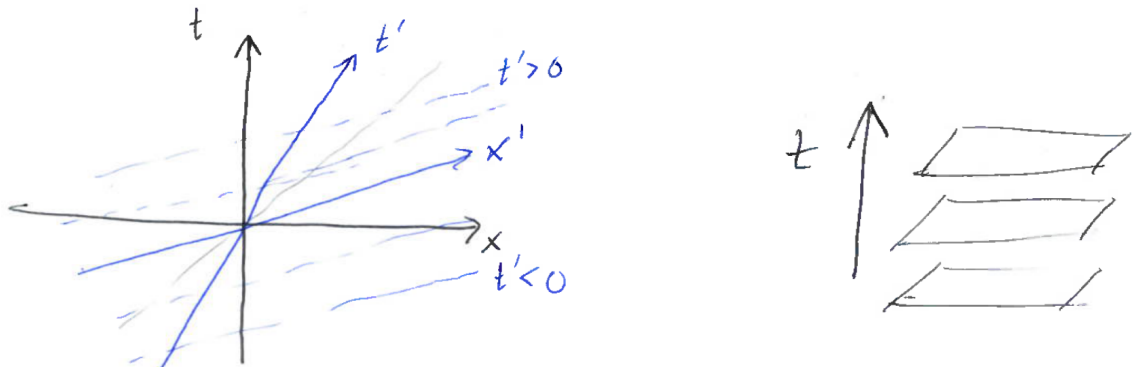

Figure 25: Relativistic constant-t' slices tilt with the observer (left). Newtonian spatial slices are labeled by a common absolute time (right).

Special relativity changes this conclusion for spatially separated events. Let $A = (0, 0)$ and $B = (0, L)$ with $L > 0$. They are simultaneous in the unprimed frame. In the primed frame,

$$t'_A = 0, \qquad t'_B = -\gamma v L \tag{68}$$

in units $c = 1$. A positive boost makes B occur before A, whereas a negative boost makes B occur after A. Restoring units replaces the second expression by $-\gamma v L/c^2$.

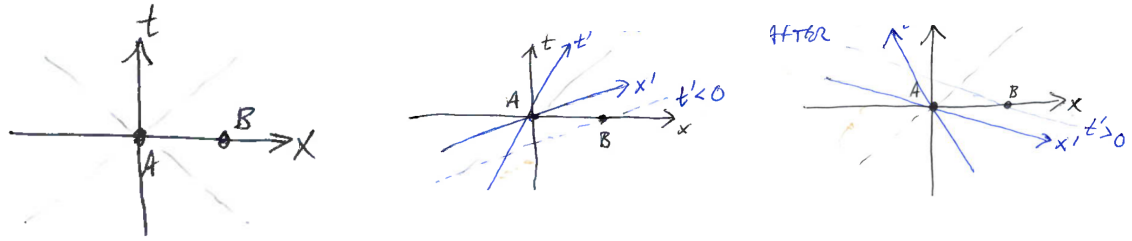

Figure 26: The same events are simultaneous in the unprimed frame (left), ordered with B earlier under a positive boost (middle), and ordered with B later under a negative boost (right).

This reordering does not allow a causal influence to arrive before it was sent. The two simultaneous, spatially distinct events cannot influence one another by a signal moving at or below the speed of light. To distinguish reorderable events from causally connected events, we need their invariant interval.

20.3 The invariant interval and its three causal types

For events $P_1 = (t_1, x_1)$ and $P_2 = (t_2, x_2)$, define

$$\Delta t = t_2 - t_1, \qquad \Delta x = x_2 - x_1.$$

Their squared Minkowski interval is

$$\Delta s^2 = -(\Delta t)^2 + (\Delta x)^2 \tag{69}$$

when $c = 1$. The notation Δs^2 denotes a signed quadratic form, not the square of a nonnegative Euclidean distance. In three spatial dimensions,

$$\Delta s^2 = -c^2(\Delta t)^2 + \Delta \boldsymbol{x} \cdot \Delta \boldsymbol{x}. \tag{70}$$

We classify distinct events as follows:

Interval	Separation	Condition with $c = 1$
$\Delta s^2 < 0$	Timelike	$\lvert\Delta \boldsymbol{x}\rvert < \lvert\Delta t\rvert$
$\Delta s^2 = 0$	Null or lightlike	$\lvert\Delta \boldsymbol{x}\rvert = \lvert\Delta t\rvert$
$\Delta s^2 > 0$	Spacelike	$\lvert\Delta \boldsymbol{x}\rvert > \lvert\Delta t\rvert$

All inertial observers agree on this classification because the interval is Lorentz invariant. In a $1+1$ diagram, the lines $x = \pm t$ through an event are its null directions. Points in the future or past wedges are timelike separated from the event; points in the side wedges are spacelike separated.

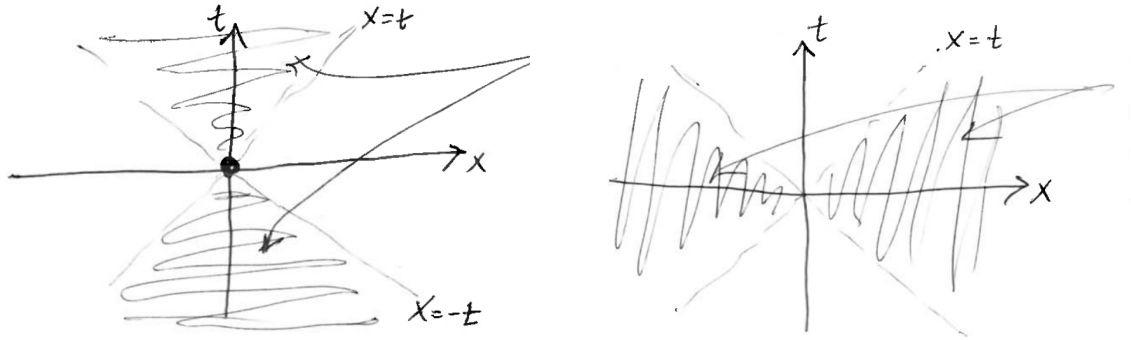

Figure 27: Timelike separation occupies the future and past wedges (left); spacelike separation occupies the side wedges (right). The boundaries are null.

20.4 Which events can change their time ordering?

Consider events C and D at the same unprimed spatial position, with $t_D > t_C$. Their separation is timelike. A boost gives

$$t'_D - t'_C = \gamma(t_D - t_C) > 0,$$

so their ordering is preserved.

The same conclusion holds for every future-directed timelike separation. In one spatial dimension, $\Delta t' = \gamma(\Delta t - v\Delta x)$. If $\Delta t > |\Delta x|$ and $|v| < 1$, then

$$\Delta t - v\Delta x \geq \Delta t - |v|\,|\Delta x| > 0.$$

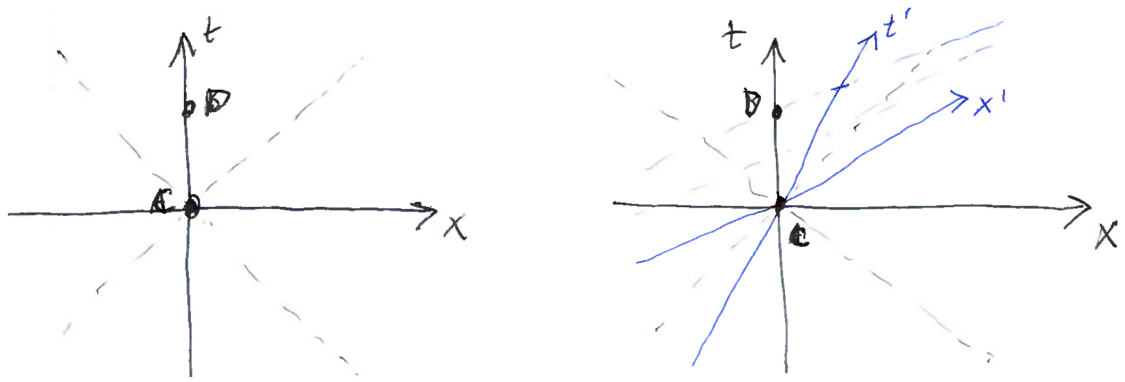

Figure 28: Two events at the same unprimed spatial position retain their temporal order in a boosted frame.

For a future-directed nonzero null separation, $\Delta t = |\Delta x| > 0$, the result remains strictly positive: $\Delta t - v\Delta x \geq (1 - |v|)\Delta t > 0$. Boosts therefore preserve time ordering for timelike and null separated events. Rotations do not change time, so the same holds for all proper orthochronous Lorentz transformations.

There is also a continuity argument. Suppose a continuously varying boost reversed the sign of Δt for a timelike pair. At an intermediate rapidity $\Delta t'$ would vanish. The interval would then be $(\Delta x')^2 \geq 0$, contradicting its negative invariant value. For distinct null events, a simultaneous frame would require both $\Delta t' = 0$ and $\Delta x' = 0$, contradicting invertibility. This excludes time reversal itself, which is not a boost connected to the identity.

For a spacelike separation, $|\Delta t/\Delta x| < 1$ in $1+1$ dimensions. The allowed boost $v_0 = \Delta t/\Delta x$ makes $\Delta t' = 0$. Boosts slightly above or below v_0 give opposite time orderings. Conversely, a timelike separation has an allowed rest-frame boost $v_{\text{rest}} = \Delta x/\Delta t$ that makes $\Delta x' = 0$. Spacelike separated events can be simultaneous in some frame, while timelike separated events can occur at the same place in some frame. Null separation admits neither such inertial frame.

20.5 Particle trajectories and the principle of causality

A constant-velocity path is a straight line on a spacetime diagram. If its endpoints obey $v_{\text{path}} = \Delta x/\Delta t$, then, with $c = 1$,

$$\left|\frac{\Delta t}{\Delta x}\right| = \frac{1}{|v_{\text{path}}|}$$

when $\Delta x \neq 0$. A subluminal trajectory has slope magnitude greater than one and timelike separated endpoints. A light ray has slope magnitude one and null separated endpoints. A hypothetical superluminal trajectory has slope magnitude less than one and spacelike separated endpoints.

The first two cases have invariant time ordering. In the third, some observers assign an arrival time earlier than the departure time. Interpreting such a spacelike trajectory as a freely controllable cause-and-effect signal is incompatible with the

ordinary Lorentz-invariant causal ordering. The causal principle used in relativity is that physical influences propagate within or on the light cone.

In 3+1 dimensions, an event can be influenced only by events in its causal past and can influence only events in its causal future. The interiors of these cones correspond to timelike separation; the boundaries correspond to light signals. The null boundary is included when discussing the allowed region for causal influences.

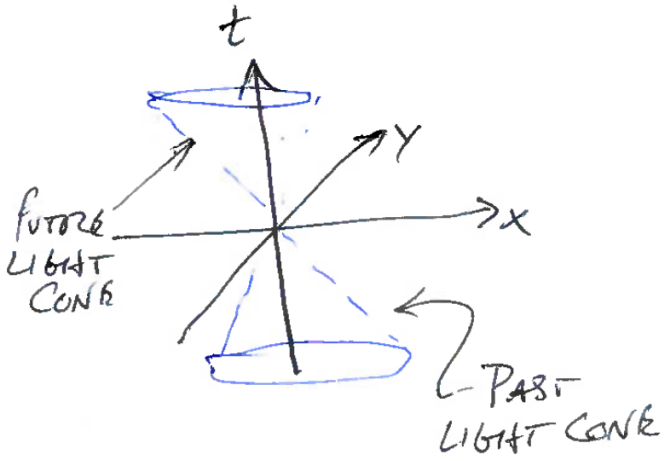

Figure 29: Future and past light cones, drawn with two spatial directions. The null boundary and timelike interior define the allowed causal directions.

21 The Minkowski metric and the time measured by a clock

Lecture 12

21.1 Metric components and the summation convention

For a D-dimensional spatial geometry, a line element can be written

$$\mathrm{d}\ell^2 = g_{ab}(x)\,\mathrm{d}x^a \mathrm{d}x^b, \qquad a, b = 1, \ldots, D. \tag{71}$$

Here the Latin indices label coordinates in the spatial geometry under discussion. Repeated upper and lower indices are summed. Explicitly, the right-hand side denotes $\sum_{a=1}^{D}\sum_{b=1}^{D} g_{ab}(x)\mathrm{d}x^a\mathrm{d}x^b$. The notation $g_{ab}(x)$ emphasizes that the metric components generally depend on position.

For a spacetime, the same construction gives

$$\mathrm{d}s^2 = g_{\mu\nu}(x)\,\mathrm{d}x^\mu \mathrm{d}x^\nu. \tag{72}$$

Greek indices here run from 0 to $D-1$ in D spacetime dimensions; in the physical four-dimensional case they run from 0 to 3. The coordinate $x^0 = ct$ is the time coordinate in an inertial Cartesian chart, and the remaining coordinates are spatial. We use the $(-,+,+,+)$ signature.

In Euclidean Cartesian coordinates the spatial metric is $g_{ab} = \delta_{ab}$, where

$$\delta_{ab} = \begin{cases} 1, & a = b, \\ 0, & a \neq b. \end{cases}$$

Consequently

$$\mathrm{d}\ell^2 = (\mathrm{d}x^1)^2 + \cdots + (\mathrm{d}x^D)^2 = \delta_{ab}\mathrm{d}x^a\mathrm{d}x^b.$$

In Cartesian Minkowski coordinates the spacetime metric is given the special symbol $\eta_{\mu\nu}$:

$$\eta_{\mu\nu} = \begin{cases} -1, & \mu = \nu = 0, \\ +1, & \mu = \nu = 1, 2, 3, \\ 0, & \mu \neq \nu, \end{cases} \qquad (\eta_{\mu\nu}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{73}$$

Thus

$$\begin{aligned} \mathrm{d}s^2 &= -(\mathrm{d}x^0)^2 + (\mathrm{d}x^1)^2 + (\mathrm{d}x^2)^2 + (\mathrm{d}x^3)^2 \\ &= -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 = \eta_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu. \end{aligned}$$

The array of components can be displayed as a matrix, but the metric itself is a coordinate-independent tensor. Its component array changes when the coordinates or basis change. Even a flat metric can have position-dependent components in curvilinear coordinates, as the Milne line element already illustrates; position dependence alone does not establish curvature.

21.2 Proper time on a straight timelike path

What does the Minkowski interval measure physically? Begin with a straight timelike path in $1 + 1$ dimensions and choose its rest frame. The object then follows a line of constant spatial position, say $x = 0$. For two events A and B on this line,

$$\Delta s^2 = -(t_B - t_A)^2$$

in units $c = 1$. A clock carried by the object measures the elapsed rest-frame time $t_B - t_A$. We call this elapsed time its **proper time**:

$$\Delta\tau = \sqrt{-\Delta s^2} = t_B - t_A \qquad (t_B > t_A,\ c = 1). \tag{74}$$

With units restored, $\Delta\tau = \sqrt{-\Delta s^2}/c$.

All inertial observers assign the same proper time to this path because the interval is Lorentz invariant. In another frame the object moves spatially, but the combination of its spatial and temporal coordinate differences gives the same result. For constant speed v,

$$\Delta x = v\Delta t \quad \Longrightarrow \quad \Delta\tau = \Delta t\sqrt{1 - \frac{v^2}{c^2}}.$$

The moving clock therefore records less time than the coordinate-time separation of its endpoints in that frame.

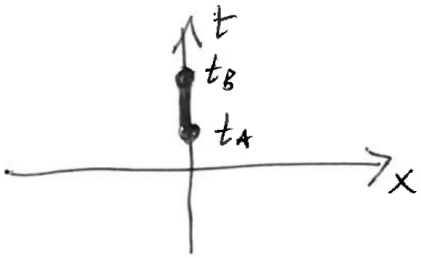

Figure 30: A clock at rest follows a vertical timelike worldline. Its elapsed proper time equals the coordinate time in this rest frame.

Physical clocks operate through laws that must be compatible with Lorentz invariance. Proper time is the invariant interval along their timelike histories that agrees with ordinary time in their instantaneous rest frames. An ideal clock measures this quantity; the coordinate time of a chosen distant observer is not itself the clock reading.

21.3 Proper time on a general timelike curve

An accelerated observer need not follow a straight line. Describe its worldline by

$$t = t(\lambda), \qquad x = x(\lambda),$$

where λ is any smooth monotone parameter. The tangent must be timelike along the path. For a future-directed parametrization this means $\dot{t} > 0$ and $\dot{t}^{\,2} - \dot{x}^{\,2} > 0$, where a dot denotes $\mathrm{d}/\mathrm{d}\lambda$.

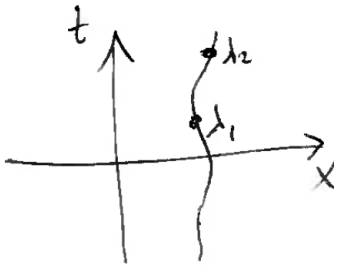

Figure 31: A timelike curve with two parameter values marking the portion over which proper time is accumulated.

Divide the curve into infinitesimal segments. Each has

$$\mathrm{d}\tau^2 = -\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}x^2$$

when $c = 1$. Using $\mathrm{d}t = (\mathrm{d}t/\mathrm{d}\lambda)\mathrm{d}\lambda$ and $\mathrm{d}x = (\mathrm{d}x/\mathrm{d}\lambda)\mathrm{d}\lambda$ gives

$$\boxed{\Delta\tau = \int_{\lambda_1}^{\lambda_2} \sqrt{\left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2 - \left(\frac{\mathrm{d}x}{\mathrm{d}\lambda}\right)^2}\,\mathrm{d}\lambda.} \tag{75}$$

This is the Lorentzian counterpart of adding infinitesimal Euclidean arc lengths, with the sign chosen to make the timelike integrand real.

If t increases along the curve, it can itself be used as the parameter. Writing $v(t) = \mathrm{d}x/\mathrm{d}t$ then yields

$$\Delta\tau = \int_{t_1}^{t_2} \sqrt{1 - v(t)^2}\,\mathrm{d}t \qquad (c = 1). \tag{76}$$

Segments traveled with speed close to one contribute comparatively little proper time. Proper time belongs to the entire path, not just to its endpoints: different timelike paths joining the same pair of events need not give the same integral.

The expression does not depend on the choice of monotone parameter. For an increasing reparametrization $\lambda = \lambda(\sigma)$,

$$\frac{\mathrm{d}x^\mu}{\mathrm{d}\sigma} = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}\sigma}.$$

The square root in the proper-time integral acquires the factor $\mathrm{d}\lambda/\mathrm{d}\sigma$, which cancels the change of integration measure. With either orientation the corresponding absolute Jacobian gives the positive elapsed time.

21.4 Proper time and proper length in four dimensions

For a curve $x^\mu(\lambda)$ in $3+1$ dimensional Minkowski spacetime, the same calculation gives, in units $c = 1$,

$$\Delta\tau = \int_{\lambda_1}^{\lambda_2} \sqrt{(\dot{x}^0)^2 - (\dot{x}^1)^2 - (\dot{x}^2)^2 - (\dot{x}^3)^2}\,\mathrm{d}\lambda \tag{77}$$

$$= \boxed{\int_{\lambda_1}^{\lambda_2} \sqrt{-\eta_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}\,\mathrm{d}\lambda.} \tag{78}$$

With $x^0 = ct$ and c explicit, the right-hand side is divided by c. Choosing coordinate time as parameter gives

$$\Delta\tau = \int_{t_1}^{t_2} \sqrt{1 - \frac{|\boldsymbol{v}(t)|^2}{c^2}}\,\mathrm{d}t. \tag{79}$$

For a spacelike curve, the quadratic form has the opposite sign. Its **proper length** is

$$\boxed{\Delta\ell = \int_{\lambda_1}^{\lambda_2} \sqrt{\eta_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}\,\mathrm{d}\lambda.} \tag{80}$$

The curve must be spacelike along the part being integrated. On a constant-time curve in an inertial frame, this reduces to ordinary spatial arc length. A null curve has $\mathrm{d}s^2 = 0$ and accumulates no proper time; proper time therefore cannot serve as its nontrivial parameter.

Part III
Vector Spaces and Tensor Algebra

22 Real vector spaces and the meaning of a vector

22.1 Why vectors and tensors are needed

The language of vector and tensor fields makes it possible to express geometry and physical laws independently of a choice of coordinates. Electric and magnetic fields already supply familiar examples of spatial vector fields. In general relativity, tensor fields will describe the spacetime metric, matter, and curvature. We first examine the linear algebra at one point, temporarily setting aside the position dependence of a field.

A real vector space V is a collection of objects called vectors equipped with addition and multiplication by real numbers. If $u, v \in V$ and $a, b \in \mathbb{R}$, then

$$au + bv \in V. \tag{81}$$

The familiar algebraic rules include

$$\begin{aligned} (u+v)+w &= u+(v+w), & u+v &= v+u, \\ a(bu) &= (ab)u, & 1u &= u, \\ a(u+v) &= au+av, & (a+b)u &= au+bu. \end{aligned}$$

There is a unique zero vector 0 such that $u+0=u$, and every vector has an additive inverse $-u$. These axioms describe a linear structure; they do not yet assign lengths, angles, or an inner product.

Vector spaces can be finite dimensional or infinite dimensional. Spaces of functions are common infinite-dimensional examples, with addition and scalar multiplication defined pointwise. The tangent spaces needed for spacetime geometry are finite dimensional.

22.2 Linear dependence and dimension

Vectors $v_1, \dots, v_k$ are **linearly independent** if

$$a_1 v_1 + \cdots + a_k v_k = 0$$

forces all coefficients to vanish. They are **linearly dependent** if the same equation has a solution in which at least one coefficient is nonzero. The nontriviality condition matters: the choice $a_1 = \cdots = a_k = 0$ is possible for every collection of vectors.

If $\dim V = n$, any $n+1$ vectors in V are linearly dependent. To see why, expand them in a basis with n elements. Their components form an $n \times (n+1)$ matrix. A linear relation among the vectors is a solution of the homogeneous matrix equation for its $n+1$ coefficients. There are more unknowns than the rank can reach, so the null space contains a nonzero vector. Conversely, a basis contains exactly n independent vectors and spans the space.

22.3 Bases and components

Lecture 13

Choose a basis $e_{(a)}$, $a = 1, \ldots, n$. Here the parenthesized index labels the basis element; it is not a coordinate component of one particular vector. Every $v \in V$ has a unique expansion

$$v = v^1 e_{(1)} + \cdots + v^n e_{(n)} = \sum_{a=1}^{n} v^a e_{(a)} = v^a e_{(a)}. \tag{82}$$

The real numbers v^a are the **components** of v in this basis. In the final expression the repeated index is summed by the Einstein convention. Generic vector-space indices a, b in this part run over the basis dimension; they are not restricted to spatial coordinates.

