

Semi-Riemann Geometry and General Relativity

Shlomo Sternberg

September 24, 2003

0.1 Introduction

This book represents course notes for a one semester course at the undergraduate level giving an introduction to Riemannian geometry and its principal physical application, Einstein's theory of general relativity. The background assumed is a good grounding in linear algebra and in advanced calculus, preferably in the language of differential forms.

Chapter I introduces the various curvatures associated to a hypersurface embedded in Euclidean space, motivated by the formula for the volume for the region obtained by thickening the hypersurface on one side. If we thicken the hypersurface by an amount h in the normal direction, this formula is a polynomial in h whose coefficients are integrals over the hypersurface of local expressions. These local expressions are elementary symmetric polynomials in what are known as the principal curvatures. The precise definitions are given in the text. The chapter culminates with Gauss' *Theorema egregium* which asserts that if we thicken a two dimensional surface evenly on *both* sides, then the these integrands depend only on the intrinsic geometry of the surface, and not on how the surface is embedded. We give two proofs of this important theorem. (We give several more later in the book.) The first proof makes use of "normal coordinates" which become so important in Riemannian geometry and, as "inertial frames," in general relativity. It was this theorem of Gauss, and particularly the very notion of "intrinsic geometry", which inspired Riemann to develop his geometry.

Chapter II is a rapid review of the differential and integral calculus on manifolds, including differential forms, the d operator, and Stokes' theorem. Also vector fields and Lie derivatives. At the end of the chapter are a series of sections in exercise form which lead to the notion of parallel transport of a vector along a curve on a embedded surface as being associated with the "rolling of the surface on a plane along the curve".

Chapter III discusses the fundamental notions of linear connections and their curvatures, and also Cartan's method of calculating curvature using frame fields and differential forms. We show that the geodesics on a Lie group equipped with a bi-invariant metric are the translates of the one parameter subgroups. A short exercise set at the end of the chapter uses the Cartan calculus to compute the curvature of the Schwarzschild metric. A second exercise set computes some geodesics in the Schwarzschild metric leading to two of the famous predictions of general relativity: the advance of the perihelion of Mercury and the bending of light by matter. Of course the theoretical basis of these computations, i.e. the theory of general relativity, will come later, in Chapter VII.

Chapter IV begins by discussing the bundle of frames which is the modern setting for Cartan's calculus of "moving frames" and also the jumping off point for the general theory of connections on principal bundles which lie at the base of such modern physical theories as Yang-Mills fields. This chapter seems to present the most difficulty conceptually for the student.

Chapter V discusses the general theory of connections on fiber bundles and then specialize to principal and associated bundles.

Chapter VI returns to Riemannian geometry and discusses Gauss's lemma which asserts that the radial geodesics emanating from a point are orthogonal (in the Riemann metric) to the images under the exponential map of the spheres in the tangent space centered at the origin. From this one concludes that geodesics (defined as self parallel curves) locally minimize arc length in a Riemann manifold.

Chapter VII is a rapid review of special relativity. It is assumed that the students will have seen much of this material in a physics course.

Chapter VIII is the high point of the course from the theoretical point of view. We discuss Einstein's general theory of relativity from the point of view of the Einstein-Hilbert functional. In fact we borrow the title of Hilbert's paper for the Chapter heading. We also introduce the principle of general covariance, first introduced by Einstein, Infeld, and Hoffmann to derive the "geodesic principle" and give a whole series of other applications of this principle.

Chapter IX discusses computational methods deriving from the notion of a Riemannian submersion, introduced and developed by Robert Hermann and perfected by Barrett O'Neill. It is the natural setting for the generalized Gauss-Codazzi type equations. Although technically somewhat demanding at the beginning, the range of applications justifies the effort in setting up the theory. Applications range from curvature computations for homogeneous spaces to cosmogeny and eschatology in Friedman type models.

Chapter X discusses the Petrov classification, using complex geometry, of the various types of solutions to the Einstein equations in four dimensions. This classification led Kerr to his discovery of the rotating black hole solution which is a topic for a course in its own. The exposition in this chapter follows joint work with Kostant.

Chapter XI is in the form of a enlarged exercise set on the star operator. It is essentially independent of the entire course, but I thought it useful to include, as it would be of interest in any more advanced treatment of topics in the course.

Contents

0.1	Introduction	2
1	The principal curvatures.	11
1.1	Volume of a thickened hypersurface	11
1.2	The Gauss map and the Weingarten map.	13
1.3	Proof of the volume formula.	16
1.4	Gauss's theorema egregium.	19
	1.4.1 First proof, using inertial coordinates.	22
	1.4.2 Second proof. The Brioschi formula.	25
1.5	Problem set - Surfaces of revolution.	27
2	Rules of calculus.	31
2.1	Superalgebras.	31
2.2	Differential forms.	31
2.3	The d operator.	32
2.4	Derivations.	33
2.5	Pullback.	34
2.6	Chain rule.	35
2.7	Lie derivative.	35
2.8	Weil's formula.	36
2.9	Integration.	38
2.10	Stokes theorem.	38
2.11	Lie derivatives of vector fields.	39
2.12	Jacobi's identity.	40
2.13	Left invariant forms.	41
2.14	The Maurer Cartan equations.	43
2.15	Restriction to a subgroup	43
2.16	Frames.	44
2.17	Euclidean frames.	45
2.18	Frames adapted to a submanifold.	47
2.19	Curves and surfaces - their structure equations.	48
2.20	The sphere as an example.	48
2.21	Ribbons	50
2.22	Developing a ribbon.	51
2.23	Parallel transport along a ribbon.	52

2.24	Surfaces in \mathbf{R}^3 .	53
3	Levi-Civita Connections.	57
3.1	Definition of a linear connection on the tangent bundle.	57
3.2	Christoffel symbols.	58
3.3	Parallel transport.	58
3.4	Geodesics.	60
3.5	Covariant differential.	61
3.6	Torsion.	63
3.7	Curvature.	63
3.8	Isometric connections.	65
3.9	Levi-Civita's theorem.	65
3.10	Geodesics in orthogonal coordinates.	67
3.11	Curvature identities.	68
3.12	Sectional curvature.	69
3.13	Ricci curvature.	69
3.14	Bi-invariant metrics on a Lie group.	70
3.14.1	The Lie algebra of a Lie group.	70
3.14.2	The general Maurer-Cartan form.	72
3.14.3	Left invariant and bi-invariant metrics.	73
3.14.4	Geodesics are cosets of one parameter subgroups.	74
3.14.5	The Riemann curvature of a bi-invariant metric.	75
3.14.6	Sectional curvatures.	75
3.14.7	The Ricci curvature and the Killing form.	75
3.14.8	Bi-invariant forms from representations.	76
3.14.9	The Weinberg angle.	78
3.15	Frame fields.	78
3.16	Curvature tensors in a frame field.	79
3.17	Frame fields and curvature forms.	79
3.18	Cartan's lemma.	82
3.19	Orthogonal coordinates on a surface.	83
3.20	The curvature of the Schwarzschild metric	84
3.21	Geodesics of the Schwarzschild metric.	85
3.21.1	Massive particles.	88
3.21.2	Massless particles.	93
4	The bundle of frames.	95
4.1	Connection and curvature forms in a frame field.	95
4.2	Change of frame field.	96
4.3	The bundle of frames.	98
4.3.1	The form ϑ .	99
4.3.2	The form ϑ in terms of a frame field.	99
4.3.3	The definition of $\bar{\omega}$.	99
4.4	The connection form in a frame field as a pull-back.	100
4.5	Gauss' theorems.	103
4.5.1	Equations of structure of Euclidean space.	103

4.5.2	Equations of structure of a surface in \mathbf{R}^3 .	104
4.5.3	<i>Theorema egregium</i> .	104
4.5.4	Holonomy.	104
4.5.5	Gauss-Bonnet.	105
5	Connections on principal bundles.	107
5.1	Submersions, fibrations, and connections.	107
5.2	Principal bundles and invariant connections.	111
5.2.1	Principal bundles.	111
5.2.2	Connections on principal bundles.	113
5.2.3	Associated bundles.	115
5.2.4	Sections of associated bundles.	116
5.2.5	Associated vector bundles.	117
5.2.6	Exterior products of vector valued forms.	119
5.3	Covariant differentials and covariant derivatives.	121
5.3.1	The horizontal projection of forms.	121
5.3.2	The covariant differential of forms on P .	122
5.3.3	A formula for the covariant differential of basic forms.	122
5.3.4	The curvature is $\mathbf{d}\bar{\omega}$.	123
5.3.5	Bianchi's identity.	123
5.3.6	The curvature and \mathbf{d}^2 .	123
6	Gauss's lemma.	125
6.1	The exponential map.	125
6.2	Normal coordinates.	126
6.3	The Euler field \mathcal{E} and its image \mathcal{P} .	127
6.4	The normal frame field.	128
6.5	Gauss' lemma.	129
6.6	Minimization of arc length.	131
7	Special relativity	133
7.1	Two dimensional Lorentz transformations.	133
7.1.1	Addition law for velocities.	135
7.1.2	Hyperbolic angle.	135
7.1.3	Proper time.	136
7.1.4	Time dilatation.	137
7.1.5	Lorentz-Fitzgerald contraction.	137
7.1.6	The reverse triangle inequality.	138
7.1.7	Physical significance of the Minkowski distance.	138
7.1.8	Energy-momentum	139
7.1.9	Psychological units.	140
7.1.10	The Galilean limit.	142
7.2	Minkowski space.	142
7.2.1	The Compton effect.	143
7.2.2	Natural Units.	146
7.2.3	Two-particle invariants.	147

7.2.4	Mandlestam variables.	150
7.3	Scattering cross-section and mutual flux.	154
8	Die Grundlagen der Physik.	157
8.1	Preliminaries.	157
8.1.1	Densities and divergences.	157
8.1.2	Divergence of a vector field on a semi-Riemannian manifold.	160
8.1.3	The Lie derivative of a semi-Riemann metric.	162
8.1.4	The covariant divergence of a symmetric tensor field.	163
8.2	Varying the metric and the connection.	167
8.3	The structure of physical laws.	169
8.3.1	The Legendre transformation.	169
8.3.2	The passive equations.	172
8.4	The Hilbert “function”.	173
8.5	Schrodinger’s equation as a passive equation.	175
8.6	Harmonic maps.	175
9	Submersions.	179
9.1	Submersions.	179
9.2	The fundamental tensors of a submersion.	181
9.2.1	The tensor T	181
9.2.2	The tensor A	182
9.2.3	Covariant derivatives of T and A	183
9.2.4	The fundamental tensors for a warped product.	185
9.3	Curvature.	186
9.3.1	Curvature for warped products.	190
9.3.2	Sectional curvature.	193
9.4	Reductive homogeneous spaces.	194
9.4.1	Bi-invariant metrics on a Lie group.	194
9.4.2	Homogeneous spaces.	197
9.4.3	Normal symmetric spaces.	197
9.4.4	Orthogonal groups.	198
9.4.5	Dual Grassmannians.	200
9.5	Schwarzschild as a warped product.	202
9.5.1	Surfaces with orthogonal coordinates.	203
9.5.2	The Schwarzschild plane.	204
9.5.3	Covariant derivatives.	205
9.5.4	Schwarzschild curvature.	206
9.5.5	Cartan computation.	207
9.5.6	Petrov type.	209
9.5.7	Kerr-Schild form.	210
9.5.8	Isometries.	211
9.6	Robertson Walker metrics.	214
9.6.1	Cosmogeny and eschatology.	216

10 Petrov types.	217
10.1 Algebraic properties of the curvature tensor	217
10.2 Linear and antilinear maps.	219
10.3 Complex conjugation and real forms.	221
10.4 Structures on tensor products.	223
10.5 Spinors and Minkowski space.	224
10.6 Traceless curvatures.	225
10.7 The polynomial algebra.	225
10.8 Petrov types.	226
10.9 Principal null directions.	227
10.10 Kerr-Schild metrics.	230
11 Star.	233
11.1 Definition of the star operator.	233
11.2 Does $\star : \wedge^k V \rightarrow \wedge^{n-k} V$ determine the metric?	235
11.3 The star operator on forms.	240
11.3.1 For \mathbf{R}^2	240
11.3.2 For \mathbf{R}^3	241
11.3.3 For $\mathbf{R}^{1,3}$	242
11.4 Electromagnetism.	243
11.4.1 Electrostatics.	243
11.4.2 Magnetoquasistatics.	244
11.4.3 The London equations.	246
11.4.4 The London equations in relativistic form.	248
11.4.5 Maxwell's equations.	249
11.4.6 Comparing Maxwell and London.	249

Chapter 1

The principal curvatures.

1.1 Volume of a thickened hypersurface

We want to consider the following problem: Let $Y \subset \mathbf{R}^n$ be an oriented hypersurface, so there is a well defined unit normal vector, $\nu(y)$, at each point of Y . Let Y_h denote the set of all points of the form

$$y + t\nu(y), \quad 0 \leq t \leq h.$$

We wish to compute $V_n(Y_h)$ where V_n denotes the n -dimensional volume. We will do this computation for small h , see the discussion after the examples.

Examples in three dimensional space.

1. Suppose that Y is a bounded region in a plane, of area A . Clearly

$$V_3(Y_h) = hA$$

in this case.

2. Suppose that Y is a right circular cylinder of radius r and height ℓ with outwardly pointing normal. Then Y_h is the region between the right circular cylinders of height ℓ and radii r and $r + h$ so

$$\begin{aligned} V_3(Y_h) &= \pi\ell[(r+h)^2 - r^2] \\ &= 2\pi\ell rh + \pi\ell h^2 \\ &= hA + h^2 \cdot \frac{1}{2r} \cdot A \\ &= A \left(h + \frac{1}{2} \cdot kh^2 \right), \end{aligned}$$

where $A = 2\pi r\ell$ is the area of the cylinder and where $k = 1/r$ is the curvature of the generating circle of the cylinder. For small h , this formula is correct, in fact,

whether we choose the normal vector to point out of the cylinder or into the cylinder. Of course, in the inward pointing case, the curvature has the opposite sign, $k = -1/r$.

For inward pointing normals, the formula breaks down when $h > r$, since we get multiple coverage of points in space by points of the form $y + t\nu(y)$.

3. Y is a sphere of radius R with outward normal, so Y_h is a spherical shell, and

$$\begin{aligned} V_3(Y_h) &= \frac{4}{3}\pi[(R+h)^3 - R^3] \\ &= h4\pi R^2 + h^2 4\pi R + h^3 \frac{4}{3}\pi \\ &= hA + h^2 \frac{1}{R}A + h^3 \frac{1}{3R^2}A \\ &= \frac{1}{3} \cdot A \cdot \left[3h + 3\frac{1}{R} \cdot h^2 + \frac{1}{R^2}h^3 \right], \end{aligned}$$

where $A = 4\pi R^2$ is the area of the sphere.

Once again, for inward pointing normals we must change the sign of the coefficient of h^2 and the formula thus obtained is only correct for $h \leq \frac{1}{R}$.

So in general, we wish to make the assumption that h is such that the map

$$Y \times [0, h] \rightarrow \mathbf{R}^n, \quad (y, t) \mapsto y + t\nu(y)$$

is injective. For Y compact, there always exists an $h_0 > 0$ such that this condition holds for all $h < h_0$. This can be seen to be a consequence of the implicit function theorem. But so not to interrupt the discussion, we will take the injectivity of the map as an hypothesis, for the moment.

In a moment we will define the notion of the various averaged curvatures, H_1, \dots, H_{n-1} , of a hypersurface, and find for the case of the sphere with outward pointing normal, that

$$H_1 = \frac{1}{R}, \quad H_2 = \frac{1}{R^2},$$

while for the case of the cylinder with outward pointing normal that

$$H_1 = \frac{1}{2r}, \quad H_2 = 0,$$

and for the case of the planar region that

$$H_1 = H_2 = 0.$$

We can thus write all three of the above the above formulas as

$$V_3(Y_h) = \frac{1}{3}A [3h + 3H_1h^2 + H_2h^3].$$

1.2 The Gauss map and the Weingarten map.

In order to state the general formula, we make the following definitions: Let Y be an (immersed) oriented hypersurface. At each $x \in Y$ there is a unique (positive) unit normal vector, and hence a well defined **Gauss map**

$$\nu : Y \rightarrow S^{n-1}$$

assigning to each point $x \in Y$ its unit normal vector, $\nu(x)$. Here S^{n-1} denotes the unit sphere, the set of all unit vectors in \mathbf{R}^n .

The normal vector, $\nu(x)$ is orthogonal to the tangent space to Y at x . We will denote this tangent space by TY_x . For our present purposes, we can regard TY_x as a subspace of \mathbf{R}^n : If $t \mapsto \gamma(t)$ is a differentiable curve lying on the hypersurface Y , (this means that $\gamma(t) \in Y$ for all t) and if $\gamma(0) = x$, then $\gamma'(0)$ belongs to the tangent space TY_x . Conversely, given any vector $v \in TY_x$, we can always find a differentiable curve γ with $\gamma(0) = x$, $\gamma'(0) = v$. So a good way to think of a tangent vector to Y at x is as an “infinitesimal curve” on Y passing through x .

Examples:

1. Suppose that Y is a portion of an $(n-1)$ dimensional linear or affine subspace sitting in \mathbf{R}^n . For example suppose that $Y = \mathbf{R}^{n-1}$ consisting of those points in \mathbf{R}^n whose last coordinate vanishes. Then the tangent space to Y at every point is just this same subspace, and hence the normal vector is a constant. The Gauss map is thus a constant, mapping all of Y onto a single point in S^{n-1} .
2. Suppose that Y is the sphere of radius R (say centered at the origin). The Gauss map carries every point of Y into the corresponding (parallel) point of S^{n-1} . In other words, it is multiplication by $1/R$:

$$\nu(y) = \frac{1}{R}y.$$

3. Suppose that Y is a right circular cylinder in \mathbf{R}^3 whose base is the circle of radius r in the x^1, x^2 plane. Then the Gauss map sends Y onto the equator of the unit sphere, S^2 , sending a point x into $(1/r)\pi(x)$ where $\pi : \mathbf{R}^3 \rightarrow \mathbf{R}^2$ is projection onto the x^1, x^2 plane.

Another good way to think of the tangent space is in terms of a **local parameterization** which means that we are given a map $X : M \mapsto \mathbf{R}^n$ where M is some open subset of \mathbf{R}^{n-1} and such that $X(M)$ is some neighborhood of x in Y . Let y^1, \dots, y^{n-1} be the standard coordinates on \mathbf{R}^{n-1} . Part of the requirement that goes into the definition of parameterization is that the map X be **regular**, in the sense that its Jacobian matrix

$$dX := \left(\frac{\partial X}{\partial y^1}, \dots, \frac{\partial X}{\partial y^{n-1}} \right)$$

whose columns are the partial derivatives of the map X has rank $n - 1$ everywhere. The matrix dX has n rows and $n - 1$ columns. The regularity condition amounts to the assertion that for each $z \in M$ the vectors,

$$\frac{\partial X}{\partial y^1}(z), \dots, \frac{\partial X}{\partial y^{n-1}}(z)$$

span a subspace of dimension $n - 1$. If $x = X(y)$ then the tangent space TY_x is precisely the space spanned by

$$\frac{\partial X}{\partial y^1}(y), \dots, \frac{\partial X}{\partial y^{n-1}}(y).$$

Suppose that F is a differentiable map from Y to \mathbf{R}^m . We can then define its differential, $dF_x : TY_x \mapsto \mathbf{R}^m$. It is a linear map assigning to each $v \in TY_x$ a value $dF_x(v) \in \mathbf{R}^m$: In terms of the “infinitesimal curve” description, if $v = \gamma'(0)$ then

$$dF_x(v) = \frac{dF \circ \gamma}{dt}(0).$$

(You must check that this does not depend on the choice of representing curve, γ .)

Alternatively, to give a linear map, it is enough to give its value at the elements of a basis. In terms of the basis coming from a parameterization, we have

$$dF_x \left(\frac{\partial X}{\partial y^i}(y) \right) = \frac{\partial F \circ X}{\partial y^i}(y).$$

Here $F \circ X : M \rightarrow \mathbf{R}^m$ is the composition of the map F with the map X . You must check that the map dF_x so determined does not depend on the choice of parameterization. Both of these verifications proceed by the chain rule.

One immediate consequence of either characterization is the following important property. Suppose that F takes values in a submanifold $Z \subset \mathbf{R}^m$. Then

$$dF_x : TY_x \rightarrow TZ_{F(x)}.$$

Let us apply all this to the Gauss map, ν , which maps Y to the unit sphere, S^{n-1} . Then

$$d\nu_x : TY_x \rightarrow TS_{\nu(x)}^{n-1}.$$

But the tangent space to the unit sphere at $\nu(x)$ consists of all vectors perpendicular to $\nu(x)$ and so can be identified with TY_x . We define the **Weingarten map** to be the differential of the Gauss map, regarded as a map from TY_x to itself:

$$W_x := d\nu_x, \quad W_x : TY_x \rightarrow TY_x.$$

The **second fundamental form** is defined to be the bilinear form on TY_x given by

$$II_x(v, w) := (W_x v, w).$$

In the next section we will show, using local coordinates, that this form is symmetric, i.e. that

$$(W_x u, v) = (u, W_x v).$$

This implies, from linear algebra, that W_x is diagonalizable with real eigenvalues. These eigenvalues, $k_1 = k_1(x), \dots, k_{n-1} = k_{n-1}(x)$, of the Weingarten map are called the **principal curvatures** of Y at the point x .

Examples:

1. For a portion of $(n - 1)$ space sitting in \mathbf{R}^n the Gauss map is constant so its differential is zero. Hence the Weingarten map and thus all the principal curvatures are zero.
2. For the sphere of radius R the Gauss map consists of multiplication by $1/R$ which is a linear transformation. The differential of a linear transformation is that same transformation (regarded as acting on the tangent spaces). Hence the Weingarten map is $1/R \times \text{id}$ and so all the principal curvatures are equal and are equal to $1/R$.
3. For the cylinder, again the Gauss map is linear, and so the principal curvatures are 0 and $1/r$.

We let H_j denote the j th normalized elementary symmetric functions of the principal curvatures. So

$$\begin{aligned} H_0 &= 1 \\ H_1 &= \frac{1}{n-1}(k_1 + \dots + k_{n-1}) \\ H_{n-1} &= k_1 \cdot k_2 \cdot \dots \cdot k_{n-1} \end{aligned}$$

and, in general,

$$H_j = \binom{n-1}{j}^{-1} \sum_{1 \leq i_1 < \dots < i_j \leq n-1} k_{i_1} \cdot \dots \cdot k_{i_j}. \quad (1.1)$$

H_1 is called the **mean curvature** and H_{n-1} is called the **Gaussian curvature**. All the principal curvatures are functions of the point $x \in Y$. For notational simplicity, we will frequently suppress the dependence on x . Then the formula for the volume of the thickened hypersurface (we will call this the “volume formula” for short) is:

$$V_n(Y_h) = \frac{1}{n} \sum_{i=1}^n \binom{n}{i} h^i \int_Y H_{i-1} d^{n-1}A \quad (1.2)$$

where $d^{n-1}A$ denotes the $(n - 1)$ dimensional (area) measure on Y .

A immediate check shows that this gives the answers that we got above for the the plane, the cylinder, and the sphere.

1.3 Proof of the volume formula.

We recall that the Gauss map, ν assigns to each point $x \in Y$ its unit normal vector, and so is a map from Y to the unit sphere, S^{n-1} . The Weingarten map, W_x , is the differential of the Gauss map, $W_x = d\nu_x$, regarded as a map of the tangent space, TY_x to itself. We now describe these maps in terms of a local parameterization of Y . So let $X : M \rightarrow \mathbf{R}^n$ be a parameterization of class C^2 of a neighborhood of Y near x , where M is an open subset of \mathbf{R}^{n-1} . So $x = X(y)$, $y \in M$, say. Let

$$N := \nu \circ X$$

so that $N : M \rightarrow S^{n-1}$ is a map of class C^1 . The map

$$dX_y : \mathbf{R}^{n-1} \rightarrow TY_x$$

gives a **frame** of TY_x . The word “frame” means an isomorphism of our “standard” $(n-1)$ -dimensional space, \mathbf{R}^{n-1} with our given $(n-1)$ -dimensional space, TY_x . Here we have identified $T(\mathbf{R}^{n-1})_y$ with \mathbf{R}^{n-1} , so the frame dX_y gives us a particular isomorphism of \mathbf{R}^{n-1} with TY_x .

Giving a frame of a vector space is the same as giving a basis of that vector space. We will use these two different ways of using the word “frame” interchangeably. Let e_1, \dots, e_{n-1} denote the standard basis of \mathbf{R}^{n-1} , and for X and N , let the subscript i denote the partial derivative with respect to the i th Cartesian coordinate. Thus

$$dX_y(e_i) = X_i(y)$$

for example, and so $X_1(y), \dots, X_{n-1}(y)$ “is” the frame determined by dX_y (when we regard TY_x as a subspace of \mathbf{R}^n). For the sake of notational simplicity we will drop the argument y . Thus we have

$$\begin{aligned} dX(e_i) &= X_i, \\ dN(e_i) &= N_i, \\ \text{and so} & \\ W_x X_i &= N_i. \end{aligned}$$

Recall the definition, $II_x(v, w) = (W_x v, w)$, of the second fundamental form. Let (L_{ij}) denote the matrix of the second fundamental form with respect to the basis X_1, \dots, X_{n-1} of TY_x . So

$$\begin{aligned} L_{ij} &= II_x(X_i, X_j) \\ &= (W_x X_i, X_j) \\ &= (N_i, X_j) \end{aligned}$$

so

$$L_{ij} = -\left(N, \frac{\partial^2 X}{\partial y_i \partial y_j}\right), \quad (1.3)$$

the last equality coming from differentiating the identity

$$(N, X_j) \equiv 0$$

in the i th direction. In particular, it follows from (1.3) and the equality of cross derivatives that

$$(W_x X_i, X_j) = (X_i, W_x X_j)$$

and hence, by linearity that

$$(W_x u, v) = (u, W_x v) \quad \forall u, v \in TY_x.$$

We have proved that the second fundamental form is symmetric, and hence the Weingarten map is diagonalizable with real eigenvalues.

Recall that the principal curvatures are, by definition, the eigenvalues of the Weingarten map. We will let

$$W = (W_{ij})$$

denote the matrix of the Weingarten map with respect to the basis X_1, \dots, X_{n-1} . Explicitly,

$$N_i = \sum_j W_{ji} X_j.$$

If we write $N_1, \dots, N_{n-1}, X_1, \dots, X_{n-1}$ as column vectors of length n , we can write the preceding equation as the matrix equation

$$(N_1, \dots, N_{n-1}) = (X_1, \dots, X_{n-1})W. \quad (1.4)$$

The matrix multiplication on the right is that of an $n \times (n-1)$ matrix with an $(n-1) \times (n-1)$ matrix. To understand this abbreviated notation, let us write it out in the case $n = 3$, so that X_1, X_2, N_1, N_2 are vectors in \mathbf{R}^3 :

$$X_1 = \begin{pmatrix} X_{11} \\ X_{12} \\ X_{13} \end{pmatrix}, \quad X_2 = \begin{pmatrix} X_{21} \\ X_{22} \\ X_{23} \end{pmatrix}, \quad N_1 = \begin{pmatrix} N_{11} \\ N_{12} \\ N_{13} \end{pmatrix}, \quad N_2 = \begin{pmatrix} N_{21} \\ N_{22} \\ N_{23} \end{pmatrix}.$$

Then (1.4) is the matrix equation

$$\begin{pmatrix} N_{11} & N_{21} \\ N_{12} & N_{22} \\ N_{13} & N_{23} \end{pmatrix} = \begin{pmatrix} X_{11} & X_{21} \\ X_{12} & X_{22} \\ X_{13} & X_{23} \end{pmatrix} \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}.$$

Matrix multiplication shows that this gives

$$N_1 = W_{11}X_1 + W_{21}X_2, \quad N_2 = W_{12}X_1 + W_{22}X_2,$$

and more generally that (1.4) gives $N_i = \sum_j W_{ji}X_j$ in all dimensions.

Now consider the region Y_h , the thickened hypersurface, introduced in the preceding section except that we replace the full hypersurface Y by the portion $X(M)$. Thus the region in space that we are considering is

$$\{X(y) + \lambda N(y), y \in M, 0 < \lambda \leq h\}.$$

It is the image of the region $M \times (0, h] \subset \mathbf{R}^{n-1} \times \mathbf{R}$ under the map

$$(y, \lambda) \mapsto X(y) + \lambda N(y).$$

We are assuming that this map is injective. By (1.4), it has Jacobian matrix (differential)

$$J = (X_1 + \lambda N_1, \dots, X_{n-1} + \lambda N_{n-1}, N) = (X_1, \dots, X_{n-1}, N) \begin{pmatrix} (I_{n-1} + \lambda W) & 0 \\ 0 & 1 \end{pmatrix}. \quad (1.5)$$

The right hand side of (1.5) is now the product of two n by n matrices.

The change of variables formula in several variables says that

$$V_n(h) = \int_M \int_0^h |\det J| dh dy_1 \cdots dy_{n-1}. \quad (1.6)$$

Let us take the determinant of the right hand side of (1.5). The determinant of the matrix (X_1, \dots, X_{n-1}, N) is just the (oriented) n dimensional volume of the parallelepiped spanned by X_1, \dots, X_{n-1}, N . Since N is of unit length and is perpendicular to the X 's, this is the same as the (oriented) $n-1$ dimensional volume of the parallelepiped spanned by X_1, \dots, X_{n-1} . Thus, “by definition”,

$$|\det (X_1, \dots, X_{n-1}, N)| dy_1 \cdots dy_{n-1} = d^{n-1} A. \quad (1.7)$$

(We will come back shortly to discuss why this is the right definition.) The second factor on the right hand side of (1.5) contributes

$$\det(1 + \lambda W) = (1 + \lambda k_1) \cdots (1 + \lambda k_{n-1}).$$

For sufficiently small λ , this expression is positive, so we need not worry about the absolute value sign if h small enough. Integrating with respect to λ from 0 to h gives (1.2).

We proved (1.2) if we define $d^{n-1} A$ to be given by (1.7). But then it follows from (1.2) that

$$\frac{d}{dh} V_n(Y_h)|_{h=0} = \int_Y d^{n-1} A. \quad (1.8)$$

A moment's thought shows that the left hand side of (1.8) is exactly what we want to mean by “area”: it is the “volume of an infinitesimally thickened region”. This justifies taking (1.7) as a definition. Furthermore, although the definition (1.7) is only valid in a coordinate neighborhood, and seems to depend on the choice of local coordinates, equation (1.8) shows that it is independent of the local description by coordinates, and hence is a well defined object on Y . The functions H_j have been defined independent of any choice of local coordinates. Hence (1.2) works globally: To compute the right hand side of (1.2) we may have to break Y up into patches, and do the integration in each patch, summing the pieces. But we know in advance that the final answer is independent of how we break Y up or which local coordinates we use.

1.4 Gauss's theorema egregium.

Suppose we consider the two sided region about the surface, that is

$$V_n(Y_h^+) + V_n(Y_h^-)$$

corresponding to the two different choices of normals. When we replace $\nu(x)$ by $-\nu(x)$ at each point, the Gauss map ν is replaced by $-\nu$, and hence the Weingarten maps W_x are also replaced by their negatives. The principal curvatures change sign. Hence, in the above sum the coefficients of the even powers of h cancel, since they are given in terms of products of the principal curvatures with an odd number of factors. For $n = 3$ we are left with a sum of two terms, the coefficient of h which is the area, and the coefficient of h^3 which is the integral of the Gaussian curvature. It was the remarkable discovery of Gauss that this curvature depends only on the intrinsic geometry of the surface, and not on how the surface is embedded into three space. Thus, for both the cylinder and the plane the cubic terms vanish, because (locally) the cylinder is isometric to the plane. We can wrap the plane around the cylinder without stretching or tearing.

It was this fundamental observation of Gauss that led Riemann to investigate the intrinsic metric geometry of higher dimensional space, eventually leading to Einstein's general relativity which derives the gravitational force from the curvature of space time. A first objective will be to understand this major theorem of Gauss.

An important generalization of Gauss's result was proved by Hermann Weyl in 1939. He showed: if Y is any k dimensional submanifold of n dimensional space (so for $k = 1$, $n = 3$ Y is a curve in three space), let $Y(h)$ denote the "tube" around Y of radius h , the set of all points at distance h from Y . Then, for small h , $V_n(Y(h))$ is a polynomial in h whose coefficients are integrals over Y of intrinsic expressions, depending only on the notion of distance within Y .

Let us multiply both sides of (1.4) on the left by the matrix $(X_1, \dots, X_{n-1})^T$ to obtain

$$L = QW$$

where $L_{ij} = (X_i, N_j)$ as before, and

$$Q = (Q_{ij}) := (X_i, X_j)$$

is called the matrix of the **first fundamental form** relative to our choice of local coordinates. All three matrices in this equality are of size $(n-1) \times (n-1)$. If we take the determinant of the equation $L = QW$ we obtain

$$\det W = \frac{\det L}{\det Q}, \tag{1.9}$$

an expression for the determinant of the Weingarten map (a geometrical property of the embedded surface) as the quotient of two local expressions. For the case $n-1 = 2$, we thus obtain a local expression for the Gaussian curvature, $K = \det W$.

The first fundamental form encodes the intrinsic geometry of the hypersurface in terms of local coordinates: it gives the Euclidean geometry of the tangent space in terms of the basis X_1, \dots, X_{n-1} . If we describe a curve $t \mapsto \gamma(t)$ on the surface in terms of the coordinates y^1, \dots, y^{n-1} by giving the functions $t \mapsto y^j(t)$, $j = 1, \dots, n-1$ then the chain rule says that

$$\gamma'(t) = \sum_{j=1}^{n-1} X_j(y(t)) \frac{dy^j}{dt}(t)$$

where

$$y(t) = (y^1(t), \dots, y^{n-1}(t)).$$

Therefore the (Euclidean) square length of the tangent vector $\gamma'(t)$ is

$$\|\gamma'(t)\|^2 = \sum_{i,j=1}^{n-1} Q_{ij}(y(t)) \frac{dy^i}{dt}(t) \frac{dy^j}{dt}(t).$$

Thus the length of the curve γ given by

$$\int \|\gamma'(t)\| dt$$

can be computed in terms of $y(t)$ as

$$\int \sqrt{\sum_{i,j=1}^{n-1} Q_{ij}(y(t)) \frac{dy^i}{dt}(t) \frac{dy^j}{dt}(t)} dt$$

(so long as the curve lies within the coordinate system).

So two hypersurfaces have the same local intrinsic geometry if they have the same Q in any local coordinate system.

In order to conform with a (somewhat variable) classical literature, we shall make some slight changes in our notation for the case of surfaces in three dimensional space. We will denote our local coordinates by u, v instead of y_1, y_2 and so X_u will replace X_1 and X_v will replace X_2 , and we will denote the scalar product of two vectors in three dimensional space by a \cdot instead of $(,)$. We write

$$Q = \begin{pmatrix} E & F \\ F & G \end{pmatrix} \tag{1.10}$$

where

$$E := X_u \cdot X_u \tag{1.11}$$

$$F := X_u \cdot X_v \tag{1.12}$$

$$G := X_v \cdot X_v \tag{1.13}$$

so

$$\det Q = EG - F^2. \tag{1.14}$$

We can write the equations (1.11)-(1.13) as

$$Q = (X_u, X_v)^\dagger (X_u, X_v). \quad (1.15)$$

Similarly, let us set

$$e := N \cdot X_{uu} \quad (1.16)$$

$$f := N \cdot X_{uv} \quad (1.17)$$

$$g := N \cdot X_{vv} \quad (1.18)$$

so

$$L = - \begin{pmatrix} e & f \\ f & g \end{pmatrix} \quad (1.19)$$

and

$$\det L = eg - f^2.$$

Hence (1.9) specializes to

$$K = \frac{eg - f^2}{EG - F^2}, \quad (1.20)$$

an expression for the Gaussian curvature in local coordinates. We can make this expression even more explicit, using the notion of vector product. Notice that the unit normal vector, N is given by

$$N = \frac{1}{\|X_u \times X_v\|} X_u \times X_v$$

and

$$\|X_u \times X_v\| = \sqrt{\|X_u\|^2 \|X_v\|^2 - (X_u \cdot X_v)^2} = \sqrt{EG - F^2}.$$

Therefore

$$\begin{aligned} e &= N \cdot X_{uu} \\ &= \frac{1}{\sqrt{EG - F^2}} X_{uu} \cdot (X_u \times X_v) \\ &= \frac{1}{\sqrt{EG - F^2}} \det(X_{uu}, X_u, X_v), \end{aligned}$$

This last determinant, is the the determinant of the three by three matrix whose columns are the vectors X_{uu}, X_u and X_v . Replacing the first column by X_{uv} gives a corresponding expression for f , and replacing the first column by X_{vv} gives the expression for g . Substituting into (1.20) gives

$$K = \frac{\det(X_{uu}, X_u, X_v) \det(X_{vv}, X_u, X_v) - \det(X_{uv}, X_u, X_v)^2}{[(X_u \cdot X_u)(X_v \cdot X_v) - (X_u \cdot X_v)^2]^2}. \quad (1.21)$$

This expression is rather complicated for computation by hand, since it involves all those determinants. However a symbolic manipulation program such as maple or mathematica can handle it with ease. Here is the instruction for mathematica, taken from a recent book by Gray (1993), in terms of a function $X[u,v]$ defined in mathematica:

```

gcurvature[X.][u.,v.]:=Simplify[
(Det[D[X[uu,vv],uu,uu],D[X[uu,vv],uu],D[X[uu,vv],vv]]*
Det[D[X[uu,vv],vv,vv],D[X[uu,vv],uu],D[X[uu,vv],vv]]-
Det[D[X[uu,vv],uu,vv],D[X[uu,vv],uu],D[X[uu,vv],vv]]^2)/
(D[X[uu,vv],uu].D[X[uu,vv],uu]*
D[X[uu,vv],vv].D[X[uu,vv],vv]-
D[X[uu,vv],uu].D[X[uu,vv],vv]^2) /. uu->u,vv->v

```

We are now in a position to give two proofs, both correct but both somewhat unsatisfactory of Gauss's *Theorema egregium* which asserts that the Gaussian curvature is an intrinsic property of the metrical character of the surface. However each proof does have its merits.

1.4.1 First proof, using inertial coordinates.

For the first proof, we analyze how the first fundamental form changes when we change coordinates. Suppose we pass from local coordinates u, v to local coordinates u', v' where $u = u(u', v')$, $v = v(u', v')$. Expressing X as a function of u', v' and using the chain rule gives,

$$\begin{aligned}
X_{u'} &= \frac{\partial u}{\partial u'} X_u + \frac{\partial v}{\partial u'} X_v \\
X_{v'} &= \frac{\partial u}{\partial v'} X_u + \frac{\partial v}{\partial v'} X_v \text{ or} \\
(X_{u'}, X_{v'}) &= (X_u, X_v) J \text{ where}
\end{aligned}$$

$$J := \begin{pmatrix} \frac{\partial u}{\partial u'} & \frac{\partial u}{\partial v'} \\ \frac{\partial v}{\partial u'} & \frac{\partial v}{\partial v'} \end{pmatrix}$$

so

$$\begin{aligned}
Q' &= (X_{u'}, X_{v'})^\dagger (X_{u'}, X_{v'}) \\
&= J^\dagger Q J.
\end{aligned}$$

This gives the rule for change of variables of the first fundamental form from the unprimed to the primed coordinate system, and is valid throughout the range where the coordinates are defined. Here J is a matrix valued function of u', v' .

Let us now concentrate attention on a single point, P . The first fundamental form is a symmetric positive definite matrix. By linear algebra, we can always find a matrix R such that $R^\dagger Q(u_P, v_P) R = I$, the two dimensional identity matrix. Here (u_P, v_P) are the coordinates describing P . With no loss of generality we may assume that these coordinates are $(0, 0)$. We can then make the linear change of variables whose $J(0, 0)$ is R , and so find coordinates such that $Q(0, 0) = I$ in this coordinate system. But we can do better. We claim that we can choose coordinates so that

$$Q(0) = I, \quad \frac{\partial Q}{\partial u}(0, 0) = \frac{\partial Q}{\partial v}(0, 0) = 0. \tag{1.22}$$

Indeed, suppose we start with a coordinate system with $Q(0) = I$, and look for a change of coordinates with $J(0) = I$, hoping to determine the second derivatives so that (1.22) holds. Writing $Q' = J^\dagger Q J$ and using Leibniz's formula for the derivative of a product, the equations become

$$\frac{\partial(J + J^\dagger)}{\partial u'}(0) = -\frac{\partial Q}{\partial u}(0) \quad \frac{\partial(J + J^\dagger)}{\partial v'}(0) = -\frac{\partial Q}{\partial v}(0),$$

when we make use of $J(0) = I$. Writing out these equations gives

$$\begin{pmatrix} 2\frac{\partial^2 u}{(\partial u')^2} & \frac{\partial^2 u}{\partial u' \partial v'} + \frac{\partial^2 v}{(\partial u')^2} \\ \frac{\partial^2 u}{\partial u' \partial v'} + \frac{\partial^2 v}{(\partial u')^2} & 2\frac{\partial^2 v}{\partial u' \partial v'} \end{pmatrix} (0) = -\frac{\partial Q}{\partial u}(0)$$

$$\begin{pmatrix} 2\frac{\partial^2 u}{\partial u' \partial v'} & \frac{\partial^2 u}{(\partial v')^2} + \frac{\partial^2 v}{\partial u' \partial v'} \\ \frac{\partial^2 u}{(\partial v')^2} + \frac{\partial^2 v}{\partial u' \partial v'} & 2\frac{\partial^2 v}{(\partial v')^2} \end{pmatrix} (0) = -\frac{\partial Q}{\partial v}(0).$$

The lower right hand corner of the first equation and the upper left hand corner of the second equation determine

$$\frac{\partial^2 v}{\partial u' \partial v'}(0) \quad \text{and} \quad \frac{\partial^2 u}{\partial u' \partial v'}(0).$$

All of the remaining second derivatives are then determined (consistently since Q is a symmetric matrix). We may now choose u and v as functions of u', v' , which vanish at $(0, 0)$ together with all their first partial derivatives, and with the second derivatives as above. For example, we can choose the u and v as homogeneous polynomials in u' and v' with the above partial derivatives. A coordinate system in which (1.22) holds (at a point P having coordinates $(0, 0)$) is called an **inertial coordinate system** based at P . Obviously the collection of all inertial coordinate systems based at P is intrinsically associated to the metric, since the definition depends only on properties of Q in the coordinate system. We now claim the following

Proposition 1 *If u, v is an inertial coordinate system of an embedded surface based at P then the Gaussian curvature is given by*

$$K(P) = F_{uv} - \frac{1}{2}G_{uu} - \frac{1}{2}E_{vv} \quad (1.23)$$

the expression on the right being evaluated at $(0, 0)$.

As the collection of inertial systems is intrinsic, and as (1.23) expresses the curvature in terms of a local expression for the metric in an inertial coordinate system, the proposition implies the *Theorema egregium*.

To prove the proposition, let us first make a rotation and translation in three dimensional space (if necessary) so that $X(P)$ is at the origin and the tangent plane to the surface at P is the x, y plane. The fact that $Q(0) = I$ implies that the vectors $X_u(0), X_v(0)$ form an orthonormal basis of the x, y plane, so

by a further rotation, if necessary, we may assume that X_u is the unit vector in the positive x - direction and by replacing v by $-v$ if necessary, that X_v is the unit vector in the positive y direction. These Euclidean motions we used do not change the value of the determinant of the Weingarten map and so have no effect on the curvature. If we replace v by $-v$, E and G are unchanged and G_{uu} or E_{vv} are also unchanged. Under the change $v \mapsto -v$ F goes to $-F$, but the cross derivative F_{uv} picks up an additional minus sign. So F_{uv} is unchanged.

We have arranged that we need prove (1.23) under the assumptions that

$$X(u, v) = \begin{pmatrix} u + r(u, v) \\ v + s(u, v) \\ f(u, v) \end{pmatrix},$$

where r, s , and f are functions which vanish together with their first derivatives at the origin in u, v space. So far we have only used the property $Q(0) = I$, not the full strength of the definition of an inertial coordinate system. We claim that if the coordinate system is inertial, all the second partials of r and s also vanish at the origin. To see this, observe that

$$\begin{aligned} E &= (1 + r_u)^2 + s_u^2 + f_u^2 \\ F &= r_v + r_u r_v + s_u + s_u s_v + f_u f_v \\ G &= r_v^2 + (1 + s_v)^2 + f_v^2 \quad \text{so} \\ E_u(0) &= 2r_{uu}(0) \\ E_v(0) &= 2r_{uv}(0) \\ F_u(0) &= r_{uv}(0) + s_{uu}(0) \\ F_v(0) &= r_{vv}(0) + s_{uv}(0) \\ G_u(0) &= 2s_{uv}(0) \\ G_v(0) &= 2s_{vv}(0). \end{aligned}$$

The vanishing of all the first partials of E, F , and G at 0 thus implies the vanishing of second partial derivatives of r and s .

By the way, turning this argument around gives us a geometrically intuitive way of constructing inertial coordinates for an embedded surface: At any point P choose orthonormal coordinates in the tangent plane to P and use them to parameterize the surface. (In the preceding notation just choose $x = u$ and $y = v$ as coordinates.)

Now $N(0)$ is just the unit vector in the positive z - direction and so

$$\begin{aligned} e &= f_{uu} \\ f &= f_{uv} \\ g &= f_{vv} \\ \text{so} \\ K &= f_{uu}f_{vv} - f_{uv}^2 \end{aligned}$$

(all the above meant as values at the origin) since $EG - F^2 = 1$ at the origin. On the other hand, taking the partial derivatives of the above expressions for

E, F and G and evaluating at the origin (in particular discarding terms which vanish at the origin) gives

$$\begin{aligned} F_{uv} &= r_{uvv} + s_{uuv} + f_{uu}f_{vv} + f_{uv}^2 \\ E_{vv} &= 2[r_{uvv} + f_{uv}^2] \\ G_{uu} &= 2[s_{uuv} + f_{uv}^2] \end{aligned}$$

when evaluated at $(0, 0)$. So (1.23) holds by direct computation.

1.4.2 Second proof. The Brioschi formula.

Since the Gaussian curvature depends only on the metric, we should be able to find a general formula expressing the Gaussian curvature in terms of a metric, valid in any coordinate system, not just an inertial system. This we shall do by massaging (1.21). The numerator in (1.21) is the difference of products of two determinants. Now $\det B = \det B^\dagger$ so $\det A \det B = \det AB^\dagger$ and we can write the numerator of (1.21) as

$$\det \begin{pmatrix} X_{uu} \cdot X_{vv} & X_{uu} \cdot X_u & X_{uu} \cdot X_v \\ X_u \cdot X_{vv} & X_u \cdot X_u & X_u \cdot X_v \\ X_v \cdot X_{vv} & X_v \cdot X_u & X_v \cdot X_v \end{pmatrix} - \det \begin{pmatrix} X_{uv} \cdot X_{uv} & X_{uv} \cdot X_u & X_{uv} \cdot X_v \\ X_u \cdot X_{iv} & X_u \cdot X_u & X_u \cdot X_v \\ X_v \cdot X_{uv} & X_v \cdot X_u & X_v \cdot X_v \end{pmatrix}.$$

All the terms in these matrices except for the entries in the upper left hand corner of each is either a term of the form E, F , or G or expressible as in terms of derivatives of E, F and G . For example, $X_{uu} \cdot X_u = \frac{1}{2}E_u$ and $F_u = X_{uu} \cdot X_v + X_u \cdot X_{uv}$ so $X_{uu} \cdot X_v = F_u - \frac{1}{2}E_v$ and so on. So if not for the terms in the upper left hand corners, we would already have expressed the Gaussian curvature in terms of E, F and G . So our problem is how to deal with the two terms in the upper left hand corner. Notice that the lower right hand two by two block in these two matrices are the same. So (expanding both matrices along the top row, for example) the difference of the two determinants would be unchanged if we replace the upper left hand term, $X_{uu} \cdot X_{vv}$ in the first matrix by $X_{uu} \cdot X_{vv} - X_{uv} \cdot X_{uv}$ and the upper left hand term in the second matrix by 0. We now show how to express $X_{uu}X_{vv} - X_{uv} \cdot X_{uv}$ in terms of E, F and G and this will then give a proof of the *Theorema egregium*. We have

$$\begin{aligned} X_{uu} \cdot X_{vv} - X_{uv} \cdot X_{uv} &= (X_u \cdot X_{vv})_u - X_u \cdot X_{vvu} \\ &\quad - (X_u \cdot X_{uv})_u + X_u \cdot X_{uvv} \\ &= (X_u \cdot X_{vv})_u - (X_u \cdot X_{uv})_v \\ &= ((X_u \cdot X_v)_u - X_{uv} \cdot X_v)_u - \frac{1}{2}(X_u \cdot X_u)_{vv} \\ &= (X_u \cdot X_v)_{vu} - \frac{1}{2}(X_v \cdot X_v)_{uu} - \frac{1}{2}(X_u \cdot X_u)_{vv} \\ &= -\frac{1}{2}E_{vv} + F_{uv} - \frac{1}{2}G_{uu}. \end{aligned}$$

We thus obtain **Brioschi's formula**

$$\begin{aligned}
 K &= \frac{\det A - \det B}{(EG - F)^2} \quad \text{where} & (1.24) \\
 A &= \begin{pmatrix} \frac{1}{2}E_{vv} + F_{uv} - \frac{1}{2}G_{uu} & \frac{1}{2}E_u & F_u - \frac{1}{2}E_v \\ F_v - \frac{1}{2}G_u & E & F \\ \frac{1}{2}G_v & F & G \end{pmatrix} \\
 B &= \begin{pmatrix} 0 & \frac{1}{2}E_v & \frac{1}{2}G_u \\ \frac{1}{2}E_v & E & F \\ \frac{1}{2}G_u & F & G \end{pmatrix}.
 \end{aligned}$$

Brioschi's formula is not fit for human use but can be fed to machine if necessary. It does give a proof of Gauss' theorem. Notice that if we have coordinates which are inertial at some point, P , then Brioschi's formula reduces to (1.23) since $E = G = 1, F = 0$ and all first partials vanish at P . We will reproduce a mathematica program for Brioschi's formula from Gray at the end of this section.

In case we have **orthogonal coordinates**, a coordinate system in which $F \equiv 0$, Brioschi's formula simplifies and becomes useful: If we set $F = F_u = F_v = 0$ in Brioschi's formula and expand the determinants we get

$$\begin{aligned}
 &\frac{1}{(EG)^2} \left[\left(-\frac{1}{2}E_{vv} - \frac{1}{2}G_{uu} \right) EG + \frac{1}{4}E_u G_u G + \frac{1}{4}E_v G_v E + \frac{1}{4}E_v^2 G + \frac{1}{4}G_u^2 E \right] \\
 &= \left[-\frac{1}{2} \frac{E_{vv}}{EG} + \frac{1}{4} \frac{E_v^2}{E^2 G} + \frac{1}{4} \frac{E_v G_v}{EG^2} \right] + \left[-\frac{1}{2} \frac{G_{uu}}{EG} + \frac{1}{4} \frac{G_u^2}{EG^2} + \frac{1}{4} \frac{E_u G_u}{E^2 G} \right].
 \end{aligned}$$

We claim that the first bracketed expression can be written as

$$-\frac{1}{\sqrt{EG}} \frac{\partial}{\partial v} \left(\frac{1}{\sqrt{G}} \frac{\partial \sqrt{E}}{\partial v} \right).$$

Indeed,

$$\begin{aligned}
 \frac{1}{\sqrt{EG}} \frac{\partial}{\partial v} \left(\frac{1}{\sqrt{G}} \frac{\partial \sqrt{E}}{\partial v} \right) &= \frac{1}{\sqrt{EG}} \left(-\frac{G_v}{2G^{\frac{3}{2}}} \frac{\partial \sqrt{E}}{\partial v} + \frac{1}{\sqrt{G}} \frac{\partial^2 \sqrt{E}}{(\partial v)^2} \right) \\
 &= \frac{1}{\sqrt{EG}} \left(-\frac{E_v G_v}{4G^{\frac{3}{2}} \sqrt{E}} + \frac{1}{2\sqrt{G}} \frac{\partial}{\partial v} (E^{\frac{1}{2}} E_v) \right) \\
 &= \frac{1}{\sqrt{EG}} \left(-\frac{E_v G_v}{4G^{\frac{3}{2}} \sqrt{E}} + \frac{1}{2\sqrt{G}} \left[-\frac{E_v^2}{2E^{\frac{3}{2}}} + \frac{E_v v}{\sqrt{E}} \right] \right) \\
 &= -\frac{G_v E_v}{4G^2 E} - \frac{E_v^2}{4E^2 G} + \frac{E_{vv}}{2EG}.
 \end{aligned}$$

Doing a similar computation for the second bracketed term gives

$$K = \frac{-1}{\sqrt{EG}} \left[\frac{\partial}{\partial u} \left(\frac{1}{\sqrt{E}} \frac{\partial \sqrt{G}}{\partial u} \right) + \frac{\partial}{\partial v} \left(\frac{1}{\sqrt{G}} \frac{\partial \sqrt{E}}{\partial v} \right) \right] \quad (1.25)$$

as the expression for the Gaussian curvature in orthogonal coordinates. We shall give a more direct proof of this formula and of Gauss' theorem egregium once we develop the Cartan calculus.

1.5 Problem set - Surfaces of revolution.

The simplest (non-trivial) case is when $n = 2$ - the study of a curve in the plane. For the case of a curve $X(t) = (x(t), y(t))$ in the plane, we have

$$X'(t) = (x'(t), y'(t)), \quad N(t) = \pm \frac{1}{(x'(t)^2 + y'(t)^2)^{1/2}} (-y'(t), x'(t)),$$

where the \pm reflects the two possible choices of normals. Equation (1.3) says that the one by one matrix L is given by

$$L_{11} = -(N, X'') = \mp \frac{1}{x'^2 + y'^2} (-y'x'' + x'y'').$$

The first fundamental form is the one by one matrix given by

$$Q_{11} = \|X'\|^2.$$

So the curvature is

$$\pm \frac{1}{(x'^2 + y'^2)^{3/2}} (x''y' - y''x').$$

Verify that a straight line has curvature zero and that the curvature of a circle of radius r is $\pm 1/r$ with the plus sign when the normal points outward.

1. What does this formula reduce to in the case that x is used as a parameter, i.e. $x(t) = t, y = f(x)$?

We want to study a surface in three space obtained by rotating a curve, γ , in the x, z plane about the z -axis. Such a surface is called a **surface of revolution**. Surfaces of revolution form one of simplest yet very important classes of surfaces. The sphere, torus, paraboloid, ellipsoid with two equal axes are all surfaces of revolution. Because of modes of production going back to the potter's wheel, the surfaces of many objects of daily life are surfaces of revolution. We will find that the geometry of famous Schwarzschild black hole can be considered as a particular analogue of a surface of revolution in four dimensional space-time.

Let us temporarily assume that the curve γ is given by a function $x = f(z) > 0$ so that we can use z, θ as coordinates, where the surface is given by

$$X(z, \theta) = \begin{pmatrix} f(z) \cos \theta \\ f(z) \sin \theta \\ z \end{pmatrix},$$

and we choose the normal to point away from the z -axis.

2. Find $\nu(z, \theta)$ and show that the Weingarten map is diagonal in the X_z, X_θ basis, in fact

$$N_z = \kappa X_z, \quad N_\theta = \frac{d}{f} X_\theta$$

where κ is the curvature of the curve γ and where d is the distance of the normal vector, ν , from the z -axis. Therefore the Gaussian curvature is given by

$$K = \frac{d\kappa}{f}. \quad (1.26)$$

Check that the Gaussian curvature of a cylinder vanishes and that of a sphere of radius R is $1/R^2$.

Notice that (1.26) makes sense even if we can't use z as a parameter everywhere on γ . Indeed, suppose that γ is a curve in the x, z plane that does not intersect the z -axis, and we construct the corresponding surface of revolution. At points where the tangent to γ is horizontal (parallel to the x -axis) the normal vector to the surface of revolution is vertical, so $d = 0$. Also the Gaussian curvature vanishes, since the Gauss map takes the entire circle of revolution into the north or south pole. So (1.26) is correct at these points. At all other points we can use z as a parameter. But we must watch the sign of κ . Remember that the Gaussian curvature of a surface does not depend on the choice of normal vector, but the curvature of a curve in the plane does. In using (1.26) we must be sure that the sign of κ is the one determined by the normal pointing away from the z -axis.

3. For example, take γ to be a circle of radius r centered at a point at distance $D > r$ from the z -axis, say

$$x = D + r \cos \phi, \quad z = r \sin \phi$$

in terms of an angular parameter, ϕ . The corresponding surface of revolution is a torus. Notice that in using (1.26) we have to take κ as negative on the semicircle closer to the z -axis. So the Gaussian curvature is negative on the "inner" half of the torus and positive on the outer half. Using (1.26) and ϕ, θ as coordinates on the torus, express K as a function on ϕ, θ . Also, express the area element dA in terms of $d\phi d\theta$. Without any computation, show that the total integral of the curvature vanishes, i.e. $\int_T K dA = 0$.

Recall our definitions of $E, F,$ and G given in equations (1.11)-(1.13). In the classical literature, one write the first fundamental form as

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2.$$

the meaning of this expression is as follows: let $t \mapsto (u(t), v(t))$ describe the curve

$$C : t \mapsto X(u(t), v(t))$$

on the surface. Then ds gives the element of arc length of this curve if we substitute $u = u(t), v = v(t)$ into the expression for the first fundamental form. So the first fundamental form describes the intrinsic metrical properties of the surface in terms of the local coordinates. Recall equation (1.25) which says that if u, v is an orthogonal coordinate system then the expression for the Gaussian curvature is

$$K = \frac{-1}{\sqrt{EG}} \left[\frac{\partial}{\partial u} \left(\frac{1}{\sqrt{E}} \frac{\partial \sqrt{G}}{\partial u} \right) + \frac{\partial}{\partial v} \left(\frac{1}{\sqrt{G}} \frac{\partial \sqrt{E}}{\partial v} \right) \right].$$

4. Show that the z, θ coordinates introduced in problem 2 for a surface of revolution is an orthogonal coordinate system, find E and G and verify (??) for this case.

A curve $s \mapsto C(s)$ on a surface is called a **geodesic** if its acceleration, C'' , is everywhere orthogonal to the surface. Notice that

$$\frac{d}{ds}(C'(s), C'(s)) = 2(C''(s), C'(s))$$

and this = 0 if C is a geodesic. The term geodesic refers to a parametrized curve and the above equation shows that the condition to be a geodesic implies that $\|C'(s)\|$ is a constant; i.e that the curve is parametrized by a constant multiple of arc length. If we use a different parameterization, say $s = s(t)$ with dot denoting derivative with respect to t , then the chain rule implies that

$$\dot{C} = C' \dot{s}, \quad \ddot{C} = C'' \dot{s}^2 + C' \ddot{s}.$$

So if use a parameter other than arc length, the projection of the acceleration onto the surface is proportional to the tangent vector if C is a geodesic. In other words, the acceleration is in the plane spanned by the tangent vector to the curve and the normal vector to the surface. Conversely, suppose we start with a curve C which has the property that its acceleration lies in the plane spanned by the tangent vector to the curve and the normal vector to the surface at all points. Let us reparametrize this curve by arc length. Then $(C'(s), C'(s)) \equiv 1$ and hence $(C'', C') \equiv 0$. As we are assuming that \ddot{C} lies in the plane spanned by \dot{C} and the normal vector to the surface at each point of the curve, and that \dot{s} is nowhere 0 we conclude that C , in its arc length parametrization is a geodesic. Standard usage calls a curve which is a geodesic “up to reparametrization” a *pregeodesic*. I don't like this terminology but will live with it.

5. Show that the curves $\theta = \text{const}$ in the terminology of problem 2 are all pregeodesics. Show that the curves $z = \text{const}$. are pregeodesics if and only if z is a critical point of f , (i.e. $f'(s) = 0$).

The general setting for the concept of surfaces of revolution is that of a Riemannian submersion, which will be the subject of Chapter 8.

Chapter 2

Rules of calculus.

2.1 Superalgebras.

A (commutative associative) *superalgebra* is a vector space

$$A = A_{\text{even}} \oplus A_{\text{odd}}$$

with a given direct sum decomposition into even and odd pieces, and a map

$$A \times A \rightarrow A$$

which is bilinear, satisfies the associative law for multiplication, and

$$\begin{aligned} A_{\text{even}} \times A_{\text{even}} &\rightarrow A_{\text{even}} \\ A_{\text{even}} \times A_{\text{odd}} &\rightarrow A_{\text{odd}} \\ A_{\text{odd}} \times A_{\text{even}} &\rightarrow A_{\text{odd}} \\ A_{\text{odd}} \times A_{\text{odd}} &\rightarrow A_{\text{even}} \\ \omega \cdot \sigma &= \sigma \cdot \omega \text{ if either } \omega \text{ or } \sigma \text{ are even,} \\ \omega \cdot \sigma &= -\sigma \cdot \omega \text{ if both } \omega \text{ and } \sigma \text{ are odd.} \end{aligned}$$

We write these last two conditions as

$$\omega \cdot \sigma = (-1)^{\text{deg}\sigma \text{deg}\omega} \sigma \cdot \omega.$$

Here $\text{deg } \tau = 0$ if τ is even, and $\text{deg } \tau = 1 \pmod{2}$ if τ is odd.

2.2 Differential forms.

A *linear* differential form on a manifold, M , is a rule which assigns to each $p \in M$ a linear function on TM_p . So a linear differential form, ω , assigns to each p an element of TM_p^* . We will, as usual, only consider linear differential forms which are smooth.

The superalgebra, $\Omega(M)$ is the superalgebra generated by smooth functions on M (taken as even) and by the linear differential forms, taken as odd.

Multiplication of differential forms is usually denoted by \wedge . The number of differential factors is called the *degree* of the form. So functions have degree zero, linear differential forms have degree one.

In terms of local coordinates, the most general *linear* differential form has an expression as $a_1 dx_1 + \cdots + a_n dx_n$ (where the a_i are functions). Expressions of the form

$$a_{12} dx_1 \wedge dx_2 + a_{13} dx_1 \wedge dx_3 + \cdots + a_{n-1,n} dx_{n-1} \wedge dx_n$$

have degree two (and are even). Notice that the multiplication rules require

$$dx_i \wedge dx_j = -dx_j \wedge dx_i$$

and, in particular, $dx_i \wedge dx_i = 0$. So the most general sum of products of two linear differential forms is a differential form of degree two, and can be brought to the above form, locally, after collections of coefficients. Similarly, the most general differential form of degree $k \leq n$ in n dimensional manifold is a sum, locally, with function coefficients, of expressions of the form

$$dx_{i_1} \wedge \cdots \wedge dx_{i_k}, \quad i_1 < \cdots < i_k.$$

There are $\binom{n}{k}$ such expressions, and they are all even, if k is even, and odd if k is odd.

2.3 The d operator.

There is a linear operator d acting on differential forms called *exterior* differentiation, which is completely determined by the following rules: It satisfies Leibniz' rule in the "super" form

$$d(\omega \cdot \sigma) = (d\omega) \cdot \sigma + (-1)^{\deg \omega} \omega \cdot (d\sigma).$$

On functions it is given by

$$df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n$$

and, finally,

$$d(dx_i) = 0.$$

Since functions and the dx_i generate, this determines d completely. For example, on linear differential forms

$$\omega = a_1 dx_1 + \cdots + a_n dx_n$$

we have

$$\begin{aligned}
d\omega &= da_1 \wedge dx_1 + \cdots + da_n \wedge dx_n \\
&= \left(\frac{\partial a_1}{\partial x_1} dx_1 + \cdots + \frac{\partial a_1}{\partial x_n} dx_n \right) \wedge dx_1 + \cdots \\
&\quad \left(\frac{\partial a_n}{\partial x_1} dx_1 + \cdots + \frac{\partial a_n}{\partial x_n} dx_n \right) \wedge dx_n \\
&= \left(\frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2} \right) dx_1 \wedge dx_2 + \cdots + \left(\frac{\partial a_n}{\partial x_{n-1}} - \frac{\partial a_{n-1}}{\partial x_n} \right) dx_{n-1} \wedge dx_n.
\end{aligned}$$

In particular, equality of mixed derivatives shows that $d^2 f = 0$, and hence that $d^2 \omega = 0$ for any differential form. Hence the rules to remember about d are:

$$\begin{aligned}
d(\omega \cdot \sigma) &= (d\omega) \cdot \sigma + (-1)^{\deg \omega} \omega \cdot (d\sigma) \\
d^2 &= 0 \\
df &= \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n.
\end{aligned}$$

2.4 Derivations.

A linear operator $\ell : A \rightarrow A$ is called an *odd derivation* if, like d , it satisfies

$$\ell : A_{\text{even}} \rightarrow A_{\text{odd}}, \quad \ell : A_{\text{odd}} \rightarrow A_{\text{even}}$$

and

$$\ell(\omega \cdot \sigma) = (\ell\omega) \cdot \sigma + (-1)^{\deg \omega} \omega \cdot \ell\sigma.$$

A linear map $\ell : A \rightarrow A$,

$$\ell : A_{\text{even}} \rightarrow A_{\text{even}}, \quad \ell : A_{\text{odd}} \rightarrow A_{\text{odd}}$$

satisfying

$$\ell(\omega \cdot \sigma) = (\ell\omega) \cdot \sigma + \omega \cdot (\ell\sigma)$$

is called an *even derivation*. So the Leibniz rule for derivations, even or odd, is

$$\ell(\omega \cdot \sigma) = (\ell\omega) \cdot \sigma + (-1)^{\deg \ell \deg \omega} \omega \cdot \ell\sigma.$$

Knowing the action of a derivation on a set of generators of a superalgebra determines it completely. For example, the equations

$$d(x_i) = dx_i, \quad d(dx_i) = 0 \quad \forall i$$

implies that

$$dp = \frac{\partial p}{\partial x_1} dx_1 + \cdots + \frac{\partial p}{\partial x_n} dx_n$$

for any polynomial, and hence determines the value of d on any differential form with polynomial coefficients. The local formula we gave for df where f is any

differentiable function, was just the natural extension (by continuity, if you like) of the above formula for polynomials.

The sum of two even derivations is an even derivation, and the sum of two odd derivations is an odd derivation.

The composition of two derivations will not, in general, be a derivation, but an instructive computation from the definitions shows that the *commutator*

$$[\ell_1, \ell_2] := \ell_1 \circ \ell_2 - (-1)^{\deg \ell_1 \deg \ell_2} \ell_2 \circ \ell_1$$

is again a derivation which is even if both are even or both are odd, and odd if one is even and the other odd.