Existence of the expansion is the spanning property. Its uniqueness follows from linear independence: if $v = v^a e_{(a)} = \widetilde{v}^a e_{(a)}$, subtraction gives $(v^a - \widetilde{v}^a) e_{(a)} = 0$, so $v^a = \widetilde{v}^a$ for every a. The components depend on the basis; the vector itself does not.

The basis vectors are not assumed to be unit vectors or mutually orthogonal. At this point there is no inner product on V, so those words have not yet been assigned a meaning. A basis is simply an independent spanning collection.

23 Tangent spaces and vector fields

23.1 A vector at a point is a tangent to a curve

In differential geometry and general relativity, there is a vector space at each point p of spacetime. This is its **tangent space**, denoted $T_p\mathcal{M}$ or simply T_p. If the spacetime is four dimensional, the tangent space is a four-dimensional real vector space.

A tangent vector can be represented by a parametrized curve passing through p. If the curve is $x^\mu(s)$ and p corresponds to $s = s_0$, its tangent components are

$$v^\mu = \left.\frac{\mathrm{d}x^\mu}{\mathrm{d}s}\right|_{s_0}. \tag{83}$$

The tangent records the local rate and direction of change of the coordinates along the curve. Different curves may determine different tangent vectors, although curves that agree to first order at p determine the same one. Reparametrizing a curve rescales the tangent by the derivative of the old parameter with respect to the new one.

The chain rule explains the transformation of its components. If $x'^\mu = x'^\mu(x)$, then

$$v'^\mu = \left.\frac{\partial x'^\mu}{\partial x^\nu}\right|_p v^\nu. \tag{84}$$

Thus the component list changes with coordinates in a precisely controlled way. The tangent vector is the geometric object represented by these different lists.

Figure 32: A parametrized curve through p supplies a tangent vector at that point.

23.2 Tangent vectors do not connect arbitrary points

In Euclidean space, after choosing an origin, one often identifies a point with its position vector and represents a displacement by an arrow joining two points. This construction uses the global affine structure of flat space. The accompanying sketch illustrates the familiar position-vector description.

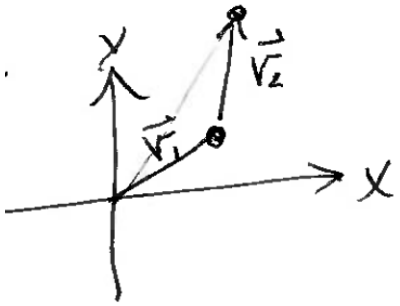

Figure 33: Position vectors in a flat coordinate plane rely on a chosen origin and the affine structure of the plane.

On a general curved spacetime, a vector lives at one point in its tangent space. It is not intrinsically an arrow joining that point to some other point of spacetime. There is no preferred global position vector, no natural subtraction of arbitrary spacetime points, and no automatic identification of $T_p\mathcal{M}$ with $T_q\mathcal{M}$ when $p \neq q$.

A sphere provides a useful picture: the tangent planes at different points are distinct planes, even when the sphere is drawn inside three-dimensional Euclidean space. Vectors in one tangent plane can be added and multiplied by scalars within that plane. Combining vectors at different points requires additional geometric structure and a rule for comparison.

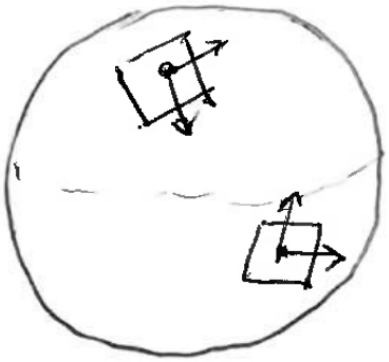

Figure 34: Distinct points of a sphere have distinct tangent spaces. The vectors belong to the local tangent planes.

23.3 The tangent bundle and a vector field

The collection of tangent spaces at all points is the **tangent bundle**,

$$T\mathcal{M} = \coprod_{p \in \mathcal{M}} T_p\mathcal{M}. \tag{85}$$

The disjoint-union notation keeps track of the base point attached to each vector. A vector in $T_p\mathcal{M}$ and an equal-looking component list in $T_q\mathcal{M}$ are not identified merely because their numerical entries match. The bundle projection is

$$\pi : T\mathcal{M} \longrightarrow \mathcal{M}, \qquad \pi(v_p) = p.$$

A **vector field** assigns a tangent vector $X_p \in T_p\mathcal{M}$ to each point p, with the desired continuity or smoothness. In bundle language this is a **section**: a map $X : \mathcal{M} \to T\mathcal{M}$ satisfying $\pi \circ X = \mathrm{id}_{\mathcal{M}}$. The section condition says exactly that the vector assigned to p belongs to the tangent space at p. Physicists' vector fields and mathematicians' sections of the tangent bundle are therefore the same objects described in different language.

24 Dual vectors, cotangent spaces, and tensors

24.1 The dual vector space

Starting with a real vector space V, construct the **dual space** V^* of all linear maps from V to $\mathbb{R}$. An element $\omega \in V^*$ is a **dual vector**, **covector**, or **linear functional**. It takes a vector as input and returns a real number:

$$\omega : V \longrightarrow \mathbb{R}, \qquad v \longmapsto \omega(v). \tag{86}$$

Its defining property is linearity:

$$\omega(au + bv) = a\omega(u) + b\omega(v) \qquad (a, b \in \mathbb{R}). \tag{87}$$

Covectors themselves can be added and multiplied by scalars. If $\omega, \eta \in V^*$, define

$$\sigma = a\omega + b\eta, \qquad \sigma(v) = a\omega(v) + b\eta(v).$$

The result is again a linear functional. Indeed,

$$\begin{aligned}\sigma(cu + dv) &= a\omega(cu + dv) + b\eta(cu + dv)\\ &= ac\omega(u) + ad\omega(v) + bc\eta(u) + bd\eta(v)\\ &= c\sigma(u) + d\sigma(v).\end{aligned}$$

The zero covector sends every vector to zero, and the remaining vector-space axioms follow pointwise from those of $\mathbb{R}$. Thus V^* is itself a real vector space.

For spacetime, the dual of $T_p\mathcal{M}$ is the **cotangent space** $T_p^*\mathcal{M}$. Collecting these spaces over all points gives the **cotangent bundle** $T^*\mathcal{M}$. A covector field assigns an element of $T_p^*\mathcal{M}$ to each point; equivalently it is a section of the cotangent bundle.

24.2 Constructing the dual basis

Let $e_{(a)}$, $a = 1, \ldots, n$, be a basis of V. Define covectors $\theta^{(a)}$ by their values on these basis vectors:

$$\boxed{\theta^{(a)}(e_{(b)}) = \delta^a{}_b,} \qquad \delta^a{}_b = \begin{cases} 1, & a = b,\\ 0, & a \neq b.\end{cases} \tag{88}$$

A linear map is completely determined by its values on a basis. Consequently this prescription determines one and only one $\theta^{(a)}$ for each a.

Its action on an arbitrary vector follows step by step:

$$\theta^{(a)}(v) = \theta^{(a)}\big(v^b e_{(b)}\big) \tag{89}$$

$$= v^b \theta^{(a)}(e_{(b)}) \tag{90}$$

$$= v^b \delta^a{}_b = v^a. \tag{91}$$

The dual basis element $\theta^{(a)}$ extracts the ath component of the input vector.

To see that the $\theta^{(a)}$ form a basis of V^*, first suppose $c_a\theta^{(a)} = 0$. Evaluating this functional on $e_{(b)}$ gives $c_a\delta^a{}_b = c_b = 0$, proving independence. Next, for any $\omega \in V^*$, define

$$\omega_a = \omega(e_{(a)}). \tag{92}$$

On an arbitrary vector $v = v^b e_{(b)}$,

$$\omega(v) = v^b\omega(e_{(b)}) = v^b\omega_b = \big(\omega_a\theta^{(a)}\big)(v).$$

The two functionals agree on every vector, so

$$\boxed{\omega = \omega_a\theta^{(a)}.} \tag{93}$$

They span the dual space. There are n dual basis elements, proving $\dim V^* = \dim V = n$.

24.3 The natural pairing of a vector and a covector

Combining the expansions of a vector and a covector gives

$$\omega(v) = \omega_a \theta^{(a)}\big(v^b e_{(b)}\big) \tag{94}$$

$$= \omega_a v^b \theta^{(a)}(e_{(b)}) \tag{95}$$

$$= \omega_a v^b \delta^a{}_b = \boxed{\omega_a v^a.} \tag{96}$$

The result is a scalar. No inner product or metric was needed. The upper index on v^a and the lower index on ω_a indicate their different roles and their natural contraction.

For example, if $v = 2e_{(1)} - 3e_{(2)}$ and $\omega = 4\theta^{(1)} + 5\theta^{(2)}$, then $\omega(v) = 4(2) + 5(-3) = -7$. The component rule is simply the evaluation of a linear functional, not an assumed Euclidean dot product between two vectors.

Although V and V^* have the same finite dimension, a preferred identification between them has not yet been supplied. Such an identification can be constructed from a nondegenerate metric, but the dual space and its evaluation pairing exist before any metric is chosen.

24.4 A vector as a linear map on covectors

The pairing can be read in the opposite order. Given $v \in V$, define a map $J(v)$ on the dual space by

$$J(v) : V^* \longrightarrow \mathbb{R}, \qquad J(v)(\omega) = \omega(v). \tag{97}$$

This is linear in ω:

$$J(v)(a\omega + b\eta) = a\omega(v) + b\eta(v) = aJ(v)(\omega) + bJ(v)(\eta).$$

Thus $J(v)$ belongs to the double dual $(V^*)^* = V^{**}$. The assignment $J : V \to V^{**}$ is itself linear because

$$J(au + bv)(\omega) = \omega(au + bv) = aJ(u)(\omega) + bJ(v)(\omega).$$

It is injective. If $J(v) = 0$, every covector annihilates v; in particular $\theta^{(a)}(v) = v^a = 0$ for all a, so $v = 0$. Since the domain and codomain both have dimension n, injectivity implies surjectivity. We therefore obtain the canonical identification

$$\boxed{V^{**} \simeq V} \tag{98}$$

for finite-dimensional V. Unlike an identification of V with V^*, this map requires no metric and no choice of basis. With the identification understood, one may write

$$v(\omega) \equiv \omega(v) = v^a \omega_a. \tag{99}$$

Vectors and covectors are thus both linear maps: a covector acts on vectors, while a vector acts on covectors.

24.5 Tensors as multilinear maps

The evaluation viewpoint extends naturally from one input to several. A **tensor of type** (k, ℓ) on a finite-dimensional real vector space V is a multilinear map

$$T : \underbrace{V^* \times \cdots \times V^*}_{k \text{ factors}} \times \underbrace{V \times \cdots \times V}_{\ell \text{ factors}} \longrightarrow \mathbb{R}. \tag{100}$$

Here k, ℓ are nonnegative integers, and $\times$ denotes the Cartesian product of input spaces. The collection of these tensors is itself a vector space, conventionally denoted $\mathcal{T}^{(k,\ell)}(V)$.

Multilinear means linear in each input separately, with all other inputs held fixed. For a vector slot, for example,

$$T(\ldots, au + bv, \ldots) = aT(\ldots, u, \ldots) + bT(\ldots, v, \ldots),$$

and the same condition applies to each covector slot. There are k covector inputs and ℓ vector inputs; this order explains the upper and lower index counts used for tensor components.

The preceding constructions are the simplest cases. A covector is a $(0, 1)$ tensor because it takes one vector as input. A vector, viewed through the double-dual identification, is a $(1, 0)$ tensor because it takes one covector as input. A scalar is a $(0, 0)$ tensor. At a spacetime point p, the appropriate choice is $V = T_p\mathcal{M}$ and $V^* = T_p^*\mathcal{M}$. A tensor field assigns such a multilinear map smoothly at each point. This definition supplies the coordinate-independent object whose components will be developed next.

25 Tensors as multilinear maps

Lecture 14

The construction of a tensor begins with a real vector space V and its dual V^*. Throughout the algebraic discussion, let $\dim V = D < \infty$. Lowercase Latin indices $a, b, \ldots$ label an arbitrary basis of this D-dimensional space. When the space is a spacetime tangent space, we use Greek indices $\mu, \nu, \ldots = 0, 1, 2, 3$. No metric is needed to define a tensor.

25.1 Vectors and covectors as maps

A covector $\omega \in V^*$ is a linear map

$$\omega : V \longrightarrow \mathbb{R}, \qquad \omega(au + bv) = a\omega(u) + b\omega(v).$$

A vector can also be viewed as a linear map, but its input is a covector. The canonical double-dual identification established in Section 24.4 is the evaluation map

$$v(\omega) = \omega(v), \qquad v : V^* \longrightarrow \mathbb{R}. \tag{101}$$

It does not require a basis or an inner product. The distinction between V and V^* remains important: the canonical identification is between V and its *double* dual.

25.2 The definition and the meaning of the slots

Recall the definition in Section 24.5: a tensor of type (k, ℓ) on V is a multilinear map

$$T : \underbrace{V^* \times \cdots \times V^*}_{k \text{ factors}} \times \underbrace{V \times \cdots \times V}_{\ell \text{ factors}} \longrightarrow \mathbb{R}. \tag{102}$$

The vector space of all such tensors is denoted $\mathcal{T}^{(k,\ell)}(V)$. We now unpack its multilinearity and construct the associated component description.

The symbol $\times$ in the domain is a Cartesian product. An element of $V \times V$ is an ordered pair (u, v), while an element of $V^* \times V$ is an ordered pair (ω, v). The order and type of the inputs are part of the definition of the map.

Multilinearity means linearity in each argument separately, with the remaining arguments held fixed. For a tensor of type $(0, 2)$, the two requirements are

$$T(au + bv, w) = aT(u, w) + bT(v, w), \tag{103}$$
$$T(u, av + bw) = aT(u, v) + bT(u, w). \tag{104}$$

For example, applying both rules gives

$$\begin{aligned} T(au + bv, cw + dz) = {} & acT(u, w) + adT(u, z) \\ & + bcT(v, w) + bdT(v, z). \end{aligned}$$

In particular, scaling both arguments by a multiplies the value by a^2. Multilinearity is therefore not the same as linearity in the ordered pair regarded as one argument.

A type $(2, 1)$ tensor has three slots:

$$T(\omega, \sigma, v) \in \mathbb{R}, \qquad \omega, \sigma \in V^*, \quad v \in V.$$

The first two slots accept covectors, and the third accepts a vector. For instance,

$$T(a\omega + b\eta, \sigma, v) = aT(\omega, \sigma, v) + bT(\eta, \sigma, v),$$

and analogous rules hold in its second and third slots. The upper and lower numbers in the tensor type record these different kinds of input, not merely the total number of arguments.

25.3 Scalars, vectors, covectors, and tensor fields

The simplest tensor types reproduce familiar objects:

- A covector is a tensor of type $(0, 1)$, since it maps one vector to a real number: $\mathcal{T}^{(0,1)}(V) = V^*$.
- A vector is a tensor of type $(1, 0)$, using (101): $\mathcal{T}^{(1,0)}(V) = V$.
- A tensor of type $(0, 0)$ is a scalar.

The last statement can be understood as a map with no arguments. The Cartesian product of zero factors consists of one empty tuple, and specifying a map from that singleton to $\mathbb{R}$ amounts to choosing one real number. Thus $T = 5$ is a perfectly good type $(0, 0)$ tensor.

On a manifold, these constructions are performed at each point. A tensor field of type (k, ℓ) assigns to each $p \in \mathcal{M}$ a tensor on $T_p\mathcal{M}$ and $T_p^*\mathcal{M}$, with smooth dependence on p. A type $(0, 0)$ tensor field is therefore an ordinary smooth function $f : \mathcal{M} \to \mathbb{R}$. A tensor at one point and a tensor field are distinct objects; the latter is a smoothly varying collection of the former.

25.4 The vector space of tensors

Tensors of a fixed type can be added and multiplied by real numbers by adding or scaling their values. For example, if T_1, T_2 have type $(2, 1)$, define

$$T_3 = aT_1 + bT_2$$

through

$$T_3(\omega, \sigma, u) = aT_1(\omega, \sigma, u) + bT_2(\omega, \sigma, u).$$

This again takes real values and is linear in every argument. In the first argument, for example,

$$\begin{aligned} T_3(c\omega + d\eta, \sigma, u) &= a\big[cT_1(\omega, \sigma, u) + dT_1(\eta, \sigma, u)\big] \\ &\quad + b\big[cT_2(\omega, \sigma, u) + dT_2(\eta, \sigma, u)\big] \\ &= cT_3(\omega, \sigma, u) + dT_3(\eta, \sigma, u). \end{aligned}$$

The same calculation works in the other slots. The zero map is the zero tensor, and additive inverses are obtained by multiplying by -1. Hence $\mathcal{T}^{(k,\ell)}(V)$ is a vector space. To construct a useful basis for it, we first need the tensor product.

26 Tensor products, bases, and components

Lecture 15

26.1 The tensor product of two tensors

Let T have type (k, ℓ) and let S have type (m, n). Their tensor product $T \otimes S$ has type $(k+m, \ell+n)$. Its definition separates the inputs into the slots belonging to each factor:

$$\begin{aligned} &(T \otimes S)(\omega_1, \ldots, \omega_{k+m}, v_1, \ldots, v_{\ell+n}) \\ &\quad = T(\omega_1, \ldots, \omega_k, v_1, \ldots, v_\ell) \\ &\qquad \times S(\omega_{k+1}, \ldots, \omega_{k+m}, v_{\ell+1}, \ldots, v_{\ell+n}). \end{aligned} \tag{105}$$

The multiplication on the right is ordinary multiplication of real numbers. The first k covectors and first ℓ vectors are given to T; the remaining inputs are given to S.

Each argument enters just one of the two factors. Thus, if one argument is replaced by a linear combination, multilinearity of that factor and distributivity of real

multiplication prove multilinearity of $T \otimes S$. This checks that the definition really produces a tensor of the claimed type.

Example 26.1 (A vector times a covector)**.** Let $v \in V$ and $\omega \in V^*$. Then $v \otimes \omega$ has type $(1,1)$ and acts by

$$(v \otimes \omega)(\sigma, u) = v(\sigma)\omega(u) = \sigma(v)\omega(u). \tag{106}$$

If $\sigma(v) = 2$ and $\omega(u) = -3$, the value is -6. The output is a scalar, even though the two factors are geometric objects of different kinds.

The tensor product is bilinear in its tensor factors:

$$\begin{aligned}(aT + bU) \otimes S &= a(T \otimes S) + b(U \otimes S),\\ T \otimes (aS + bW) &= a(T \otimes S) + b(T \otimes W).\end{aligned}$$

Both identities follow by evaluating the two sides on arbitrary allowed inputs. Repeated tensor products are associative under the usual identification of their slots. The order of the factors within a set of slots nevertheless matters. For example,

$$(\omega \otimes \sigma)(u, v) = \omega(u)\sigma(v), \qquad (\sigma \otimes \omega)(u, v) = \sigma(u)\omega(v),$$

which need not agree.

26.2 Building a basis for the tensor space

Choose a basis $\{e_a\}_{a=1}^{D}$ for V and its dual basis $\{\theta^a\}_{a=1}^{D}$ for V^*. Their products

$$e_{a_1} \otimes \cdots \otimes e_{a_k} \otimes \theta^{b_1} \otimes \cdots \otimes \theta^{b_\ell} \tag{107}$$

form a basis of $\mathcal{T}^{(k,\ell)}(V)$. Every index in (107) independently takes D possible values.

To derive the expansion, write the input covectors and vectors as

$$\omega_j = (\omega_j)_a \theta^a, \qquad v_j = (v_j)^b e_b.$$

Multilinearity gives

$$\begin{aligned}&T(\omega_1, \ldots, \omega_k, v_1, \ldots, v_\ell)\\ &\quad = (\omega_1)_{a_1} \cdots (\omega_k)_{a_k} (v_1)^{b_1} \cdots (v_\ell)^{b_\ell}\\ &\qquad \times T(\theta^{a_1}, \ldots, \theta^{a_k}, e_{b_1}, \ldots, e_{b_\ell}).\end{aligned} \tag{108}$$

Define the components by

$$T^{a_1 \cdots a_k}{}_{b_1 \cdots b_\ell} = T(\theta^{a_1}, \ldots, \theta^{a_k}, e_{b_1}, \ldots, e_{b_\ell}). \tag{109}$$

The tensor built from these numbers and the basis products has exactly the values in (108) on every input. Therefore

$$\begin{aligned}T = T^{a_1 \cdots a_k}{}_{b_1 \cdots b_\ell} e_{a_1} \otimes \cdots \otimes e_{a_k}\\ \otimes \theta^{b_1} \otimes \cdots \otimes \theta^{b_\ell}.\end{aligned} \tag{110}$$

Repeated upper and lower indices are summed. The summation convention saves writing $k+\ell$ nested sums in this formula.

For completeness, suppose a linear combination of the products (107) vanishes. Evaluate it on a particular tuple $(\theta^{c_1},\dots,\theta^{c_k},e_{d_1},\dots,e_{d_\ell})$. The dual-basis identities turn all the factors into Kronecker deltas and select exactly one coefficient. That coefficient must vanish. Since the tuple was arbitrary, every coefficient vanishes, proving linear independence. Consequently,

$$\dim \mathcal{T}^{(k,\ell)}(V) = D^{k+\ell}. \tag{111}$$

The k upper indices in the component array are vector-like; the ℓ lower indices are covector-like.

26.3 Evaluating a tensor using its components

The elementary evaluations are

$$e_a(\omega) = \omega(e_a) = \omega_a, \qquad \theta^b(v) = v^b.$$

Insert (110) into the multilinear map and apply the definition of the tensor product. Each vector basis factor extracts a covector component, and each dual-basis factor extracts a vector component. The result is

$$\begin{aligned} &T(\omega_1,\dots,\omega_k,v_1,\dots,v_\ell) \\ &\qquad = T^{a_1\cdots a_k}{}_{b_1\cdots b_\ell}(\omega_1)_{a_1}\cdots(\omega_k)_{a_k}(v_1)^{b_1}\cdots(v_\ell)^{b_\ell}. \end{aligned} \tag{112}$$

All the indices on the right are contracted, so the result is a real number, as required by the definition of a tensor.