A derivation followed by a multiplication is again a derivation: specifically, let ℓ be a derivation (even or odd) and let τ be an even or odd element of A . Consider the map

$$\omega \mapsto \tau \ell \omega.$$

We have

$$\begin{aligned} \tau \ell(\omega \sigma) &= (\tau \ell \omega) \cdot \sigma + (-1)^{\deg \ell \deg \omega} \tau \omega \cdot \ell \sigma \\ &= (\tau \ell \omega) \cdot \sigma + (-1)^{(\deg \ell + \deg \tau) \deg \omega} \omega \cdot (\tau \ell \sigma) \end{aligned}$$

so $\omega \mapsto \tau \ell \omega$ is a derivation whose degree is

$$\deg \tau + \deg \ell.$$

2.5 Pullback.

Let $\phi : M \rightarrow N$ be a smooth map. Then the pullback map ϕ^* is a linear map that sends differential forms on N to differential forms on M and satisfies

$$\begin{aligned} \phi^*(\omega \wedge \sigma) &= \phi^* \omega \wedge \phi^* \sigma \\ \phi^* d\omega &= d\phi^* \omega \\ (\phi^* f) &= f \circ \phi. \end{aligned}$$

The first two equations imply that ϕ^* is completely determined by what it does on functions. The last equation says that on functions, ϕ^* is given by “substitution”: In terms of local coordinates on M and on N ϕ is given by

$$\begin{aligned} \phi(x^1, \dots, x^m) &= (y^1, \dots, y^n) \\ y^i &= \phi^i(x^1, \dots, x^m) \quad i = 1, \dots, n \end{aligned}$$

where the ϕ_i are smooth functions. The local expression for the pullback of a function $f(y^1, \dots, y^n)$ is to substitute ϕ^i for the y^i 's as into the expression for f so as to obtain a function of the x 's.

It is important to observe that the pull back on differential forms is defined for any smooth map, not merely for diffeomorphisms. This is the great advantage of the calculus of differential forms.

2.6 Chain rule.

Suppose that $\psi : N \rightarrow P$ is a smooth map so that the composition

$$\phi \circ \psi : M \rightarrow P$$

is again smooth. Then the *chain rule* says

$$(\phi \circ \psi)^* = \psi^* \circ \phi^*.$$

On functions this is essentially a tautology - it is the associativity of composition: $f \circ (\phi \circ \psi) = (f \circ \phi) \circ \psi$. But since pull-back is completely determined by what it does on functions, the chain rule applies to differential forms of any degree.

2.7 Lie derivative.

Let ϕ_t be a one parameter group of transformations of M . If ω is a differential form, we get a family of differential forms, $\phi_t^* \omega$ depending differentiably on t , and so we can take the derivative at $t = 0$:

$$\frac{d}{dt} (\phi_t^* \omega) |_{t=0} = \lim_{t \rightarrow 0} \frac{1}{t} [\phi_t^* \omega - \omega].$$

Since $\phi_t^*(\omega \wedge \sigma) = \phi_t^* \omega \wedge \phi_t^* \sigma$ it follows from the Leibniz argument that

$$\ell_\phi : \omega \mapsto \frac{d}{dt} (\phi_t^* \omega) |_{t=0}$$

is an even derivation. We want a formula for this derivation.

Notice that since $\phi_t^* d = d\phi_t^*$ for all t , it follows by differentiation that

$$\ell_\phi d = d\ell_\phi$$

and hence the formula for ℓ_ϕ is completely determined by how it acts on functions.

Let X be the vector field generating ϕ_t . Recall that the geometrical significance of this vector field is as follows: If we fix a point x , then

$$t \mapsto \phi_t(x)$$

is a curve which passes through the point x at $t = 0$. The tangent to this curve at $t = 0$ is the vector $X(x)$. In terms of local coordinates, X has coordinates $X = (X^1, \dots, X^n)$ where $X^i(x)$ is the derivative of $\phi^i(t, x^1, \dots, x^n)$ with respect to t at $t = 0$. The chain rule then gives, for any function f ,

$$\begin{aligned} \ell_\phi f &= \frac{d}{dt} f(\phi^1(t, x^1, \dots, x^n), \dots, \phi^n(t, x^1, \dots, x^n)) |_{t=0} \\ &= X^1 \frac{\partial f}{\partial x_1} + \dots + X^n \frac{\partial f}{\partial x_n}. \end{aligned}$$

For this reason we use the notation

$$X = X^1 \frac{\partial}{\partial x_1} + \cdots + X^n \frac{\partial}{\partial x_n}$$

so that the differential operator

$$f \mapsto Xf$$

gives the action of ℓ_ϕ on functions.

As we mentioned, this action of ℓ_ϕ on functions determines it completely. In particular, ℓ_ϕ depends only on the vector field X , so we may write

$$\ell_\phi = L_X$$

where L_X is the even derivation determined by

$$L_X f = Xf, \quad L_X d = dL_X.$$

2.8 Weil's formula.

But we want a more explicit formula L_X . For this it is useful to introduce an odd derivation associated to X called the *interior product* and denoted by $i(X)$. It is defined as follows: First consider the case where

$$X = \frac{\partial}{\partial x_j}$$

and define its interior product by

$$i\left(\frac{\partial}{\partial x_j}\right) f = 0$$

for all functions while

$$i\left(\frac{\partial}{\partial x_j}\right) dx_k = 0, \quad k \neq j$$

and

$$i\left(\frac{\partial}{\partial x_j}\right) dx_j = 1.$$

The fact that it is a derivation then gives an easy rule for calculating $i(\partial/\partial x_j)$ when applied to any differential form: Write the differential form as

$$\omega + dx_j \wedge \sigma$$

where the expressions for ω and σ do not involve dx_j . Then

$$i\left(\frac{\partial}{\partial x_j}\right) [\omega + dx_j \wedge \sigma] = \sigma.$$

The operator

$$X^j i \left(\frac{\partial}{\partial x_j} \right)$$

which means first apply $i(\partial/\partial x_j)$ and then multiply by the function X^j is again an odd derivation, and so we can make the definition

$$i(X) := X^1 i \left(\frac{\partial}{\partial x_1} \right) + \cdots + X^n i \left(\frac{\partial}{\partial x_n} \right). \quad (2.1)$$

It is easy to check that this does not depend on the local coordinate system used.

Notice that we can write

$$Xf = i(X)df.$$

In particular we have

$$\begin{aligned} L_X dx_j &= dL_X x_j \\ &= dX_j \\ &= di(X)dx_j. \end{aligned}$$

We can combine these two formulas as follows: Since $i(X)f = 0$ for any function f we have

$$L_X f = di(X)f + i(X)df.$$

Since $ddx_j = 0$ we have

$$L_X dx_j = di(X)dx_j + i(X)ddx_j.$$

Hence

$$L_X = di(X) + i(X)d = [d, i(X)] \quad (2.2)$$

when applied to functions or to the forms dx_j . But the right hand side of the preceding equation is an even derivation, being the commutator of two odd derivations. So if the left and right hand side agree on functions and on the differential forms dx_j they agree everywhere. This equation, (2.2), known as *Weil's formula*, is a basic formula in differential calculus.

We can use the interior product to consider differential forms of degree k as k -multilinear functions on the tangent space at each point. To illustrate, let σ be a differential form of degree two. Then for any vector field, X , $i(X)\sigma$ is a linear differential form, and hence can be evaluated on any vector field, Y to produce a function. So we define

$$\sigma(X, Y) := [i(X)\sigma](Y).$$

We can use this to express exterior derivative in terms of ordinary derivative and Lie bracket: If θ is a linear differential form, we have

$$\begin{aligned} d\theta(X, Y) &= [i(X)d\theta](Y) \\ i(X)d\theta &= L_X\theta - d(i(X)\theta) \\ d(i(X)\theta)(Y) &= Y[\theta(X)] \\ [L_X\theta](Y) &= L_X[\theta(Y)] - \theta(L_X(Y)) \\ &= X[\theta(Y)] - \theta([X, Y]) \end{aligned}$$

where we have introduced the notation $L_X Y =: [X, Y]$ which is legitimate since on functions we have

$$(L_X Y)f = L_X(Yf) - YL_X f = X(Yf) - Y(Xf)$$

so $L_X Y$ as an operator on functions is exactly the commutator of X and Y . (See below for a more detailed geometrical interpretation of $L_X Y$.) Putting the previous pieces together gives

$$d\theta(X, Y) = X\theta(Y) - Y\theta(X) - \theta([X, Y]), \quad (2.3)$$

with similar expressions for differential forms of higher degree.

2.9 Integration.

Let

$$\omega = f dx_1 \wedge \cdots \wedge dx_n$$

be a form of degree n on \mathbf{R}^n . (Recall that the most general differential form of degree n is an expression of this type.) Then its integral is defined by

$$\int_M \omega := \int_M f dx_1 \cdots dx_n$$

where M is any (measurable) subset. This, of course is subject to the condition that the right hand side converges if M is unbounded. There is a lot of hidden subtlety built into this definition having to do with the notion of orientation. But for the moment this is a good working definition.

The *change of variables formula* says that if $\phi : M \rightarrow \mathbf{R}^n$ is a smooth differentiable map which is one to one whose Jacobian determinant is everywhere positive, then

$$\int_M \phi^* \omega = \int_{\phi(M)} \omega.$$

2.10 Stokes theorem.

Let U be a region in \mathbf{R}^n with a chosen orientation and smooth boundary. We then orient the boundary according to the rule that an outward pointing normal

vector, together with the a positive frame on the boundary give a positive frame in \mathbf{R}^n . If σ is an $(n - 1)$ -form, then

$$\int_{\partial U} \sigma = \int_U d\sigma.$$

A manifold is called *orientable* if we can choose an atlas consisting of charts such that the Jacobian of the transition maps $\phi_\alpha \circ \phi_\beta^{-1}$ is always positive. Such a choice of an atlas is called an orientation. (Not all manifolds are orientable.) If we have chosen an orientation, then relative to the charts of our orientation, the transition laws for an n -form (where $n = \dim M$) and for a density are the same. In other words, given an orientation, we can identify densities with n -forms and n -form with densities. Thus we may integrate n -forms. The change of variables formula then holds for orientation preserving diffeomorphisms as does Stokes theorem.

2.11 Lie derivatives of vector fields.

Let Y be a vector field and ϕ_t a one parameter group of transformations whose “infinitesimal generator” is some other vector field X . We can consider the “pulled back” vector field ϕ_t^*Y defined by

$$\phi_t^*Y(x) = d\phi_{-t}\{Y(\phi_tx)\}.$$

In words, we evaluate the vector field Y at the point $\phi_t(x)$, obtaining a tangent vector at $\phi_t(x)$, and then apply the differential of the (inverse) map ϕ_{-t} to obtain a tangent vector at x .

If we differentiate the one parameter family of vector fields ϕ_t^*Y with respect to t and set $t = 0$ we get a vector field which we denote by L_XY :

$$L_XY := \frac{d}{dt}\phi_t^*Y|_{t=0}.$$

If ω is a linear differential form, then we may compute $i(Y)\omega$ which is a function whose value at any point is obtained by evaluating the linear function $\omega(x)$ on the tangent vector $Y(x)$. Thus

$$i(\phi_t^*Y)\phi_t^*\omega(x) = \langle d\phi_t^*\omega(\phi_tx), d\phi_{-t}Y(\phi_tx) \rangle = \{i(Y)\omega\}(\phi_tx).$$

In other words,

$$\phi_t^*\{i(Y)\omega\} = i(\phi_t^*Y)\phi_t^*\omega.$$

We have verified this when ω is a differential form of degree one. It is trivially true when ω is a differential form of degree zero, i.e. a function, since then both sides are zero. But then, by the derivation property, we conclude that it is true for forms of all degrees. We may rewrite the result in shorthand form as

$$\phi_t^* \circ i(Y) = i(\phi_t^*Y) \circ \phi_t^*.$$

Since $\phi_t^* d = d\phi_t^*$ we conclude from Weil's formula that

$$\phi_t^* \circ L_Y = L_{\phi_t^* Y} \circ \phi_t^*.$$

Until now the subscript t was superfluous, the formulas being true for any fixed diffeomorphism. Now we differentiate the preceding equations with respect to t and set $t = 0$. We obtain, using Leibniz's rule,

$$L_X \circ i(Y) = i(L_X Y) + i(Y) \circ L_X$$

and

$$L_X \circ L_Y = L_{L_X Y} + L_Y \circ L_X.$$

This last equation says that Lie derivative (on forms) with respect to the vector field $L_X Y$ is just the commutator of L_X with L_Y :

$$L_{L_X Y} = [L_X, L_Y].$$

For this reason we write

$$[X, Y] := L_X Y$$

and call it the Lie bracket (or commutator) of the two vector fields X and Y . The equation for interior product can then be written as

$$i([X, Y]) = [L_X, i(Y)].$$

The Lie bracket is antisymmetric in X and Y . We may multiply Y by a function g to obtain a new vector field gY . From the definitions we have

$$\phi_t^*(gY) = (\phi_t^* g)\phi_t^* Y.$$

Differentiating at $t = 0$ and using Leibniz's rule we get

$$[X, gY] = (Xg)Y + g[X, Y] \tag{2.4}$$

where we use the alternative notation Xg for $L_X g$. The antisymmetry then implies that for any differentiable function f we have

$$[fX, Y] = -(Yf)X + f[X, Y]. \tag{2.5}$$

From both this equation and from Weil's formula (applied to differential forms of degree greater than zero) we see that the Lie derivative with respect to X at a point x depends on more than the value of the vector field X at x .

2.12 Jacobi's identity.

From the fact that $[X, Y]$ acts as the commutator of X and Y it follows that for any three vector fields X, Y and Z we have

$$[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0.$$

This is known as **Jacobi's identity**. We can also derive it from the fact that $[Y, Z]$ is a natural operation and hence for any one parameter group ϕ_t of diffeomorphisms we have

$$\phi_t^*([Y, Z]) = [\phi_t^*Y, \phi_t^*Z].$$

If X is the infinitesimal generator of ϕ_t then differentiating the preceding equation with respect to t at $t = 0$ gives

$$[X, [Y, Z]] = [[X, Y], Z] + [Y, [X, Z]].$$

In other words, X acts as a derivation of the "multiplication" given by Lie bracket. This is just Jacobi's identity when we use the antisymmetry of the bracket. In the future we will have occasion to take cyclic sums such as those which arise on the left of Jacobi's identity. So if F is a function of three vector fields (or of three elements of any set) with values in some vector space (for example in the space of vector fields) we will define the cyclic sum $\mathcal{Cyc} F$ by

$$\mathcal{Cyc} F(X, Y, Z) := F(X, Y, Z) + F(Y, Z, X) + F(Z, X, Y).$$

With this definition Jacobi's identity becomes

$$\mathcal{Cyc} [X, [Y, Z]] = 0. \tag{2.6}$$

Exercises

2.13 Left invariant forms.

Let G be a group and M be a set. A **left action** of G on M consists of a map

$$\phi : G \times M \rightarrow M$$

satisfying the conditions

$$\phi(a, \phi(b, m)) = \phi(ab, m)$$

(an associativity law) and

$$\phi(e, m) = m, \quad \forall m \in M$$

where e is the identity element of the group. When there is no risk of confusion we will write am for $\phi(a, m)$. (But in much of the beginning of the following exercises there *will* be a risk of confusion since there will be several different actions of the same group G on the set M). We think of an action as assigning to each element $a \in G$ a transformation, ϕ_a , of M :

$$\phi_a : M \rightarrow M, \quad \phi_a : m \mapsto \phi(a, m).$$

So we also use the notation

$$\phi_a m = \phi(a, m).$$

For example, we may take M to be the group G itself and let the action be left multiplication, L , so

$$L(a, m) = am.$$

We will write

$$L_a : G \rightarrow G, \quad L_a m = am.$$

We may also consider the (left) action of right multiplication:

$$R : G \times G \rightarrow G, \quad R(a, m) = ma^{-1}.$$

(The inverse is needed to get the order right in $R(a, R(b, m)) = R(ab, m)$.) So we will write

$$R_a : G \rightarrow G, \quad R_a m = ma^{-1}.$$

We will be interested in the case that G is a Lie group, which means that G is a manifold and the multiplication map $G \times G \rightarrow G$ and the inverse map $G \rightarrow G$, $a \mapsto a^{-1}$ are both smooth maps. Then the differential, $(dL_a)_m$ maps the tangent space to G at m , to the tangent space to G at am :

$$dL_a : TG_m \rightarrow TG_{am}$$

and similarly

$$dR_a : TG_m \rightarrow TG_{ma}.$$

In particular,

$$dL_{a^{-1}} : TG_a \rightarrow TG_e.$$

Let $G = Gl(n)$ be the group of all invertible $n \times n$ matrices. It is an open subset (hence a submanifold) of the n^2 dimensional space $\text{Mat}(n)$ of all $n \times n$ matrices. We can think of the tautological map which sends every $A \in G$ into itself thought of as an element of $\text{Mat}(n)$ as a matrix valued function on G . Put another way, A is a matrix of functions on G , each of the matrix entries A_{ij} of A is a function on G . Hence $dA = (dA_{ij})$ is a matrix of differential forms (or, we may say, a matrix valued differential form). So we may consider

$$A^{-1}dA$$

which is also a matrix valued differential form on G . Let B be a fixed element of G .

1. Show that

$$L_B^*(A^{-1}dA) = A^{-1}dA. \quad (2.7)$$

So each of the entries of $A^{-1}dA$ is *left invariant*.

2. Show that

$$R_B^*(A^{-1}dA) = B(A^{-1}dA)B^{-1}. \quad (2.8)$$

So the entries of $A^{-1}dA$ are not right invariant (in general), but (2.8) shows how they are transformed into one another by right multiplication.

For any two matrix valued differential forms $R = (R_{ij})$ and $S = (S_{ij})$ define their matrix exterior product $R \wedge S$ by the usual formula for matrix product, but with exterior multiplication of the entries instead of ordinary multiplication, so

$$(R \wedge S)_{ik} := \sum_j R_{ij} \wedge S_{jk}.$$

Also, if $R = (R_{ij})$ is a matrix valued differential form, define dR by applying d to each of the entries. So

$$(dR)_{ij} := (dR_{ij}).$$

Finally, if $\psi : X \rightarrow Y$ is a smooth map and $R = (R_{ij})$ is a matrix valued form on Y then we define its pullback by pulling back each of the entries:

$$(\psi^* R)_{ij} := (\psi^* R_{ij}).$$

2.14 The Maurer Cartan equations.

3. In elementary calculus we have the formula $d(1/x) = -dx/x^2$. What is the generalization of this formula for the matrix function A^{-1} . In other words, what is the formula for $d(A^{-1})$?

4. Show that if we set $\omega = A^{-1}dA$ then

$$d\omega + \omega \wedge \omega = 0. \tag{2.9}$$

Here is another way of thinking about $A^{-1}dA$: Since $G = Gl(n)$ is an open subset of the vector space $\text{Mat}(n)$, we may identify the tangent space TG_A with the vector space $\text{Mat}(n)$. That is we have an isomorphism between TG_A and $\text{Mat}(n)$. If you think about it for a minute, it is the form dA which effects this isomorphism at every point. On the other hand, left multiplication by A^{-1} is a linear map. Under this identification, the differential of a linear map L looks just like L . So in terms of this identification, $A^{-1}dA$, when evaluated at the tangent space TG_A is just the isomorphism $dL_A^{-1} : TG_A \rightarrow TG_I$ where I is the identity matrix.

2.15 Restriction to a subgroup

Let H be a Lie subgroup of G . This means that H is a subgroup of G and it is also a submanifold. In other words we have an embedding

$$\iota : H \rightarrow G$$

which is a(n injective) group homomorphism. Let

$$\mathfrak{h} = TH_I$$

denote the tangent space to H at the identity element.

5. Conclude from the preceding discussion that if we now set

$$\omega = \iota^*(A^{-1}dA)$$

then ω takes values in \mathfrak{h} . In other words, when we evaluate ω on any tangent vector at any point of H we get a matrix belonging to the subspace \mathfrak{h} .

6. Show that on a group, the only transformations which commute with all the right multiplications, R_b , $b \in G$, are the left multiplications, L_a .

For any vector $\xi \in TH_I$, define the vector field X by

$$X(A) = dR_{A^{-1}}\xi.$$

(Recall that $R_{A^{-1}}$ is right multiplication by A and so sends I into A .) For example, if we take H to be the full group $G = Gl(n)$ and identify the tangent space at every point with $Mat(n)$ then the above definition becomes

$$X(A) = \xi A.$$

By construction, the vector field X is right invariant, i.e. is invariant under all the diffeomorphisms R_B .

7. Conclude that the flow generated by X is left multiplication by a one parameter subgroup. Also conclude that in the case $H = Gl(n)$ the flow generated by X is left multiplication by the one parameter group

$$\exp t\xi.$$

Finally conclude that for a general subgroup H , if $\xi \in \mathfrak{h}$ then all the $\exp t\xi$ lie in H .

8. What is the space \mathfrak{h} in the case that H is the group of Euclidean motions in three dimensional space, thought of as the set of all four by four matrices of the form

$$\begin{pmatrix} A & v \\ 0 & 1 \end{pmatrix}, \quad AA^\dagger = I, \quad v \in \mathbf{R}^3?$$

2.16 Frames.

Let V be an n dimensional vector space. Recall that **frame** on V is, by definition, an isomorphism $\mathbf{f} : \mathbf{R}^n \rightarrow V$. Giving \mathbf{f} is the same as giving each of the

vectors $f_i = \mathbf{f}(\delta_i)$ where the δ_i range over the standard basis of \mathbf{R}^n . So giving a frame is the same as giving an ordered basis of V and we will sometimes write

$$\mathbf{f} = (f_1, \dots, f_n).$$

If $A \in Gl(n)$ then A is an isomorphism of \mathbf{R}^n with itself, so $\mathbf{f} \circ A^{-1}$ is another frame. So we get an action, $R : Gl(n) \times \mathbf{F} \rightarrow \mathbf{F}$ where $\mathbf{F} = \mathbf{F}(V)$ denotes the space of all frames:

$$R(A, \mathbf{f}) = \mathbf{f} \circ A^{-1}. \quad (2.10)$$

If \mathbf{f} and \mathbf{g} are two frames, then $\mathbf{g}^{-1} \circ \mathbf{f} = M$ is an isomorphism of \mathbf{R}^n with itself, i.e. a matrix. So given any two frames, \mathbf{f} and \mathbf{g} , there is a unique $M \in Gl(n)$ so that $\mathbf{g} = \mathbf{f} \circ M^{-1}$. Once we fix an \mathbf{f} , we can use this fact to identify \mathbf{F} with $Gl(n)$, but the identification depends on the choice of \mathbf{f} . But in any event the (non-unique) identification shows that \mathbf{F} is a manifold and that (2.10) defines an action of $Gl(n)$ on \mathbf{F} . Each of the f_i (the i -th basis vector in the frame) can be thought of as a V valued function on \mathbf{F} . So we may write

$$df_j = \sum \omega_{ij} f_i \quad (2.11)$$

where the ω_{ij} are ordinary (number valued) linear differential forms on \mathbf{F} . We think of this equation as giving the expansion of an infinitesimal change in f_j in terms of the basis $\mathbf{f} = (f_1, \dots, f_n)$. If we use the “row” representation of \mathbf{f} as above, we can write these equations as

$$d\mathbf{f} = \mathbf{f}\omega \quad (2.12)$$

where $\omega = (\omega_{ij})$.

9. Show that the ω defined by (2.12) satisfies

$$R_B^* \omega = B\omega B^{-1}. \quad (2.13)$$

To see the relation with what went on before, notice that we *could* take $V = \mathbf{R}^n$ itself. Then \mathbf{f} is just an invertible matrix, A and (2.12) becomes our old equation $\omega = A^{-1}dA$. So (2.13) reduces to (2.8).

If we take the exterior derivative of (2.12) we get

$$0 = d(d\mathbf{f}) = d\mathbf{f} \wedge \omega + \mathbf{f}d\omega = \mathbf{f}(\omega \wedge \omega + d\omega)$$

from which we conclude

$$d\omega + \omega \wedge \omega = 0. \quad (2.14)$$

2.17 Euclidean frames.

We specialize to the case where $V = \mathbf{R}^n, n = d + 1$ so that the set of frames becomes identified with the group $Gl(n)$ and restrict to the subgroup, H , of

Euclidean motions which consist of all $n \times m$ matrices of the form

$$\begin{pmatrix} A & v \\ 0 & 1 \end{pmatrix}, \quad A \in O(d), \quad v \in \mathbf{R}^d.$$

Such a matrix, when applied to a vector

$$\begin{pmatrix} w \\ 1 \end{pmatrix}$$

sends it into the vector

$$\begin{pmatrix} Aw + v \\ 1 \end{pmatrix}$$

and $Aw + v$ is the orthogonal transformation A applied to w followed by the translation by v . The corresponding *Euclidean frames* (consisting of the columns of the elements of H) are thus defined to be the frames of the form

$$f_i = \begin{pmatrix} e_i \\ 0 \end{pmatrix}, \quad i = 1, \dots, d,$$

where the e_i form an orthonormal basis of \mathbf{R}^d and

$$f_n = \begin{pmatrix} v \\ 1 \end{pmatrix},$$

where $v \in \mathbf{R}^d$ is an arbitrary vector. The idea is that v represents a choice of origin in d dimensional space and $\mathbf{e} = (e_1, \dots, e_d)$ is an orthonormal basis. We can write this in shorthand notation as

$$\mathbf{f} = \begin{pmatrix} \mathbf{e} & v \\ 0 & 1 \end{pmatrix}.$$

If ι denotes the embedding of H into G , we know from the exercise 5 that

$$\iota^*\omega = \begin{pmatrix} \Omega & \theta \\ 0 & 0 \end{pmatrix},$$

where

$$\Omega_{ij} = -\Omega_{ji}.$$

So the pull back of (2.12) becomes

$$d \begin{pmatrix} \mathbf{e} & v \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{e}\Omega & \mathbf{e}\theta \\ 0 & 0 \end{pmatrix} \quad (2.15)$$

or, in more expanded notation,

$$de_j = \sum_i \Omega_{ij} e_i, \quad dv = \sum_i \theta_i e_i.$$

Let (\cdot, \cdot) denote the Euclidean scalar product. Then we can write

$$\theta_i = (dv, e_i) \quad (2.16)$$

and

$$(de_j, e_i) = \Omega_{ij}.$$

If we set

$$\Theta = -\Omega$$

this becomes

$$(de_i, e_j) = \Theta_{ij}. \quad (2.17)$$

Then (2.14) becomes

$$d\theta = \Theta \wedge \theta, \quad d\Theta = \Theta \wedge \Theta. \quad (2.18)$$

Or, in more expanded notation,

$$d\theta_i = \sum_j \Theta_{ij} \wedge \theta_j, \quad d\Theta_{ik} = \sum_j \Theta_{ij} \wedge \Theta_{jk}. \quad (2.19)$$

Equations (2.16)-(2.18) or (2.19) are known as the **structure equations of Euclidean geometry**.

2.18 Frames adapted to a submanifold.

Let M be a k dimensional submanifold of \mathbf{R}^d . This determines a submanifold of the manifold, H , of all Euclidean frames by the following requirements:

- i) $v \in M$ and
- ii) $e_i \in TM_v$ for $i \leq k$. We will usually write m instead of v to emphasize the first requirement - that the frames be based at points of M . The second requirement says that the first k vectors in the frame based at m be tangent to M (and hence that the last $n - k$ vectors in the frame are normal to M). We will denote this manifold by $\mathcal{O}(M)$. It has dimension

$$k + \frac{k(k-1)}{2} + \frac{(d-k-1)(d-k)}{2}.$$

The first term comes from the point m varying on M , the second is the dimension of the orthogonal group $O(k)$ corresponding to the choices of the first k vectors in the frame, and the third term is $\dim O(d-k)$ correspond to the last $(n-k)$ vectors. We have an embedding of $\mathcal{O}(M)$ into H , and hence the forms θ and Θ pull back to $\mathcal{O}(M)$. As we are running out of letters, we will continue to denote these pull backs by the same letters. So the pulled back forms satisfy the same structure equations (2.16)-(2.18) or (2.19) as above, but they are supplemented by

$$\theta_i = 0, \quad \forall i > k. \quad (2.20)$$

2.19 Curves and surfaces - their structure equations.

We will be particularly interested in curves and surfaces in three dimensional Euclidean space. For a curve, C , the manifold of frames is two dimensional, and we have

$$dC = \theta_1 e_1 \quad (2.21)$$

$$de_1 = \Theta_{12} e_2 + \Theta_{13} e_3 \quad (2.22)$$

$$de_2 = \Theta_{21} e_1 + \Theta_{23} e_3 \quad (2.23)$$

$$de_3 = \Theta_{31} e_1 + \Theta_{32} e_2. \quad (2.24)$$

One can visualize the manifold of frames as a sort of tube: about each point of the curve there is a circle in the plane normal to the tangent line corresponding the possible choices of e_2 .

For the case of a surface the manifold of frames is three dimensional: we can think of it as a union of circles each centered at a point of S and in the plane tangent to S at that point. Then equation (2.21) is replaced by

$$dX = \theta_1 e_1 + \theta_2 e_2 \quad (2.25)$$

but otherwise the equations are as above, including the structure equations (2.19). These become

$$d\theta_1 = \Theta_{12} \wedge \theta_2 \quad (2.26)$$

$$d\theta_2 = -\Theta_{12} \wedge \theta_1 \quad (2.27)$$

$$0 = \Theta_{31} \wedge \theta_1 + \Theta_{32} \wedge \theta_2 \quad (2.28)$$

$$d\Theta_{12} = \Theta_{13} \wedge \Theta_{32} \quad (2.29)$$

$$d\Theta_{13} = \Theta_{12} \wedge \Theta_{23} \quad (2.30)$$

$$d\Theta_{23} = \Theta_{21} \wedge \Theta_{13} \quad (2.31)$$

Equation (2.29) is known as Gauss' equation, and equations (2.30) and (2.31) are known as the Codazzi-Mainardi equations.

2.20 The sphere as an example.

In computations with local coordinates, we may find it convenient to use a "cross-section" of the manifold of frames, that is a map which assigns to each point of neighborhood on the surface a preferred frame. If we are given a parametrization $m = m(u, v)$ of the surface, one way of choosing such a cross-section is to apply the Gram-Schmidt orthogonalization procedure to the tangent vector fields m_u and m_v , and take into account the chosen orientation.

For example, consider the sphere of radius R . We can parameterize the sphere with the north and south poles (and one longitudinal semi-circle) removed

by the $(u, v) \in (0, 2\pi) \times (0, \pi)$ by $X = X(u, v)$ where

$$X(u, v) = \begin{pmatrix} R \cos u \sin v \\ R \sin u \sin v \\ R \cos v \end{pmatrix}.$$

Here v denotes the angular distance from the north pole, so the excluded value $v = 0$ corresponds to the north pole and the excluded value $v = \pi$ corresponds to the south pole. Each constant value of v between 0 and π is a circle of latitude with the equator given by $v = \frac{\pi}{2}$. The parameter u describes the longitude from the excluded semi-circle.

In any frame adapted to a surface in \mathbf{R}^3 , the third vector e_3 is normal to the surface at the base point of the frame. There are two such choices at each base point. In our sphere example let us choose the outward pointing normal, which at the point $m(u, v)$ is

$$e_3(m(u, v)) = \begin{pmatrix} \cos u \sin v \\ \sin u \sin v \\ \cos v \end{pmatrix}.$$

We will write the left hand side of this equation as $e_3(u, v)$. The coordinates u, v are orthogonal, i.e. X_u and X_v are orthogonal at every point, so the orthonormalization procedure amounts only to normalization: Replace each of these vectors by the unit vectors pointing in the same direction at each point. So we get

$$e_1(u, v) = \begin{pmatrix} -\sin u \\ \cos u \\ 0 \end{pmatrix}, \quad e_2(u, v) = \begin{pmatrix} \cos u \cos v \\ \sin u \cos v \\ -\sin v \end{pmatrix}.$$

We thus obtain a map ψ from $(0, 2\pi) \times (0, \pi)$ to the manifold of frames,

$$\psi(u, v) = (X(u, v), e_1(u, v), e_2(u, v), e_3(u, v)).$$

Since $X_u \cdot e_1 = R \sin v$ and $X_v \cdot e_2 = R$ we have

$$dX(u, v) = (R \sin v du) e_1(u, v) + (R dv) e_2(u, v).$$

Thus we see from (2.25) that

$$\psi^* \theta_1 = R \sin v du, \quad \psi^* \theta_2 = R dv$$

and hence that

$$\psi^*(\theta_1 \wedge \theta_2) = R^2 \sin v du \wedge dv.$$

Now $R^2 \sin v du dv$ is just the area element of the sphere expressed in u, v coordinates. The choice of e_1, e_2 determines an orientation of the tangent space to the sphere at the point $X(u, v)$ and so $\psi^*(\theta_1 \wedge \theta_2)$ is the pull-back of the corresponding oriented area form.

10. Compute $\psi^*\Theta_{12}$, $\psi^*\Theta_{13}$, and $\psi^*\Theta_{23}$ and verify that

$$\psi^*(d\Theta_{12}) = -\psi^*(K)\psi^*(\theta_1 \wedge \theta_2)$$

where $K = 1/R^2$ is the curvature of the sphere.

We will generalize this equation to an arbitrary surface in \mathbf{R}^3 in section ??.

2.21 Ribbons

The idea here is to study a curve on a surface, or rather a curve with an “infinitesimal” neighborhood of a surface along it. So let C be a curve and $\mathcal{O}(C)$ its associated two dimensional manifold of frames. We have a projection $\pi : \mathcal{O}(C) \rightarrow C$ sending every frame into its origin. By a **ribbon** based on C we mean a section $n : C \rightarrow \mathcal{O}(C)$, so n assigns a unique frame to each point of the curve in a smooth way. We will only be considering curves with non-vanishing tangent vector everywhere. With no loss of generality we may assume that we have parametrized the curve by arc length, and the choice of e_1 determines an orientation of the curve, so $\theta = ds$. The choice of e_2 at every point then determines e_3 up to a \pm sign. So a good way to visualize s is to think of a rigid metal ribbon determined by the curve and the vectors e_2 perpendicular to the curve (determined by n) at each point. The forms Θ_{ij} all pull back under n to function multiples of ds :

$$n^*\Theta_{12} = kds, \quad n^*\Theta_{23} = -\tau ds, \quad n^*\Theta_{13} = wds \quad (2.32)$$

where k, τ and w are functions of s . We can write equations (2.21)- (2.24) above as

$$\frac{dC}{ds} = e_1,$$

and

$$\frac{de_1}{ds} = ke_2 + we_3, \quad \frac{de_2}{ds} = -ke_1 - \tau e_3, \quad \frac{de_3}{ds} = -we_1 + \tau e_3. \quad (2.33)$$

For later applications we will sometimes be sloppy and write Θ_{ij} instead of $n^*\Theta_{ij}$ for the pull back to the curve, so along the ribbon we have $\Theta_{12} = kds$ etc. Also it will sometimes be convenient in computations (as opposed to proving theorems) to use parameters other than arc length.

11. Show that two ribbons (defined over the same interval of s values) are congruent (that is there is a Euclidean motion carrying one into the other) if and only if the functions k, τ , and w are the same.

A ribbon is really just a curve in the space, H , of all Euclidean frames, having the property that the base point, that is the v of the frame (v, e_1, e_2, e_3) has non-vanishing derivative. The previous exercise says that two curves, $i : I \rightarrow H$ and $j : I \rightarrow H$ in H differ by an overall left translation (that is satisfy $j = L_h \circ i$) if and only if the forms $\theta, \Theta_{12}, \Theta_{13}, \Theta_{23}$ pull back to the same forms on I . The

form $i^*\theta$ is just the arc length form ds as we mentioned above. It is absolutely crucial for the rest of this course to understand the meaning of the form $i^*\Theta_{12}$.

Consider a circle of latitude on a sphere of radius R . To fix the notation, suppose that the circle is at angular distance v from the north pole and that we use u as angular coordinates along the circle. Take the ribbon adapted to the sphere, so e_1 is the unit tangent vector to the circle of latitude and e_2 is the unit tangent vector to the circle of longitude chosen as above. Problem **10** then implies that $i^*\Theta_{12} = -\cos v du$.

12. Let C be a straight line (say a piece of the z -axis) parametrized according to arc length and let e_2 be rotating at a rate $f(s)$ about C (so, for example, $e_2 = \cos f(s)\mathbf{i} + \sin f(s)\mathbf{j}$ where \mathbf{i} and \mathbf{j} are the unit vectors in the x and y directions). What is $i^*\Theta_{12}$?

To continue our understanding of Θ_{12} , let us consider what it means for two ribbons, $i: I \rightarrow H$ and $j: I \rightarrow H$ to have the same value of the pullback of Θ_{12} at some point $s_0 \in I$ (where I is some interval on the real line). So

$$(i^*\Theta_{12})|_{s=s_0} = (j^*\Theta_{12})|_{s=s_0}.$$

There is a (unique) left multiplication, that is a unique Euclidean motion, which carries $i(s_0)$ to $j(s_0)$. Let assume that we have applied this motion so we assume that $i(s_0) = j(s_0)$. Let us write

$$i(s) = (C(s), e_1(s), e_2(s), e_3(s)), \quad j(s) = (D(s), f_1(s), f_2(s), f_3(s))$$

and we are assuming that $C(s_0) = D(s_0)$, $C'(s_0) = e_1(s_0) = f_1(s_0) = D'(s_0)$ so the curves C and D are tangent at s_0 , and that $e_2(s_0) = f_2(s_0)$ so that the planes of the ribbon (spanned by the first two orthonormal vectors) coincide. Then our condition about the equality of the pullbacks of Θ_{12} asserts that

$$((e'_2 - f'_2)(s_0), e_1(s_0)) = 0$$

and of course $((e'_2 - f'_2)(s_0), e_2(s_0)) = 0$ automatically since $e_2(s)$ and $f_2(s)$ are unit vectors. So the condition is that the relative change of e_2 and f_2 (and similarly e_1 and f_1) at s_0 be normal to the common tangent plane to the ribbon.

2.22 Developing a ribbon.

We will now drop one dimension, and consider ribbons in the plane (or, if you like, ribbons lying in a fixed plane in three dimensional space). So all we have is θ and Θ_{12} . Also, the orientation of the curve and of the plane completely determines e_2 as the unit vector in the plane perpendicular to the curve and such that e_1, e_2 give the correct orientation. so a ribbon in the plane is the same as an oriented curve.

13. Let $k = k(s)$ be any continuous function of s . Show that there is a ribbon in the plane whose base curve is parametrized by arc length and for which

$j^*\Theta_{12} = kds$. Furthermore, show that this planar ribbon (curve) is uniquely determined up to a planar Euclidean motion.

It follows from the preceding exercise, that we have a way of associating a curve in the plane (determined up to a planar Euclidean motion) to any ribbon in space. It consists of rocking and rolling the ribbon along the plane in such a way that infinitesimal change in the e_1 and e_2 are always normal to the plane. Mathematically, it consists in solving problem **13** for the $k = k(s)$ where $i^*\Theta_{12} = kds$ for the ribbon. We call this operation *developing* the ribbon onto a plane. In particular, if we have a curve on a surface, we can consider the ribbon along the curve induced by the surface. In this way, we may talk of developing the surface on a plane along the given curve. Intuitively, if the surface were convex, this amounts to rolling the surface on a plane along the curve.

noindent14. What are results of developing the ribbons of Problem **12** and the ribbon we associated to a circle of latitude on the sphere?

2.23 Parallel transport along a ribbon.

Recall that a ribbon is a curve in the space, H , of all Euclidean frames, having the property that the base point, that is the C of the frame (C, e_1, e_2, e_3) has non-vanishing derivative at all points. So C defines a curve in Euclidean three space with nowhere vanishing tangent. We will parameterize this curve (and the ribbon) by arc length. By a unit vector field tangent to the ribbon we will mean a curve, $v(s)$ of unit vectors everywhere tangent to the ribbon, so

$$v(s) = \cos \alpha(s) e_1(s) + \sin \alpha(s) e_2(s). \quad (2.34)$$

We say that the vector field is *parallel* along the ribbon if the infinitesimal change in v is always normal to the ribbon, i.e. if

$$(v'(s), e_1(s)) \equiv (v'(s), e_2(s)) \equiv 0.$$

Recall the form $\Theta_{12} = kds$ from before.

15. Show that the vector field as given above is parallel if and only if the function α satisfies the differential equation

$$\alpha' + k = 0.$$

Conclude that the notion of parallelism depends only on the form Θ_{12} . Also conclude that given any unit vector, v_0 at some point s_0 , there is a unique parallel vector field taking on the value v_0 at s_0 . The value $v(s_1)$ at some second point is called the *parallel transport* of v_0 (along the ribbon) from s_0 to s_1 .

16. What is the condition on a ribbon that the tangent vector to the curve itself, i.e. the vector field e_1 , be parallel? Which circles on the sphere are such that the associated ribbon has this property?

Suppose the ribbon is closed, i.e. $C(s+L) = C(s)$, $e_1(s+L) = e_1(s)$, $e_2(s+L) = e_2(s)$ for some length L . We can then start with a vector v_0 at point s_0 and transport it all the way around the ribbon until we get back to the same point, i.e. transport from s_0 to $s_0 + L$. The vector v_1 we so obtain will make some angle, call it Φ with the vector v_0 . The angle Φ is called the *holonomy* of the (parallel transport of the) ribbon.

17. Show that Φ is independent of the choice of s_0 and v_0 . What is its expression in terms of Θ_{12} ?

18. What is the holonomy for a circle on the sphere in terms of its latitude.

19. Show that if the ribbon is planar (so e_1 and e_2 lie in a fixed plane) a vector field is parallel if and only if it is parallel in the usual sense of Euclidean geometry (say makes a constant angle with the x-axis). But remember that the curve is turning. So the holonomy of a circle in the plane is $\pm 2\pi$ depending on the orientation. Similarly for the sum of the exterior angles of a triangle (think of the corners as being rounded out).

Convince yourself of the following fact which is not so easy unless you know the trick: Show that for any smooth simple closed curve (i.e. one with no self intersections) in the plane the holonomy is always $\pm 2\pi$.

Exercises **15,17**, and **19**, together with the results above give an alternative interpretation of parallel transport: develop the ribbon onto the plane and then just translate the vector v_0 in the Euclidean plane so that its origin lies at the image of s_1 . Then consider the corresponding vector field along the ribbon.

The function k in $\Theta_{12} = kds$ is called the geodesic curvature of the ribbon. The integral $\int \Theta_{12} = \int kds$ is called the *total geodesic curvature* of the ribbon. It gives the total change in angle (including multiples of 2π) between the tangents to the initial and final points of the developed curve.

2.24 Surfaces in \mathbf{R}^3 .

We let M be a two dimensional submanifold of \mathbf{R}^3 and \mathcal{O} its bundle of adapted frames. We have a “projection” map

$$\pi : \mathcal{O} \rightarrow M, \quad (m, e_1, e_2, e_3) \mapsto m,$$

which we can also write

$$\pi = m.$$

Suppose that we consider the “truncated” version of the adapted bundle of frames $\tilde{\mathcal{O}}$ where we forget about e_3 . That is, let consist of all (m, e_1, e_2) where $m \in M$ and e_1, e_2 is an orthonormal basis of the tangent space TM_m to M at m . Notice that the definition we just gave was *intrinsic*. The concept of an orthonormal basis of TM_m depends only on the scalar product on TM_m . The differential of the map $m : \tilde{\mathcal{O}} \rightarrow M$ at a point (m, e_1, e_2) sends a tangent vector ξ to $\tilde{\mathcal{O}}$ at (m, e_1, e_2, e_3) to a tangent vector to M at m , and the scalar product of this image vector with e_1 is a linear function of ξ . We have just given an intrinsic of θ_1 . (By abuse of language I am using this same letter θ_1 for the form (dm, e_1) on $\tilde{\mathcal{O}}$ as e_3 does not enter into its definition.) Similarly, we see that θ_2 is an intrinsically defined form. From their very definitions, the forms θ_1 and θ_2 are linearly independent at every point of $\tilde{\mathcal{O}}$. Therefore the forms $d\theta_1$ and $d\theta_2$ are intrinsic, and this proves that the form Θ_{12} is intrinsic. Indeed, if we had two linear differential forms σ and τ on \mathcal{O} which satisfied

$$\begin{aligned} d\theta_1 &= \sigma \wedge \theta_2, \\ d\theta_1 &= \tau \wedge \theta_2 \\ d\theta_2 &= -\sigma \wedge \theta_1 \\ d\theta_2 &= -\tau \wedge \theta_1 \end{aligned}$$

then the first two equations give

$$(\sigma - \tau) \wedge \theta_2 \equiv 0$$

which implies that $(\sigma - \tau)$ is a multiple of θ_2 and the last two equations imply that $\sigma - \tau$ is a multiple of θ_1 so $\sigma = \tau$. The next few problems will give a (third) proof of Gauss’s theorem egregium. They will show that

$$d\Theta_{12} = -\pi^*(K)\theta_1 \wedge \theta_2$$

where K is the Gaussian curvature.

This assertion is local (in M), so we may temporarily make the assumption that M is orientable - this allows us to look at the sub-bundle $\bar{\mathcal{O}} \subset \mathcal{O}$ of oriented frames, consisting of those frames for which e_1, e_2 form an oriented basis of TM_m and where e_1, e_2, e_3 an oriented frame on \mathbf{R}^3 .

Let dA denote the (oriented) area form on the surface M . (A bad but standard notation, since we the area form is not the differential of a one form, in general.) Recall that when evaluated on any pair of tangent vectors, η_1, η_2 at $m \in M$ it is the oriented area of the parallelogram spanned by η_1 and η_2 , and this is just the determinant of the matrix of scalar products of the η ’s with any oriented orthonormal basis. Conclude

20. Explain why

$$\pi^*dA = \theta_1 \wedge \theta_2.$$

The third component, e_3 of any frame is completely determined by the point on the surface and the orientation as the unit normal, n to the surface. Now n

can be thought of as a map from M to the unit sphere, S in \mathbf{R}^3 . Let dS denote the oriented area form of the unit sphere. So n^*dS is a two form on M and we can define the function K by

$$n^*dS = KdA.$$

21 Show that the function K is Gaussian curvature of the surface.

22. Show that

$$n^*dS = \Theta_{31} \wedge \Theta_{32}$$

and

23. Conclude that

$$d\Theta_{12} = -\pi^*(KdA).$$

We are going to want to apply Stokes' theorem to this formula. But in order to do so, we need to integrate over a two dimensional region. So let U be some open subset of M and let

$$\psi : U \rightarrow \pi^{-1}U \subset \mathcal{O}$$

be a map satisfying

$$\pi \circ \psi = id.$$

So ψ assigns a frame to each point of U in a differentiable manner. Let C be a curve on M and suppose that C lies in U . Then the surface determines a ribbon along this curve, namely the choice of frames from which e_1 is tangent to the curve (and pointing in the positive direction). So we have a map $R : C \rightarrow \mathcal{O}$ coming from the geometry of the surface, and (with now necessarily different notation from the preceding section) $R^*\Theta_{12} = kds$ is the geodesic curvature of the ribbon as studied above. Since the ribbon is determined by the curve (as M is fixed) we can call it the geodesic curvature of the curve. On the other hand, we can consider the form $\psi^*\Theta_{12}$ pulled back to the curve. Let

$$\psi \circ C(s) = (C(s), f_1(s), f_2(s), n(s))$$

and let $\phi(s)$ be the angle that $e_1(s)$ makes with $f_1(s)$ so

$$e_1(s) = \cos \phi(s)f_1(s) + \sin \phi(s)f_2(s), \quad e_2(s) = -\sin \phi(s)f_1(s) + \cos \phi(s)f_2(s).$$

24. Let $C^*\psi^*\Theta_{12}$ denote the pullback of $\psi^*\Theta_{12}$ to the curve. Show that

$$kds = d\phi + C^*\psi^*\Theta_{12}.$$

Conclude that

Proposition 2 *The*

*total geodesic curvature = $\phi(b) - \phi(a) + \int_C \psi * \Theta_{12}$ where $\phi(b) - \phi(a)$ denotes the total change of angle around the curve.*

How can we construct a ψ ? Here is one way that we described earlier: Suppose that U is a coordinate chart and that x_1, x_2 are coordinates on this chart. Then $\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}$ are linearly independent vectors at each point and we can apply Gram Schmidt to orthonormalize them. This give a ψ and the angle ϕ above is just the angle that the vector e_1 makes with the x -axis in this coordinate system. Suppose we take C to be the boundary of some nice region, D , in U . For example, suppose that C is a triangle or some other polygon with its edges rounded to make a smooth curve. Then the total change in angle is 2π and so

25. Conclude that for such a curve

$$\int \int_D K dA + \int_C k ds = 2\pi.$$

The integral of $K dA$ is called the total Gaussian curvature.

26. Show that as the curve actually approaches the polygon, the contribution from the rounded corners approaches the exterior angle of the polygon. Conclude that if a region in a coordinate neighborhood on the surface is bounded by continuous piecewise differentiable arcs making exterior angles at the corners

Proposition 3 *the total Gaussian curvature + \sum total geodesic curvatures + \sum exterior angles = 2π .*

27. Suppose that we have subdivided a compact surface into polygonal regions, each contained in a coordinate neighborhood, with f faces, e edges, and v vertices. Let $\xi = f - e + v$. show that

$$\int_M K dA = 2\pi\xi.$$

Chapter 3

Levi-Civita Connections.

3.1 Definition of a linear connection on the tangent bundle.

A **linear connection** ∇ on a manifold M is a rule which assigns a vector field $\nabla_X Y$ to each pair of vector fields X and Y which is bilinear (over \mathbf{R}) subject to the rules

$$\nabla_{fX} Y = f \nabla_X Y \quad (3.1)$$

and

$$\nabla_X(gY) = (Xg)Y + g(\nabla_X Y). \quad (3.2)$$

While condition (3.2) is the same as the corresponding condition

$$L_X(gY) = [X, gY] = (Xg)Y + gL_X Y$$

for Lie derivatives, condition (3.1) is quite different from the corresponding formula

$$L_{fX} Y = [fX, Y] = -(Yf)X + fL_X Y$$

for Lie derivatives. In contrast to the Lie derivative, condition (3.1) implies that the value of $\nabla_X Y$ at $x \in M$ depends only on the value $X(x)$.

If $\xi \in TM_x$ is a tangent vector at $x \in M$, and Y is a vector field defined in some neighborhood of x we use the notation

$$\nabla_\xi Y := (\nabla_X Y)(x), \quad \text{where } X(x) = \xi. \quad (3.3)$$

By the preceding comments, this does not depend on how we choose to extend ξ to X so long as $X(x) = \xi$.

While the Lie derivative is an intrinsic notion depending only on the differentiable structure, a connection is an additional piece of geometric structure.

3.2 Christoffel symbols.

These give the expression of a connection in local coordinates: Let x^1, \dots, x^n be a coordinate system, and let us write

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

for the corresponding vector fields. Then

$$\nabla_{\partial_i} \partial_j = \sum_k \Gamma_{ij}^k \partial_k$$

where the functions Γ_{ij}^k are called the **Christoffel symbols**. We will frequently use the shortened notation

$$\nabla_i := \nabla_{\partial_i}.$$

So the definition of the Christoffel symbols is written as

$$\nabla_i \partial_j = \sum_k \Gamma_{ij}^k \partial_k. \quad (3.4)$$

If

$$Y = \sum_j Y^j \partial_j$$

is the local expression of a general vector field Y then (3.2) implies that

$$\nabla_i Y = \sum_k \left\{ \frac{\partial Y^k}{\partial x^i} + \sum_j \Gamma_{ij}^k Y^j \right\} \partial_k. \quad (3.5)$$

3.3 Parallel transport.

Let $C : I \rightarrow M$ be a smooth map of an interval I into M . We refer to C as a parameterized curve. We will say that this curve is non-singular if $C'(t) \neq 0$ for any t where $C'(t)$ denotes the tangent vector at $t \in I$. By a **vector field Z along C** we mean a rule which smoothly attaches to each $t \in I$ a tangent vector $Z(t)$ to M at $C(t)$. We will let $\mathcal{V}(C)$ denote the set of all smooth vector fields along C . For example, if V is a vector field on M , then the restriction of V to C , i.e. the rule

$$V_C(t) := V(C(t))$$

is a vector field along C . Since the curve C might cross itself, or be closed, it is clear that not every vector field along C is the restriction of a vector field.

On the other hand, if C is non-singular, then the implicit function theorem says that for any $t_0 \in I$ we can find an interval J containing t_0 and a system of coordinates about $C(t_0)$ in M such that in terms of these coordinates the curve is given by

$$x^1(t) = t, \quad x^i(t) = 0, \quad i > 1$$

for $t \in J$. If Z is a smooth vector field along C then for $t \in J$ we may write

$$Z(t) = \sum_j Z^j(t) \partial_j(t, 0, \dots, 0).$$

We may then define the vector field Y on this coordinate neighborhood by

$$Y(x^1, \dots, x^n) = \sum_j Z^j(x^1) \partial_j$$

and it is clear that Z is the restriction of Y to C on J . In other words, *locally*, every vector field along a non-singular curve is the restriction of a vector field of M . If $Z = Y_C$ is the restriction of a vector field Y to C we can define its “derivative” Z' , also a vector field along C by

$$Y'_C(t) := \nabla_{C'(t)} Y. \quad (3.6)$$

If g is a smooth function defined in a neighborhood of the image of C , and h is the pull back of g to I via C , so

$$h(t) = g(C(t))$$

then the chain rule says that

$$h'(t) = \frac{d}{dt} g(C(t)) = C'(t)g,$$

the derivative of g with respect to the tangent vector $C'(t)$. Then if

$$Z = Y_C$$

for some vector field Y on M (and $h = g(C(t))$) equation (3.2) implies that

$$(hZ)' = h'Z + hZ'. \quad (3.7)$$

We claim that there is a unique linear map $Z \mapsto Z'$ defined on all of $\mathcal{V}(C)$ such that (3.7) and (3.6) hold. Indeed, to prove uniqueness, it is enough to prove uniqueness in a coordinate neighborhood, where

$$Z(t) = \sum_j Z^j(t) (\partial_j)_C.$$

Equations (3.7) and (3.6) then imply that

$$Z'(t) = \sum_j \left(Z^{j'}(t) (\partial_j)_C + Z^j(t) \nabla_{C'(t)} \partial_j \right). \quad (3.8)$$

In other words, any notion of “derivative along C ” satisfying (3.7) and (3.6) must be given by (3.8) in any coordinate system. This proves the uniqueness. On the other hand, it is immediate to check that (3.8) satisfies (3.7) and (3.6) if the

curve lies entirely in a coordinate neighborhood. But the uniqueness implies that on the overlap of two neighborhoods the two formulas corresponding to (3.8) must coincide, proving the global existence.

We can make formula (3.8) even more explicit in local coordinates using the Christoffel symbols which tell us that

$$\nabla_{C'(t)} \partial_j = \sum_k \Gamma_{ij}^k \frac{dx^i \circ C}{dt} (\partial_k)_C.$$

Substituting into (3.8) gives

$$Z' = \sum_k \left(\frac{dZ^k}{dt} + \sum_{ij} \Gamma_{ij}^k \frac{dx^i \circ C}{dt} Z^j \right) (\partial_k)_C. \quad (3.9)$$

A vector field Z along C is said to be **parallel** if

$$Z'(t) \equiv 0.$$

Locally this amounts to the Z^i satisfying the system of linear differential equations

$$\frac{dZ^k}{dt} + \sum_{ij} \Gamma_{ij}^k \frac{dx^i \circ C}{dt} Z^j = 0. \quad (3.10)$$

Hence the existence and uniqueness theorem for linear homogeneous differential equations (in particular existence over the entire interval of definition) implies that

Proposition 4 *For any $\zeta \in TM_{C(0)}$ there is a unique parallel vector field Z along C with $Z(0) = \zeta$.*

The rule $t \mapsto C'(t)$ is a vector field along C and hence we can compute its derivative, which we denote by C'' and call the **acceleration** of C . Whereas the notion of tangent vector, C' , makes sense on any manifold, the acceleration only makes sense when we are given a connection.

3.4 Geodesics.

A curve with acceleration zero is called a **geodesic**. In local coordinates we substitute $Z^k = x^{k'}$ into (3.10) to obtain the equation for geodesics in local coordinates:

$$\frac{d^2 x^k}{dt^2} + \sum_{ij} \Gamma_{ij}^k \frac{dx^i}{dt} \frac{dx^j}{dt} = 0, \quad (3.11)$$

where we have written x^k instead of $x^k \circ C$ in (3.11) to unburden the notation. The existence and uniqueness theorem for ordinary differential equations implies that

Proposition 5 *For any tangent vector ξ at any point $x \in M$ there is an interval I about 0 and a unique geodesic C such that $C(0) = x$ and $C'(0) = \xi$.*

By the usual arguments we can then extend the domain of definition of the geodesic through ξ to be maximal.

This is the first of many definitions (or characterizations, if take this to be the basic definition) that we shall have of geodesics - the notion of being self-parallel. (In the case that all the $\Gamma_{ij}^k = 0$ we get the equations for straight lines.)

Suppose that $C : I \rightarrow M$ is a (non-constant) geodesic, and we consider a “reparametrization” of C , i.e. consider the curve $B = C \circ h : J \rightarrow M$ where $h : J \rightarrow I$ is a diffeomorphism of the interval J onto the interval I . We write $t = h(s)$ so that

$$\frac{dB}{ds} = \frac{dC}{dt} \frac{dh}{ds}$$

and hence

$$\frac{d^2B}{ds^2} = \frac{d^2C}{dt^2} \left(\frac{dh}{ds}\right)^2 + \frac{dC}{ds} \frac{d^2h}{ds^2} = \frac{dC}{ds} \frac{d^2h}{ds^2}$$

since $C'' = 0$ as C is a geodesic. The fact that C is not constant (and the uniqueness theorem for differential equations) says that C' is never zero. Hence B is a geodesic if and only if

$$\frac{d^2h}{ds^2} \equiv 0$$

or

$$h(s) = as + b$$

where a and b are constants with $a \neq 0$. In short, the fact of being a non-constant geodesic determines the parameterization up to an affine change of parameter.

3.5 Covariant differential.

We can extend the notion of covariant derivative with respect to a vector field X (which has been defined on functions by $f \mapsto Xf$ and on vector fields by $Y \mapsto \nabla_X Y$) to all tensor fields: We first extend to linear differential forms by the rule

$$(\nabla_X \theta)(Y) = X(\theta(Y)) - \theta(\nabla_X Y) \tag{3.12}$$

Replacing Y by gY has the effect of pulling out a factor of g since the two terms on the right involving Xg cancel. This shows that $\nabla_X \theta$ is again a linear differential form. Notice that

$$\nabla_{fX} \theta = f \nabla_X \theta$$

and

$$\nabla_X (g\theta) = (Xg)\theta + g \nabla_X \theta.$$

We now extend ∇_X to be a “tensor derivation” requiring that

$$\nabla_X(\alpha \otimes \beta) = (\nabla_X \alpha) \otimes \beta + \alpha \otimes \nabla_X \beta$$

for any pair of tensor fields α and β . For example

$$\nabla_X(\theta \otimes Z) = \nabla_X \theta \otimes Z + \theta \otimes \nabla_X Z.$$

This then defines ∇_X on all tensor fields which are sums of products of one forms and vector fields. Notice that if we define the “contraction”

$$C : \theta \otimes Z \mapsto \theta(Z)$$

then the definition (3.12) of $\nabla_X \theta$ implies that

$$\nabla_X(C(\theta \otimes Z)) = \nabla_X(\theta(Z)) = C(\nabla_X \theta \otimes Z + \theta \otimes \nabla_X Z) = C(\nabla_X(\theta \otimes Z)).$$

in other words, ∇_X commutes with contraction

$$\nabla_X \circ C = C \circ \nabla_X. \quad (3.13)$$

This was checked in the special case that we had a tensor of type (1,1) which was the tensor product of a one form and a vector field. But if we have a tensor of type (r,s) which is a product of one forms and vector fields, then we may form the contraction of any one-form factor with any vector field factor to obtain a tensor of type (r-1,s-1) and (3.13) continues to hold.

If γ is a general tensor field of type (r,s), it is completely determined by evaluation on all tensor fields ρ of type (s,r) which are products of one forms and vector fields. We then define $\nabla_X \gamma$ by

$$(\nabla_X \gamma)(\rho) = X(\gamma(\rho)) - \gamma(\nabla_X \rho).$$

In the case that γ is itself a sum of products of one-forms and vector fields this coincides with our old definition. Again this implies that $\nabla_X \gamma$ is a tensor. Furthermore, contraction in any two positions in γ is dual (locally) to insertion of $\sum \theta^i \otimes E_i$ into the corresponding positions in a tensor of type (s-1,r-1) where the E_i form a basis locally of the vector fields at each point and the θ^i form the dual basis. But

$$\nabla_X(\theta^i \otimes E_i) = 0$$

since if the functions a_j^i are defined by $\nabla_X E_j = \sum_i a_j^i E_i$ then $\nabla_X \theta^i = -\sum_j a_j^i \theta^j$ as follows from (3.12). This shows that (3.13) holds in general.

We can think of the covariant derivative as assigning to each tensor field γ of type (r,s) a tensor field $\nabla \gamma$ of type (r,s+1), given by

$$\nabla \gamma(\rho \otimes X) = \nabla_X \gamma(\rho).$$

The tensor $\nabla \gamma$ is called the **covariant differential** of γ .

3.6 Torsion.

Let ∇ be a connection, X and Y vector fields and f and g functions. Using (3.1), (3.2), and the corresponding equations for Lie brackets we find

$$\nabla_{fX}(gY) - \nabla_{gY}(fX) - [fX, gY] = fg(\nabla_X Y - \nabla_Y X - [X, Y]),$$

In other words the value of

$$\tau(X, Y) := \nabla_X Y - \nabla_Y X - [X, Y]$$

at any point x depends only on the values $X(x), Y(x)$ of the vector fields at x . So τ defines a tensor field of type (1,2) in the sense that it assigns to any pair of tangent vectors at a point, a third tangent vector at that point. This tensor field is called the **torsion tensor** of the connection. So a connection has *zero torsion* if and only if

$$\nabla_X Y - \nabla_Y X = [X, Y] \quad (3.14)$$

for all pairs of vector fields X and Y . In terms of local coordinates, $[\partial_i, \partial_j] = 0$. So

$$\tau(\partial_i, \partial_j) = \nabla_i \partial_j - \nabla_j \partial_i = \sum_k (\Gamma_{ij}^k - \Gamma_{ji}^k) \partial_k.$$

Thus a connection has zero torsion if and only if its Christoffel symbols are symmetric in i and j .

3.7 Curvature.

The **curvature** $R = R(\nabla)$ of the connection ∇ is defined to be the map $\mathcal{V}(M)^3 \rightarrow \mathcal{V}(M)$ assigning to three vector fields X, Y, Z the value

$$R_{XY}Z := [\nabla_X, \nabla_Y]Z - \nabla_{[X, Y]}Z. \quad (3.15)$$

The expression $[\nabla_X, \nabla_Y]$ occurring on the right in (3.15) is the commutator of the two operators ∇_X and ∇_Y , that is $[\nabla_X, \nabla_Y] = \nabla_X \circ \nabla_Y - \nabla_Y \circ \nabla_X$. We first observe that R is a tensor, i.e. that the value of $R_{XY}Z$ at a point depends only on the values of X, Y , and Z at that point. To see this we must show that

$$R_{fXgYhZ} = fghR_{XY}Z$$

for any three smooth functions f, g and h . For this it suffices to check this one at a time, i.e. when two of the three functions are identically equal to one. For example, if $f \equiv 1 \equiv h$ we have

$$\begin{aligned} -R_{X,gY}Z &= \nabla_{[X,gY]}Z - \nabla_X \nabla_{gY}Z + \nabla_{gY} \nabla_X Z \\ &= (Xg)\nabla_Y Z + g\nabla_{[X,Y]}Z - (Xg)\nabla_X \nabla_Y Z - g\nabla_X \nabla_Y Z + g\nabla_Y \nabla_X Z \\ &= gR_{XY}Z. \end{aligned}$$

Since R is anti-symmetric in X and Y we conclude that $R_{fXY}Z = fR_{XY}Z$. Finally,

$$\begin{aligned}
-R_{XY}(hZ) &= ([X, Y]h)Z + h\nabla_{[X, Y]}Z - \nabla_X((Yh)Z + h\nabla_Y Z) + \\
&\quad \nabla_Y((Xh)Z + h\nabla_X Z) \\
&= hR_{XY}Z + ([X, Y]h - (XY - YX)h)Z - Xh\nabla_Y Z \\
&\quad - Yh\nabla_X Z + Yh\nabla_X Z + Xh\nabla_Y Z \\
&= hR_{XY}Z.
\end{aligned}$$

Thus we get a curvature **tensor** (of type (1,3)) which assigns to every three tangent vectors ξ, η, ζ at a point x the value

$$R_{\xi\eta}\zeta := (R_{XY}Z)(x)$$

where X, Y, Z are any three vector fields with $X(x) = \xi, Y(x) = \eta, Z(x) = \zeta$. Alternatively, we speak of the curvature **operator** at the point x defined by

$$R_{\xi\eta} : TM_x \rightarrow TM_x, \quad R_{\xi\eta} : \zeta \mapsto R_{\xi\eta}\zeta.$$

As we mentioned, the curvature operator is anti-symmetric in ξ and η :

$$R_{\xi\eta} = -R_{\eta\xi}.$$

The classical expression of the curvature tensor in terms of the Christoffel symbols is obtained as follows: Since $[\partial_k, \partial_\ell] = 0$,

$$\begin{aligned}
R_{\partial_k \partial_\ell} \partial_j &= \nabla_k(\nabla_\ell \partial_j) - \nabla_\ell(\nabla_k \partial_j) \\
&= -\nabla_\ell \left(\sum_m \Gamma_{kj}^m \partial_m \right) + \nabla_k \left(\sum_m \Gamma_{\ell j}^m \partial_m \right) \\
&= -\sum_m \left(\frac{\partial}{\partial x^\ell} \Gamma_{kj}^m \partial_m + \sum_{m,r} \Gamma_{kj}^m \Gamma_{\ell r}^m \partial_r \right) \partial_r + \sum_m \left(\frac{\partial}{\partial x^k} \Gamma_{\ell j}^m \partial_m - \sum_{m,r} \Gamma_{\ell j}^m \Gamma_{km}^r \partial_r \right) \partial_r \\
&= \sum_i R_{jkl}^i \partial_i
\end{aligned}$$

where

$$R_{jkl}^i = -\frac{\partial}{\partial x^\ell} \Gamma_{kj}^i + \frac{\partial}{\partial x^k} \Gamma_{\ell j}^i - \sum_m \Gamma_{\ell m}^i \Gamma_{kj}^m + \sum_m \Gamma_{km}^i \Gamma_{\ell j}^m. \quad (3.16)$$

If the connection has zero torsion we claim that

$$R_{\xi\eta}\zeta + R_{\eta\zeta}\xi + R_{\zeta\xi}\eta = 0, \quad (3.17)$$

or, using the cyclic sum notation we introduced with the Jacobi identity, that

$$\mathcal{Cyc} R_{\xi\eta}\zeta = 0.$$

To prove this, we may extend ξ, η , and ζ to vector fields whose brackets all commute (say by using vector fields with constant coefficients in a coordinate neighborhood). Then

$$R_{XY}Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z.$$

Therefore

$$\begin{aligned} \text{Cyc } R_{XY}Z &= \text{Cyc } \nabla_X \nabla_Y Z - \text{Cyc } \nabla_Y \nabla_X Z \\ &= \text{Cyc } \nabla_X \nabla_Z Y - \text{Cyc } \nabla_X \nabla_Z Y \end{aligned}$$

since making a cyclic permutation in an expression $\text{Cyc } F(X, Y, Z)$ does not affect its value. But the fact that the connection is torsion free means that we can write the last expression as

$$\text{Cyc } \nabla_X [Y, Z] = 0$$

by our assumption that all Lie brackets vanish. QED

3.8 Isometric connections.

Suppose that M is a **semi-Riemannian manifold**, meaning that we are given a smoothly varying non-degenerate scalar product $\langle \cdot, \cdot \rangle_x$ on each tangent space TM_x . Given two vector fields X and Y , we let $\langle X, Y \rangle$ denote the function

$$\langle X, Y \rangle(x) := \langle X(x), Y(x) \rangle_x.$$

We say that a connection ∇ is **isometric** for $\langle \cdot, \cdot \rangle$ if

$$X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle \quad (3.18)$$

for any three vector fields X, Y, Z . It is a sort of Leibniz's rule for scalar products. If we go back to the definition of the derivative of a vector field along a curve arising from the connection ∇ , we see that (3.18) implies that

$$\frac{d}{dt} \langle Y, Z \rangle = \langle Y', Z \rangle + \langle Y, Z' \rangle$$

for any pair of vector fields along a curve C . In particular, if Y and Z are parallel along the curve, so that $Y' = Z' = 0$, we see that $\langle Y, Z \rangle$ is constant. This is the key meaning of the condition that a connection be isometric: parallel translation along any curve is an isometry of the tangent spaces.

3.9 Levi-Civita's theorem.

This asserts that on any semi-Riemannian manifold there exists a unique connection which is isometric and is torsion free. It is characterized by the **Koszul formula**

$$2\langle \nabla_V W, X \rangle = \\ V\langle W, X \rangle + W\langle X, V \rangle - X\langle V, W \rangle - \langle V, [W, X] \rangle + \langle W, [X, V] \rangle + \langle X, [V, W] \rangle \quad (3.19)$$

for any three vector fields X, V, W . To prove Koszul's formula, we apply the isometric condition to each of the first three terms occurring on the right hand side of (3.19). For example the first term becomes $\langle \nabla_V W, X \rangle + \langle W, \nabla_V X \rangle$. We apply the torsion free condition to each of the last three terms. For example the last term becomes $\langle X, \nabla_V W - \nabla_W V \rangle$. There will be a lot of cancellation leaving the left hand side. Since the vector field $\nabla_V W$ is determined by knowing its scalar product $\langle \nabla_V W, X \rangle$ for all vector fields X , the Koszul formula proves the uniqueness part of Levi-Civita's theorem.

On the other hand, the right hand side of the Koszul formula is function linear in X , i.e.

$$\langle \nabla_V W, fX \rangle = f\langle \nabla_V W, X \rangle$$

as can be checked using the properties of ∇ and Lie bracket. So we obtain a well defined vector field, $\nabla_V W$ and it is routine to check that this satisfies the conditions for a connection and is torsion free and isometric.

We can use the Koszul identity to derive a formula for the Christoffel symbols in terms of the metric. First some standard notations: We will use the symbol \mathbf{g} to stand for the metric, so \mathbf{g} is just another notation for $\langle \cdot, \cdot \rangle$. In a local coordinate system we write

$$g_{ij} := \langle \partial_i, \partial_j \rangle$$

so

$$\mathbf{g} = \sum_{ij} g_{ij} dx^i \otimes dx^j.$$

Here the g_{ij} are functions on the coordinate neighborhood, but we are suppressing the functional dependence on the points in the notation. The metric \mathbf{g} is a (symmetric) tensor of type (0,2). It induces an isomorphism (at each point) of the tangent space with the cotangent space, each tangent vector ξ going into the linear function $\langle \xi, \cdot \rangle$ consisting of scalar product by ξ . By the above formula the map is given by

$$\partial_i \mapsto \sum_j g_{ij} dx^j.$$

This isomorphism induces a scalar product on the cotangent space at each point, and so a tensor of type (2,0) which we shall denote by $\hat{\mathbf{g}}$ or sometimes by $\mathbf{g} \uparrow \uparrow$. We write

$$g^{ij} := \langle dx^i, dx^j \rangle$$

so

$$\hat{\mathbf{g}} = \sum_{ij} g^{ij} \partial_i \otimes \partial_j.$$

(The transition from the two lower indices to the two upper indices is the reason for the vertical arrows notation.) The metric on the cotangent spaces induces a map into its dual space which is the tangent space given by

$$dx^i \mapsto \sum g^{ij} \partial_j$$

and the two maps - from tangent spaces to cotangent spaces and vice versa - are inverses of one another so

$$\sum_k g^{ik} g_{kj} = \delta_j^i,$$

the “matrices” (g_{ij}) and (g^{ij}) are inverses.

Now let us substitute $X = \partial_m, V = \partial_i, W = \partial_j$ into the Koszul formula (3.19). All brackets on the right vanish and we get

$$2\langle \nabla_i \partial_j, \partial_m \rangle = \partial_i(g_{jm}) + \partial_j(g_{im}) - \partial_m(g_{ij}).$$

Since

$$\nabla_i \partial_j = \sum_k \Gamma_{ij}^k \partial_k$$

is the definition of the Christoffel symbols, the preceding equation becomes

$$2 \sum_a \Gamma_{ij}^a g_{am} = \partial_i(g_{jm}) + \partial_j(g_{im}) - \partial_m(g_{ij}).$$

Multiplying this equation by g^{mk} and summing over m gives

$$\Gamma_{ij}^k = \frac{1}{2} \sum_m g^{km} \left\{ \frac{\partial g_{jm}}{\partial x^i} + \frac{\partial g_{im}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^m} \right\}. \quad (3.20)$$

In principle, we should substitute this formula into (3.11) and solve to obtain the geodesics. In practice this is a mess for a general coordinate system and so we will spend a good bit of time developing other means (usually group theoretical) for finding geodesics. However the equations are manageable in orthogonal coordinates.

3.10 Geodesics in orthogonal coordinates.

A coordinate system is called **orthogonal** if

$$g_{ij} = 0, \quad i \neq j.$$

If we are lucky enough to have an orthogonal coordinate system the equations for geodesics take on a somewhat simpler form. First notice that (3.20) becomes

$$\Gamma_{ij}^k = \frac{1}{2} g^{kk} \left\{ \frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k} \right\}$$

So (3.11) becomes

$$\frac{d^2 x^k}{dt^2} + g^{kk} \sum_i \frac{\partial g_{kk}}{\partial x^i} \frac{dx^k}{dt} \frac{dx^i}{dt} - \frac{1}{2} g^{kk} \sum_i \frac{\partial g_{ii}}{\partial x^k} \frac{dx^i}{dt} \frac{dx^i}{dt} = 0.$$

If we multiply this equation by g_{kk} and bring the negative term to the other side we obtain

$$\frac{d}{dt} \left(g_{kk} \frac{dx^k}{dt} \right) = \frac{1}{2} \sum_i \frac{\partial g_{ii}}{\partial x^k} \left(\frac{dx^i}{dt} \right)^2 \quad (3.21)$$

as the equations for geodesics in orthogonal coordinates.

3.11 Curvature identities.

The curvature of the Levi-Civita connection satisfies several additional identities beyond the two curvature identities that we have already discussed. Let us choose vector fields X, Y, V with vanishing brackets. We have

$$\begin{aligned} -\langle R_{XY}V, V \rangle &= -\langle \nabla_X \nabla_Y V, V \rangle + \langle \nabla_Y \nabla_X V, V \rangle \\ &= Y \langle \nabla_X V, V \rangle - \langle \nabla_X V, \nabla_Y V \rangle - X \langle \nabla_Y V, V \rangle + \langle \nabla_Y V, \nabla_X V \rangle \\ &= \frac{1}{2} YX \langle V, V \rangle - \frac{1}{2} XY \langle V, V \rangle \\ &= \frac{1}{2} [X, Y] \langle V, V \rangle \\ &= 0. \end{aligned}$$

This implies that for any three tangent vectors we have

$$\langle R_{\xi\eta}\zeta, \zeta \rangle = 0$$

and hence by polarization that for any four tangent vectors we have

$$\langle R_{\xi\eta}v, \zeta \rangle = -\langle v, R_{\xi\eta}\zeta \rangle. \quad (3.22)$$

This equation says that the curvature operator $R_{\xi\eta}$ acts as an infinitesimal orthogonal transformation on the tangent space.

The last identity we want to discuss is the symmetry property

$$\langle R_{\xi\eta}v, \zeta \rangle = \langle R_{v\zeta}\xi, \eta \rangle. \quad (3.23)$$

The proof consists of starting with the identity

$$\text{Cyc} R_{\eta,v}\xi = 0$$

and taking the scalar product with ζ to obtain

$$\langle \text{Cyc} R_{\eta,v}\xi, \zeta \rangle = 0.$$

This is an equation involving three terms. Take the cyclic permutation of the four vectors to obtain four equations like this involving twelve terms in all. When we add the four equations eight of the terms cancel in pairs and the remaining terms give (3.23). We summarize the symmetry properties of the Riemann curvature:

- $R_{\xi\eta} = -R_{\eta\xi}$
- $\langle R_{\xi\eta}v, \zeta \rangle = -\langle vR_{\xi\eta}\zeta \rangle$
- $R_{\xi\eta}\zeta + R_{\eta\zeta}\xi + R_{\zeta\xi}\eta = 0$
- $\langle R_{\xi\eta}v, \zeta \rangle = \langle R_{v\zeta}\xi, \eta \rangle$.

3.12 Sectional curvature.

Let V is a vector space with a non-degenerate symmetric bilinear form $\langle \cdot, \cdot \rangle$. A subspace is called **non-degenerate** if the restriction of the $\langle \cdot, \cdot \rangle$ to this subspace is non-degenerate. (If $\langle \cdot, \cdot \rangle$ is positive definite, then all subspaces are non-degenerate.) A two dimensional subspace Π is non-degenerate if and only if for any basis v, w of Π the “area ”

$$Q(v, w) := \langle v, v \rangle \langle w, w \rangle - \langle v, w \rangle^2$$

does not vanish.

Let Π be a non-degenerate plane (=two dimensional subspace) of the tangent space TM_x of a semi-Riemannian manifold. Then its **sectional curvature** is defined as

$$K(\Pi) := \frac{-\langle R_{vw}v, w \rangle}{Q(v, w)}. \quad (3.24)$$

It is easy to check that this is independent of the choice of basis v, w .

3.13 Ricci curvature.

If we hold $\xi \in TM_x$ and $\eta \in TM_x$ fixed in $R_{\xi v}\eta$ then the map

$$v \mapsto R_{\xi v}\eta \quad v \in TM_x$$

is a linear map of TM_x into itself. Its trace (which is biinear in ξ and η) is known as the **Ricci curvature tensor**.

$$Ric(\xi, \eta) := \text{tr}[v \mapsto R_{x,v}\eta]. \quad (3.25)$$

Ricci curvature plays a key role in general relativity because it is the Ricci curvature rather than the full Riemann curvature which enters into the Einstein field equations.

3.14 Bi-invariant metrics on a Lie group.

The simplest example of a Riemann manifold is Euclidean space, where the geodesics are straight lines and all curvatures vanish. We may think of Euclidean space as a commutative Lie group under addition, and view the straight lines as translates of one parameter subgroups (lines through the origin). An easy but important generalization of this is when we consider bi-invariant metrics on a Lie group, a concept we shall explain below. In this case also, the geodesics are the translates of one parameter subgroups.

3.14.1 The Lie algebra of a Lie group.

Let G be a Lie group. This means that G is a group, and is a smooth manifold such that the multiplication map $G \times G \rightarrow G$ is smooth, as is the map $\text{inv} : G \rightarrow G$ sending every element into its inverse:

$$\text{inv} : a \mapsto a^{-1}, \quad a \in G.$$

Until now the Lie groups we studied were given as subgroups of $GL(n)$. We can continue in this vein, or work with the more general definition just given. We have the left action of G on itself

$$L_a : G \rightarrow G, \quad b \mapsto ab$$

and the right action

$$R_a : G \rightarrow G, \quad b \mapsto ba^{-1}.$$

We let \mathfrak{g} denote the tangent space to G at the identity:

$$\mathfrak{g} = TG_e.$$

We identify \mathfrak{g} with the space of all left invariant vector fields on G , so $\xi \in \mathfrak{g}$ is identified with the vector field X which assigns to every $a \in G$ the tangent vector

$$d(L_a)_e \xi \in TG_a.$$

We will alternatively use the notation X, Y or ξ, η for elements of \mathfrak{g} .

The left invariant vector field X generates a one parameter group of transformations which commutes with all left multiplications and so must consist of a one parameter group of right multiplications. In the case of a subgroup of $GL(n)$, where \mathfrak{g} was identified with a subspace of the space of all $n \times n$ matrices, we saw that this was the one parameter group of transformations

$$A \mapsto A \exp tX,$$

i.e. the one parameter group

$$R_{\exp -tX}.$$

So we might as well use this notation in general: $\exp tX$ denotes the one parameter subgroup of G obtained by looking at the solution curve through e of the left

invariant vector field X , and then the one parameter group of transformations generated by the vector field X is $R_{\exp -tX}$.

Let X and Y be elements of \mathfrak{g} thought of as left invariant vector fields, and let us compute their Lie bracket as vector fields. So let

$$\phi_t = R_{\exp -tX}$$

be the one parameter group of transformations generated by X . According to the general definition, the Lie bracket $[X, Y]$ is obtained by differentiating the time dependent vector field $\phi_t^* Y$ at $t = 0$. By definition, the pull-back $\phi_t^* Y$ is the vector field which assigns to the point a the tangent vector

$$(d\phi_t)_a^{-1} Y(\phi_t(a)) = (dR_{\exp tX})_a Y(a(\exp tX)). \quad (3.26)$$

In the case that G is a subgroup of the general linear group, this is precisely the left invariant vector field

$$a \mapsto a(e \exp tX) Y(\exp -tX).$$

Differentiating with respect to t and setting $t = 0$ shows that the vector field $[X, Y]$ is precisely the left invariant vector field corresponding to the commutator of the two matrices X and Y .

We can mimic this computation for a general Lie group, not necessarily given as a subgroup of $Gl(n)$: First let us record the special case of (3.26) when we take $a = e$:

$$(d\phi_t)_e^{-1} Y(\phi_t(e)) = (dR_{\exp tX})_Y(\exp tX). \quad (3.27)$$

For any $a \in G$ we let A_a denote conjugation by the element $a \in G$, so

$$A_a : G \rightarrow G, A_a(b) = aba^{-1}.$$

We have $A_a(e) = e$ and A_a carries one-parameter subgroups into one parameter subgroups. In particular the differential of A_a at $TG_e = \mathfrak{g}$ is a linear transformation of \mathfrak{g} which we shall denote by Ad_a :

$$d(A_a)_e =: \text{Ad}_a : TG_e \rightarrow TG_e.$$

We have

$$A_a = L_a \circ R_a = R_a \circ L_a.$$

So if Y is the left invariant vector field on G corresponding to $\eta \in TG_e = \mathfrak{g}$, we have $dL_a(\eta) = Y(a)$ and so

$$d(A_a)_e \eta = d(R_a)_a \circ d(L_a)_e \eta = d(R_a)_a Y(a).$$

Set $a = \exp tX$, and compare this with (3.27). Differentiate with respect to t and set $t = 0$. We see that the left invariant vector field $[X, Y]$ corresponds to the element of TG_e obtained by differentiating $\text{Ad}_{\exp tX} \eta$ with respect to t and setting $t = 0$. In symbols, we can write this as

$$\frac{d}{dt} \text{Ad}_{\exp tX} \Big|_{t=0} = \text{ad}(X) \quad \text{where} \quad \text{ad}(X) : \mathfrak{g} \rightarrow \mathfrak{g}, \quad \text{ad}(X)Y = [X, Y]. \quad (3.28)$$

Now $\text{ad}(X)$ as defined above is a linear transformation of \mathfrak{g} . So we can consider the corresponding one parameter group $\exp t \text{ad}(X)$ of linear transformations of \mathfrak{g} (using the usual formula for the exponential of a matrix). But (3.28) says that $\text{Ad}_{\exp tX}$ is a one parameter group of linear transformations with the same derivative, $\text{ad}(X)$ at $t = 0$. The uniqueness theorem for linear differential equations then implies the important formula

$$\exp(t \text{ad}(X)) = \text{Ad}_{\exp tX}. \quad (3.29)$$

3.14.2 The general Maurer-Cartan form.

If $v \in TG_a$ is tangent vector at the point $a \in G$, there will be a unique left invariant vector field X such that $X(a) = v$. In other words, there is a linear map

$$\omega_a : TG_a \rightarrow \mathfrak{g}$$

sending the tangent vector v to the element $\xi = \omega_a(v) \in \mathfrak{g}$ where the left invariant vector field X corresponding to ξ satisfies $X(a) = v$. So we have defined a \mathfrak{g} valued linear differential form ω identified the tangent space at any $a \in G$ with \mathfrak{g} . If

$$dL_b v = w \in TG_{ba}$$

then $X(ba) = w$ since $X(v) = v$ and X is left invariant. In other words,

$$\omega_{L_b a} \circ dL_b = \omega_a,$$

or, what amounts to the same thing

$$L_b^* \omega = \omega$$

for all $b \in G$. The form ω is left invariant. When we proved this for a subgroup of $Gl(n)$ this was a computation. But in the general case, as we have just seen, it is a tautology. We now want to establish the generalization of the Maurer-Cartan equation (2.9) which said that for subgroups of $Gl(n)$ we have

$$d\omega + \omega \wedge \omega = 0.$$

Since we no longer have, in general, the notion of matrix multiplication which enters into the definition of $\omega \wedge \omega$, we must first rewrite $\omega \wedge \omega$ in a form which generalizes to an arbitrary Lie group.

So let us temporarily consider the case of a subgroup of $Gl(n)$. Recall that for any two form τ and a pair of vector fields X and Y we write $\tau(X, Y) = i(Y)i(X)\tau$. Thus

$$(\omega \wedge \omega)(X, Y) = \omega(X)\omega(Y) - \omega(Y)\omega(X),$$

the commutator of the two matrix valued functions, $\omega(X)$ and $\omega(Y)$. Consider the commutator of two matrix valued one forms, ω and σ ,

$$\omega \wedge \sigma + \sigma \wedge \omega$$

(according to our usual rules of superalgebra). We denote this by

$$[\omega \wedge, \sigma].$$

In particular we may take $\omega = \sigma$ to obtain

$$[\omega \wedge, \omega] = 2\omega \wedge \omega.$$

So we can rewrite the Maurer-Cartan equation for a subgroup of $Gl(n)$ as

$$d\omega + \frac{1}{2}[\omega \wedge, \omega] = 0. \quad (3.30)$$

Now for a general Lie group we *do* have the Lie bracket map

$$\mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}.$$

So we can define the two form $[\omega \wedge, \omega]$. It is a \mathfrak{g} valued two form which satisfies

$$i(X)[\omega \wedge, \omega] = [X, \omega] - [\omega, X]$$

for any left invariant vector field X . Hence

$$\begin{aligned} [\omega \wedge, \omega](X, Y) &:= i(Y)i(X)[\omega \wedge, \omega] = i(Y)([X, \omega] - [\omega, X]) \\ &= [X, Y] - [Y, X] = 2[X, Y] \end{aligned}$$

for any pair of left invariant vector fields X and Y . So to prove (3.30) in general, we must verify that for any pair of left invariant vector fields we have

$$d\omega(X, Y) = -\omega([X, Y]).$$

But this is a consequence of our general formula (2.3) for the exterior derivative which in our case says that

$$d\omega(X, Y) = X\omega(Y) - Y\omega(X) - \omega([X, Y]).$$

In our situation the first two terms on the right vanish since, for example, $\omega(Y) = Y = \eta$ a constant element of \mathfrak{g} so that $X\omega(Y) = 0$ and similarly $Y\omega(X) = 0$.

3.14.3 Left invariant and bi-invariant metrics.

Any non-degenerate scalar product, $\langle \cdot, \cdot \rangle$, on \mathfrak{g} determines (and is equivalent to) a left invariant semi-Riemann metric on G via the left-identification $dL_a : \mathfrak{g} = TG_e \rightarrow TG_a, \forall a \in G$,

Since $A_a = L_a \circ R_a$, the left invariant metric, $\langle \cdot, \cdot \rangle$ is right invariant if and only if it is A_a invariant for all $a \in G$, which is the same as saying that $\langle \cdot, \cdot \rangle$ is invariant under the adjoint representation of G on \mathfrak{g} , i.e. that

$$\langle Ad_a Y, Ad_a Z \rangle = \langle Y, Z \rangle, \quad \forall Y, Z \in \mathfrak{g}, a \in G.$$

Setting $a = \exp tX$, $X \in \mathfrak{g}$, differentiating with respect to t and setting $t = 0$ gives

$$\langle [X, Y], Z \rangle + \langle Y, [X, Z] \rangle = 0, \quad \forall X, Y, Z \in \mathfrak{g}. \quad (3.31)$$

If every element of G can be written as a product of elements of the form $\exp \xi$, $\xi \in \mathfrak{g}$ (which will be the case if G is connected), this condition implies that $\langle \cdot, \cdot \rangle$ is invariant under Ad and hence is invariant under right and left multiplication. Such a metric is called **bi-invariant**.

Let inv denote the map sending every element into its inverse:

$$\text{inv} : a \mapsto a^{-1}, \quad a \in G.$$

Since $\text{inv} \exp tX = \exp(-tX)$ we see that

$$d \text{inv}_e = -\text{id}.$$

Also

$$\text{inv} = R_{a^{-1}} \circ \text{inv} \circ L_{a^{-1}}$$

since the right hand side sends $b \in G$ into

$$b \mapsto a^{-1}b \mapsto b^{-1}a \mapsto b^{-1}.$$

Hence $d \text{inv}_a : TG_a \rightarrow TG_{a^{-1}}$ is given, by the chain rule, as

$$dR_{a^{-1}} \circ d \text{inv}_e \circ dL_{a^{-1}} = -dR_{a^{-1}} \circ dL_{a^{-1}}$$

implying that a bi-invariant metric is invariant under the map inv . Conversely, if a left invariant metric is invariant under inv then it is also right invariant, hence bi-invariant since

$$R_a = \text{inv} \circ L_a^{-1} \circ \text{inv}.$$

3.14.4 Geodesics are cosets of one parameter subgroups.

The Koszul formula simplifies considerably when applied to left invariant vector fields and bi-invariant metrics since all scalar products are constant, so their derivatives vanish, and we are left with

$$2\langle \nabla_X Y, Z \rangle = -\langle X, [Y, Z] \rangle - \langle Y, [X, Z] \rangle + \langle Z, [X, Y] \rangle$$

and the first two terms cancel by (9.50). We are left with

$$\nabla_X Y = \frac{1}{2}[X, Y]. \quad (3.32)$$

Conversely, if $\langle \cdot, \cdot \rangle$ is a left invariant metric for which (9.51) holds, then

$$\begin{aligned} \langle X, [Y, Z] \rangle &= 2\langle X, \nabla_Y Z \rangle \\ &= -2\langle \nabla_Y X, Z \rangle \\ &= -\langle [Y, X], Z \rangle \\ &= \langle [X, Y], Z \rangle \end{aligned}$$

so the metric is bi-invariant.

Let α be an integral curve of the left invariant vector field X . Equation (9.51) implies that $\alpha'' = \nabla_X X = 0$ so α is a geodesic. Thus the one-parameter groups are the geodesics through the identity, and all geodesics are left cosets of one parameter groups. (This is the reason for the name exponential map in Riemannian geometry which we shall study in Chapter V.)

In Chapter VIII we will study Riemannian submersions. It will emerge from this study that if we have a quotient space $B = G/H$ of a group with a bi-invariant metric (satisfying some mild conditions), then the geodesics on B in the induced metric are orbits of certain one parameter subgroups. For example, the geodesics on spheres are the great circles.

3.14.5 The Riemann curvature of a bi-invariant metric.

We compute the Riemann curvature of a bi-invariant metric by applying the definition (3.15) to left invariant vector fields:

$$R_{XY}Z = \frac{1}{4}[X, [Y, Z]] - \frac{1}{4}[Y, [X, Z]] - \frac{1}{2}[[X, Y], Z]$$

Jacobi's identity implies the first two terms add up to $\frac{1}{4}[[X, Y], Z]$ and so

$$R_{XY}Z = -\frac{1}{4}[[X, Y], Z]. \quad (3.33)$$

3.14.6 Sectional curvatures.

In particular

$$\langle R_{XY}X, Y \rangle = -\frac{1}{4}\langle [[X, Y], X], Y \rangle = -\frac{1}{4}\langle [X, Y], [X, Y] \rangle$$

so

$$K(X, Y) = \frac{1}{4} \frac{\| [X, Y] \|^2}{\| X \wedge Y \|^2}. \quad (3.34)$$

3.14.7 The Ricci curvature and the Killing form.

Recall that for each $X \in \mathfrak{g}$ the linear transformation of \mathfrak{g} consisting of bracketing on the left by X is called $\text{ad } X$. So

$$\text{ad } X : \mathfrak{g} \rightarrow \mathfrak{g}, \quad \text{ad } X(V) := [X, V].$$

We can thus write our formula for the curvature as

$$R_{XV}Y = \frac{1}{4}(\text{ad } Y)(\text{ad } X)V.$$

Now the Ricci curvature was defined as

$$\text{Ric}(X, Y) = \text{tr}[V \mapsto R_{XV}Y].$$

We thus see that for any bi-invariant metric, the Ricci curvature is always given by

$$\text{Ric} = \frac{1}{4}B \quad (3.35)$$

where B , the **Killing form**, is defined by

$$B(X, Y) := \text{tr}(\text{ad } X)(\text{ad } Y). \quad (3.36)$$

The Killing form is symmetric, since $\text{tr}(AC) = \text{tr}CA$ for any pair of linear operators. It is also invariant. Indeed, let $\mu : \mathfrak{g} \rightarrow \mathfrak{g}$ be any automorphism of \mathfrak{g} , so $\mu([X, Y]) = [\mu(X), \mu(Y)]$ for all $X, Y \in \mathfrak{g}$. We can read this equation as saying

$$\text{ad}(\mu(X))(\mu(Y)) = \mu(\text{ad}(X)(Y))$$

or

$$\text{ad}(\mu(X)) = \mu \circ \text{ad } X \mu^{-1}.$$

Hence

$$\text{ad}(\mu(X))\text{ad}(\mu(Y)) = \mu \circ \text{ad } X \text{ad } Y \mu^{-1}.$$

Since trace is invariant under conjugation, it follows that

$$B(\mu(X), \mu(Y)) = B(X, Y).$$

Applied to $\mu = \exp(t\text{ad } Z)$ and differentiating at $t = 0$ shows that $B([Z, X], Y) + B(X, [Z, Y]) = 0$.

So the Killing form defines a bi-invariant symmetric bilinear form on G . Of course it need not, in general, be non-degenerate. For example, if the group is commutative, it vanishes identically. A group G is called *semi-simple* if its Killing form is non-degenerate. So on a semi-simple Lie group, we can always choose the Killing form as the bi-invariant metric. For such a choice, our formula above for the Ricci curvature then shows that the group manifold with this metric is **Einstein**, i.e. the Ricci curvature is a multiple of the scalar product.

Suppose that the adjoint representation of G on \mathfrak{g} is irreducible. Then \mathfrak{g} can not have two invariant non-degenerate scalar products unless one is a multiple of the other. In this case, we can also conclude from our formula that the group manifold is Einstein.

3.14.8 Bi-invariant forms from representations.

Here is a way to construct invariant scalar products on a Lie algebra \mathfrak{g} of a Lie group G . Let ρ be a representation of G . This means that ρ is a smooth homomorphism of G into $Gl(n, \mathbf{R})$ or $Gl(n, \mathbf{C})$. This induces a representation $\dot{\rho}$ of \mathfrak{g} by

$$\dot{\rho}(X) := \frac{d}{dt}\rho(\exp tX)|_{t=0}.$$

So

$$\dot{\rho} : \mathfrak{g} \rightarrow gl(n)$$

where $gl(n)$ is the Lie algebra of $Gl(n)$, and

$$\dot{\rho}(X.Y) = [\dot{\rho}(X), \dot{\rho}(Y)]$$

where the bracket on the right is in $gl(n)$. More generally, a linear map $\dot{\rho} : \mathfrak{g} \rightarrow gl(n, \mathbf{C})$ or $gl(n, \mathbf{R})$ satisfying the above identity is called a representation of the Lie algebra \mathfrak{g} . Every representation of G gives rise to a representation of \mathfrak{g} but not every representation of \mathfrak{g} need come from a representation of G in general.

If $\dot{\rho}$ is a representation of \mathfrak{g} , with values in $gl(n, \mathbf{R})$, we may define

$$\langle X, Y \rangle_{\mathfrak{g}} := \text{tr } \dot{\rho}(X)\dot{\rho}(Y).$$

This is real valued, symmetric in X and Y , and

$$\langle [X, Y], Z \rangle_{\mathfrak{g}} + \langle Y, X, Z \rangle_{\mathfrak{g}} =$$

$$\text{tr } (\dot{\rho}(X)\dot{\rho}(Y)\dot{\rho}(Z) - \dot{\rho}(Y)\dot{\rho}(X)\dot{\rho}(Z) + \dot{\rho}(Y)\dot{\rho}(X)\dot{\rho}(Z) - \dot{\rho}(Y)\dot{\rho}(Z)\dot{\rho}(X)) = 0.$$

So this is invariant. Of course it need not be non-degenerate.

A case of particular interest is when the representation $\dot{\rho}$ takes values in $u(n)$, the Lie algebra of the unitary group. An element of $u(n)$ is a skew adjoint matrix, i.e. a matrix of the form iA where $A = A^*$ is self adjoint. If $A = A^*$ and $A = (a_{ij})$ then

$$\text{tr } A^2 = \text{tr } AA^* = \sum_{i,j} a_{ij}a_{ji} = \sum_{i,j} a_{ij}\overline{a_{ij}} = \sum_{ij} |a_{ij}|^2$$

which is positive unless $A = 0$. So

$$- \text{tr}(iA)(iA)$$

is positive unless $A = 0$. This implies that if $\dot{\rho} : \mathfrak{g} \rightarrow u(n)$ is injective, then the form

$$\langle X, Y \rangle = - \text{tr } \dot{\rho}(X)\dot{\rho}(Y)$$

is a positive definite invariant scalar product on \mathfrak{g} .

For example, let us consider the Lie algebra $\mathfrak{g} = u(2)$ and the representation $\dot{\rho}$ of \mathfrak{g} on the exterior algebra of \mathbf{C}^2 . We may decompose

$$\wedge(\mathbf{C}^2) = \wedge^0(\mathbf{C}^2) \oplus \wedge^1(\mathbf{C}^2) \oplus \wedge^2(\mathbf{C}^2)$$

and each of the summands is invariant under our representation. Every element of $u(2)$ acts trivially on $\wedge^0(\mathbf{C}^2)$ and acts in its standard fashion on $\wedge^1(\mathbf{C}^2) = \mathbf{C}^2$. Every element of $u(2)$ acts via multiplication by its trace on $\wedge^2(\mathbf{C}^2)$ so in particular all elements of $su(2)$ act trivially there. Thus restricted to $su(2)$, the induced scalar product is just

$$\langle X, Y \rangle = - \text{tr } XY, \quad X, Y \in su(2),$$

while on scalar matrices, i.e. matrices of the form $S = riI$ we have

$$\langle S, S \rangle = - \text{tr } \dot{\rho}(S)^2 = 2r^2 + (2r)^2 = 6r^2 = -3 \text{tr } S^2 = -\frac{3}{2}(\text{tr } S)^2.$$

3.14.9 The Weinberg angle.

The preceding example illustrates the fact that if the adjoint representation of \mathfrak{g} is not irreducible, there may be more than a one parameter family of invariant scalar products on \mathfrak{g} . Indeed the algebra $u(2)$ decomposes as a sum

$$u(2) = su(2) \oplus u(1)$$

of subalgebras, where $u(1)$ consists of the scalar matrices (which commute with all elements of $u(2)$). It follows from the invariance condition that $u(1)$ must be orthogonal to $su(2)$ under any invariant scalar product. Each of these summands is irreducible under the adjoint representation, so the restriction of any invariant scalar product to each summand is determined up to positive scalar multiple, but these multiples can be chosen independently for each summand. So there is a two parameter family of choices.

In the physics literature it is conventional to write the most general invariant scalar product on $u(2)$ as

$$\langle A, B \rangle = -\frac{2}{g_2^2} \operatorname{tr} \left(A - \frac{1}{2}(\operatorname{tr} A)I \right) \left(B - \frac{1}{2}(\operatorname{tr} B)I \right) + -\frac{1}{g_1^2} \operatorname{tr} A \operatorname{tr} B$$

where g_1 and g_2 are sometimes called “coupling strengths”. The first summand vanishes on $u(1)$ and the second summand vanishes on $su(2)$. The **Weinberg angle** θ_W is defined by

$$\sin \theta_W := \frac{g_1}{\sqrt{g_1^2 + g_2^2}}$$

and plays a key role in Electro-Weak theory which unifies the electromagnetic and weak interactions. In the current state of knowledge, there is no broadly agreed theory that predicts the Weinberg angle. It is an input derived from experiment. The data as of July 2002 from the Particle Data Group gives a value of

$$\sin^2 \theta_W = 0.23113\dots$$

Notice that the computation that we did from the exterior algebra has

$$g_1^2 = \frac{2}{3} \quad \text{and} \quad g_2^2 = 2$$

so

$$\sin^2 \theta_W = \frac{\frac{2}{3}}{\frac{2}{3} + 2} = .25$$

Of course several quite different representations will give the same metric or Weinberg angle.

3.15 Frame fields.

By a frame field we mean an n -tuple $E = (E_1, \dots, E_n)$ of vector fields (defined on some neighborhood) whose values at every point form a basis of the tangent

space at that point. These then define a dual collection of differential forms

$$\theta = \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \end{pmatrix}$$

whose values at every point form the dual basis. For example, a coordinate system x^1, \dots, x^n provides the frame field $\partial_1, \dots, \partial_n$ with dual forms dx^1, \dots, dx^n . But the use of more general frame fields allows for flexibility in computation.

A frame field is called “**orthonormal**” if $\langle E_i, E_j \rangle \equiv 0$ for $i \neq j$ and $\langle E_i, E_i \rangle \equiv \epsilon_i$ where $\epsilon_i = \pm 1$. For example, applying the Gram-Schmidt procedure to an arbitrary frame field for a positive definite metric yields an orthonormal one.

3.16 Curvature tensors in a frame field.

In terms of a frame field the curvature tensor is given as

$$\sum R_{j k \ell}^i E_i \theta^k \theta^\ell \theta^j \quad \text{where} \quad R_{j k \ell}^i = \theta^i(R_{E_k E_\ell} E_j).$$

The Ricci tensor, which as we mentioned, plays a key role in general relativity, takes the form

$$\text{Ric} = \sum R_{ij} \theta^i \theta^j \quad \text{where} \quad R_{ij} = \text{Ric}(E_i, E_j) := \sum R_{imj}^m.$$

If the frame is orthonormal then for any pair of vector fields V, W we have

$$\text{Ric}(V, W) = \sum \epsilon_m \langle R_{V E_m} E_m, W \rangle.$$

A manifold is called **Ricci flat** if its Ricci curvature vanishes.

The **scalar curvature** S is defined as

$$S := \sum g^{ij} R_{ij}.$$

3.17 Frame fields and curvature forms.

Let M be a semi-Riemannian manifold. Let E_1, \dots, E_n be an “orthonormal” frame field defined on some open subset of M . (In order not to clutter up the notation we will not introduce a specific name for the domain of definition of our frame field.) This means the E_i are vector fields and

$$\langle E_i, E_j \rangle \equiv 0, \quad i \neq j$$

while

$$\langle E_i, E_i \rangle \equiv \epsilon_i, \quad \epsilon_i = \pm 1.$$

Thus $E_1(p), \dots, E_n(p)$ form an “orthonormal” basis of the tangent space TM_p at each point p in the domain of definition. The dual basis of the cotangent space then provides a family of linear differential forms, $\theta^1, \dots, \theta^n$. It follows from the definition, that if $v \in TM_p$ then

$$\langle v, v \rangle = \epsilon_1 (\theta^1(v))^2 + \dots + \epsilon_n (\theta^n(v))^2.$$

This equation, true at all points in the domain of definition of the frame field is usually written as

$$ds^2 = \epsilon_1 (\theta^1)^2 + \dots + \epsilon_n (\theta^n)^2. \quad (3.37)$$

Conversely, if $\theta^1, \dots, \theta^n$ is a collection of linear differential forms satisfying (3.37) (defined on some open set) then the dual vector fields constitute an “orthonormal” frame field.

On any manifold, we have the tautological tensor field of type (1,1) which assigns to each tangent space the identity linear transformation. We will denote this tautological tensor field by id . Thus for any $p \in M$ and any $v \in TM_p$,

$$\text{id}(v) = v.$$

In terms of a frame field we have

$$\text{id} = E_1 \otimes \theta^1 + \dots + E_n \otimes \theta^n$$

in the sense that both sides yield v when applied to any tangent vector v in the domain of definition of the frame field. We can say that the θ^i give the expression for id in terms of the frame field and also introduce the “vector of differential forms”

$$\theta := \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \end{pmatrix}$$

as a shorthand for the collection of the θ^i .

For each i the Levi-Civita connection yields a tensor field ∇E_i , the covariant differential of E_i with respect to the connection, and hence linear differential forms ω_j^i defined by

$$\omega_j^i(\xi) = \theta^i(\nabla_\xi E_j). \quad (3.38)$$

So

$$\nabla_\xi E_j = \sum_m \omega_j^m(\xi) E_m.$$

The **first structure equation of Cartan** asserts that

$$d\theta^i = - \sum_m \omega_m^i \wedge \theta^m. \quad (3.39)$$

To prove this, we apply the formula (2.3) which says that

$$d\theta(X, Y) = X(\theta(Y)) - Y(\theta(X)) - \theta([X, Y])$$

holds for any linear differential form θ and vector fields X and Y . We apply this to θ^i, E_a, E_b to obtain

$$d\theta^i(E_a, E_b) = E_a\theta^i(E_b) - E_b\theta^i(E_a) - \theta^i([E_a, E_b]).$$

Since $\theta^i(E_b)$ and $\theta^i(E_a) = 0$ or 1 are constants, the first two terms vanish and so the left hand side of (3.39) when evaluated on (E_a, E_b) becomes

$$-\theta^i([E_a, E_b]).$$

As to the right hand side we have

$$\begin{aligned} \left[-\sum \omega_m^i \wedge \theta^m\right](E_a, E_b) &= \left(i(E_a)\left[-\sum \omega_m^j \wedge \theta^m\right]\right)(E_b) \\ &= \left[-\sum \omega_m^i(E_a)\theta^m + \sum \theta^m(E_a)\omega_m^i\right](E_b) \\ &= -\omega_b^i(E_a) + \omega_a^i(E_b) \\ &= -\theta^i(\nabla_{E_a} E_b) - \nabla_{E_b} E_a \\ &= -\theta^i([E_a, E_b]). \quad \text{QED} \end{aligned}$$

Notice that

$$\omega_j^i(\xi) = \theta^i(\nabla_\xi E_j) = \epsilon_i \langle \nabla_\xi E_j, E_i \rangle.$$

Since

$$0 = d\langle E_i, E_j \rangle$$

we have

$$\epsilon_j \omega_i^j = -\epsilon_i \omega_j^i. \quad (3.40)$$

In particular $\omega_i^i = 0$. If we introduce the “matrix of linear differential forms”

$$\omega := (\omega_j^i)$$

we can write the first structural equations as

$$d\theta + \omega \wedge \theta = 0.$$

For tangent vectors $\xi, \eta \in TM_p$ let $(\Omega_j^i(\xi, \eta))$ be the matrix of the curvature operator $R_{\xi\eta}$ with respect to the basis $E_1(p), \dots, E_n(p)$. So

$$R_{\xi\eta}(E_j)(p) = \sum_i \Omega_j^i(\xi, \eta) E_i.$$

Since $R_{\eta, \xi} = -R_{\xi, \eta}$, $\Omega_j^i(\xi, \eta) = -\Omega_j^i(\eta, \xi)$ so the Ω_j^i are exterior differential forms of degree two.

Cartan’s **second structural equation** asserts that

$$\Omega_j^i = d\omega_j^i + \sum_m \omega_m^i \wedge \omega_j^m. \quad (3.41)$$

We have

$$R_{E_a E_b}(E_j) = \sum_i \Omega_j^i(E_a, E_b) E_i$$

by definition. We must show that the right hand side of (3.41) yields the same result when we substitute E_a, E_b into the differential forms, multiply by E_i and sum over i .

Write

$$R_{E_a E_b}(E_j) = \nabla_{E_a}(\nabla_{E_b} E_j) - \nabla_{E_b}(\nabla_{E_a} E_j) - \nabla_{[E_a, E_b]} E_j.$$

Since $\nabla_{E_b}(E_j) = \sum_i \omega_j^i(E_b) E_i$ we get

$$\begin{aligned} \nabla_{E_a}(\nabla_{E_b} E_j) &= \sum E_a[\omega_j^i(E_b)] E_i + \sum \omega_j^m(E_b) \nabla_{E_a} E_m \\ &= \sum E_a[\omega_j^i(E_b)] E_i + \sum_{i,m} \omega_j^m(E_b) \omega_m^i(E_a) E_i \\ \nabla_{E_b}(\nabla_{E_a} E_j) &= \sum E_b[\omega_j^i(E_a)] E_i + \sum_{i,m} \omega_j^m(E_a) \omega_m^i(E_b) E_i \\ \nabla_{[E_a, E_b]} E_j &= \sum \omega_j^i([E_a, E_b]) E_j \quad \text{so} \\ R_{E_a E_b} E_j &= \sum_i [E_a \omega_j^i(E_b) - E_b \omega_j^i(E_a) - \omega_j^i([E_a, E_b])] E_i \\ &\quad + \sum_{m,i} [\omega_m^i(E_a) \omega_j^m(E_b) - \omega_m^i(E_b) \omega_j^m(E_a)] E_i. \end{aligned}$$

The first expression in square brackets is the value on E_a, E_b of $d\omega_j^i$ by (2.3) while the second expression in square brackets is the value on E_a, E_b of $\sum \omega_m^i \wedge \omega_j^m$. This proves Cartan's second structural equation.

We can write the two structural equations as

$$d\theta + \omega \wedge \theta = 0 \tag{3.42}$$

$$d\omega + \omega \wedge \omega = \Omega \tag{3.43}$$

3.18 Cartan's lemma.

We will show that the equations (3.42) and (3.40) determine the ω_j^i . First a result in exterior algebra:

Lemma 1 *Let x_1, \dots, x_p be linearly independent elements of a vector space, V , and suppose that $y_1, \dots, y_p \in V$ satisfy*

$$x_1 \wedge y_1 + \dots + x_p \wedge y_p = 0.$$

Then

$$y_j = \sum_{k=1}^p A_{jk} x_k \quad \text{with } A_{jk} = A_{kj}.$$

Proof. Choose x_{p+1}, \dots, x_n if $p < n$ so as to obtain a basis of V and write

$$y_i = \sum_{j=1}^p A_{ij}x_j + \sum_{k=p+1}^n B_{ik}x_k.$$

Substituting into the equation of the lemma gives

$$\sum_{i < j \leq p} (A_{ij} - A_{ji})x_i \wedge x_j + \sum_{i \leq p < k} B_{ik}x_i \wedge x_k = 0.$$

Since the $x_i \wedge x_\ell$, $i < \ell$ form a basis of $\wedge^2(V)$, we conclude that $B_{ik} = 0$ and $A_{ij} = A_{ji}$ which is the content of the lemma.

Suppose that ω and ω' are two matrices of one forms which satisfy (3.39). Then their difference, $\sigma := \omega - \omega'$ satisfies $\sigma \wedge \theta = 0$. Applying the lemma we conclude that

$$\sigma_k^i = \sum A_{jk}^i \theta^j, \quad A_{jk}^i = A_{kj}^i.$$

If we set

$$B_{jk}^i = \epsilon_i A_{jk}^i$$

and if both ω and ω' satisfy (3.40) so that σ does as well, then

$$B_{jk}^i = B_{kj}^i \quad \text{and} \quad B_{jk}^i = -B_{ki}^j.$$

We claim that these two equations imply that all the $B_{jk}^i = 0$ and hence that $\sigma = 0$. Indeed,

$$\begin{aligned} B_{jk}^i = B_{kj}^i &= -B_{ij}^k = B_{ki}^j \\ &= B_{ik}^j = -B_{jk}^i. \end{aligned}$$

The upshot is that if we have found ω satisfying (3.42) and (3.40) then we know that it is the matrix of connection forms.

3.19 Orthogonal coordinates on a surface.

If $n = 2$ there is only one independent linear differential form in ω namely

$$\omega_1^2 = -\epsilon_1 \epsilon_2 \omega_2^1.$$

Suppose that (u, v) are orthogonal coordinates on the surface which means that

$$\langle \partial_u, \partial_v \rangle \equiv 0.$$

Set $\mathbf{e} := \sqrt{|E|}$ and $\mathbf{g} := \sqrt{|G|}$ where

$$E := \langle \partial_u, \partial_u \rangle := \epsilon_1 \mathbf{e}^2, \quad G := \langle \partial_v, \partial_v \rangle := \epsilon_2 \mathbf{g}^2.$$

The frame field

$$E_1 := \frac{1}{\mathbf{e}} \partial_u, \quad E_2 := \frac{1}{\mathbf{g}} \partial_v$$

is “orthonormal” with dual frame given by

$$\theta^1 = \mathbf{e}du, \quad \theta^2 = \mathbf{g}dv.$$

Taking exterior derivatives yields

$$\begin{aligned} d\theta^1 &= \mathbf{e}_v dv \wedge du = -(\mathbf{e}_v/\mathbf{g})du \wedge \theta^2 \\ d\theta^2 &= \mathbf{g}_u du \wedge dv = -(\mathbf{g}_u/\mathbf{e})du \wedge \theta^1. \end{aligned}$$

Hence

$$\omega_2^1 = (\mathbf{e}_v/\mathbf{g})du - \epsilon_1\epsilon_2(\mathbf{g}_u/\mathbf{e})dv$$

by the uniqueness of the solution to the Cartan equations. In two dimensions the second structural equation reduces to

$$\Omega_2^1 = d\omega_2^1$$

and we compute

$$d\omega_2^1 = -[(\mathbf{e}_v/\mathbf{g})_v + \epsilon_1\epsilon_2(\mathbf{g}_u/\mathbf{e})_u]du \wedge dv = -\frac{1}{\mathbf{e}\mathbf{g}}[(\mathbf{e}_v/\mathbf{g})_v + \epsilon_1\epsilon_2(\mathbf{g}_u/\mathbf{e})_u]\theta^1 \wedge \theta^2.$$

The sectional curvature (=the Gaussian curvature) is then given by

$$K = \epsilon_1\Omega_2^1(E_1, E_2) = -\frac{1}{\mathbf{e}\mathbf{g}}[(\mathbf{e}_v/\mathbf{g})_v + \epsilon_1\epsilon_2(\mathbf{g}_u/\mathbf{e})_u]. \quad (3.44)$$

We obtained this formula in the positive definite case by much more complicated means in the first chapter.

Exercises. 1.

3.20 The curvature of the Schwartzschild metric

We use polar coordinates on space and t for time so coordinates t, r, ϑ, ϕ and introduce the shorthand notation

$$S := \sin \vartheta, \quad C := \cos \vartheta.$$

We fix a positive real number M and assume that

$$r > 2M.$$

The Schwartzschild metric is given as

$$\begin{aligned} ds^2 &= -(\theta^0)^2 + (\theta^1)^2 + (\theta^2)^2 + (\theta^3)^2 \quad \text{where} \\ \theta^0 &= \sqrt{h}dt, \quad h := 1 - \frac{2M}{r} \\ \theta^1 &= \frac{1}{\sqrt{h}}dr \\ \theta^2 &= r d\vartheta \\ \theta^3 &= rSd\phi \end{aligned}$$

1. Compute $d\sqrt{h}$ and then each of the $d\theta^i$, $i = 0, 1, 2, 3$.
2. Find the connection form matrix ω .
3. Find the curvature form matrix $\Omega = d\omega + \omega \wedge \omega$.
4. Show that the Schwarzschild metric is Ricci flat.
5. Find the sectional curvatures of the “coordinate planes”, i.e. the planes spanned by any two of $\partial_t, \partial_r, \partial_\vartheta, \partial_\phi$.
6. The space of the Scharzschild metric is the “twisted product” of the “Schwarzschild plane” \mathcal{B} spanned by r, t with the metric given by $-(\theta^0)^2 + (\theta^1)^2$ and the unit sphere \mathcal{S} in the sense that the metric has the form

$$\mathbf{g} = \mathbf{g}_B + r^2 \mathbf{g}_S.$$

From this fact alone (i.e. using none of the preceding computations) together with Koszul’s formula show that

$$\langle \nabla_X Y, Z \rangle = 0$$

if X and Y are vector fields on B and Z is a vector field on S (all thought of as vector fields on the full space).

Exercises 2.

3.21 Geodesics of the Schwarzschild metric.

The purpose of this problem set is to go through the details of two of the famous results general relativity, the explanation of the advance of the perihelion of Mercury and the deflection of light passing near the sun. (Einstein, 1915). In order to get results in useful form, we shall explicitly include Newton’s gravitational constant G

The equations for geodesics in a local coordinate system on a semi-Riemannian manifold are

$$\frac{d^2 x^k}{ds^2} + \sum_{i,j} \Gamma_{ij}^k \frac{dx^i}{ds} \frac{dx^j}{ds} = 0 \quad (3.45)$$

where

$$\Gamma_{ij}^k := \frac{1}{2} \sum_m g^{km} \left(\frac{\partial g_{jm}}{\partial x^i} + \frac{\partial g_{im}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^m} \right). \quad (3.46)$$

One of the postulates of general relativity is that a “small” particle will move along a geodesic in a four dimensional Lorentzian manifold whose metric is determined by the matter distribution over the manifold. Here the word

“small” is taken to mean that the effect of the mass of the particle on the metric itself can be ignored. We can ignore the mass of a planet when the metric is determined by the mass distribution of the stars. This notion of “small” or “passive” is similar to that involved in the equations of motion of a charged particle in an electromagnetic field. General electromagnetic theory says that the particle itself affects the electromagnetic field, but for “small” particles we ignore this and treat the particles as passively responding to the field. Similarly here. We will have a lot to say about the philosophical underpinnings of the postulate “small particles move along geodesics” when we have enough mathematical machinery. The theory also specifies that if the particle is massive then the geodesic is timelike, while if the particle has mass zero then the geodesic is a null geodesic, i.e. lightlike.

The second component of the theory is how the distribution of matter determines the metric. This is given by the Einstein field equations: Matter distribution is described by a (possibly degenerate) symmetric bilinear form on the tangent space at each point called the *stress energy tensor*, T . The Einstein equations take the form $\mathcal{G} = 8\pi T$ where \mathcal{G} is related to the Ricci curvature. In particular, in empty space, the Einstein equations become $\mathcal{G} = 0$.

Although the study of these equations is a huge enterprise, the solution for the equations $\mathcal{G} = 0$ in the exterior of a star of mass M which is “spherically symmetric”, “stationary” and tends to the Minkowski metric at large distances was found almost immediately by Schwarzschild. (The words in quotes need to be more carefully defined.) This is the metric

$$ds^2 := -h dt^2 + h^{-1} dr^2 + r^2 d\sigma^2 \quad (3.47)$$

where

$$h(r) := 1 - \frac{2GM}{r} \quad (3.48)$$

where G is Newton’s gravitational constant and $d\sigma^2$ is the invariant metric on the ordinary unit sphere,

$$d\sigma^2 = d\theta^2 + \sin^2 \theta d\phi^2. \quad (3.49)$$

To be more precise, let $P_I \subset \mathbf{R}^2$ consist of those pairs, (t, r) with

$$r > 2GM.$$

Let

$$N = P_I \times S^2,$$

the set of all (t, r, q) , $r > 2GM$, $q \in S^2$. The coordinates (θ, ϕ) can be used on the sphere with the north and south pole removed, and (3.49) is the local expression for the invariant metric of the unit sphere in terms of these coordinates. Then the metric we are considering on N is given by (3.47) as above.

Notice that the structure of N is like that of a surface of revolution, with the interval on the z -axis replaced by the two dimensional region, N , the circle replaced by the sphere, and the radius of revolution, f , replaced by r^2 . I

If we set $x^0 := t, x^1 := r, x^2 := \theta, x^3 := \phi$ then

$$g_{ij} = 0, \quad i \neq j \quad (3.50)$$

while

$$g_{00} = -h, g_{11} = h^{-1}, g_{22} = r^2, g_{33} = r^2 \sin^2 \theta.$$

Recall that a system of coordinates in which a metric satisfies (3.50) is called an orthogonal coordinate system. In such a coordinate system we have seen that the geodesic equations are

$$\frac{d}{ds} \left(g_{kk} \frac{dx^k}{ds} \right) = \frac{1}{2} \sum_j \frac{\partial g_{jj}}{\partial x^k} \left(\frac{dx^j}{ds} \right)^2. \quad (3.51)$$

1. Show that for the Schwarzschild metric, (3.47), the equation involving g_{22} on the left is

$$\frac{d}{ds} \left[r^2 \frac{d\theta}{ds} \right] = r^2 \sin \theta \cos \theta \left(\frac{d\phi}{ds} \right)^2.$$

Conclude from the uniqueness theorem for solutions of differential equations that if $\theta(0) = \pi/2, \dot{\theta}(0) = 0$ then $\theta(s) \equiv \pi/2$ along the whole geodesic. Conclude from rotational invariance that all geodesics must lie in a plane, i.e. by suitable choice of poles of the sphere we can arrange that $\theta \equiv \pi/2$.

2. With the above choice of spherical coordinates along the geodesic, show that the g_{00} and g_{33} equations become

$$h \frac{dt}{ds} = E$$

$$r^2 \frac{d\phi}{ds} = L$$

where E and L are constants. These constants are called the “energy” and the “angular momentum”. Notice that for $L > 0$, as we shall assume, $d\phi/ds > 0$, so we can use ϕ as a parameter on the orbit if we like.

General principles of mechanics imply that there is a “constant of motion” associated to every one parameter group of symmetries of the system. The Schwarzschild metric is invariant under time translations $t \mapsto t + c$ and under rotations $\phi \mapsto \phi + \alpha$. Under the general principles mentioned above, it turns out that E corresponds to time translation and that L corresponds to $\phi \mapsto \phi + \alpha$.

We now consider separately the case of a massive particle where we can choose the parameter s so that $\langle \gamma'(s), \gamma'(s) \rangle \equiv 1$ and massless particles for which $\langle \gamma'(s), \gamma'(s) \rangle \equiv 0$.

3.21.1 Massive particles.

We can write the tangent vector, $\gamma'(s)$ to the geodesic γ at the point s as

$$\gamma'(s) = \dot{x}_0(s) \left(\frac{\partial}{\partial x^0} \right)_{\gamma(s)} + \dot{x}_1(s) \left(\frac{\partial}{\partial x^1} \right)_{\gamma(s)} + \dot{x}_2(s) \left(\frac{\partial}{\partial x^2} \right)_{\gamma(s)} + \dot{x}_3(s) \left(\frac{\partial}{\partial x^3} \right)_{\gamma(s)}.$$

Let us assume that we use proper time as the parameterization of our geodesic so that

$$\langle \gamma'(s), \gamma'(s) \rangle_{\gamma(s)} \equiv -1.$$

vskip.2in

3. Using this last equation and the results of problem **2**, show that

$$E^2 = \left(\frac{dr}{ds} \right)^2 + \left(1 + \frac{L^2}{r^2} \right) h(r) \quad (3.52)$$

along any geodesic.

Orbit Types.

We can write (3.52) as

$$E^2 = \left(\frac{dr}{ds} \right)^2 + V(r) \quad (3.53)$$

where the *effective potential* V is given as

$$V(r) := 1 - \frac{2GM}{r} + \frac{L^2}{r^2} - \frac{2GML^2}{r^3}.$$

The behavior of the orbit depends on the the relative size of L and GM . In particular, (3.53) implies that on any orbit, r is restricted to an interval

$$I \subset \{r : V(r) \leq E^2\} \quad \text{such that} \quad r(0) \in I.$$

If we differentiate (3.53) we get

$$2 \left(\frac{d^2r}{ds^2} \right) \left(\frac{dr}{ds} \right) = -V'(r) \left(\frac{dr}{ds} \right). \quad (3.54)$$

In particular, a critical point of V , i.e. a point r_0 for which $V'(r_0) = 0$, gives rise to a circular orbit $r \equiv r_0$. If R is a non-critical point of V for which $V(R) = E^2$, then R is a turning point - the orbit reaches the end point R of the interval I and then turns around to move along I in the opposite direction.

Observe that $V(2GM) = 0$ and $V(r) \rightarrow 1$ as $r \rightarrow \infty$. To determine how V goes from 0 to 1 on $[2GM, \infty)$ we compute

$$V'(r) = \frac{2}{r^4} (GMr^2 - L^2r + 3GML^2) \quad (3.55)$$

and the quadratic polynomial in r given by the expression in parenthesis has discriminant

$$L^2(L^2 - 12G^2M^2).$$

If this discriminant is negative, there are no critical points, so V increase monotonically from 0 to ∞ . If this discriminant is positive there are two critical points, $r_1 < r_2$. Since $V'(2GM) > 0$, we see that r_1 a local maximum and r_2 a local minimum. (We will ignore the exceptional case of discriminant zero.) In the positive discriminant case we must distinguish between the cases where the local maximum at r_1 is not a global maximum, and when it is. Since $V(r) \rightarrow 1$ as $r \rightarrow \infty$ these two cases are distinguished by $V(r_1) < 1$ and $V(r_1) > 1$. Ignoring non-generic cases we thus can classify the behavior of $r(s)$ as:

- $L^2 < 12G^2M^2$ so V has no critical points and hence is monotone increasing on the interval $[2GM, \infty)$. The behavior of $r(s)$ for $s \geq 0$ subdivides into four cases, all leading to “crashing” (i.e. reaching the Schwarzschild boundary $2GM$ in finite s) or escape to infinity. The four possibilities have to do with the sign of $\dot{r}(0)$ and whether $E^2 < 1$ or $E^2 > 1$.
 1. $E^2 < 1$, $\dot{r}(0) < 0$. Since V decreases as r decreases, (3.53) implies that $\dot{r}s, \dot{r}(0) < 0$ for all $s > 0$ where it is defined. The particle crashes into the barrier at $2GM$ in finite time.
 2. $E^2 < 1$, $\dot{r} > 0$. The orbit initially moves in the direction of increasing r , reaches its maximum value where $V(r) = E^2$, turns around and crashes.
 3. $E^2 > 1$, $\dot{r} > 0$. The particle escapes to infinity.
 4. $E^2 > 1$, $\dot{r} < 0$. The particle crashes.
- $12G^2M^2 < L^2 < 16G^2M^2$. Here there are two critical points, but the maximum value at r_1 is < 1 . There are now four types of intervals I , depending on the value of E :
 1. $E^2 < V(r_1)$, $r < r_1$. Here the interval I lies to the left of the local maximum. The behavior will be like the first two cases above - “crash” if $\dot{r}(0) < 0$ and turn around then crash if $\dot{r} > 0$.
 2. $E^2 < V(r_1)$, $r > r_1$. the interval I now lies in a well to the right of r_1 , and so the value of r has two turning points corresponding to the end points of this interval. In other words the value of r is bounded along the entire orbit. We call this a **bound orbit**. In the “non-relativistic” approximation, this corresponds to Kepler’s ellipses. In problems 4 and 5 below we will examine more closely how this approximation works and derive Einstein’s famous calculation of the advance of the perihelion of Mercury.
 3. $V(r_1) < E^2 < 1$. The interval I is bounded on the right by the curve and extends all the way to the left (up to the barrier at $2GM$). The behavior is again either direct crash or turn around and then crash according to the sign of $\dot{r}(0)$,

4. $E^2 > 1$. Now the possible behaviors are “crash” if $\dot{r}(0) < 0$ or escape to infinity if $\dot{r}(0) > 0$.
- $L > 4M$. Now $V(r_1) > 1$. Again there will be four possible intervals:
 1. $E^2 < V(r_1)$, $r(0) < r_1$. This is an interval lying to the left of the “potential barrier” and so yield either a crash or turn around then crash orbit.
 2. $1 < E^2 < V(r_1)$. Now I lies to the right of the barrier, but below its peak, extending out to ∞ on the right. The orbit will escape to infinity if $\dot{r}(0) > 0$ or turn around and then escape if $\dot{r}(0) < 0$.
 3. $E^2 > 1$. The interval I extends from $2GM$ to infinity and the orbit is either crash or escape depending on the sign of $\dot{r}(0)$.
 4. $V(r_2) < E^2 < 1$. The interval now lies in a “well” to the right of the peak at r_1 . We have again a bound orbit.

We are interested in the bound orbits with $L > 0$. According to problem 2 we can use ϕ as a parameter on such an orbit and by the second equation in that problem we have

$$\dot{r} := \frac{dr}{ds} = \frac{dr/d\phi}{ds/d\phi} = \frac{L}{r^2} \frac{dr}{d\phi}.$$

Substituting this and the definition of h into (3.52) we get

$$E^2 = \frac{L^2}{r^4} \left(\frac{dr}{d\phi} \right)^2 + \left(1 + \frac{L^2}{r^2} \right) \left(1 - \frac{2GM}{r} \right).$$

It is now convenient to introduce the variable

$$u := \frac{1}{r}$$

instead of r . We have

$$\frac{du}{d\phi} = -\frac{1}{r^2} \left(\frac{dr}{d\phi} \right) = -u^2 \frac{dr}{d\phi}$$

so

$$E^2 = L^2 \left(\frac{du}{d\phi} \right)^2 + (1 + L^2 u^2)(1 - 2GMu). \quad (3.56)$$

We can rewrite this as

$$\left(\frac{du}{d\phi} \right)^2 = 2GMQ, \quad Q := u^3 - \frac{1}{2GM}u^2 + \beta_1 u + \beta_0 \quad (3.57)$$

where β_0 and β_1 are constants, combinations of E , L , and GM :

$$\beta_1 = \frac{1}{L^2}, \quad \beta_0 = \frac{1 - E^2}{2GML}.$$

Perihelion advance.

We will be interested in the case of bound orbits. In this case, a maximum value, u_1 along the orbit must be a root of the cubic polynomial, Q , as must be a minimum, u_2 , since these are turning points where the left hand side of (3.57) vanishes. Notice that these values do not depend on ϕ , being roots of a given polynomial with constant coefficients. Since two of the roots of Q are real, so is the third, and all three roots must add up to $\frac{1}{2GM}$, the negative of the coefficient of u^2 . Thus the third root is

$$\frac{1}{2GM} - u_1 - u_2.$$

We thus have

$$\left(\frac{du}{d\phi}\right)^2 = 2GM(u - u_1)(u - u_2)\left(u - \frac{1}{2GM} + u_1 + u_2\right).$$

Since the first factor on the right is non-positive and the second non-negative, the third is non-positive as the product must equal the non-negative expression on the left. Furthermore, we will be interested in the region where $r \gg 2GM$ so

$$2GM(u + u_1 + u_2) < 6GMu_1 \ll 1.$$

We therefore have the following expressions for $|d\phi/du|$:

$$\left|\frac{d\phi}{du}\right| = \frac{1}{\sqrt{(u_1 - u)(u - u_2)}} \cdot [1 - 2GM(u + u_1 + u_2)]^{-\frac{1}{2}} \quad (3.58)$$

$$\doteq \frac{1 + GM(u + u_1 + u_2)}{\sqrt{(u_1 - u)(u - u_2)}} \quad (3.59)$$

$$\doteq \frac{1}{\sqrt{(u_1 - u)(u - u_2)}} \quad (3.60)$$

Here (3.59) is obtained from (3.58) by ignoring terms which are quadratic or higher in $2GM(u + u_1 + u_2)$ and (3.60) is obtained from (3.58) by ignoring terms which are linear in $2GM(u + u_1 + u_2)$.

The strategy now is to observe that (3.60) is really the equation of an ellipse, whose Apollonian parameters, the latus rectum and the eccentricity, are expressed in terms of u_1 and u_2 . Then (3.59) is used to approximate the advance in the perihelion of Keplerian motion associate to this ellipse.

4. Show that the ellipse

$$u = \frac{1}{\ell}(1 + e \cos \phi)$$

is a solution of (3.60) where e and ℓ are determined from

$$u_1 = \frac{1}{\ell}(1 + e), \quad u_2 = \frac{1}{\ell}(1 - e)$$

so that the mean distance

$$a := \frac{1}{2} \left(\frac{1}{u_1} + \frac{1}{u_2} \right) = \frac{\ell}{1 - e^2}.$$

This is the approximating ellipse with the same maximum and minimum distance to the sun as the true orbit, if we choose our angular coordinate ϕ so that the x -axis is aligned with the axis of the ellipse.

In principle (3.58) is in solved form; if we integrate the right hand side from u_1 to u_2 and then back again, we will get the total change in ϕ across a complete cycle. Instead, we will approximate this integral by replacing (3.58) by (3.59) and then also make the approximate change of variables $u = \ell^{-1}(1 + e \cos \phi)$.