For a type $(0,2)$ tensor this specializes to

$$T(u,v) = T_{ab}u^a v^b,$$

while a type $(1,1)$ tensor satisfies

$$T(\omega,v) = T^a{}_b\omega_a v^b.$$

The expression on the left is independent of a choice of basis; the expression on the right makes the tensor types and all the component operations explicit. They describe the same evaluation.

Example 26.2 (A complete two-dimensional evaluation)**.** Let $D=2$ and suppose

$$(\omega_1,\omega_2) = (5,-3), \qquad (v^1,v^2) = (-2,1), \qquad (T^a{}_b) = \begin{pmatrix} 4 & 2 \\ -1 & 0 \end{pmatrix}.$$

The upper index labels the row and the lower index labels the column. Expanding the two summed indices gives four terms:

$$\begin{aligned} T(\omega,v) &= T^1{}_1\omega_1 v^1 + T^1{}_2\omega_1 v^2 + T^2{}_1\omega_2 v^1 + T^2{}_2\omega_2 v^2 \\ &= 4\cdot 5\cdot(-2) + 2\cdot 5\cdot 1 + (-1)\cdot(-3)\cdot(-2) + 0\cdot(-3)\cdot 1 \\ &= -40 + 10 - 6 + 0 = -36. \end{aligned}$$

An equivalent matrix calculation is

$$\begin{pmatrix} 5 & -3 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 & -3 \end{pmatrix} \begin{pmatrix} -6 \\ 2 \end{pmatrix} = -36.$$

27 The metric tensor and index lowering

Lectures 16–18

27.1 A symmetric bilinear form at each point

The metric is a symmetric, nondegenerate tensor of type $(0, 2)$. At a fixed point, it is a bilinear map

$$g : V \times V \longrightarrow \mathbb{R}, \qquad g(u, v) = g(v, u).$$

Its basis expansion and its action are

$$g = g_{ab}\theta^a \otimes \theta^b, \qquad g(u, v) = g_{ab}u^a v^b, \qquad g_{ab} = g_{ba}. \tag{113}$$

Writing only g_{ab} is a common abbreviation for the metric together with an understood basis. On spacetime, the metric is a tensor field: the bilinear form and its components vary from one point to another.

Nondegeneracy means

$$g(u, v) = 0 \text{ for every } v \in V \quad \Longrightarrow \quad u = 0. \tag{114}$$

Equivalently, the matrix (g_{ab}) is invertible. This condition does not require $g(u, u)$ to be positive for nonzero u. In relativity it is essential that the metric is indefinite.

The squared length associated with a vector is

$$g(v, v) = g_{ab}v^a v^b.$$

It is often called a squared norm, although for a Lorentzian metric it is not the square of a positive-definite norm in the usual mathematical sense.

27.2 Minkowski inner products and causal character

In a Cartesian inertial basis of Minkowski spacetime,

$$(\eta_{\mu\nu}) = \operatorname{diag}(-1, 1, 1, 1). \tag{115}$$

Thus $\eta_{00} = -1$, $\eta_{11} = \eta_{22} = \eta_{33} = 1$, and all off-diagonal components vanish. The same geometric metric may be written as the abstract tensor η, as $\eta_{\mu\nu}\theta^\mu \otimes \theta^\nu$, or through its component matrix.

For vectors u and v,

$$\eta(u, v) = -u^0v^0 + u^1v^1 + u^2v^2 + u^3v^3, \tag{116}$$

$$\eta(v, v) = -(v^0)^2 + (v^1)^2 + (v^2)^2 + (v^3)^2. \tag{117}$$

With the signature $(-,+,+,+)$, a nonzero vector is classified by

$$\begin{array}{lcl} g(v,v) < 0 & : & \text{timelike,} \\ g(v,v) = 0 & : & \text{null or lightlike,} \\ g(v,v) > 0 & : & \text{spacelike.} \end{array} \tag{118}$$

The zero vector has zero squared length but is excluded when classifying a direction as null. For example, $(1,0,0,0)$ is timelike, $(1,1,0,0)$ is null, and $(0,1,0,0)$ is spacelike in the basis of (115). The tangent vectors to the worldline of a massive object are timelike; a light ray has null tangents.

A null vector does not contradict nondegeneracy. The condition $g(v,v) = 0$ fixes only the pairing of v with itself. It does not say that $g(v,w) = 0$ for every w.

27.3 The metric maps vectors to covectors

Fix one argument of the metric. For $u \in V$, the map

$$\omega = g(u, \cdot\,)$$

still has one vector slot open. Bilinearity of g implies

$$\omega(av + bw) = g(u, av + bw) = a\omega(v) + b\omega(w),$$

so ω is a covector. In components, for arbitrary v,

$$\omega(v) = \omega_b v^b = g(u,v) = g_{ab} u^a v^b.$$

Since this holds for every v, the coefficients of v^b agree:

$$\omega_b = g_{ab} u^a. \tag{119}$$

Symmetry permits us to write the same rule as

$$u_a = g_{ab} u^b. \tag{120}$$

We use the same letter for the vector and its associated covector, while the index position distinguishes their components.

Any bilinear form determines a unique covector from each vector by this procedure. Nondegeneracy supplies the stronger property that different vectors determine different covectors and that every covector arises this way. Indeed, if $g(u,\cdot) = g(v,\cdot)$, then $g(u - v, w) = 0$ for every w; (114) gives $u = v$. In finite dimension, injectivity and equal dimensions then give an isomorphism $V \to V^*$. Unlike the identification $V \simeq V^{**}$, this identification depends on the chosen metric.

Example 27.1 (Lowering an index in Minkowski spacetime)**.** Using (115),

$$\begin{aligned} v_0 &= \eta_{00} v^0 + \eta_{01} v^1 + \eta_{02} v^2 + \eta_{03} v^3 = -v^0, \\ v_1 &= \eta_{10} v^0 + \eta_{11} v^1 + \eta_{12} v^2 + \eta_{13} v^3 = v^1. \end{aligned}$$

Similarly $v_2 = v^2$ and $v_3 = v^3$. Therefore

$$(v_0, v_1, v_2, v_3) = (-v^0, v^1, v^2, v^3).$$

In this particular basis, lowering changes the sign of the time component and leaves the spatial components unchanged. This is a property of the Cartesian Minkowski component matrix, not a general rule for all metrics or all coordinate systems.

Example 27.2 (Lowering an index in plane polar coordinates)**.** The flat Euclidean plane has line element

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2.$$

In the coordinate basis $(\partial_r, \partial_\theta)$,

$$(g_{ab}) = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}.$$

For $v = v^r\partial_r + v^\theta\partial_\theta$, the associated covector has components

$$\begin{aligned} v_r &= g_{rr}v^r + g_{r\theta}v^\theta = v^r, \\ v_\theta &= g_{\theta r}v^r + g_{\theta\theta}v^\theta = r^2 v^\theta. \end{aligned}$$

Thus

$$g(v, \cdot) = v^r\mathrm{d}r + r^2 v^\theta \mathrm{d}\theta.$$

The factor r^2 reflects the length of the angular coordinate basis vector. These expressions apply on a polar coordinate patch with $r > 0$; the polar chart degenerates at the origin.

Once lowering has been defined, inner products have the equivalent forms

$$g(u, v) = g_{ab}u^a v^b = u_b v^b = u^a v_a, \qquad g(v, v) = v_a v^a. \tag{121}$$

The mixed upper-lower contraction is the natural pairing between a covector and a vector. The metric has been used to turn one of the original vectors into the required covector.

28 Mixed tensors and the inverse metric

28.1 A type $(1, 1)$ tensor as a linear map

A mixed tensor

$$T = T^a{}_b\, e_a \otimes \theta^b$$

can be used in two complementary ways. Fixing a vector v in its vector slot leaves a linear map on covectors, which is a vector:

$$u = T(\,\cdot\,, v), \qquad u^a = T^a{}_b v^b. \tag{122}$$

Indeed, for every covector ω, $\omega(u) = \omega_a T^a{}_b v^b = T(\omega, v)$. Conversely, fixing a covector leaves a covector:

$$\sigma = T(\omega, \cdot), \qquad \sigma_b = \omega_a T^a{}_b. \tag{123}$$

The second operation is the dual action of the first. No metric is needed for either operation.

In two dimensions, regard the components of a vector as a column. Then (122) is ordinary matrix multiplication:

$$\begin{pmatrix} u^1 \\ u^2 \end{pmatrix} = \begin{pmatrix} T^1{}_1 & T^1{}_2 \\ T^2{}_1 & T^2{}_2 \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix},$$

or explicitly

$$u^1 = T^1{}_1 v^1 + T^1{}_2 v^2, \qquad u^2 = T^2{}_1 v^1 + T^2{}_2 v^2.$$

Covectors are naturally displayed as rows, so the dual action is

$$\begin{pmatrix} \sigma_1 & \sigma_2 \end{pmatrix} = \begin{pmatrix} \omega_1 & \omega_2 \end{pmatrix} \begin{pmatrix} T^1{}_1 & T^1{}_2 \\ T^2{}_1 & T^2{}_2 \end{pmatrix}.$$

For instance,

$$\sigma_1 = \omega_1 T^1{}_1 + \omega_2 T^2{}_1, \qquad \sigma_2 = \omega_1 T^1{}_2 + \omega_2 T^2{}_2.$$

Thus a type $(1,1)$ tensor represents an endomorphism of V, with its associated dual map on V^*.

28.2 The Kronecker tensor is the identity

The identity map has tensor representation

$$I = e_a \otimes \theta^a = \delta^a{}_b \, e_a \otimes \theta^b, \qquad \delta^a{}_b = \begin{cases} 1, & a = b, \\ 0, & a \neq b. \end{cases} \tag{124}$$

It returns both vectors and covectors unchanged:

$$\delta^a{}_b v^b = v^a, \qquad \omega_a \delta^a{}_b = \omega_b.$$

As a multilinear map, $I(\omega, v) = \omega(v)$. In four spacetime dimensions its matrix is

$$(\delta^\mu{}_\nu) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

It is a type $(1,1)$ tensor, whereas the metric has type $(0,2)$. Even when a Euclidean metric has an identity matrix in a particular basis, the two tensors have different slots and should be distinguished.

28.3 The inverse metric

Because the metric is nondegenerate, its component matrix has an inverse. Denote its entries by g^{ab}. The defining identity is

$$g^{ab}g_{bc} = \delta^a{}_c. \tag{125}$$

These components define a tensor of type $(2,0)$,

$$g^{-1} = g^{ab}e_a \otimes e_b. \tag{126}$$

One may write $(g^{-1})^{ab}$ when emphasizing that the inverse metric is a different tensor, but the standard notation is g^{ab}. The inverse of a symmetric matrix is symmetric, since

$$(g^{-1})^{\mathsf{T}} = (g^{\mathsf{T}})^{-1} = g^{-1}.$$

Thus $g^{ab} = g^{ba}$.

For a diagonal metric, the inverse entries are the reciprocals of the nonzero diagonal entries. The polar metric gives

$$(g^{ab}) = \begin{pmatrix} 1 & 0 \\ 0 & r^{-2} \end{pmatrix}, \qquad g^{rr} = 1, \quad g^{\theta\theta} = r^{-2}. \tag{127}$$

For the Cartesian Minkowski metric,

$$(\eta^{\mu\nu}) = \operatorname{diag}(-1,1,1,1). \tag{128}$$

Here the inverse matrix has the same numerical entries as the original matrix because $(-1)^{-1} = -1$ and $1^{-1} = 1$. This equality of entries is special to this basis; it does not remove the distinction between covariant and contravariant tensors.

28.4 Raising an index and undoing index lowering

The inverse metric has two covector slots. Fixing one covector ω gives a vector $g^{-1}(\cdot,\omega)$. In components,

$$u^a = g^{ab}\omega_b. \tag{129}$$

This is called raising an index. If the covector was obtained by lowering a vector, the two operations cancel:

$$\begin{aligned} \omega_b &= g_{bc}v^c, \\ g^{ab}\omega_b &= g^{ab}g_{bc}v^c = \delta^a{}_c v^c = v^a. \end{aligned} \tag{130}$$

Conversely, if $u^b = g^{bc}\omega_c$, then

$$g_{ab}u^b = g_{ab}g^{bc}\omega_c = \delta_a{}^c\omega_c = \omega_a.$$

In index-free language, these statements say that the maps $v \mapsto g(\cdot, v)$ and $\omega \mapsto g^{-1}(\cdot,\omega)$ are inverse isomorphisms.

For the polar metric, the inverse rule reads

$$v^r = v_r, \qquad v^\theta = \frac{v_\theta}{r^2},$$

which exactly reverses the factors in Example 27.2. In Minkowski coordinates, raising again changes the sign of the time component.

The same operation may be applied to any selected index of a higher-rank tensor. For instance,

$$T^a{}_b = g^{ac}T_{cb}, \qquad T_a{}^b = T_{ac}g^{cb}, \qquad T^{ab} = g^{ac}g^{bd}T_{cd}. \tag{131}$$

The position in the sequence of slots is retained when an index is raised or lowered. For a general nonsymmetric tensor, raising the first slot and raising the second slot are different operations.

29 Contractions and traces

29.1 Contracting an upper index with a lower index

Contraction is the operation that pairs one vector-like index with one covector-like index. It maps a tensor of type (k, ℓ) to one of type $(k-1, \ell-1)$, provided $k, \ell \geq 1$. For a type $(1,1)$ tensor it produces the scalar

$$\operatorname{tr} T = T^a{}_a. \tag{132}$$

In four dimensions this is

$$T^\mu{}_\mu = T^0{}_0 + T^1{}_1 + T^2{}_2 + T^3{}_3.$$

The repeated index is summed, not left as an additional free index. This operation is also called tracing over a pair of indices.

There is a basis-independent reason why the trace is a scalar. Using the basis and dual basis,

$$\operatorname{tr} T = \sum_a T(\theta^a, e_a).$$

Under a basis change $e'_a = e_b M^b{}_a$, the dual basis is $\theta'^a = (M^{-1})^a{}_c \theta^c$. Therefore

$$\begin{aligned}\sum_a T(\theta'^a, e'_a) &= \sum_{a,b,c} (M^{-1})^a{}_c M^b{}_a \, T(\theta^c, e_b) \\ &= \sum_{b,c} \delta^b{}_c \, T(\theta^c, e_b) = \sum_b T(\theta^b, e_b).\end{aligned}$$

The inserted basis and dual basis cancel, so the result is unchanged. The same argument applies when other slots remain uncontracted. For example,

$$C^a{}_c = T^{ab}{}_{cb}$$

contracts the second upper slot with the second lower slot of a type $(2,2)$ tensor and leaves a type $(1,1)$ tensor.

Example 29.1 (The trace of the identity)**.** In four spacetime dimensions,

$$\delta^{\mu}{}_{\mu} = \delta^{0}{}_{0} + \delta^{1}{}_{1} + \delta^{2}{}_{2} + \delta^{3}{}_{3} = 4.$$

In an arbitrary D-dimensional vector space,

$$\delta^{a}{}_{a} = D. \tag{133}$$

The trace counts the dimension; it does not depend on the metric signature.

29.2 Taking traces with the metric

A type $(0,2)$ tensor does not have one upper and one lower index to contract directly. First raise one index:

$$T_{a}{}^{b} = T_{ac}g^{cb}.$$

Then contract to obtain

$$T_{a}{}^{a} = T_{ac}g^{ca} = g^{ab}T_{ab}. \tag{134}$$

Likewise, for a type $(2,0)$ tensor, lower an index and contract:

$$T_{a}{}^{b} = g_{ac}T^{cb}, \qquad T_{a}{}^{a} = g_{ac}T^{ca} = g_{ab}T^{ab}. \tag{135}$$

These traces use the metric; the trace of a type $(1,1)$ tensor does not require one.

For example, tracing the metric with its inverse gives

$$g^{ab}g_{ab} = \delta^{a}{}_{a} = D. \tag{136}$$

In Minkowski coordinates this becomes

$$\eta^{\mu\nu}\eta_{\mu\nu} = (-1)(-1) + 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 = 4.$$

By contrast, the ordinary diagonal sum of the displayed matrix $(\eta_{\mu\nu})$ is 2. That ordinary matrix sum has not paired an upper index with a lower index and is not the metric trace. The index positions specify the geometric operation.

30 Symmetrization and antisymmetrization

30.1 Two-index symmetry and its projections

A tensor T_{ab} of type $(0,2)$ is symmetric if $T_{ab} = T_{ba}$, and antisymmetric if $T_{ab} = -T_{ba}$. For instance, symmetry requires $T_{13} = T_{31}$; antisymmetry requires $T_{13} = -T_{31}$. Any tensor of this type determines a symmetric tensor

$$T_{(ab)} = \frac{1}{2}(T_{ab} + T_{ba}) \tag{137}$$

and an antisymmetric tensor

$$T_{[ab]} = \frac{1}{2}(T_{ab} - T_{ba}). \tag{138}$$

Round brackets denote symmetrization; square brackets denote antisymmetrization. The factor $1/2$ is part of the convention.

The claimed symmetries follow directly:

$$T_{(ba)} = \frac{1}{2}(T_{ba} + T_{ab}) = T_{(ab)},$$
$$T_{[ba]} = \frac{1}{2}(T_{ba} - T_{ab}) = -\frac{1}{2}(T_{ab} - T_{ba}) = -T_{[ab]}.$$

If the original tensor is already symmetric,

$$T_{(ab)} = \frac{1}{2}(T_{ab} + T_{ab}) = T_{ab}, \qquad T_{[ab]} = 0.$$

If it is already antisymmetric,

$$T_{[ab]} = \frac{1}{2}\big(T_{ab} - (-T_{ab})\big) = T_{ab}, \qquad T_{(ab)} = 0.$$

Thus these operations are projections: applying either one twice has the same effect as applying it once.

Adding (137) and (138) gives the decomposition

$$T_{ab} = T_{(ab)} + T_{[ab]}. \tag{139}$$

It is unique. A tensor that is simultaneously symmetric and antisymmetric satisfies $T_{ab} = T_{ba} = -T_{ab}$ and must vanish. Therefore the intersection of the two subspaces contains only the zero tensor.

The same constructions apply to upper indices: $S^{(ab)} = (S^{ab} + S^{ba})/2$ and $S^{[ab]} = (S^{ab} - S^{ba})/2$. One can state the symmetry of a tensor compactly by writing

$$S^{ab} = S^{(ab)}, \qquad A_{ab} = A_{[ab]}.$$

The slots being exchanged must have the same tensor type. For a mixed tensor, a metric can first be used to put the selected indices at the same level.

30.2 Products that are symmetric or antisymmetric

Example 30.1 (A repeated covector)**.** Let $T_{ab} = v_a v_b$. Then

$$T_{ba} = v_b v_a = v_a v_b = T_{ab},$$

because components are real numbers and commute under ordinary multiplication. Thus $v \otimes v$ is symmetric. Likewise $v^a v^b$ is symmetric in its upper indices.

Example 30.2 (The antisymmetric product of two covectors)**.** Let

$$A_{ab} = v_a w_b - w_a v_b.$$

Interchanging the indices gives

$$A_{ba} = v_b w_a - w_b v_a = w_a v_b - v_a w_b = -A_{ab}.$$

Thus the expression is antisymmetric. It is twice the antisymmetric part of $v_a w_b$:

$$A_{ab} = 2v_{[a}w_{b]}.$$

In particular, every diagonal component of an antisymmetric tensor vanishes: for a fixed a with no summation, $A_{aa} = -A_{aa}$ implies $A_{aa} = 0$. This is a statement about the components, distinct from taking a trace with a metric.

30.3 Symmetric–antisymmetric contractions vanish

Proposition 30.3. *If $S^{ab} = S^{ba}$ and $A_{ab} = -A_{ba}$, then*

$$S^{ab}A_{ab} = 0. \tag{140}$$

Proof. Using antisymmetry, then symmetry, then relabeling the dummy indices $a \leftrightarrow b$, we find

$$S^{ab}A_{ab} = -S^{ab}A_{ba} = -S^{ba}A_{ba} = -S^{ab}A_{ab}.$$

The contraction is a real number equal to its own negative, so it is zero. □

Taking $S^{ab} = v^a v^b$ gives the useful special case

$$A_{ab}v^a v^b = 0. \tag{141}$$

Equivalently, an antisymmetric bilinear form vanishes when both arguments are equal. Taking $S^{ab} = g^{ab}$ shows that the metric trace of an antisymmetric tensor vanishes. More generally,

$$S^{ab}T_{ab} = S^{ab}T_{(ab)}, \qquad A^{ab}T_{ab} = A^{ab}T_{[ab]}.$$

Only the part with the matching symmetry contributes to the contraction.

30.4 Symmetrizing more indices or only selected slots

Symmetrization of p indices averages over all $p!$ permutations. Antisymmetrization uses the same average, with a plus sign for an even permutation and a minus sign for an odd permutation. In terms of the permutation group S_p,

$$T_{(a_1\cdots a_p)} = \frac{1}{p!}\sum_{\pi\in S_p} T_{a_{\pi(1)}\cdots a_{\pi(p)}}, \tag{142}$$

$$T_{[a_1\cdots a_p]} = \frac{1}{p!}\sum_{\pi\in S_p} \operatorname{sgn}(\pi) T_{a_{\pi(1)}\cdots a_{\pi(p)}}. \tag{143}$$

The normalization ensures that a tensor already possessing the specified symmetry is unchanged.

For a type $(0,3)$ tensor, the full expressions contain six terms:

$$\begin{aligned} T_{(abc)} = \frac{1}{6}\big(&T_{abc} + T_{acb} + T_{bac} \\ &+ T_{bca} + T_{cab} + T_{cba}\big), \end{aligned} \tag{144}$$

$$\begin{aligned} T_{[abc]} = \frac{1}{6}\big(&T_{abc} - T_{acb} - T_{bac} \\ &+ T_{bca} + T_{cab} - T_{cba}\big). \end{aligned} \tag{145}$$

The cyclic orderings (abc), (bca), and (cab) are even permutations, while the other three orderings are odd.

One need not symmetrize every slot. For a type $(0,5)$ tensor, for example,

$$T_{(ab)cde} = \frac{1}{2}(T_{abcde} + T_{bacde}). \tag{146}$$

Only the first two indices are exchanged; c, d, e remain in their original positions. The analogous expression with square brackets uses a minus sign. This permits different symmetry requirements on different subsets of the indices of a tensor.

30.5 Mixed symmetry at higher rank

Although (139) holds for every two-index tensor, a general three-index tensor does *not* satisfy

$$T_{abc} = T_{(abc)} + T_{[abc]}.$$

There can be a remaining part with neither complete symmetry nor complete antisymmetry.