5. By making these approximations and substitutions show that the integral becomes

$$\int_0^{2\pi} [1 + GM\ell^{-1}(3 + \cos \phi)]d\phi = 2\pi + \frac{6\pi GM}{\ell}$$

so the perihelion advance in one revolution is

$$\frac{6\pi GM}{a(1 - e^2)}.$$

We have done these computations in units where the speed of light is one. If we are given the various constants in conventional units, say

$$G = 6.67 \times 10^{-11} \text{m}^3/\text{kg sec},$$

and the mass of the sun in kilograms

$$M = 1.99 \times 10^{30} \text{kg}$$

we must replace G by G/c^2 where c is the speed of light, $c = 3 \times 10^8$ m/sec. Then $2GM/c^2 \doteq 1.5\text{km}$. We may divide by the period of the planet to get the rate of advance as

$$\frac{6\pi GM}{c^a(1 - e^2)T}.$$

If we substitute, for Mercury, the mean distance $a = 5.768 \times 10^{10}\text{m}$, eccentricity $e = 0.206$ and period $T = 88$ days, and use the conversions

$$\begin{aligned} \text{century} &= 36524 \text{ days} \\ \text{radian} &= [360/2\pi]\text{degrees} \\ \text{degree} &= 3600'' \end{aligned}$$

we get the famous value of $43.1''/\text{century}$ for the advance of the perihelion of Mercury. This advance had been observed in the middle of the last century.

Up until recently, this observational verification of general relativity was not conclusive. The reason is that Newton's theory is based on the assumption that the mass of the sun is concentrated at a point. A famous theorem of Newton says that the attraction due to a homogeneous ball (on a particle outside) is the same as if all the mass is concentrated at a point. But if the sun is not a perfect sphere, or if its mass is not uniformly distributed, one would expect some deviation from Kepler's laws. The small effect of the advance of the perihelion of Mercury might have an explanation in terms of Newtonian mechanics. In the recent years, measurements from pulsars indicate large perihelion advances of the order of degrees per year (instead of arc seconds per century) yielding a striking confirmation of Einstein's theory.

3.21.2 Massless particles.

We now have

$$\gamma'(s) = \dot{x}_0(s) \left(\frac{\partial}{\partial x^0} \right)_{\gamma(s)} + \dot{x}_1(s) \left(\frac{\partial}{\partial x^1} \right)_{\gamma(s)} + \dot{x}_2(s) \left(\frac{\partial}{\partial x^2} \right)_{\gamma(s)} + \dot{x}_3(s) \left(\frac{\partial}{\partial x^3} \right)_{\gamma(s)}.$$

$$\langle \gamma'(s), \gamma'(s) \rangle_{\gamma(s)} \equiv 0.$$

6. Using problem 2 verify that

$$E^2 = \left(\frac{dr}{ds} \right)^2 + \left(\frac{L^2}{r^2} \right) h$$

and then

$$\frac{d^2 u}{d\phi^2} + u = 3GMu^2. \quad (3.61)$$

We will be interested in orbits which go out to infinity in both directions. For large values of r , the right hand side is negligibly small, so we should compare (3.61) with

$$\frac{d^2 u_0}{d\phi^2} + u_0 = 0$$

whose solutions are

$$u_0 = a \cos \phi + b \sin \phi$$

or

$$1 = ax + by, \quad x = r \cos \phi, \quad y = r \sin \phi,$$

in other words straight lines. We might as well choose our angular coordinate ϕ so that this straight line is parallel to the y -axis, i.e.

$$u_0 = r_0^{-1} \cos \phi$$

where r_0 is the distance of closest approach to the origin. Suppose we are interested in light rays passing the sun. The radius of the sun is about 7×10^5 km while $2GM$ is about 1.5km. Hence in units where r_0 is of order 1 the expression $3GM$ is a very small quantity, call it ϵ . So write our equation as

$$u'' + u = \epsilon u^2, \quad \epsilon = 3GM. \quad (3.62)$$

We solve this by the method of perturbation theory: look for a solution of the form

$$u = u_0 + \epsilon v + \dots$$

where the error is of order ϵ^2 . We choose u_0 as above to solve the equation obtained by equating the zero-th order terms in ϵ .

7. Compare coefficients of ϵ to obtain the equation

$$v'' + v = \frac{1}{2r_0^2}(1 + \cos 2\phi)$$

and try a solution of the form $v = a + b \cos 2\phi$ to find the solution of this equation and so obtain the first order approximation

$$u = \frac{1}{r_0} \cos \phi - \frac{\epsilon}{3r_0^2} \cos^2 \phi + \frac{2\epsilon}{3r_0^2} \quad (3.63)$$

to (3.62).

The asymptotes as $r \rightarrow \infty$ or $u \rightarrow 0$ will be straight lines with angles obtained by setting $u = 0$ in (3.63). This gives a quadratic equation for $\cos \phi$.

8. Remembering that cosine must be ≤ 1 show that up through order ϵ we have

$$\cos \phi = -\frac{2\epsilon}{3r_0} = -\frac{2GM}{r_0}.$$

Writing $\phi = \pi/2 + \delta$ this gives $\sin \delta = 2GM/r_0$ or approximately $\delta = 2GM/r_0$. This was for one asymptote. The same calculation gives the same result for the other asymptote. Adding the two and passing to conventional units gives

$$\Delta = \frac{4GM}{c^2 r_0} \quad (3.64)$$

for the deflection. For light just grazing the sun this predicts a deflection of $1.75''$. This was approximately observed in the expedition to the solar eclipse of 1919.

Recent, remarkable, photographs from the Hubble space telescope have given strong confirmation to Einstein's theory from the deflection of light by dark matter.

Chapter 4

The bundle of frames.

4.1 Connection and curvature forms in a frame field.

Let $E = (E_1, \dots, E_n)$ be a(n orthonormal) frame field and

$$\theta = \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \end{pmatrix}$$

the dual frame field so

$$\text{id} = E_1 \otimes \theta^1 + \dots + E_n \otimes \theta^n$$

or

$$\text{id} = (E_1, \dots, E_n) \otimes \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \end{pmatrix}$$

where id is the tautological tensor field which assigns the identity map to each tangent space. We write this more succinctly as

$$\text{id} = E\theta.$$

The (matrix of) connection form(s) in terms of the frame field is then determined by

$$d\theta + \omega \wedge \theta = 0$$

and the curvature by

$$d\omega + \omega \wedge \omega = \Omega.$$

We now repeat an argument that we gave when discussing the general Maurer Cartan form: Recall that for any two form τ and a pair of vector fields X and Y we write $\tau(X, Y) = i(Y)i(X)\tau$. Thus

$$(\omega \wedge \omega)(X, Y) = \omega(X)\omega(Y) - \omega(Y)\omega(X),$$

the commutator of the two matrix valued functions, $\omega(X)$ and $\omega(Y)$. Consider the commutator of two matrix valued one forms, ω and σ ,

$$\omega \wedge \sigma + \sigma \wedge \omega$$

(according to our usual rules of superalgebra). We denote this by

$$[\omega \wedge, \sigma].$$

In particular we may take $\omega = \sigma$ to obtain

$$[\omega \wedge, \omega] = 2\omega \wedge \omega.$$

We can thus also write the curvature as

$$\Omega = d\omega + \frac{1}{2}[\omega \wedge, \omega].$$

This way of writing the curvature has useful generalizations when we want to study connections on principal bundles later on in this chapter.

4.2 Change of frame field.

Suppose that E' is a second frame field whose domain of definition overlaps with the domain of definition of E . On the intersection of their domains of definition we must have

$$E' = EC$$

is another frame field where C is a(n orthogonal) matrix valued function. Let θ' be the dual frame field of E' . On the common domain of definition we have

$$EC\theta' = E'\theta' = \text{id} = E\theta$$

so

$$\theta = C\theta'.$$

Let ω' be the connection form associated to θ' , so ω' is determined (using Cartan's lemma) by the anti-symmetry condition and

$$d\theta' + \omega' \wedge \theta' = 0.$$

Then

$$d\theta = d(C\theta') = dC \wedge \theta' + Cd\theta' = dCC^{-1} \wedge \theta - C\omega' C^{-1} \wedge \theta$$

implying that

$$\omega = -dCC^{-1} + C\omega' C^{-1}$$

or

$$\omega' = C^{-1}\omega C + C^{-1}dC. \quad (4.1)$$

We have

$$\omega' \wedge \omega' = C^{-1}\omega \wedge \omega C + C^{-1}\omega \wedge dC + C^{-1}dCC^{-1} \wedge \omega C + C^{-1}dC \wedge C^{-1}dC$$

while

$$d\omega' = d(C^{-1}) \wedge \omega C + C^{-1}d\omega C - C^{-1}\omega \wedge dC + d(C^{-1}) \wedge dC.$$

Now it follows from

$$C^{-1}C \equiv I$$

that

$$d(C^{-1}) = -C^{-1}dCC^{-1}$$

and hence from the expression

$$\Omega' = \omega' \wedge \omega' + d\omega'$$

we get

$$\Omega' = C^{-1}\Omega C. \quad (4.2)$$

Notice that this equation contains the assertion that the curvature is a tensor. Indeed, recall that for any pair of tangent vectors $\xi, \eta \in TM_p$ the matrix $\Omega(\xi, \eta)$ gives the matrix of the operator $R_{\xi\eta} : TM_p \rightarrow TM_p$ relative to the orthonormal basis $E_1(p), \dots, E_n(p)$. Let $\zeta \in TM_p$ be a tangent vector at p and let z^i be the coordinates of ζ relative to this basis so $\zeta = z^1 E_1 + \dots + z^n E_n$ which we can write as

$$\zeta = E(p)z \quad \text{where} \quad z = \begin{pmatrix} z^1 \\ \vdots \\ z^n \end{pmatrix}.$$

Then

$$R_{\xi\eta}\zeta = E(p)\Omega(\xi, \eta)z.$$

If we use a different frame field $E' = EC$ then $\zeta = E'(p)z'$ where $z' = C^{-1}(p)z$. Equation (4.2) implies that

$$\Omega'(\xi, \eta)z' = C^{-1}(p)\Omega(\xi, \eta)z$$

which shows that

$$E'(p)\Omega'(\xi, \eta)z' = E(p)\Omega(\xi, \eta)z.$$

Thus the transformation $\zeta \mapsto E(p)\Omega(\xi, \eta)z$ is a well defined linear transformation. So if we did not yet know that $R_{\xi\eta}$ is a well defined linear transformation, we could conclude this fact from (4.2).

4.3 The bundle of frames.

We will now make a reinterpretation of the arguments of the preceding section which will have far reaching consequences. Let $\mathcal{O}(M)$ denote the set of all “orthonormal” bases of all TM_p . So a point, \mathcal{E} , of $\mathcal{O}(M)$ is an “orthonormal” basis of TM_p for some point $p \in M$, and we will denote this point by $\pi(\mathcal{E})$. So

$$\pi : \mathcal{O}(M) \rightarrow M, \quad \mathcal{E} \text{ is an o.n. basis of } TM_{\pi(\mathcal{E})}$$

assigns to each \mathcal{E} the point at which it is the orthonormal basis.

Suppose that E is a frame field defined on an open set $U \subset M$. If $p \in U$, and $\pi(\mathcal{E}) = p$, then there is a unique “orthogonal” matrix A such that

$$\mathcal{E} = E(p)A.$$

We will denote this matrix A by $\phi(\mathcal{E})$. (If we want to make the dependence on the frame field explicit, we will write ϕ_E instead of ϕ .) Thus

$$\mathcal{E} = E(\pi(\mathcal{E}))\phi(\mathcal{E}).$$

This gives an identification

$$\psi : \pi^{-1}(U) \rightarrow U \times G, \quad \psi(\mathcal{E}) = (\pi(\mathcal{E}), \phi(\mathcal{E})) \quad (4.3)$$

where G denotes the group of all “orthogonal” matrices. It follows from the definition that

$$\phi(\mathcal{E}B) = \phi(\mathcal{E})B, \quad \forall B \in G. \quad (4.4)$$

Let E' be a second frame field defined on an open set U' . We have a map

$$C : U \cap U' \rightarrow G$$

such that

$$E' = EC$$

as in the last section. Thus

$$\mathcal{E} = E\phi_E(\mathcal{E}) = EC(\pi(\mathcal{E}))\phi_{E'}(\mathcal{E})$$

so

$$\phi_E\phi_{E'}^{-1} = C \circ \pi. \quad (4.5)$$

This shows that the identifications given by (4.3) define, in a consistent way, a manifold structure on $\mathcal{O}(M)$. The manifold $\mathcal{O}(M)$ together with the action of the “orthogonal group” G by “multiplication on the right”

$$R_A : \mathcal{E} \mapsto \mathcal{E} \circ A^{-1}$$

and the differentiable map $\pi : \mathcal{O}(M) \rightarrow M$ is called the **bundle of (orthonormal) frames**.

We will now define forms

$$\vartheta = \begin{pmatrix} \vartheta^1 \\ \vdots \\ \vartheta^n \end{pmatrix} \quad \text{and} \quad \bar{\omega} = (\bar{\omega}_j^i)$$

on $\mathcal{O}(M)$:

4.3.1 The form ϑ .

Let $\xi \in T(\mathcal{O}(M))_{\mathcal{E}}$ be a tangent vector at the point $\mathcal{E} \in \mathcal{O}(M)$. Then $d\pi_{\mathcal{E}}(\xi)$ is a tangent vector to M at the point $\pi(\mathcal{E})$:

$$d\pi_{\mathcal{E}}(\xi) \in TM_{\pi(\mathcal{E})}.$$

As such, the vector $d\pi_{\mathcal{E}}(\xi)$ has coordinates relative to the basis, \mathcal{E} of $TM_{\pi(\mathcal{E})}$ and these coordinates depend linearly on ξ . So we may write

$$d\pi_{\mathcal{E}}(\xi) = \vartheta^1(\xi)\mathcal{E}_1 + \cdots + \vartheta^n(\xi)\mathcal{E}_n$$

defining the forms ϑ^i . As usual, we write this more succinctly as

$$d\pi = \mathcal{E}\vartheta.$$

4.3.2 The form ϑ in terms of a frame field.

Let $v \in T(\mathcal{O}(M))_{\mathcal{E}}$ be a tangent vector at the point $\mathcal{E} \in \mathcal{O}(M)$. Assume that $\pi(\mathcal{E})$ lies in the domain of definition of a frame field E and that $\mathcal{E} = E(p)A$ where $p = \pi(\mathcal{E})$. Let us write $d\pi(v)$ instead of $d\pi_{\mathcal{E}}(v)$ so as not to overburden the notation. We have

$$d\pi(v) = E(p)\theta(d\pi(v)) = \mathcal{E}\vartheta(v) = E(p)A\vartheta(v)$$

so

$$A\vartheta(v) = \theta(d\pi(v)).$$

We can write this as

$$\vartheta = \psi^* [A^{-1}\theta] \tag{4.6}$$

where

$$A^{-1}\theta$$

is the one form defined on $U \times G$ by

$$A^{-1}\theta(\eta + \zeta) = A^{-1}(\theta(\eta)), \quad \eta \in TM_x, \quad \zeta \in TG_A.$$

Here we have made the standard identification of $T(U \times G)_{(x,A)}$ as a direct sum,

$$T(U \times G)_{(x,A)} \sim TM_x \oplus TG_A,$$

valid on any product space.

4.3.3 The definition of $\bar{\omega}$.

Next we will define $\bar{\omega}$ in terms of the identification

$$\psi : \pi^{-1}(U) \rightarrow U \times G$$

given by a local frame field, and check that it satisfies

$$d\vartheta + \bar{\omega} \wedge \vartheta = 0, \quad \epsilon_i \bar{\omega}_j^i = -\epsilon_j \bar{\omega}_i^j.$$

By Cartan's lemma, this uniquely determines $\bar{\omega}$, so the definition must be independent of the choice of frame field, and so $\bar{\omega}$ is globally defined on $\mathcal{O}(M)$.

Let ω be the connection form (of the Levi-Civita connection) of the frame field E .

Define

$$\bar{\omega} := \psi^* [A^{-1}\omega A + A^{-1}dA] \quad (4.7)$$

where the expression in brackets on the right is a matrix valued one form defined on $U \times G$. Then on $U \times G$ we have

$$\begin{aligned} d[A^{-1}\theta] &= -A^{-1}dA \wedge A^{-1}\theta + A^{-1}d\theta \\ &= -A^{-1}dA \wedge A^{-1}\theta - A^{-1}\omega A \wedge A^{-1}\theta \text{ so} \\ 0 &= d[A^{-1}\theta] + [A^{-1}\omega A + A^{-1}dA] \wedge A^{-1}\theta. \end{aligned}$$

Applying ψ^* yields

$$d\vartheta + \bar{\omega} \wedge \vartheta = 0.$$

as desired. The antisymmetry condition says that ω takes values in the Lie algebra of G . Hence so does $A\omega A^{-1}$ for any $A \in G$. We also know that $A^{-1}dA$ takes values in the Lie algebra of G . Hence so does $\bar{\omega}$.

4.4 The connection form in a frame field as a pull-back.

We now have a reinterpretation of the connection form, ω , associated to a frame field. Indeed, the form $\bar{\omega}$ is a matrix valued linear differential form defined on all of $\mathcal{O}(M)$. A frame field, E , defined on an open set U , can be thought of as a map, $x \mapsto E(x)$ from U to $\mathcal{O}(M)$:

$$E : U \rightarrow \mathcal{O}(M), \quad x \mapsto E(x).$$

Then the pull-back of $\bar{\omega}$ under this map is exactly ω , the connection form associated to the frame field! In symbols

$$E^*\bar{\omega} = \omega.$$

To see this, observe that under the map $\psi : \pi^{-1}(U) \rightarrow U \times G$, we have $\psi(E(x)) = (x, I)$ where I is the identity matrix. Thus

$$\psi \circ E = (id, I)$$

where $id : U \rightarrow U$ is the identity map and I means the constant map sending every point x into the identity matrix. By the chain rule

$$\begin{aligned} E^*\bar{\omega} &= E^*\psi^* [A^{-1}\omega A + A^{-1}dA] \\ &= (\psi \circ E)^* [A^{-1}\omega A + A^{-1}dA] \\ &= \omega. \end{aligned}$$

Thus, for example, the frame field E is parallel relative to a vector field, X on M if and only if $\nabla_X(E) \equiv 0$ which is the same as

$$i(X)\omega \equiv 0$$

where ω is the connection form of the frame field. In view of the preceding result this is the same as

$$i[dE(X)]\bar{\omega} \equiv 0.$$

Here $dE(X)$ denotes the vector field along the map $E : U \rightarrow \mathcal{O}(M)$ which assigns to each $x \in U$ the vector $dE_x(X(x))$.

Let me repeat this important point in a slightly different version. Suppose that $C : [0, 1] \rightarrow M$ is a curve on M , and we start with an initial frame $\mathcal{E}(0)$ at $C(0)$. We know that there is a unique curve $t \mapsto \mathcal{E}(t)$ in $\mathcal{O}(M)$ which gives the parallel transport of $\mathcal{E}(0)$ along the curve C . We have “lifted” the curve C on M to the curve $\gamma : t \mapsto \mathcal{E}(t)$ on $\mathcal{E}(M)$. The curve γ is completely determined by

- its initial value $\gamma(0)$,
- the fact that it is a lift of C , i.e. that $\pi(\gamma(t)) = C(t)$ for all t , and
-

$$i(\gamma'(t))\bar{\omega} = 0. \tag{4.8}$$

We now want to describe two important properties of the form $\bar{\omega}$. For $B \in G$, recall that r_B denotes the transformation

$$r_B : \mathcal{O}(M) \rightarrow \mathcal{O}(M), \quad \mathcal{E} \mapsto \mathcal{E}B^{-1}.$$

We will use the same letter, r_B to denote the transformation

$$r_B : U \times G \rightarrow U \times G, \quad (x, A) \mapsto (x, AB^{-1}).$$

Because of (4.4), we may use this ambiguous notation since

$$\psi \circ r_B = r_B \circ \psi.$$

It then follows from the local definition (4.7) that

$$r_B^*\bar{\omega} = B\bar{\omega}B^{-1}. \tag{4.9}$$

Indeed

$$r_b^*(\bar{\omega}) = r_b^*\psi^* [A^{-1}\omega A + A^{-1}dA] = \psi^*r_b^* [A^{-1}\omega A + A^{-1}dA]$$

and

$$r_B^*(A^{-1}\omega A) = B(A^{-1}\omega A)B^{-1}$$

since ω does not depend on G and

$$r_B^*(A^{-1}dA) = B(A^{-1}dA)B^{-1}.$$

We can write (4.9) as

$$r_B^*\bar{\omega} = \text{Ad}_B(\bar{\omega}). \quad (4.10)$$

For the second property, we introduce some notation. Let ξ be a matrix which is “antisymmetric” in the sense that

$$\epsilon_i \xi_j^i = -\epsilon_j \xi_i^j.$$

This implies that the one parameter group

$$t \mapsto \exp -t\xi = I - t\xi + \frac{1}{2}t^2\xi^2 - \frac{1}{3!}t^3\xi^3 + \dots$$

lies in our group G for all t . Then the one parameter group of transformations

$$r_{\exp -t\xi} : \mathcal{O}(M) \rightarrow \mathcal{O}(M)$$

has as its infinitesimal generator a vector field, which we shall denote by X_ξ . The one parameter group of transformations

$$r_{\exp -t\xi} : U \times G \rightarrow U \times G$$

also has an infinitesimal generator: Identifying the tangent space to the space of matrices with the space of matrices, we see that the vector field generating this one parameter group of transformations of $U \times G$ is

$$Y_\xi : (x, A) \mapsto A\xi.$$

So the vector field Y_ξ takes values at each point in the TG component of the tangent space to $U \times G$ and assigns to each point (x, A) the matrix $A\xi$. In particular $\omega(Y_\xi) = 0$ since ω is only sensitive to the TU component. Also dA is by definition the tautological matrix valued differential form which assigns to any tangent vector Z the matrix Z . Hence

$$A^{-1}dA(Y_\xi) = \xi.$$

From

$$r_B \circ \psi = \psi \circ r_B$$

it follows that

$$\psi^*(Y_\xi) = X_\xi$$

and hence that

$$\bar{\omega}(X_\xi) \equiv \xi. \quad (4.11)$$

Finally, the curvature form from the point of view of the bundle of frames is given as usual as

$$\bar{\Omega} := d\bar{\omega} + \frac{1}{2}[\bar{\omega} \wedge, \bar{\omega}]. \quad (4.12)$$

4.5 Gauss' theorems.

We pause with this section to go back to classical differential geometry using the language we have developed so far.

4.5.1 Equations of structure of Euclidean space.

Suppose we take $M = \mathbf{R}^n$ with its standard Euclidean scalar product. The Levi-Civita connection is then derived from the identification of the tangent space at every point with \mathbf{R}^n itself - a vector field becomes identified with an \mathbf{R}^n valued function which we can then differentiate. A point of $\mathcal{O}(\mathbf{R}^n)$ can then be written as $(x, \mathcal{E}_1, \dots, \mathcal{E}_n)$ where $x \in \mathbf{R}^n$ and $\mathcal{E}_i \in \mathbf{R}^n$ with $\mathcal{E}_1, \dots, \mathcal{E}_n$ forming an orthonormal basis of \mathbf{R}^n . We then have

$$\pi(x, \mathcal{E}_1, \dots, \mathcal{E}_n) = x$$

and

$$\vartheta^i = \langle dx, \mathcal{E}_i \rangle,$$

the right hand side being the scalar product of the vector valued differential form, dx and the vector valued function \mathcal{E}_i . We have

$$dx = \mathcal{E}\vartheta.$$

Differentiating this equation gives

$$0 = d\mathcal{E} \wedge \vartheta + \mathcal{E}d\vartheta.$$

We have

$$d\mathcal{E}_j = \sum \bar{\omega}_j^i \mathcal{E}_i \quad \text{where } \bar{\omega}_j^i := \langle d\mathcal{E}_j, \mathcal{E}_i \rangle$$

or

$$d\mathcal{E} = \mathcal{E}\bar{\omega}.$$

We thus get

$$d\vartheta + \bar{\omega} \wedge \vartheta = 0$$

showing that $\bar{\omega}$ is indeed the connection form. Taking the exterior derivative of the equation $d\mathcal{E} = \mathcal{E} \wedge \bar{\omega}$ gives

$$d\bar{\omega} + \bar{\omega} \wedge \bar{\omega} = 0$$

showing that the curvature does indeed vanish. To summarize, the equations of structure of Euclidean geometry are

$$\vartheta^i := \langle dx, \mathcal{E}_i \rangle \tag{4.13}$$

$$\bar{\omega}_j^i := \langle d\mathcal{E}_j, \mathcal{E}_i \rangle \tag{4.14}$$

$$\bar{\omega}_j^i = -\bar{\omega}_i^j \tag{4.15}$$

$$dx = \mathcal{E}\vartheta \tag{4.16}$$

$$d\mathcal{E} = \mathcal{E}\bar{\omega} \tag{4.17}$$

$$d\vartheta + \bar{\omega} \wedge \vartheta = 0 \tag{4.18}$$

$$d\bar{\omega} + \bar{\omega} \wedge \bar{\omega} = 0. \tag{4.19}$$

4.5.2 Equations of structure of a surface in \mathbf{R}^3 .

We specialize to $n = 3$. Let S be a surface in \mathbf{R}^3 . This picks out a three dimensional submanifold of the six dimensional $\mathcal{O}(\mathbf{R}^3)$, call it $\mathcal{F}(S)$ consisting of all $(x, \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3)$ where

$$x \in S$$

and

$$\mathcal{E}_1 \text{ and } \mathcal{E}_2 \text{ are tangent to } S.$$

Of course this implies that \mathcal{E}_3 is normal to S . We will use a subscript S to denote the pull back of all functions and forms from $\mathcal{O}(\mathbf{R}^3)$ to $\mathcal{F}(S)$. For example, the vector valued differential form dx_S takes values in the tangent space TS_x regarded as a subspace of \mathbf{R}^3 . Hence $\vartheta_S^3 \equiv 0$. The set of all $(x_S, \mathcal{E}_{1S}, \mathcal{E}_{2S})$ constitutes $\mathcal{O}(S)$, the bundle of frames of S thought of as a two dimensional Riemann manifold. Since \mathcal{E}_{3S} is determined up to a \pm sign by the point x_S , we can think of $\mathcal{F}(S)$ as a two fold cover of $\mathcal{O}(S)$. [From a local point of view we can always make a choice of the sign, and also from the global point of view if the surface is orientable.]

From the equations of structure of Euclidean space we obtain

$$dx_S = \vartheta_S^1 \mathcal{E}_{1S} + \vartheta_S^2 \mathcal{E}_{2S} \quad (4.20)$$

$$d\mathcal{E}_{3S} = \bar{\omega}_{3S}^1 \mathcal{E}_{1S} + \bar{\omega}_{3S}^2 \mathcal{E}_{2S} \quad (4.21)$$

$$d\vartheta_S^1 + \bar{\omega}_{2S}^1 \wedge \vartheta_S^2 = 0 \quad (4.22)$$

$$d\vartheta_S^2 - \bar{\omega}_{2S}^1 \wedge \vartheta_S^1 = 0 \quad (4.23)$$

$$d\bar{\omega}_{2S}^1 + \bar{\omega}_{3S}^1 \wedge \bar{\omega}_{3S}^2 = 0 \quad (4.24)$$

the last equation following from $\bar{\omega}_2^1 = -\bar{\omega}_1^2$ and $\bar{\omega}_1^1 = 0$.

Equations (4.22) and (4.23) show that $\bar{\omega}_{2S}^1$ and $\bar{\omega}_{1S}^2 = -\bar{\omega}_{2S}^1$ are the connection forms of $\mathcal{O}(S)$ if we (locally) identify it with $\mathcal{F}(S)$. In particular, $\bar{\omega}_{2S}^1$ is intrinsically defined - it gives the Levi-Civita connection of the induced Riemann metric on S .

4.5.3 *Theorema egregium.*

Equation (4.21) shows that

$$\bar{\omega}_{3S}^1 \wedge \bar{\omega}_{3S}^2 = K \vartheta_S^1 \wedge \vartheta_S^2 \quad (4.25)$$

where K is the Gaussian curvature. Gauss's *theorema egregium* now follows immediately from (4.24).

4.5.4 Holonomy.

Let S be any two dimensional Riemann manifold (not necessarily embedded in three space). The connection matrix is a two by two antisymmetric matrix

$$\bar{\omega} = \begin{pmatrix} 0 & \bar{\omega}_2^1 \\ \bar{\omega}_1^2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \bar{\omega}_2^1 \\ -\bar{\omega}_2^1 & 0 \end{pmatrix}.$$

Let $E = (E_1, E_2)$ be a frame field on S , and let

$$\omega_2^1 = E^* \bar{\omega}_{2S}^1$$

be the corresponding form on S . Let γ be a curve on S lying entirely in the domain of definition of the frame field, and let $t \mapsto v(t) \in TS_{\gamma(t)}$ be a field of unit vectors along the curve. We can write

$$v(t) = \cos \phi(t) E_1(\gamma(t)) + \sin \phi(t) E_2(\gamma(t))$$

where $\phi(t)$ is the angle that the unit vector $v(t)$ makes with the first basis vector, $E_1(\gamma(t))$, of the frame at $\gamma(t)$. Then

$$\begin{aligned} v' &= -\phi' \sin \phi E_1 + \omega_1^2(\gamma') \cos \phi E_2 + \phi' \cos \phi E_2 + \omega_2^1(\gamma') E_1 \\ &= (\phi' - \omega_2^1(\gamma'))[-\sin \phi E_1 + \cos \phi E_2]. \end{aligned}$$

In particular, v is parallel along γ if and only if

$$\phi'(t) \equiv \omega_2^1(\gamma'(t)). \quad (4.26)$$

So

$$[\phi] = \int_{\gamma} \omega_2^1 \quad (4.27)$$

gives the change in ϕ of a parallel vector field along γ . Of course the angle is relative to a choice of frame field, and so has no intrinsic meaning. But suppose that γ is a closed curve, so $[\phi]$ measures the rotation involved in transporting a tangent vector all the way around the curve back to the starting point. This is well defined, independent of the frame field, and (4.27) is valid for any closed curve on the surface. In particular, suppose that

$$\gamma = \partial D$$

i.e. suppose that γ is the boundary curve of some oriented two dimensional region. We then may apply Stokes' theorem and (4.25) to conclude that

$$[\phi] = \int_D K dA. \quad (4.28)$$

4.5.5 Gauss-Bonnet.

Suppose that D is contained in the domain of a frame field, say a frame field obtained by orthonormalizing the basic fields of a coordinate patch, to fix the ideas. Let ψ denote the angle that the vector field makes with γ' rather than with E_1 . The tangent vector γ' turns through an angle of 2π relative to the frame field as we traverse the curve. (this requires some proof in general, but is obvious if D is convex in some coordinate chart, since then the angle that γ' makes with the x_1 -axis is steadily increasing. So we can restrict to this case to avoid calling in additional arguments.) Thus

$$[\psi] = [\phi] - 2\pi.$$

If γ is only piecewise differentiable, like the boundary of a polygon, then the change in ψ will come from two sources, the contribution of the smooth portions and the exterior angles at the corners. So we can write

$$[\psi] = [\text{edge contributions}] - \sum \text{exterior angles}.$$

We get

$$2\pi = \int_D K dA + \sum \text{exterior angles} - [\text{edge contributions}].$$

Now suppose we subdivide the surface into such “polygonal” regions, and sum the preceding equation over all regions. The edge contributions will cancel, since each edge will contribute twice, traversed in opposite directions. Thus

$$2\pi f = \int_S K dA + \sum \text{exterior angles}$$

where f is the number of regions, D , or “faces”. Now we can write each exterior angle as

$$\pi - \text{interior angle}.$$

The sum of all the interior angles at each corner from the regions impinging on it add up to 2π . Each edge contributes to two corners. So if we let e denote the number of edges and v the number of “vertices” or corners we obtain the Gauss-Bonnet formula

$$f - e + v = \frac{1}{2\pi} \int_S K dA. \quad (4.29)$$

The amazing property of this formula is that the left hand side does not depend on the choice of metric, while the right hand side does not depend on the choice of subdivision (and is not obviously an integer on the face of it). So we obtain Euler’s theorem that $f - e + v$ is independent of the choice of subdivision, and also that the integral of the curvature is independent of the choice of metric, and is an integer equal to the Euler number $f - e + v$.

Chapter 5

Connections on principal bundles.

According to the current “standard model” of elementary particle physics, every fundamental force is associated with a kind of curvature. But the curvatures involved are not only the geometric curvatures of space-time, but curvatures associated with the notion of a connection on a geometrical object (a “principal bundle”) which is a generalization of the bundle of frames studied in the preceding chapter. We develop the necessary geometrical facts in this chapter.

5.1 Submersions, fibrations, and connections.

A smooth map $\pi : Y \rightarrow X$ is called a **submersion** if $d\pi_y : TY_y \rightarrow TX_{\pi(y)}$ is surjective for every $y \in Y$. Suppose that X is n -dimensional and that Y is $n + k$ dimensional. The implicit function theorem implies the following for a submersion:

If $\pi : Y \rightarrow X$ is a submersion, then about any $y \in Y$ there exist coordinates $z^1, \dots, z^n; y^1, \dots, y^k$ (such that y has coordinates $(0, \dots, 0; 0 \dots, 0)$) and coordinates x^1, \dots, x^n about $\pi(y)$ such that in terms of these coordinates π is given by

$$\pi(z^1, \dots, z^n; y^1, \dots, y^k) = (z^1, \dots, z^n).$$

In other words, locally in Y , a submersion looks like the standard projection from \mathbf{R}^{n+k} to \mathbf{R}^n near the origin. For the rest of this section we will let $\pi : Y \rightarrow X$ denote a submersion.

For each $y \in Y$ we define the **vertical** subspace Vert_y of the tangent space TY_y to consist of those $\eta \in TY_y$ such that

$$d\pi_y(\eta) = 0.$$

In terms of the local description, the vertical subspace at any point in the coordinate neighborhood of y given above is spanned by the values of the vector fields

$$\frac{\partial}{\partial y^1}, \dots, \frac{\partial}{\partial y^k}$$

at the point in question. This shows that the Vert_y fit together to form a smooth sub-bundle, call it Vert , of the tangent bundle TY .

A **general connection** on the given submersion is a choice of complementary subbundle Hor to Vert . This means that at each $y \in Y$ we are given a subspace $\text{Hor}_y \subset TY_y$ such that

$$\text{Vert}_y \oplus \text{Hor}_y = TY_y$$

and that the Hor_y fit together smoothly to form a sub-bundle of TY . It follows from the definition that Hor_y has the same dimension as $TX_{\pi(y)}$ and, in fact, that the restriction of $d\pi_y$ to Hor_y is an isomorphism of Hor_y with $TX_{\pi(y)}$. We should emphasize that the vertical bundle Vert comes along with the notion of the submersion π . A connection Hor , on the other hand, is an additional piece of geometrical data above and beyond the submersion itself.

Let us describe a connection in terms of the local coordinates given above. The local coordinates x^1, \dots, x^n on X give rise to the vector fields

$$\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}$$

which form a basis of the tangent spaces to X at every point in the coordinate neighborhood on X . Since $d\pi$ restricted to Hor is a bijection at every point of Y , we conclude that there are functions a_{ri} , $r = 1, \dots, k$, $i = 1, \dots, n$ on the coordinate neighborhood on Y such that

$$\frac{\partial}{\partial z^1} + \sum_{r=1}^k a_{r1} \frac{\partial}{\partial y^r}, \dots, \frac{\partial}{\partial z^n} + \sum_{r=1}^k a_{rn} \frac{\partial}{\partial y^r}$$

span Hor at every point of the neighborhood.

Let $C : [0, 1] \rightarrow X$ be a smooth curve on X . We say that a smooth curve γ on Y is a **horizontal lift** of C if

- $\pi \circ \gamma = C$ and
- $\gamma'(t) \in \text{Hor}_{\gamma(t)}$ for all t .

For the first condition to hold, each point $C(t)$ must lie in the image of π . (The condition of being a submersion does not imply, without some additional hypotheses, that π is surjective.) Let us examine the second condition in terms of our local coordinate description. Suppose that $x = C(0)$, that $x = \pi(y)$, and we look for a horizontal lift with $\gamma(0) = y$. We can write

$$C(t) = (x^1(t), \dots, x^n(t))$$

in terms of the local coordinate system on X . So if γ is any lift (horizontal or not) of C , we have

$$\gamma(t) = (x^1(t), \dots, x^n(t); y^1(t), \dots, y^k(t))$$

in terms of the local coordinate system. For γ to be horizontal, we must have

$$\gamma'(t) = \sum_{i=1}^n x^{i'}(t) \frac{\partial}{\partial z^i} + \sum_r \sum_i a_{ri}(\gamma(t)) x^{i'}(t) \frac{\partial}{\partial y^r}.$$

Thus the condition that γ be a horizontal lift amount to the system of ordinary differential equations

$$\frac{dy^r}{dt} = \sum_r a_{ri}(x^1(t), \dots, x^n(t); y^1(t), \dots, y^k(t)) x^{i'}(t)$$

where the x^i and $x^{i'}$ are given functions of t . This is a system of (possibly) non-linear ordinary differential equations. The existence and uniqueness theorem for ordinary differential equations says that for a given initial condition $\gamma(0)$ there is some $\epsilon > 0$ for which there exists a unique solution of this system of differential equations for $0 \leq t < \epsilon$. Standard examples in the theory of differential equations show that the solutions can “blow up” in a finite amount of time; that in general one can not conclude the existence of the horizontal lift γ over the entire interval of definition of the curve C .

In the case of linear differential equations, we do have existence for all time, and therefore in the case of linear connections, or the connection that we studied on the bundle of orthogonal frames, there was global lifting.

We will now impose some restrictive conditions. We will say that the map $\pi : Y \rightarrow X$ is a locally trivial **fibration** if there exists a manifold F such that every $x \in X$ has a neighborhood U such that there exists a diffeomorphism

$$\psi_U \pi^{-1}(U) \rightarrow U \times F$$

such that

$$\pi_1 \circ \psi = \pi$$

where

$$\pi_1 : U \times F \rightarrow U$$

is projection onto the first factor. The implicit function theorem asserts that a submersion $\pi : Y \rightarrow X$ looks like a projection onto a first factor locally in Y . The more restrictive condition of being a fibration requires that π look like projection onto the first factor locally on X , with a second factor F which is fixed up to a diffeomorphism. If the map $\pi : Y \rightarrow X$ is a surjective submersion and is proper (meaning that the inverse image of a compact set is compact) then we shall prove below that π is a fibration if X is connected.

A second condition that we will impose is on the connection Hor . We will assume that every smooth curve C has a global horizontal lift γ . We saw that

this is the case when local coordinates can be chosen so that the equations for the lifting are linear, we shall see that it is also true when π is proper. But let us take this global lifting condition as a hypothesis for the moment.

Let $C : [a, b] \rightarrow X$ be a smooth curve. For any $y \in \pi^{-1}(C(a))$ we have a unique lifting $\gamma : [a, b] \rightarrow Y$ with $\gamma(a) = y$, and this lifting depends smoothly on y by the smooth dependence of solutions of differential equations on initial conditions. We thus have a smooth diffeomorphism associated with any smooth curve $C : [a, b] \rightarrow X$ sending

$$\pi^{-1}(C(a)) \rightarrow \pi^{-1}(C(b)).$$

If $c \in [a, b]$ it follows from the definition (and the existence and uniqueness theorem for differential equations) that the composite of the map

$$\pi^{-1}(C(a)) \rightarrow \pi^{-1}(C(c))$$

associated with the restriction of C to $[a, c]$ with the map

$$\pi^{-1}(C(c)) \rightarrow \pi^{-1}(C(b))$$

associated with the restriction of the curve C to $[c, b]$ is exactly the map

$$\pi^{-1}(C(a)) \rightarrow \pi^{-1}(C(b))$$

above. This then allows us to define a map $\pi^{-1}(C(a)) \rightarrow \pi^{-1}(C(b))$ associated to any piecewise differentiable curve, and the diffeomorphism associated to the concatenation of two curves which form a piecewise differentiable curve is the composite diffeomorphism.

Suppose that X has a smooth retraction to a point. This means that there is a smooth map $\phi : [0, 1] \times X \rightarrow X$ satisfying the following conditions where

$$\phi_t : X \rightarrow X$$

denotes the map

$$\phi_t(x) = \phi(t, x)$$

as usual. Here are the conditions:

- $\phi_0 = \text{id}$.
- $\phi_1(x) = x_0$, a fixed point of X .
- $\phi_t(x_0) = x_0$ for all $t \in [0, 1]$.

Suppose also that the submersion $\pi : Y \rightarrow X$ is surjective and has a connection with global lifting. We claim that this implies that the submersion is a trivial fibration; that there is a manifold F and a diffeomorphism

$$\Phi : Y \rightarrow X \times F \quad \text{with } \pi_1 \circ \Phi = \pi$$

where π_1 is projection onto the first factor. Indeed, take

$$F = \pi^{-1}(x_0).$$

For each $x \in X$ define

$$\Phi_x : \pi^{-1}(x) \rightarrow F$$

to be given by the lifting of the curve

$$t \mapsto \phi_t(x).$$

Then define

$$\Phi(y) = (\pi(y), \Phi_{\pi(y)}(y)).$$

The fact that Φ is a diffeomorphism follows from the fact that we can construct the inverse of Φ by doing the lifting in the opposite direction on each of the above curves. Every point on a manifold has a neighborhood which is diffeomorphic to a ball around the origin in Euclidean space. Such a ball is retractible to the origin by shrinking along radial lines. This proves that any surjective submersion which has a connection with a global lifting is locally trivial, i.e. is a fibration.

For any submersion we can always construct a connection. Simply put a Riemann metric on Y and let Hor be the orthogonal complement to Vert relative to this metric.

So to prove that if $\pi : Y \rightarrow X$ is a surjective submersion which is proper then it is a fibration, it is more than enough to prove that every connection has the global lifting property in this case. So let $C : [0, 1] \rightarrow X$ be a smooth curve. Extend C so it is defined on some slightly larger interval, say $[-a, 1+a]$, $a > 0$. For any $y \in \pi^{-1}(C(t))$, $t \in [0, 1]$ we can find a neighborhood U_y and an $\epsilon > 0$ such that the lifting of $C(s)$ exists for all $z \in U_y$ and $t-\epsilon < s < t+\epsilon$. This is what the local existence theorem for differential equations gives. But $C([0, 1])$ is a compact subset of X , and hence $\pi^{-1}(C([0, 1]))$ is compact since π is proper. This means that we can cover $\pi^{-1}(C([0, 1]))$ by finitely many such neighborhoods, and hence choose a fixed $\epsilon > 0$ that will work for all $y \in \pi^{-1}(C([0, 1]))$. But this clearly implies that we have global lifting, since we can do the lifting piecemeal over intervals of length less than ϵ and patch the local liftings together.

5.2 Principal bundles and invariant connections.

5.2.1 Principal bundles.

Let G be a Lie group with Lie algebra \mathfrak{g} . Let P be a space on which G acts. To tie in with our earlier notation, and also for later convenience, we will denote this action by

$$(p, a) \mapsto pa^{-1}, \quad p \in P, \quad a \in G$$

so $a \in G$ acts on P by a diffeomorphism that we will denote by r_a :

$$r_a : P \rightarrow P, \quad r_a(p) = pa^{-1}.$$

If $\xi \in \mathfrak{g}$, then $\exp(-t\xi)$ is a one parameter subgroup of G , and hence

$$r_{\exp(-t\xi)}$$

is a one parameter group of diffeomorphisms of P , and for each $p \in P$, the curve

$$r_{\exp(-t\xi)}p = p(\exp t\xi)$$

is a smooth curve starting at t at $t = 0$. The tangent vector to this curve at $t = 0$ is a tangent vector to P at p . In this way we get a linear map

$$u_p : \mathfrak{g} \rightarrow TP_p, \quad u_p(\xi) = \left. \frac{d}{dt} p(\exp t\xi) \right|_{t=0}. \quad (5.1)$$

For example, if we take $P = G$ with G acting on itself by right multiplication, and if we assumed that G is a subgroup of $Gl(n)$, so that we may identify TP_p as a subspace of the the space of all $n \times n$ matrices, then we have seen that

$$u_p(\xi) = p\xi$$

where the meaning of $p\xi$ on the right hand side is the product of the matrix p with the matrix ξ . For this case, if $r_a(p) = p$ for some $p \in P$, the $a = e$, the identity element.

In general, we say that the group action of G on P is **free** if no point of P is fixed by any element of G other than the identity. So “free” means that if $r_a(p) = p$ for some $p \in G$ then $a = e$. Clearly, if the action is free, then the map u_p is injective for all $p \in P$.

If we have an action of G on P and on Q , then we automatically get an action of G (diagonally) on $P \times Q$, and if the action of P is free then so is the action on $P \times Q$.

For example (to change the notation slightly), if X is a space on which G acts trivially, and if we let G act on itself by right multiplication, then we get a free action of G on $X \times G$. This is what we encountered when we began to construct the manifold structure on the bundle of orthogonal frames out of a local frame field. We now generalize this construction:

If we are given an action of G on P we have a projection $\pi : P \rightarrow P/G$ which sends each $p \in P$ to its G -orbit. We make the following assumptions:

- The action of G on P is free.
- The space P/G is a differentiable manifold M and the projection $\pi : P \rightarrow M$ is a smooth fibration.
- The fibration π is locally trivial consistent with the G action in the sense that every $m \in M$ has a neighborhood U such that there exists a diffeomorphism

$$\psi_U \pi^{-1}(U) \rightarrow U \times G$$

such that

$$\pi_1 \circ \psi = \pi$$

where

$$\pi_1 : U \times F \rightarrow U$$

is projection onto the first factor and if $\psi(p) = (m, b)$ then

$$\psi(r_a p) = (m, ba^{-1}).$$

When all this happens, we say that $\pi : P \rightarrow M$ is a **principal fiber bundle** over M with **structure group** G .

Suppose that $\pi : P \rightarrow M$ is a principal fiber bundle with structure group G . Since π is a submersion, we have the sub-bundle Vert of the tangent bundle TP , and from its construction, the subspace $\text{Vert}_p \subset TP_p$ is spanned by the tangents to the curves $p(\exp t\xi)$, $\xi \in \mathfrak{g}$. In other words, u_p is a surjective map from \mathfrak{g} to Vert_p . Since the action of G on P is free, we know that u_p is injective. Putting these two facts together we conclude that

Proposition 6 *If $\pi : P \rightarrow M$ is a principal fiber bundle with structure group G then u_p is an isomorphism of \mathfrak{g} with Vert_p for every $p \in P$.*

Let us compare the isomorphism u_p with the isomorphism $u_{r_b(p)} = u_{pb^{-1}}$. The action of $b \in G$ on P preserves the fibration and hence

$$d(r_b)_p : \text{Vert}_p \rightarrow \text{Vert}_{pb^{-1}}.$$

Let $v = u_p(\xi) \in \text{Vert}_p$. This means that

$$v = \frac{d}{dt}(p \exp t\xi)_{t=0}.$$

By definition

$$d(r_b)_p v = \frac{d}{dt}(r_b(p \exp t\xi))_{t=0} = \frac{d}{dt}((p \exp t\xi)b^{-1})_{t=0}.$$

We have

$$\begin{aligned} p(\exp t\xi)b^{-1} &= pb^{-1}(b(\exp t\xi)b^{-1}) \\ &= pb^{-1} \exp t \text{Ad}_b \xi \end{aligned}$$

where Ad is the conjugation, or adjoint, action of G on its Lie algebra. We have thus shown that

$$d(r_b)_p u_p(\xi) = u_{r_b(p)}(\text{Ad}_b \xi). \quad (5.2)$$

5.2.2 Connections on principal bundles.

Let $\pi : P \rightarrow M$ be a principal bundle with structure group G . Recall that in the general setting, we defined a (general) connection to be a sub-bundle Hor of the tangent bundle TP which is complementary to the vertical sub-bundle Vert . Given the group action of G , we can demand that Hor be invariant under

G . So by a **connection on a principal bundle** we will mean a sub-bundle Hor of the tangent bundle such that

$$TP_p = \text{Vert}_p \oplus \text{Hor}_p \quad \text{at all } p \in P$$

and

$$d(r_b)_p(\text{Hor}_p) = \text{Hor}_{r_b(p)} \quad \forall b \in G, \quad p \in P. \quad (5.3)$$

At any p we can define the projection

$$\mathbf{V}_p : TP_p \rightarrow \text{Vert}_p$$

along Hor_p , i.e. \mathbf{V}_p is the identity on Vert_p and sends all elements of Hor_p to 0. Giving Hor_p is the same as giving \mathbf{V}_p and condition (5.3) is the same as the condition

$$d(r_b)_p \circ \mathbf{V}_p = \mathbf{V}_{r_b(p)} \circ d(r_b)_p \quad \forall b \in G, \quad p \in P. \quad (5.4)$$

Let us compose $u_p^{-1} : \text{Vert}_p \rightarrow \mathfrak{g}$ with \mathbf{V}_p . So we define the \mathfrak{g} valued form $\bar{\omega}$ by

$$\bar{\omega}_p := u_p^{-1} \circ \mathbf{V}_p. \quad (5.5)$$

Then it follows from (5.2) and (5.4) that

$$r_b^* \bar{\omega} = \text{Ad}_b \bar{\omega}. \quad (5.6)$$

Let ξ_P be the vector field on P which is the infinitesimal generator of $r_{\exp t\xi}$. In view of definition of u_p as identifying ξ with the tangent vector to the curve $t \mapsto p(\exp t\xi) = r_{\exp -t\xi} p$ at $t = 0$, we see that

$$i(\xi_P) \bar{\omega} = -\xi. \quad (5.7)$$

The infinitesimal version of (5.6) is

$$D_{\xi_P} \bar{\omega} = [\xi, \bar{\omega}]. \quad (5.8)$$

Define the curvature by our formula

$$\bar{\Omega} := d\bar{\omega} + \frac{1}{2}[\bar{\omega}, \bar{\omega}]. \quad (5.9)$$

It follows from (5.6) that

$$r_b^* \bar{\Omega} = \text{Ad}_b \bar{\Omega} \quad \forall b \in G. \quad (5.10)$$

Now

$$i(\xi_P) d\bar{\omega} = D_{\xi_P} \bar{\omega} - di(\xi_P) \bar{\omega}$$

by Weil's formula for the Lie derivative. By (5.7) the second term on the right vanishes because it is the differential of the constant $-\xi$. So

$$i(\xi_P) d\bar{\omega} = [\xi, \bar{\omega}].$$

On the other hand

$$i(\xi_P)[\bar{\omega}, \bar{\omega}] = [i(\xi_P)\bar{\omega}, \bar{\omega}] - [\bar{\omega}, i(\xi_P)\bar{\omega}] = -2[\xi, \bar{\omega}]$$

where we used (5.7) again. So

$$i(v)\bar{\Omega} = 0 \quad \text{if } v \in \text{Vert}_p. \quad (5.11)$$

To understand the meaning of $\bar{\Omega}$ when evaluated on a pair of horizontal vectors, let X and Y be pair of horizontal vector fields, that is vector fields whose values at every point are elements of Hor . Then $i(X)\bar{\omega} = 0$ and $i(Y)\bar{\omega} = 0$. So

$$\bar{\Omega}(X, Y) = i(Y)i(X)\bar{\Omega} = i(Y)i(X)d\bar{\omega} = d\bar{\omega}(X, Y).$$

But by our general formula for the exterior derivative we have

$$d\bar{\omega}(X, Y) = X(i(Y)\bar{\omega}) - Y(i(X)\bar{\omega}) - \bar{\omega}([X, Y]).$$

The first two terms vanish and so

$$\bar{\Omega} = -\bar{\omega}([X, Y]). \quad (5.12)$$

This shows how the curvature measures the failure of the bracket of two horizontal vector fields to be horizontal.

5.2.3 Associated bundles.

Let $\pi : P \rightarrow M$ be a principal bundle with structure group G , and let F be some manifold on which G acts. We will write this action as multiplication on the left; i.e. we will denote the action of an element $a \in G$ on an element $f \in F$ as af . We then have the diagonal action of G on $P \times F$: For $a \in G$ we define

$$\text{diag}(a) : P \times F \rightarrow P \times F, \quad \text{diag}(a)(p, f) = (pa^{-1}, af).$$

Since the action of G on P is free, so is its diagonal action on $P \times F$. We can form the quotient space of this action, i.e. identify all elements of $P \times F$ which lie on the same orbit; so we identify the points (p, f) and (pa^{-1}, af) . The quotient space under this identification will be denoted by

$$P \times_G F$$

or by

$$F(P).$$

It is a manifold and the projection map $\pi : P \rightarrow M$ descends to a projection of $F(P) \rightarrow M$ which we will denote by π_F or simply by π when there is no danger of confusion. The map $\pi_F : F(P) \rightarrow M$ is a fibration. The bundle $F(P)$ is called the bundle associated to P by the G -action on F .

Let

$$\rho : P \times F \rightarrow F(P)$$

be the map which send (p, f) into its equivalence class.

Suppose that we are given a connection on the principal bundle P . Recall that this means that at each $p \in P$ we are given a subspace $\text{Hor}_p \subset TP_p$ which is complementary to the vertical, and that this assignment is invariant under the action of G in the sense that

$$\text{Hor}_{pa^{-1}} = dr_a(\text{Hor}_p).$$

Given an $f \in F$, we can consider Hor_p as the subspace

$$\text{Hor}_p \times \{0\} \subset T(P \times F)_{(p,f)} = TP_p \oplus TF_f$$

and then form

$$d\rho_{(p,f)} \text{Hor}_p \subset T(F(P))_{\rho(p,f)}$$

which is complementary to the vertical subspace $V(F(P))_{\rho(p,f)} \subset T(F(P))_{\rho(p,f)}$. The invariance condition of Hor implies that $d\rho_{(p,f)}(\text{Hor}_p)$ is independent of the choice of (p, f) in its equivalence class.

So a connection on a principal bundle induces a connection on each of its associated bundles.

5.2.4 Sections of associated bundles.

If $\pi : Y \rightarrow X$ is a submersion, then a **section** of this submersion is a map

$$s : X \rightarrow Y$$

such that

$$\pi \circ s = \text{id}.$$

In other words, s is a map which associates to each $x \in X$ an element

$$s(x) \in Y_x = \pi^{-1}(x).$$

Naturally, we will be primarily interested in sections which are smooth.

For example, we might consider the tangent bundle TM . A section of the tangent bundle then associates to each $x \in M$ a tangent vector $s(x) \in TM_x$. In other words, s is a vector field. Similarly, a linear differential form on M is a section of the cotangent bundle T^*M .

Suppose that $\pi = \pi_F : F(P) \rightarrow M$ is an associated bundle of a principal bundle P , and that $s : M \rightarrow F(P)$ is a section of this bundle. Let x be a point of M , and let $p \in P_x = \pi^{-1}(x)$ be a point in the fiber of the principal bundle $P \rightarrow M$ lying in the fiber over x . Then there is a unique $f \in F$ such that

$$\rho((p, f)) = s(x).$$

We thus get a function $\phi_s : P \rightarrow F$ by assigning to p this element $f \in F$. In other words, ϕ_s is uniquely determined by

$$\rho((p, \phi_s(p))) = s(\pi(p)). \quad (5.13)$$

Suppose we replace p by $r_a(p) = pa^{-1}$. Since $\rho((pa^{-1}, af)) = \rho((p, f))$ we see that ϕ_s satisfies the condition

$$\phi \circ r_a = a\phi \quad \forall a \in G. \quad (5.14)$$

Conversely, suppose that $\phi : P \rightarrow F$ satisfies (5.14). Then

$$\rho((p, \phi(p))) = \rho((pa^{-1}, \phi(pa^{-1})))$$

and so defines an element $s(x)$, $x = \pi(p)$. So a $\phi : P \rightarrow F$ satisfying (5.14) determines a section $s : M \rightarrow F(P)$ with $\phi = \phi_s$. It is routine to check that s is smooth if and only if ϕ is smooth. We have thus proved

Proposition 7 *There is a one to one correspondence between (smooth) sections $s : M \rightarrow F(P)$ and (smooth) functions $\phi : P \rightarrow F$ satisfying (5.14). The correspondence is given by (5.13).*

An extremely special case of this proposition is where we take F to be the real numbers with the trivial action of G on \mathbf{R} . Then $\mathbf{R}(P) = M \times \mathbf{R}$ since the map ρ does not identify two distinct elements of \mathbf{R} but merely identifies all elements of P_x . A section s of $M \times \mathbf{R}$ is of the form $s(x) = (x, f(x))$ where f is a real valued function. the proposition then asserts that we can identify real valued functions on M with real valued functions on P which are constant on the fibers P_x .

5.2.5 Associated vector bundles.

We now specialize to the case that F is a vector space, and the action of G on F is linear. In other words, we are given a linear representation of G on the vector space F . If $x \in M$ we can add two elements v_1 and v_2 of $F(P)_x$ by choosing $p \in P_x$ which then determines f_1 and f_2 in F such that

$$\rho((p, f_1)) = v_1 \quad \text{and} \quad \rho((p, f_2)) = v_2.$$

We then define

$$v_1 + v_2 := \rho((p, f_1 + f_2)).$$

The fact that the action of G on F is linear guarantees that this definition is independent of the choice of p . In a similar way, we define multiplication of an element of $F(P)_x$ by a scalar and verify that all the conditions for $F(P)_x$ to be a vector space are satisfied.

Let $V \rightarrow M$ be a vector bundle. So $V \rightarrow M$ is a fibration for which each V_x has the structure of a vector space. (As a class of examples of vector bundles we can consider the associated vector bundles $F(P)$ just considered.) We can then consider V valued differential forms on M . For example, a V valued linear differential form τ will be a rule which assigns a linear map

$$\tau_x : TM_x \rightarrow V_x$$

for each $x \in M$, and similarly we can talk of V valued k -forms.

For the case that $V = F(P)$ is an associated vector bundle we have a generalization of Proposition 7 to the case of differential forms. That is, we can describe $F(P)$ valued differential forms as certain kinds of F -valued forms on P . To see how this works, suppose that τ is an $F(P)$ -valued k -form on M . Let $x \in M$ and let $p \in P_x$. Now

$$\tau_x : \wedge^k(TM_x) \rightarrow F(P)_x$$

and p gives an identification map which we will denote by

$$\text{ident}_p$$

of $F(P)_x$ with F - the element $f \in F$ being identified with $\rho((p, f)) \in F(P)_x$. Also,

$$d\pi_p : TP_p \rightarrow TM_x$$

and so induces map (which we shall also denote by $d\pi_p$)

$$d\pi_p : \wedge^k(TP_p) \rightarrow \wedge^k(TM_x).$$

So

$$\sigma_p := \text{ident}_p \circ \tau_x \circ d\pi_p$$

maps $\wedge^k(TP_p) \rightarrow F$. Thus we have defined an F -valued k -form σ on P . If v is a vertical tangent vector at any point p of P we have $d\pi_p(v) = 0$, so

$$i(v)\sigma = 0 \quad \text{if } v \in \text{Vert}(P). \quad (5.15)$$

Let us see what happens when we replace p by $r_a(p) = pa^{-1}$ in the expression for σ . since $\pi \circ r_a = \pi$, we conclude that

$$d\pi_{pa^{-1}} \circ d(r_a)_p = d\pi_p.$$

Also,

$$\text{ident}_{pa^{-1}} = a \circ \text{ident}_p$$

where the a on the right denotes the action of a on F . We thus conclude that

$$r_a^*\sigma = a \circ \sigma. \quad (5.16)$$

Conversely, suppose that σ is an F -valued k -form on P which satisfies (5.15) and (5.16). It defines an $F(P)$ valued k -form τ on M as follows: At each $x \in M$ choose a $p \in P_x$. For any k tangent vectors $v_1, \dots, v_k \in TM_x$ choose tangent vectors $w_1, \dots, w_k \in TP_p$ such that

$$d\pi_p(w_j) = v_j, \quad j = 1, \dots, k.$$

Then consider

$$\sigma_p(w_1 \wedge \dots \wedge w_k) \in F.$$

Condition (5.15) guarantees that this value is independent of the choice of the w_i with $d\pi_p(w_j) = v_j$. In this way we define a map

$$\wedge^k(TM_x) \rightarrow F.$$

If we now apply $\rho(p, \cdot)$ to the image, we get a map

$$\wedge^k(TM_x) \rightarrow F(P)_x$$

and condition (5.16) guarantees that this map is independent of the choice of $p \in P_x$. From the construction it is clear that the assignments $\tau \rightarrow \sigma$ and $\sigma \rightarrow \tau$ are inverses of one another. We have thus proved:

Proposition 8 *There is one to one correspondence between $F(P)$ valued forms on M and F valued forms on P which satisfy (5.15) and (5.16).*

Forms on P which satisfy (5.15) and (5.16) are called **basic** forms because (according to the proposition) F -valued forms on P forms on P which (5.15) and (5.16) correspond to forms on the base manifold M with values in the associated bundle $F(P)$.

For example, equations (5.10) and (5.11) say that the curvature of a connection on a principal bundle is a basic \mathfrak{g} valued form relative to the adjoint action of G on \mathfrak{g} . According to the proposition, we can consider this curvature as a two form on the base M with values in $\mathfrak{g}(P)$, the vector bundle associated to P by the adjoint action of G on its Lie algebra.

Here is another important illustration of the concept. Equation (5.6) says that a connection form $\bar{\omega}$ satisfies (5.16), but it certainly does *not* satisfy (5.15). Indeed, the interior product of a vertical vector with the linear differential form $\bar{\omega}$ is given by (5.7). However, suppose that we are given two connection forms $\bar{\omega}_1$ and $\bar{\omega}_2$. Then their difference $\bar{\omega}_1 - \bar{\omega}_2$ *does* satisfy (5.15) and, of course, (5.16). We can phrase this by saying that the difference of two connections is a basic \mathfrak{g} valued one-form.

5.2.6 Exterior products of vector valued forms.

Suppose that F_1 and F_2 are two vector spaces on which G acts, and suppose that we are given a bilinear map

$$\mathbf{b} : F_1 \times F_2 \rightarrow F_3$$

into a third vectors space F_3 on which G acts, and suppose that \mathbf{b} is consistent with the actions of G in the sense that

$$\mathbf{b}(af_1, af_2) = a\mathbf{b}(f_1, f_2).$$

Examples of such a situation that we have come across before are:

1. G is a subgroup of $Gl(n)$ and F_1, F_2 and F_3 are all the vector space of $n \times n$ matrices, and \mathbf{b} is matrix multiplication.

2. G is a subgroup of $Gl(n)$, F_1 is the space of all $n \times n$ matrices, F_2 and F_3 are \mathbf{R}^n and \mathbf{b} is multiplication of a matrix times a vector.
3. G is a general Lie group, $F_1 = F_2 = F_3 = \mathfrak{g}$, the Lie algebra of G and \mathbf{b} is Lie bracket.

In each of these cases we have had occasion to form the exterior product of an F_1 valued differential form with an F_2 valued differential form to obtain an F_3 valued form.

We can do this construction in general: form the exterior product of an F_1 valued k -form with an F_2 -valued ℓ form to get an F_3 valued $k + \ell$ form. For example, if f_1^1, \dots, f_m^1 is a basis of F_1 and f_1^2, \dots, f_n^2 is a basis of F_2 then the most general F_1 -valued k -form α can be written as

$$\alpha = \sum \alpha^i f_i^1$$

where the α^i are real valued k -forms, and the most general F_2 -valued ℓ -form β can be written as

$$\beta = \sum \beta^j f_j^2$$

where the β^j are real valued ℓ forms. Let f_1^3, \dots, f_q^3 be a basis of F_3 and define the numbers B_{ij}^k by

$$\mathbf{b}(f_i^1, f_j^2) = \sum_k B_{ij}^k f_k^3.$$

Then you can check that $\alpha \wedge \beta$ defined by

$$\alpha \wedge \beta := \sum B_{ij}^k (\alpha^i \wedge \beta^j) f_k^3$$

is independent of the choice of bases. In a similar way we can define the exterior derivative of a vector valued form, the interior product of a vector valued form with a vector field, the pull back of a vector valued form under a map etc. There should be little problem in understanding the concept involved. There is a bit of a notational problem - how explicit do we want to make the map \mathbf{b} in writing down a symbol for this exterior product. In example 1) we simply wrote \wedge for the exterior product of two matrix valued forms. This forced us to use the rather ugly $[\alpha, \wedge \beta]$ for the exterior product of two Lie algebra valued forms, where the \mathbf{b} was commutator or Lie bracket. We shall retain this ugly notation for the sake of the clarity it gives.

A situation that we will want to discuss in the next section is: we are given an action of G on a vector space F , and unless forced to be more explicit, we have chosen to denote the action of an element $a \in G$ on an element $f \in F$ simply by af . This determines a bilinear map

$$\mathbf{b} : \mathfrak{g} \times F \rightarrow F$$

by

$$\mathbf{b}(\xi, f) := \frac{d}{dt}(\exp t\xi)f|_{t=0}.$$

We therefore get an exterior multiplication of a \mathfrak{g} -valued form with an F -valued form. We shall denote this particular type of exterior multiplication by \bullet . So if α is a \mathfrak{g} -valued k -form and β is an F -valued ℓ form then $\alpha \bullet \beta$ is an F -valued $(k + \ell)$ -form.

We point out that conditions (5.15) and (5.16) make perfectly good sense for vector valued forms, and so we can talk of basic vector valued forms on P , and the exterior product of two basic vector valued forms is again basic.

5.3 Covariant differentials and covariant derivatives.

In this section we consider a fixed connection on a principal bundle P . This means that we are given a projection \mathbf{V} of TP onto the vertical bundle and therefore a connection form $\bar{\omega}$. Of course we also have a projection

$$\mathbf{id} - \mathbf{V}$$

onto the horizontal bundle Hor of the connection, where \mathbf{id} is the identity operator. This projection kills all vertical vectors.

5.3.1 The horizontal projection of forms.

If α is a (possibly vector valued) k -form on P , we will define the horizontal projection $\mathbf{H}\alpha$ of α by

$$\mathbf{H}\alpha(v_1, \dots, v_k) = \alpha((\mathbf{id} - \mathbf{V})v_1, \dots, (\mathbf{id} - \mathbf{V})v_k). \quad (5.17)$$

The following properties of \mathbf{H} follow immediately from its definition and the invariance of the horizontal bundle under the action of G :

1. $\mathbf{H}(\alpha \wedge \beta) = \mathbf{H}\alpha \wedge \mathbf{H}\beta$.
2. $r_a^* \circ \mathbf{H} = \mathbf{H} \circ r_a^* \quad \forall a \in G$.
3. If α has the property that $i(w)\alpha = 0$ for any horizontal vector w then $\mathbf{H}\alpha = 0$. In particular,
4. $\mathbf{H}\bar{\omega} = 0$.
5. If α has the property that $i(v)\alpha = 0$ for any vertical vector v then $\mathbf{H}\alpha = \alpha$. In particular,
6. \mathbf{H} is the identity on basic forms.

In 1) α and β could be vector valued forms if we have the bilinear map \mathbf{b} which allows us to multiply them.

The map \mathbf{H} is clearly a projection in the sense that

$$\mathbf{H}^2 = \mathbf{H}.$$

5.3.2 The covariant differential of forms on P .

Define \mathbf{d} mapping k -forms into $(k + 1)$ forms by

$$\mathbf{d} := \mathbf{H} \circ d. \quad (5.18)$$

The following facts are immediate:

- $\mathbf{d}(\alpha \wedge \beta) = \mathbf{d}\alpha \wedge \mathbf{H}\beta + (-1)^k \mathbf{H}\alpha \wedge \mathbf{d}\beta$ if α is a k -form.
- $i(v)\mathbf{d} = 0$ for any vertical vector v .
- $r_a^* \circ \mathbf{d} = \mathbf{d} \circ r_a^* \quad \forall a \in G$.

It follows from the second and third items that \mathbf{d} carries basic forms into basic forms.

If F is a vector space on which G acts linearly, we can form the associated vector bundle $F(P)$, and we know from Proposition 8 that k -forms on M with values in $F(P)$ are the same as basic k -forms on P with values in F . So giving a connection on P induces an operator \mathbf{d} mapping k -forms on M with values in $F(P)$ to $(k + 1)$ -forms on M with values in $F(P)$. For example, a section s of $F(P)$ is just a zero form on M with values in the vector bundle $F(M)$. Giving the connection on P allows us to construct the one form $\mathbf{d}s$ with values in $F(P)$. If X is a vector field on M , then we can define

$$\nabla_X s := i(X)\mathbf{d}s,$$

the covariant derivative of s in the direction X .

5.3.3 A formula for the covariant differential of basic forms.

Let α be a basic form on P with values in the vector space F on which G acts linearly. Let \mathbf{d} be the covariant differential associated with the connection form $\bar{\omega}$. We claim that

$$\mathbf{d}\alpha = d\alpha + \bar{\omega} \bullet \alpha. \quad (5.19)$$

In order to prove this formula, it is enough to prove that when we apply $i(v)$ to the right hand side we get zero, if v is vertical. For then applying \mathbf{H} does not change the right hand side. But applying \mathbf{H} to the right hand side yields $\mathbf{d}\alpha$ since $\mathbf{d}\alpha := \mathbf{H}(d\alpha)$ and

$$\mathbf{H}\bar{\omega} = 0$$

so

$$\mathbf{H}(\bar{\omega} \bullet \alpha) = 0.$$

So it is enough to show that for any $\xi \in \mathfrak{g}$ we have

$$i(\xi_P)d\alpha = -i(\xi_P)(\bar{\omega} \bullet \alpha).$$

Since α is basic, we have $i(\xi_P)\alpha = 0$, so by Weil's identity we have

$$i(\xi_P)\alpha = D_{\xi_P}\alpha = \xi_P \bullet \alpha$$

by the infinitesimal version of the invariance condition (5.16). On the other hand, since $i(\xi_P)\alpha = 0$ and $i(\xi_P)\omega = -\xi$, we have proved our formula.

There are a couple of special cases of (5.19) worth mentioning. If F is \mathbf{R} with the trivial representation then (5.19) says that $\mathbf{d} = d$. This implies, that if s is a section of an associated vector bundle $F(P)$, and if ϕ is a function on M , so that ϕs is again a section of $F(P)$ then

$$\mathbf{d}(\phi s) = (d\phi) \wedge s + s \mathbf{d}s$$

implying that for any vector field X on M we have

$$\nabla_X(\phi s) = (X\phi)s + \phi(\nabla_X s).$$

Another important special case is where we take $F = \mathfrak{g}$ with the adjoint action. Then (5.19) says that

$$\mathbf{d}\alpha = d\alpha + [\bar{\omega} \wedge, \alpha].$$

5.3.4 The curvature is $\mathbf{d}\bar{\omega}$.

We wish to prove that

$$\mathbf{d}\bar{\omega} = d\omega + \frac{1}{2}[\omega \wedge, \omega]. \quad (5.20)$$

Both sides vanish when we apply $i(v)$ where v is a vertical vector - this is true for the left hand side by definition, and we have already verified this for the right hand side, see equation (5.11). But if we apply \mathbf{H} to both sides, we get $\mathbf{d}\bar{\omega}$ on the left, and also on the right since $\mathbf{H}\bar{\omega} = 0$. \square

5.3.5 Bianchi's identity.

In our setting this says that

$$\mathbf{d}\bar{\Omega} = 0. \quad (5.21)$$

Proof. We have

$$d\bar{\Omega} = d(d\bar{\omega}) + d\frac{1}{2}[\bar{\omega} \wedge, \bar{\omega}] = [d\bar{\omega} \wedge, \bar{\omega}].$$

Applying \mathbf{H} yields zero because $\mathbf{H}\bar{\omega} = 0$. \square

5.3.6 The curvature and \mathbf{d}^2 .

We wish to show that

$$\mathbf{d}^2\alpha = \bar{\Omega} \bullet \alpha. \quad (5.22)$$

In this equation α is a basic form on P with values in the vector space F where G acts, and we know that $\bar{\Omega}$ is a basic form with values in \mathfrak{g} , so the right hand side makes sense and is a basic F valued form. To prove this we use our formula

$$\mathbf{d}\alpha = d\alpha + \bar{\omega} \bullet \alpha$$

and apply it again to get

$$\mathbf{d}^2\alpha = d(d\alpha + \bar{\omega} \bullet \alpha) + \bar{\omega} \bullet (d\alpha + \bar{\omega}).$$

We have $d^2 = 0$ so the first expression (under the d) becomes

$$d(\bar{\omega} \bullet \alpha) = d\bar{\omega} \bullet \alpha - \bar{\omega} \bullet d\alpha.$$

The second term on the right here cancels the term $\bar{\omega} \bullet d\alpha$ so we get

$$\mathbf{d}^2\alpha = d\bar{\omega} \bullet \alpha + \bar{\omega} \bullet (\bar{\omega} \bullet \alpha).$$

So to complete the proof we must check that

$$\frac{1}{2}[\bar{\omega} \wedge, \bar{\omega}] \bullet \alpha = \bar{\omega} \bullet (\bar{\omega} \bullet \alpha).$$

This is a variant of a computation we have done several times before. Since interior product with vertical vectors sends α to zero, while interior product with horizontal vectors sends $\bar{\omega}$ to zero, it suffices to verify that the above equation is true after we take the interior product of both sides with two vertical vectors, say η_P and ξ_P . Now

$$i(\xi_P) = [\bar{\omega} \wedge, \bar{\omega}] = -[\xi, \bar{\omega}] + [\bar{\omega}, \xi] = -2[\xi, \bar{\omega}]$$

and so

$$i(\eta_P)i(\xi_P)\left(\frac{1}{2}[\bar{\omega} \wedge, \bar{\omega}] \bullet \alpha\right) = [\xi, \eta] \bullet \alpha.$$

A similar computation shows that

$$i(\eta_P)i(\xi_P)(\bar{\omega} \bullet (\bar{\omega} \bullet \alpha)) = \xi \bullet (\eta \bullet \alpha) - \eta \bullet (\xi \bullet \alpha).$$

But the equality of these two expressions follows from the fact that we have an action of G on F which implies that for any $\xi, \eta \in \mathfrak{g}$ and any $f \in F$ we have

$$[\xi, \eta]f = \xi(\eta f) - \eta(\xi f). \quad \square$$

Chapter 6

Gauss's lemma.

We have defined geodesics as being curves which are self parallel. But there are several other characterizations of geodesics which are just as important: for example, in a Riemann manifold geodesics locally minimize arc length: “a straight line is the shortest distance between two points”. We want to give one explanation of this fact here, using the “exponential map,” a concept introduced by al Biruni (973-1048) but unappreciated for about 1000 years. The key result, known as Gauss' lemma asserts that radial geodesics are orthogonal to the images of spheres under the exponential map, and this will allow us to relate geodesics to extremal properties of arc length.

6.1 The exponential map.

Suppose that M is a manifold with a connection ∇ . Let m_0 be a point of M and $\xi \in TM_{m_0}$. Then there is a unique (maximal) geodesic γ_ξ with $\gamma_\xi(0) = m_0$, $\gamma'_\xi(0) = \xi$. It is found by solving a system of second order ordinary differential equations. The existence and uniqueness theorem for solutions of such equations implies that the solutions depend smoothly on ξ . In other words, there exists a neighborhood \mathcal{N} of ξ in the tangent bundle TM and an interval I about 0 in \mathbf{R} such that $(\eta, s) \mapsto \gamma_\eta(s)$ is smooth on $\mathcal{N} \times I$.

If we take $\xi = 0$, the zero tangent vector, the corresponding “geodesic” defined for all t is the constant curve $\gamma_0(t) \equiv m_0$. The continuity thus implies that for ξ in some neighborhood of the origin in TM_{m_0} , the geodesic γ_ξ is defined for $t \in [0, 1]$. Let \mathcal{D}_0 be the set of vectors ξ in TM_{m_0} such that the maximal geodesic through ξ is defined on $[0, 1]$. By the preceding remarks this contains some neighborhood of the origin. Define the exponential map

$$\exp = \exp_{m_0} : \mathcal{D}_0 \rightarrow M, \quad \exp(\xi) = \gamma_\xi(1). \quad (6.1)$$

For $\xi \in TM_{m_0}$ and fixed $t \in \mathbf{R}$ the curve

$$s \mapsto \gamma_\xi(ts)$$

is a geodesic whose tangent vector at $s = 0$ is $t\xi$. So the exponential map carries straight lines through the origin in TM_{m_0} into geodesics through m in M :

$$\exp : t\xi \mapsto \gamma_\xi(t).$$

Now the tangent vector to the line $t \mapsto t\xi$ at $t = 0$ is just ξ under the standard identification of the tangent space to a vector space with the vector space itself. Also, the tangent vector to the curve $t \mapsto \gamma_\xi(t)$ at $t = 0$ is ξ , by the definition of γ_ξ . So taking the derivatives of both sides shows that the differential of the exponential map is the identity:

$$d\exp_0 : T(TM_{m_0})_0 \rightarrow TM_{m_0} = \text{id}$$

under the standard identification of the tangent space $T(TM_{m_0})_0$ with TM_{m_0} .

From the inverse function theorem it follows that the exponential map is a diffeomorphism in some neighborhood of the origin. Let \mathcal{U} be a star shaped neighborhood of the origin in TM_{m_0} on which \exp is a diffeomorphism, and let $U := \exp(\mathcal{U})$ be its image in M under the exponential map. Then U is called a normal neighborhood of m_0 . By construction (and the uniqueness theorem for differential equations) for every $m \in U$ there exists a unique geodesic which joins m_0 to m and lies entirely in U .

6.2 Normal coordinates.

Suppose that we choose a basis $e = (e_1, \dots, e_n)$ of TM_{m_0} and let ℓ^1, \dots, ℓ^n be the dual basis. We then get a coordinate system on U defined by

$$\exp^{-1}(m) = \sum x^i(m)e_i$$

or, what is the same,

$$x^i = \ell^i \circ \exp^{-1}.$$

These coordinates are known as **normal** coordinates, or sometimes as **inertial** coordinates for the following reason:

Let $\xi = \sum a^i e_i$ be an element of $\mathcal{U} \subset TM_{m_0}$. Since $\exp(t\xi) = \gamma_\xi(t)$ the coordinates of $\gamma_\xi(t)$ are given by

$$x^i(\gamma_\xi(t)) = \ell^i(t\xi) = t\ell^i(\xi) = ta^i.$$

Thus the second derivative of $x^i(\gamma_\xi(t))$ with respect to t vanishes and the geodesic equations (satisfied by $\gamma_\xi(t)$) becomes

$$\sum_{ij} \Gamma_{ij}^k(\gamma_\xi(t)) a^i a^j = 0, \quad \forall k.$$

In particular, evaluating at $t = 0$ we get

$$\sum_{ij} \Gamma_{ij}^k(0) a^i a^j = 0, \quad \forall k.$$

But this must hold for all (sufficiently small) values of the a^i and hence for all values of the a^i . If the connection has zero torsion, so that the Γ_{ij}^k are symmetric in i and j , this implies that

$$\Gamma_{ij}^k(0) = 0. \quad (6.2)$$

In a normal coordinate system, the Christoffel symbols of a torsionless connection vanish at the origin. Hence at this one point, the equations for a geodesic look like the equations of a straight line in terms of these coordinates. This was Einstein's resolution of Mach's problem: How can the laws of physics - particularly mechanics - involve rectilinear motion in absence of forces, as this depends on the coordinate system. According to Einstein the distribution of matter in the universe determines the metric which then determines the connection which picks out the inertial frame.

6.3 The Euler field \mathcal{E} and its image \mathcal{P} .

The multiplicative group \mathbf{R}^+ acts on any vector space: $r \in \mathbf{R}^+$ sends any vector v into rv . We set

$$r = e^t.$$

The vector field corresponding \mathcal{E} corresponding to the one parameter group

$$v \mapsto e^t v$$

is known as the Euler operator. From its definition, if q is a homogeneous polynomial of degree k , then

$$\mathcal{E}q = kq,$$

an equation which is known as Euler's equation. Also from its definition, differentiating the curve

$$t \mapsto e^t v$$

at $t = 0$ shows that the value of \mathcal{E} at any vector v is

$$\mathcal{E}(v) = v$$

under the natural identification of the tangent space at v of the vector space with the vector space itself.

We want to consider the Euler field on the tangent space TM_{m_0} (and its restriction to the star shaped neighborhood \mathcal{U}) and its image under the exponential map, call it \mathcal{P} . So \mathcal{P} is a vector field defined on U . Since

$$\exp(r\xi) = \gamma_\xi(r)$$

we have

$$\begin{aligned} \mathcal{P}(\exp \xi) &= \frac{d}{dt} (\exp(e^t \xi))|_{t=0} \\ &= \frac{d}{dt} \gamma_\xi(e^t)|_{t=0} \end{aligned}$$

so

$$\mathcal{P}(\exp \xi) = \dot{\gamma}_\xi(1)$$

where we are using the dot to denote differentiation of the geodesic $r \mapsto \gamma_\xi(r)$ with respect to r . Applied to the vector $s\xi$ we obtain

$$\mathcal{P}(\gamma_\xi(s)) = s\dot{\gamma}_\xi(s).$$

We claim that

$$\nabla_{\mathcal{P}} \mathcal{P} = \mathcal{P}. \quad (6.3)$$

Indeed,

$$\begin{aligned} \nabla_{\mathcal{P}} \mathcal{P}(\gamma_\xi(t)) &= t \nabla_{\dot{\gamma}_\xi(t)} (t \dot{\gamma}_\xi(t)) \\ &= t \frac{dt}{dt} \dot{\gamma}_\xi(t) + t^2 \nabla_{\dot{\gamma}_\xi(t)} \dot{\gamma}_\xi(t) \\ &= t \dot{\gamma}_\xi(t) \quad \text{since } \gamma_\xi \text{ is a geodesic} \\ &= \mathcal{P}(\gamma_\xi(t)). \end{aligned}$$

Since the points of the form $\gamma_\xi(t)$ fill out the normal neighborhood, we conclude that (6.3) holds.

Suppose that we have chosen a basis $e = (e_1, \dots, e_n)$ of TM_{m_0} and so the corresponding normal coordinates x^1, \dots, x^n on U . Each vector e_i determines the “constant” vector field on TM_{m_0} which assigns to each vector ξ the value e_i (under the identification of $(TM_{m_0})_\xi$ with TM_{m_0}). Let us temporarily introduce the notation \tilde{e}_i to denote this vector field. As ℓ^1, \dots, ℓ^n form the dual basis, then each of the ℓ^j is a linear function on TM_{m_0} , and the derivative of the function ℓ^j with respect to the vector field \tilde{e}_i is given by

$$\tilde{e}_i \ell^j = 0, \quad i \neq j, \quad \tilde{e}_i \ell^i = 1.$$

Now $x^j = \ell^j \circ \exp^{-1}$ so we conclude that under the exponential map the vector field \tilde{e}_i is carried over into ∂_i in terms of the normal coordinates.

Now

$$\mathcal{E}(\xi) = \xi = \sum \ell^i(\xi) e_i = \sum \ell^i(\xi) \tilde{e}_i(\xi)$$

and $\ell^i(\xi) = x^i(m)$ if $\xi = \exp^{-1}(m)$. We conclude that the expression for \mathcal{P} in normal coordinates is given by

$$\mathcal{P} = \sum x^i \partial_i. \quad (6.4)$$

Thus in normal coordinates, the expression for \mathcal{P} is the same as the expression for the Euler operator \mathcal{E} in linear coordinates.

6.4 The normal frame field.

Let E_i be the vector field obtained from $\partial_i(m_0) = e_i$ by parallel translation along the $\gamma_\xi(t)$. By the existence and uniqueness theorem for differential equations,

we know that E_i is a smooth vector field on our normal neighborhood N . By the definition of \mathcal{P} we have

$$\nabla_{\mathcal{P}}(E_i) = 0. \quad (6.5)$$

Notice that at the single point m_0 we have $E_i(m_0) = \partial_i(m_0)$ but this equality need not hold at any other point. But $E = (E_1, \dots, E_m)$ is a frame field which is covariant constant with respect to \mathcal{P} . We call it the **normal frame field** (associated to the basis (e_1, \dots, e_n)). We then also construct the dual frame field θ which is also covariantly constant with respect to \mathcal{P} .

We claim that, remarkably,

$$\mathcal{P} = \sum_i x^i E_i. \quad (6.6)$$

Indeed, the coefficients $\theta^i(\mathcal{P})$ of \mathcal{P} with respect to the E_i are smooth functions on our normal neighborhood. Our first claim is that these functions are (in terms of our normal coordinates) homogeneous functions of order one. To show this it is enough, by Euler's theorem, to show that they satisfy the equation $\mathcal{P}f = f$. But we have

$$\mathcal{P}\theta^i(\mathcal{P}) = (\nabla_{\mathcal{P}}\theta^i) + \theta^i(\nabla_{\mathcal{P}}\mathcal{P}) = \theta^i(\mathcal{P})$$

since $\nabla_{\mathcal{P}}\theta^i = 0$ and $\nabla_{\mathcal{P}}\mathcal{P} = P$. So each of the $\theta^i(\mathcal{P})$ is a homogeneous linear function in terms of the normal coordinates.

This means that we can write $\theta^i(\mathcal{P}) = \sum a_{ij}x^j$ for some constants a_{ij} . Thus

$$\mathcal{P} = \sum_{ij} a_{ij}x^j E_i = \sum_k x^k \partial_k,$$

We want to show that $a_{ij} = \delta_{ij}$. By definition, $E_i(0) = \partial_i(0)$. If we write

$$|x|^2 = \sum_i x^{i2}$$

we have

$$\sum_{ij} a_{ij}x^j E_j = \sum_{ij} x^j \partial_i + O(|x|^2)$$

and also

$$\sum_{ij} a_{ij}x^j E_i = \mathcal{P} = \sum_j x^j \partial_j.$$

The only way that two linear expressions can agree up to terms quadratic or higher is if they are equal. so we have proved that (6.6) holds.

6.5 Gauss' lemma.

Now suppose that M is a semi-Riemannian manifold, and ∇ is the corresponding Levi-Civita connection. We choose our basis $e = (e_1, \dots, e_n)$ of TM_{m_0} to be "orthonormal", so that $E = (E_1, \dots, E_n)$ is an "orthonormal" frame field.

Since the E_i form an orthonormal frame at each point, it follows from (6.6) that

$$\langle \mathcal{P}, \mathcal{P} \rangle = \sum_i \epsilon_i x_i^2. \quad (6.7)$$

We claim that we also have

$$\langle \mathcal{P}, \partial_i \rangle = \epsilon_i x^i. \quad (6.8)$$

To prove this observe that

$$\langle \mathcal{P}, \partial_i \rangle = \sum_j x^j \langle \partial_j, \partial_i \rangle = \epsilon_i x^i + O(|x|^2).$$

So it is enough to show that

$$\mathcal{P} \langle \mathcal{P}, \partial_i \rangle = \epsilon_i x^i$$

in order to conclude (6.8). Now $[\mathcal{P}, \partial_i] = -\partial_i$ from the formula (6.4) for \mathcal{P} , and hence

$$\nabla_{\mathcal{P}} \partial_i = \nabla_{\partial_i} \mathcal{P} - \partial_i,$$

since the torsion of the Levi-Civita connection vanishes. Hence

$$\begin{aligned} \mathcal{P} \langle \mathcal{P}, \partial_i \rangle &= \langle \nabla_{\mathcal{P}} \mathcal{P}, \partial_i \rangle + \langle \mathcal{P}, \nabla_{\mathcal{P}} \partial_i \rangle \\ &= \langle \mathcal{P}, \partial_i \rangle + \langle \mathcal{P}, \nabla_{\partial_i} \mathcal{P} \rangle - \langle \mathcal{P}, \partial_i \rangle \\ &= \frac{1}{2} \partial_i \langle \mathcal{P}, \mathcal{P} \rangle \\ &= \frac{1}{2} \partial_i \sum_i \epsilon_i x_i^2 \\ &= \epsilon_i x^i. \end{aligned}$$

In particular, it follows from (6.8) that

$$\langle \mathcal{P}, \epsilon_j x^i \partial_j - \epsilon_i x^j \partial_i \rangle = 0. \quad (6.9)$$

Now the vector fields

$$\epsilon_j x^i \partial_j - \epsilon_i x^j \partial_i$$

correspond, under the exponential map, to the vector fields

$$\epsilon_j \ell^i \tilde{e}_j - \epsilon_i \ell^j \tilde{e}_i$$

which generate the one parameter group of “rotations” in the e_i, e_j plane in TM_{m_0} . These rotations, acting in the tangent space, when applied to any point, sweep out the “pseudo-sphere” centered at the origin and passing through that point. Let S_ξ be the pseudo-sphere in the tangent space TM_{m_0} passing through the point $\xi \in TM_{m_0}$ and let $\Sigma_p = \exp(S_\xi)$ be its image under the exponential map. Then we can restate equation (6.9) as

Proposition 9 *The radial geodesic through the point $p = \exp(\xi)$ is orthogonal in the Riemann metric to the hypersurface Σ_p .*

This result is known as Gauss' lemma.

6.6 Minimization of arc length.

We now specialize to the Riemannian case so that S_ξ is an actual sphere in the Euclidean sense. Let $\gamma : [0, 1] \rightarrow M$ be any curve joining m_0 to a point m in the normal neighborhood. In particular, for small values of t the points $\gamma(t)$ all lie in the normal neighborhood. Let

$$d = |x(m)|$$

i.e. $d^2 = \sum_i x^i{}^2$ in terms of the normal coordinates. We know that d is the length of the geodesic emanating from m_0 and ending at m by the definition of the exponential map and normal coordinates. We wish to show that

$$\text{length of } \gamma \geq d.$$

In other words, that the geodesic joining m_0 to m is the shortest curve joining m_0 to m . Since $\gamma(1) = m$ we have $|\gamma(1)| = d$.

Let T be the first time that $|x(\gamma(t))| \geq d$. (That is, T is the greatest lower bound of the set of all t for which $\gamma(t)$ does not lie strictly inside the sphere of radius d in normal coordinates.) Then $\gamma(T)$ must lie on the surface Σ , the image of the sphere of radius d under the exponential map. It is enough to prove that curve $\gamma : [0, T] \rightarrow M$ has length $\geq d$, where now $x(\gamma(t))$ lies inside the sphere of radius d for all $0 \leq t < T$. By the same argument, we may assume that $|x(\gamma(t))| > 0$ for all $t > 0$. Then

$$d = \int_0^T \frac{d|x(\gamma(t))|}{dt} dt.$$

Let u denote the unit vector field in the radial direction, defined outside the origin in the normal coordinates. So $u(x) = \frac{1}{|x|} \mathcal{P}(x)$. Decompose the tangent vector, $\gamma'(t)$ into its component along u and its component, τ along the plane spanned by the vector fields $x^i \partial_j - x^j \partial_i$. So

$$\gamma'(t) = c(t)u(t) + \tau(t).$$

Then

$$\int_0^T \frac{d|x(\gamma(t))|}{dt} dt = \int_0^T c(t) dt \leq \int_0^T |c(t)| dt.$$

On the other hand, $u(t)$ and $\tau(t)$ are orthogonal relative to the Riemann metric, and hence

$$\|\gamma'(t)\|^2 = |c(t)|^2 + \|\tau(t)\|^2$$

so

$$|\gamma'(t)| \geq |c(t)|$$

and hence

$$\text{length } \gamma = \int_0^T \|\gamma'(t)\| dt \geq d$$

as was to be proved.

Chapter 7

Special relativity

7.1 Two dimensional Lorentz transformations.

We study a two dimensional vector space with scalar product $\langle \cdot, \cdot \rangle$ of signature $+ -$. A Lorentz transformation is a linear transformation which preserves the scalar product. In particular it preserves

$$\|\mathbf{u}\|^2 := \langle \mathbf{u}, \mathbf{u} \rangle$$

(where with the usual abuse of notation this expression can be positive negative or zero). In particular, every such transformation must preserve the “light cone” consisting of all \mathbf{u} with $\|\mathbf{u}\|^2 = 0$.

All such two dimensional spaces are isomorphic. In particular, we can choose our vector space to be \mathbf{R}^2 with metric given by

$$\left\| \begin{pmatrix} u \\ v \end{pmatrix} \right\|^2 = uv.$$

The light cone consists of the coordinate axes, so every Lorentz transformation must carry the axes into themselves or interchange the axes. A transformation which preserves the axes is just a diagonal matrix. Hence the (connected component of) the Lorentz group consists of all matrices of the form

$$\begin{pmatrix} r & 0 \\ 0 & r^{-1} \end{pmatrix}, \quad r > 0.$$

So the group is isomorphic to the multiplicative group of the positive real num-

bers. We introduce (t, x) coordinates by

$$\begin{aligned} u &= t + x \\ v &= t - x \end{aligned}$$

or

$$\begin{aligned} \begin{pmatrix} u \\ v \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} \\ \text{so } \left\| \begin{pmatrix} t \\ x \end{pmatrix} \right\|^2 &= t^2 - x^2. \end{aligned}$$

Notice that

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^2 = 2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

so if

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix}$$

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} r & 0 \\ 0 & r^{-1} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} t' \\ x' \end{pmatrix}$$

$$\text{then } \begin{pmatrix} t' \\ x' \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} r & 0 \\ 0 & r^{-1} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix}$$

Multiplying out the matrices gives

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = \gamma \begin{pmatrix} 1 & w \\ w & 1 \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} \quad (7.1)$$

where

$$\gamma := \frac{r + r^{-1}}{2} \quad (7.2)$$

$$w := \frac{r - r^{-1}}{r + r^{-1}}. \quad (7.3)$$

The parameter w is called the “velocity” and is, of course, restricted by

$$|w| < 1. \quad (7.4)$$

We have

$$\begin{aligned} 1 - w^2 &= \frac{r^2 + 2 + r^{-2} - r^2 + 2 - r^{-2}}{(r + r^{-1})^2} \\ &= \frac{4}{(r + r^{-1})^2} \end{aligned}$$

so

$$\gamma = \frac{1}{\sqrt{1-w^2}}. \quad (7.5)$$

Thus w determines γ . Similarly, we can recover r from w :

$$r = \sqrt{\frac{1+w}{1-w}}.$$

So we can use w to parameterize the Lorentz transformations. We write

$$L_w := \gamma \begin{pmatrix} 1 & w \\ w & 1 \end{pmatrix}$$

7.1.1 Addition law for velocities.

It is useful to express the multiplication law in terms of the velocity parameter. If

$$w_1 = \frac{r - r^{-1}}{r + r^{-1}}$$

$$w_2 = \frac{s - s^{-1}}{s + s^{-1}}$$

then

$$\frac{rs - (rs)^{-1}}{rs + (rs)^{-1}} = \frac{\frac{r-r^{-1}}{r+r^{-1}} + \frac{s-s^{-1}}{s+s^{-1}}}{1 + \frac{s-s^{-1}}{s+s^{-1}} \cdot \frac{r-r^{-1}}{r+r^{-1}}}$$

so we obtain

$$L_{w_1} \circ L_{w_2} = L_w \quad \text{where } w = \frac{w_1 + w_2}{1 + w_1 w_2}. \quad (7.6)$$

This is known as the “addition law for velocities”.

7.1.2 Hyperbolic angle.

One also introduces the “hyperbolic angle”, actually a real number, ϕ by

$$r = e^\phi$$

so

$$\gamma = \cosh \phi = \frac{1}{\sqrt{1-w^2}}$$

and

$$L_w = \begin{pmatrix} \cosh \phi & \sinh \phi \\ \sinh \phi & \cosh \phi \end{pmatrix}.$$

Here

$$w = \tanh \phi.$$

For any $\begin{pmatrix} t \\ x \end{pmatrix}$ with $t > 0$ and $t^2 - x^2 = 1$, we must have $t > x$ and $t - x = (t + x)^{-1}$ so

$$\begin{pmatrix} t+x \\ t-x \end{pmatrix} = \begin{pmatrix} r & 0 \\ 0 & r^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad r = t+x.$$

This shows that the group of all (one dimensional proper) Lorentz transformations, $\{L_w\}$, acts simply transitively on the hyperbola

$$\left\| \begin{pmatrix} t \\ x \end{pmatrix} \right\|^2 = 1, \quad t > 0.$$

This means that if $\begin{pmatrix} t \\ x \end{pmatrix}$ and $\begin{pmatrix} t' \\ x' \end{pmatrix}$ are two points on this hyperbola, there is a unique L_w with

$$L_w \begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} t' \\ x' \end{pmatrix}.$$

If

$$\begin{pmatrix} t \\ x \end{pmatrix} = L_z \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

this means that

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = L_w L_z \begin{pmatrix} 1 \\ 0 \end{pmatrix} = L_z L_w \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and so

$$\left\langle \begin{pmatrix} t \\ x \end{pmatrix}, \begin{pmatrix} t' \\ x' \end{pmatrix} \right\rangle = tt' - xx' = \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, L_w \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle.$$

Writing $w = \tanh \phi$ as above we have

$$\langle \mathbf{u}, \mathbf{u}' \rangle = \cosh \phi, \quad \mathbf{u} = \begin{pmatrix} t \\ x \end{pmatrix} \mathbf{u}' = \begin{pmatrix} t' \\ x' \end{pmatrix},$$

and ϕ is called the hyperbolic angle between \mathbf{u} and \mathbf{u}' .

More generally, if we don't require $\|\mathbf{u}\| = \|\mathbf{u}'\| = 1$ but merely $\|\mathbf{u}\| > 0$, $\|\mathbf{u}'\| > 0$, $t > 0$, $t' > 0$ we define the hyperbolic angle between them to be the hyperbolic angle between the corresponding unit vectors so

$$\langle \mathbf{u}, \mathbf{u}' \rangle = \|\mathbf{u}\| \|\mathbf{u}'\| \cosh \phi.$$

7.1.3 Proper time.

A **material particle** is a curve $\alpha : \tau \mapsto \alpha(\tau)$ whose tangent vector $\alpha'(\tau)$ has positive t coordinate everywhere and satisfies

$$\|\alpha'(\tau)\| \equiv 1.$$

Of course, this fixes the parameter τ up to an additive constant. τ is called the **proper time** of the material particle. It is to be thought of as the “internal clock” of the material particle. For an unstable particle, for example, it is this internal clock which tells the particle that its time is up. Let ∂_0 denote unit vector in the t direction,

$$\partial_0 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

7.1.4 Time dilatation.

Let us write $t(\tau)$ for the t coordinate of $\alpha(\tau)$ and $x(\tau)$ for its x coordinate so that

$$\alpha(\tau) = \begin{pmatrix} t(\tau) \\ x(\tau) \end{pmatrix} \quad \alpha' := \frac{d\alpha}{d\tau} = \begin{pmatrix} dt/d\tau \\ dx/d\tau \end{pmatrix}.$$

We have

$$\begin{aligned} \frac{dt}{d\tau} &= \langle \partial_0, \alpha' \rangle, \\ &= \cosh \phi \\ &= \frac{1}{\sqrt{1-w^2}} \\ &\geq 1 \end{aligned}$$

where

$$w := \frac{dx}{dt} = \frac{dx/d\tau}{dt/d\tau} \tag{7.7}$$

is the “velocity” of the particle measured in the t, x coordinate system. Thus the internal clock of a moving particle appears to run slow in any coordinate system where it is not at rest. This phenomenon, known as “time dilatation” is observed all the time in elementary particle physics. For example, fast moving muons make it from the upper atmosphere to the ground before decaying due to this effect.

7.1.5 Lorentz-Fitzgerald contraction.

Let α and β be material particles whose trajectories are parallel straight lines. Once we have chosen a Minkowski basis, we have a notion of “simultaneity” relative to that basis, meaning that we can adjust the arbitrary additive constant in the definition of the proper time of each particle so that the two parallel straight lines are given by

$$\tau \mapsto \begin{pmatrix} a\tau \\ b\tau + c \end{pmatrix}, \text{ and } \tau \mapsto \begin{pmatrix} a\tau \\ b\tau + c + \ell \end{pmatrix}.$$

We can then think of the configuration as the motion of the end points of a “rigid rod” of length ℓ . The length ℓ depends on our notion of simultaneity. For example, suppose we apply a Lorentz transformation L_w to obtain $a = 1, b = 0$

(and readjust the additive constants in the clocks to achieve simultaneity). The corresponding frame is called the rest frame of the rod and its length, ℓ_{rest} , called the rest length of the rod is related to our “laboratory frame” by

$$\ell_{\text{rest}} = (\cosh \phi) \ell_{\text{lab}}$$

or

$$\ell_{\text{lab}} = \sqrt{1 - w^2} \ell_{\text{rest}}, \quad (7.8)$$

a moving object “contracts” in the direction of its motion. This is the Lorentz-Fitzgerald contraction which was discovered before special relativity in the context of electromagnetic theory, and can be considered as a forerunner of special relativity. As an effect in the laboratory, it is not nearly as important as time dilatation.

7.1.6 The reverse triangle inequality.

Consider any interval, say $[0, T]$, on the t axis, and let $0 < s < T$. The curve $t^2 - x^2 = s^2$ bends away from the origin. In other words, all other vectors with t coordinate equal to s have smaller Minkowski length:

$$\left\| \begin{pmatrix} s \\ x \end{pmatrix} \right\|^2 < s^2, \quad x \neq 0.$$

The length of any timelike vector $\mathbf{u} := \begin{pmatrix} s \\ x \end{pmatrix}$ is $< s$ if $x \neq 0$. Similarly, the Minkowski length of the (timelike) vector, \mathbf{v} , joining $\begin{pmatrix} s \\ x \end{pmatrix}$ to $\begin{pmatrix} T \\ 0 \end{pmatrix}$ is $< T - s$. We conclude that

$$\|\mathbf{u} + \mathbf{v}\| \geq \|\mathbf{u}\| + \|\mathbf{v}\| \quad (7.9)$$

with equality holding only if \mathbf{u} and \mathbf{v} actually lie on the t axis. There is nothing special in this argument about the t axis, or the fact that we are in two dimensions. It holds for any pair of forward timelike vectors, with equality holding if and only if the vectors are collinear. Inequality (7.9) is known as the **reverse triangle inequality**. The classical way of putting this is to say that the time measured by a clock moving along a (timelike) straight line path joint the events P and Q is *longer* than the time measured along any (timelike forward) broken path joining P to Q . It is also called the “twin effect”. The twin moving along the broken path (if he survives the bumps) will be younger than the twin who moves along the uniform path. This was known as the *twin paradox*. It is no paradox, just an immediate corollary of the reverse triangle inequality.

7.1.7 Physical significance of the Minkowski distance.

We wish to give an interpretation of the Minkowski square length (due originally to Robb (1936)) in terms of signals and clocks. Consider points $\begin{pmatrix} t_1 \\ 0 \end{pmatrix}$ and

$\begin{pmatrix} t_2 \\ 0 \end{pmatrix}$ on the t axis which are joined to the point $\begin{pmatrix} t \\ x \end{pmatrix}$ by light rays (lines parallel to $t = x$ or $t = -x$). Then (assuming $t_2 > t > t_1$)

$$\begin{aligned} t - t_1 &= x \quad \text{so} \\ t_1 &= t - x \quad \text{and} \\ t_2 - t &= x \quad \text{so} \\ t_2 &= t + x \end{aligned}$$

hence

$$t_1 t_2 = t^2 - x^2. \quad (7.10)$$

This equation has the following significance: Point $P = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ at rest or in uniform motion wishes to communicate with point $Q = \begin{pmatrix} t \\ x \end{pmatrix}$. It records the time, t_1 on its clock when a light signal was sent to Q and the time t_2 when the answer was received (assuming an instantaneous response.) Even though the individual times depend on the coordinates, their product, $t_1 t_2$ gives the square of the Minkowski norm of the vector joining P to Q .

7.1.8 Energy-momentum

In classical mechanics, a momentum vector is usually considered to be an element of the cotangent space, i.e the dual space to the tangent space. Thus in our situation, where we identify all tangent spaces with the Minkowski plane itself, a “momentum” vector will be a row vector of the form $\mu = (E, p)$. For a material particle the associated momentum vector, called the “energy momentum vector” in special relativity, is a row vector with the property that the evaluation map

$$v \mapsto \mu(v)$$

for any vector v is a positive multiple of the scalar product evaluation

$$v \mapsto \langle v, \alpha'(\tau) \rangle.$$

In other words, evaluation under μ is the same as scalar product with $m\alpha'$ where m , is an invariant of the material particle known as the **rest mass**. The rest mass is an invariant of the particle in question, constant throughout its motion. So in the rest frame of the particle, where $\alpha' = \partial_0$, the energy momentum vector has the form $(m, 0)$. Here m is identified (up to a choice of units, and we will have more to say about units later) with the usual notion of mass, as determined by collision experiments, for example. In a general frame we will have

$$\mu = (E, p), \quad E^2 - p^2 = m^2. \quad (7.11)$$

In this frame we have

$$p/E = w \quad (7.12)$$

where w is the velocity as defined in (7.7). We can solve equations (7.11) and (7.12) to obtain

$$E = \frac{m}{\sqrt{1-w^2}} \quad (7.13)$$

$$p = \frac{mw}{\sqrt{1-w^2}}. \quad (7.14)$$

For small values of w we have the Taylor expansion

$$\frac{1}{\sqrt{1-w^2}} = 1 + \frac{1}{2}w^2 + \dots$$

and so we have

$$E \doteq m + \frac{1}{2}mw^2 + \dots \quad (7.15)$$

$$p \doteq mw + \frac{1}{2}mw^3 + \dots \quad (7.16)$$

The first term in (7.16) looks like the classical expression $p = mw$ for the momentum in terms of the velocity if we think of m as the classical mass, and the second term in (7.15) looks like the classical expression for the kinetic energy. We are thus led to the following modification of the classical definitions of energy and momentum. Associated to any object there is a definite value of m called its rest mass. If the object is at rest in a given frame, its rest mass coincides with the classical notion of mass; when it is in motion relative to a given frame, its energy momentum vector is of the form (E, p) where E and p are determined by equations (7.13) and (7.14). We have been implicitly assuming that $m > 0$ which implies that $|w| < 1$. We can supplement these particles by particles of rest mass 0 whose energy momentum vector satisfy (7.11), so have the form $(E, \pm E)$. These correspond to particles which move along light rays $x = \pm t$. The law of conservation of energy momentum says that in any collision the total energy momentum vector is conserved.

7.1.9 Psychological units.

Our description of two dimensional Minkowski geometry has been in terms of “natural units” where the speed of light is one. Points in our two dimensional space time are called *events*. They record when and where something happens. If we record the total events of a single human consciousness (say roughly 70 years measured in seconds) and several thousand meters measured in seconds, we get a set of events which is enormously stretched out in one particular time direction compared to space direction, by a factor of something like 10^{18} . Being very skinny in the space direction as opposed to the time direction we tend to have a preferred splitting of spacetime with space and time directions picked out, and to measure distances in space with much smaller units, such as meters, than the units we use (such as seconds) to measure time. Of course, if we use a

small unit, the corresponding numerical value of the measurement will be large; in terms of human or “ordinary units” space distances will be greatly magnified in comparison with time differences. This suggests that we consider variables T and X related to the natural units t and x by $T = c^{-1}t$, $X = x$ or

$$\begin{pmatrix} T \\ X \end{pmatrix} = \begin{pmatrix} c^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix}.$$

The light cone $|t| = |x|$ goes over to $|X| = c|T|$ and we say that the “speed of light is c in ordinary units”. Similarly, the time-like hyperbolas $t^2 - x^2 = k > 0$ become very flattened out and are almost the vertical lines $T = \text{const.}$, lines of “simultaneity”. To find the expression for the Lorentz transformations in ordinary units, we must conjugate the Lorentz transformation, L , by the matrix $\begin{pmatrix} c^{-1} & 0 \\ 0 & 1 \end{pmatrix}$ so

$$\begin{aligned} M &= \begin{pmatrix} c^{-1} & 0 \\ 0 & 1 \end{pmatrix} L \begin{pmatrix} c & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \cosh \phi & c^{-1} \sinh \phi \\ c \sinh \phi & \cosh \phi \end{pmatrix} \\ &= \gamma \begin{pmatrix} 1 & c^{-1}w \\ cw & 1 \end{pmatrix}, \end{aligned}$$

where $L = L_w$. Of course w is a pure number in natural units. In psychological units we must write $w = v/c$, the ratio of a velocity (in units like meters per second) to the speed of light. Then

$$M = M_v = \gamma \begin{pmatrix} 1 & \frac{v}{c^2} \\ v & 1 \end{pmatrix}, \quad \gamma = \frac{1}{(1 - \frac{v^2}{c^2})^{1/2}}. \quad (7.17)$$

Since we have passed to new coordinates in which

$$\left\| \begin{pmatrix} T \\ X \end{pmatrix} \right\|^2 = c^2 T^2 - X^2,$$

the corresponding metric in the dual space will have the energy component divided by c . As we have used cap for energy and lower case for momentum, we shall continue to denote the energy momentum vector in psychological units by (E, p) and we have

$$\|(E, p)\|^2 = \frac{E^2}{c^2} - p^2.$$

We still must see how these units relate to our conventional units of mass. For this, observe that we want the second term in (7.15) to look like kinetic energy when E is replaced by E/c , so we must rescale by $m \mapsto mc$. Thus we get

$$\|(E, p)\|^2 = \frac{E^2}{c^2} - p^2 = m^2 c^2. \quad (7.18)$$

So in psychological coordinates we rewrite (7.11)-(7.15) as (7.18) together with

$$\frac{p}{E} = \frac{v}{c^2} \quad (7.19)$$

$$E = \frac{mc^2}{(1 - v^2/c^2)^{1/2}} \quad (7.20)$$

$$p = \frac{mv}{(1 - v^2/c^2)^{1/2}} \quad (7.21)$$

$$E \doteq mc^2 + \frac{1}{2}mv^2 + \dots \quad (7.22)$$

$$p \doteq mv + \frac{1}{2}m\frac{v^3}{c^2} + \dots \quad (7.23)$$

Of course at velocity zero we get the famous Einstein formula $E = mc^2$.

7.1.10 The Galilean limit.

In “the limit” $c \rightarrow \infty$ the transformations M_v become

$$G_v = \begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix}$$

which preserve T and send $X \mapsto X + vT$. These are known as Galilean transformations. They satisfy the more familiar addition rule for velocities:

$$G_{v_1} \circ G_{v_2} = G_{v_1+v_2}.$$

7.2 Minkowski space.

Since our everyday space is three dimensional, the correct space for special relativity is a four dimensional Lorentzian vector space. This key idea is due to Minkowski. In a famous lecture at Cologne in September 1908 he says

Henceforth space by itself, and time by itself are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.

Much of what we did in the two dimensional case goes over unchanged to four dimensions. Of course, velocity, w or v , become vectors, \mathbf{w} and \mathbf{v} as does momentum, \mathbf{p} instead of p . So in any expression a term such as v^2 must be replaced by $\|\mathbf{v}\|^2$, the three dimensional norm squared, etc.. With this modification the key formulas of the preceding section go through. We will not rewrite them. The reverse triangle inequality and so the twin effect go through unchanged.

Of course there are important differences: the light cone is really a cone, and not two light rays, the space-like vectors form a connected set, the Lorentz group is ten dimensional instead of one dimensional. We will study the Lorentz

group in four dimensions in a later section. In this section we will concentrate on two-particle collisions, where the relative angle between the momenta gives an additional ingredient in four dimensions.

7.2.1 The Compton effect.

We consider a photon (a “particle” of mass zero) impinging on a massive particle (say an electron) at rest. After the collision the photon moves at an angle, θ , to its original path. The frequency of the light is changed as a function of the angle: If λ is the incoming wave length and λ' the wave length of the scattered light then

$$\lambda' = \lambda + \frac{h}{mc}(1 - \cos \theta), \quad (7.24)$$

where h is Planck’s constant and m is the mass of the target particle. The expression

$$\frac{h}{mc}$$

is known as the Compton wave length of a particle of mass m .

Compton derived (7.24) from the conservation of energy momentum as follows: We will work in natural units where $c = 1$. Assume Einstein’s formula

$$E_{\text{photon}} = h\nu \quad (7.25)$$

for the energy of the photon, where ν is the frequency, or equivalently,

$$E_{\text{photon}} = \frac{h}{\lambda} \quad (7.26)$$

where λ is the wave length. Work in the rest frame of the target particle, so its energy momentum vector is $(m, 0, 0, 0)$. Take the x -axis to be the direction of the incoming photon, so its energy momentum vector is $(\frac{h}{\lambda}, \frac{h}{\lambda}, 0, 0)$. Assume that the collision is *elastic* so that the outgoing photon still has mass zero and the recoiling particle still has mass m . Choose the y -axis so that the outgoing photon and the recoiling particle move in the x, y plane. Then the outgoing photon has energy momentum $(\frac{h}{\lambda'}, \frac{h}{\lambda'} \cos \theta, \frac{h}{\lambda'} \sin \theta, 0)$ while the recoiling particle has energy momentum $(E, p_x, p_y, 0)$ and conservation of energy momentum together with the assumed elasticity of the collision yield

$$\begin{aligned} \frac{h}{\lambda} + m &= \frac{h}{\lambda'} + E \\ \frac{h}{\lambda} &= \frac{h}{\lambda'} \cos \theta + p_x \\ 0 &= \frac{h}{\lambda'} \sin \theta + p_y \\ m^2 &= E^2 - p_x^2 - p_y^2. \end{aligned}$$

Substituting the second and third equations into the last gives

$$E^2 = m^2 + \frac{h^2}{\lambda^2} + \frac{h^2}{\lambda'^2} - 2\frac{h^2}{\lambda\lambda'} \cos \theta$$

while the first equation yields

$$E^2 = m^2 + \frac{h^2}{\lambda^2} + \frac{h^2}{\lambda'^2} + 2 \left[m \frac{h}{\lambda} - m \frac{h}{\lambda'} - \frac{h^2}{\lambda\lambda'} \right].$$

Comparing these two equations gives Compton's formula, (7.24).

Notice that Compton's formula makes three startling predictions: that the shift in wavelength is independent of the wavelength of the incoming radiation, the explicit nature of the dependence of this shift on the scattering angle, and an experimental determination of h/mc , in particular, if h and c are known, of the mass, m , of the scattering particle. These were the results of Compton's experiment.

It is worth recalling the historical importance of Compton's experiment (1923). At the end of the nineteenth century, statistical mechanics, which had been enormously successful in explaining many aspects of thermodynamics, yielded wrong, and even non-sensical, predictions when it came to the study of the electromagnetic radiation emitted by a hot body - the study of "blackbody radiation". In 1900 Planck showed that the paradoxes could be resolved and a an excellent fit to the experimental data achieved if one assumed that the electromagnetic radiation is emitted in packets of energy given by (7.25) where h is a constant, now called Planck's constant, with value

$$h = 6.26 \times 10^{-27} \text{erg s.}$$

For Planck, this quantization of the energy of radiation was a property of the emission process in blackbody radiation. In 1905 Einstein proposed the radical view that (7.25) was a property of the electromagnetic field itself, and not of any particular emission process. Light, according to Einstein, is quantized according to (7.25). He used this to explain the *photoelectric effect*: When light strikes a metallic surface, electrons are emitted. According to Einstein, an incoming light quantum of energy $h\nu$ strikes an electron in the metal, giving up all its energy to the electron, which then uses up a certain amount of energy, w , to escape from the surface. The electron may also use up some energy to reach the surface. In any event, the escaping electron has energy

$$E \leq h\nu - w$$

where w is an empirical property of the material. The startling consequence here is that the maximum energy of the emitted electron depends only on the frequency of the radiation, but not on the intensity of the light beam. Increasing the intensity will increase the number of electrons emitted, but not their maximum energy. Einstein's theory was rejected by the entire physics community. With the temporary exception of Stark (who later became a vicious nazi and attacked the theory of relativity as a Jewish plot) physicists could not accept the idea of a corpuscular nature to light, for this seemed to contradict the well established interference phenomena which implied a wave theory, and also contradicted Maxwell's equations, which were the cornerstone of all of theoretical physics. For a typical view, let us quote at length from Millikan (of oil

drop fame) whose experimental result gave the best confirmation of Einstein's predictions for the photoelectric effect. In his Nobel lecture (1924) he writes

After ten years of testing and changing and learning and sometimes blundering, all efforts being directed from the first toward the accurate experimental measurement of the energies of emission of photoelectrons, now as a function of temperature, now of wavelength, now of material (contact e.m.f. relations), this work resulted, contrary to my own expectation, in the first direct experimental proof in 1914 of the exact validity, within narrow limits of experimental error, of the Einstein equation, and the first direct photoelectric determination of Planck's h .

But despite Millikan's own experimental verification of Einstein's formula for the photoelectric effect, he did not regard this as confirmation of Einstein's theory of quantized radiation. On the contrary, in his paper, "A direct Photoelectric Determination of Planck's h " *Phy. Rev.* 7 (1916)355-388 where he presents his experimental results he writes:

... the semi-corpuseular theory by which Einstein arrived at his equation seems at present to wholly untenable....[Einstein's] bold, not to say reckless [hypothesis] seems a violation of the very conception of electromagnetic disturbance...[it] flies in the face of the thoroughly established facts of interference.... Despite... the apparently complete success of the Einstein equation, the physical theory of which it was designed to be the symbolic expression is found so untenable that Einstein himself, I believe, no longer holds to it, and we are in the position of having built a perfect structure and then knocked out entirely the underpinning without causing the building to fall. It stands complete and apparently well tested, but without any visible means of support. These supports must obviously exist, and the most fascinating problem of modern physics is to find them. Experiment has outrun theory, or, better, guided by an erroneous theory, it has discovered relationships which seem to be of the greatest interest and importance, but the reasons for them are as yet not at all understood.

Of course, Millikan was mistaken when he wrote that Einstein himself had abandoned his own theory. In fact, Einstein extended his theory in 1916 to include the quantization of the momentum of the photon. But for Millikan, as for most physicists, Einstein's hypothesis of the light quantum was clearly "an erroneous theory".

By the way, it is amusing to compare Millikan's actual state of mind in 1916 (which was the accepted view of the entire physics community outside of Einstein) with his fallacious account of it in his autobiography (1950) pp. 100-101, where he writes about his experimental verification of Einstein's equation for the photoelectric effect:

This seemed to me, as it did to many others, a matter of very great importance, for it rendered what I will call Planck's 1912 explosive or trigger approach to the problem of quanta completely untenable and proved simply and irrefutedly, I thought, *that the emitted electron that escapes with the energy $h\nu$ gets that energy by the direct transfer of $h\nu$ units of energy from the light to the electron* and hence scarcely permits of any interpretation than that which Einstein had originally suggested, namely that of the semi-corpuseular or photon theory of light itself.

Self-delusion or outright mendacity? In general I have found that one can not trust the accounts given by scientists of their own thought processes, especially those given many years after the events.

In any event, it was only with the Compton experiment, that Einstein's formula, (7.25) was accepted as a property of light itself.

For a detailed history see the book *The Compton Effect* by Roger H. Stuewer, Science History Publications, New York 1975, from which I have taken the above quotes.

7.2.2 Natural Units.

In this section I will make the paradoxical argument that Planck's constant and (7.25) have a purely classical interpretation: Like c , Planck's constant, h , may be viewed as a conversion factor from natural units to conventional units.

For this I will again briefly call on a higher theory, symplectic geometry. In that theory, conserved quantities are associated to continuous symmetries. More precisely, if G is a Lie group of symmetries with Lie algebra g , the moment map, Φ for a Hamiltonian action takes values in g^* , the *dual space* of the Lie algebra. A basis of g determines a dual basis of g^* . In the case at hand, the Lie algebra in question is the algebra of translation, and the moment map yields the (total) energy-momentum vector. Hence if we measure translations in units of length, then the corresponding units for energy momentum should be inverse length. In this sense the role of Planck's constant in (7.26) is a conversion factor from natural units of inverse length to the conventional units of energy. So we interpret $h = 6.626 \times 10^{-27}$ erg s as the conversion factor from the natural units of inverse seconds to the conventional units of ergs.

In order to emphasize this point, let us engage in some historical science fiction: Suppose that mechanics had developed before the invention of clocks. So we could observe trajectories of particles, their collisions and deflections, but not their velocities. For instance, we might be able to observe tracks in a bubble chamber or on a photographic plate. If our theory is invariant under the group of translations in space, then linear momentum would be an invariant of the particle; if our theory is invariant under the group of three dimensional Euclidean motions, the symplectic geometry tells that $\|\mathbf{p}\|$, the length of the linear momentum is an invariant of the particle. In the absence of a notion of velocity, we might not be able to distinguish between a heavy particle moving

slowly or a light particle moving fast. Without some way of relating momentum to length, we would introduce “independent units” of momentum, perhaps by combining particles in various ways and by performing collision experiments. But symplectic geometry tells us that the “natural” units of momentum should be inverse length, and that de Broglie’s equation

$$\|\mathbf{p}\| = \frac{h}{\lambda} \quad (7.27)$$

gives Planck’s constant as a conversion factor from natural units to conventional units. In fact, the crucial experiment was the photo-electric effect, carried out in detail by Millikan.

The above discussion does not diminish, even in retrospect, from the radical character of Einstein’s 1905 proposal. Even in terms of “natural units” the startling proposal is that it is a single particle, the photon, which interacts with a single particle, the electron to produce the photoelectric effect. It is this “corpuscular” picture which was so difficult to accept. Furthermore, it is a bold hypothesis to identify the “natural units” of the photon momentum with the inverse wave length.

For reasons of convenience physicists frequently prefer to use $\hbar := h/2\pi$ as the conversion factor.

One way of choosing natural units is to pick some particular particle and use its mass as the mass unit. Suppose we pick the proton. Then m_P , the mass of the proton is the basic unit of mass, and ℓ_P , the Compton wave length of the proton is the basic unit of length. Also t_P , the time it takes for light to travel the distance of one Compton wave length, is the basic unit of time. The conversion factors to the cgs system (using \hbar) are:

$$\begin{aligned} m_P &= 1.672 \times 10^{-24}g \\ \ell_P &= .211 \times 10^{-13}cm \\ t_P &= 0.07 \times 10^{-23}sec. \end{aligned}$$

We will oscillate between using natural units and familiar units. Usually, we will derive the formulas we want in natural units, where the computations are cleaner and then state the results in conventional units which are used in the laboratory.

7.2.3 Two-particle invariants.

Suppose that A and B are particles with energy momentum vectors p_A and p_B . In any particular frame they have the expression $p_A = (E_A, \mathbf{p}_A)$ and $p_B = (E_B, \mathbf{p}_B)$. We have the three invariants

$$\begin{aligned} p_A^2 &= m_A^2 = E_A^2 - (\mathbf{p}_A, \mathbf{p}_A) \\ p_B^2 &= m_B^2 = E_B^2 - (\mathbf{p}_B, \mathbf{p}_B) \\ p_A \cdot p_B &= E_A E_B - (\mathbf{p}_A, \mathbf{p}_B). \end{aligned}$$

For the purpose of this section our notation is that (\cdot, \cdot) refers to the three dimensional scalar product, a symbol such as p_A denotes the energy momentum (four) vector of particle A , $p_A \cdot p_B$ denotes the four dimensional scalar product and we write p_A^2 for $p_A \cdot p_A$. These are all standard notations. The left hand sides are all invariants in the sense that their computation does not depend on the choice of frame. Many computations become transparent by choosing a frame in which some of the expressions on the right take on a particularly simple form. It is intuitively obvious (and also a theorem) that these are the only invariants - that any other invariant expression involving the two momenta vectors must be a function of these three. For example,

$$(p_A + p_B)^2 = p_A^2 + 2p_A \cdot p_B + p_B^2$$

and

$$(p_A - p_B)^2 = p_A^2 - 2p_A \cdot p_B + p_B^2.$$

Here are some examples:

Decay at rest.

Particle A , at rest, decays into particles B and C , symbolically $A \rightarrow B + C$. Find the energies, and the magnitudes of the momenta and velocities of the outgoing particles in the rest frame of particle A . Conservation of energy momentum gives

$$\begin{aligned} p_A &= p_B + p_C \quad \text{or} \\ p_C &= p_A - p_B, \\ \text{so} \\ p_C^2 &= p_A^2 + p_B^2 - 2p_A \cdot p_B \quad \text{or} \\ m_C^2 &= m_A^2 + m_B^2 - 2m_A E_B \quad \text{since} \\ p_A &= (m_A, 0, 0, 0). \end{aligned}$$

Solving gives

$$E_B = \frac{m_A^2 + m_B^2 - m_C^2}{2m_A}.$$

Interchanging B and C gives the formula for E_C . We have $\mathbf{p}_B = -\mathbf{p}_C$ and $E_B^2 - \|\mathbf{p}_B\|^2 = m_B^2$. Substituting into the above expression for E_B gives

$$\|\mathbf{p}_B\|^2 = \frac{1}{4m_A^2} (m_A^4 + m_B^4 + m_C^4 - 2m_A^2 m_C^2 - 2m_B^2 m_C^2 + 2m_A^2 m_B^2 - 4m_A^2 m_B^2)$$

so

$$\|\mathbf{p}_B\| = \|\mathbf{p}_C\| = \frac{\sqrt{\lambda(m_A^2, m_B^2, m_C^2)}}{2m_A} \quad (7.28)$$

where λ is the “triangle function”

$$\lambda(x, y, z) := x^2 + y^2 + z^2 - 2xy - 2xz - 2yz. \quad (7.29)$$

If we now redo the computation in ordinary units keeping track of dividing by c^2 in the energy part of the scalar product and multiplying all m 's by c we get

$$E_B = \frac{m_A^2 + m_B^2 - m_C^2}{2m_A} c^2 \quad (7.30)$$

$$\|\mathbf{p}_B\| = \|\mathbf{p}_C\| = \frac{\sqrt{\lambda(m_A^2, m_B^2, m_C^2)}}{2m_A} c. \quad (7.31)$$

Equation (7.12) becomes

$$\mathbf{v} = \frac{c^2}{E} \mathbf{p}. \quad (7.32)$$

Taking the magnitudes and using the above formulas for the energies and magnitudes of the momenta give the magnitudes of the velocities of the outgoing particles. Since our energies are all non-negative, (7.30) shows that decay from rest can not occur unless $m_A \geq m_B + m_C$. (Similarly the expression for $\|\mathbf{p}_B\|$ would become imaginary if $m_A < m_B + m_C$.)

Energy, momenta, and velocities in the center of momentum system.

Suppose we have particles B and C with energy momenta p_B and p_C in some coordinate system, and we want to find the expression for their energy, momenta, and velocity in their center of momentum system.

To find these values for particles B and C we can apply the following trick. Consider an imaginary particle, A , whose energy momentum vector is $p_B + p_C$. Its mass² is given by

$$m_A^2 = m_B^2 + m_C^2 + 2p_B \cdot p_C.$$

Plugging this value for m_A into the formulas of the previous subsection gives the desired answers.

Colliding beam versus stationary target.

A beam of particles of type A smashes into a stationary target of particles of type A , in the hope of producing the reaction

$$A + A \mapsto A + A + A + \bar{A}$$

where \bar{A} denotes the antiparticle of A . (All we have to know about the antiparticle is that it has the same mass as A .) What is kinetic energy needed to produce this reaction?

Let m denote the mass of A . In the laboratory frame, the stationary, target particle has energy momentum vector $(m, 0, 0, 0)$ while the incoming particle has energy momentum vector (E, \mathbf{p}) . Thus before the collision, we have $p_{\text{tot}} = (E + m, \mathbf{p})$ and hence

$$p_{\text{tot}}^2 = (E + m)^2 - \|\mathbf{p}\|^2.$$

In the center of momentum frame, the threshold for production of the four particles will be when there is no energy left over for motion, so they are all four

at rest. The total energy momentum vector, call it q , for the four particles will then be $q = (4m, 0, 0, 0)$ in the center of momentum system, hence $q^2 = (4m)^2$ and so $p_{\text{tot}}^2 = q^2$ implies

$$(E + m)^2 - \|\mathbf{p}\|^2 = (4m)^2.$$

But $E^2 - \|\mathbf{p}\|^2 = m^2$ so we get

$$2mE + m^2 = 15m^2$$

or

$$E = 7m.$$

In ordinary units we would write this as

$$E = 7mc^2.$$

Now $E = mc^2 + \text{kinetic energy} + \dots$ so approximately $6mc^2$ of kinetic energy must be supplied.

On the other hand, if we shoot two beams of particles of type A against one another, then for the collision, the laboratory frame and the center of momentum frame coincide, and the incoming total energy momentum vector is $(2E, 0, 0, 0)$ and our conservation equation becomes $4E = 4m$. We thus must supply kinetic energy equal to about m to each particle, or a total energy of about $2mc^2$ in ordinary units. Comparing the two experiments we see that the colliding beam experiment is more energy efficient (by a factor of three). Today virtually all new machines for collision experiments are colliders for this reason.

7.2.4 Mandelstam variables.

We consider a two body scattering event with a two body outcome, so

$$A + B \rightarrow C + D.$$

Both the incoming and the outgoing particles can exist in various states, and it is the role of any quantum mechanical theory to yield a probability amplitude for a pair of incoming states to scatter into a pair of outgoing states. In general, the states are characterized by various “internal” parameters such as spin, isospin etc., in addition to their momentum. However we shall consider the situation where the only important parameters describing the states are their momenta. So the quantum mechanical theory is to provide the transition amplitude, $T(p_A, p_B, p_C, p_D)$, a complex number such that $|T|^2$ gives the relative probability of two entering states with energy momentum vectors p_A and p_B to scatter to the outgoing states with energy momenta p_C and p_D . This looks like a function of four vectors, i.e. of sixteen variables, but Lorentz invariance and conservation of energy momentum implies that there are only two free variables. Indeed, Lorentz invariance implies that T should be a function of various scalar products $p_A^2, p_A \cdot p_B$ etc., of which there are ten in all. Of

these, $p_A^2 = m_A^2, p_B^2 = m_B^2, p_C^2 = m_C^2$, and $p_D^2 = m_D^2$ are parameters of the particles, and hence do not vary, leaving the six products of the form $p_A \cdot p_B$ etc. as variables. But these are constrained by conservation of energy momentum, $p_A + p_B = p_C + p_D$ which provides four equations, leaving only two of the products independent. It turns out, for reasons of “crossing symmetry” that is convenient to use two of the three **Mandelstam variables** defined by

$$s := c^{-2}(p_A + p_B)^2 \quad (7.33)$$

$$t := c^{-2}(p_A - p_C)^2 \quad (7.34)$$

$$u := c^{-2}(p_A - p_D)^2 \quad (7.35)$$

as independent variables. Now conservation of energy momentum implies that $p_A \cdot (p_C + p_D) = p_A \cdot (p_A + p_B)$. Hence

$$s + t + u = m_A^2 + m_B^2 + m_C^2 + m_D^2 \quad (7.36)$$

gives the relation between the three Mandelstam variables. Although the Mandelstam variables are important for theoretical work, the parameters that are measured in the laboratory are incoming and outgoing energies and scattering angle. It therefore becomes useful to express these laboratory parameters in terms of the Mandelstam variables.

Energies in terms of s .

By the definition of s , we see that the total energy in the center of momentum system is given by

$$E_A^{\text{CM}} + E_B^{\text{CM}} = E_C^{\text{CM}} + E_D^{\text{CM}} = c^2 \sqrt{s}. \quad (7.37)$$

To find the energy of A in the center of momentum system, we again employ the trick of thinking of a fictitious particle of energy momentum $p_C + p_D$ (hence at rest in the CM system and with mass \sqrt{s}) decaying into particles A and B with energy momenta p_A and p_B and apply (7.30) to obtain

$$E_A^{\text{CM}} = \frac{(s + m_A^2 - m_B^2)c^2}{2\sqrt{s}}. \quad (7.38)$$

To find the laboratory energy of particle A (where particle B is at rest) we go through an argument similar to that used in deriving (7.30). As usual we will do the derivation in a system of where $c = 1$. We have $p_B = (m_B, 0, 0, 0)$ and $p_A = p_C + p_D - p_B$ so

$$\begin{aligned} m_A^2 &= (p_C + p_D)^2 + m_B^2 - 2p_B \cdot (p_C + p_D) \\ &= s + m_B^2 - 2p_B(p_A + p_B) \\ &= s - m_B^2 - 2m_B E_A \end{aligned}$$

Solving gives

$$E_A = \frac{s - m_B^2 - m_A^2}{2m_B}.$$

Reverting to general coordinates gives

$$E_A^{\text{Lab}} = \frac{(s - m_A^2 - m_B^2)c^2}{2m_B}. \quad (7.39)$$

Angles in terms of Mandelstam variables.

We will study the special case where $m_C = m_A$ and $m_D = m_B$, for example when the outgoing and incoming particles are the same. The variable t is called the **momentum transfer**. Conservation of energy momentum says that

$$q := p_A - p_C = p_D - p_B$$

and the definition that

$$t = q^2.$$

(We work in units where $c = 1$.) Squaring both sides of $p_D = p_B + q$ and using the assumption that $m_B = m_D$ gives

$$t = -2p_B \cdot q.$$

In the Laboratory frame where B is at rest so $p_B = (m_B, 0, 0, 0)$ this becomes

$$t = -2m_B(E_A - E_C).$$

Suppose that A is a very light particle, practically of mass zero, so that $p_A \doteq (||\mathbf{k}_A||, \mathbf{k}_A)$ and $p_C \doteq (||\mathbf{k}_C||, \mathbf{k}_C)$. Then

$$\begin{aligned} t &= q^2 \\ &= (||\mathbf{k}_A|| - ||\mathbf{k}_C||)^2 - ||\mathbf{k}_A - \mathbf{k}_C||^2 \\ &= -2||\mathbf{k}_A||||\mathbf{k}_C||(1 - \cos \theta) \\ &= -4||\mathbf{k}_A||||\mathbf{k}_C|| \sin^2 \theta/2, \end{aligned}$$

where θ , the scattering angle, is the angle between \mathbf{k}_A and \mathbf{k}_C . Substituting $t = -2m_B(||\mathbf{k}_A|| - ||\mathbf{k}_C||)$ into the above expression gives

$$2m_B(||\mathbf{k}_A|| - ||\mathbf{k}_C||) = 4||\mathbf{k}_A||||\mathbf{k}_C|| \sin^2 \theta/2$$

or

$$\frac{||\mathbf{k}_A||}{||\mathbf{k}_C||} = 1 + 2\frac{||\mathbf{k}_A||}{m_B} \sin^2 \theta/2.$$

If we assume that $||\mathbf{k}_A||$ is small in comparison to m_B then

$$||\mathbf{k}_A|| \doteq ||\mathbf{k}_C||$$

and we get

$$t \doteq -4||\mathbf{k}_A||^2 \sin^2 \theta/2. \quad (7.40)$$

This formula is for a light particle of moderate energy scattering off a massive particle.