To see this explicitly, use a basis in dimension at least two and let the only nonzero component be $T_{112} = 1$. In the six-term symmetrization of the 112 component, the value T_{112} occurs twice, while the other four contributions vanish. Hence

$$T_{(112)} = \frac{1}{3}.$$

Complete antisymmetrization vanishes whenever two indices are equal, so $T_{[112]} = 0$. Therefore

$$T_{112} = 1 \neq \frac{1}{3} + 0.$$

This example isolates the additional possibility of mixed symmetry. For a tensor with more than two indices, specifying which slots are symmetrized is essential.

Part IV
Differential Geometry

31 Manifolds and global topology

Lecture 19

The mathematical model of spacetime in general relativity is a smooth four-dimensional manifold equipped with a Lorentzian metric. These are two separate structures. The manifold provides points, neighborhoods, coordinates, and a consistent notion of differentiation. The metric provides inner products of tangent vectors and hence lengths, proper times, angles where appropriate, and causal character. We can therefore develop the manifold and its calculus before introducing a metric.

31.1 What it means to look locally like Euclidean space

An n-dimensional manifold is locally modeled on $\mathbb{R}^n$: every point p has an open neighborhood U that can be put into one-to-one, continuous correspondence with an open subset of $\mathbb{R}^n$, with continuous inverse. The usual manifold assumptions also require the space to be Hausdorff and to have a countable basis of open sets. The first separates distinct points by neighborhoods; the second excludes pathological spaces that cannot be described by a countable collection of local coordinate data.

The coordinates label points in a neighborhood, not necessarily on the whole manifold. The dimension is the number of independent local coordinates. Neither a global coordinate system nor an embedding into a larger Euclidean space is required by this definition. A drawing of a curved surface in three-dimensional space is useful intuition, but the geometry can be defined intrinsically.

The word *locally* is essential. An observer restricted to a sufficiently small arc of a circle can use a real coordinate along it. Globally, the arc closes into a loop, and one coordinate cannot describe the whole circle with the properties required of a chart. Likewise, small regions of a sphere look like regions of a plane, although the complete sphere does not have the topology of $\mathbb{R}^2$.

31.2 Why coordinate patches are open

Consider the circle S^1. An angle in the range $0 \le \theta < 2\pi$ labels every point once, but this does not give one global chart: the range is not open in $\mathbb{R}$, and the inverse angle function is discontinuous where the endpoints are identified. Allowing an interval slightly longer than 2π makes the range open but repeats some points, so the map is no longer one-to-one.

One can instead use two angular patches. Let

$$U = S^1 \setminus \{(-1,0)\}, \qquad V = S^1 \setminus \{(1,0)\}.$$

Use $-\pi < \theta_U < \pi$ on U and $0 < \theta_V < 2\pi$ on V. Each patch is open in the circle, and $U \cup V = S^1$. The overlap has two connected components, on which the transition

functions are

$$\theta_V = \begin{cases} \theta_U, & 0 < \theta_U < \pi, \\ \theta_U + 2\pi, & -\pi < \theta_U < 0. \end{cases}$$

Each branch is smooth; there is no requirement that disconnected pieces of an overlap be described by the same algebraic formula. Alternatively, two slightly overlapping arcs centered on opposite points can be used. These are the two patches indicated in the sketch.

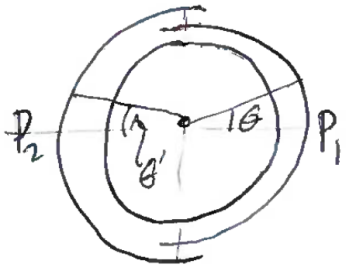

Figure 35: Two overlapping angular patches cover a circle. Each coordinate is local even though the underlying circle is one global object.

The circle need not be geometrically round. A deformation of its shape preserves the manifold structure and topology. Without a metric, there is no notion of a circle's radius, circumference, or roundness.

31.3 The sphere and stereographic coordinates

The same distinction appears for the two-sphere S^2. One can wrap a plane around the sphere so that it covers all points except one pole, but a second patch is needed to cover that remaining point. Stereographic projection makes this statement explicit.

Use the unit sphere in $\mathbb{R}^3$ only as a convenient model:

$$X^2 + Y^2 + Z^2 = 1.$$

Projection from the north pole $N = (0, 0, 1)$ onto the equatorial plane gives coordinates

$$u = \frac{X}{1 - Z}, \qquad v = \frac{Y}{1 - Z}, \qquad S^2 \setminus \{N\} \longrightarrow \mathbb{R}^2.$$

To obtain the inverse, set $s^2 = u^2 + v^2$. The sphere equation implies $s^2(1-Z)^2 + Z^2 = 1$, whose non-pole solution gives

$$X = \frac{2u}{1 + s^2}, \qquad Y = \frac{2v}{1 + s^2}, \qquad Z = \frac{s^2 - 1}{1 + s^2}.$$

Every finite (u, v) corresponds to exactly one point away from the north pole. Approaching that pole sends s to infinity, explaining why it is absent from this chart.

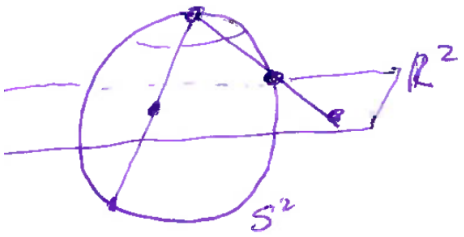

Figure 36: Stereographic projection gives planar coordinates on the sphere away from the projection pole. A chart projected from the opposite pole completes the atlas.

A second chart, projected from the south pole, has coordinates

$$\widetilde{u} = \frac{X}{1+Z}, \qquad \widetilde{v} = \frac{Y}{1+Z}.$$

On the overlap, where $s \neq 0$, the coordinate change is

$$(\widetilde{u}, \widetilde{v}) = \frac{(u, v)}{u^2 + v^2}.$$

It is smooth and has a smooth inverse of the same form. Thus the two patches define a smooth atlas. The manifold S^2 refers to the topology and smooth structure; a choice of round metric is additional information.

31.4 Examples and nonexamples

For connected one-dimensional manifolds without boundary, the two possibilities are the line $\mathbb{R}$ and the circle S^1, up to diffeomorphism. A crossing made by two intersecting lines is not a one-dimensional manifold at its intersection. Removing the central point from a small crossing leaves four branches, whereas removing a point from an open interval leaves only two. Consequently no neighborhood of the crossing can be homeomorphic to an open interval. It is also not a two-dimensional manifold: its neighborhoods do not look like open disks.

For connected, compact, orientable surfaces without boundary, topology is classified by the genus g, the number of handles. The sphere has genus zero, the torus genus one, and adding handles gives genus two, three, and so on. Compactness matters in this statement. The plane $\mathbb{R}^2$ is a noncompact surface, and the classification of noncompact surfaces requires more than a finite count of handles.

Orientability asks whether consistent orientations can be chosen throughout the manifold. The Klein bottle is a compact nonorientable surface without boundary. A Möbius strip is nonorientable and has a boundary; its interior is a noncompact surface without boundary. Both illustrate why orientability is a separate topological issue from local dimension.

Higher-dimensional examples include spheres S^n and tori T^n. The n-torus can be constructed from $\mathbb{R}^n$ by identifying

$$(x^1, \dots, x^n) \sim (x^1 + L_1 m_1, \dots, x^n + L_n m_n), \qquad m_i \in \mathbb{Z}.$$

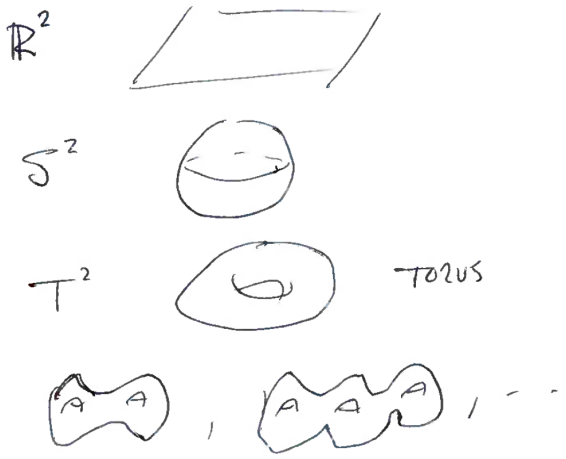

Figure 37: Examples of two-dimensional manifolds: the plane, sphere, torus, and surfaces with more handles. The plane is noncompact; the closed orientable examples are distinguished by genus.

For T^2, a fundamental parallelogram or rectangle has its opposite edges identified. Small neighborhoods crossing an identified edge continue smoothly through the opposite edge; the edge is not a physical boundary.

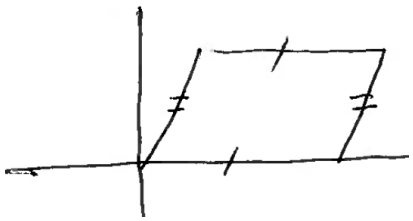

Figure 38: A fundamental parallelogram for a torus. Matching marks indicate the pairs of edges to be identified.

The local-coordinate definition applies in every dimension, while the global classification problem becomes much more involved beyond surfaces. In particular, the simple genus classification does not extend unchanged to three- or higher-dimensional manifolds.

32 Charts, atlases, and tangent spaces

Lecture 20

32.1 Formal coordinate maps and their domains

Let $A, B \subset \mathcal{M}$ be open sets in an n-dimensional manifold. A coordinate chart is an open set together with a coordinate map, for example

$$\rho : A \longrightarrow \rho(A) \subset \mathbb{R}^n, \qquad \phi : B \longrightarrow \phi(B) \subset \mathbb{R}^n.$$

The images are open, and the maps are invertible onto their images. Their inverses have the precise domains

$$\rho^{-1} : \rho(A) \longrightarrow A, \qquad \phi^{-1} : \phi(B) \longrightarrow B.$$

They need not be defined on all of $\mathbb{R}^n$. An atlas is a collection of such charts whose domains cover $\mathcal{M}$.

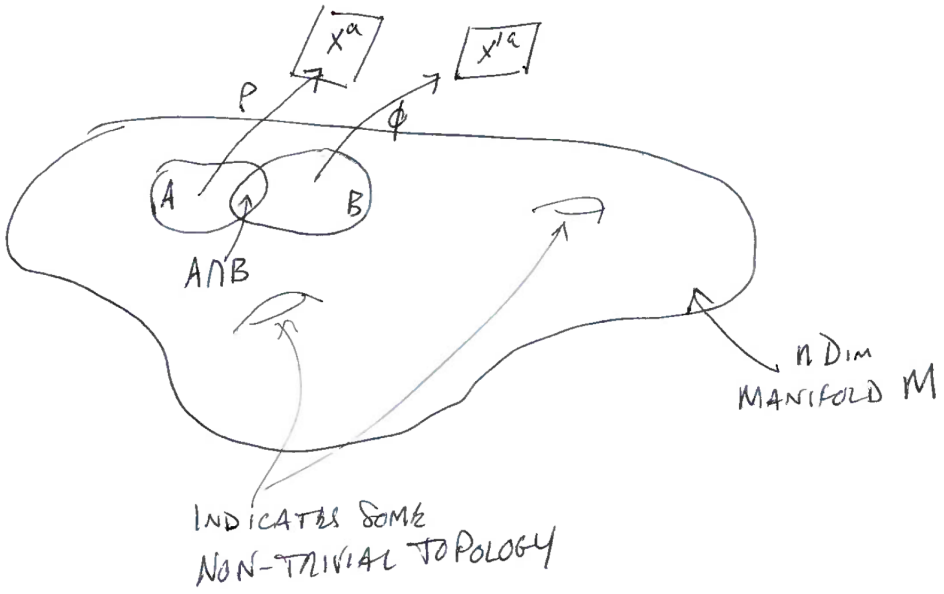

Figure 39: Overlapping coordinate patches on a manifold. Each chart maps a neighborhood into a Euclidean coordinate region; the topology of the manifold can remain nontrivial.

On an overlap $A \cap B$, the two coordinate descriptions are related by

$$\phi \circ \rho^{-1} : \rho(A \cap B) \longrightarrow \phi(A \cap B),$$
$$\rho \circ \phi^{-1} : \phi(A \cap B) \longrightarrow \rho(A \cap B).$$

In coordinates these are $x'^a = x'^a(x^b)$ and their inverses $x^a = x^a(x'^b)$. The map on the right in a composition acts first: starting with coordinates x, apply ρ^{-1} to recover the manifold point, then apply ϕ to obtain x'.

A smooth atlas requires every transition map to be C^∞ on its overlap. This is what makes differentiation consistent between charts. If a scalar function f is represented by $f \circ \rho^{-1}$ in one chart and $f \circ \phi^{-1}$ in another, the two representations are related by a smooth change of variables, so ordinary multivariable calculus relates their coordinate derivatives. One often performs calculations within a single chart, but the transformation laws retain the meaning of the result independently of that chart.

Throughout this part, indices $a, b, \ldots = 1, \ldots, n$ label coordinates on a general manifold. For spacetime these can instead be written $\mu, \nu, \ldots = 0, 1, 2, 3$; the tensor rules are identical.

32.2 Tangent vectors from curves

Fix a point $p \in \mathcal{M}$ and a parametrized curve $\gamma(\lambda)$ passing through it at $\lambda = \lambda_0$. In local coordinates, the curve is described by $x^a(\lambda)$, and its tangent components are

$$V^a = \left.\frac{\mathrm{d}x^a}{\mathrm{d}\lambda}\right|_{\lambda_0}.$$

These numbers are not a coordinate-independent definition of the vector: they change when coordinates change. The geometric object must include its basis.

A convenient intrinsic definition is its action on a smooth scalar function:

$$V(f) = \left.\frac{\mathrm{d}}{\mathrm{d}\lambda} f\big(\gamma(\lambda)\big)\right|_{\lambda_0}. \tag{147}$$

The chain rule gives

$$\begin{aligned} V(f) &= \left.\frac{\mathrm{d}x^a}{\mathrm{d}\lambda}\frac{\partial f}{\partial x^a}\right|_p \\ &= V^a \left.\frac{\partial}{\partial x^a}\right|_p f. \end{aligned}$$

Since this holds for every f, we identify the tangent vector itself as

$$V = V^a \left.\frac{\partial}{\partial x^a}\right|_p, \qquad e_a = \left.\frac{\partial}{\partial x^a}\right|_p. \tag{148}$$

The operators e_a are the coordinate-basis vectors. They are tangent to the coordinate curves along which only x^a varies.

The vector V is a linear map from smooth functions to real numbers at p. It also satisfies the product rule

$$V(fg) = f(p)V(g) + g(p)V(f).$$

Such a map is called a derivation at p. Its action depends on the first-order behavior of a function near p, not merely on the value of the function at that point. Applying a vector to a coordinate function recovers a component:

$$V(x^a) = V^b \partial_b x^a = V^a.$$

32.3 Why tangent vectors form an n-dimensional vector space

All tangent vectors at p form the tangent space $T_p\mathcal{M}$. The curve and its parametrization matter: reparametrizing the same geometric curve changes the scale of its tangent, and reversing the parameter reverses its direction. Considering all parametrized curves through p produces every possible set of components V^a.

To see this explicitly, choose any n real numbers v^a. For sufficiently small λ, the coordinate curve

$$x^a(\lambda) = x^a(p) + \lambda v^a$$

lies inside the chart. Its tangent at $\lambda = 0$ is $v^a \partial_a$. Thus every linear combination of the coordinate derivatives occurs as a tangent to some curve.

If $U = u^a \partial_a$ and $V = v^a \partial_a$, then

$$\alpha U + \beta V = (\alpha u^a + \beta v^a)\partial_a$$

is again such a tangent vector. The basis is linearly independent because $w^a \partial_a = 0$, applied to x^b, gives $w^b = 0$. It spans by (148), so $\dim T_p\mathcal{M} = n$. Different curves can determine the same tangent at p: they agree to first order there even if they separate at higher order. The tangent vector captures this common first-order information.

32.4 How coordinate bases transform

On a chart overlap, write $x'^a = x'^a(x^b)$ with inverse $x^b = x^b(x'^a)$. For any function f, the chain rule gives

$$\frac{\partial f}{\partial x'^a} = \frac{\partial x^b}{\partial x'^a}\frac{\partial f}{\partial x^b}.$$

Because this is an identity of differential operators acting on arbitrary functions,

$$e'_a = \frac{\partial}{\partial x'^a} = \frac{\partial x^b}{\partial x'^a}\, e_b. \tag{149}$$

The inverse relation is

$$e_a = \frac{\partial x'^b}{\partial x^a}\, e'_b.$$

Define the Jacobian matrices

$$J^a{}_b = \frac{\partial x'^a}{\partial x^b}, \qquad K^a{}_b = \frac{\partial x^a}{\partial x'^b}.$$

Differentiating $x^a = x^a(x'(x))$ proves that they are inverses:

$$K^a{}_c J^c{}_b = \delta^a{}_b, \qquad J^a{}_c K^c{}_b = \delta^a{}_b. \tag{150}$$

Consequently transforming the basis and transforming back returns the original differential operator.

32.5 How vector components transform

The chain rule along the curve gives

$$V'^a = \frac{\mathrm{d}x'^a}{\mathrm{d}\lambda} = \frac{\partial x'^a}{\partial x^b}\frac{\mathrm{d}x^b}{\mathrm{d}\lambda}.$$

Thus

$$V'^a = J^a{}_b V^b, \qquad e'_a = K^b{}_a e_b. \tag{151}$$

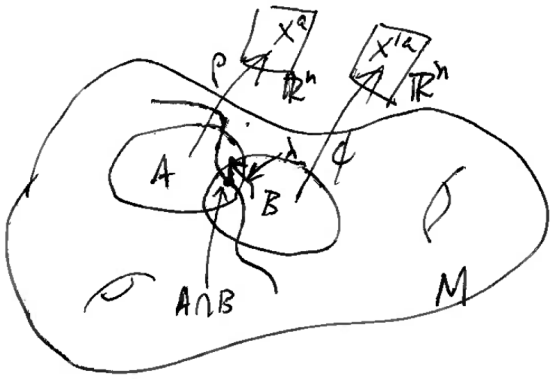

Figure 40: The same curve and tangent vector can be described in two overlapping charts. The coordinate labels and basis vectors change together.

The components and basis transform inversely. Their contraction is unchanged:

$$\begin{aligned} V'^a e'_a &= (J^a{}_b V^b)(K^c{}_a e_c) \\ &= V^b (K^c{}_a J^a{}_b) e_c \\ &= V^b \delta^c{}_b e_c = V^b e_b. \end{aligned}$$

This verifies in components that V is the same geometric vector in either coordinate description. The argument applies equally to a redefinition of coordinates within one chart and to a transition between two distinct charts.

33 Cartesian and polar coordinates

Lectures 21–22

33.1 Components, Jacobians, and inverse transformations

On a local polar patch of the plane with $r > 0$, let

$$x = r\cos\theta, \qquad y = r\sin\theta.$$

A vector has two descriptions:

$$V = V^x \partial_x + V^y \partial_y = V^r \partial_r + V^\theta \partial_\theta.$$

The derivatives of (x, y) with respect to (r, θ) are particularly simple:

$$\frac{\partial(x,y)}{\partial(r,\theta)} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}.$$

Therefore

$$V^x = \cos\theta\, V^r - r\sin\theta\, V^\theta, \tag{152}$$

$$V^y = \sin\theta\, V^r + r\cos\theta\, V^\theta. \tag{153}$$

The determinant of this Jacobian is r, which is nonzero on the polar patch. Its inverse gives

$$\begin{pmatrix} V^r \\ V^\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta/r & \cos\theta/r \end{pmatrix} \begin{pmatrix} V^x \\ V^y \end{pmatrix}. \tag{154}$$

Equivalently,

$$V^r = \frac{x}{r}V^x + \frac{y}{r}V^y, \qquad V^\theta = -\frac{y}{r^2}V^x + \frac{x}{r^2}V^y.$$

The coordinate singularity at $r = 0$ is visible in these formulas; it does not make the Euclidean plane singular. It means this particular chart does not include the origin.

33.2 Coordinate bases are not necessarily unit bases

The basis transformation is

$$\partial_r = \cos\theta\,\partial_x + \sin\theta\,\partial_y,$$
$$\partial_\theta = -r\sin\theta\,\partial_x + r\cos\theta\,\partial_y.$$

In the Euclidean metric, ∂_r has unit length, while ∂_θ has length r. Moving through angle $\mathrm{d}\theta$ at fixed radius traverses physical distance $r\,\mathrm{d}\theta$.

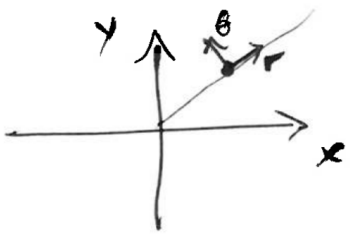

Figure 41: Radial and angular directions in the plane. The coordinate basis vector ∂_θ has length r, so it differs from the unit angular vector.

If $\widehat{e}_r = \partial_r$ and $\widehat{e}_\theta = r^{-1}\partial_\theta$ are the unit basis vectors, then

$$V = V^r\widehat{e}_r + (rV^\theta)\widehat{e}_\theta.$$

Thus a textbook using a unit polar basis reports angular component $V^{\widehat{\theta}} = rV^\theta$. The factor of r is a basis choice, not a contradiction between the vector transformation laws.

34 Vector fields and their commutator

34.1 Vector fields as first-order differential operators

A smooth vector field specifies a vector in $T_p\mathcal{M}$ at every point p, with components that vary smoothly in each chart. Write

$$U = U^a(x)\partial_a, \qquad V = V^a(x)\partial_a.$$

Acting on a scalar function produces another scalar function: $V(f) = V^a \partial_a f$. We can then apply U to $V(f)$. The resulting operator UV generally includes second derivatives, so it is not itself a vector field. The difference $UV - VU$ has a special cancellation.

Index placement alone does not make a collection of quantities a tensor. One must either construct a coordinate-independent geometric object or verify that its components satisfy the appropriate transformation law. The commutator provides a detailed example of both approaches.

34.2 Deriving the components and canceling second derivatives

Define the commutator, or Lie bracket, by its action on scalar functions:

$$[U,V](f) = U\big(V(f)\big) - V\big(U(f)\big). \tag{155}$$

Expand the two terms using the product rule:

$$\begin{aligned} U(Vf) &= U^b \partial_b (V^a \partial_a f) \\ &= U^b (\partial_b V^a)\partial_a f + U^b V^a \partial_b \partial_a f, \\ V(Uf) &= V^b \partial_b (U^a \partial_a f) \\ &= V^b (\partial_b U^a)\partial_a f + V^b U^a \partial_b \partial_a f. \end{aligned}$$

Subtracting gives

$$\begin{aligned} [U,V](f) = {} & (U^b \partial_b V^a - V^b \partial_b U^a)\partial_a f \\ & + (U^b V^a - V^b U^a)\partial_b \partial_a f. \end{aligned}$$

In the second term, interchange the dummy labels a and b in one of the summands. Smoothness implies $\partial_a \partial_b f = \partial_b \partial_a f$, so

$$U^b V^a \partial_b \partial_a f - U^a V^b \partial_b \partial_a f = U^b V^a (\partial_b \partial_a f - \partial_a \partial_b f) = 0.$$

Only a first-order differential operator remains:

$$[U,V] = W^a \partial_a, \qquad W^a = U^b \partial_b V^a - V^b \partial_b U^a. \tag{156}$$

The construction uses only the action of vectors on functions, so it is coordinate independent. It also proves directly that the commutator is a vector field.

34.3 Why either directional component derivative fails separately

It is instructive to verify covariance from the component formula itself. Define the two coordinate expressions

$$W_1^a = U^b \partial_b V^a, \qquad W_2^a = V^b \partial_b U^a, \qquad W^a = W_1^a - W_2^a.$$

Let $J^a{}_b = \partial x'^a/\partial x^b$ and $K^a{}_b = \partial x^a/\partial x'^b$. Then

$$U'^b = J^b{}_c U^c, \qquad V'^a = J^a{}_e V^e, \qquad \partial'_b = K^d{}_b \partial_d.$$

Substitute these into $W_1'^a$ and apply the derivative to both factors:

$$\begin{aligned} W_1'^a &= U'^b \partial'_b V'^a \\ &= (J^b{}_c U^c) K^d{}_b \partial_d (J^a{}_e V^e) \\ &= U^d \left[J^a{}_e \partial_d V^e + (\partial_d J^a{}_e) V^e \right] \\ &= J^a{}_e W_1^e + \frac{\partial^2 x'^a}{\partial x^d \partial x^e} U^d V^e. \end{aligned}$$

The first term is the desired vector transformation. The second contains second derivatives of the coordinate change and is not allowed in a vector transformation law. Thus $U^b \partial_b V^a$ does not, by itself, define vector components under a general nonlinear coordinate change.

Interchanging U and V gives

$$W_2'^a = J^a{}_e W_2^e + \frac{\partial^2 x'^a}{\partial x^d \partial x^e} V^d U^e.$$

Their difference is

$$\begin{aligned} W'^a &= J^a{}_e W^e + \frac{\partial^2 x'^a}{\partial x^d \partial x^e} (U^d V^e - V^d U^e) \\ &= J^a{}_e W^e. \end{aligned}$$

The extra term vanishes because the second derivative is symmetric in d, e, whereas the expression in parentheses is antisymmetric. This is the explicit transformation-law proof that the bracket is a vector.

34.4 Simple checks on the bracket

Coordinate basis fields commute:

$$[\partial_a, \partial_b] = 0,$$

because their action on any smooth scalar is the difference of mixed partial derivatives. General fields need not commute because their components vary. For instance, on a plane,

$$U = \partial_x, \qquad V = x\partial_y \quad \Longrightarrow \quad [U, V] = \partial_y.$$

The component formula also gives

$$[U, fV] = f[U, V] + U(f)V.$$

Thus the bracket is linear over constant real coefficients, but it is not a bilinear tensorial operation over arbitrary scalar functions. It measures the failure of the two directional derivative operations to commute.

35 Cotangent vectors and differential one-forms

35.1 The differential of a function

The dual space of $T_p\mathcal{M}$ is the cotangent space $T_p^*\mathcal{M}$. Its elements are linear maps from tangent vectors to real numbers. Given a smooth scalar f, define its differential at p by

$$(\mathrm{d}f)_p(V) = V(f). \tag{157}$$

This is a covector because

$$\begin{aligned}\mathrm{d}f(\alpha U + \beta V) &= (\alpha U + \beta V)(f)\\ &= \alpha U(f) + \beta V(f)\\ &= \alpha\,\mathrm{d}f(U) + \beta\,\mathrm{d}f(V).\end{aligned}$$

The differential is therefore a geometric object, independently of how the function is described in coordinates.

35.2 The dual coordinate basis

Take the function $f = x^a$ and the vector $e_b = \partial_b$. Then

$$\mathrm{d}x^a(e_b) = e_b(x^a) = \frac{\partial x^a}{\partial x^b} = \delta^a{}_b.$$

This is precisely the defining pairing of a dual basis. Thus the basis of $T_p^*\mathcal{M}$ dual to $\{\partial_a\}$ is $\{\mathrm{d}x^a\}$.

An arbitrary covector can be expanded as

$$\omega = \omega_a\,\mathrm{d}x^a.$$

Its components are recovered by evaluation on the tangent basis:

$$\omega(e_b) = \omega_a \mathrm{d}x^a(e_b) = \omega_a \delta^a{}_b = \omega_b.$$

For $\omega = \mathrm{d}f$, this gives

$$\mathrm{d}f = (\partial_a f)\mathrm{d}x^a. \tag{158}$$

The familiar total differential from multivariable calculus is exactly the coordinate expansion of the covector $\mathrm{d}f$. In particular, the lower-index quantities $\partial_a f$ are covector components.

No metric has been used. It is $\mathrm{d}f$ that is naturally defined by the function. A vector called the gradient requires a metric to raise the index: $(\mathrm{grad}\, f)^a = g^{ab}\partial_b f$. The distinction between the differential and the gradient becomes important away from Cartesian Euclidean coordinates.

35.3 Covector components and bases transform inversely

For a scalar f, the chain rule gives

$$\partial'_a f = \frac{\partial x^b}{\partial x'^a} \partial_b f.$$

The same lower-index transformation law applies to any covector:

$$\omega'_a = \frac{\partial x^b}{\partial x'^a} \omega_b = K^b{}_a \omega_b. \tag{159}$$

Meanwhile, applying the total-differential formula to the coordinate functions $x'^a(x)$ gives the transformation of the dual basis,

$$\mathrm{d}x'^a = \frac{\partial x'^a}{\partial x^b} \mathrm{d}x^b = J^a{}_b \mathrm{d}x^b. \tag{160}$$

Their product is invariant:

$$\begin{aligned} \omega'_a \mathrm{d}x'^a &= (K^b{}_a \omega_b)(J^a{}_c \mathrm{d}x^c) \\ &= \omega_b \delta^b{}_c \mathrm{d}x^c = \omega_b \mathrm{d}x^b. \end{aligned}$$

The four transformation laws can be kept together:

$$\begin{aligned} V'^a &= J^a{}_b V^b, & e'_a &= K^b{}_a e_b, \\ \omega'_a &= K^b{}_a \omega_b, & \mathrm{d}x'^a &= J^a{}_b \mathrm{d}x^b. \end{aligned}$$

Vector components use derivatives of new coordinates with respect to old ones; covector components use derivatives of old coordinates with respect to new ones. Each basis transforms inversely to the corresponding components.

35.4 The scalar pairing is invariant

For a vector V and covector ω, their pairing is

$$\omega(V) = \omega_a V^a.$$

Compute it in the primed chart:

$$\begin{aligned} \omega'_a V'^a &= (K^b{}_a \omega_b)(J^a{}_c V^c) \\ &= \omega_b \delta^b{}_c V^c = \omega_b V^b. \end{aligned}$$

It is therefore a scalar. This contraction does not require a metric: a covector is already defined as a linear map on vectors. A metric instead provides a way to associate a covector with a vector, allowing two vectors to be paired.

36 General tensors and the metric

36.1 Tensor-product bases

At a point p, a tensor of type (k, ℓ) belongs to

$$(T_p\mathcal{M})^{\otimes k} \otimes (T_p^*\mathcal{M})^{\otimes \ell}.$$

Equivalently, it is a multilinear map taking k covectors and ℓ vectors as arguments and returning a real number. A tensor field assigns such a tensor smoothly at every point.

Using the coordinate bases, its expansion is

$$T = T^{a_1 \cdots a_k}{}_{b_1 \cdots b_\ell}\, e_{a_1} \otimes \cdots \otimes e_{a_k} \otimes \mathrm{d}x^{b_1} \otimes \cdots \otimes \mathrm{d}x^{b_\ell}.$$

An upper index corresponds to a tangent-vector factor; a lower index corresponds to a cotangent-vector factor. The tensor itself does not depend on the choice of these bases.

36.2 The full tensor transformation law

Substitute the basis transformations into the primed expansion of T and compare coefficients in the unprimed basis. There must be one factor of J for each upper component index and one factor of K for each lower component index:

$$T'^{a_1 \cdots a_k}{}_{b_1 \cdots b_\ell} = J^{a_1}{}_{c_1} \cdots J^{a_k}{}_{c_k} K^{d_1}{}_{b_1} \cdots K^{d_\ell}{}_{b_\ell} T^{c_1 \cdots c_k}{}_{d_1 \cdots d_\ell}. \tag{161}$$

For example, a $(1,1)$ tensor transforms as

$$A'^{a}{}_{b} = J^{a}{}_{c} K^{d}{}_{b} A^{c}{}_{d}.$$

Each contracted upper–lower pair produces a factor $KJ = \mathrm{id}$. Consequently contractions transform as tensors with those two indices removed. For example, $A^a{}_a$ is a scalar.

The transformation rule can also be used as a practical criterion: a collection of coordinate arrays with these compatibility properties on overlaps specifies a tensor field. This criterion is meaningful only when the rule is imposed under general allowed coordinate changes. An array that happens to have the right number or placement of indices need not satisfy it.

36.3 The metric as a symmetric bilinear form

The metric is a smooth, symmetric, nondegenerate $(0,2)$ tensor field:

$$g = g_{ab}\, \mathrm{d}x^a \otimes \mathrm{d}x^b, \qquad g_{ab} = g_{ba}.$$

It acts on two vectors by

$$g(U, V) = g_{ab} U^a V^b.$$

The line element is conventionally written

$$\mathrm{d}s^2 = g_{ab}\,\mathrm{d}x^a \mathrm{d}x^b,$$

with the tensor-product symbol suppressed. Under a change of coordinates,

$$g'_{ab} = K^c{}_a K^d{}_b g_{cd} = \frac{\partial x^c}{\partial x'^a}\frac{\partial x^d}{\partial x'^b} g_{cd}. \tag{162}$$

This is also obtained directly by replacing $\mathrm{d}x^c = (\partial x^c/\partial x'^a)\mathrm{d}x'^a$ in the line element:

$$\begin{aligned}\mathrm{d}s^2 &= g_{cd}\,\mathrm{d}x^c \mathrm{d}x^d \\ &= g_{cd} K^c{}_a K^d{}_b\,\mathrm{d}x'^a \mathrm{d}x'^b = g'_{ab}\,\mathrm{d}x'^a \mathrm{d}x'^b.\end{aligned}$$

Thus the earlier coordinate-change rule for line elements is a special case of the general tensor transformation law.

37 Inverse metrics, identity tensors, and signature

Lecture 23

37.1 The inverse metric is a contravariant tensor

Nondegeneracy means that g_{ab} has an inverse, denoted g^{ab}:

$$g^{ac} g_{cb} = \delta^a{}_b.$$

Since g is symmetric, its inverse is symmetric. In matrix notation the metric transformation is $g' = K^{\mathsf{T}} g K$, so taking the inverse yields

$$(g')^{-1} = J g^{-1} J^{\mathsf{T}}.$$

In indices,

$$g'^{ab} = J^a{}_c J^b{}_d g^{cd}. \tag{163}$$

It is therefore a $(2,0)$ tensor. A direct check gives

$$\begin{aligned} g'^{ac} g'_{cb} &= J^a{}_e J^c{}_f g^{ef} K^h{}_c K^j{}_b g_{hj} \\ &= J^a{}_e g^{ef} \delta^h{}_f g_{hj} K^j{}_b \\ &= J^a{}_e \delta^e{}_j K^j{}_b = \delta^a{}_b. \end{aligned}$$

The metric and inverse metric lower and raise indices:

$$V_a = g_{ab} V^b, \qquad \omega^a = g^{ab} \omega_b.$$

These operations are inverse to each other, and their tensor transformation properties follow by contracting the laws for g, g^{-1}, and the original tensor.

37.2 Why only the mixed Kronecker tensor is the identity

The identity map on $T_p\mathcal{M}$ is the $(1,1)$ tensor

$$\mathrm{id} = e_a \otimes \mathrm{d}x^a, \qquad \mathrm{id}(V) = V.$$

Its components are $\delta^a{}_b$ in every coordinate basis. Indeed,

$$\begin{aligned} \delta'^a{}_b &= J^a{}_c K'^d{}_b \delta^c{}_d \\ &= J^a{}_c K'^c{}_b = \delta^a{}_b. \end{aligned}$$

This statement relies on one upper and one lower index. A numerical array δ_{ab} with two lower indices is not the same coordinate-invariant identity tensor. In Euclidean Cartesian coordinates it agrees with the metric components, but in polar coordinates that metric is

$$g_{ab} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}, \qquad g^{ab} = \begin{pmatrix} 1 & 0 \\ 0 & r^{-2} \end{pmatrix}.$$

Neither array is the identity matrix in all coordinate systems. Lowering the upper index of $\delta^a{}_b$ with the metric gives g_{ab}, not a fixed numerical two-lower-index array.

37.3 Canonical form and metric signature

At any specified point, a real symmetric nondegenerate metric can be brought by an invertible basis change to a diagonal form with entries $+1$ and -1. One way to see this is to diagonalize the symmetric matrix, obtaining real nonzero eigenvalues λ_i, and then rescale each eigenvector by $|\lambda_i|^{-1/2}$. The entries become $\operatorname{sgn}\lambda_i$. A coordinate change can realize this prescribed invertible linear change of basis at the point.

If there are r positive and s negative entries, then $n = r + s$ and the signature is (r, s). The counts of positive and negative directions are invariant under real invertible changes of coordinates. They cannot change continuously on a connected manifold while the metric remains nondegenerate: a change of sign would require an eigenvalue to pass through zero.

A positive-definite metric has signature $(n, 0)$ and defines a Riemannian geometry. A Lorentzian metric in n spacetime dimensions has one negative and $n - 1$ positive directions. In four spacetime dimensions, our convention is

$$\eta_{\mu\nu} = \mathrm{diag}(-1, +1, +1, +1),$$

so the count of positive and negative entries is $(3, 1)$. At a point one can always choose coordinates with this metric matrix. That pointwise statement does not imply that the metric has the same constant form throughout an open region; curvature is the obstruction to doing this in general.

38 Why ordinary partial derivatives fail

38.1 Ordinary derivatives are not tensor components

Tensor equations retain their meaning in every coordinate system. Derivatives are indispensable in physical equations, so we need a derivative operation that takes tensors

to tensors. The ordinary partial derivative of scalar functions has this property as a covector, but the ordinary partial derivatives of vector or higher-tensor components generally do not.

Let $V^b(x)$ be a vector field and form the coordinate array

$$A^b{}_a = \partial_a V^b.$$

If this were a $(1,1)$ tensor, it would transform as $A'^b{}_a = K^c{}_a J^b{}_d A^d{}_c$. Instead, direct differentiation gives

$$\begin{aligned} A'^b{}_a &= \partial'_a V'^b \\ &= K^c{}_a \partial_c (J^b{}_d V^d) \\ &= K^c{}_a J^b{}_d \partial_c V^d + K^c{}_a (\partial_c J^b{}_d) V^d \\ &= K^c{}_a J^b{}_d A^d{}_c + \frac{\partial x^c}{\partial x'^a} \frac{\partial^2 x'^b}{\partial x^c \partial x^d} V^d. \end{aligned} \tag{164}$$

The second term spoils the tensor law. For a constant linear coordinate change, such as a Lorentz transformation between inertial Cartesian frames, these second derivatives vanish. General relativity must also accommodate nonlinear coordinate changes, for which the extra term is present.

In particular, $\partial_a V^b = 0$ in one chart does not imply $\partial'_a V'^b = 0$ in another. Equations involving such an array alone would not express a coordinate-independent physical condition. The problem is that changing vector components can reflect both a changing vector and a changing coordinate basis.

38.2 A constant vector in polar coordinates

Consider the Euclidean plane with the constant vector field

$$V = C\partial_y, \qquad V^x = 0, \quad V^y = C,$$

where C is constant. Every Cartesian component derivative vanishes.

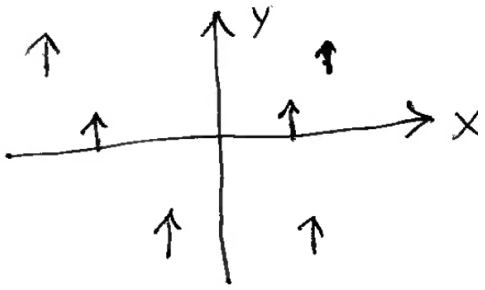

Figure 42: A vector field with constant Cartesian components. Its geometric direction is unchanged from point to point, although its polar coordinate components vary.

Using (154), its polar components are

$$V^r = C\sin\theta, \qquad V^\theta = \frac{C\cos\theta}{r}.$$

Their derivatives are

$$\partial_r V^r = 0, \qquad \partial_\theta V^r = C\cos\theta,$$
$$\partial_r V^\theta = -\frac{C\cos\theta}{r^2}, \qquad \partial_\theta V^\theta = -\frac{C\sin\theta}{r}.$$

The same field therefore has nonzero partial component derivatives in the polar chart. The changing radial and angular basis vectors account for this variation; it is not a physical change of the constant vector field.

38.3 The defining properties of a covariant derivative

A covariant derivative ∇ is designed to map tensor fields to tensor fields while accounting for the change of basis. Acting on a tensor of type (k, ℓ), it produces one of type $(k, \ell + 1)$; its extra lower index specifies the derivative direction. The required properties begin with:

1. **Linearity.** For tensor fields of the same type and constant real numbers α, β,

$$\nabla(\alpha T + \beta S) = \alpha\nabla T + \beta\nabla S.$$

2. **The tensor-product rule.** Differentiation satisfies the Leibniz rule,

$$\nabla_a(T \otimes S) = (\nabla_a T) \otimes S + T \otimes (\nabla_a S).$$

   For example,

$$\nabla_a(\omega_b V^c) = (\nabla_a \omega_b)V^c + \omega_b \nabla_a V^c.$$

3. **Scalar functions.** For a scalar f, the covariant derivative agrees with the ordinary derivative:

$$\nabla_a f = \partial_a f.$$

   This already transforms correctly as the covector $\mathrm{d}f$.

Combining the scalar property with the product rule gives

$$\nabla_a(fT) = (\partial_a f)T + f\nabla_a T.$$

Thus a covariant derivative is linear over constant coefficients but differentiates position-dependent coefficients. To construct its action on vectors and other tensors, one must specify how the local basis changes. That additional geometric data is a connection.

39 Connections and covariant differentiation

Lecture 24

In this part, lower-case indices $a, b, c, \ldots$ denote general coordinate indices on the manifold under discussion. They range over two values for a surface and four values for spacetime. The construction of a covariant derivative applies in either case.

39.1 A constant vector field in polar coordinates

Recall the constant Cartesian field of Section 38.2. Its components $V^x = 0$, $V^y = C$ become

$$V^r = C\sin\theta, \qquad V^\theta = \frac{C\cos\theta}{r}, \tag{165}$$

so ordinary component derivatives need not vanish even when the field is constant. A covariant derivative must account for the changing coordinate basis as well as the changing components.

The covariant derivative introduced in Section 38.3 is designed to avoid this problem. It maps a tensor of type (k, ℓ) to one of type $(k, \ell + 1)$, is linear, obeys the tensor-product Leibniz rule, and reduces to the ordinary derivative on scalar functions:

$$\nabla_a f = \partial_a f, \qquad \nabla_a (T \otimes S) = (\nabla_a T) \otimes S + T \otimes (\nabla_a S).$$

The derivative of a scalar already transforms as a covector. What remains is to define the derivative of fields with tensor indices.

39.2 The connection coefficients

For a vector field, introduce coefficients $\Gamma^b{}_{ac}$ by

$$\nabla_a V^b = \partial_a V^b + \Gamma^b{}_{ac} V^c. \tag{166}$$

The correction is linear in V, just as the unwanted term in the transformation of $\partial_a V^b$ is. The coefficients are called a *connection*, or connection coefficients. They are not themselves the components of a tensor.

To obtain their transformation law, let $x^{a'}$ be new coordinates and write

$$J^{a'}{}_b = \frac{\partial x^{a'}}{\partial x^b}, \qquad K^a{}_{b'} = \frac{\partial x^a}{\partial x^{b'}}, \qquad JK = KJ = I.$$

Here I denotes the identity matrix. Differentiating $V^{b'} = J^{b'}{}_d V^d$ gives

$$\partial_{a'} V^{b'} = K^c{}_{a'} J^{b'}{}_d \, \partial_c V^d + K^c{}_{a'} (\partial_c J^{b'}{}_d) V^d.$$

We require

$$\nabla_{a'} V^{b'} = K^c{}_{a'} J^{b'}{}_d \, \nabla_c V^d.$$

Substitute (166) on both sides, cancel the terms containing ∂V, and equate the coefficients of the arbitrary components V^d. Multiplication by $K^d{}_{c'}$ gives

$$\Gamma^{b'}{}_{a'c'} = J^{b'}{}_e K^d{}_{a'} K^f{}_{c'} \Gamma^e{}_{df} - K^d{}_{a'} K^f{}_{c'} \frac{\partial^2 x^{b'}}{\partial x^d \partial x^f}. \tag{167}$$

The first term has the form of a tensor transformation; the second, inhomogeneous term cancels the unwanted second derivative of the coordinate transformation in ∂V. Equivalently,

$$\Gamma^{b'}{}_{a'c'} = J^{b'}{}_{e} K^{d}{}_{a'} K^{f}{}_{c'} \Gamma^{e}{}_{df} + J^{b'}{}_{e} \frac{\partial^2 x^e}{\partial x^{a'} \partial x^{c'}}.$$

These expressions agree by differentiating the inverse-coordinate identity. The connection law guarantees that the full quantity $\nabla_a V^b$, although built from nontensorial pieces, is a tensor.