The general expression is a bit more messy, but not much. We wish to find the angle, θ , between the incoming momentum \mathbf{p}_A and the outgoing momentum \mathbf{p}_C in the rest frame of B in terms of the Mandelstam variables and the masses. In this frame we have

$$p_A \cdot p_C = E_A E_C - \|\mathbf{p}_A\| \|\mathbf{p}_C\| \cos \theta.$$

So we will proceed in two steps: first to express $E_A, E_C, \|\mathbf{p}_A\|, \|\mathbf{p}_C\|$ in terms of the four dimensional scalar products (and the masses) and then to express the scalar products in terms of the Mandelstam variables. We have

$$\begin{aligned} E_A &= \frac{p_A \cdot p_B}{m_B} \\ E_C &= \frac{p_C \cdot p_B}{m_B} \\ \|\mathbf{p}_A\| &= \sqrt{E_A^2 - m_A^2} \\ \|\mathbf{p}_C\| &= \sqrt{E_C^2 - m_C^2} \text{ so} \\ \cos \theta &= \frac{E_A E_C - p_A \cdot p_C}{\sqrt{(E_A^2 - m_A^2)(E_C^2 - m_C^2)}}. \end{aligned}$$

If we denote the common value of m_A and m_C by m we have

$$\cos \theta = \frac{(p_A \cdot p_B)(p_C \cdot p_B) - m_B^2 p_A \cdot p_C}{\sqrt{[(p_A \cdot p_B)^2 - m_B^2 m^2][(p_C \cdot p_B)^2 - m_B^2 m^2]}}. \quad (7.41)$$

To complete the program observe that it follows from the definitions that

$$\begin{aligned} 2p_A \cdot p_B &= s - m_A^2 - m_B^2 \\ 2p_A \cdot p_C &= m_A^2 + m_C^2 - t \text{ and} \\ 2p_C \cdot p_B &= m_A^2 + m_B^2 - u. \end{aligned}$$

Substituting these values into (7.41) gives us our desired expression.

By the way, equation (7.41) has a nice interpretation in terms of the scalar product induced on the space of exterior two vectors. We have

$$\begin{aligned} (p_A \wedge p_B) \cdot (p_C \wedge p_B) &= (p_A \cdot p_C)(p_B \cdot p_B) - (p_A \cdot p_B)(p_C \cdot p_B) \\ &= -[(p_A \cdot p_B)(p_C \cdot p_B) - m_B^2 p_A \cdot p_C] \text{ while} \\ \|p_A \wedge p_B\|^2 &= m_B^2 m_A^2 - (p_A \cdot p_B)^2 \text{ and} \\ \|p_C \wedge p_B\|^2 &= m_B^2 m_C^2 - (p_C \cdot p_B)^2. \end{aligned}$$

(The two last expressions are negative, since the two plane spanned by two timelike vectors has signature $+$ - .) We can thus write (7.41) as

$$\cos \theta = -\frac{(p_A \wedge p_B) \cdot (p_C \wedge p_B)}{\sqrt{\|p_A \wedge p_B\|^2 \|p_C \wedge p_B\|^2}}. \quad (7.42)$$

7.3 Scattering cross-section and mutual flux.

Let us go back to the expression $\|p_A \wedge p_B\|^2 = m_B^2 m_A^2 - (p_A \cdot p_B)^2$ from the end of the last section. In terms of a given space time splitting with unit timelike vector ∂_0 , we can write

$$\begin{aligned} p_A &= E_A \partial_0 + \mathbf{p}_A, \\ p_B &= E_B \partial_0 + \mathbf{p}_B \text{ so} \\ p_A \wedge p_B &= \mathbf{p}_A \wedge \mathbf{p}_B + \partial_0 \wedge (E_A \mathbf{p}_B - E_B \mathbf{p}_A) \text{ and hence} \\ \|p_A \wedge p_B\|^2 &= \|\mathbf{p}_A \times \mathbf{p}_B\|_{\mathbf{R}^3}^2 - \|E_A \mathbf{p}_B - E_B \mathbf{p}_A\|_{\mathbf{R}^3}^2 \end{aligned}$$

where \times denotes the cross product in \mathbf{R}^3 and the norms in the last expression are the three dimensional norms. In a frame where the momenta are aligned, such as a the CM frame where $\mathbf{p}_A = -\mathbf{p}_B$ or the laboratory frame where $\mathbf{p}_B = 0$, we have $\mathbf{p}_A \times \mathbf{p}_B = 0$. Recall our relativistic definition of velocity as \mathbf{p}/E . So in a frame where the momenta are aligned we have

$$\begin{aligned} \|E_A \mathbf{p}_B - E_B \mathbf{p}_A\|_{\mathbf{R}^3} &= E_A E_B \|\mathbf{v}\|_{\mathbf{R}^3} \text{ where} \\ \mathbf{v} &= \frac{1}{E_A} \mathbf{p}_A - \frac{1}{E_B} \mathbf{p}_B \end{aligned}$$

is the **mutual velocity**. So in such a frame we have

$$-\|p_A \wedge p_B\|^2 = E_A^2 E_B^2 \|\mathbf{v}\|_{\mathbf{R}^3}^2. \quad (7.43)$$

We want to apply this to the following situation which we first study in a fixed frame where the momenta are aligned.

A beam of particles of type A impacts on a target of particles of type B and some events of type f are observed. Let n_f denote the number of events of type f per unit time, so n_f has dimensions $(\text{time})^{-1}$. We assume that the target density is ρ_B with dimensions $(\text{vol})^{-1}$ and the beam density is ρ_A . We assume that the beam is well collimated and that all of its particles have approximately the same momentum, \mathbf{p}_A . The mutual flux per unit time (at time t) is

$$v \int \rho_A(t, \mathbf{x}) \rho_B(t, \mathbf{x}) d^3 \mathbf{x}$$

where

$$v := \|\mathbf{v}\|_{\mathbf{R}^3}$$

and \mathbf{v} is the mutual velocity. So the mutual flux has dimensions

$$\frac{(\text{distance})}{(\text{time})(\text{vol.})} = \frac{1}{(\text{time})(\text{area})}.$$

Thus

$$\frac{n_f}{v \int \rho_A(t, \mathbf{x}) \rho_B(t, \mathbf{x}) d^3 \mathbf{x}}$$

has the dimensions of area. Similarly, if we integrate the numerator and denominator with respect to time, the corresponding quotient will have the dimensions of area. Let $N_f := \int n_f dt$ be the total number of events of type f . Then, integrating the denominator as well,

$$\sigma_f := \frac{N_f}{v \int \rho_A(x) \rho_B(x) d^4x} \quad (7.44)$$

is called the **total cross-section** for events of type f . So it has the dimensions of area. The convenient unit is the *barn* (as in “he can’t hit the side of a barn”) where

$$1 \text{ barn} = 10^{-24} \text{ cm}^2.$$

The denominator in the expression for the total cross-section is called the **mutual flux**. It has a more invariant expression as follows: In the frame where the target particles are at rest, the “current” of the target particles (a three form) has the expression

$$J_B = \rho_B dx \wedge dy \wedge dz$$

while the current for the beam will have an expression of the form

$$J_A = \rho_A dx \wedge dy \wedge dz + dt \wedge (j_{A_z} dx \wedge dy + j_{A_y} dz \wedge dx + j_{A_x} dy \wedge dz).$$

So

$$\rho_A \rho_B = -J_A \cdot J_B.$$

Also, in this frame, $E_A E_B = p_A \cdot p_B$. So by (7.43) we have

$$\text{mutual flux} = \left| \frac{\|p_A \wedge p_B\|}{p_A \cdot p_B} \int J_A \cdot J_B d^4x \right|. \quad (7.45)$$

It is the function of any dynamical theory in quantum mechanics to make some predictions about the expected number of events of type f .

Chapter 8

Die Grundlagen der Physik.

This was the title of Hilbert's 1915 paper. It sounds a bit audacious, but let us try to put the ideas in a general context. We need to do a few computations in advance, so as not to disrupt the flow of the argument.

8.1 Preliminaries.

8.1.1 Densities and divergences.

If we regard \mathbf{R}^n as a differentiable manifold, the law for the change of variables for an integral involves the absolute value of the Jacobian determinant. This is different from the law of change of variables of a function (which is just substitution). [But it is close to the transition law for an n -form which involves the Jacobian determinant (not its absolute value).] For this reason we can not expect to integrate functions on a manifold. The objects that we *can* integrate are known as **densities**. We briefly recall two equivalent ways of defining these objects:

- **Coordinate chart description.** A density ρ is a rule which assigns to each coordinate chart (U, α) on M (where U is an open subset of M and $\alpha : U \rightarrow \mathbf{R}^n$) a function ρ_α defined on $\alpha(U)$ subject to the following transition law: If (W, β) is a second chart then

$$\rho_\alpha(v) = \rho_\beta(\beta \circ \alpha^{-1}(v)) \cdot |\det J_{\beta \circ \alpha^{-1}}|(v) \quad \text{for } v \in \alpha(U \cap W) \quad (8.1)$$

where $J_{\beta \circ \alpha^{-1}}$ denotes the Jacobian matrix of the diffeomorphism

$$\beta \circ \alpha^{-1} : \beta(U \cap V) \rightarrow \alpha(U \cap V).$$

Of course (8.1) is just the change of variables formula for an integrand in \mathbf{R}^n .

- **Tangent space description.** If V is an n -dimensional vector space, let $|\wedge V^*|$ denote the space of (real or complex valued) functions of n -tuplets

of vectors which satisfy

$$\sigma(Av_1, \dots, Av_n) = |\det A| \sigma(v_1, \dots, v_n). \quad (8.2)$$

The space $|\wedge V^*|$ is clearly a one-dimensional vector space. A density ρ is then a rule which assigns to each $x \in M$ an element of $|\wedge TM_x^*|$.

The relation between these two descriptions is the following: Let ρ be a density according to the tangent space description. Thus $\rho_x \in |\wedge TM_x^*|$ for every $x \in M$. Let (U, α) be a coordinate chart with coordinates x^1, \dots, x^n . Then on U we have the vector fields

$$\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}.$$

We can then evaluate ρ_x on the values of these vector fields at any $x \in U$, and so define

$$\rho_\alpha(\alpha(x)) = \rho_x \left(\left(\frac{\partial}{\partial x^1} \right)_x, \dots, \left(\frac{\partial}{\partial x^n} \right)_x \right).$$

If (W, β) is a second coordinate chart with coordinates y^1, \dots, y^n then on $U \cap W$ we have

$$\frac{\partial}{\partial x^j} = \sum \frac{\partial y^i}{\partial x^j} \frac{\partial}{\partial y^i}$$

and

$$J_{\beta \circ \alpha^{-1}} = \left(\frac{\partial y^i}{\partial x^j} \right)$$

so (8.1) follows from (8.2).

If (U, α) is a coordinate chart with coordinates x^1, \dots, x^n then the density defined on U by $\rho_\alpha \equiv 1$, that is by

$$\rho_x \left(\left(\frac{\partial}{\partial x^1} \right)_x, \dots, \left(\frac{\partial}{\partial x^n} \right)_x \right) = 1 \quad \forall x \in U$$

is denoted by dx . Every other density then has the local description Gdx on U where G is a function.

If $\phi : N \rightarrow M$ is a diffeomorphism and if ρ is a density on M , then the pull back $\phi^*\rho$ is the density on N defined by

$$(\phi^*\rho)_z(v_1, \dots, v_n) := \rho_{\phi(z)}(d\phi_z(v_1), \dots, d\phi_z(v_n)) \quad z \in N, \quad v_1, \dots, v_n \in TN_z.$$

(It is easy to check that this is indeed a density, i.e. that (8.2) holds at each $z \in N$.)

In particular, if X is a vector field on M generating a one parameter group

$$t \mapsto \phi_t = \exp tX$$

of diffeomorphisms, we can form the Lie derivative

$$D_X \rho := \frac{d}{dt} \phi_t^* \rho|_{t=0}.$$

We will need a local description of this Lie derivative. We can derive such a local description from Weil's formula for the Lie derivative of a differential form by the following device: Suppose that the manifold M is orientable and that we have chosen an orientation of M . This means that we have chosen a system of coordinate charts such that all the Jacobian determinants $\det J_{\beta \circ \alpha^{-1}}$ are positive. Relative to this system of charts, we can drop the absolute value sign in (8.1) since $\det J_{\beta \circ \alpha^{-1}} > 0$. But (8.1) without the absolute value sign is just the transition law for an n -form on the n -dimensional manifold M . In other words, *once we have chosen an orientation* on an orientable manifold M we can identify densities with n -forms. A fixed chart (U, α) carries the orientation coming from \mathbf{R}^n and our identification amounts to identifying the density dx with the n -form $dx^1 \wedge \cdots \wedge dx^n$.

If τ is an n -form on an n -dimensional manifold then Weil's formula

$$D_X \tau = i(X)d\tau + di(X)\tau$$

reduces to

$$D_X \tau = di(X)\tau$$

since $d\tau = 0$ as there are no non-zero $(n+1)$ forms on an n -dimensional manifold. If

$$X = X^1 \frac{\partial}{\partial x^1} + \cdots + X^n \frac{\partial}{\partial x^n}$$

and

$$\tau = G dx^1 \wedge \cdots \wedge dx^n$$

in terms of local coordinates then an immediate computation gives

$$di(X)\tau = \left(\sum_{i=1}^n \partial_i(GX^i) \right) dx^1 \wedge \cdots \wedge dx^n \quad (8.3)$$

where

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

It is useful to express this formula somewhat differently. It makes no sense to talk about a numerical value of a density ρ at a point x since ρ is not a function. But it *does* make sense to say that ρ does not vanish at x , since if $\rho_\alpha(\alpha(x)) \neq 0$ then (8.1) implies that $\rho_\beta(\beta(x)) \neq 0$. Suppose that ρ is a density which does not vanish anywhere. Then any other density on M is of the form $f \cdot \rho$ where f is a function. If X is a vector field, so that $D_X \rho$ is another density, then $D_X \rho$ is of the form $f\rho$ where f is a function, called the **divergence** of the vector field X relative to the non-vanishing density ρ and denoted by $\text{div}_\rho(X)$. In symbols,

$$D_X \rho = (\text{div}_\rho(X)) \cdot \rho.$$

We can then rephrase (8.3) as saying that

$$\text{div}_\rho(X) = \frac{1}{G} \sum_{i=1}^n \partial_i(GX^i) \quad (8.4)$$

in a local coordinate system where

$$X = X^1 \partial_1 + \cdots + X^n \partial_n$$

is the local expression for X and

$$Gdx$$

is the local expression for ρ .

8.1.2 Divergence of a vector field on a semi-Riemannian manifold.

Suppose that \mathbf{g} is a semi-Riemann metric on an n -dimensional manifold, M . Then \mathbf{g} determines a density, call it g , which assigns to every n tangent vectors, ξ_1, \dots, ξ_n at a point x the “volume” of the parallelepiped that they span:

$$g : \xi_1, \dots, \xi_n \mapsto |\det(\langle \xi_i, \xi_j \rangle)|^{\frac{1}{2}}. \quad (8.5)$$

If we replace the ξ_i by $A\xi_i$ where $A : TM_x \rightarrow TM_x$ the determinant is replaced by

$$\det(\langle A\xi_i, A\xi_j \rangle) = \det(A(\langle \xi_i, \xi_j \rangle)A^*) = (\det A)^2 \det(\langle \xi_i, \xi_j \rangle)$$

so we see that (8.2) is satisfied. So g is indeed a density, and since the metric is non-singular, the density g does not vanish at any point.

So if X is a vector field on M , we can consider its divergence $\operatorname{div}_g(X)$ with respect to g . Since g will be fixed for the rest of this subsection, we may drop the subscript g and simply write $\operatorname{div} X$. So

$$\operatorname{div} X \cdot g = D_X g. \quad (8.6)$$

On the other hand, we can form the covariant differential of X with respect to the connection determined by \mathbf{g} ,

$$\nabla X.$$

It assigns an element of $\operatorname{Hom}(TM_p, TM_p)$ to each $p \in M$ according to the rule

$$\xi \mapsto \nabla_\xi X.$$

The trace of this operator is a number, assigned to each point, p , i.e. a function known as the “contraction” of ∇X , so

$$C(\nabla X) := f, \quad f(p) := \operatorname{tr}(\xi \mapsto \nabla_\xi X).$$

We wish to prove the following formula

$$\operatorname{div} X = C(\nabla X). \quad (8.7)$$

We will prove this by computing both sides in a coordinate chart with coordinates, say, x^1, \dots, x^n . Let $dx = dx^1 dx^2 \cdots dx^n$ denote the standard density (the one which assigns constant value one to the $\partial_1, \dots, \partial_n$, $\partial_i := \partial/\partial x^i$). Then

$$g = Gdx, \quad G = |\det(\langle \partial_i, \partial_j \rangle)|^{\frac{1}{2}} = (\epsilon \det(\langle \partial_i, \partial_j \rangle))^{\frac{1}{2}}$$

where

$$\epsilon := \operatorname{sgn} \det(\langle \partial_i, \partial_j \rangle).$$

Recall the local formula (8.4) for the divergence:

$$\operatorname{div} X = \frac{1}{G} \sum_i \partial_i (X^i G).$$

Write

$$\Delta := \det(\langle \partial_i, \partial_j \rangle)$$

so

$$\begin{aligned} \frac{1}{G} \partial_i G &= \frac{1}{\sqrt{\epsilon \Delta}} \frac{1}{2\sqrt{\epsilon \Delta}} \frac{\partial(\epsilon \Delta)}{\partial x^i} \\ &= \frac{1}{2} \frac{1}{\Delta} \frac{\partial \Delta}{\partial x^i} \end{aligned}$$

independent of whether $\epsilon = 1$ or -1 . To compute this partial derivative, let us use the standard notation

$$g_{ij} := \langle \partial_i, \partial_j \rangle$$

so

$$\Delta = \det(g_{ij}) = \sum_j g_{ij} \Delta^{ij}$$

where we have expanded the determinant along the i -th row and the Δ^{ij} are the corresponding cofactors. If we think of Δ as a function of the n^2 variables, g_{ij} then, since none of the Δ_{ik} (for a fixed i) involve g_{ij} , we conclude from the above cofactor expansion that

$$\frac{\partial \Delta}{\partial g_{ij}} = \Delta^{ij} \tag{8.8}$$

and hence by the chain rule that

$$\frac{\partial \Delta}{\partial x^k} = \sum_{ij} \Delta^{ij} \frac{\partial g_{ij}}{\partial x^k}.$$

But

$$\frac{1}{\Delta} (\Delta^{ij}) = (g_{ij})^{-1},$$

the inverse matrix of (g_{ij}) , which is usually denoted by

$$(g^{kl})$$

so we have

$$\frac{\partial \Delta}{\partial x^k} = \Delta \sum_{ij} g^{ij} \frac{\partial g_{ij}}{\partial x^k}$$

or

$$\frac{1}{G} \partial_k G = \frac{1}{2} \sum_{ij} g^{ij} \frac{\partial g_{ij}}{\partial x^k}.$$

Recall that

$$\Gamma_{bc}^a := \frac{1}{2} \sum_r g^{ar} \left(\frac{\partial g_{rb}}{\partial x^c} + \frac{\partial g_{rc}}{\partial x^b} - \frac{\partial g_{bc}}{\partial x^r} \right)$$

so

$$\sum \Gamma_{ba}^a = \frac{1}{2} \sum_{ar} g^{ar} \frac{\partial g_{ar}}{\partial x^b}$$

or

$$\sum_a \Gamma_{ka}^a = \frac{1}{G} \frac{\partial G}{\partial x^k}. \quad (8.9)$$

On the other hand, we have

$$\nabla_{\partial_i} X = \sum \frac{\partial X^j}{\partial x^i} \partial_j + \sum \Gamma_{ik}^j X^k \partial_k$$

so

$$C(\nabla X) = \sum \frac{\partial X^j}{\partial x^j} + \frac{1}{G} \sum X^j \frac{\partial G}{\partial x^j},$$

proving (8.7).

For later use let us go over one step of this proof. From (8.8) we can conclude, as above, that

$$\frac{\partial G}{\partial g_{ij}} = \frac{1}{2} G g^{ij}. \quad (8.10)$$

8.1.3 The Lie derivative of of a semi-Riemann metric.

We wish to prove

$$L_V \mathbf{g} = \mathcal{S} \nabla (V \downarrow). \quad (8.11)$$

The left hand side of this equation is the Lie derivative of the metric \mathbf{g} with respect to the vector field V . It is a rule which assigns a symmetric bilinear form to each tangent space. By definition, it is the rule which assigns to any pair of vector fields, X and Y , the value

$$(L_V \mathbf{g})(X, Y) = V \langle X, Y \rangle - \langle [V, X], Y \rangle - \langle X, [V, Y] \rangle.$$

The right hand side of (8.11) means the following: $V \downarrow$ denotes the linear differential form whose value at any vector field Y is

$$(V \downarrow)(Y) := \langle V, Y \rangle.$$

In tensor calculus terminology, \downarrow is the “lowering operator”, and it commutes with covariant differential. Since \downarrow commutes with ∇ , we have

$$\nabla (V \downarrow)(X, Y) = \nabla_X (V \downarrow)(Y) = \langle \nabla_X V, Y \rangle.$$

The symbol \mathcal{S} in (8.11) denotes symmetric sum, so that the right hand side of (8.11) when applied to X, Y is

$$\langle \nabla_X V, Y \rangle + \langle \nabla_Y V, X \rangle.$$

But now (8.11) follows from the identities

$$\begin{aligned} L_V \langle X, Y \rangle = V \langle X, Y \rangle &= \langle \nabla_V X, Y \rangle + \langle X, \nabla_V Y \rangle \\ \nabla_V X - [V, X] &= \nabla_X V \\ \nabla_V Y - [V, Y] &= \nabla_Y V. \end{aligned}$$

8.1.4 The covariant divergence of a symmetric tensor field.

Let \mathbf{T} be a symmetric “contravariant ” tensor field (of second order), so that in any local coordinate system \mathbf{T} has the expression

$$\mathbf{T} = \sum T^{ij} \partial_i \partial_j, \quad T^{ij} = T^{ji}.$$

If θ is a linear differential form, then we can “contract” \mathbf{T} with θ to obtain a vector field, $\mathbf{T} \cdot \theta$: In local coordinates, if

$$\theta = \sum a_i dx^i$$

then

$$\mathbf{T} \cdot \theta = \sum T^{ij} a_j \partial_i.$$

We can form the covariant differential, $\nabla \mathbf{T}$ which then assigns to every linear differential form a linear transformation of the tangent space at each point, and then form the contraction, $C(\nabla \mathbf{T})$. (Since \mathbf{T} is symmetric, we don't have to specify on “which of the upper indices” we are contracting.) We define

$$\operatorname{div} \mathbf{T} := C(\nabla \mathbf{T}),$$

called the covariant divergence of \mathbf{T} . It is a vector field. The purpose of this section is to explain the geometrical significance of the condition

$$\operatorname{div} \mathbf{T} = 0. \tag{8.12}$$

If \mathbf{S} is a “covariant ” symmetric tensor field so that

$$\mathbf{S} = \sum S_{ij} dx^i dx^j$$

in local coordinates, let $\mathbf{S} \bullet \mathbf{T}$ denote the double contraction. It is a function, given in local coordinates by

$$\mathbf{S} \bullet \mathbf{T} = \sum S_{ij} T^{ij}.$$

Thus \mathbf{T} can be regarded as a linear function on the space of all covariant symmetric tensors of compact support by the rule

$$\mathbf{S} \mapsto \int_M \mathbf{S} \bullet \mathbf{T}g,$$

where g is the volume density associated to \mathbf{g} . Let V be a vector field of compact support. Then $L_V \mathbf{g}$ is a symmetric tensor of compact support. We claim

Proposition 10 Equation (8.12) is equivalent to

$$\int_M (L_V \mathbf{g}) \bullet \mathbf{T}g = 0 \quad (8.13)$$

for all vector fields V of compact support.

Proof. Let $\theta := V \downarrow$ so $\mathbf{T} \cdot \theta$ is a vector field of compact support, and so

$$\int_M C(\nabla(\mathbf{T} \cdot \theta))g = \int_M L_{\mathbf{T} \cdot \theta}g = 0$$

by the divergence theorem. (Recall our notation: the symbol \cdot denotes a “single” contraction, so that $\mathbf{T} \cdot \theta$ is a vector field.)

On the other hand,

$$\nabla(\mathbf{T} \cdot \theta) = (\nabla \mathbf{T}) \cdot \theta + \mathbf{T} \cdot \nabla \theta.$$

Apply the contraction, C :

$$\begin{aligned} 2C(\mathbf{T} \cdot \nabla \theta) &= 2\mathbf{T} \bullet \nabla \theta \\ &= \mathbf{T} \bullet L_V \mathbf{g}, \end{aligned}$$

using the fact that \mathbf{T} is symmetric and (8.11). So

$$2\mathbf{T} \bullet \nabla V \downarrow = \mathbf{T} \bullet L_V \mathbf{g}$$

and hence $\int_M \mathbf{T} \bullet L_V \mathbf{g}g = 0$ for all V of compact support if and only if $\int_M (\operatorname{div} \mathbf{T} \cdot \theta)g = 0$ for all θ of compact support. If $\operatorname{div} \mathbf{T} \neq 0$, we can find a point p and a linear differential form θ such that $\operatorname{div} \mathbf{T} \cdot \theta(p) > 0$ at some point, p . Multiplying θ by a blip function ϕ if necessary, we can arrange that θ has compact support and $\operatorname{div} \mathbf{T} \geq 0$ so that $\int_M (\operatorname{div} \mathbf{T} \cdot \theta)g > 0$.

Let us write $\ell_{\mathbf{T}}$ for the linear function on the space of smooth covariant tensors of compact support given by

$$\ell_{\mathbf{T}}(\mathbf{S}) := \int_M \mathbf{S} \bullet \mathbf{T}g.$$

We can rewrite (8.13) as

$$\ell(L_V \mathbf{g}) = 0 \quad \forall V \text{ of compact support} \quad (8.14)$$

when $\ell = \ell_{\mathbf{T}}$.

We can ask about condition (8.14) for different types of linear functions, ℓ . For example, consider a “delta tensor concentrated at a point”, that is a linear function of the form

$$\ell(\mathbf{S}) = \mathbf{S}(p) \bullet t$$

where t is a (“contravariant”) symmetric tensor defined at the point $p \in M$. We claim that no (non-zero) linear function of this form can satisfy (8.14). Indeed, let W be a vector field of compact support and let ϕ be a smooth function which vanishes at p . Set $V = \phi W$. Then

$$\nabla V \downarrow = d\phi \otimes W \downarrow + \phi \nabla W \downarrow$$

and the second term vanishes at p . Therefore condition (8.14) says that

$$0 = t \bullet (d\phi(p) \otimes W \downarrow(p)) = [t \cdot W \downarrow(p)] \cdot d\phi(p).$$

This says that the tangent vector $t \cdot (W \downarrow)(p)$ yields zero when applied to the function ϕ :

$$t \cdot (W \downarrow)(p) \phi = 0.$$

This is to hold for all ϕ vanishing at p , which implies that

$$t \cdot (W \downarrow)(p) = 0.$$

Now given any tangent vector, $w \in TM_p$ we can always find a vector field W of compact support such that $W(p) = w$. Hence the preceding equation implies that $t \cdot w \downarrow = 0 \forall w \in TM_p$ which implies that $t = 0$.

Let us turn to the next simplest case, a “delta tensor concentrated on a curve”. That is, let $\gamma : I \rightarrow M$ be a smooth curve and let τ be a continuous function which assigns to each $s \in I$ a symmetric contravariant tensor, $\tau(s)$ at the point $\gamma(s)$. Define the linear function ℓ_τ on the space of covariant symmetric tensor fields of compact support by

$$\ell_\tau(\mathbf{S}) = \int_I \mathbf{S}(\gamma(s)) \bullet \tau(s) ds.$$

Let us examine the implications of (8.14) for $\ell = \ell_\tau$. Once again, let us choose $V = \phi W$, this time with $\phi = 0$ on γ . We then get that

$$\int_I [\tau(s) \cdot W \downarrow(s)] \phi(s) ds = 0$$

for all vector fields W and all functions ϕ of compact support vanishing on γ . This implies that for each s , the tangent vector $\tau(s) \cdot w \downarrow$ is tangent to the curve γ for any tangent vector w at $\gamma(s)$. (For otherwise we could find a function ϕ which vanished on γ and for which $[\tau(s) \cdot w] \phi \neq 0$. By extending w to a vector field W with $W(\gamma(s)) = w$ and modifying ϕ if necessary so as to vanish outside a small neighborhood of $\gamma(s)$ we could then arrange that the integral on the left hand side of the preceding equation would not vanish.)

The symmetry of $\tau(s)$ then implies that $\tau(s) = c(s)\gamma'(s) \otimes \gamma'(s)$ for some scalar function, c . (Indeed, in local coordinates suppose that $\gamma'(s) = \sum v^i \partial_{i\gamma(s)}$ and $\tau(s) = \sum t^{ij} \partial_{i\gamma(s)} \partial_{j\gamma(s)}$. Applied successively to the basis vectors $w = \partial_{i\gamma(s)}$ we conclude that $t^{ij} = c^i v^j$ and hence from $t^{ij} = t^{ji}$ that $t^{ij} = ct^{ji}$.)

Let us assume that $\tau(s) \neq 0$ so $c(s) \neq 0$. Changing the parameterization means multiplying $\gamma'(s)$ by a scalar factor, and hence multiplying τ by a positive factor. So by reparametrizing the curve we can arrange that $\tau = \pm \gamma' \otimes \gamma'$. To avoid carrying around the \pm sign, let us assume that $\tau = \gamma' \otimes \gamma'$. Since multiplying τ by -1 does not change the validity of (8.14), we may make this choice without loss of generality.

Again let us choose $V = \phi W$, but this time with no restriction on ϕ , but let us use the fact that $\tau(s) = \gamma'(s) \otimes \gamma'(s)$. We get

$$\begin{aligned} \tau \cdot \nabla V \downarrow &= \tau \cdot [d\phi \otimes W \downarrow + \phi \nabla W \downarrow] \\ &= (\gamma' \phi) \langle \gamma', W \rangle + \phi \langle \nabla_{\gamma'} W, \gamma' \rangle \\ &= \gamma' (\phi \langle \gamma', W \rangle) - \phi \langle W, \nabla_{\gamma'} \gamma' \rangle. \end{aligned}$$

The integral of this expression must vanish for every vector field and every function ϕ of compact support. We claim that this implies that $\nabla_{\gamma'} \gamma' \equiv 0$, that γ is a geodesic! Indeed, suppose that $\nabla_{\gamma'(s)} \gamma'(s) \neq 0$ for some value, s_0 , of s . We could then find a tangent vector w at $\gamma(s_0)$ such that $\langle w, \nabla_{\gamma'(s_0)} \gamma'(s_0) \rangle = 1$ and then extend w to a vector field W , and so $\langle W \nabla_{\gamma'(s)} \gamma'(s) \rangle > 0$ for all s near s_0 . Now choose $\phi \geq 0$ with $\phi(s_0) = 1$ and of compact support. Indeed, choose ϕ to have support contained in a small neighborhood of $\gamma(s_0)$, so that

$$\int_I \gamma' (\phi \langle \gamma', W \rangle) = (\phi \langle \gamma', W \rangle) (\gamma(b)) - (\phi \langle \gamma', W \rangle) (\gamma(a)) = 0$$

where $a < s_0 < b$ are points in I with $\gamma(b)$ and $\gamma(a)$ outside the support of ϕ . We are thus left with

$$\ell_\tau(L_{\phi W}(\mathbf{g})) = - \int_a^b \phi \langle W, \nabla_{\gamma'} \gamma' \rangle ds < 0.$$

Conversely, if γ is a geodesic and $\tau = \gamma' \otimes \gamma'$ then

$$\tau \bullet \nabla V \downarrow = \langle \nabla_{\gamma'} V, \gamma' \rangle = \gamma' \langle V, \gamma' \rangle - \langle V, \nabla_{\gamma'} \gamma' \rangle.$$

The second term vanishes since γ is a geodesic, and the integral of the first term vanishes so long as γ extends beyond the support of V or if γ is a closed curve. We have thus proved a remarkable theorem of Einstein, Infeld and Hoffmann (1938)

Theorem 1 *If τ is a continuous (contravariant second order) symmetric tensor field along a curve γ whose associated linear function, ℓ_τ satisfies (8.14) then we can reparametrize γ so that it becomes a geodesic and so that $\tau = \pm \gamma' \otimes \gamma'$. Conversely, if τ is of this form and if γ is unbounded or closed then ℓ_τ satisfies (8.14).*

(Here “unbounded” means that for any compact region, K , there are real numbers a and b such that $\gamma(s) \notin K$, $\forall s > b$ or $< a$.)

8.2 Varying the metric and the connection.

We will regard the space of smooth covariant symmetric tensor fields \mathbf{S} such as those we considered in the preceding section as the “compactly supported piece” of the tangent space to a given metric \mathbf{g} . This is to be interpreted in the following sense: Let \mathcal{M} denote the space of all semi-Riemann metrics on a manifold, M , say all with a fixed signature. If $\mathbf{g} \in \mathcal{M}$ is a particular metric, and if \mathbf{S} is a compactly supported symmetric tensor field, then

$$\mathbf{g} + t\mathbf{S}$$

is again a metric of the same signature for sufficiently small $|t|$. So we can regard \mathbf{S} as the infinitesimal variation in \mathbf{g} along this “line segment” of metrics. On the other hand, if \mathbf{g}_t is any curve of metrics depending smoothly on t , and with the property that $\mathbf{g}_t = \mathbf{g}$ outside some fixed compact set, K , then

$$\mathbf{S} := \left. \frac{d\mathbf{g}_t}{dt} \right|_{t=0}$$

is a symmetric tensor field of compact support.

So we will denote the space of all compactly supported smooth fields of symmetric covariant two tensors by

$$T\mathcal{M}_{\text{compact}}.$$

Notice that we have identified this fixed vector space as the (compactly supported) tangent space at every point, \mathbf{g} in the space of metrics. We have “trivialized” the tangent bundle to \mathcal{M} .

The space of all (symmetric) connections also has a natural trivialization. Indeed, let ∇ and ∇' be two connections. Then

$$\nabla_f X Y - \nabla'_f X Y = f \nabla_X Y - f \nabla'_X Y.$$

In other words, the map

$$A : (X, Y) = \nabla_X Y - \nabla'_X Y$$

is a tensor; its value at any point p depends only on the values of X and Y at p . We can say that

$$A = \nabla - \nabla'$$

is a tensor field, of type $T^* \otimes T^* \otimes T$ (one which assigns to every tangent vector at $p \in M$ an element of $\text{Hom}(TM_p, TM_p)$).

Conversely, if A is any such tensor field and if ∇' is any connection then $\nabla = \nabla' + A$ is another connection. Thus the space of all connections is an *affine space* whose associated *linear space* is the space of all A 's. We will be interested in symmetric connections, in which case the A 's are restricted to being symmetric: $A_X Z = A_Z X$. (Check this as an exercise.) Let \mathcal{A} denote the space of all such (smooth) symmetric A and let \mathcal{C} denote the space of all symmetric

connections. Then we can identify the “tangent space” to \mathcal{C} at any connection ∇ with the space \mathcal{A} , because in any affine space we can identify the tangent space at any point with the associated linear space. In symbols, we may write

$$T\mathcal{C}_\nabla = \mathcal{A},$$

independent of the particular ∇ . Once again we will be interested in variations of compact support in the connection, so we will want to consider the space

$$\mathcal{A}_{\text{compact}}$$

consisting of tensor fields of our given type of compact support.

The Levi-Civita map assigns to every Riemann metric a symmetric connection. So it can be considered as a map, call it $L.C.$, from metrics to connections:

$$L.C. : \mathcal{M} \rightarrow \mathcal{C}.$$

The value of $L.C.(\mathbf{g})$ at any point depends only on g_{ij} and its first derivatives at the point, and hence the differential of the Levi-Civita map can be considered as a linear map

$$d(L.C.)_{\mathbf{g}} : T\mathcal{M}_{\text{compact}} \rightarrow \mathcal{A}_{\text{compact}}.$$

(The spaces on both sides are independent of \mathbf{g} but the differential definitely depends on \mathbf{g} .) In what follows, we will let A denote the value of this differential at a given \mathbf{g} and $\mathbf{S} \in T\mathcal{M}_{\text{compact}}$:

$$A := d(L.C.)_{\mathbf{g}}[\mathbf{S}].$$

As an exercise, you should compute the expression for A in terms of $\nabla\mathbf{S}$ where $\nabla = L.C.(\mathbf{g})$ is the Levi-Civita connection associated to \mathbf{g} .

The map R associates to every metric its Riemann curvature tensor. The map Ric associates to every metric its Ricci curvature. For reasons that will soon become apparent, we need to compute the differentials of these maps.

The curvature is expressed in terms of the connection:

$$R_{X,Y} = \nabla_{[X,Y]} - [\nabla_X, \nabla_Y].$$

So we may think of the right hand side of this equation as defining a map, curv , from the space of connections to the space of tensors of curvature type. Differentiating this expression using Leibniz’s rule gives, for any $A \in \mathcal{A}$,

$$(d\text{curv}_\nabla[A])(X, Y) = A_{[X,Y]} - A_X \nabla_Y - \nabla_X A_Y + A_Y \nabla_X + \nabla_Y A_X.$$

We have

$$[X, Y] = \nabla_X Y - \nabla_Y X$$

so

$$A_{[X,Y]} = A_{\nabla_X Y} - A_{\nabla_Y X}.$$

On the other hand, the covariant differential, ∇A of the tensor field A with respect to the connection, ∇ is given by

$$(\nabla A)(X, Y)Z = \nabla_X(A_Y Z) - A_{\nabla_X Y} Z - A_Y \nabla_X Z$$

or, more succinctly,

$$(\nabla A)(X, Y) = \nabla_X A_Y - A_{\nabla_X Y} - A_Y \nabla_X.$$

From this we see that

$$(d\text{curv}_{\nabla}[A])(X, Y) = (\nabla A)(Y, X) - (\nabla A)(X, Y).$$

If we let $\tilde{\nabla}A$ denote the tensor obtained from ∇A by $\tilde{\nabla}A(X, Y) = \nabla A(Y, X)$ we can write this equation even more succinctly as

$$d\text{curv}_{\nabla}[A] = \tilde{\nabla}A - \nabla A.$$

If we substitute $A = d(L.C.)_{\mathbf{g}}[\mathbf{S}]$ into this equation we get, by the chain rule, the value of $dR_{\mathbf{g}}[\mathbf{S}]$. Taking the contraction, C , which yields the Ricci tensor from the Riemann tensor, we obtain

$$d\text{Ric}_{\mathbf{g}}[\mathbf{S}] = C(\tilde{\nabla}A - \nabla A).$$

Let $\hat{\mathbf{g}}$ denote the contravariant symmetric tensor corresponding to \mathbf{g} , the scalar product induced by \mathbf{g} on the cotangent space at each point. Thus, for example, the scalar curvature, S , is obtained from the Ricci curvature by contraction with $\hat{\mathbf{g}}$:

$$S = \hat{\mathbf{g}} \bullet \text{Ric}.$$

Contracting the preceding equation with $\hat{\mathbf{g}}$ and using the fact that ∇ commutes with contraction with $\hat{\mathbf{g}}$ and with C we obtain

$$\hat{\mathbf{g}} \bullet d\text{Ric}_{\mathbf{g}}[\mathbf{S}] = C(\nabla V)$$

where V is the vector field

$$V := C(A) \uparrow - \hat{\mathbf{g}} \cdot A.$$

We have $C(\nabla V) = \text{div}V$. Also V has compact support since \mathbf{S} does. Hence we obtain, from the divergence theorem, the following important result:

$$\int_M \hat{\mathbf{g}} \bullet d\text{Ric}_{\mathbf{g}}[\mathbf{S}] g = 0. \quad (8.15)$$

8.3 The structure of physical laws.

8.3.1 The Legendre transformation.

Let f be a function of one real variable. We can consider the map $t \mapsto f'(t)$ which is known as the Legendre transformation, or the “point slope transformation”,

$\mathcal{L}(f)$, associated to f . For example, if $f = \frac{1}{2}kt^2$ then the associated Legendre transformation is the linear map $t \mapsto kt$. As for any transformation, we might be interested in computing its inverse. That is, find the (or a) point t with a given value of $f'(t)$.

For a function, f , of two variables we can make the same definition and pose the same question: Define $\mathcal{L}(f)$ as the map

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto (\partial f / \partial x, \partial f / \partial y).$$

Given (a, b) we may ask to solve the equations

$$\begin{aligned} \partial f / \partial x &= a \\ \partial f / \partial y &= b \end{aligned}$$

for $\begin{pmatrix} x \\ y \end{pmatrix}$.

The general situation is as follows: Suppose that \mathcal{M} is a manifold whose tangent bundle is trivialized, i.e. that we are given a smooth identification of $T\mathcal{M}$ with $\mathcal{M} \times \mathcal{V}$, all the tangent spaces are identified with a fixed vector space, \mathcal{V} . Of course this also gives an identification of all the cotangent spaces with the fixed vector space \mathcal{V}^* . In this situation, if F is a function on \mathcal{M} , the associated Legendre transformation is the map

$$\mathcal{L}(F) : \mathcal{M} \rightarrow \mathcal{V}^*, \quad \mathbf{x} \mapsto dF_{\mathbf{x}}.$$

In particular, given $\ell \in \mathcal{V}^*$, we may ask to find $\mathbf{x} \in \mathcal{M}$ which solves the equation

$$dF_{\mathbf{x}} = \ell. \tag{8.16}$$

This is the “source equation” of physics, with the caveat that the function F need not be completely defined. Nevertheless, its differential might be defined, provided that we restrict to “variations with compact support” as is illustrated by the following example:

In Newtonian physics, the background is Euclidean geometry and the objects are conservative force fields which are linear differential forms that are closed. With a mild loss of generality let us consider “potentials” instead of force fields, so the objects are functions, ϕ on Euclidean three space. Our space \mathcal{M} consists of all (smooth) functions. Since \mathcal{M} is a vector space, its tangent space is automatically identified with \mathcal{M} itself, so $\mathcal{V} = \mathcal{M}$. The force field associated with the potential ϕ is $-d\phi$, and its “energy density” at a point is one half the Euclidean length². That is, the density is given by

$$\frac{1}{2}(\phi_x^2 + \phi_y^2 + \phi_z^2)$$

where subscript denotes partial derivative. We would like to define the function F to be the “total energy”

$$F(\phi) = \frac{1}{2} \int_{\mathbf{R}^3} (\phi_x^2 + \phi_y^2 + \phi_z^2) dx dy dz$$

but there is no reason to believe that this integral need converge. However, suppose that s is a smooth function of compact support, K . Thus s vanishes outside the closed bounded set, K . For any bounded set, B , the integral

$$F^B(\phi) := \frac{1}{2} \int_B (\phi_x^2 + \phi_y^2 + \phi_z^2) dx dy dz$$

converges, and the derivative

$$\left. \frac{dF^B[\phi + ts]}{dt} \right|_{t=0} = \int_{\mathbf{R}^3} (\phi_x s_x + \phi_y s_y + \phi_z s_z) dx dy dz$$

exists and is independent of B so long as $B \supset K$. So it is reasonable to define the right hand side of this equation as dF_ϕ evaluated at s :

$$dF_\phi[s] := \int_{\mathbf{R}^3} (\phi_x s_x + \phi_y s_y + \phi_z s_z) dx dy dz$$

even though the function F itself is not well defined. Of course to do so, we must not take $\mathcal{V} = \mathcal{M}$ but take \mathcal{V} to be of the subspace consisting of functions of compact support.

A linear function on \mathcal{V} is just a density, but in Euclidean space, with Euclidean volume density $dx dy dz$ we may identify densities with functions. Suppose that ρ is a smooth function, and we let ℓ_ρ be the corresponding element of \mathcal{V}^* ,

$$\ell_\rho(s) = \int_{\mathbf{R}^3} s \rho dx dy dz.$$

Equation (8.16) with $\ell = \ell_\rho$ becomes

$$\int_{\mathbf{R}^3} (\phi_x s_x + \phi_y s_y + \phi_z s_z) dx dy dz = \int_{\mathbf{R}^3} s \rho dx dy dz \quad \forall s \in \mathcal{V},$$

which is to be regarded as an equation for ϕ where ρ is given. We have

$$\phi_x s_x + \phi_y s_y + \phi_z s_z = (\phi_x s)_x + (\phi_y s)_y + (\phi_z s)_z - s \Delta \phi$$

where Δ is the Euclidean Laplacian,

$$\Delta \phi = \phi_{xx} + \phi_{yy} + \phi_{zz}.$$

Thus, since the total derivatives $(s\phi_x)_x$ etc. contribute zero to the integral, equation (8.16) is the Poisson equation

$$\Delta \phi = \rho.$$

As we know, a solution to this equation is given by convolution with the $1/r$ potential:

$$\phi(x, y, z) = \frac{1}{4\pi} \int \frac{\rho(\xi, \eta, \zeta)}{\sqrt{(x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2}} d\xi d\eta d\zeta$$

if ρ has compact support, for example, so that this integral converges. In this sense Euclidean geometry determines the $1/r$ potential.

8.3.2 The passive equations.

Symmetries of the function F may lead to constraints on the right hand side of (8.16). In our example of a function of two variables, suppose that the function f on the plane is invariant under rotations. Thus f would have to be a function of the radius, r , and hence the right hand side of (8.16) would have to be proportional to dr , and in particular, vanish on vectors tangent to the circle through the point $\begin{pmatrix} x \\ y \end{pmatrix}$.

More generally, suppose that \mathcal{G} is group acting on \mathcal{M} , and that the function F is invariant under the action of this group, i.e.

$$F(a \cdot \mathbf{x}) = F(\mathbf{x}) \quad \forall a \in \mathcal{G}.$$

Let $\mathcal{O} = \mathcal{G} \cdot \mathbf{x}$ denote the orbit through \mathbf{x} , so $\mathcal{G} \cdot \mathbf{x}$ consists of all points of the form $a\mathbf{x}$, $a \in \mathcal{G}$. Then the function F is constant on \mathcal{O} and so $dF_{\mathbf{x}}$ must vanish when evaluated on the tangent space to \mathcal{O} . We may write this symbolically as

$$\ell \in (T\mathcal{O})_{\mathbf{x}}^0 \tag{8.17}$$

if (8.16) holds. Of course, in the infinite dimensional situations where we want to apply this equation, we must use some imagination to understand what is meant by the tangent space to the orbit.

We want to consider what happens when we modify ℓ by adding to it a “small” element, $\mu \in \mathcal{V}^*$. Presumably the solution \mathbf{x} to our “source equation” (8.16) would then be modified by a small amount and so the tangent space to the orbit would change. We would then have to apply (8.17) to $\ell + \mu$ using the modified tangent space. [One situation where disregarding this change in \mathbf{x} could be justified is when $\ell = 0$. Presumably the modification of \mathbf{x} will be of first order in μ , and hence the change in (8.17) will be a second order effect which can be ignored if μ is small.]

A *passive equation* of physics is where we apply (8.17) but disregard the change in the tangent space and so obtain the equation

$$\mu \in (T\mathcal{M})_{\mathbf{x}}^0. \tag{8.18}$$

The justification for ignoring the non-linear effect of μ of \mathbf{x} may be problematical from our abstract point of view, but the equation we have just obtained for the passive reaction of μ to the presence of \mathbf{x} is a powerful principle of physics. About half the laws of physics are of this form

We have enunciated two principles of physics, a source equation (8.16) which amounts to inverting a Legendre transformation, and the passive equation (8.18) which is a consequence of symmetry. We now turn to how Hilbert and Einstein implemented these principles for gravity.

8.4 The Hilbert “function”.

The space \mathcal{M} is the space of Lorentzian metrics on a given manifold, M . Hilbert chooses as his function

$$F(\mathbf{G}) = - \int_M Sg, \quad S = \hat{\mathbf{g}} \cdot \text{Ric}(\mathbf{g}).$$

As discussed above, this “function” need not be defined since the integral in question need not converge. But the differential

$$dF_{\mathbf{g}}[\mathbf{S}]$$

will be defined when evaluated on a variation of compact support. The integral defining F involves \mathbf{g} at three locations: in the definition of the density g , in the dual metric $\hat{\mathbf{g}}$ and in Ric . Thus, by Leibniz’s rule

$$-dF_{\mathbf{g}}[\mathbf{S}] = \int_M \hat{\mathbf{g}} \cdot \text{Ric}(\mathbf{g}) dg[\mathbf{S}] + \int_M d\hat{\mathbf{g}}[\mathbf{S}] \cdot \text{Ric}(\mathbf{g}) + \int_M \hat{\mathbf{g}} \cdot d\text{Ric}_{\mathbf{g}}[\mathbf{S}].$$

We have already done the hard work involved in showing that the third integral vanishes, equation (8.15). So we are left with the first two terms.

As to the first term, the coordinate free way of rewriting (8.10) is

$$dg_{\mathbf{g}}[\mathbf{S}] = \frac{1}{2} \hat{\mathbf{g}} \cdot \mathbf{S}g.$$

As to the second term, recall that in local coordinates, $\hat{\mathbf{g}}$ is given by $\sum g^{ij} \partial_i \partial_j$ where (g^{ij}) is the inverse matrix of g_{ij} . So we recall a formula we derived for the differential of the inverse function of a matrix. If inv denotes the inverse function, so

$$\text{inv}(B) = B^{-1},$$

then it follows from differentiating the identity

$$BB^{-1} \equiv I$$

using Leibniz’s rule that

$$d \text{inv}_B[C] = -B^{-1}CB^{-1}.$$

It follows that the differential of the function $\mathbf{g} \mapsto \hat{\mathbf{g}}$ when evaluated at \mathbf{S} is $\mathbf{S} \uparrow \uparrow$, the contravariant symmetric tensor obtained from \mathbf{S} by applying the raising operator (coming from \mathbf{g}) twice. Now

$$(\mathbf{S} \uparrow \uparrow) \cdot \text{Ric} = \mathbf{S} \cdot \text{Ric} \uparrow \uparrow.$$

So if we define

$$\text{RIC} := \text{Ric} \uparrow \uparrow$$

to be the contravariant form of the Ricci tensor we obtain

$$dF_{\mathbf{g}}[\mathbf{S}] = \int_M (\text{RIC} - \frac{1}{2}S\hat{\mathbf{g}}) \cdot \mathbf{S}g. \quad (8.19)$$

This is left hand side of the source equation (8.16). The right hand side is a linear function on the space $T(\mathcal{M})_{\text{compact}}$. We know that if \mathbf{T} is a smooth symmetric tensor field, then it defines a linear function on $T(\mathcal{M})_{\text{compact}}$ given by

$$\ell_{\mathbf{T}}(\mathbf{S}) = \int_M \mathbf{S} \cdot \mathbf{T}g.$$

Thus for $\ell = \ell_{\mathbf{T}}$ equation (8.16) becomes the celebrated Einstein field equations

$$\text{RIC} - \frac{1}{2}S\hat{\mathbf{g}} = \mathbf{T} \quad (8.20)$$

So if we regard the physical objects as semiRiemann metrics, and if we believe that matter determines the metric, by a source type equation, then matter should be considered as a linear function on $T(\mathcal{M})_{\text{compact}}$. In particular a “smooth” matter distribution is a contravariant symmetric tensor field. If we believe that the laws of physics are described by the function given by Hilbert, we get the Einstein field equations. Modifying the function would change the source equations. For example, if we replace S by $S + c$ where c is a constant, this would have the effect of adding a term $\frac{1}{2}c\hat{\mathbf{g}}$ to the left hand side of the field equations. This is the notorious “cosmological constant” term.

We will take our group of symmetries to be the group of diffeomorphisms of M of compact support - diffeomorphisms which are the identity outside some compact set. Such transformations preserve the function F .

If V is a vector field of compact support which generates a one parameter group, ϕ_s of transformations, then these transformations have compact support, and the fact that the function F is invariant under these transformations translates into the assertion that $dF_{\mathbf{g}}[L_V\mathbf{g}] = 0$.

In other words, the “tangent space to the orbit through \mathbf{g} ” is the subspace of $T(\mathcal{M})_{\text{compact}}$ consisting of all $L_V\mathbf{g}$ where V is a vector field of compact support. From the results obtained above we now know that the passive equation translates into

$$\text{div}\mathbf{T} = 0$$

for a smooth tensor field and into

$$T = \pm\gamma' \otimes \gamma', \quad \gamma \text{ a geodesic}$$

for a continuous tensor field concentrated along a curve. These results are independent of the choice of F .

8.5 Schrodinger's equation as a passive equation.

In quantum mechanics, the background is a complex Hilbert space. In order to avoid technicalities, let us assume that \mathcal{H} is a finite dimensional complex vector space with an Hermitian scalar product. Let \mathcal{M} denote space of all self adjoint operators on \mathcal{H} . Let \mathcal{G} be the group of all unitary operators, and let \mathcal{G} act on \mathcal{M} by conjugation: $U \in \mathcal{G}$ acts on \mathcal{M} by sending

$$A \mapsto UAU^{-1}.$$

Since \mathcal{M} is a vector space, its tangent bundle is automatically trivialized. We may also identify the space of linear functions on \mathcal{M} with \mathcal{M} by assigning to $B \in \mathcal{M}$ the linear function ℓ_B defined by

$$\ell_B(A) = \text{tr } AB.$$

If C is a self adjoint matrix, the tangent to the curve

$$\exp(itC)A \exp(-itC)$$

at $t = 0$ is $i[C, A]$. So the "tangent space to the orbit through A " consists of all $i[C, A]$

Show that the passive equation (8.18) becomes

$$[A, B] = 0$$

for $\mu = \ell_B$. A linear function is called a *pure state* if it is of the form ℓ_B where B is projection onto a one dimensional subspace. This means that there is a unit vector $\phi \in \mathcal{H}$ (determined up to phase) so that

$$Bu = (u, \phi)\phi \quad \forall u \in \mathcal{H}$$

where $(,)$ denotes the scalar product on \mathcal{H} .

Show that a pure state satisfies (8.18) if and only if ϕ is an eigenvector of H :

$$H\phi = \lambda\phi$$

for some real number λ . This is the (time independent) Schrodinger equation.

8.6 Harmonic maps.

Let us return to equation (8.18) in the setting of the group of diffeomorphisms of compact support of a manifold M acting on the semi-Riemannian metrics. In the case that we our linear function μ was given by a "delta function tensor field supported along a curve" we saw that condition (8.18) implies that the curve γ is a geodesic and the tensor field is $\pm\gamma' \otimes \gamma'$ (under suitable reparametrization of the curve and assuming that the tensor field does not vanish anywhere on the

curve). We now examine what condition (8.18) says for a “delta function tensor field” on a more general submanifold. So we are interested in the condition

$$\mu(\mathcal{S}(\nabla V \downarrow)) = 0$$

for all V of compact support where μ is provided by the following data:

1. A k dimensional manifold Q and a proper map $f : Q \rightarrow M$,
2. A smooth section \mathbf{t} of $f^\#S^2(TM)$, so \mathbf{t} assigns to each $q \in Q$ an element $\mathbf{t}(q) \in S^2TM_{f(q)}$, and
3. A density ω on Q .

For any section s of S^2T^*M and any $q \in Q$ we can form the “double contraction” $s(q) \bullet \mathbf{t}(q)$ since $s(q)$ and $\mathbf{t}(q)$ take values in dual vector spaces, and since f is proper, if s has compact support then so does the function $q \mapsto s(q) \bullet \mathbf{t}(q)$ on Q . We can then form the integral

$$\mu[s] := \int_Q s(\cdot) \bullet \mathbf{t}(\cdot) \omega. \quad (8.21)$$

We observe (and this will be important in what follows) that μ depends on the tensor product $\mathbf{t} \otimes \omega$ as a section of $f^\#S^2TM \otimes \mathbf{D}$ where \mathbf{D} denotes the line bundle of densities of Q rather than on the individual factors.

We apply the equation $\mu(\mathcal{S}(\nabla V \downarrow)) = 0$ to this μ and to $v = \phi W$ where ϕ is a function of compact support and W a vector field of compact support on M . Since

$$\nabla(\phi W) = d\phi \otimes W + \phi \nabla W$$

and \mathbf{t} is symmetric, this becomes

$$\int_Q \mathbf{t} \bullet (d\phi \otimes W \downarrow + \phi \nabla W \downarrow) \omega = 0. \quad (8.22)$$

We first apply this to a ϕ which vanishes on $f(Q)$, so that the term $\phi \nabla W$ vanishes when restricted to Q . We conclude that the “single contraction” $\mathbf{t} \cdot \theta$ must be tangent to $f(Q)$ at all points for all linear differential forms θ and hence that

$$\mathbf{t} = df_* \mathbf{h}$$

for some section \mathbf{h} of $S^2(TQ)$.

Again, let us apply condition (8.22), but no longer assume that ϕ vanishes on $f(Q)$. For any vector field Z on Q let us, by abuse of language, write

$$Z\phi \quad \text{for} \quad Zf^*\phi,$$

for any function ϕ on M , write

$$\langle Z, W \rangle \quad \text{for} \quad \langle df_* Z, W \rangle_M$$

where W is a vector field on M , and

$$\nabla_Z W \quad \text{for} \quad \nabla_{df_* Z} W.$$

Write

$$\mathbf{h} = \sum h^{ij} e_i e_j$$

in terms of a local frame field e_1, \dots, e_k on Q . Then

$$\mathbf{t} \bullet (\nabla V \downarrow) = \sum h^{ij} [e_i(\phi) \langle e_j, W \rangle + \phi \langle \nabla_{e_i} W, e_j \rangle].$$

Now

$$\langle \nabla_{e_i} W, e_j \rangle = e_i \langle W, e_j \rangle - \langle W, \nabla_{e_i} e_j \rangle$$

so

$$\mathbf{t} \bullet \nabla V \downarrow = \sum_{ij} [h^{ij} e_i(\phi \langle e_j, W \rangle) - \phi \langle W, h^{ij} \nabla_{e_i} e_j \rangle].$$

Also,

$$\int_Q \sum h^{ij} e_i(\phi \langle e_j, W \rangle) \omega = - \int_Q \phi \langle e_j, W \rangle L_{\sum_i h^{ij} e_i} \omega.$$

Let us write

$$z^j = \operatorname{div}_\omega(\sum h^{ij} e_i)$$

so

$$L_{\sum_i h^{ij} e_i} \omega = z^j \omega.$$

If we set

$$Z := \sum z^j e_j$$

then condition (8.22) becomes

$$\sum_{ij} h^{ij} {}^M \nabla_{e_i} e_j = -Z, \tag{8.23}$$

where we have used ${}^M \nabla$ to emphasize that we are using the covariant derivative with respect to the Levi-Civita connection on M , i.e.

$${}^M \nabla_{e_i} e_j := \nabla_{df_* e_i} (df_* e_j).$$

To understand (8.23) suppose that we assume that \mathbf{h} is non-degenerate, and so induces a semi-Riemannian metric $\check{\mathbf{h}}$ on Q , and let us *assume* that ω is the volume form associated with $\check{\mathbf{h}}$. (In all dimensions except $k = 2$ this second assumption is harmless, since we can rescale \mathbf{h} to arrange it to be true.) Let ${}^{\mathbf{h}} \nabla$ denote covariant differential with respect to $\check{\mathbf{h}}$. Let us choose the frame field e_1, \dots, e_k to be “orthonormal” with respect to $\check{\mathbf{h}}$, i.e.

$$h^{ij} = \epsilon_j \delta_{ij}, \quad \text{where } \epsilon_j = \pm 1$$

so that

$$\sum_i h^{ij} e_i = \epsilon_j e - j.$$

Then

$$L_{e_j}\omega = C({}^{\mathbf{h}}\nabla e_j)\omega$$

and

$$C({}^{\mathbf{h}}\nabla e_j) = \sum_i \epsilon_i \langle {}^{\mathbf{h}}\nabla_{e_i} e_j, e_i \rangle_{\check{\mathbf{h}}} = - \sum_i \langle e_j, \epsilon_i {}^{\mathbf{h}}\nabla_{e_i} e_i \rangle_{\check{\mathbf{h}}},$$

so

$$Z = - \sum_j \sum_i \epsilon_j \langle e_j, \epsilon_i {}^{\mathbf{h}}\nabla_{e_i} e_i \rangle_{\check{\mathbf{h}}} e_j = - \sum_i \epsilon_i {}^{\mathbf{h}}\nabla_{e_i} e_i = - \sum_{ij} {}^{\mathbf{h}}\nabla_{e_i} e_j.$$

Given a metric $\check{\mathbf{h}}$ on Q , a metric \mathbf{g} on M , the **second fundamental form of a map** $f : Q \rightarrow M$, is defined as

$$B_f(X, Y) := \mathbf{g}\nabla_{df(X)}(df(Y)) - df({}^{\mathbf{h}}\nabla_X Y). \quad (8.24)$$

Here X and Y are vector fields on Q and $df(X)$ denotes the “vector field along f ” which assigns to each $q \in Q$ the vector $df_q(X_q) \in TM_{f(q)}$.

The **tension field** $\tau(f)$ of the map f (relative to a given \mathbf{g} and $\check{\mathbf{h}}$) is the trace of the second fundamental form so

$$\tau(f) = \sum_{ij} h^{ij} \mathbf{g}\nabla_{df(e_i)}(df(e_j)) - df({}^{\mathbf{h}}\nabla_{e_i} e_j)$$

in terms of local frame field.

A map f such that $\tau(f) \equiv 0$ is called **harmonic**. We thus see that under the above assumptions about \mathbf{h} and ω

Theorem 2 *Condition (8.22) says that f is harmonic relative to \mathbf{g} and $\check{\mathbf{h}}$.*

Suppose that we make the further assumption that $\check{\mathbf{h}}$ is the metric induced from \mathbf{g} by the map f . Then

$$df({}^{\mathbf{h}}\nabla_X Y) = (\mathbf{g}\nabla_{df(X)} df(Y))^{\text{tan}},$$

the tangential component of $\mathbf{g}\nabla_{df(X)} df(Y)$ and hence

$$B_f(X, Y) = (\mathbf{g}\nabla_{df(X)} df(Y))^{\text{nor}},$$

the normal component of $\mathbf{g}\nabla_{df(X)} df(Y)$. This is just the classical second fundamental form vector of Q regarded as an immersed submanifold of M . Taking its trace gives kH where H is the mean curvature vector of the immersion. Thus if in addition to the above assumptions we make the assumption that the metric $\check{\mathbf{h}}$ is induced by the map f , then we conclude that (8.18) says that $H = 0$, i.e. that the immersion f must be a minimal immersion.

Chapter 9

Submersions.

The treatment here is that of a 1966 paper by O’Neill (Michigan Journal of Math.) following earlier basic work by Hermann. In a sense, the subject can be regarded as the appropriate generalization of the notion of a “surface of revolution”

9.1 Submersions.

Let M and B be differentiable manifolds, and $\pi : M \rightarrow B$ be a submersion, which means that $d\pi_m : TM_m \rightarrow TB_{\pi(m)}$ is surjective for all $m \in M$. The implicit function theorem then guarantees that $\pi^{-1}(b)$ is a submanifold of M for all $b \in B$. These submanifolds are called the *fibers* of the submersion. By the implicit function theorem, the tangent space to the fiber through $m \in M$ is just the kernel of the differential of the projection, π . Call this space $V(M)_m$. So

$$V(M)_m := \ker d\pi_m.$$

The set of such tangent vectors at m is called the set of vertical vectors, and a vector field on M whose values at every point are vertical will be called a vertical vector field. We will denote the set of vertical vector fields by $\mathcal{V}(M)$.

If ϕ is a smooth function on B , and V is a vertical vector field, then $V\pi^*\phi = 0$. Conversely, if $V\pi^*\phi = 0$ for all smooth functions, ϕ on B , then V is vertical. In particular, if U and V are vertical vector fields, then so is $[U, V]$.

Now suppose that both M and B are (semi-)Riemann manifolds. Let

$$H(M)_m := V(M)_m^\perp.$$

We assume the following:

$$d\pi_m : H(M)_m \rightarrow TB_{\pi(m)}$$

is an isometric isomorphism, i.e. is bijective and preserves the scalar product of tangent vectors. Notice that this implies that $V(M)_m \cap H(M)_m = \{0\}$ so that

the restriction of the scalar product to $V(M)_m$ is non-singular. (Of course in the Riemannian case this is automatic.)

We let $\mathcal{H} : T(M)_m \rightarrow H(M)_m$ denote the orthogonal projection at each point and also let $\mathcal{H}(M)$ denote the set of “horizontal” vector fields (vector fields which belong to $H(M)_m$ at each point). Similarly, we let \mathcal{V} denote orthogonal projection onto $V(M)_m$ at each point. So if E is a vector field on M , then $\mathcal{V}E$ is a vertical vector field and $\mathcal{H}E$ is a horizontal vector field. We will reserve the letters U, V, W for vertical vector fields, and X, Y, Z for horizontal vector fields.

Among the horizontal vector fields, there is a subclass, the *basic* vector fields. They are defined as follows: Let X_B be a vector field on B . If $m \in M$, there is a unique tangent vector, call it $X(m) \in H(M)_m$ such that $d\pi_m X(m) = X_B(\pi(m))$. This defines the basic vector field, X , corresponding to X_B . Notice that if X is the basic vector field corresponding to X_B , and if ϕ is a smooth function on B , then

$$X\pi^*\phi = \pi^*(X_B\phi).$$

Also, by definition,

$$\langle X, Y \rangle_M = \pi^*\langle X_B, Y_B \rangle_B$$

for basic vector fields X and Y . In general, if X and Y are horizontal, or even basic vector fields, their Lie bracket, $[X, Y]$ need not be horizontal. But if X and Y are basic, then we can compute the horizontal component of $[X, Y]$ as follows: If ϕ is any function on B and if X and Y are basic vector fields, then

$$\begin{aligned} (\mathcal{H}[X, Y])\pi^*\phi &= [X, Y]\pi^*\phi \\ &= XY\pi^*\phi - YX\pi^*\phi \\ &= \pi^*(X_BY_B\phi - Y_BX_B\phi) \\ &= \pi^*([X_B, Y_B]\phi) \end{aligned}$$

so $\mathcal{H}[X, Y]$ is the basic vector field corresponding to $[X_B, Y_B]$.