39.3 Contractions and the derivative of a covector

A further requirement is that covariant differentiation commute with contraction. For example, the scalar $u_b V^b$ must give the same answer whether the contraction is performed before or after differentiation:

$$\nabla_a (u_b V^b) = (\nabla_a u_b) V^b + u_b \nabla_a V^b = \partial_a (u_b V^b). \tag{168}$$

Expanding the ordinary derivative and substituting the vector derivative,

$$\begin{aligned} (\nabla_a u_b) V^b &= (\partial_a u_b) V^b + u_b \partial_a V^b - u_b (\partial_a V^b + \Gamma^{b}{}_{ac} V^c) \\ &= \big(\partial_a u_b - \Gamma^{c}{}_{ab} u_c\big) V^b, \end{aligned}$$

where dummy indices were relabeled in the final term. Since V^b is arbitrary,

$$\boxed{\nabla_a u_b = \partial_a u_b - \Gamma^{c}{}_{ab} u_c.} \tag{169}$$

The connection contributes a plus sign for an upper index and a minus sign for a lower index. This is precisely what allows the correction terms to cancel in a contraction.

The identity tensor is covariantly constant:

$$\nabla_a \delta^b{}_c = 0. \tag{170}$$

One way to see this is to write a contraction as multiplication by $\delta^b{}_c$ and compare differentiation before and after contraction.

Exercise 39.1. Verify (170) directly from the connection formula.

There is one correction for each index:

$$\nabla_a \delta^b{}_c = \partial_a \delta^b{}_c + \Gamma^{b}{}_{ad} \delta^d{}_c - \Gamma^{d}{}_{ac} \delta^b{}_d = 0 + \Gamma^{b}{}_{ac} - \Gamma^{b}{}_{ac} = 0.$$

39.4 The derivative of an arbitrary tensor

Applying the tensor-product rule to products of vectors and covectors gives the general expression

$$\begin{aligned} \nabla_a T^{b_1 \ldots b_k}{}_{c_1 \ldots c_\ell} &= \partial_a T^{b_1 \ldots b_k}{}_{c_1 \ldots c_\ell} \\ &\quad + \sum_{i=1}^{k} \Gamma^{b_i}{}_{ad} \, T^{b_1 \ldots b_{i-1} d \, b_{i+1} \ldots b_k}{}_{c_1 \ldots c_\ell} \\ &\quad - \sum_{j=1}^{\ell} \Gamma^{d}{}_{ac_j} \, T^{b_1 \ldots b_k}{}_{c_1 \ldots c_{j-1} d \, c_{j+1} \ldots c_\ell}. \end{aligned} \tag{171}$$

Thus there is one connection term for every tensor index. For a mixed rank-two tensor, for example,

$$\nabla_a T^b{}_c = \partial_a T^b{}_c + \Gamma^b{}_{ad} T^d{}_c - \Gamma^d{}_{ac} T^b{}_d.$$

Contraction of b with c cancels the last two contributions after relabeling their dummy indices, as required. At this stage no metric has been used, and these rules alone do not select a unique connection.

40 The Levi–Civita connection

Lectures 25–26

40.1 The freedom in choosing a connection

If $\Gamma^c{}_{ab}$ is a connection and $A^c{}_{ab}$ is any tensor of type $(1,2)$, then

$$\widetilde{\Gamma}^c{}_{ab} = \Gamma^c{}_{ab} + A^c{}_{ab}$$

is another connection. In particular,

$$\widetilde{\nabla}_a V^b = \nabla_a V^b + A^b{}_{ac} V^c$$

is the sum of two tensors. Conversely, the difference of two connections is a tensor: subtracting their transformation laws cancels the inhomogeneous second-derivative term.

General relativity uses the unique connection satisfying two additional conditions: vanishing torsion and compatibility with the metric. It is called the *Levi–Civita connection*; its coefficients are the *Christoffel symbols*.

40.2 Torsion and symmetry of the lower indices

In a coordinate basis, define the torsion tensor by

$$T^c{}_{ab} = \Gamma^c{}_{ab} - \Gamma^c{}_{ba}. \tag{172}$$

Although each connection coefficient is nontensorial, their antisymmetric combination is a tensor. The second derivative in (167) is symmetric in the two lower indices, so it drops out on antisymmetrization. Therefore $T^c{}_{ab}$ transforms homogeneously with one upper and two lower indices.

The torsion-free requirement is

$$T^c{}_{ab} = 0 \quad \Longleftrightarrow \quad \Gamma^c{}_{ab} = \Gamma^c{}_{ba}. \tag{173}$$

Because torsion is a tensor, if it vanishes in one coordinate system it vanishes in every coordinate system. The symmetry statement is being made in a coordinate basis; the connection formulas here all use such bases.

40.3 Metric compatibility and moving indices

Metric compatibility means

$$\nabla_a g_{bc} = 0. \tag{174}$$

The metric is covariantly constant, although its coordinate components can vary with position. The inverse metric is also covariantly constant. To prove this, differentiate $g^{bd} g_{dc} = \delta^b{}_c$:

$$0 = \nabla_a \delta^b{}_c = (\nabla_a g^{bd}) g_{dc} + g^{bd} \nabla_a g_{dc} = (\nabla_a g^{bd}) g_{dc}.$$

Multiplying by g^{ce} gives $\nabla_a g^{be} = 0$.

It follows that covariant differentiation commutes with raising and lowering indices:

$$\nabla_a V_b = \nabla_a (g_{bc} V^c) = g_{bc} \nabla_a V^c, \qquad \nabla_a u^b = \nabla_a (g^{bc} u_c) = g^{bc} \nabla_a u_c.$$

This property is indispensable when comparing tensor equations with indices in different positions.

40.4 Deriving the Christoffel formula

Write metric compatibility with three permutations of the indices:

$$\begin{aligned} \partial_a g_{bc} &= \Gamma^d{}_{ab} g_{dc} + \Gamma^d{}_{ac} g_{bd}, \\ \partial_b g_{ca} &= \Gamma^d{}_{bc} g_{da} + \Gamma^d{}_{ba} g_{cd}, \\ \partial_c g_{ab} &= \Gamma^d{}_{ca} g_{db} + \Gamma^d{}_{cb} g_{ad}. \end{aligned}$$

Add the first two equations and subtract the third. The terms containing $\Gamma^d{}_{ac}$ cancel those containing $\Gamma^d{}_{ca}$, and the $\Gamma^d{}_{bc}$ terms cancel the $\Gamma^d{}_{cb}$ terms. These cancellations use torsion freedom. The remaining two terms are equal:

$$\partial_a g_{bc} + \partial_b g_{ac} - \partial_c g_{ab} = 2 g_{cd} \Gamma^d{}_{ab}.$$

Multiplying by $\frac{1}{2} g^{ec}$ gives

$$\boxed{\Gamma^e{}_{ab} = \frac{1}{2} g^{ec} \big(\partial_a g_{bc} + \partial_b g_{ac} - \partial_c g_{ab} \big).} \tag{175}$$

This formula is manifestly symmetric in a, b. Substituting it into $\nabla_a g_{bc}$ verifies metric compatibility. The derivation proves uniqueness, because any connection satisfying both requirements must have this expression; the substitution proves existence.

In flat space with Cartesian coordinates the metric components are constant, so every Christoffel symbol vanishes and $\nabla = \partial$ on tensor components. This conclusion does not hold for curvilinear coordinates on the same flat space. Nonzero Christoffel symbols alone therefore do not establish nonzero curvature.

40.5 Christoffel symbols of the two-sphere

For the two-sphere of radius R,

$$\mathrm{d}\ell^2 = R^2\big(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\big),$$

so

$$g_{\theta\theta} = R^2, \qquad g_{\phi\phi} = R^2\sin^2\theta, \qquad g_{\theta\phi} = 0,$$

and

$$g^{\theta\theta} = \frac{1}{R^2}, \qquad g^{\phi\phi} = \frac{1}{R^2\sin^2\theta}, \qquad g^{\theta\phi} = 0.$$

The only nonzero derivative of a metric component is $\partial_\theta g_{\phi\phi} = 2R^2\sin\theta\cos\theta$. We can therefore compute every independent Christoffel symbol explicitly.

For two lower θ indices,

$$\begin{aligned}
\Gamma^{\theta}{}_{\theta\theta} &= \frac{1}{2}g^{\theta\theta}\partial_\theta g_{\theta\theta} = 0,\\
\Gamma^{\phi}{}_{\theta\theta} &= \frac{1}{2}g^{\phi\phi}\big(2\partial_\theta g_{\theta\phi} - \partial_\phi g_{\theta\theta}\big) = 0.
\end{aligned}$$

The mixed symbol with upper θ is

$$\Gamma^{\theta}{}_{\theta\phi} = \Gamma^{\theta}{}_{\phi\theta} = \frac{1}{2}g^{\theta\theta}\partial_\phi g_{\theta\theta} = 0.$$

For the remaining upper-θ symbol,

$$\begin{aligned}
\Gamma^{\theta}{}_{\phi\phi} &= \frac{1}{2}g^{\theta\theta}\big(2\partial_\phi g_{\phi\theta} - \partial_\theta g_{\phi\phi}\big)\\
&= -\frac{1}{2R^2}\partial_\theta(R^2\sin^2\theta) = -\sin\theta\cos\theta.
\end{aligned}$$

The mixed upper-ϕ symbols are

$$\begin{aligned}
\Gamma^{\phi}{}_{\theta\phi} = \Gamma^{\phi}{}_{\phi\theta} &= \frac{1}{2}g^{\phi\phi}\partial_\theta g_{\phi\phi}\\
&= \frac{1}{2R^2\sin^2\theta}\partial_\theta(R^2\sin^2\theta) = \frac{\cos\theta}{\sin\theta}.
\end{aligned}$$

Finally,

$$\Gamma^{\phi}{}_{\phi\phi} = \frac{1}{2}g^{\phi\phi}\partial_\phi g_{\phi\phi} = 0.$$

Thus the complete nonzero list is

$$\boxed{\Gamma^{\theta}{}_{\phi\phi} = -\sin\theta\cos\theta, \qquad \Gamma^{\phi}{}_{\theta\phi} = \Gamma^{\phi}{}_{\phi\theta} = \cot\theta.} \tag{176}$$

The coordinate singularities of $\cot\theta$ at the poles reflect the failure of the angular coordinate chart there. They do not indicate singularities of the sphere.

40.6 All four components of the vector derivative on the sphere

For a vector field $V = V^\theta \partial_\theta + V^\phi \partial_\phi$, substitution into (166) gives

$$\nabla_\theta V^\theta = \partial_\theta V^\theta, \tag{177}$$

$$\nabla_\theta V^\phi = \partial_\theta V^\phi + \cot\theta\, V^\phi, \tag{178}$$

$$\nabla_\phi V^\theta = \partial_\phi V^\theta - \sin\theta\cos\theta\, V^\phi, \tag{179}$$

$$\nabla_\phi V^\phi = \partial_\phi V^\phi + \cot\theta\, V^\theta. \tag{180}$$

For example, V^θ can vanish identically while $\nabla_\phi V^\theta$ remains nonzero because of the V^ϕ term. Likewise, $V^\phi = 0$ does not imply $\nabla_\phi V^\phi = 0$. Covariant differentiation accounts for how the coordinate basis changes; it is not merely differentiation of an individual component function.

40.7 Why some partial-derivative combinations are tensors

The commutator of two vector fields and the antisymmetrized derivative of a covector already have tensorial expressions involving ordinary derivatives:

$$[u,v]^a = u^b\partial_b v^a - v^b\partial_b u^a, \qquad \partial_{[a}w_{b]} = \frac{1}{2}(\partial_a w_b - \partial_b w_a).$$

The connection explains why these combinations are special.

For the commutator, substitute $\partial_b v^a = \nabla_b v^a - \Gamma^a{}_{bc}v^c$:

$$\begin{aligned}[u,v]^a &= u^b\nabla_b v^a - v^b\nabla_b u^a - \Gamma^a{}_{bc}u^b v^c + \Gamma^a{}_{bc}v^b u^c \\ &= u^b\nabla_b v^a - v^b\nabla_b u^a + (\Gamma^a{}_{cb} - \Gamma^a{}_{bc})u^b v^c.\end{aligned}$$

The last term vanishes for a torsion-free connection. Hence

$$[u,v]^a = u^b\nabla_b v^a - v^b\nabla_b u^a. \tag{181}$$

The right side is explicitly a vector, agreeing with the direct coordinate transformation argument.

Similarly,

$$\begin{aligned}\nabla_{[a}w_{b]} &= \frac{1}{2}\big(\partial_a w_b - \Gamma^c{}_{ab}w_c - \partial_b w_a + \Gamma^c{}_{ba}w_c\big) \\ &= \partial_{[a}w_{b]}.\end{aligned}$$

Thus

$$\nabla_{[a}w_{b]} = \partial_{[a}w_{b]}. \tag{182}$$

The cancellation depends on symmetry of the connection in its lower indices. It explains the tensorial character of the covector curl without asserting that an arbitrary ordinary derivative is a tensor.

41 Parallel transport

41.1 Constant direction along a curve

Let a curve C be specified by $x^a(\lambda)$, with tangent vector

$$u^a = \frac{\mathrm{d}x^a}{\mathrm{d}\lambda}, \qquad u = u^a \partial_a = \frac{\mathrm{d}}{\mathrm{d}\lambda}.$$

A vector field along the curve is *parallel transported* if

$$u^a \nabla_a V^b = 0. \tag{183}$$

In flat space with Cartesian coordinates this becomes $u^a \partial_a V^b = \mathrm{d}V^b/\mathrm{d}\lambda = 0$, so the Cartesian components are constant along the curve. The vector is transported parallel to itself, as in elementary Euclidean geometry.

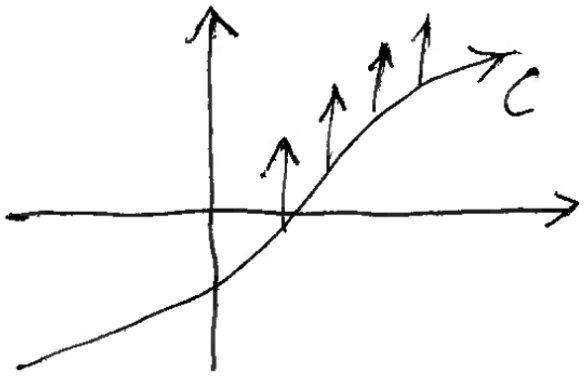

Figure 43: Parallel transport along a curve in Cartesian coordinates keeps the vector components constant, even when the tangent to the curve changes.

The condition $u^a \partial_a V^b = 0$ alone would not be covariant. The same geometric vector may acquire changing polar or spherical components. Equation (183) is a tensor equation, so if it holds in one coordinate system it holds in all of them.

There is also a distinction between transporting a vector along one specified curve and finding a vector field satisfying $\nabla_a V^b = 0$ in every direction throughout an open region. In Euclidean space, constant Cartesian fields have the latter property. In a curved space there need not be any nonzero field with that property; parallel transport along a chosen curve is still well defined.

41.2 Parallel transport as a linear initial-value problem

Expanding the covariant derivative gives

$$0 = u^a \partial_a V^b + \Gamma^b{}_{ac} u^a V^c = \frac{\mathrm{d}V^b}{\mathrm{d}\lambda} + \Gamma^b{}_{ac} \frac{\mathrm{d}x^a}{\mathrm{d}\lambda} V^c.$$

Thus the components satisfy

$$\boxed{\frac{\mathrm{d}V^b}{\mathrm{d}\lambda} = -\Gamma^b{}_{ac}(x(\lambda))\frac{\mathrm{d}x^a}{\mathrm{d}\lambda}V^c.} \tag{184}$$

For a specified curve, the quantities multiplying V^c are known functions of λ. This is a system of coupled linear first-order ordinary differential equations. Specifying the vector at one initial point determines its parallel transport along the curve, as long as the curve and connection remain regular in the charts used.

The geodesic equation will be the special case in which the transported vector is the curve's own tangent:

$$u^a \nabla_a u^b = 0.$$

Parallel transport around a closed curve can return a vector different from the original one. This failure to return will provide a geometric measure of curvature.

41.3 A flat-space circle in polar coordinates

Consider a circle of radius R in the Euclidean plane, parameterized by

$$r = R, \qquad \theta = \lambda, \qquad 0 \le \lambda \le 2\pi.$$

At $\lambda = 0$, choose the unit vector in the positive Cartesian x direction. Its initial polar components are

$$V^r(0) = 1, \qquad V^\theta(0) = 0.$$

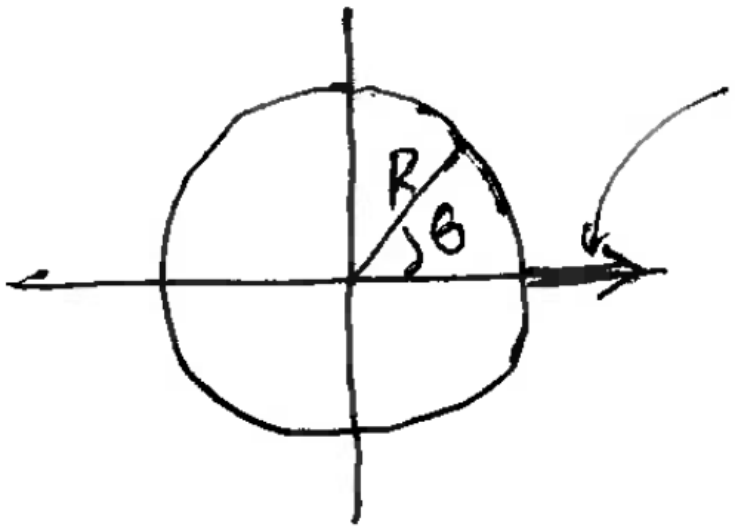

Figure 44: The circular path $r = R$ with its initial vector directed along the positive x axis. The parameter is the polar angle.

The plane metric is $\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2$. From (175), its only nonzero connection coefficients are

$$\Gamma^r{}_{\theta\theta} = -r, \qquad \Gamma^\theta{}_{r\theta} = \Gamma^\theta{}_{\theta r} = \frac{1}{r}.$$

Since $\dot{r} = 0$ and $\dot{\theta} = 1$, the two transport equations become

$$\frac{\mathrm{d}V^r}{\mathrm{d}\lambda} = RV^\theta, \qquad \frac{\mathrm{d}V^\theta}{\mathrm{d}\lambda} = -\frac{1}{R}V^r. \tag{185}$$

Differentiate the first and use the second:

$$\frac{\mathrm{d}^2V^r}{\mathrm{d}\lambda^2} = R\frac{\mathrm{d}V^\theta}{\mathrm{d}\lambda} = -V^r.$$

This is a harmonic-oscillator equation with unit frequency. Its general solution and the corresponding angular component are

$$V^r = A\cos(\lambda + B), \qquad V^\theta = \frac{1}{R}\frac{\mathrm{d}V^r}{\mathrm{d}\lambda} = -\frac{A}{R}\sin(\lambda + B).$$

The initial conditions give $A = 1$ and $B = 0$, so

$$\boxed{V^r(\theta) = \cos\theta, \qquad V^\theta(\theta) = -\frac{\sin\theta}{R}.} \tag{186}$$

The components are not constant, but the vector is parallel transported.

To verify the geometric interpretation, transform back to Cartesian components:

$$V^x = \cos\theta\, V^r - r\sin\theta\, V^\theta, \qquad V^y = \sin\theta\, V^r + r\cos\theta\, V^\theta.$$

Along $r = R$, the solution gives

$$\begin{aligned} V^x &= \cos^2\theta + \sin^2\theta = 1, \\ V^y &= \sin\theta\cos\theta - \cos\theta\sin\theta = 0. \end{aligned}$$

Thus it is exactly the constant Cartesian unit vector expected from flat geometry.

The same polar connection also resolves the earlier constant-y example. For (165), the four components are

$$\begin{aligned} \nabla_r V^r &= 0, \\ \nabla_r V^\theta &= -\frac{C\cos\theta}{r^2} + \frac{1}{r}\frac{C\cos\theta}{r} = 0, \\ \nabla_\theta V^r &= C\cos\theta - r\frac{C\cos\theta}{r} = 0, \\ \nabla_\theta V^\theta &= -\frac{C\sin\theta}{r} + \frac{1}{r}C\sin\theta = 0. \end{aligned}$$

The covariant derivative therefore vanishes in polar coordinates just as it does in Cartesian coordinates.

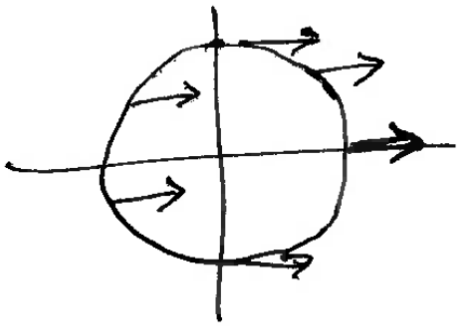

Figure 45: The transported vector retains its Cartesian direction all the way around the circle. Its polar components vary because the polar basis varies.

42 Geodesics and the variational principle

Lecture 27

42.1 Parallel transport of the tangent vector

Geodesics generalize straight lines to curved geometry. A straight line in Cartesian coordinates has a constant tangent vector; the corresponding coordinate-independent condition is that the tangent be parallel transported along the curve itself. If $u^a = \mathrm{d}x^a/\mathrm{d}\lambda$, an affinely parameterized geodesic satisfies

$$\boxed{u^a \nabla_a u^b = 0.} \tag{187}$$

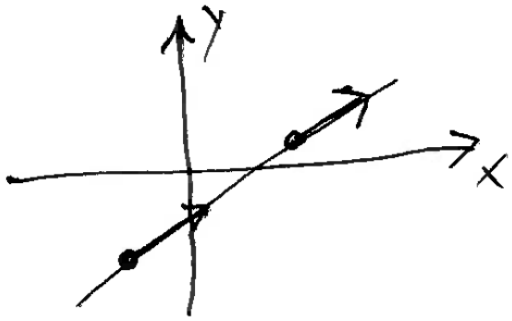

Figure 46: Along an affinely parameterized straight line, the tangent vectors have the same Cartesian components and are parallel to one another.

Expanding the derivative gives

$$u^b \nabla_b u^a = u^b \partial_b u^a + \Gamma^a{}_{bc} u^b u^c = \frac{\mathrm{d}u^a}{\mathrm{d}\lambda} + \Gamma^a{}_{bc} u^b u^c.$$

Since $u^a = \dot{x}^a$, this is equivalently

$$\boxed{\frac{\mathrm{d}^2 x^a}{\mathrm{d}\lambda^2} + \Gamma^a{}_{bc}(x) \frac{\mathrm{d}x^b}{\mathrm{d}\lambda} \frac{\mathrm{d}x^c}{\mathrm{d}\lambda} = 0.} \tag{188}$$

This is a second-order system for the curve. It differs from parallel transport along a given curve: here the curve itself is unknown, the connection is evaluated at its unknown position, and the velocity appears quadratically. The geodesic initial-value problem is consequently nonlinear in general, even though parallel transport along a prescribed curve is linear in the transported vector.

42.2 Stationary length and the local nature of the problem

In a Riemannian space, a sufficiently short geodesic segment is a length-minimizing path between its endpoints. This provides a second description of geodesics and leads to the same differential equation. The variational statement compares a reference curve with nearby curves whose endpoints are fixed.

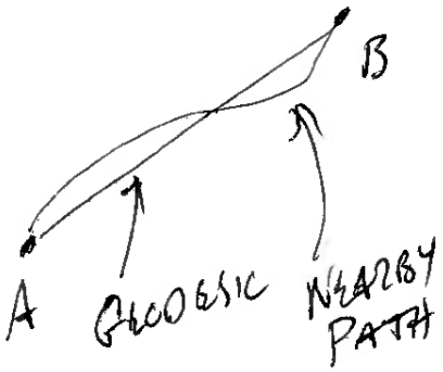

Figure 47: A reference path from A to B and a nearby path with the same endpoints. The first variation tests the change of length under such deformations.

Stationarity and global minimality must be distinguished. On a sphere, both the shorter and the longer arcs of the equator connecting two nonantipodal points satisfy the geodesic equation. The shorter arc has the smaller length; the longer one is still a geodesic, but is not a minimizing path between those endpoints. More generally, the geodesic equation expresses stationarity and is local, whereas whether an entire segment minimizes length depends on the geometry and the extent of the segment.

42.3 The length functional and fixed-endpoint variations

Parameterize a curve from A to B by $\lambda_A \le \lambda \le \lambda_B$. Its Riemannian length is

$$\ell[x] = \int_{\lambda_A}^{\lambda_B} \sqrt{g_{ab}(x)\dot{x}^a\dot{x}^b}\,\mathrm{d}\lambda. \tag{189}$$

The dot denotes differentiation with respect to λ. Let

$$x^a(\lambda) \longmapsto x^a(\lambda) + \delta x^a(\lambda), \qquad \delta x^a(\lambda_A) = \delta x^a(\lambda_B) = 0.$$

The endpoints remain fixed, but the path between them can vary arbitrarily.

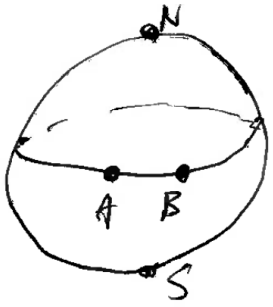

Figure 48: Both equatorial arcs are geodesic segments. For distinct, nonantipodal endpoints A and B, only the short arc minimizes distance.

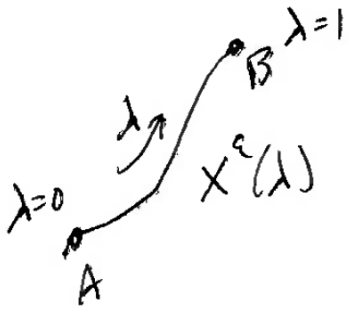

Figure 49: A parameterized curve with endpoints fixed at A and B.

The length of the varied curve has an expansion

$$\ell[x+\delta x] = \ell[x] + \delta\ell + O((\delta x)^2).$$

A necessary condition for an extremum is $\delta\ell = 0$ for every admissible variation. This is the functional analogue of the condition $f'(x_0) = 0$ at an ordinary stationary point.

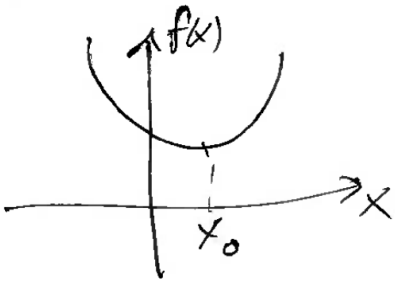

Figure 50: The ordinary calculus analogy: the first derivative vanishes at a stationary point of a function.

Define

$$q(\lambda) = g_{ab}(x)\dot{x}^a\dot{x}^b.$$

Using

$$\delta g_{ab} = (\partial_c g_{ab})\delta x^c, \qquad \delta\dot{x}^a = \frac{\mathrm{d}}{\mathrm{d}\lambda}\delta x^a,$$

the variation of q is

$$\delta q = (\partial_c g_{ab})\dot{x}^a\dot{x}^b\delta x^c + 2g_{ab}\dot{x}^b\frac{\mathrm{d}(\delta x^a)}{\mathrm{d}\lambda}.$$

The factor two in the second term uses symmetry of the metric. Expanding the square root to first order gives

$$\delta\ell = \frac{1}{2}\int_{\lambda_A}^{\lambda_B}\frac{\mathrm{d}\lambda}{\sqrt{q}}\left[(\partial_c g_{ab})\dot{x}^a\dot{x}^b\delta x^c + 2g_{ab}\dot{x}^b\frac{\mathrm{d}(\delta x^a)}{\mathrm{d}\lambda}\right]. \tag{190}$$

42.4 Integration by parts and the geodesic equation

Choose a constant-speed parameter on the reference curve, so that $q = v_0^2$ is constant. Arclength is the special case $v_0 = 1$. This choice is made for the reference curve when evaluating the first variation; the varied paths still have arbitrary fixed-endpoint variations.

The term involving $\mathrm{d}(\delta x^a)/\mathrm{d}\lambda$ can be integrated by parts:

$$\int_{\lambda_A}^{\lambda_B} g_{ab}\dot{x}^b \frac{\mathrm{d}(\delta x^a)}{\mathrm{d}\lambda}\,\mathrm{d}\lambda = \left[g_{ab}\dot{x}^b\delta x^a\right]_{\lambda_A}^{\lambda_B} - \int_{\lambda_A}^{\lambda_B} \frac{\mathrm{d}}{\mathrm{d}\lambda}(g_{ab}\dot{x}^b)\delta x^a\,\mathrm{d}\lambda.$$

The boundary term vanishes because δx^a vanishes at both endpoints. Furthermore,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}(g_{ab}\dot{x}^b) = (\partial_c g_{ab})\dot{x}^c\dot{x}^b + g_{ab}\ddot{x}^b.$$

Relabeling the index on the first variation term so that every term multiplies δx^a, we obtain

$$\delta\ell = \frac{1}{2v_0}\int_{\lambda_A}^{\lambda_B} \left[(\partial_a g_{cb})\dot{x}^c\dot{x}^b - 2(\partial_c g_{ab})\dot{x}^c\dot{x}^b - 2g_{ab}\ddot{x}^b\right]\delta x^a\,\mathrm{d}\lambda. \tag{191}$$

Because $\dot{x}^b\dot{x}^c$ is symmetric in b, c,

$$2(\partial_c g_{ab})\dot{x}^c\dot{x}^b = (\partial_c g_{ab} + \partial_b g_{ac})\dot{x}^b\dot{x}^c.$$

Consequently,

$$\delta\ell = -\frac{1}{v_0}\int_{\lambda_A}^{\lambda_B} \left[g_{ab}\ddot{x}^b + \frac{1}{2}(\partial_c g_{ab} + \partial_b g_{ac} - \partial_a g_{bc})\dot{x}^b\dot{x}^c\right]\delta x^a\,\mathrm{d}\lambda. \tag{192}$$

The variations are arbitrary in the interior, so the expression in square brackets must vanish. Multiplying by g^{da} yields

$$\ddot{x}^d + \frac{1}{2}g^{da}(\partial_b g_{ca} + \partial_c g_{ba} - \partial_a g_{bc})\dot{x}^b\dot{x}^c = 0.$$

Recognizing the Christoffel symbol gives

$$\ddot{x}^d + \Gamma^d{}_{bc}\dot{x}^b\dot{x}^c = 0.$$

Thus stationary length, with a constant-speed parameter, produces exactly the parallel-tangent equation (187). In compact form the first variation is

$$\delta\ell = -\frac{1}{v_0}\int g_{ab}(u^c\nabla_c u^b)\,\delta x^a\,\mathrm{d}\lambda.$$

42.5 Affine parameters and constant tangent norm

The homogeneous geodesic equations (187) and (188) use an *affine parameter*. Metric compatibility gives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}(g_{ab}u^a u^b) &= u^c\nabla_c(g_{ab}u^a u^b)\\ &= 2g_{ab}u^a(u^c\nabla_c u^b) = 0.\end{aligned}$$

Thus the squared norm of the tangent is constant along an affinely parameterized geodesic. In a Riemannian space it can be normalized to one by using arclength. In spacetime it can be normalized to -1 for a timelike geodesic parameterized by proper time, or to $+1$ for a spacelike geodesic parameterized by proper length, in units $c = 1$. For a null geodesic the norm is zero and cannot determine the scale of the affine parameter.

43 Causal character and geodesic parameters

Lecture 28

43.1 Stationarity, minima, and maxima

The condition $\delta\ell = 0$ determines a stationary curve. Distinguishing a minimum, a maximum, and a saddle requires the second variation. In Riemannian geometry, sufficiently short geodesic segments minimize length; longer segments need not remain minimizing.

For a timelike curve in a spacetime with signature $(-,+,+,+)$, the proper time is

$$\Delta\tau = \int_{\lambda_A}^{\lambda_B} \sqrt{-g_{ab}\dot{x}^a\dot{x}^b}\,\mathrm{d}\lambda \qquad (c=1). \tag{193}$$

The variational calculation proceeds as above with the sign inside the square root changed. In a proper-time parameter it again gives the geodesic equation. A sufficiently short timelike geodesic maximizes proper time among nearby timelike curves with the same endpoints.

The flat-spacetime example makes the sign of the extremum transparent. For $A = (0,0)$ and $B = (T,0)$ in an inertial frame,

$$\Delta\tau = \int_0^T \sqrt{1 - v(t)^2}\,\mathrm{d}t \le T.$$

The stationary inertial worldline $x(t) = 0$ attains T. Introducing nearly lightlike segments can make the elapsed proper time much smaller: a zigzag path with speed magnitude approaching one spends very little proper time between the same events.

For spacelike curves in a Lorentzian spacetime, stationarity of proper length does not generally imply a minimum under unrestricted spacetime variations. For example, in Minkowski space near the spacelike segment $t = 0$, $y = 0$, parameterized by x, let $t = \epsilon f(x)$ and $y = \epsilon h(x)$ with fixed endpoints. Then

$$\ell = \int \sqrt{1 + \epsilon^2 h'(x)^2 - \epsilon^2 f'(x)^2}\,\mathrm{d}x = \ell_0 + \frac{\epsilon^2}{2}\int [h'(x)^2 - f'(x)^2]\,\mathrm{d}x + O(\epsilon^4).$$

Figure 51: An almost lightlike zigzag between timelike-separated events can have much less proper time than the timelike geodesic joining them.

Spatial and temporal deformations change the length with opposite signs. The length-minimizing interpretation applies directly to the Riemannian spatial problem; the Lorentzian geodesic equation expresses stationarity for nonnull curves.

A lightlike curve has $g_{ab}u^a u^b = 0$ and zero proper-time interval. Its geodesic equation is defined using parallel transport of its tangent,

$$u^a \nabla_a u^b = 0$$

for an affine parameter. Proper time cannot serve as that parameter, since it is constant along a null curve.

43.2 Straight lines in flat space and spacetime

In Cartesian coordinates on the Euclidean plane, all connection coefficients vanish. Equation (188) reduces to

$$\frac{\mathrm{d}^2 x}{\mathrm{d}\lambda^2} = 0, \qquad \frac{\mathrm{d}^2 y}{\mathrm{d}\lambda^2} = 0.$$

Its solution is

$$x = a\lambda + b, \qquad y = c\lambda + d, \tag{194}$$

the parameterized equation of a straight line. In inertial coordinates in Minkowski spacetime, the same reasoning gives $x^\mu(\lambda) = A^\mu \lambda + B^\mu$. For a timelike geodesic this is constant-velocity motion.

43.3 Great circles on the sphere

A great circle is the intersection of a sphere with a plane through its centre. By rotating the spherical coordinate system, any great circle can be made to pass through the north and south poles, so that a segment away from the coordinate singularities has constant ϕ.

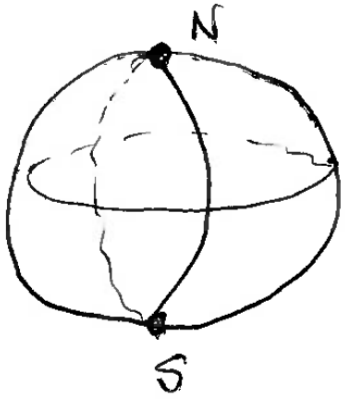

Figure 52: A great circle through the north and south poles. After rotating the angular coordinates, any great circle can be described locally in this form.

For a unit sphere, consider

$$\theta = \lambda, \qquad \phi = \phi_0.$$

Then $\dot{\theta} = 1$, $\dot{\phi} = 0$, and both second derivatives vanish. Using (176), the two geodesic equations are

$$\ddot{\theta} - \sin\theta\cos\theta\,\dot{\phi}^2 = 0, \tag{195}$$

$$\ddot{\phi} + 2\cot\theta\,\dot{\theta}\dot{\phi} = 0. \tag{196}$$

The proposed meridian satisfies both equations. Rotational invariance then establishes that great circles, including the equator, are geodesics. For the equator, this can also be checked directly by setting $\theta = \pi/2$ and ϕ linear in the affine parameter. The angular coordinate chart must be changed when crossing a pole; the geometric great circle remains smooth.

43.4 Length does not depend on the choice of parameter

A curve is a geometric path, whereas its parameter specifies how that path is traversed. Consider a smooth, orientation-preserving reparameterization $\lambda = \lambda(\sigma)$ with $\mathrm{d}\lambda/\mathrm{d}\sigma > 0$. The chain rule gives

$$\frac{\mathrm{d}x^a}{\mathrm{d}\sigma} = \frac{\mathrm{d}x^a}{\mathrm{d}\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}\sigma}.$$

The length written in the new parameter is

$$\begin{aligned}\ell &= \int \sqrt{g_{ab}\frac{\mathrm{d}x^a}{\mathrm{d}\sigma}\frac{\mathrm{d}x^b}{\mathrm{d}\sigma}}\,\mathrm{d}\sigma \\ &= \int \frac{\mathrm{d}\lambda}{\mathrm{d}\sigma}\sqrt{g_{ab}\frac{\mathrm{d}x^a}{\mathrm{d}\lambda}\frac{\mathrm{d}x^b}{\mathrm{d}\lambda}}\,\mathrm{d}\sigma \\ &= \int \sqrt{g_{ab}\dot{x}^a\dot{x}^b}\,\mathrm{d}\lambda.\end{aligned}$$

Thus the length is unchanged. For an orientation reversal, the absolute value in the square root and the reversal of integration limits yield the same positive length. The stationary path is consequently independent of parameterization, even though the homogeneous differential equation uses a special class of parameters.

43.5 Deriving the nonaffine geodesic equation

Let λ be an affine parameter, set $u^a = \mathrm{d}x^a/\mathrm{d}\lambda$, and use a new parameter σ. Define

$$w^a = \frac{\mathrm{d}x^a}{\mathrm{d}\sigma} = \alpha u^a, \qquad \alpha = \frac{\mathrm{d}\lambda}{\mathrm{d}\sigma}.$$

Then

$$\begin{aligned}w^b\nabla_b w^a &= \alpha u^b\nabla_b(\alpha u^a) \\ &= \alpha(u^b\partial_b\alpha)u^a + \alpha^2 u^b\nabla_b u^a \\ &= \frac{\mathrm{d}\alpha}{\mathrm{d}\sigma}u^a = \frac{1}{\alpha}\frac{\mathrm{d}\alpha}{\mathrm{d}\sigma}w^a.\end{aligned}$$

The second term vanishes because λ is affine. Therefore the geodesic equation in a general parameter is

$$\boxed{w^b\nabla_b w^a = f(\sigma)w^a, \qquad f(\sigma) = \frac{\mathrm{d}^2\lambda/\mathrm{d}\sigma^2}{\mathrm{d}\lambda/\mathrm{d}\sigma}.} \tag{197}$$

Writing the same coefficient in terms of $\sigma(\lambda)$ gives

$$f(\sigma) = -\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\lambda}\right)^{-2}\frac{\mathrm{d}^2\sigma}{\mathrm{d}\lambda^2}. \tag{198}$$

The coordinate form is

$$\boxed{\frac{\mathrm{d}^2x^a}{\mathrm{d}\sigma^2} + \Gamma^a{}_{bc}\frac{\mathrm{d}x^b}{\mathrm{d}\sigma}\frac{\mathrm{d}x^c}{\mathrm{d}\sigma} = f(\sigma)\frac{\mathrm{d}x^a}{\mathrm{d}\sigma}.} \tag{199}$$

The additional acceleration is parallel to the tangent. It changes the speed with which the curve is parameterized without bending its direction.

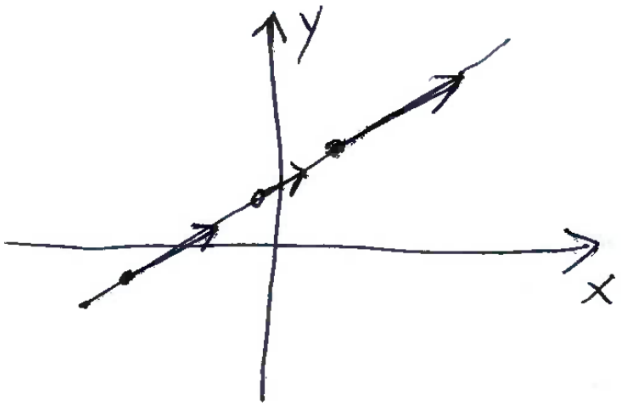

Figure 53: A nonaffine parameter changes the magnitudes of the tangent vectors along a straight line while preserving their direction.

Affine parameters have $f = 0$. If both λ and σ are affine, (198) implies $\mathrm{d}^2\sigma/\mathrm{d}\lambda^2 = 0$, so

$$\sigma = A\lambda + B, \qquad A \neq 0.$$

Thus affine parameters are determined only up to a nonzero constant rescaling and a translation.

For a nonnull curve, the coefficient can also be read from its speed. Metric compatibility and (197) give

$$\frac{\mathrm{d}}{\mathrm{d}\sigma}(g_{ab}w^a w^b) = 2f(\sigma)\, g_{ab}w^a w^b.$$

Writing $v = \sqrt{|g_{ab}w^a w^b|}$ yields $f = \mathrm{d}(\log v)/\mathrm{d}\sigma$. The tangent norm is constant for affine parameterization. For null curves the norm vanishes for every parameter, so this norm condition alone cannot distinguish affine from nonaffine parameters.

44 Parallel transport around closed curves

44.1 Curvature as failure to return

Parallel transport around a closed curve in the Euclidean plane brings a vector back unchanged, as the circle calculation in Section 41.3 demonstrated. In a curved space, the final vector can differ from its initial value even though the transport equation has been obeyed at every point. The comparison is meaningful because the initial and final vectors belong to the same tangent space.

The resulting linear map on the tangent space is called *holonomy*. For a sufficiently small closed loop, its departure from the identity measures curvature and will lead to the Riemann tensor. Flat geometry has trivial transport around loops contained in an ordinary simply connected flat coordinate neighborhood. Local curvature is detected by shrinking loops, rather than by confusing coordinate variation with a geometric change of the transported vector.

44.2 Parallel transport preserves lengths and inner products

Suppose V^a is parallel transported along a curve with tangent u^a. Then

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}(V^b V_b) &= u^a \nabla_a (g_{bc} V^b V^c) \\ &= (u^a \nabla_a g_{bc}) V^b V^c + g_{bc}(u^a \nabla_a V^b) V^c + g_{bc} V^b (u^a \nabla_a V^c) = 0. \end{aligned} \tag{200}$$

The first term vanishes by metric compatibility, and the other two by parallel transport. More generally, if V and W are both transported, the same product-rule calculation gives

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}(g_{ab} V^a W^b) = 0.$$

Thus holonomy preserves the tangent-space metric. On a Riemannian surface it acts locally as a rotation. In a Lorentzian spacetime it is a Lorentz isometry, so its action can include boosts as well as spatial rotations. In either case, the change around a loop preserves the appropriate inner products.

44.3 Transport around a circle of constant latitude

Consider a circle on the unit two-sphere at fixed polar angle $\theta = \theta_0$, with $0 < \theta_0 < \pi$. Parameterize it by

$$\phi = \lambda, \qquad 0 \le \lambda \le 2\pi.$$

The tangent has components $u^\theta = 0$, $u^\phi = 1$.

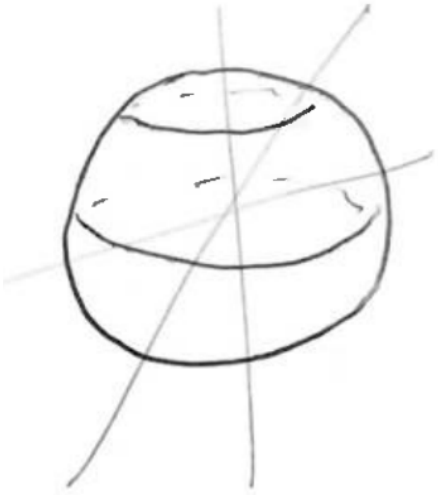

Figure 54: A closed constant-latitude circle on the two-sphere. The polar angle θ_0 is fixed while the azimuthal angle increases through 2π.

Using (176), the parallel-transport equations are

$$\frac{\mathrm{d}V^\theta}{\mathrm{d}\lambda} = \sin\theta_0 \cos\theta_0 \, V^\phi, \tag{201}$$

$$\frac{\mathrm{d}V^\phi}{\mathrm{d}\lambda} = -\frac{\cos\theta_0}{\sin\theta_0} \, V^\theta. \tag{202}$$

These equations also follow directly from the components of $\nabla_\phi V^a$ in (180). Differentiating once more gives

$$\frac{\mathrm{d}^2V^\theta}{\mathrm{d}\lambda^2} = -\cos^2\theta_0 \, V^\theta, \qquad \frac{\mathrm{d}^2V^\phi}{\mathrm{d}\lambda^2} = -\cos^2\theta_0 \, V^\phi. \tag{203}$$

Thus each component satisfies a harmonic-oscillator equation. For convenient sign bookkeeping, set $\omega = \cos\theta_0$, so the oscillator frequency magnitude is $|\omega|$. Away from the equator one may write

$$V^\theta = A\cos(\omega\lambda + B), \qquad V^\phi = -\frac{A}{\sin\theta_0}\sin(\omega\lambda + B).$$

The second relation follows by substituting into the first equation of (202). It remains well behaved at the equator by continuity, even though solving for V^ϕ by division by $\cos\theta_0$ would not be appropriate there.