We claim that

$$\mathcal{H}(\nabla_X Y) \text{ is the basic vector field corresponding to } \nabla_{X_B}^B(Y_B) \quad (9.1)$$

where ∇^B denotes the Levi-Civita covariant derivative on B and ∇ denotes the covariant derivative on M . Indeed, let X_B, Y_B, Z_B be vector fields on B and X, Y, Z the corresponding basic vector fields on M . Then

$$X\langle Y, Z \rangle_M = X(\pi^*\langle X_B, Y_B \rangle_B) = \pi^*(X_B\langle Y_B, Z_B \rangle_B)$$

while

$$\langle X, [Y, Z] \rangle = \langle X, \mathcal{H}[Y, Z] \rangle = \pi^*(\langle X_B, [Y_B, Z_B] \rangle_B)$$

since $\mathcal{H}[Y, Z]$ is the basic vector field corresponding to $[Y_B, Z_B]$. From the Koszul formula it then follows that

$$\langle \nabla_X Y, Z \rangle_M = \pi^*\langle \nabla_{X_B}^B Y_B, Z_B \rangle_B.$$

Therefore $d\pi_m(\nabla_X Y(m)) = \nabla_{X_B}^B Y_B(\pi(m))$ for all points m which implies (9.1).

Suppose that γ is a horizontal geodesic, so that $\pi\gamma$ is a regular curve, so an integral curve of a vector field X_B on B . Let X be the corresponding basic vector field, so γ is an integral curve of X . The fact that γ is a geodesic implies that $\nabla_X X = 0$ along γ , and hence by (9.1) $\nabla_{X_B} X_B = 0$ along $\pi\gamma$ so $\pi\gamma$ is a geodesic. We have proved

$$\pi(\gamma) \text{ is a geodesic if } \gamma \text{ is a horizontal geodesic.} \quad (9.2)$$

If V and W are vertical vector fields, then we may consider their restriction to each fiber as a vector field along that fiber, and may also consider the Levi-Civita connection on the fiber considered as a semi-Riemann manifold in its own right. We will denote the covariant derivative of W with respect to V relative to the connection induced by the metric on each fiber by $\nabla_V^\mathcal{V} W$. It follows from the Koszul formula, and the fact that $[V, W]$ is vertical if V and W are that

$$\nabla_V^\mathcal{V} W = \mathcal{V}(\nabla_V W) \quad (9.3)$$

for vertical vector fields. Here ∇ is the Levi-Civita covariant derivative on M , so that $\nabla_V W$ has both a horizontal and a vertical component.

9.2 The fundamental tensors of a submersion.

9.2.1 The tensor T .

For arbitrary vector fields E and F on M define

$$T_E F := \mathcal{H}[\nabla_{\mathcal{V}E}(\mathcal{V}F)] + \mathcal{V}[\nabla_{\mathcal{V}E}(\mathcal{H}F)],$$

where, in this equation, ∇ denotes the Levi-Civita covariant derivative determined by the metric on M .

If f is any differentiable function on M , then $\mathcal{V}fF = f\mathcal{V}F$ and $\nabla_{\mathcal{V}E}(f\mathcal{V}F) = [(\mathcal{V}E)f]\mathcal{V}F + f\nabla_{\mathcal{V}E}(\mathcal{V}F)$ so

$$\mathcal{H}[\nabla_{\mathcal{V}E}(\mathcal{V}(fF))] = f\mathcal{H}[\nabla_{\mathcal{V}E}(\mathcal{V}F)].$$

Similarly f pulls out of the second term in the definition of T . Also $\mathcal{V}(fE) = f\mathcal{V}E$ and $\nabla_{f\mathcal{V}E} = f\nabla_{\mathcal{V}E}$ by a defining property of ∇ .

This proves that T is a tensor of type $(1, 2)$: $T_{fE}F = T_E(fF) = fT_EF$.

By definition, $T_E = T_{\mathcal{V}E}$ depends only on the vertical component, $\mathcal{V}E$ of E . If U and V are vertical vector fields, then

$$\begin{aligned} T_U V &= \mathcal{H}\nabla_U V \\ &= \mathcal{H}\nabla_V U + \mathcal{H}([U, V]) \\ &= \mathcal{H}\nabla_V U \end{aligned}$$

since $[U, V]$ is vertical. Thus

$$T_U V = T_V U \quad (9.4)$$

for vertical vector fields. Also notice that if U is a vertical vector field then $T_E U$ is horizontal, while if X is a horizontal vector field, then $T_E X$ is vertical.

1. Show that

$$\langle T_E F_1, F_2 \rangle = -\langle F_1, T_E F_2 \rangle \quad (9.5)$$

for any pair of vector fields F_1, F_2 .

9.2.2 The tensor A .

This is defined by interchanging the role of horizontal and vertical in T , so

$$A_E F := \mathcal{V}\nabla_{\mathcal{H}E}(\mathcal{H}F) + \mathcal{H}\nabla_{\mathcal{H}E}(\mathcal{V}F).$$

The same proof as above shows that A is a tensor, that A_E sends horizontal vector fields into vertical vector fields and vice versa, and the your solution of problem 1 will also show that

$$\langle A_E F_1, F_2 \rangle = -\langle F_1, A_E F_2 \rangle$$

for any pair of vector fields F_1, F_2 .

Notice that the any horizontal vector field can be written (locally) as a function combination of basic vector fields, and if V is vertical and X basic, then

$$[V, X]\pi^*\phi = V\pi^*(X_B\phi) - XV\pi^*\phi = 0,$$

so the Lie bracket of a vertical vector field and a basic vector field is vertical.

2. Show that $A_X X = 0$ for any horizontal vector field, X , and hence that

$$A_X Y = -A_Y X \quad (9.6)$$

for any pair of horizontal vector fields X, Y . Since

$$\mathcal{V}[X, Y] = \mathcal{V}(\nabla_X Y - \nabla_Y X) = A_X Y - A_Y X$$

it then follows that

$$A_X Y = \frac{1}{2}\mathcal{V}[X, Y]. \quad (9.7)$$

(Hint, it suffices to show that $A_X X = 0$ for basic vector fields, and for this, that $\langle V, A_X X \rangle = 0$ for all vertical vector fields since $A_X X$ is vertical. Use Koszul's formula.)

We can express the relations between covariant derivatives of horizontal and

vertical vector fields and the tensors T and A :

$$\nabla_V W = T_V W + \nabla_V^\mathcal{V} W \quad (9.8)$$

$$\nabla_V X = \mathcal{H}\nabla_V X + T_V X \quad (9.9)$$

$$\nabla_X V = A_X V + \mathcal{V}\nabla_X V \quad (9.10)$$

$$\nabla_X Y = \mathcal{H}\nabla_X Y + A_X Y \quad (9.11)$$

If X is a basic vector field, then $\nabla_V X = \nabla_X V + [V, X]$ and $[V, X]$ is vertical. Hence

$$\mathcal{H}\nabla_V X = A_X V \text{ if } X \text{ is basic.} \quad (9.12)$$

9.2.3 Covariant derivatives of T and A .

The definition of covariant derivative of a tensor field gives

$$(\nabla_{E_1} A)_{E_2} E_3 = \nabla_{E_1} (A_{E_2} E_3) - A_{\nabla_{E_1} E_2} E_3 - A_{E_2} (\nabla_{E_1} E_3)$$

for any three vector fields E_1, E_2, E_3 . Suppose, in this equation, we take $E_1 = V$ and $E_2 = W$ to be vertical, and $E_3 = E$ to be a general vector field. Then $A_{E_2} = A_W = 0$ so the first and third terms on the right vanish. In the middle term we have

$$A_{\nabla_V W} = A_{\mathcal{H}\nabla_V W} = A_{T_V W}$$

so that we get

$$(\nabla_V A)_W = -A_{T_V W}. \quad (9.13)$$

If we take $E_1 = X$ to be horizontal and $E_2 = W$ to be vertical, again only the middle term survives and we get

$$(\nabla_X A)_W = -A_{A_X W}. \quad (9.14)$$

Similarly,

$$(\nabla_X T)_Y = -T_{A_X Y} \quad (9.15)$$

$$(\nabla_V T)_Y = -T_{T_V Y}. \quad (9.16)$$

3. Show that

$$\langle (\nabla_U A)_X V, W \rangle = \langle T_U V, A_X W \rangle - \langle T_U W, A_X V \rangle \quad (9.17)$$

$$\langle (\nabla_E A)_X Y, V \rangle = -\langle (\nabla_E A)_Y X, V \rangle \quad (9.18)$$

$$\langle (\nabla_E T)_V W, X \rangle = \langle (\nabla_E T)_W V, X \rangle \quad (9.19)$$

where U, V, W are vertical, X, Y are horizontal and E is a general vector field.

We also claim that

$$\mathcal{S}\langle (\nabla_Z A)_X Y, V \rangle = \mathcal{S}\langle A_X Y, T_V Z \rangle \quad (9.20)$$

where V is vertical, X, Y, Z horizontal and \mathcal{S} denotes cyclic sum over the horizontal vectors.

Proof. This is a tensor equation, so we may assume that X, Y, Z are basic and that the corresponding vector fields X_B, Y_B, Z_B have all their Lie brackets vanish at $b = \pi(m)$ where m is the point at which we want to check the equation. Thus all Lie brackets of X, Y, Z are vertical at m . We have $\frac{1}{2}[X, Y] = A_X Y$ by (9.7), so

$$\frac{1}{2}[[X, Y], Z] = [A_X Y, Z] = \nabla_{A_X Y}(Z) - \nabla_Z(A_X Y)$$

and the cyclic sum of the leftmost side vanishes by the Jacobi identity. So

$$\mathcal{S}[\nabla_{A_X Y}(Z)] = \mathcal{S}[\nabla_Z(A_X Y)]. \quad (9.21)$$

Taking scalar product with the vertical vector $V(m)$, we have (at the point m) by repeated use of (9.4) and (9.5)

$$\begin{aligned} \langle \nabla_{A_X Y}(Z), V \rangle &= \langle T_{A_X Y}(Z), V \rangle \\ &= -\langle Z, T_{A_X Y}(V) \rangle \\ &= -\langle Z, T_V(A_X Y) \rangle \\ &= \langle T_V Z, A_X Y \rangle \end{aligned}$$

We record this fact for later use as

$$\langle T_{A_X Y}(Z), V \rangle = \langle T_V Z, A_X Y \rangle. \quad (9.22)$$

Using (9.21) we obtain

$$\mathcal{S}\langle \nabla_Z(A_X Y), V \rangle = \mathcal{S}\langle T_V Z, A_X Y \rangle. \quad (9.23)$$

Now

$$\langle \nabla_Z(A_X Y), V \rangle - \langle (\nabla_Z A)_X Y, V \rangle = \langle A_{\nabla_Z X}(Y), V \rangle + \langle A_X(\nabla_Z Y), V \rangle$$

while

$$A_{\nabla_Z X}(Y) = -A_Y(\mathcal{H}\nabla_Z X) = -A_Y(\mathcal{H}\nabla_X Z)$$

using (9.6) for the first equations and the fact that $[X, Z]$ is vertical for the second equation. Taking scalar product with V gives

$$-\langle A_Y(\mathcal{H}\nabla_X Z), V \rangle = -\langle A_Y(\nabla_X Z), V \rangle$$

since $A_Y U$ is horizontal for any vertical vector, and hence

$$\langle A_Y(\mathcal{V}\nabla_X Z), V \rangle = 0.$$

We thus obtain

$$\langle \nabla_Z(A_X Y), V \rangle - \langle (\nabla_Z A)_X Y, V \rangle = \langle A_X(\nabla_Z Y), V \rangle - \langle A_Y(\nabla_X Z), V \rangle.$$

The cyclic sum of the right hand side vanishes. So, taking cyclic sum and applying (9.23) establishes (9.20).

9.2.4 The fundamental tensors for a warped product.

A very special case of a semi-Riemann submersion is that of a “warped product” following O’Neill’s terminology. Here $M = B \times F$ as a manifold, so π is just projection onto the first factor. We are given a positive function, f on B and metrics $\langle \cdot, \cdot \rangle_B$ and $\langle \cdot, \cdot \rangle_F$ on each factor. At each point $m = (b, q)$, $b \in B, q \in F$ we have the direct sum decomposition

$$TM_m = TB_b \oplus TF_q$$

as vector spaces, and the warped product metric is defined as the direct sum

$$\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_B \oplus f^2 \langle \cdot, \cdot \rangle_F.$$

O’Neill writes $M = B \times_f F$ for the warped product, the metrics on B and F being understood. The notion of warped product can itself be considered as a generalization of a surface of revolution, where B is a plane curve not meeting the axis of revolution, where f is the distance to the axis, and where $F = S^1$, the unit circle with its standard metric.

On a warped product, the basic vector fields are just the vector fields of B considered as vector fields of $B \times F$ in the obvious way, having no F component. In particular, the Lie bracket of two basic vector fields, X and Y on M is just the Lie bracket of the corresponding vector fields X_B and Y_B on B , considered as a vector field on M via the direct product. In particular, $[X, Y]$ has no vertical component, so $A_X Y = 0$. In fact, we can be more precise. For each fixed $q \in F$, the projection π restricted to $B \times \{q\}$ is an isometry of $B \times \{q\}$ with B . Thus

$$\nabla_X Y = \text{the basic vector field corresponding to } \nabla_{X_B}^B Y_B.$$

On a warped product, there is a special class of vertical vector fields, those that are vector fields on F considered as vector fields on $B \times F$ via the direct product decomposition. Let us denote the collection of these vector fields by $\mathcal{L}(F)$, the “lifts” of vector fields on F to use O’Neill’s terminology. If $V \in \mathcal{L}(F)$ and X is a basic vector field, then $[X, V] = 0$ since they “depend on different variables” and hence $\nabla_X V = \nabla_V X$. The vector field $\nabla_X V$ is vertical, since $\langle \nabla_X V, Y \rangle = -\langle V, \nabla_X Y \rangle = 0$ for any basic vector field, Y , as $\nabla_X Y$ is horizontal. This shows that $A_X V = 0$ as well, so $A = 0$. We claim that once again we can be more precise:

$$\nabla_X V = \nabla_V X = \frac{Xf}{f} V \quad \forall \text{ basic } X, \text{ and } \forall V \in \mathcal{L}(F). \quad (9.24)$$

Indeed, the only term that survives in the Koszul formula for $2\langle \nabla_X V, W \rangle$, $W \in \mathcal{L}(F)$ is $X\langle V, W \rangle$. We have

$$\langle V, W \rangle = f^2 \langle V_F, W_F \rangle_F$$

where we have written f instead of $\pi^* f$ by the usual abuse of language for a direct product. Now $\langle V_F, W_F \rangle_F$ is a function on F (pulled back to $B \times F$) and

so is annihilated by X . Hence

$$X\langle V, W \rangle = 2f(Xf)\langle V_F, W_F \rangle_F = \frac{2Xf}{f}\langle V, W \rangle,$$

proving (9.24). Notice that (9.24) gives us a piece of T , namely

$$T_V X = \frac{Xf}{f}V.$$

We can also derive the “horizontal” piece of T , namely

$$T_V W = -\frac{\langle V, W \rangle}{f}\text{grad } f. \quad (9.25)$$

Indeed

$$\begin{aligned} \langle \nabla_V W, X \rangle &= -\langle W, \nabla_V X \rangle \\ &= -\frac{Xf}{f}\langle V, W \rangle \quad \text{and} \\ Xf &= \langle \text{grad } f, X \rangle. \end{aligned}$$

In this formula, it doesn't matter whether we consider f as a function on M and compute its gradient there, or think of f as a function on B and compute its gradient relative to B and then take the horizontal lift. The answer is the same since f has no F dependence. Finally, the vertical component of $\nabla_V W$, $V, W \in \mathcal{L}(F)$ is just the same as the extension to M of $\nabla_{V_F}^F W_F$ since the metric on each fiber differs from that of F by a constant factor, which has no influence on the covariant derivative.

9.3 Curvature.

We want equations relating the curvature of the base and the curvature of the fibers to the curvature of M and the tensors T , A , and their covariant derivatives.

So we will be considering expressions of the form

$$\langle R_{E_1 E_2} E_3, E_4 \rangle$$

where R is the curvature of M and the E 's are either horizontal or vertical. We let $n = 0, 1, 2, 3$, or 4 denote the number of horizontal vectors, the remaining being vertical. This gives five cases. So we will get five equations for curvature. For example, $n = 0$ corresponds to all vectors vertical, so we are asking for the relation between the curvature of the fiber and the full curvature. Let $R^\mathcal{V}$ denote the curvature tensor of the fiber (as a semi-Riemann submanifold).

The case $\mathbf{n} = \mathbf{0}$ is the **Gauss equation** of each fiber:

$$\langle R_{UV} W, F \rangle =$$

$$\langle R_{UV}^\mathcal{V}W, F \rangle - \langle T_UW, T_VF \rangle + \langle T_VW, T_UF \rangle, \quad U, V, W, F \in \mathcal{V}(M). \quad (9.26)$$

We recall the proof (O'Neill 100). We may assume $[U, V] = 0$ so

$$R_{UV} = -\nabla_U \nabla_V + \nabla_V \nabla_U$$

and, using (9.3) and the definition of T , if we have

$$\begin{aligned} \langle \nabla_U \nabla_V W, F \rangle &= \langle \mathcal{V} \nabla_U \nabla_V^\mathcal{V} W, F \rangle + \langle \nabla_U (T_V W), F \rangle \\ &= \langle \nabla_U^\mathcal{V} \nabla_V^\mathcal{V} W, F \rangle + U \langle T_V W, F \rangle - \langle T_V W, \nabla_U F \rangle \\ &= \langle \nabla_U^\mathcal{V} \nabla_V^\mathcal{V} W, F \rangle - \langle T_V W, T_U F \rangle. \end{aligned}$$

Substituting the above expression for R_{UV} into $\langle R_{UV}W, F \rangle$ then proves (9.26).

The case $\mathbf{n} = 1$ is the **Codazzi equation** for each fiber: Let U, V, W be vertical vector fields and X a horizontal vector field. Then

$$\langle R_{UV}W, X \rangle = \langle (\nabla_V T)_U W, X \rangle - \langle (\nabla_U T)_V W, X \rangle \quad (9.27)$$

This is also in O'Neill, page 115. We recall the proof. We assume that $[U, V] = 0$ so $R_{UV} = -\nabla_U \nabla_V + \nabla_V \nabla_U$ as before. We have

$$\begin{aligned} \langle \nabla_U \nabla_V W, X \rangle &= \langle \nabla_U \nabla_V^\mathcal{V} W, X \rangle + \langle \nabla_U (T_V W), X \rangle \\ &= \langle T_U (\nabla_V^\mathcal{V} W), X \rangle + \langle \nabla_U (T_V W), X \rangle. \end{aligned}$$

We write

$$\nabla_U (T_V W) = (\nabla_U T)_V W + T_{\nabla_U V} W + T_V \nabla_U W$$

and

$$\langle T_V \nabla_U W, X \rangle = \langle T_V \nabla_U^\mathcal{V} W, X \rangle$$

so

$$\langle \nabla_U \nabla_V W, X \rangle = \langle (\nabla_U T)_V W, X \rangle + \langle T_U (\nabla_V^\mathcal{V} W), X \rangle + \langle T_V (\nabla_U W), X \rangle + \langle T_{\nabla_U V} W, X \rangle.$$

Interchanging U and V and subtracting, using $\nabla_U V = \nabla_V U$ proves (9.27).

We now turn to the opposite extreme, $n = 4$ and $n = 3$ but first some notation. We let $R^\mathcal{H}$ denote the horizontal lift of the curvature tensor of B : If $h_i \in H(M)_m$ with $v_i := d\pi_m h_i$ define $R_{h_1 h_2}^\mathcal{H} h_3$ to be the unique horizontal vector such that

$$d\pi_m (R_{h_1 h_2}^\mathcal{H} h_3) = R_{v_1 v_2}^B v_3.$$

The case $\mathbf{n} = 4$ is given by

$$\langle R_{XY}Z, H \rangle = \langle R_{XY}^{\mathcal{H}}Z, H \rangle - 2\langle A_X Y, A_Z H \rangle + \langle A_Y Z, A_X H \rangle + \langle A_Z X, A_Y H \rangle \quad (9.28)$$

for any four horizontal vector fields X, Y, Z, H . As usual, we may assume X, Y, Z are basic and all their brackets are vertical. We will massage each term on the right of

$$R_{XY}Z = \nabla_{[X, Y]}Z - \nabla_X \nabla_Y Z + \nabla_Y \nabla_X Z.$$

Since $[X, Y]$ is vertical, $[X, Y] = 2A_X Y$. So

$$\nabla_{[X, Y]}Z = 2\mathcal{H}\nabla_{A_X Y}Z + 2T_{A_X Y}(Z).$$

Since Z is basic we can apply (9.12) to the first term giving

$$\nabla_{[X, Y]}Z = 2A_Z(A_X Y) + 2T_{A_X Y}(Z).$$

Let us write

$$\nabla_Y Z = \mathcal{H}\nabla_Y Z + A_Y Z$$

and apply equation (9.1) which we write, by abuse of language as

$$\mathcal{H}\nabla_Y Z = \nabla_Y^B Z.$$

Then

$$\nabla_X \nabla_Y Z = \nabla_X^B \nabla_Y^B Z + A_X(\nabla_Y^B Z) + A_X A_Y Z + \mathcal{V}\nabla_X(A_Y Z).$$

Separating the horizontal and vertical components in the definition of R gives

$$\mathcal{H}R_{XY}Z = -[\nabla_X^B, \nabla_Y^B]Z + 2A_Z A_X Y - A_X A_Y Z + A_Y A_X Z \quad (9.29)$$

$$\begin{aligned} \mathcal{V}R_{XY}Z &= 2T_{A_X Y}(Z) - \mathcal{V}\nabla_X(A_Y Z) + \\ &\quad + \mathcal{V}\nabla_Y(A_X Z) - A_X(\nabla_Y^B Z) + A_Y(\nabla_X^B Z) \end{aligned} \quad (9.30)$$

As we have chosen X, Y such that $[X_B, Y_B] = 0$, the first term on the right of (9.29) is just $R_{XY}^{\mathcal{H}}Z$. Taking the scalar product of (9.29) with a horizontal vector field (and using the fact that A_E is skew adjoint relative to the metric and $A_X Z = -A_Z X$) proves (9.28).

If we take the scalar product of (9.30) with a vertical vector field, V we get an expression for $\langle R_{XY}Z, V \rangle$ (and we can drop the projections \mathcal{V}). Let us examine what happens when we take the scalar product of the various terms on the right of (9.30) with V . The first term gives

$$\langle T_{A_X Y}(Z), V \rangle = \langle T_V Z, A_X Y \rangle$$

by (9.22). The next two terms give

$$\begin{aligned} \langle \nabla_Y(A_X Z), V \rangle - \langle \nabla_X(A_Y Z), V \rangle &= \langle (\nabla_Y A)_X Z, V \rangle - \langle (\nabla_X A)_Y Z, V \rangle \\ &\quad + \langle A_X(\nabla_Y Z), V \rangle - \langle A_Y(\nabla_X Z), V \rangle \end{aligned}$$

since $\nabla_X Y - \nabla_Y X = [X, Y]$ is vertical by assumption. The last two terms cancel the terms obtained by taking the scalar product of the last two terms in (9.30) with V and we obtain

$$\langle R_{XY}Z, V \rangle = 2\langle A_X Y, T_V Z \rangle + \langle (\nabla_Y A)_X Z, V \rangle - \langle (\nabla_X A)_Y Z, V \rangle. \quad (9.31)$$

We can simplify this a bit using (9.18) and (9.20). Indeed, by (9.18) we can replace the second term on the right by $-\langle (\nabla_X A)_Y Z, V \rangle$ and then apply (9.20) to get, for $\mathbf{n} = \mathbf{3}$,

$$\langle R_{XY}Z, V \rangle = \langle (\nabla_Z A)_X Y, V \rangle + \langle A_X Y, T_Z V \rangle - \langle A_Y Z, T_V X \rangle - \langle A_Z X, T_V Y \rangle. \quad (9.32)$$

Finally we give an expression for the case $\mathbf{n} = \mathbf{2}$:

$$\langle R_{XV}Y, W \rangle = \langle (\nabla_X T)_V W, Y \rangle + \langle (\nabla_V A)_X Y, W \rangle - \langle T_V X, T_W Y \rangle + \langle A_X V, A_Y W \rangle. \quad (9.33)$$

To prove this, write

$$R_{XV} = \nabla_{\nabla_X V} - \nabla_{\nabla_V X} - \nabla_X \nabla_V + \nabla_V \nabla_X$$

and

$$\begin{aligned} \langle \nabla_{\nabla_X V} Y, W \rangle &= -\langle Y, T_{\nabla_X V} W \rangle + \langle A_{\nabla_X V} Y, W \rangle \\ -\langle \nabla_{\nabla_V X} Y, W \rangle &= -\langle T_{\nabla_V X} Y, W \rangle - \langle A_{\nabla_V X} Y, W \rangle \\ -\langle \nabla_X \nabla_V Y, W \rangle &= -\langle \nabla_X (T_V Y), W \rangle + \langle \nabla_V Y, A_X W \rangle \\ \langle \nabla_V \nabla_X Y, W \rangle &= \langle \nabla_V A_X Y, W \rangle - \langle \nabla_X Y, T_V W \rangle \end{aligned}$$

where, for example, in the last equation we have written $\nabla_X Y = A_X Y + \mathcal{H} \nabla_X Y$ and

$$\langle \nabla_V \mathcal{H} \nabla_X Y, W \rangle = -\langle \mathcal{H} \nabla_X Y, \nabla_V W \rangle = \langle \mathcal{H} \nabla_X Y, T_V W \rangle = \langle \nabla_X Y, T_V W \rangle.$$

We have

$$\begin{aligned} \langle (\nabla_X T)_V W, Y \rangle &= -\langle W, (\nabla_X T)_V Y \rangle \\ &= -\langle W, \nabla_X (T_V Y) \rangle + \langle W, T_{\nabla_X V} Y \rangle + \langle W, T_V \nabla_X Y \rangle \\ \langle (\nabla_V A)_X Y, W \rangle &= \langle \nabla_V (A_X Y), W \rangle - \langle A_{\nabla_V X} Y, W \rangle - \langle A_X \nabla_V Y, W \rangle. \end{aligned}$$

The six terms on the right of the last two equations equal six of the eight terms on the right of the preceding four leaving two remaining terms,

$$\langle A_{\nabla_X V} Y, W \rangle - \langle T_{\nabla_V X} Y, W \rangle.$$

But

$$\begin{aligned} \langle A_{\nabla_X V} Y, W \rangle &= -\langle A_Y \mathcal{H} \nabla_X V, W \rangle \\ &= -\langle A_Y A_X V, W \rangle \\ &= \langle A_X V, A_Y W \rangle \end{aligned}$$

and a similar argument deals with the second term.

We repeat our equations. In terms of increasing values on n we have

$$\begin{aligned} \langle R_{UV} W, F \rangle &= \langle R_{UV}^\vee W, F \rangle - \langle T_U W, T_V F \rangle + \langle T_V W, T_U F \rangle, \\ \langle R_{UV} W, X \rangle &= \langle (\nabla_V T)_U W, X \rangle - \langle (\nabla_U T)_V W, X \rangle \\ \langle R_{XV} Y, W \rangle &= \langle (\nabla_X T)_V W, Y \rangle + \langle (\nabla_V A)_X, W \rangle - \langle T_V X, T_W Y \rangle + \langle A_X V, A_Y W \rangle, \\ \langle R_{XY} Z, V \rangle &= \langle (\nabla_Z A)_X Y, V \rangle + \langle A_X Y, T_Z V \rangle - \langle A_Y Z, T_V X \rangle - \langle A_Z X, T_V Y \rangle, \\ \langle R_{XY} Z, H \rangle &= \langle R_{XY}^H Z, H \rangle - 2\langle A_X Y, A_Z H \rangle + \langle A_Y Z, A_X H \rangle + \langle A_Z X, A_Y H \rangle. \end{aligned}$$

We have stated the formula for $n = 2$, i.e. two vertical and two horizontal fields for the case $\langle R_{XV} Y, W \rangle$, i.e. where one horizontal and one vertical vector occur in the subscript $R_{E_1 E_2}$. But it is easy to check that all other arrangements of two horizontal and two vertical fields can be reduced to this one by curvature identities. Similarly for $n = 1$ and $n = 3$.

9.3.1 Curvature for warped products.

The curvature formulas simplify considerably in the case of a warped product where $A = 0$ and

$$T_V X = \frac{Xf}{f} V, \quad T_V W = -\frac{\langle V, W \rangle}{f} \text{grad} f.$$

We will give the formulas where X, Y, Z, H are basic and $U, V, W, F \in \mathcal{L}(F)$. We have $Vf = 0$ and $\langle \nabla_V \text{grad} f, X \rangle = VXf - \langle \text{grad} f, \nabla_V X \rangle = 0$. We conclude that the right hand side of (9.27) vanishes, so $R_{UV} W$ is vertical and we conclude from (9.26) that

$$R_{UV} W = R_{UV}^F W - \frac{\langle \text{grad} f, \text{grad} f \rangle}{f^2} (\langle U, W \rangle V - \langle V, W \rangle U) \quad (9.34)$$

The Hessian of a function f on a semi-Riemann manifold is defined to be the bilinear form on the tangent space at each point defined by

$$H^f(X, Y) = \langle \nabla_X \text{grad} f, Y \rangle.$$

In fact, we have

$$\begin{aligned}\langle \nabla_X \text{grad } f, Y \rangle &= XYf - \langle \text{grad } f, \nabla_X Y \rangle \\ &= \nabla_X(df(Y)) - df(\nabla_X Y) \\ &= [\nabla \nabla f](X, Y)\end{aligned}$$

which gives an alternative definition of the Hessian as

$$H^f = \nabla \nabla f$$

and shows that it is indeed a (0,2) type tensor field. Also

$$\begin{aligned}H^f(X, Y) &= XYf - (\nabla_X Y)f \\ &= [X, Y]f + YXf - [\nabla_X Y - \nabla_Y X + \nabla_Y X]f \\ &= YXf - (\nabla_Y X)f \\ &= H^f(Y, X)\end{aligned}$$

showing that H^f is a symmetric tensor field.

We have

$$\nabla_X V = \frac{Xf}{f}V = T_V X, \quad T_V W = -\frac{\langle V, W \rangle}{f} \text{grad } f,$$

if X is basic and $V, W \in \mathcal{L}(F)$. So

$$\begin{aligned}(\nabla_X T)_V W &= \nabla_X(T_V W) - T_{\nabla_X V} - T_V(\nabla_X W) \\ &= -\langle V, W \rangle \left(\frac{Xf}{f^2} \text{grad } f + \frac{1}{f} \nabla_X \text{grad } f \right) + 2\langle V, W \rangle \frac{Xf}{f^2}\end{aligned}$$

and $\langle \text{grad } f, Y \rangle = Yf$. Therefore the case $n = 2$ above yields

$$\langle R_{XV} Y, W \rangle = -\frac{H^f(X, Y)}{f} \langle V, W \rangle. \quad (9.35)$$

The case $n = 3$ gives

$$\langle R_{XY} Z, V \rangle = 0$$

and hence by a symmetry property of the curvature tensor, $\langle R_{XY} Z, V \rangle = \langle R_{ZV} X, Y \rangle = 0$, or, changing notation,

$$\langle R_{XV} Y, Z \rangle = 0.$$

Thus

$$R_{VX} Y = \frac{H^f(X, Y)}{f} V. \quad (9.36)$$

We have $\langle R_{UV} X, W \rangle = -\langle R_{UV} X, W \rangle = 0$ and by (9.36) and the first Bianchi identity $\langle R_{UV} X, Y \rangle = (H^f(X, Y)/f) \times (\langle U, V \rangle - \langle V, U \rangle) = 0$ so

$$R_{UV} X = 0. \quad (9.37)$$

If we use this fact, the symmetry of the curvature tensor and (9.35) we see that

$$R_{XV}W = \frac{\langle V, W \rangle}{f} \nabla_X \text{grad } f. \quad (9.38)$$

It follows from the case $n = 3$ and $n = 4$ that

$$R_{XY}Z = R_{XY}^{\mathcal{H}}Z, \quad (9.39)$$

the basic vector field corresponding to the vector field $R_{X_B Y_B}^B Z_B$. Hence $\langle R_{XY}V, Z \rangle = 0$. We also have $\langle R_{XY}V, W \rangle = \langle R_{VW}X, Y \rangle = 0$ so

$$R_{XY}V = 0. \quad (9.40)$$

Ricci curvature of a warped product.

Recall that the Ricci curvature, $\text{Ric}(X, Y)$ defined as the trace of the map $V \mapsto R_{XV}Y$ is given in terms of an “orthonormal” frame field E_1, \dots, E_n by

$$\text{Ric}(X, Y) = \sum \epsilon_i \langle R_{X E_i} Y, E_i \rangle, \quad \epsilon_i = \langle E_i, E_i \rangle.$$

We will apply this to a frame field whose first $\dim B$ vectors lie in $\text{Vect } B$ and whose last $d = \dim F$ vectors lie in $\text{Vect } F$. We will assume that $d > 1$ and that $X, Y \in \text{Vect } B$ and $U, V \in \text{Vect } F$. We get

$$\text{Ric}(X, Y) = \text{Ric}^B(X, Y) - \frac{d}{f} \text{Hess}^B(f)(X, Y) \quad (9.41)$$

$$\text{Ric}(X, V) = 0 \quad (9.42)$$

$$\text{Ric}(V, W) = \text{Ric}^F(V, W) - \langle V, W \rangle f^\# \quad \text{where} \quad (9.43)$$

$$f^\# := \frac{\Delta f}{f} + (d-1) \frac{\langle \text{grad } f, \text{grad } f \rangle}{f^2} \quad (9.44)$$

where Δf is the Laplacian of f which is the same as the contraction of the Hessian of f .

Geodesics for a warped product

We now compute the equations for a geodesic on $B \times_f F$. Let $\gamma(s) = (\alpha(s), \beta(s))$ be a curve on $B \times_f F$ and suppose temporarily the neither $\alpha'(s) = 0$ nor $\beta'(s) = 0$ in an interval we are studying. So we can embed the tangent vectors along both projected curves in vector fields, X on B and V on F , so that γ is a solution curve to $X + V$ on $B \times_f F$. The condition that γ be a geodesic is then that $\nabla_{X+V}(X + V) = 0$ along γ . But

$$\begin{aligned} \nabla_{X+V}(X + V) &= \nabla_X X + \nabla_X V + \nabla_V X + \nabla_V V \\ &= \nabla_X^B X + 2 \frac{Xf}{f} V - \frac{\langle V, V \rangle}{f} \text{grad } f + \nabla_V^F V. \end{aligned}$$

Separating the vertical and horizontal components, and using the fact that $\nabla_X^B X = \alpha''$ along α and $\beta' = V$, $\nabla_V^F V = \beta''$ along β shows that the geodesic equations take the form

$$\alpha'' = \langle \beta', \beta' \rangle_F (f \circ \alpha) \text{grad } f \text{ on } B \quad (9.45)$$

$$\beta'' = -\frac{2}{f \circ \alpha} \frac{d(f \circ \alpha)}{ds} \beta' \text{ on } F \quad (9.46)$$

A limiting argument [O-208] shows that these equations hold for all geodesics.

We repeat all the important equations of this subsection:

$$\nabla_X Y = \nabla_X^B Y$$

$$\nabla_X V = \frac{Xf}{f} V = \nabla_V X$$

$$\begin{aligned} \mathcal{H}\nabla_V W &= T_U V \\ &= -\frac{1}{f} \langle V, W \rangle \text{grad } f \end{aligned}$$

$$\begin{aligned} \text{vert } \nabla_V W &= \nabla_V^F W \end{aligned}$$

geodesic eqns

$$\begin{aligned} \alpha'' &= \langle \beta', \beta' \rangle_F (f \circ \alpha) \text{grad } f \text{ on } B \\ \beta'' &= -\frac{2}{f \circ \alpha} \frac{d(f \circ \alpha)}{ds} \beta' \text{ on } F \end{aligned}$$

curvature

$$\begin{aligned} R_{XY} Z &= R_{XY}^{\mathcal{H}} Z \\ R_{VX} Y &= \frac{\text{Hess}^B(f)(X, Y)}{f} V \\ R_{XV} W &= \frac{\langle V, W \rangle}{f} \nabla_X \text{grad } f \\ R_{UV} W &= R_{UV}^F W - \frac{\langle \text{grad } f, \text{grad } f \rangle}{f^2} (\langle U, W \rangle V - \langle V, W \rangle U) \end{aligned}$$

Ricci curv

$$\begin{aligned} \text{Ric}(X, Y) &= \text{Ric}^B(X, Y) - \frac{d}{f} \text{Hess}^B(f)(X, Y) \\ \text{Ric}(X, V) &= 0 \\ \text{Ric}(V, W) &= \text{Ric}^F(V, W) - \langle V, W \rangle f^\# \text{ where} \\ f^\# &:= \frac{\Delta f}{f} + (d-1) \frac{\langle \text{grad } f, \text{grad } f \rangle}{f^2}. \end{aligned}$$

9.3.2 Sectional curvature.

We return to the general case of a submersion, and recall that the sectional curvature of the plane, $P_{ab} \subset TM_m$, spanned by two independent vectors, $a, b \in TM_m$ is defined as

$$K(P_{ab}) := \frac{\langle R_{ab} a, b \rangle}{\langle a, a \rangle \langle b, b \rangle - \langle a, b \rangle^2}.$$

We can write the denominator more simply as $\|a \wedge b\|^2$.

In the following formulas, all pairs of vectors are assumed to be independent, with u, v vertical and x, y horizontal, and where x_B denotes $d\pi_m(x)$ and $y_B := d\pi_m(y)$. Substituting into our formulas for the curvature gives

$$K(P_{vw}) = K^{\mathcal{V}}(P_{vw}) - \frac{\langle T_v v, T_w w \rangle - \|T_v w\|^2}{\|v \wedge w\|^2} \quad (9.47)$$

$$K(P_{xv}) = \frac{\langle (\nabla_x T)_v v, x \rangle + \|A_x v\|^2 - \|T_v x\|^2}{\|x\|^2 \|v\|^2} \quad (9.48)$$

$$K(P_{xy}) = K^B(P_{x_B y_B}) - \frac{3\|A_x y\|^2}{\|x \wedge y\|^2}. \quad (9.49)$$

9.4 Reductive homogeneous spaces.

9.4.1 Bi-invariant metrics on a Lie group.

Let G be a Lie group with Lie algebra, \mathfrak{g} , which we identify with the left invariant vector fields on G . Any non-degenerate scalar product, $\langle \cdot, \cdot \rangle$, on \mathfrak{g} thus determines (and is equivalent to) a left invariant semi-Riemann metric on G . We let A_a denote conjugation by the element $a \in G$, so

$$A_a : G \rightarrow G, A_a(b) = aba^{-1}.$$

We have $A_a(e) = e$ and

$$d(A_a) = \text{Ad}_a : TG_e \rightarrow TG_e.$$

Since $A_a = L_a \circ R_{a^{-1}}$, the left invariant metric, $\langle \cdot, \cdot \rangle$ is right invariant if and only if it is A_a invariant for all $a \in G$, which is the same as saying that $\langle \cdot, \cdot \rangle$ is invariant under the adjoint representation of G on \mathfrak{g} , i.e. that

$$\langle \text{Ad}_a Y, \text{Ad}_a Z \rangle = \langle Y, Z \rangle, \quad \forall Y, Z \in \mathfrak{g}, a \in G.$$

Setting $a = \exp tX$, $X \in \mathfrak{g}$, differentiating with respect to t and setting $t = 0$ gives

$$\langle [X, Y], Z \rangle + \langle Y, [X, Z] \rangle = 0, \quad \forall X, Y, Z \in \mathfrak{g}. \quad (9.50)$$

If G is connected, this condition implies that $\langle \cdot, \cdot \rangle$ is invariant under Ad and hence is invariant under right and left multiplication. Such a metric is called *bi-invariant*.

Let inv denote the map sending every element into its inverse:

$$\text{inv} : a \mapsto a^{-1}, \quad a \in G.$$

Since $\text{inv} \exp tX = \exp(-tX)$ we see that

$$d \text{inv}_e = -\text{id}.$$

Also

$$\text{inv} = R_{a^{-1}} \circ \text{inv} \circ L_{a^{-1}}$$

since the right hand side sends $b \in G$ into

$$b \mapsto a^{-1}b \mapsto b^{-1}a \mapsto b^{-1}.$$

Hence $d \text{inv}_a : TG_a \rightarrow TG_{a^{-1}}$ is given, by the chain rule, as

$$dR_{a^{-1}} \circ d \text{inv}_e \circ dL_{a^{-1}} = -dR_{a^{-1}} \circ dL_{a^{-1}}$$

implying that a bi-invariant metric is invariant under the map inv . Conversely, if a left invariant metric is invariant under inv then it is also right invariant, hence bi-invariant since

$$R_a = \text{inv} \circ L_a^{-1} \circ \text{inv}.$$

The Koszul formula simplifies considerably when applied to left invariant vector fields and bi-invariant metrics since all scalar products are constant, so their derivatives vanish, and we are left with

$$2\langle \nabla_X Y, Z \rangle = -\langle X, [Y, Z] \rangle - \langle Y, [X, Z] \rangle + \langle Z, [X, Y] \rangle$$

and the first two terms cancel by (9.50). We are left with

$$\nabla_X Y = \frac{1}{2}[X, Y]. \quad (9.51)$$

Conversely, if $\langle \cdot, \cdot \rangle$ is a left invariant bracket for which (9.51) holds, then

$$\begin{aligned} \langle X, [Y, Z] \rangle &= 2\langle X, \nabla_Y Z \rangle \\ &= -2\langle \nabla_Y X, Z \rangle \\ &= -\langle [Y, X], Z \rangle \\ &= \langle [X, Y], Z \rangle \end{aligned}$$

so the metric is bi-invariant.

Let α be an integral curve of the left invariant vector field X . Condition (9.51) implies that $\alpha'' = \nabla_X X = 0$ so α is a geodesic. Thus the one-parameter groups are the geodesics through the identity, and all geodesics are left cosets of one parameter groups. (This is the reason for the name exponential map in Riemannian geometry.)

We compute the curvature of a bi-invariant metric by applying the definition to left invariant vector fields:

$$R_{XY}Z = \frac{1}{2}[[X, Y], Z] - \frac{1}{4}[X, [Y, Z]] + \frac{1}{4}[Y, [X, Z]].$$

Jacobi's identity implies the last two terms add up to $-\frac{1}{4}[[X, Y], Z]$ and so

$$R_{XY}Z = \frac{1}{4}[[X, Y], Z]. \quad (9.52)$$

In particular

$$\langle R_{XY}X, Y \rangle = \frac{1}{4} \langle [[X, Y], X], Y \rangle = \frac{1}{4} \langle [X, Y], [X, Y] \rangle$$

so

$$K(X, Y) = \frac{1}{4} \frac{\| [X, Y] \|^2}{\| X \wedge Y \|^2}. \quad (9.53)$$

For each $X \in \mathfrak{g}$ the linear transformation of \mathfrak{g} consisting of bracketing on the left by X is called $\text{ad } X$. So

$$\text{ad } X : \mathfrak{g} \rightarrow \mathfrak{g}, \quad \text{ad } X(V) := [X, V].$$

We can thus write our formula for the curvature as

$$R_{XV}Y = -\frac{1}{4}(\text{ad } Y)(\text{ad } X)V.$$

Now the Ricci curvature was defined as

$$\text{Ric}(X, Y) = \text{tr}[V \mapsto R_{XV}Y].$$

We thus see that for any bi-invariant metric, the Ricci curvature is always given by

$$\text{Ric} = -\frac{1}{4}B \quad (9.54)$$

where B , the *Killing form*, is defined by

$$B(X, Y) := \text{tr}(\text{ad } X)(\text{ad } Y). \quad (9.55)$$

The Killing form is symmetric, since $\text{tr}(AB) = \text{tr}BA$ for any pair of linear operators. It is also invariant. Indeed, let $\mu : \mathfrak{g} \rightarrow \mathfrak{g}$ be any automorphism of \mathfrak{g} , so $\mu([X, Y]) = [\mu(X), \mu(Y)]$ for all $X, Y \in \mathfrak{g}$. We can read this equation as saying

$$\text{ad}(\mu(X))(\mu(Y)) = \mu(\text{ad}(X)(Y))$$

or

$$\text{ad}(\mu(X)) = \mu \circ \text{ad } X \mu^{-1}.$$

Hence

$$\text{ad}(\mu(X))\text{ad}(\mu(Y)) = \mu \circ \text{ad } X \text{ad } Y \mu^{-1}.$$

Since trace is invariant under conjugation, it follows that

$$B(\mu(X), \mu(Y)) = B(X, Y).$$

Applied to $\mu = \exp(t\text{ad } Z)$ and differentiating at $t = 0$ shows that $B([Z, X], Y) + B(X, [Z, Y]) = 0$.

So the Killing form defines a bi-invariant scalar product on \mathfrak{g} . Of course it need not, in general, be non-degenerate. For example, if the group is commutative, it vanishes identically. A group G is called *semi-simple* if its Killing form

is non-degenerate. So on a semi-simple Lie group, we can always choose the Killing form as the bi-invariant metric. For such a choice, our formula above for the Ricci curvature then shows that the group manifold with this metric is *Einstein*, i.e. the Ricci curvature is a multiple of the scalar product.

Suppose that the adjoint representation of G on \mathfrak{g} is irreducible. Then \mathfrak{g} can not have two invariant non-degenerate scalar products unless one is a multiple of the other. In this case, we can also conclude from our formula that the group manifold is Einstein.

9.4.2 Homogeneous spaces.

Now suppose that $B = G/H$ where H is a subgroup with Lie algebra, \mathfrak{h} such that \mathfrak{h} has an H invariant complementary subspace, $\mathfrak{m} \subset \mathfrak{g}$. In fact, for simplicity, let us assume that \mathfrak{g} has a non-degenerate bi-invariant scalar product, whose restriction to \mathfrak{h} is non-degenerate, and let $\mathfrak{m} = \mathfrak{h}^\perp$. This defines a G invariant metric on B , and the projection $\rightarrow G/H = B$ is a submersion. The left invariant horizontal vector fields are exactly the vector fields $X \in \mathfrak{m}$, and so

$$A_X Y = \frac{1}{2} \mathcal{V}[X, Y], \quad X, Y \in \mathfrak{m}.$$

On the other hand, the fibers are cosets of H , hence totally geodesic since the geodesics are one parameter subgroups. Hence $T = 0$. We can read (9.49) backwards to determine $K_B(P_B)$ as

$$K_B(P_{X_B Y_B}) = K(P_{XY}) + \frac{3}{4} \frac{\|\mathcal{V}[X, Y]\|^2}{\|X \wedge Y\|^2}$$

or

$$K_B(P_{X_B Y_B}) = \frac{\frac{1}{4} \|\mathcal{H}[X, Y]\|^2 + \|\mathcal{V}[X, Y]\|^2}{\|X \wedge Y\|^2}, \quad X, Y \in \mathfrak{m}. \quad (9.56)$$

See O'Neill pp. 313-15 for a slightly more general formulation of this result.

It follows from (9.2) that the geodesics emanating from the point $H \in B = G/H$ are just the curves $(\exp tX)H$, $X \in \mathfrak{m}$.

9.4.3 Normal symmetric spaces.

Formula (9.56) simplifies if all brackets of basic vector fields are vertical. So we assume that $[\mathfrak{m}, \mathfrak{m}] \subset \mathfrak{h}$. Then we get

$$K_B(P_{X_B Y_B}) = \frac{\|[X, Y]\|^2}{\|X \wedge Y\|^2} = \frac{\langle [[X, Y], X], Y \rangle}{\|X \wedge Y\|^2}. \quad (9.57)$$

For examples where this holds, we need to search for a Lie group G whose Lie algebra \mathfrak{g} has an Ad-invariant non-degenerate scalar product, $\langle \cdot, \cdot \rangle$ and a

decomposition $g = h + m$ such that

$$\begin{aligned} h &\perp m \\ [h, h] &\subset h \\ [h, m] &\subset m \\ [m, m] &\subset h. \end{aligned}$$

Let $\theta : g \rightarrow g$ be the linear map determined by

$$\theta X = -X, \quad X \in m, \quad \theta U = U, \quad U \in h.$$

Then

- θ is an isometry of $\langle \cdot, \cdot \rangle$
- $\theta[E, F] = [\theta E, \theta F] \quad \forall E, F \in g$
- $\theta^2 = \text{id}$.

Conversely, suppose we start with a θ satisfying these conditions. Since $\theta^2 = \text{id}$, we can write g as the linear direct sum of the $+1$ and -1 eigenspaces of θ , i.e. define $h := \{U | \theta(U) = U\}$ and $m := \{X | \theta(X) = -X\}$. Since θ preserves the scalar product, eigenspaces corresponding to different eigenvalues must be orthogonal, and the bracket conditions on h and m follow automatically from their definition.

One way of finding such a θ is to find a diffeomorphism $\sigma : G \rightarrow G$ such that

- G has a bi-invariant metric which is also preserved by σ ,
- σ is an automorphism of G , i.e. $\sigma(ab) = \sigma(a)\sigma(b)$,
- $\sigma^2 = \text{id}$.

If we have such a σ , then $\theta := d\sigma_e$ satisfies our requirements. Furthermore, the set of fixed points of σ ,

$$F := \{a \in G | \sigma(a) = a\}$$

is clearly a subgroup, which we could take as our subgroup, H . In fact, let F_0 denote the connected component of the identity in F , and let H be any subgroup satisfying $F_0 \subset H \subset F$. Then $M = G/H$ satisfies all our requirements. Such a space is called a normal symmetric space. We construct a large collection of examples of such spaces in the next two subsections.

9.4.4 Orthogonal groups.

We begin by constructing an explicit model for the spaces $\mathbf{R}^{p,q}$ and the orthogonal groups $O(p, q)$. We let \bullet denote the standard Euclidean (positive definite) scalar product on \mathbf{R}^n . For any matrix, M , square or rectangular, we let ${}^t M$ denote its transpose. For a given choice of (p, q) with $p + q = n$ we let ϵ denote the

diagonal matrix with +1 in the first p positions and -1 in the last q positions. Then

$$\langle u, v \rangle := (\epsilon u) \bullet v = u \bullet (\epsilon v)$$

is a scalar product on \mathbf{R}^n of type (p, q) .

The condition that a matrix A belong to $O(p, q)$ is then that

$$\epsilon Av \bullet Aw = \epsilon v \bullet w, \quad \forall v, w \in \mathbf{R}^n$$

which is the same as

$$({}^t A \epsilon A v) \bullet w = (\epsilon v) \bullet w \quad \forall v, w \in \mathbf{R}^n$$

which is the as the condition

$${}^t A \epsilon A v = \epsilon v \quad \forall v \in \mathbf{R}^n.$$

So ${}^t A \epsilon A = \epsilon$ or

$$A \in O(p, q) \Leftrightarrow {}^t A = \epsilon A^{-1} \epsilon. \quad (9.58)$$

Now suppose that $A = \exp tM$, $M \in \mathfrak{g} := \mathfrak{o}(p, q)$. Then, since the exponential of the transpose of a matrix is the transpose of its exponential, we have

$$\exp s {}^t M = \epsilon \exp(-sM) \epsilon = \exp(-s \epsilon M \epsilon)$$

since $\epsilon^{-1} = \epsilon$. Differentiation at $s = 0$ gives

$${}^t M = -\epsilon M \epsilon \quad (9.59)$$

as the condition for a matrix to belong to the Lie algebra $\mathfrak{o}(p, q)$. If we write M in “block” form

$$M = \begin{pmatrix} a & x \\ y & b \end{pmatrix}$$

then

$${}^t M = \begin{pmatrix} {}^t a & {}^t y \\ {}^t x & {}^t b \end{pmatrix}$$

and the condition to belong to $\mathfrak{o}(p, q)$ is that

$${}^t a = -a, \quad {}^t b = -b, \quad y = {}^t x$$

so the most general matrix in $\mathfrak{o}(p, q)$ has the form

$$M = \begin{pmatrix} a & x \\ {}^t x & b \end{pmatrix}, \quad {}^t a = -a, \quad {}^t b = -b. \quad (9.60)$$

Consider the symmetric bilinear form $X, Y \mapsto \text{tr } XY$, called the “trace form”. It is clearly invariant under conjugation, hence, restricted to X, Y both belonging to $\mathfrak{o}(p, q)$, it is an invariant bilinear form. Let us show that is non-degenerate. Indeed, suppose that

$$X = \begin{pmatrix} a & x \\ {}^t x & b \end{pmatrix}, \quad Y = \begin{pmatrix} c & y \\ {}^t y & d \end{pmatrix}$$

are elements of $o(p, q)$. Then

$$\operatorname{tr} XY = \operatorname{tr} (ac + bd + 2x^t y).$$

this shows that the subalgebra $h := o(p) \oplus o(q)$ consisting of all “block diagonal” matrices is orthogonal to the subspace m consisting of all matrices with zero entries on the diagonal, i.e. of the form

$$M = \begin{pmatrix} 0 & x \\ {}^t x & 0 \end{pmatrix}.$$

For matrices of the latter form, we have

$$\operatorname{tr} x^t x = \sum_{ij} x_{ij}^2$$

and so is positive definite. On the other hand, since ${}^t a = -a$ and ${}^t b = -b$ we have $a \mapsto \operatorname{tr} a^2 = -\sum_{ij} a_{ij}^2$ is negative definite, and similarly for b . Hence the restriction of the trace form to h is negative definite.

9.4.5 Dual Grassmannians.

Suppose we consider the space \mathbf{R}^{p+q} , the positive definite Euclidean space, with orthogonal group $O(p+q)$. Its Lie algebra consists of all anti-symmetric matrices of size $p+q$ and the restriction of the trace form to $o(p+q)$ is negative definite. So we can choose a positive definite invariant scalar product on $g = o(p+q)$ by setting

$$\langle X, Y \rangle := -\frac{1}{2} \operatorname{tr} XY.$$

Let ϵ be as in the preceding subsection, so ϵ is diagonal with p plus 1's and q minus 1's on the diagonal. Notice that ϵ is itself an orthogonal transformation (for the positive definite scalar product on \mathbf{R}^{p+q}), and hence conjugation by ϵ is an automorphism of $O(p+q)$ and also of $SO(p+q)$ the subgroup consisting of orthogonal matrices with determinant one.

Let us take $G = SO(p+q)$ and σ to be conjugation by ϵ . So

$$\sigma \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix}$$

and hence the fixed point subgroup is $F = S(O(p) \times O(q))$. We will take $H = SO(p) \times S(q)$. The subspace m consists of all matrices of the form

$$X = \begin{pmatrix} 0 & -{}^t x \\ x & 0 \end{pmatrix}$$

and

$$\operatorname{tr} X^2 = -2\operatorname{tr} {}^t x x$$

so

$$\langle X, X \rangle = \text{tr } {}^t x x,$$

which was the reason for the $\frac{1}{2}$ in our definition of $\langle \cdot, \cdot \rangle$.

Our formula for the sectional curvature of a normal symmetric space shows that the sectional curvature of

$$\tilde{G}_{p,q}$$

is non-negative. The special case $p = 1$ the quotient space is the q - dimensional sphere,

$$\tilde{G}_{1,q} = S^q$$

and the x occurring in the above formula is a column vector. Hence $[X, Y]$ where Y corresponds to the column vector y is the operator $= y \otimes {}^t x - x \otimes {}^t y \in \mathfrak{o}(q)$, and

$$\|[X, Y]\|^2 = \|X \wedge Y\|^2,$$

proving that the unit sphere has constant curvature $+1$.

Next let G be the connected component of $O(p, q)$, as described in the preceding subsection, and again take σ to be conjugation by ϵ . This time take

$$\langle X, Y \rangle = \frac{1}{2} \text{tr } XY.$$

The -1 eigenspace, m of σ consists of all matrices of the form

$$\begin{pmatrix} 0 & {}^t x \\ x & 0 \end{pmatrix}$$

and the restriction of $\langle \cdot, \cdot \rangle$ to m is positive definite, while the restriction to $H := SO(p) \times SO(q)$ is negative definite. The corresponding symmetric space G/H is denoted by G_{pq}^* . It has negative sectional curvature. In particular, the case $p = 1$ is hyperbolic space, and the same computation as above shows that it has constant sectional curvature equal to -1 . This realizes hyperbolic space as the space of timelike lines through the origin in a Lorentz space of one higher dimension.

These two classes of symmetric spaces are dual in the following sense: Suppose that (h, m) and (h^*, m^*) are the Lie algebra data of symmetric spaces G/H and G^*/H^* . Suppose we have

- a Lie algebra isomorphism $\ell : h \rightarrow h^*$ such that $\langle \ell U, \ell V \rangle^* = -\langle U, V \rangle$, $\forall U, V \in h$ and
- a linear isometry $i : m \rightarrow m^*$ which reverses the bracket:

$$[iX, iY]^* = -[X, Y] \quad \forall X, Y \in m.$$

Then it is immediate from our formula for the sectional curvature that

$$K^*(iX, iY) = -K(X, Y)$$

for any $X, Y \in m$ spanning a non-degenerate plane. We say that the symmetric spaces G/H and G^*/H^* are in duality.

In our case, $H = SO(p) \times SO(q)$ for both $\tilde{G}_{p,q}$ and $G_{p,q}^*$ so we take $\ell = \text{id}$. We define i by

$$i : \begin{pmatrix} 0 & -{}^t x \\ x & 0 \end{pmatrix} \mapsto \begin{pmatrix} 0 & {}^t x \\ x & 0 \end{pmatrix}.$$

It is easy to check that these satisfy our axioms and so $\tilde{G}_{p,q}$ and $G_{p,q}^*$ are dual. For example, the sphere and hyperbolic space are dual in this sense.

9.5 Schwarzschild as a warped product.

In the Schwarzschild model we define

$$P := \{(t, r) | r > 0, r \neq 2M\}$$

with metric

$$-h dt^2 + \frac{1}{h} dr^2, \quad h = h(r) = 1 - \frac{2M}{r}.$$

Then construct the warped product

$$P \times_r S^2$$

where S^2 is the ordinary unit sphere with its standard positive definite metric, call it $d\sigma^2$. So, following O'Neill's conventions, the total metric is of type $(3, 1)$ (timelike = negative square length) given by

$$-h dt^2 + \frac{1}{h} dr^2 + r^2 d\sigma^2.$$

We write $P = P_I \cup P_{II}$ where

$$P_I = \{(t, r) | r > 2M\}, \quad P_{II} = \{(t, r) | r < 2M\}$$

and

$$N = P_I \times_r S^2, \quad B = P_{II} \times_r S^2.$$

N is called the Schwarzschild *exterior* and B is called the *black hole*. In the exterior, ∂_t is timelike. In B , ∂_t is spacelike and ∂_r is timelike.

In either, the vector fields ∂_t, ∂_r are orthogonal and basic. So the base is a surface with orthogonal coordinates. To apply the formulas for warped products we need some preliminary computations on connections and curvature of surfaces with orthogonal coordinates.

9.5.1 Surfaces with orthogonal coordinates.

We consider a surface with coordinates u, v and metric

$$Edu^2 + Gdv^2$$

and set

$$\epsilon_1 := \operatorname{sgn} E, \quad \epsilon_2 := \operatorname{sgn} G$$

and write

$$Edu^2 + Gdv^2 = \epsilon_1(\theta^1)^2 + \epsilon_2(\theta^2)^2$$

where

$$\theta^1 := edu, \quad e := \sqrt{\epsilon_1 E}, \quad e > 0,$$

and

$$\theta^2 := gdv, \quad g := \sqrt{\epsilon_2 G}, \quad g > 0.$$

The dual orthonormal frame field is given by

$$F_1 = \frac{1}{e}\partial_u, \quad F_2 = \frac{1}{g}\partial_v.$$

The connection forms, ω_2^1 and $\omega_1^2 = -\epsilon_1\epsilon_2\omega_2^1$ are defined by

$$\omega_2^1(X) = \theta^1(\nabla_X F_2), \quad \omega_1^2(X) = \theta^2(\nabla_X F_1)$$

for any vector field, X , and are determined by the Cartan equations

$$d\theta^1 + \omega_2^1 \wedge \theta^2 = 0, \quad d\theta^2 + \omega_1^2 \wedge \theta^1 = 0.$$

The curvature form is then given by $d\omega_2^1$. We find the connection forms by straightforward computation:

$$d\theta^1 = e_v dv \wedge du = -\frac{e_v}{g} du \wedge \theta^2$$

$$d\theta^2 = g_u du \wedge dv = -\frac{g_u}{e} dv \wedge \theta^1$$

where subscripts denote partial derivatives. Thus

$$\omega_2^1 = \frac{e_v}{g} du - \epsilon_1\epsilon_2 \frac{g_u}{e} dv$$

satisfies both structure equations (with $\omega_1^2 = -\epsilon_1\epsilon_2\omega_2^1$) and is uniquely determined by them. We compute

$$\begin{aligned} d\omega_2^1 &= \left(\frac{e_v}{g}\right)_v dv \wedge du - \epsilon_1\epsilon_2 \left(\frac{g_u}{e}\right)_u du \wedge dv \\ &= -\left[\left(\frac{e_v}{g}\right)_v + \epsilon_1\epsilon_2 \left(\frac{g_u}{e}\right)_u\right] du \wedge dv. \end{aligned}$$

This is the curvature form Ω_2^1 . In general, the Riemann curvature is related to the connection form by

$$R_{vw}(F_j) = - \sum \Omega_j^i(v, w)F_i.$$

In our case there is only one term in the sum and the sectional curvature, which equals the Gauss curvature is given by

$$\begin{aligned} \langle R_{F_1 F_2} F_1, F_2 \rangle &= -\langle R_{F_1 F_2} F_2, F_1 \rangle \\ &= \langle \Omega_2^1(F_1, F_2) F_1, F_1 \rangle \\ &= \epsilon_1 \Omega_2^1(F_1, F_2) \\ &= \frac{\epsilon_1}{eg} \Omega_2^1(\partial_u, \partial_v) \\ &= -\frac{\epsilon_1}{eg} \left[\left(\frac{e_v}{g} \right)_v + \epsilon_1 \epsilon_2 \left(\frac{gu}{e} \right)_u \right] \\ &= -\frac{1}{eg} \left[\epsilon_1 \left(\frac{e_v}{g} \right)_v + \epsilon_2 \left(\frac{gu}{e} \right)_u \right]. \end{aligned}$$

So

$$K = -\frac{1}{eg} \left[\epsilon_1 \left(\frac{e_v}{g} \right)_v + \epsilon_2 \left(\frac{gu}{e} \right)_u \right] \quad (9.61)$$

is the formula for the curvature of a surface in terms of orthogonal coordinates.

9.5.2 The Schwarzschild plane.

In the case of the Schwarzschild plane, P , we have $eg = 1$ so $e_r/g = ee_r = \epsilon_1 \frac{1}{2} E_r$, and the partial derivatives with respect to t vanish. The formula simplifies to $K = \frac{1}{2} E_{rr}$ or

$$K = \frac{2M}{r^3}. \quad (9.62)$$

The connection form in the Schwarzschild plane is given by

$$\omega_2^1 = \frac{M}{r^2} dt, \quad \omega_1^2 = \frac{M}{r^2} dt$$

by the same computation since $\epsilon_1 \epsilon_2 = -1$. So

$$\begin{aligned} \nabla_{\partial_t} \partial_t &= \nabla_{\partial_t} (h^{\frac{1}{2}} F_1) \\ &= h^{\frac{1}{2}} \nabla_{\partial_t} F_1 \\ &= h^{\frac{1}{2}} \omega_1^2(\partial_t) F_2 \\ &= h^{\frac{1}{2}} \frac{M}{r^2} F_2 \\ &= \frac{Mh}{r^2} \partial_r. \end{aligned}$$

Similarly,

$$\begin{aligned}\nabla_{\partial_r}\partial_t &= \partial_r(h^{\frac{1}{2}})F_1 \\ &= h^{-\frac{1}{2}}\frac{M}{r^2}F_1 \\ &= \frac{M}{r^2h}\partial_t,\end{aligned}$$

and

$$\begin{aligned}\nabla_{\partial_r}\partial_r &= \partial_r(h^{-\frac{1}{2}})F_2 \\ &= -\frac{M}{r^2h}\partial_r.\end{aligned}$$

We will also need the Hessian of the function r . We have, by definition,

$$H^r(X, Y) = \langle \nabla_X(\text{grad } r), Y \rangle.$$

Now

$$\text{grad } r = h\partial_r$$

and

$$\begin{aligned}\nabla_{\partial_t}h\partial_r &= h\nabla_{\partial_t}\partial_r \\ &= \frac{M}{r^2}\partial_t \\ \nabla_{\partial_r}h\partial_r &= h_r\partial_r + h\nabla_{\partial_r}\partial_r \\ &= \frac{M}{r^2}\partial_r.\end{aligned}$$

Thus

$$H^r = \frac{M}{r^2}\langle \cdot, \cdot \rangle. \quad (9.63)$$

9.5.3 Covariant derivatives.

We wish to apply the formulas for covariant derivatives in warped products to the basic vector fields, ∂_t, ∂_r , and to vector fields V, W tangent to the sphere (considered as vector fields on $N \cup B$, the warped product).

The covariant derivatives of basic vector fields are the lifts of the corresponding vector fields on the base, and so from the previous subsection we get

$$\nabla_{\partial_t}\partial_t = \frac{Mh}{r^2}\partial_r \quad (9.64)$$

$$\begin{aligned}\nabla_{\partial_t}\partial_r &= \nabla_{\partial_r}\partial_t \\ &= \frac{M}{r^2h}\partial_t\end{aligned} \quad (9.65)$$

$$\nabla_{\partial_r}\partial_r = -\frac{M}{r^2h}\partial_r \quad (9.66)$$

From the formula

$$\nabla_X V = \nabla_V X = \frac{Xf}{f} V$$

for a warped product we get, taking $f = r$,

$$\nabla_{\partial_t} V = \nabla_V \partial_t = 0, \quad (9.67)$$

and

$$\nabla_{\partial_r} V = \nabla_V \partial_r = \frac{1}{r} V. \quad (9.68)$$

Applying the formula for $T_V W$ for a warped product gives

$$T_V W = -\frac{h}{r} \langle V, W \rangle \partial_r \quad (9.69)$$

since $\text{grad } r = h \partial_r$. This is the horizontal component of $\nabla_V W$. The vertical component is just the lift of $\nabla_V^S W$, the covariant derivative on the sphere.