Choose

$$V^\theta(0) = 1, \qquad V^\phi(0) = 0.$$

Then $A = 1$, $B = 0$, and

$$\boxed{V^\theta(\lambda) = \cos(\lambda\cos\theta_0), \qquad V^\phi(\lambda) = -\frac{\sin(\lambda\cos\theta_0)}{\sin\theta_0}.} \tag{204}$$

Its norm is indeed constant:

$$(V^\theta)^2 + \sin^2\theta_0 (V^\phi)^2 = 1.$$

After a complete loop,

$$V^\theta(2\pi) = \cos(2\pi\cos\theta_0), \qquad V^\phi(2\pi) = -\frac{\sin(2\pi\cos\theta_0)}{\sin\theta_0}. \tag{205}$$

For a generic latitude this differs from the initial vector. At the equator, $\cos\theta_0 = 0$, the two transport equations say that both components are constant, and the vector returns unchanged. The pole itself is outside this angular chart and the constant-latitude circle degenerates there.

44.4 The small-loop limit near the pole

To describe a genuine rotation angle, use components in the orthonormal basis along the latitude:

$$V^{\hat\theta} = V^\theta, \qquad V^{\hat\phi} = \sin\theta_0 \, V^\phi.$$

Equation (204) becomes

$$\begin{pmatrix} V^{\hat{\theta}} \\ V^{\hat{\phi}} \end{pmatrix} = \begin{pmatrix} \cos(\lambda \cos\theta_0) \\ -\sin(\lambda \cos\theta_0) \end{pmatrix}.$$

Because the basis at $\phi = 2\pi$ is the same as the basis at $\phi = 0$, the rotation after one circuit is $-2\pi \cos\theta_0$, understood modulo 2π. A representative tending continuously to zero for a small north-polar loop is

$$\Delta\alpha = 2\pi(1 - \cos\theta_0). \tag{206}$$

For $\theta_0 \ll 1$, retaining the first correction to the cosine gives

$$\cos\theta_0 = 1 - \frac{\theta_0^2}{2} + O(\theta_0^4), \qquad \Delta\alpha = \pi\theta_0^2 + O(\theta_0^4).$$

The vector almost returns to itself, as expected because a sufficiently small region of the sphere looks approximately flat. Replacing $\cos\theta_0$ by exactly one would discard the leading curvature effect; the quadratic term is essential.

The area of the north-polar cap on the unit sphere is

$$\mathcal{A} = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\theta_0} \sin\theta \, \mathrm{d}\theta = 2\pi(1 - \cos\theta_0),$$

so the holonomy angle equals the enclosed cap area for this sphere. For radius R, the same angular connection gives $\Delta\alpha = \mathcal{A}/R^2$. The change is proportional to the area of a small loop, illustrating how curvature is extracted from infinitesimal parallel transport.

45 Curvature as a local tensor

Lectures 29–30

45.1 Infinitesimal loops and the meaning of the indices

Parallel transport around a finite closed curve can change a vector even though the curve returns to its starting point. To extract a *local* measure of this effect, shrink the loop to an infinitesimal parallelogram at a point p. Two linearly independent tangent vectors u and w specify its plane. The change in a transported vector is linear in that vector and proportional, to leading order, to the oriented area of the loop. For sides ϵu and δw, the leading curvature term has the form

$$\Delta v^a \propto \epsilon\delta \, u^c w^d R_{cd}{}^a{}_b v^b,$$

with the overall sign fixed by the direction in which the loop is traversed. Reversing the orientation reverses this leading change.

There are therefore two different pairs of indices in curvature: one pair describes the two directions spanning the infinitesimal loop; the other describes the linear transformation of the vector being transported. For a Riemannian metric, this infinitesimal transformation preserves inner products and acts as a rotation. For a Lorentzian metric it is an infinitesimal Lorentz transformation, which can include a boost.

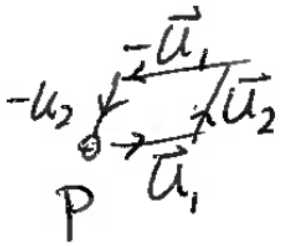

Figure 55: An infinitesimal loop spanned by two tangent directions.

45.2 Definition through a commutator

Throughout this discussion ∇ is the torsion-free, metric-compatible Levi–Civita connection. In a coordinate basis, define the Riemann tensor by its action on an arbitrary covector field v_c:

$$(\nabla_a \nabla_b - \nabla_b \nabla_a) v_c = R_{abc}{}^{d} v_d. \tag{207}$$

The order of the indices in this formula fixes the sign convention. In particular, the first two slots are the derivative directions. The expression on the left initially appears to contain second and first derivatives of v_c. These derivatives cancel, leaving a linear function of the value of v_c at the same point. That cancellation is what makes the right-hand side a tensorial local measure of curvature.

Remark 45.1 (Index placement). Define the fully covariant tensor by

$$R_{abcd} = g_{de} R_{abc}{}^{e}.$$

Raising the third slot instead gives $R_{ab}{}^{c}{}_{d} = g^{ce} R_{abed}$. The upper index occupies a different slot in these two expressions, so they must not be interchanged without accounting for the curvature symmetries.

45.3 Deriving the connection formula

The covariant derivative of a covector is

$$\nabla_b v_c = \partial_b v_c - \Gamma^{d}{}_{bc} v_d.$$

Since $\nabla_b v_c$ has two covariant indices, differentiating again requires a connection term for *each* index:

$$\begin{aligned} \nabla_a \nabla_b v_c &= \partial_a (\nabla_b v_c) - \Gamma^{d}{}_{ab} \nabla_d v_c - \Gamma^{d}{}_{ac} \nabla_b v_d \\ &= \partial_a \partial_b v_c - (\partial_a \Gamma^{d}{}_{bc}) v_d - \Gamma^{d}{}_{bc} \partial_a v_d \\ &\quad - \Gamma^{d}{}_{ab} \partial_d v_c + \Gamma^{d}{}_{ab} \Gamma^{e}{}_{dc} v_e \\ &\quad - \Gamma^{d}{}_{ac} \partial_b v_d + \Gamma^{d}{}_{ac} \Gamma^{e}{}_{bd} v_e. \end{aligned} \tag{208}$$

Subtract the expression with a and b exchanged. The second partial derivatives cancel because partial derivatives commute. The terms containing $\Gamma^{d}{}_{ab} \partial_d v_c$ cancel because

$\Gamma^d{}_{ab} = \Gamma^d{}_{ba}$. The two remaining pairs of first partial derivatives also cancel. The product terms containing $\Gamma^d{}_{ab}$ cancel for the same torsion-free reason. Relabeling the dummy indices in the remaining products gives

$$[\nabla_a, \nabla_b]v_c = \big(-\partial_a \Gamma^d{}_{bc} + \partial_b \Gamma^d{}_{ac} + \Gamma^e{}_{ac}\Gamma^d{}_{be} - \Gamma^e{}_{bc}\Gamma^d{}_{ae}\big) v_d. \tag{209}$$

Consequently,

$$\boxed{R_{abc}{}^d = -\partial_a \Gamma^d{}_{bc} + \partial_b \Gamma^d{}_{ac} + \Gamma^e{}_{ac}\Gamma^d{}_{be} - \Gamma^e{}_{bc}\Gamma^d{}_{ae}.} \tag{210}$$

Although the individual connection coefficients are not tensor components, this combination is tensorial by (207).

Recall the Levi–Civita formula

$$\Gamma^c{}_{ab} = \frac{1}{2} g^{cd}\big(\partial_a g_{bd} + \partial_b g_{ad} - \partial_d g_{ab}\big).$$

Thus curvature contains second derivatives of the metric and terms quadratic in its first derivatives. A connection can be nonzero merely because the coordinates are curved; the Riemann tensor identifies the coordinate-independent obstruction to flatness.

46 The curvature commutator on tensors

46.1 Scalars and the torsion-free condition

For a scalar f, its first covariant derivative is just $\nabla_a f = \partial_a f$. Its second derivative, however, is the covariant derivative of the covector $\partial_b f$:

$$\nabla_a \nabla_b f = \partial_a \partial_b f - \Gamma^c{}_{ab} \partial_c f.$$

Therefore

$$\begin{aligned} [\nabla_a, \nabla_b] f &= \partial_a \partial_b f - \partial_b \partial_a f - (\Gamma^c{}_{ab} - \Gamma^c{}_{ba}) \partial_c f \\ &= 0. \end{aligned} \tag{211}$$

Commutation on scalars is compatible with nonzero curvature: a scalar has no vector or covector index for curvature to act on.

46.2 Vectors: two ways to obtain the answer

First use metric compatibility to raise the covector index:

$$\begin{aligned} [\nabla_a, \nabla_b] v^c &= g^{cd} [\nabla_a, \nabla_b] v_d \\ &= g^{cd} R_{abd}{}^e v_e = R_{ab}{}^c{}_d v^d. \end{aligned} \tag{212}$$

A second derivation begins with an arbitrary scalar contraction $f = u_c v^c$. Expanding both second derivatives by the product rule, the cross terms cancel pairwise, leaving

$$0 = [\nabla_a, \nabla_b](u_c v^c) = u_c[\nabla_a, \nabla_b]v^c + ([\nabla_a, \nabla_b]u_c)v^c. \tag{213}$$

Substitute the defining covector formula and rename dummy indices:

$$0 = u_d[\nabla_a, \nabla_b]v^d + u_d R_{abc}{}^d v^c.$$

Because u_d is arbitrary,

$$\boxed{[\nabla_a, \nabla_b]v^c = -R_{abd}{}^c v^d.} \tag{214}$$

Comparing this result with (212) yields

$$R_{ab}{}^c{}_d = -R_{abd}{}^c.$$

Lowering the upper index with g_{ec} proves

$$R_{abed} = -R_{abde}. \tag{215}$$

Thus the opposite signs in the vector and covector commutators are exactly what is required for contractions and the metric to be preserved.

Remark 46.1 (The convention with an upper index first). If the curvature is instead denoted by $\mathcal{R}^c{}_{dab}$ through $[\nabla_a, \nabla_b]v^c = \mathcal{R}^c{}_{dab} v^d$, then

$$\mathcal{R}^c{}_{dab} = -R_{abd}{}^c, \qquad \mathcal{R}^c{}_{dab} = \partial_a \Gamma^c{}_{bd} - \partial_b \Gamma^c{}_{ad} + \Gamma^c{}_{ae}\Gamma^e{}_{bd} - \Gamma^c{}_{be}\Gamma^e{}_{ad}.$$

Both notations describe the same connection and the same curvature; the defining commutator determines how to compare the slots.

46.3 One curvature term for each index

For a tensor of type (k, l), the commutator acts on every slot. Using the convention of (207),

$$\begin{aligned}[\nabla_a, \nabla_b]T^{c_1\cdots c_k}{}_{d_1\cdots d_l} = &-\sum_{i=1}^{k} R_{abe}{}^{c_i} T^{c_1\cdots c_{i-1} e c_{i+1}\cdots c_k}{}_{d_1\cdots d_l} \\ &+ \sum_{j=1}^{l} R_{abd_j}{}^e T^{c_1\cdots c_k}{}_{d_1\cdots d_{j-1} e d_{j+1}\cdots d_l}.\end{aligned} \tag{216}$$

This follows by contracting the tensor with arbitrary test vectors and covectors to form a scalar, applying (211), and using the product rule. Each upper index contributes a vector term, and each lower index contributes a covector term.

Example 46.2 (A mixed tensor)**.** For $T^c{}_d$ the formula becomes

$$[\nabla_a, \nabla_b]T^c{}_d = -R_{abe}{}^c T^e{}_d + R_{abd}{}^e T^c{}_e.$$

Equivalently, the first term is $R_{ab}{}^c{}_e T^e{}_d$.

Example 46.3 (The metric)**.** Because $\nabla_a g_{cd} = 0$,

$$\begin{aligned} 0 &= [\nabla_a, \nabla_b] g_{cd} \\ &= R_{abc}{}^{e} g_{ed} + R_{abd}{}^{e} g_{ce} \\ &= R_{abcd} + R_{abdc}. \end{aligned}$$

This supplies a second proof of antisymmetry in the last pair of indices.

47 Riemann symmetries and identities

For the Levi–Civita connection, the five principal identities are

$$R_{bacd} = -R_{abcd}, \qquad \text{first-pair antisymmetry,} \tag{217}$$

$$R_{abdc} = -R_{abcd}, \qquad \text{last-pair antisymmetry,} \tag{218}$$

$$R_{abcd} = R_{cdab}, \qquad \text{pair exchange,} \tag{219}$$

$$R_{[abc]}{}^{d} = 0, \qquad \text{algebraic Bianchi identity,} \tag{220}$$

$$\nabla_{[a} R_{bc]d}{}^{e} = 0, \qquad \text{differential Bianchi identity.} \tag{221}$$

The first identity follows immediately by interchanging the two covariant derivatives in the definition. The second has already been derived both from scalar contractions and from metric compatibility.

47.1 Pair exchange and the algebraic identity

At any chosen point p, take normal coordinates, so that $\Gamma^{a}{}_{bc}(p) = 0$ and $\partial_a g_{bc}(p) = 0$. The derivatives of Γ need not vanish. Writing $g_{ab,cd}$ for $\partial_c \partial_d g_{ab}$, the connection formula reduces at p to

$$R_{abcd} = \frac{1}{2}\big(g_{bc,ad} + g_{ad,bc} - g_{bd,ac} - g_{ac,bd}\big). \tag{222}$$

This follows by differentiating the three terms of the Levi–Civita formula; the terms $g_{cd,ab}$ and $g_{cd,ba}$ cancel. Using $g_{ab} = g_{ba}$ and commuting the partial derivatives shows directly that exchanging (a, b) with (c, d) leaves the expression unchanged. Since the result is a tensor identity at an arbitrary point, pair exchange holds in every coordinate system.

The algebraic Bianchi identity is equivalently the cyclic identity

$$R_{abcd} + R_{bcad} + R_{cabd} = 0. \tag{223}$$

To verify it explicitly, twice its left-hand side in normal coordinates is

$$\begin{aligned} &(g_{bc,ad} + g_{ad,bc} - g_{bd,ac} - g_{ac,bd}) \\ &\quad + (g_{ca,bd} + g_{bd,ca} - g_{cd,ba} - g_{ba,cd}) \\ &\quad + (g_{ab,cd} + g_{cd,ab} - g_{ad,cb} - g_{cb,ad}), \end{aligned}$$

and every term cancels with one of opposite sign. Again, tensoriality makes the identity independent of the special coordinates used to prove it.

47.2 The differential Bianchi identity

The antisymmetrized derivative in (221) can be written as

$$\nabla_a R_{bcd}{}^{e} + \nabla_b R_{cad}{}^{e} + \nabla_c R_{abd}{}^{e} = 0. \tag{224}$$

At the origin of normal coordinates the covariant derivatives in this expression can be replaced by partial derivatives. The connection-product terms in (210) have vanishing first derivative there because each differentiated product still contains one undifferentiated Γ. The cyclic sum is therefore

$$\begin{aligned} &- \partial_a\partial_b\Gamma^{e}{}_{cd} + \partial_a\partial_c\Gamma^{e}{}_{bd} \\ &- \partial_b\partial_c\Gamma^{e}{}_{ad} + \partial_b\partial_a\Gamma^{e}{}_{cd} \\ &- \partial_c\partial_a\Gamma^{e}{}_{bd} + \partial_c\partial_b\Gamma^{e}{}_{ad} = 0. \end{aligned}$$

Commutation of the partial derivatives gives the required cancellation. The identity holds at every point and is especially important in relating curvature to conservation laws in general relativity.

48 Ricci curvature and scalar curvature

48.1 The Ricci contraction

The traces of the Riemann tensor isolate geometrically important lower-rank tensors. In the index convention used here, the Ricci tensor is

$$\boxed{R_{ac} = R_{abc}{}^{b}.} \tag{225}$$

Equivalently, $R_{ab} = R_{acb}{}^{c}$. Inserting (210),

$$R_{ab} = \partial_c\Gamma^{c}{}_{ab} - \partial_a\Gamma^{c}{}_{bc} + \Gamma^{c}{}_{ab}\Gamma^{d}{}_{cd} - \Gamma^{d}{}_{ac}\Gamma^{c}{}_{bd}. \tag{226}$$

The Ricci tensor is symmetric. A direct index proof uses pair exchange:

$$\begin{aligned} R_{ab} &= R_{acb}{}^{c} = g^{cd}R_{acbd} = g^{cd}R_{bdac} \\ &= R_{bda}{}^{d} = R_{ba}. \end{aligned} \tag{227}$$

48.2 The scalar contraction

Taking the trace once more defines scalar curvature:

$$\boxed{R = g^{ab}R_{ab} = R_{a}{}^{a}.} \tag{228}$$

Unlike a coordinate component such as $R_{\theta\theta}$, the value of R at a point is independent of the coordinate system. The Riemann tensor contains the full curvature information; Ricci and scalar curvature are particular contractions of that information.

49 Curvature of a two-sphere

Consider a round two-sphere of radius L with line element

$$\mathrm{d}\ell^2 = L^2\big(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\big). \tag{229}$$

Its metric and inverse are

$$(g_{ab}) = \begin{pmatrix} L^2 & 0 \\ 0 & L^2\sin^2\theta \end{pmatrix}, \qquad (g^{ab}) = \begin{pmatrix} L^{-2} & 0 \\ 0 & (L^2\sin^2\theta)^{-1} \end{pmatrix}.$$

The nonzero Christoffel symbols are

$$\Gamma^{\theta}{}_{\phi\phi} = -\sin\theta\cos\theta, \qquad \Gamma^{\phi}{}_{\theta\phi} = \Gamma^{\phi}{}_{\phi\theta} = \cot\theta. \tag{230}$$

Every factor of L^2 introduced by differentiating g_{ab} cancels against L^{-2} from the inverse metric. This explains why the Christoffel symbols do not depend on the constant radius L.

49.1 The independent Riemann component

Antisymmetry in each pair leaves only one independent fully covariant component in two dimensions:

$$R_{\theta\phi\theta\phi} = -R_{\phi\theta\theta\phi} = -R_{\theta\phi\phi\theta} = R_{\phi\theta\phi\theta}. \tag{231}$$

All components with equal indices in either antisymmetric pair vanish. It is convenient to calculate $R_{\theta\phi\theta}{}^{\phi}$ first. From (210),

$$\begin{aligned} R_{\theta\phi\theta}{}^{\phi} &= -\partial_\theta\Gamma^{\phi}{}_{\phi\theta} + \partial_\phi\Gamma^{\phi}{}_{\theta\theta} + \Gamma^{e}{}_{\theta\theta}\Gamma^{\phi}{}_{\phi e} - \Gamma^{e}{}_{\phi\theta}\Gamma^{\phi}{}_{\theta e} \\ &= -\frac{\mathrm{d}}{\mathrm{d}\theta}(\cot\theta) - \cot^2\theta \\ &= \csc^2\theta - \cot^2\theta = 1. \end{aligned} \tag{232}$$

The second and third terms vanish because $\Gamma^{\phi}{}_{\theta\theta} = 0$ and $\Gamma^{e}{}_{\theta\theta} = 0$. Only $e = \phi$ contributes to the last product. Lowering the final index gives

$$R_{\theta\phi\theta\phi} = g_{\phi\phi}R_{\theta\phi\theta}{}^{\phi} = L^2\sin^2\theta. \tag{233}$$

The component with one upper index has no explicit factor of L; the fully covariant component does. Index placement matters when reading the dimensions of a coordinate expression.

49.2 Computing every Ricci component

The $\theta\theta$ component has only one nonzero contribution:

$$R_{\theta\theta} = R_{\theta c\theta}{}^{c} = R_{\theta\phi\theta}{}^{\phi} = 1. \tag{234}$$

For the $\phi\phi$ component,

$$\begin{aligned} R_{\phi\phi} &= R_{\phi\theta\phi}{}^{\theta} = g^{\theta\theta} R_{\phi\theta\phi\theta} \\ &= g^{\theta\theta} R_{\theta\phi\theta\phi} = \frac{1}{L^2} L^2 \sin^2\theta = \sin^2\theta. \end{aligned} \tag{235}$$

The off-diagonal entries vanish:

$$R_{\theta\phi} = R_{\phi\theta} = 0.$$

For example, in $R_{\theta\phi} = R_{\theta c\phi}{}^{c}$, the $c = \theta$ term has a repeated index in the first antisymmetric pair, and the $c = \phi$ term vanishes after lowering the final index because of the last antisymmetric pair. Thus

$$(R_{ab}) = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix} = \frac{1}{L^2}(g_{ab}). \tag{236}$$

49.3 Scalar curvature and the radius

Finally,

$$\begin{aligned} R &= g^{\theta\theta} R_{\theta\theta} + g^{\phi\phi} R_{\phi\phi} \\ &= \frac{1}{L^2} \cdot 1 + \frac{1}{L^2 \sin^2\theta} \cdot \sin^2\theta = \boxed{\frac{2}{L^2}}. \end{aligned} \tag{237}$$

The coordinate-dependent factors cancel: the round sphere has the same scalar curvature at every point. A larger sphere is less curved, with curvature proportional to inverse length squared. Although the connection and the displayed Ricci components contain no L, the inverse metric in the scalar contraction restores the physical curvature scale.

The same results can be collected in the coordinate-independent formula

$$R_{abcd} = \frac{1}{L^2}(g_{ac}g_{bd} - g_{ad}g_{bc}). \tag{238}$$

For example, its $\theta\phi\theta\phi$ component is $L^{-2} g_{\theta\theta} g_{\phi\phi} = L^2 \sin^2\theta$, agreeing with the explicit calculation. Contracting it gives (236) and then (237).

50 Einstein spaces

Definition 50.1. An **Einstein space** is a manifold whose Ricci tensor is proportional to the metric with a constant proportionality factor:

$$R_{ab} = \lambda g_{ab}, \qquad \lambda = \text{constant}. \tag{239}$$

For Lorentzian signature one also speaks of an Einstein spacetime.

The round two-sphere is an example with $\lambda = L^{-2}$. More generally, tracing (239) on an n-dimensional manifold gives

$$R = g^{ab} R_{ab} = \lambda g^{ab} g_{ab} = n\lambda.$$

The scale λ has units of inverse length squared, consistent with the fact that curvature involves two derivatives of the metric.

In general relativity, such a constant is related to the cosmological constant. A positive cosmological constant is one possible description of dark energy and accelerated cosmic expansion. This supplies a physical motivation for studying geometries whose Ricci tensor is proportional to their metric.