9.5.4 Schwarzschild curvature.

From formulas (9.39) and (9.40) for warped products and our formula (9.62) for the curvature in the Schwarzschild plane we get

$$R_{\partial_t \partial_r}(\partial_t) = (-2Mh/r^3) \partial_r \quad (9.70)$$

$$R_{\partial_r \partial_t}(\partial_r) = (2M/r^3 h) \partial_t \quad (9.71)$$

$$R_{\partial_t \partial_r} V = 0. \quad (9.72)$$

From (9.36) and (9.63) we obtain

$$R_{XV} Y = -R_{VX} Y = -\frac{M}{r^3} \langle X, Y \rangle V$$

so

$$R_{\partial_t V}(\partial_t) = (Mh/r^3) V \quad (9.73)$$

$$R_{\partial_t V}(\partial_r) = 0 \quad (9.74)$$

$$R_{\partial_r V}(\partial_t) = 0 \quad (9.75)$$

$$R_{\partial_r V}(\partial_r) = (M/hr^3) V. \quad (9.76)$$

We apply (9.34) to compute R_{UV} . We have $\langle \text{grad } h, \text{grad } h \rangle = h^2 \langle \partial_r, \partial_r \rangle = h$ and the fiber over (t, r) is the sphere of radius r whose curvature is r^{-2} . we get

$$R_{VW} U = (2M/r^3) (\langle U, V \rangle W - \langle U, W \rangle V) \quad (9.77)$$

$$R_{VW}(\partial_t) = 0 \quad (9.78)$$

$$R_{VW}(\partial_r) = 0 \quad (9.79)$$

To apply (9.38) we compute

$$\begin{aligned}\nabla_{\partial_t} \text{grad } r &= h \nabla_{\partial_t} (\partial_r) \\ &= \frac{M}{r^2} \partial_t \\ \nabla_{\partial_r} (h \partial_r) &= (2M/r^2) \partial_r + h \nabla_{\partial_r} (\partial_r) \\ &= \frac{M}{r^2} \partial_r\end{aligned}$$

so

$$R_{\partial_t V}(W) = R_{\partial_t W}(V) = (M/r^3) \langle V, W \rangle \partial_t \quad (9.80)$$

$$R_{\partial_r V}(W) = R_{\partial_r W}(V) = (M/r^3) \langle V, W \rangle \partial_r. \quad (9.81)$$

We show that the Ricci curvature vanishes by applying our formulas for the Ricci curvature of a warped product, (9.41)-(9.43). For a surface, $\text{Ric}(X, Y) = K \langle X, Y \rangle$ and this is $(2M/r^3) \langle X, Y \rangle$ for vectors in the Schwarzschild plane. On the other hand, $d = 2, f = r, H^f = (M/r^2) \langle \cdot, \cdot \rangle$. This shows that $\text{Ric}(X, Y) = 0$.

For vertical vectors, we have

$$\text{Ric}^F(V, W) = r^{-2} \langle V, W \rangle$$

while

$$\begin{aligned}\Delta r &= C(\text{Hess}^r) \\ &= 2M/r^2 \\ \langle \text{grad } f, \text{grad } f \rangle &= h \text{ so} \\ f^\# &= r^2\end{aligned}$$

showing that $\text{Ric}(V, W) = 0$.

9.5.5 Cartan computation.

We have used the techniques of warped product to compute the Schwarzschild connection and curvature. However, the Cartan method is more direct:

$$ds^2 = (\theta^0)^2 - (\theta^1)^2 - (\theta^2)^2 - (\theta^3)^2 \quad \text{where}$$

$$\theta^0 = \sqrt{h}dt, \quad h := 1 - \frac{2M}{r}$$

$$\theta^1 = \frac{1}{\sqrt{h}}dr$$

$$\theta^2 = rd\vartheta$$

$$\theta^3 = rSd\phi$$

$$S = \sin \vartheta$$

$$C = \cos \vartheta$$

$$d\sqrt{h} = \frac{1}{2\sqrt{h}} \left(\frac{2M}{r^2} \right) dr$$

so

$$d\theta^0 = \frac{M}{r^2\sqrt{h}}dr \wedge dt = -\frac{M}{r^2\sqrt{h}}\theta^0 \wedge \theta^1$$

$$d\theta^1 = 0$$

$$d\theta^2 = -\frac{\sqrt{h}}{r}\theta^2 \wedge \theta^1$$

$$d\theta^3 = -\frac{\sqrt{h}}{r}\theta^3 \wedge \theta^1 - \frac{C}{rS}\theta^3 \wedge \theta^2$$

or

$$d\theta = -\omega \wedge \theta$$

where

$$\theta = \begin{pmatrix} \theta^0 \\ \theta^1 \\ \theta^2 \\ \theta^3 \end{pmatrix} \quad \omega = \begin{pmatrix} 0 & \frac{M}{r^2\sqrt{h}}\theta^0 & 0 & 0 \\ \frac{M}{r^2\sqrt{h}}\theta^0 & 0 & -\frac{\sqrt{h}}{r}\theta^2 & -\frac{\sqrt{h}}{r}\theta^3 \\ 0 & \frac{\sqrt{h}}{r}\theta^2 & 0 & -\frac{C}{Sr}\theta^3 \\ 0 & \frac{\sqrt{h}}{r}\theta^3 & \frac{C}{Sr}\theta^3 & 0 \end{pmatrix}.$$

Now

$$\begin{aligned}
\frac{M}{r^2\sqrt{h}}\theta^0 &= \frac{M}{r^2}dt \text{ so} \\
d\left(\frac{M}{r^2\sqrt{h}}\theta^0\right) &= \frac{2M}{r^3}\theta^0 \wedge \theta^1 \\
\frac{\sqrt{h}}{r}\theta^2 &= \sqrt{h}d\vartheta \text{ so} \\
d\left(\frac{\sqrt{h}}{r}\theta^2\right) &= \frac{M}{r^2\sqrt{h}}dr \wedge d\vartheta = \frac{M}{r^3}\theta^1 \wedge \theta^2 \\
d\left(\frac{\sqrt{h}}{r}\theta^3\right) &= d\left(\sqrt{h} Sd\phi\right) \\
&= \frac{M}{r^2\sqrt{h}}dr \wedge Sd\phi + \sqrt{h} Cd\theta \wedge d\phi \\
&= \frac{M}{r^3}\theta^1 \wedge \theta^3 + \frac{C\sqrt{h}}{Sr^2}\theta^2 \wedge \theta^3 \\
d\left(\frac{C}{Sr}\theta^3\right) &= d(Cd\phi) = -Sd\theta \wedge d\phi \\
&= -\frac{1}{r^2}\theta^2 \wedge \theta^3.
\end{aligned}$$

This then gives the curvature matrix in this frame as

$$\Omega := d\omega + \theta \wedge \omega = \begin{pmatrix} 0 & \frac{2M}{r^3}\theta^0 \wedge \theta^1 & -\frac{M}{r^3}\theta^0 \wedge \theta^2 & -\frac{M}{r^3}\theta^0 \wedge \theta^3 \\ \frac{2M}{r^3}\theta^0 \wedge \theta^1 & 0 & -\frac{M}{r^3}\theta^1 \wedge \theta^2 & \frac{M}{r^3}\theta^3 \wedge \theta^1 \\ -\frac{M}{r^3}\theta^0 \wedge \theta^2 & \frac{M}{r^3}\theta^1 \wedge \theta^2 & 0 & \frac{2M}{r^3}\theta^2 \wedge \theta^3 \\ -\frac{M}{r^3}\theta^0 \wedge \theta^3 & -\frac{M}{r^3}\theta^3 \wedge \theta^1 & -\frac{2M}{r^3}\theta^2 \wedge \theta^3 & 0 \end{pmatrix}.$$

The curvature tensor is given in terms of Ω as

$$R_{vw}(E_j) = \sum \Omega_j^i(v, w)E_i$$

or

$$R_{jk\ell}^i = \Omega_j^i(E_k, E_\ell).$$

Notice from the form of Ω given above, that $R_{jk\ell}^i = 0$ if $j \neq k, \ell$. Hence $R_{j\ell m}^m = 0$ if $j \neq \ell$. Looking at the columns we see that $\sum R_{imi}^m = \pm(2I - I - I) = 0$. Thus the Schwarzschild metric is Ricci flat.

9.5.6 Petrov type.

The tensor R_{cd}^{ab} is obtained from the tensor $R^abcd = \Omega_b^a(E_c, E_d)$ by “raising ” the second index. We want to consider this as the matrix of the operator $[R]$

relative to the basis $E_i \wedge E_j$. If we use ij to stand for $E_i \wedge E_j$ and omit the zero entries we see that the matrix of $[R]$ is

$$[R] = \begin{array}{cccccc} & 01 & 02 & 03 & 23 & 31 & 12 \\ 01 & 2M/r^3 & & & & & \\ 02 & & -M/r^3 & & & & \\ 03 & & & -M/r^3 & & & \\ 23 & & & & 2M/r^3 & & \\ 31 & & & & & -M/r^3 & \\ 12 & & & & & & -M/r^3 \end{array} \quad (9.82)$$

We can write this in block three by three form as

$$[R] = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}$$

where

$$A = \begin{pmatrix} 2M/r^3 & 0 & 0 \\ 0 & -M/r^3 & 0 \\ 0 & 0 & -M/r^3 \end{pmatrix},.$$

On the other hand we have

$$\begin{aligned} \star(E_0 \wedge E_1) &= E_2 \wedge E_3 \\ \star(E_0 \wedge E_2) &= E_3 \wedge E_1 \\ \star(E_0 \wedge E_3) &= E_1 \wedge E_2 \quad \text{and} \\ \star^2 &= -\text{id}. \end{aligned}$$

Thus the matrix of \star relative to the same basis is

$$[\star] = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

where I is the three by three identity matrix. Clearly the operator given by R on $\wedge^2 T(M)$ commutes with the star operator, as predicted by the general theory for any Ricci flat curvature, and we see from the form of the matrix A that it is of Petrov type D , with real eigenvalues $2M/r^3, -M/r^3 - M/r^3$.

9.5.7 Kerr-Schild form.

We will show that by making a change of variables that the metric is the sum of a flat metric, and a multiple of the square of a linear differential form, α , where $|\alpha|^2 = 0$ in the flat metric. The generalization of this construction will be important in the case of rotating black holes. We make the change of variables in two stages: Let

$$u = t + T(r)$$

where T is any function of r (determined up to additive constant) such that

$$T' = \frac{1}{h}.$$

Then

$$du = dt + \frac{1}{h} dr$$

so

$$\begin{aligned} hdt^2 &= hdu^2 - 2dudr + \frac{1}{h} dr^2 \\ -hdt^2 + \frac{1}{h} dr^2 &= -hdu^2 + 2dudr \\ &= -(du - dr)^2 + dr^2 + \frac{2M}{r} du^2. \end{aligned}$$

So if we set

$$x^0 := u - r$$

this becomes

$$-d(x^0)^2 + dr^2 - \frac{2M}{r} [dx^0 + dr]^2.$$

The form $dx^0 + dr$ has square length zero in the flat metric

$$-d(x^0)^2 + dx^2 + y^2 + dz^2, \quad r^2 = x^2 + y^2 + z^2$$

and the Schwarzschild metric is given by

$$d(x^0)^2 + dr^2 + r^2 d\sigma^2 - \frac{2M}{r} [dx^0 + dr]^2$$

which is the desired Kerr-Schild form.

9.5.8 Isometries.

A vector field X is an infinitesimal isometry or a **Killing** field if its flow preserves the metric. This equivalent to the assertion that

$$L_X \langle Y, Z \rangle = \langle [X, Y], Z \rangle + \langle Y, [X, Z] \rangle \quad (9.83)$$

for all vector fields Y and Z . Now

$$X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle$$

and

$$\nabla_X Y = \nabla_Y X + [X, Y]$$

with a similar equation for Z . So (9.83) is equivalent to

$$\langle \nabla_Y X, Z \rangle + \langle \nabla_Z X, Y \rangle = 0. \quad (9.84)$$

Let S be a submanifold, N a normal vector field to S , and Y, V, W tangential vector fields to S , all extended to vector fields in the ambient manifold. Then along S we have the decomposition

$$\nabla_V Y = \nabla_V^S Y + II(V, Y)$$

into tangential and normal components. So

$$\begin{aligned} 0 &= V\langle N, Y \rangle \\ &= \langle \nabla_V N, Y \rangle + \langle N, II(V, Y) \rangle. \end{aligned}$$

If the submanifold is totally geodesic, so $II(V, Y) \equiv 0$, we see that

$$\langle \nabla_V N, Y \rangle = 0.$$

So for any vector field,

$$\langle \nabla_V X, W \rangle = \langle \nabla_V \tan X, W \rangle$$

and hence if X is a Killing vector field and S a totally geodesic submanifold then the tangential component of X along S is a Killing field for S .

The curvature of the Schwarzschild plane is $2M/r^3$. So any isometry must preserve r since it preserves curvature. Hence it must be of the form

$$(t, r) \mapsto (\phi(t, r), r)$$

and so carries the vector fields

$$\partial_t \mapsto \frac{\partial \phi}{\partial t} \partial_t; \quad \partial_r \mapsto \frac{\partial \phi}{\partial r} \partial_t + \partial_r.$$

Comparing the lengths of ∂_r and its image we see that

$$\frac{\partial \phi}{\partial r} = 0$$

and comparing the lengths of ∂_t and its image shows that

$$\frac{\partial \phi}{\partial t} = 1.$$

So the only isometries of the Schwarzschild plane are translations in t , i.e. $(t, r) \mapsto (t + c, r)$.

Since the planes (at fixed spherical angle) are totally geodesic, this means that the tangential component of any Killing vector, Y must be a multiple of ∂_t . So the most general Killing field is of the form

$$Y = f \partial_t + V$$

where V is vertical and f is a function on S^2 . The claim is that f is a constant and V does not depend on (t, r) and is a Killing vector for the sphere. In the following, U denotes any vector field on the sphere lifted up to be a vertical vector field, and u denotes the value of this vector field at some point $q \in S^2$.

We have

$$\begin{aligned}\nabla_{\partial_r}\partial_t &= \frac{M}{r^2h}\partial_t \\ \nabla_{\partial_r}V &= \frac{1}{r}V \\ \nabla_{\partial_r}U &= \frac{1}{r}U \quad \text{so} \\ \partial_r\langle V,U\rangle &= \langle\nabla_{\partial_r}V,U\rangle + \langle V,\nabla_{\partial_r}U\rangle \\ &= \frac{2}{r}\langle V,U\rangle.\end{aligned}$$

Solving this equation for a fixed point $q \in S^2$ and fixed tangent vector u at q , we see that (at these fixed values)

$$\langle V,U\rangle = g(t)r^2.$$

Now

$$\begin{aligned}\nabla_{\partial_t}U &= 0 \quad \text{so} \\ \langle\nabla_{\partial_t}Y,U\rangle &= \partial_t\langle Y,U\rangle \\ \nabla_UY &= Uf\partial_t + (\dots)\partial_r + \text{vertical} \\ &\text{so} \\ \langle\nabla_UY,\partial_t\rangle &= -hUf \quad \text{so} \\ \langle\nabla_{\partial_t}Y,U\rangle + \langle\nabla_UY,\partial_t\rangle &= 0 \quad (\text{Killing}) \text{ implies} \\ \partial_t\langle V,U\rangle &= 0.\end{aligned}$$

Again, fixing u , this gives

$$g'(t)r^2 = h(r)Uf.$$

But no multiple of r^2 can equal any multiple of $h(r) = 1 - \frac{2M}{r}$ unless both multiples are zero. So $g' = 0$ which implies that

$$\langle V,U\rangle = k(U)r^2.$$

But the factor r^2 is what we multiply the spherical metric by in the Schwarzschild metric. Hence this last equation shows that the projection of V onto the sphere does not depend on r or t . Then it must be a Killing field on S^2 . The condition $Uf \equiv 0$ implies that f is a constant.

Conclusion: the connected group of isometries of the Schwarzschild solution is

$$\mathbf{R} \times SO(3)$$

consisting of time translations and rotations of the sphere.

9.6 Robertson Walker metrics.

These are warped products of an interval $I \subset \mathbf{R}$ (with a negative definite metric) and a three dimensional spacelike manifold, S , of constant curvature. So S is either the three sphere, Euclidean three dimensional space, or hyperbolic three space. We use arc length, t as the coordinate on I , so the total metric has the form

$$-dt^2 + f^2 d\sigma^2$$

where $d\sigma^2$ is the constant curvature metric on S and $f = f(t)$. We write

$$\partial_t f = f', \quad \text{grad } f = -f' \partial_t$$

so covariant derivatives are given by

$$\nabla_{\partial_t}(\partial_t) = 0 \tag{9.85}$$

$$\nabla_{\partial_t} V = \nabla_V(\partial_t) = (f'/f)V \tag{9.86}$$

$$T_V W = \langle V, W \rangle (f'/f) \partial_t \tag{9.87}$$

$$\nabla_V^\vee W = \nabla_V^E W. \tag{9.88}$$

We have $H^f(\partial_t, \partial_t) := f''$ and so

$$R_{V\partial_t}\partial_t = (f''/f)V \tag{9.89}$$

$$R_{VW}\partial_t = 0 \tag{9.90}$$

$$R_{\partial_t V}W = (f''/f)\langle V, W \rangle \partial_t \tag{9.91}$$

$$R_{UV}W = [(f'/f)^2 + (k/f^2)] [\langle U, W \rangle V - \langle V, W \rangle U]. \tag{9.92}$$

where $k = 1, 0$ or -1 is the constant curvature of S . The fiber dimension is $d = 3$ so

$$\text{Ric}(\partial_t, \partial_t) = -\frac{3f''}{f} \tag{9.93}$$

while $\text{Ric}(\partial_t, V) = 0$ as always in a warped product and

$$\text{Ric}(V, W) = \left(2 \left(\frac{f'}{f} \right)^2 + 2 \frac{k}{f} + \frac{f''}{f} \right) \langle V, W \rangle. \tag{9.94}$$

Taking the contraction of the Ricci tensor gives the scalar curvature as

$$S = 6 \left(\left(\frac{f'}{f} \right)^2 + \frac{k}{f^2} + \frac{f''}{f} \right) \tag{9.95}$$

and hence the Einstein tensor $T = \text{Ric} - \frac{1}{2}S\langle \cdot, \cdot \rangle$ is given by

$$T(V, W) = \wp \langle V, W \rangle, \quad \wp := - \left[\frac{2f''}{f} + \left(\frac{f'}{f} \right)^2 + \frac{k}{f^2} \right].$$

Also

$$T(\partial_t, \partial_t) = -\frac{3f''}{f} + \frac{1}{2}S = 3\left(\frac{f'}{f}\right)^2 + \frac{3k}{f^2} := \rho.$$

With these definitions of φ and ρ we can write

$$T = (\rho + \varphi)dt \otimes dt + \varphi\langle \cdot, \cdot \rangle. \quad (9.96)$$

An energy momentum tensor of type

$$T = (\rho + \varphi)\theta \otimes \theta + \varphi\mathbf{g}$$

where X is a forward timelike vector and

$$\theta(\cdot) = \langle X, \cdot \rangle$$

and where ρ and φ are functions, is called a *perfect fluid* for reasons explained in O'Neill. The function φ is called the pressure. The fluid is called a *dust* if $\varphi = 0$. A Robertson Walker model which is a dust is called a Friedman model.

Let us compute the covariant divergence of the T given by (9.96). We compute relative to a frame field whose first component is ∂_t and whose last three components U_1, U_2, U_3 are therefor vertical. The covariant divergence is defined to be

$$\sum \epsilon_i \nabla_{E_i} T(E_i, \cdot).$$

In all situations, the covariant divergent of hg is just dh since $\nabla_E \mathbf{g} = 0$ and

$$\sum dh(E_i) \epsilon_i \langle E_i, \cdot \rangle = dh.$$

Hence we obtain, for $\text{div } T$, the expression

$$-(\rho' + \varphi')dt + (\rho + \varphi) \sum \epsilon_i \nabla_{E_i} (dt)(E_i)dt + \varphi' dt.$$

Now $\nabla_{\partial_t} (dt)(\partial_t) = dt(\nabla_{\partial_t} \partial_t) = 0$, while

$$(\nabla_U dt)(U) = -dt(\nabla_U U) = -\frac{f'}{f}$$

for any unit vector orthogonal to ∂_t . Thus we obtain

$$-\left[\rho' + 3(\rho + \varphi)\frac{f'}{f}\right] dt$$

for the covariant divergence.

9.6.1 Cosmogeny and eschatology.

The function

$$H := \frac{f'}{f}$$

is called the (Hubble) expansion rate for obvious reasons. The vanishing of the covariant divergence of T yields the equation

$$\rho' = -3(\rho + \wp)H. \quad (9.97)$$

if we go back to the definitions of ρ and \wp we see that

$$-6\frac{f''}{f} = \rho + 3\wp.$$

Now in the known universe, $\rho \gg \wp > 0$, so $f'' < 0$. So the graph of f is convex down, i.e. it lies below its tangent line at any point. Let $H_0 = H(t_0)$ denote the Hubble constant at the present time, t_0 . The tangent line to the graph of f at t_0 has slope $H_0 f(t_0)$ and hence is given by the equation

$$\ell(t) = f(t_0) + H_0 f(t_0)(t - t_0).$$

At $t_0 - H_0^{-1}$ the line ℓ crosses the axis. Since $f > 0$ by definition, this shows that the model must fail at some time t_* in the past, no more than H_0^{-1} units of time ago. (The current estimates on Hubble's constant give this value as somewhere between ten and twenty billion years.) Notice also that if $f'(T) < 0$ at some time in the future then the convex downward property implies that the model will also fail at some future time T^* .

For a discussion of further details of the "big bang" and "big crunch" and more specifically Friedman models where it is assumed that $\wp = 0$ see O'Neill.

Chapter 10

Petrov types.

10.1 Algebraic properties of the curvature tensor

The Riemann curvature tensor $\langle R_{XY}Z, W \rangle$ is anti-symmetric in X, Y and in Z, W so can be thought of as a bilinear form on $\wedge^2 TM_m$ at any point m of a semi-Riemann manifold M . It is also invariant under simultaneous interchange of X, Y with Z, W so this bilinear form is symmetric. In addition, it satisfies the cyclicity condition

$$\langle R_{XY}Z, W \rangle + \langle R_{XZ}W, Y \rangle + \langle R_{XW}Y, Z \rangle = 0.$$

We want to consider the algebraic possibilities and properties of this tensor, so will replace TM_x by a general vector space V with non-degenerate scalar product and want to consider symmetric bilinear forms R on $\wedge^2 V$ which satisfy

$$R(v \wedge x, y \wedge z) + R(v \wedge y, z \wedge x) + R(v \wedge z, x \wedge y) = 0. \quad (10.1)$$

For example, if V is four dimensional, then $\wedge^2 V$ is six dimensional, and the space of symmetric bilinear forms on $\wedge^2 V$ is 21 dimensional. The cyclicity condition in this case imposes no constraint on R if v is equal to (and hence linearly dependent on) x, y or z . Hence there is only one equation on R implied by (10.1) in this case. Thus the space of possible curvature tensors at any point in a four dimensional semi-Riemannian manifold is 20 dimensional. The Ricci tensor is the contraction (say with respect to the (1,3) position) of the Riemann curvature:

$$\text{Ric}(R) = C_{13}(R), \quad \text{Ric}(R)(x, y) := \sum \epsilon_a R(e_a \wedge x, e_a \wedge y)$$

where the sum is over any “orthonormal” basis. It is a symmetric tensor on V . So we can think of Ric as a map from the space of possible curvatures to possible Ricci curvatures. If we let

$$\text{Curv}(V) \subset S^2((\wedge^2 V)^*)$$

denote the subspace of the space of symmetric bilinear forms on $\wedge^2 V$ satisfying (10.1). Then

$$\text{Ric} : \text{Curv}(V) \rightarrow S^2(V^*).$$

Let us show that if $\dim V > 2$ this map is surjective. Indeed, suppose that $A \in S^2(V^*)$. Let $A \wedge A$ denote the induced symmetric form on $\wedge^2 V$ so that

$$(A \wedge A)(u \wedge v, x \wedge y) := A(v, x)A(w, y) - A(v, y)A(w, x).$$

Holding v fixed and cyclically summing over w, x, y we get

$$A(v, x)[A(w, y) - A(w, y)] + A(v, y)[A(x, w) - A(w, x)] + A(v, w)[A(y, x) - A(x, y)] = 0.$$

Thus $A \wedge A$ satisfies (10.1). If A and B are two elements of $S^2(V^*)$ we see that

$$A \wedge B + B \wedge A := (A + B) \wedge (A + B) - A \wedge A - B \wedge B$$

also satisfies (10.1). Let $g \in S^2(V^*)$ denote the scalar product itself. We claim that

$$\text{Ric}(g \wedge g) = (n - 1)g.$$

Indeed

$$\begin{aligned} \text{Ric}(g \wedge g)(v, w) &= \sum \epsilon_a (\langle e_a, e_a \rangle \langle v, w \rangle - \langle e_a, w \rangle \langle v, e_a \rangle) \\ &= n \langle v, w \rangle - \sum \epsilon_a \langle v, e_a \rangle \langle e_a, w \rangle \\ &= (n - 1) \langle v, w \rangle. \end{aligned}$$

For any $R \in \text{Curv}(V)$ on $\wedge^2(V)$ define its “scalar curvature” $S = S(R)$ by

$$S := \sum \epsilon_a \text{Ric}(R)(e_a, e_a) = C(\text{Ric}(R)).$$

Also, for any $A \in S^2(V^*)$, we have

$$C(A) := \sum \epsilon_a \text{Ric}(R)(e_a, e_a)$$

so

$$S(R) = C(\text{Ric}(R)).$$

Then

$$\begin{aligned} \sum \epsilon_a (A(e_a, e_a) \langle v, w \rangle - A(e_a, v) \langle e_a, w \rangle) &= C(A) \langle v, w \rangle - A(v, w) \\ \sum \epsilon_a (\langle e_a, e_a \rangle A(v, w) - A(e_a, w) \langle e_a, v \rangle) &= (n - 1)A(v, w) \end{aligned}$$

so

$$\text{Ric}(g \wedge A + A \wedge g) = (n - 2)A + C(A)g \quad (10.2)$$

where $n = \dim V$. Since $\text{Ric}(g \wedge g) = (n - 1)g$ this shows that $\text{Ric} : \text{Curv}(V) \rightarrow S^2(V^*)$ is surjection. We say that R is *Ricci flat* if $\text{Ric}(R) = 0$. Thus in four

dimensions, the space of Ricci flat curvature tensors (at any point) is ten dimensional. The purpose of this chapter is to explain how the complex geometry of spinors leads to a classification of all possible Ricci flat curvatures into five types, the *Petrov* classification published in 1954 in the relatively obscure journal *Sci. Nat. Kazan State University*. In analyzing Petrov type D, Kerr was led to his discovery of the rotating black hole solutions of the Einstein equations, which generalize the Schwarzschild solution, in 1963. Unfortunately we will not have time to study the remarkable properties of this solution. It would take a whole semester.

Let us briefly go back to the general situation where $\dim V > 2$. Let $R \in \text{Curv}(V)$. Then W defined by

$$R = W + \frac{1}{n-2}(g \wedge \text{Ric}(R) + \text{Ric}(R) \wedge g) - \frac{S(R)}{(n-1)(n-2)}g \wedge g$$

satisfies

$$\text{Ric}(W) = 0.$$

It is called the Weyl curvature (or the Weyl component of the Riemann curvature.) It is, as was discovered by Hermann Weyl, a conformal invariant of the metric. In three dimensions we have $\dim \wedge^2 = 3$ and hence $\ker \text{Ric} = 0$, there are no Weyl tensors. They exist in four or more dimensions.

We now turn to the special properties of the curvature tensors in general relativity. In what follows, all vector spaces and tensor products are over the complex numbers unless otherwise specified. All vector spaces are assumed to be finite dimensional.

10.2 Linear and antilinear maps.

A map $\phi : U \rightarrow V$ between vector spaces is called *antilinear* if

$$\phi(a_1 u_1 + a_2 u_2) = \bar{a}_1 \phi(u_1) + \bar{a}_2 \phi(u_2) \quad \forall u_1, u_2 \in U, a_1, a_2 \in \mathbf{C}.$$

The composition of two antilinear maps is linear, and the composition of a linear map with an antilinear map (in either order) is antilinear.

We let $U^\#$ denote the space of all antilinear functions on U , that is the set of all antilinear maps $\phi : U \rightarrow \mathbf{C}$. As usual, we let U^* , the complex dual space of U denote the space of linear maps of $U \rightarrow \mathbf{C}$. We have a canonical linear isomorphism

$$U \rightarrow (U^\#)^\#$$

where $u \in U$ is sent to the antilinear function of $f \in U^\#$ given by

$$f \mapsto \overline{f(u)}.$$

Notice that

$$\overline{f(au)} = \bar{a} \overline{f(u)} = a \cdot \overline{f(u)},$$

so this map of $U \rightarrow U^{\#\#}$ is linear. It is injective and hence bijective since our spaces are finite dimensional.

We define

$$\bar{U} := U^{\#\#}$$

so \bar{U} consists of antilinear functions on U^* .

Given a linear function, ℓ , on an vector space, W , we get an antilinear function by composing with the standard conjugation on the complex numbers, so

$$\bar{\ell} = - \circ \ell, \quad - : \mathbf{C} \rightarrow \mathbf{C}$$

or

$$\bar{\ell}(w) = \overline{\ell(w)} \quad \forall w \in W.$$

Also, starting with an antilinear function we produce a linear function by composition with complex conjugation. Thus, for example, the most general linear function on U^* is of the form

$$\ell \mapsto \ell(u) \quad u \in U,$$

and hence the most general antilinear function on U^* is of the form

$$\ell \mapsto \overline{\ell(u)}.$$

But if we write $\ell = \bar{m} = - \circ m$ where $m \in U^{\#}$, then, considered as a function of m this is the assignment

$$m \mapsto m(u)$$

which is a linear function of m . Thus we have a canonical identification

$$\bar{U} := U^{\#\#} = U^{\#\#*}.$$

Also

$$\bar{\bar{U}} = (U^{\#\#})^{\#\#} = U.$$

We have an antilinear map $u \mapsto \bar{u}$, $U \rightarrow \bar{U}$ given by composition with conjugation on \mathbf{C} as above, where we think of U as U^{**} . So

$$\bar{u}(\ell) = \overline{\ell(u)}, \quad \forall \ell \in U^*$$

or

$$\bar{u}(m) = m(u), \quad m = \bar{\ell} \in U^{\#}.$$

So

$$\bar{\bar{u}}(m) = \overline{m(u)} = \ell(u) = u^{**}(\ell)$$

and thus

$$\bar{\bar{u}} = u$$

under the identification of $\bar{\bar{U}}$ with U .

We also have

$$\overline{U \otimes V} = \bar{U} \otimes \bar{V}$$

as a canonical identification, with

$$\overline{u \otimes v} = \bar{u} \otimes \bar{v}$$

as the map

$$— : U \otimes V \rightarrow \bar{U} \otimes \bar{V}.$$

If b is a bilinear form on U , then we can think of \bar{b} as a bilinear form on \bar{U} according to the rule

$$\bar{b}(\bar{u}, \bar{v}) := \overline{b(u, v)}.$$

Indeed,

$$\begin{aligned} \bar{b}(a\bar{u}, \bar{v}) &= \bar{b}(\overline{au}, \bar{v}) \\ &= \overline{b(au, v)} \\ &= \overline{ab(u, v)} \\ &= a\bar{b}(\bar{u}, \bar{v}), \end{aligned}$$

and similarly for $\bar{b}(\bar{u}, a\bar{v})$.

If b is symmetric or antisymmetric then so is \bar{b} .

10.3 Complex conjugation and real forms.

A complex conjugation of a complex vector space, V , is an antilinear map of V to itself whose square is the identity. Suppose that

$$\dagger : v \mapsto v^\dagger$$

is such a complex conjugation. Then the set of vectors fixed by \dagger ,

$$\{v \mid v^\dagger = v\}$$

is a real vector space. It is called the real form of the complex vector space, V , relative to the conjugation, \dagger . We denote this vector space by V_\dagger , real or simply by V_{real} when \dagger is understood. If $v \in V_{\text{real}}$ then iv satisfies the equation $w^\dagger = -w$ (and we might want to call such vectors “imaginary”). Every vector $u \in V$ can be written in a unique way as

$$u = v + iw, \quad v, w \in V_{\text{real}},$$

indeed

$$v = \frac{1}{2}(v + v^\dagger), \quad w = \frac{-i}{2}(v - v^\dagger).$$

Familiar examples are: V is the set of all $n \times n$ complex matrices and \dagger is conjugate transpose. The real vectors are then the self adjoint matrices. Another example is to start with a real vector space, E , and then complexify it by tensoring with the complex numbers:

$$V = E \otimes_{\mathbf{R}} \mathbf{C}$$

with

$$(x \otimes_{\mathbf{R}} c)^{\dagger} = x \otimes_{\mathbf{R}} \bar{c}.$$

The corresponding real subspace is then identified with our starting space, E . The above remarks about every vector being written as $u = v + iw$ shows that any complex vector space with conjugation can be identified with this example, i.e. as $V = E \otimes_{\mathbf{R}} \mathbf{C}$ where $E = V_{\text{real}}$.

We shall be interested in two other types of examples. Suppose we start with a vector space U and construct $V = U \otimes \bar{U}$. Define complex conjugation by

$$(u \otimes \bar{v})^{\dagger} := v \otimes \bar{u}.$$

So

$$\dagger = s \circ - \otimes -$$

where

$$s : \bar{U} \otimes U \mapsto U \otimes \bar{U}$$

switches the order of the factors. The real subspace is spanned by the elements of the form $u \otimes \bar{u}$.

A second example is $V = U \oplus \bar{U} \cong \bar{U} \oplus U$ with

$$(x + \bar{y})^{\dagger} = y + \bar{x}.$$

The real subspace consists of all $x + \bar{x}$ and hence can be identified with U as a real vector space. That is we can consider U as a vector space over the real numbers (forgetting about multiplication by i), and this can be identified as a real vector space with the real subspace of $U + \bar{U}$. For example, suppose that g is a symmetric (complex) bilinear form on U . We then obtain a complex symmetric bilinear form, \bar{g} on \bar{U} and hence a complex symmetric bilinear form, $g \oplus \bar{g}$ on $U \oplus \bar{U}$ by declaring U and \bar{U} to be orthogonal:

$$(g \oplus \bar{g})(x + \bar{u}, y + \bar{v}) := g(x, y) + \overline{g(u, v)}.$$

This restricts to a real bilinear form on the real subspace:

$$(g \oplus \bar{g})(x + \bar{x}, y + \bar{y}) = 2\text{Re } g(x, y).$$

So under the identification of the real subspace of $U \oplus \bar{U}$ with U , the metric $g \oplus \bar{g}$ becomes identified with the real quadratic form $2\text{Re } g$. Suppose that g is non-degenerate, and we choose a (complex) orthonormal basis, e_1, \dots, e_n for g . So $g(e_i, e_i) = 1$ and $g(e_i, e_j) = 0$ for $i \neq j$. This is always possible for non-degenerate symmetric forms on complex vector spaces. Then $e_1, \dots, e_n, ie_1, \dots, ie_n$ is an orthogonal basis for U as a real vector space with scalar product $\text{Re } g$ and $\text{Re } g(ie_k, ie_k) = -1$. So the metric $2\text{Re } g$ is of type (n, n) on the space U thought of as a $2n$ dimensional real vector space.

10.4 Structures on tensor products.

If U and V are (complex) vector spaces, then

$$\wedge^2(U \otimes V) = S^2(U) \otimes \wedge^2(V) \oplus \wedge^2(U) \otimes S^2(V).$$

Exterior multiplication is given by

$$(u_1 \otimes v_1) \wedge (u_2 \otimes v_2) = u_1 u_2 \otimes v_1 \wedge v_2 + u_1 \wedge u_2 \otimes v_1 v_2$$

where $u_1 u_2$ denotes the product of u_1 and u_2 in the symmetric algebra, and similarly for $v_1 v_2$. We will want to apply this construction to the case $V = \bar{U}$.

If U has an antisymmetric bilinear form, ω , and V has an antisymmetric form, σ , then this induces a symmetric bilinear form on $U \otimes V$ by

$$\langle u_1 \otimes v_1, u_2 \otimes v_2 \rangle = \omega(u_1, u_2) \sigma(v_1, v_2).$$

We will want to apply this construction to $V = \bar{U}$ and $\sigma = \bar{\omega}$.

The symmetric bilinear induced on $U \otimes V$ in turn induces a scalar product on $\wedge^2(U \otimes V) = S^2(U) \otimes \wedge^2(V) \oplus \wedge^2(U) \otimes S^2(V)$ according to the usual rule

$$\begin{aligned} & \langle (u_1 \otimes v_1) \wedge (u_2 \otimes v_2), (u_3 \otimes v_3) \wedge (u_4 \otimes v_4) \rangle = \\ &= \langle u_1 \otimes v_1, u_3 \otimes v_3 \rangle \langle u_2 \otimes v_2, u_4 \otimes v_4 \rangle - \langle u_1 \otimes v_1, u_4 \otimes v_4 \rangle \langle u_2 \otimes v_2, u_3 \otimes v_3 \rangle \\ &= \omega(u_1, u_3) \omega(u_2, u_4) \sigma(v_1, v_3) \sigma(v_2, v_4) - \omega(u_1, u_4) \omega(u_1, u_3) \sigma(v_1, v_4) \sigma(v_2, v_3). \end{aligned}$$

We can interpret this scalar product as follows, put scalar products on the spaces $S^2(U)$ and $\wedge^2(U)$ according to the rules

$$\langle u_1 u_2, u_3 u_4 \rangle := \frac{1}{2} (\omega(u_1, u_3) \omega(u_2, u_4) + \omega(u_1, u_4) \omega(u_2, u_3))$$

and

$$\langle u_1 \wedge u_2, u_3 \wedge u_4 \rangle := \omega(u_1, u_3) \omega(u_2, u_4) - \omega(u_1, u_4) \omega(u_2, u_3).$$

Make similar definitions for $S^2(V), \wedge^2(V)$. Put the tensor product scalar product on $S^2(U) \otimes \wedge^2(V)$ and $\wedge^2(U) \otimes S^2(V)$. Declare the spaces $S^2(U) \otimes \wedge^2(V)$ and $\wedge^2(U) \otimes S^2(V)$ in the direct sum,

$$\wedge^2(U \otimes V) = S^2(U) \otimes \wedge^2(V) \oplus \wedge^2(U) \otimes S^2(V).$$

This direct sum scalar product then coincides with the scalar product described above. In particular, when $V = \bar{U}$ and $\sigma = \bar{\omega}$, and when we think of conjugation as mapping $S^2(U) \otimes \wedge^2(\bar{U}) \mapsto \wedge^2(U) \otimes S^2(\bar{U})$, we are in the situation described above, of $g \oplus \bar{g}$, where g is the tensor product metric on $S^2(U) \otimes \wedge^2(\bar{U})$.

10.5 Spinors and Minkowski space.

Let U be a two dimensional complex vector space with an antisymmetric non-degenerate bilinear form, ω . Then we get a symmetric bilinear form on $U \otimes \bar{U}$. Let us check that the restriction of this symmetric form to the real subspace is real, and is of type (1,3). To see this, let u be any non-zero element of U , and let v be some other vector with

$$\omega(u, v) = \frac{1}{\sqrt{2}}.$$

Then $u \otimes \bar{u}$ is a null vector of $U \otimes \bar{U}$ for any u , since $\omega(u, u) = 0$. Then

$$\langle (u \otimes \bar{u} + v \otimes \bar{v}), (u \otimes \bar{u} + v \otimes \bar{v}) \rangle = 2\omega(u, v)^2 = 1.$$

Also

$$\begin{aligned} \langle u \otimes \bar{u} - v \otimes \bar{v}, u \otimes \bar{u} - v \otimes \bar{v} \rangle &= -1 \\ \langle u \otimes \bar{v} + v \otimes \bar{u}, u \otimes \bar{v} + v \otimes \bar{u} \rangle &= -1 \\ \langle i(u \otimes \bar{v} - v \otimes \bar{u}), i(u \otimes \bar{v} - v \otimes \bar{u}) \rangle &= -1 \end{aligned}$$

and the vectors $u \otimes \bar{u} + v \otimes \bar{v}$, $u \otimes \bar{u} - v \otimes \bar{v}$, $u \otimes \bar{v} + v \otimes \bar{u}$, $i(u \otimes \bar{v} - v \otimes \bar{u})$ are mutually orthogonal, and span the real subspace.

Let $\alpha := u \wedge v$. So α can be characterized as the unique element of $\wedge^2 U$ satisfying $\omega(\alpha) = \frac{1}{2}$. Then

$$u \otimes \bar{u} \wedge (u \otimes \bar{v} + v \otimes \bar{u}) = u^2 \otimes \bar{\alpha} + \alpha \otimes \bar{u}^2.$$

This element of $\wedge^2 T$, where T is the real subspace of $U \otimes \bar{U}$ is the wedge product of a null vector, $u \otimes \bar{u}$ and a spacelike vector orthogonal to the null vector. Hence it corresponds to a “null plane” containing the null vector $u \otimes \bar{u}$.

Thus each non-zero $u \in U$ determines a null vector, $u \otimes \bar{u}$, and a “null plane”, Q_u , corresponding to the decomposable element $u^2 \otimes \bar{\alpha} + \alpha \otimes \bar{u}^2$. Multiplying u by a phase factor, $e^{i\theta}$ multiplies \bar{u} by $e^{-i\theta}$ and hence does not change the null vector $u \otimes \bar{u}$. But it changes the null plane since $u^2 \mapsto e^{2i\theta} u^2$. Geometrically, this amounts to replacing v by $e^{-i\theta} v$ and so rotates the vector $u \otimes \bar{v} + v \otimes \bar{u}$ by 2θ . So $Q_{e^{i\theta} u}$ is obtained from Q_u by rotation through angle 2θ .

We can compute the star operator in terms of the orthonormal basis constructed above from u and v , and find by direct computation that $\star(u^2 \otimes \bar{\alpha}) = \pm i u^2 \otimes \bar{\alpha}$ (the same choice of sign for all u). Since the sign of the star operator is determined by the orientation, we can choose the orientation so that $\star(u^2 \otimes \bar{\alpha}) = i u^2 \otimes \bar{\alpha} \forall u \in U$, and hence the decomposition

$$\wedge^2(U \otimes \bar{U}) = S^2(U) \otimes \wedge^2(\bar{U}) \oplus \wedge^2(U) \otimes S^2(\bar{U})$$

is the decomposition into the $+i$ and $-i$ eigenspaces of (the complexification of) \star on $\wedge^2(U \otimes \bar{U}) = \wedge^2 T \otimes_{\mathbf{R}} \mathbf{C}$.

10.6 Traceless curvatures.

If we use α and $\bar{\alpha}$ to identify $\wedge^2(U)$ and $\wedge^2(\bar{U})$ with \mathbf{C} , we then can write

$$\wedge^2(U \otimes \bar{U}) = S^2(U) \oplus S^2(\bar{U}),$$

as the decomposition into the $\pm i$ eigenvalues of the star operator. Then

$$S^2(\wedge^2(U \otimes \bar{U}))_- = S^2(S^2(U)) \oplus S^2(S^2(\bar{U}))$$

is the -1 eigenspace of the induced action of \star on $S^2(\wedge^2(U \otimes \bar{U}))$. The complex conjugation is the obvious one coming from the complex conjugation $U \rightarrow \bar{U}$. Thus we may identify the space of (real) -1 eigenvectors of \star on $S^2(\wedge^2(T))$ with $S^2(S^2(U))$ considered as a real vector space.

The space $S^2(S^2(U))$ is six dimensional (over the complex numbers). It has an invariant five dimensional subspace, $S^4(U)$, the space of quartic polynomials in elements of U . We can also describe this subspace as follows: we can use the quadratic form on $S^2(U)$ and on $S^2(S^2(U))$ to define a map

$$\mu : S^2(S^2(U)) \rightarrow \text{End } S^2(U),$$

$$\langle \mu(t)s_1, s_2 \rangle = \langle t, s_1 \cdot s_2 \rangle, \quad t \in S^2(S^2(U)), \quad s_1, s_2 \in S^2(U),$$

and where $s_1 \cdot s_2 \in S^2(S^2(U))$. This identifies $S^2(S^2(U))$ with the space of all symmetric operators on $S^2(U)$, symmetric with respect to the quadratic form on $S^2(U)$. The map $t \mapsto \text{tr } \mu(t)$ is a linear form which is invariantly defined. Since $Sl(U)$ acts irreducibly on $S^4(U)$, the restriction of this linear form to $S^4(U)$ must be zero, so we can think of $S^4(U)$ as consisting of traceless operators. Up to an inessential scalar, we can consider the restriction of μ to $S^4(U)$, call it ν , characterized by

$$\langle \nu(t)s_1, s_2 \rangle = \langle t, s_1 s_2 \rangle,$$

where $s_1 s_2 \in S^4(U)$ is the product of s_1 and s_2 in the symmetric algebra, and the scalar product on the right is the scalar product in $S^4(U)$.

10.7 The polynomial algebra.

It will be convenient to deal with the entire symmetric algebra, $S := S(U)$, where S^k denote the homogeneous polynomials of degree k . For any $u \neq 0 \in U$, let us now choose w such that $\omega(u, w) = 1$, and define the derivation on S

$$\iota(u) : S^k \rightarrow S^{k-1}$$

by

$$\iota(u)z = \omega(u, z) \quad \forall z \in U$$

which defines it on generators and hence determines it on all of S . The commutator of any two derivations is a derivation, and the commutator $[i(u), i(u')]$

vanishes on S^1 and hence on S for any pair of vectors u and u' . Thus all derivations $\iota(u)$ commute, and hence $u \mapsto \iota(u)$ extends to a homomorphism

$$\iota : S \rightarrow \text{End } S.$$

This allow us to extend ω to a bilinear form on S by

$$\langle s, t \rangle := [i(s)t]_0$$

where the subscript 0 denotes the component in degree zero. So the spaces S^k and S^ℓ are orthogonal with respect to this bilinear form, and the restriction to $S^k \times S^k$ is symmetric when k is even, and antisymmetric when k is odd.

We can write the operator $\nu(t), t \in S^4$ as

$$\langle \nu(t)s_1, s_2 \rangle = \iota(t)(s_1s_2), \quad s_1, s_2 \in S^2.$$

Since every quartic homogeneous polynomial in two variables is a product of four linear polynomials, $t = u_1u_2u_3u_4$, we can use this formula and the derivation property to describe the operator $\nu(t)$.

10.8 Petrov types.

For example, suppose that $t = u^4$, i.e. all four factors are identical. Then $\iota(u^4)u^k w^{4-k} = 0$, for $k \neq 0$ and $\iota(u^4)u^4 = 12$. Hence

$$\nu(u^4)u^2 = \nu(u^4)uw = 0, \quad \nu(u^4)w^2 = 6u^2.$$

Thus for any non zero $u \in U$, the operator $\nu(u^4)$ is a rank one nilpotent operator with image $\mathbf{C}u^2$.

Suppose that three of the factors of t are the same, and the fourth linearly independent. So we may assume that $t = u^3w$ for $u, w \in U$ with $\omega(u, w) = 1$. Then

$$\iota(u^3w)u^k w^{n-k} = 0, \quad k \neq 1$$

and

$$\iota(u^3w)uw^3 = -1.$$

So

$$\nu(u^3w)u^2 = 0, \quad \nu(u^3w)uw \in \mathbf{C}u^2, \quad \nu(u^3w)w^2 \in \mathbf{C}uw.$$

Thus $\nu(u^3w)$ has kernel $\mathbf{C}u^2$ and image the plane spanned by u^2 and uw in S^2 . The image of this plane is the kernel, so $\nu(u^3w)$ is a two step nilpotent operator.

Next consider the case where $u_1 = u_2, u_3 = u_4, u_2 \neq u_3$, all not zero. The non-zero value of $\omega(u_1, u_3)$ is an invariant. But we can always multiply our element t by a scalar factor, to arrange that this value is one. So up to scalar multiple we have $t = u^2w^2$ for $0 \neq u \in U, \omega(u, w) = 1$. Then

$$\iota(u^2w^2)u^k w^k = 0, \quad k \neq 2, \quad \iota(u^2w^2)u^2w^2 = 4.$$

Our current choice of the normalization of the scalar product on S^2 yields

$$\langle u^2, w^2 \rangle = \langle w^2, u^2 \rangle = 2, \quad \langle uw, uw \rangle = -1$$

all other scalar products equal zero for the basis u^2, uw, w^2 of S^2 . Hence follows that $\iota(u^2w^2)$ is diagonalizable with eigenvalues $-4, 2, 2$:

$$\nu(u^2w^2)u^2 = 2u^2, \quad \nu(u^2w^2)w^2 = 2w^2, \quad \nu(u^2w^2)uw = -4uw.$$

Suppose that exactly two factors are equal. We can assume that the two equal factors are u . Multiplying by scalars if necessary, we can arrange that $\omega(u, u_3) = \omega(u, u_4) = 1$. So $u_3 = w + au$, $u_4 = w + bu$, $a \neq b$. Replacing w by $w - \frac{1}{2}(a+b)u$ we may write

$$t = u^2(w + ru)(w - ru) = u^2w^2 - r^2u^4, \quad r \neq 0,$$

where we have $r = \frac{1}{2}(a - b)$. Now the semisimple element $\nu(u^2w^2)$ commutes with the rank one nilpotent element $\nu(u^4)$ since

$$\nu(u^2w^2) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \nu(u^4) = \begin{pmatrix} 0 & 0 & 6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

in terms of the basis u^2, uw, w^2 of $S^2(U)$. In fact, the form of these two matrices shows that the operator $\nu(u^2w^2 - r^2u^4)$ is not diagonalizable.

Finally, the generic case of four distinct linear factors corresponds to the generic case of three distinct eigenvalues. We thus have the various Petrov types for non-zero elements:

name	# linear factors	structure of $\nu(t)$
<i>I</i>	4 distinct	distinct eigenvalues, diagonalizable
<i>II</i>	3 distinct	$2\lambda, -\lambda, -\lambda$, non-diagonalizable
<i>D</i>	$u_1 = u_2 \neq u_3 = u_4$	$2\lambda, -\lambda, -\lambda$, diagonalizable
<i>III</i>	$u_1 = u_2 = u_3 \neq u_4$	nilpotent, rank 2
<i>N</i>	$u_1 = u_2 = u_3 = u_4$	nilpotent, rank one

10.9 Principal null directions.

We have identified the space $S^2 = S^2(U)$ with the $+i$ eigenspace of \star acting on $\wedge^2 T \otimes \mathbf{C}$. The map $\alpha \mapsto \alpha - i\star\alpha$ is a real linear identification of $\wedge^2 T$ with this eigenspace, under which multiplication by i is pulled back to the star operator. So an element α corresponds to a null vector in S^2 if and only if it satisfies $\alpha \wedge \alpha = 0$ and $\langle \alpha, \alpha \rangle = 0$, and so determines a null plane which is degenerate under the restriction of the Lorentz scalar product. Such a null plane contains a unique null line. We can describe this null plane and null line in terms of S^2 as follows. The null elements of S^2 are those elements which are squares of

linear elements. Indeed, every element of S^2 can be factored into the product of two linear factors, say as uv , and if v is not a multiple of u then $\langle uv, uv \rangle \neq 0$. So the null bivectors in $\wedge^2 T$ correspond to elements of the form u^2 , and the corresponding null line in T is the line spanned by $u \otimes \bar{u}$. If u^2 satisfies

$$\langle \nu(t)u^2, u^2 \rangle = 0$$

then α is called a *principal* null bivector and its corresponding null line is called a *principal* null line, and a non-zero vector in a principal null line is called a principal null vector. If $\alpha = u^2$ then we say that u is a principal spinor.

Projectively, the two quadric curves $\langle \alpha, \alpha \rangle = 0$ and $\langle \nu(t)\alpha, \alpha \rangle = 0$ will intersect at four points, but these points may coalesce to give multiple points of intersection.

The multiplicity, m , of a principal null vector $\ell = u \otimes \bar{u}$ is defined to be

- $m = 1$ if u^2 is not an eigenvector of $\nu(t)$,
- $m = 2$ if u^2 is an eigenvector of $\nu(t)$ with non-zero eigenvalue,
- $m = 3$ if $\nu(t)u^2 = 0$, $\dim \ker \nu(t) = 1$,
- $m = 4$ if $\nu(t)u^2 = 0$ and $\dim \ker \nu(t) = 2$.

The condition for u to be a principal null spinor can be written as

$$\iota(t)u^4 = 0.$$

If we write t as a product of linear factors, $t = u_1 u_2 u_3 u_3$ we see that this is equivalent to saying that $u = u_i$ (up to a constant factor), i.e. that u be a factor of t . If we now go back to the previous section and examine each of the normal forms we constructed for each type, we see that the factorization properties defining the type of t also give the multiplicities of the principal null vectors. So type *I* has four distinct principal null vectors each of multiplicity 1, type *II* has one principal null vector of multiplicity 2 and two of multiplicity 1, type *D* has two principal null spinors each of multiplicity two, type *III* has one of multiplicity 3 and one of multiplicity 1, and type *N* has one principal null vector of multiplicity 4. In symbols:

$$\begin{aligned} I &\Leftrightarrow (1, 1, 1, 1) \\ II &\Leftrightarrow (2, 1, 1) \\ D &\Leftrightarrow (2, 2) \\ III &\Leftrightarrow (3, 1) \\ N &\Leftrightarrow 4. \end{aligned}$$

Here is another description of the multiplicity of a null vector, $k = u \otimes \bar{u}$. The element u^2 corresponds to a bivector $\alpha = k \wedge x$ where x is some spacelike vector perpendicular to k . To say that k is principal is the same as to say that $\mathbf{g}(\nu(t)\alpha, \alpha) = 0$ where \mathbf{g} is the complex scalar product pulled back to $\wedge^2 T$. The

real part of \mathbf{g} is just the original scalar product so $\langle \nu(t)\alpha, \alpha \rangle = 0$. Since we can multiply u and α by an arbitrary phase factor, the condition of being principal is that

$$\langle \nu(t)k \wedge x, k \wedge x \rangle = 0 \quad \forall x \perp k.$$

Writing

$$\langle \nu(t)k \wedge x, k \wedge x \rangle = \langle R_{kx}k, x \rangle$$

and then polarizing, we see that this is the same as saying that

$$\langle R_{kx}k, y \rangle = 0 \quad \forall x, y \perp k. \quad (10.3)$$

We claim that the null vector is principal with multiplicity ≥ 2 if and only if

$$\langle R_{kx}k, y \rangle = 0, \quad \forall x \perp k \text{ and } \forall y. \quad (10.4)$$

Proof. Suppose that $k = u \otimes \bar{u}$ is a factor of order at least two in t . This happens if and only if $\iota(t)u^3v = \iota(t)u^4 = 0$. This is the same as saying that $\nu(t)u^2$ is orthogonal to the complex two dimensional space $u^{2\perp}$ relative to the complex metric. This complex two dimensional space corresponds to a real four dimensional space, the orthogonal complement of the two dimensional subspace of $\wedge^2 T$ spanned by $k \wedge x = u^2$ and $k \wedge z = iu^2$. Here x and z are spacelike vectors orthogonal to k and to each other as above. So u is a repeated factor of t if and only if

$$\langle [R](k \wedge x), \gamma \rangle = 0$$

for all γ in this four dimensional subspace of $\wedge^2 T$ and similarly for z . The four dimensional space in question is spanned by the three dimensional space of elements of the form $k \wedge y$, $y \in T$ and the element $x \wedge z$. In particular, applied to elements of the form $x \wedge y$ we get condition (10.4) for the x we have chosen and also for x replaced by z . It is automatic with x replaced with k since $R_{kk} = 0$. This proves that (10.4) holds if u is a repeated factor.

To prove the converse, we must show that $[R](k \wedge x)$ is orthogonal to the four dimensional subspace of $\wedge^2 T$ spanned by all $k \wedge y$ and $x \wedge z$. Condition (10.4) guarantees the orthogonality for the elements of the form $k \wedge y$. So we must prove that

$$\langle R_{kx}x, z \rangle = 0.$$

This will follow from the Ricci flatness condition. Indeed, choose a null vector ℓ orthogonal to k, x and z with $\langle k, \ell \rangle = 0$. Then

$$0 = Ric[R](k, z) = \sum_{ij} g^{ij} \langle R_{ky_i z, y_j} \rangle$$

as y_i, y_j range over the elements k, ℓ, x, z . This sum reduces to

$$-\langle R_{k\ell z}, k \rangle + \langle R_{kx}z, x \rangle$$

all other terms vanishing. The first term vanishes by (10.4) and this implies the vanishing of the second.

10.10 Kerr-Schild metrics.

We want to use (10.4) to conclude that if a Ricci flat metric is obtained from a flat metric by adding the square of a null form to a flat metric, then the tracefree curvature has a repeated factor. More precisely, if the null form is α , and the corresponding vector field is N , then we will show that this repeated factor is u where $u \otimes \bar{u} = N$. For this we recall the following facts,

- A necessary condition for the new metric to be Ricci flat is that

$$\nabla_N N = \phi N$$

for some function ϕ . Thus the integral curves of N are null geodesics in the old (and new) metric but with a possibly non-affine parametrization.

- The new affine connection differs from the old affine connection by adding the a tensor $A \in T \otimes S^2 T^*$ which can be expressed in terms of the null form. That is, the new connection is

$$\nabla_X Y + A_X Y$$

where ∇ is the old connection and we can write down a formula for $A_X Y$ involving the null form and its covariant derivatives. In particular,

$$A_N = \phi N \otimes \alpha \tag{10.5}$$

i.e.

$$A_N(X) = \phi \alpha(X) N.$$

Also

$$\alpha(A_X \cdot) = \phi \alpha \otimes \alpha$$

i.e.

$$\alpha(A_X Y) = \phi \alpha(X) \alpha(Y). \tag{10.6}$$

- If the affine connection is modified by the addition of a tensor A , then the new curvature differs from the old curvature by

$$R'_{XY} = R_{XY} + [A_X, A_Y] + (\nabla A)(X, Y) - (\nabla A)(Y, X).$$

Here

$$(\nabla A)(X, Y) \in \text{Hom}(T, T)$$

is defined by

$$(\nabla A)(X, Y)Z = \nabla_X(A_Y Z) - A_{\nabla_X Y} Z - A_Y \nabla_X Z$$

where ∇ is the connection relative to the old metric.

In our case the old curvature is zero and we are interested in computing

$$\langle R'_{NX}N, Y \rangle = -\alpha(R_{NX}Y).$$

We have $A_N N = 0$ and $A_N A_X N = A_N A_N X = 0$ so the bracket term makes no contribution. Since N is a null vector field, we have $\nabla_X N \perp N$ and so

$$A_{\nabla_X N} N = A_N \nabla_X N = 0$$

for any X , and $A_N N \equiv 0$ so $\nabla_X(A_N N) = 0$ for any X . So we are left with the formula

$$\alpha(R'_{NX}Y) = \alpha((\nabla_N A)_X Y).$$

Now

$$\alpha(\nabla_N A) = \nabla_N(\alpha(A)) - (\nabla_N \alpha)(A) = -(\nabla_N \phi + \phi^2)\alpha \otimes \alpha$$

or

$$\alpha(R_{NX}Y) = -(\nabla_N \phi + \phi^2)\alpha(X)\alpha(Y). \quad (10.7)$$

In particular, if $X \perp N$ so $\alpha(X) = 0$, the preceding expression vanishes for any Y proving that N is a principal null vector of multiplicity at least two. QED

Chapter 11

Star.

11.1 Definition of the star operator.

We start with a finite dimensional vector space V over the real numbers which carries two additional pieces of structure: an orientation and a non-degenerate scalar product. The scalar product, $\langle \cdot, \cdot \rangle$ determines a scalar product on each of the spaces $\wedge^k V$ which is fixed by the requirement that it take on the values

$$\langle x_1 \wedge \cdots \wedge x_k, y_1 \wedge \cdots \wedge y_k \rangle = \det(\langle x_i, y_j \rangle)$$

on decomposable elements. This scalar product is non-degenerate. Indeed, starting from an “orthonormal” basis e_1, \dots, e_n of V , the basis $e_{i_1} \wedge \cdots \wedge e_{i_k}$, $i_1 < \cdots < i_k$ is an “orthonormal” basis of \wedge^k where

$$\langle e_{i_1} \wedge \cdots \wedge e_{i_k}, e_{i_1} \wedge \cdots \wedge e_{i_k} \rangle = (-1)^r$$

where r is the number of the i_j with $\langle e_{i_j}, e_{i_j} \rangle = -1$.

In particular, there are exactly two elements in the one dimensional space $\wedge^n V$, $n = \dim V$ which satisfy

$$\langle v, v \rangle = \pm 1.$$

Here the ± 1 is determined by the signature (p, q) (p pluses and q minuses) of the scalar product:

$$\langle v, v \rangle = (-1)^q.$$

An orientation of a vector space amounts to choosing one of the two half lines (rays) of non-zero elements in $\wedge^n V$. Hence for an oriented vector space with non-degenerate scalar product there is a well defined unique basis element

$$v \in \wedge^n V \quad \langle v, v \rangle = (-1)^q.$$

Wedge product always gives a bilinear map from $\wedge^k V \times \wedge^{n-k} V \rightarrow \wedge^n V$ But now we have a distinguished basis element for the one dimensional space, $\wedge^n V$.

The wedge product allows us to assign to each element of $\lambda \in \wedge^k V$ the linear function, ℓ_λ on $\wedge^{n-k} V$ given by

$$\lambda \wedge \omega = \ell_\lambda(\omega)v \quad \forall \omega \in \wedge^{n-k} V.$$

But since the induced scalar product on $\wedge^{n-k} V$ is non-degenerate, any linear function ℓ is given as $\ell(\omega) = \langle \tau, \omega \rangle$ for a unique $\tau = \tau(\ell)$. So there is a unique element

$$\star \lambda \in \wedge^{n-k} V$$

determined by

$$\lambda \wedge \omega = \langle \star \lambda, \omega \rangle v. \quad (11.1)$$

This is our convention with regard to the star operator. In short, we have defined a linear map

$$\star : \wedge^k V \rightarrow \wedge^{n-k} V$$

for each $0 \leq k \leq n$ which is determined by (11.1).

Let us choose an orthonormal basis of V as above, but being sure to choose our orthonormal basis to be oriented, which means that

$$v = e_1 \wedge \cdots \wedge e_n.$$

Let $I = (i_1, \dots, i_k)$ be a k -subset of $\{1, \dots, n\}$ with its elements arranged in order, $i_1 < \cdots < i_k$ so that the

$$e_I := e_{i_1} \wedge \cdots \wedge e_{i_k}$$

form an “orthonormal” basis of $\wedge^k V$. Let I^c denote the complementary set of $I \subset \{1, \dots, n\}$ with its elements arranged in increasing order. Thus e_{I^c} is one of the basis elements, $\{e_J\}$ where J ranges over all $(n-k)$ subsets of $\{1, \dots, n\}$. We have

$$e_I \wedge e_J = 0 \quad \text{if } J \neq I^c$$

while

$$e_I \wedge e_{I^c} = (-1)^\pi v$$

where $(-1)^\pi$ is the sign of the permutation required to bring the entries in $e_I \wedge e_{I^c}$ back to increasing order. Thus

$$\star e_I = (-1)^{\pi+r(I^c)} e_{I^c} \quad (11.2)$$

where $(-1)^{\pi+r} := (-1)^\pi (-1)^r$ and

$$r(J) \text{ is the number of } j \in J \text{ with } \langle e_j, e_j \rangle = -1,$$

i.e.

$$(-1)^{r(J)} = \langle e_J, e_J \rangle. \quad (11.3)$$

We should explicate the general definition of the star operator for the extreme cases $k = 0$ and $k = n$. We have $\wedge^0 V = \mathbf{R}$ for any vector space V ,

and the scalar product on \mathbf{R} is the standard one assigning to each real number its square. Taking the number 1 as a basis for \mathbf{R} thought of as a one dimensional vector space over itself, this means that $\langle 1, 1 \rangle = 1$. Wedge product by an element of $\wedge^0 V = \mathbf{R}$ is just ordinary multiplication by of a vector by a real number. So,

$$v \wedge 1 = 1 \wedge v = v$$

and the definition

$$v \wedge 1 = \langle \star v, 1 \rangle v$$

requires that

$$\star v = 1 \tag{11.4}$$

no matter what the signature of the scalar product on V is. On the other hand, $\star 1 = \pm v$. We determine the sign from

$$1 \wedge v = v = \langle \star 1, v \rangle v$$

so

$$\star 1 = \langle v, v \rangle v = (-1)^q v \tag{11.5}$$

in accordance with our general rule.

Applying \star twice gives a linear map of $\wedge^k V$ into itself for each k . We claim that

$$\star^2 = (-1)^{k(n-k)+q} \text{id}. \tag{11.6}$$

Indeed, since both sides are linear operators it suffices to verify this equation on basis elements, e.g. on elements of the form e_I , and by relabeling if necessary we may assume, without loss of generality, that $I = \{1, \dots, k\}$. Then

$$\star(e_1 \wedge \dots \wedge e_k) = (-1)^{r(I^c)} e_{k+1} \wedge \dots \wedge e_n,$$

while

$$\star(e_{k+1} \wedge \dots \wedge e_n) = (-1)^{k(n-k)+r(I)} e_1 \wedge \dots \wedge e_k$$

since there are $n - k$ transpositions needed to bring each of the e_i , $i \leq k$, past $e_{k+1} \wedge \dots \wedge e_n$. Since $r(I) + r(I^c) = q$, (11.6) follows.

11.2 Does $\star : \wedge^k V \rightarrow \wedge^{n-k} V$ determine the metric?

The star operator depends on the metric and on the orientation. Clearly, changing the orientation changes the sign of the star operator.

Let us discuss the question of when the star operator determines the scalar product. We claim, as a preliminary, that it follows from the definition that

$$\lambda \wedge \star \omega = (-1)^q \langle \lambda, \omega \rangle v \quad \forall \lambda, \omega \in \wedge^k \tag{11.7}$$

for any $0 \leq k \leq n$. Indeed, we have really already verified this formula for the case $k = 0$ or $k = n$. For any intermediate k , we observe that both sides are

bilinear in λ and ω , so it suffices to verify this equation on basis elements, i.e. when $\lambda = e_I$ and $\omega = e_K$ where I and K are k -subsets of $\{1, \dots, n\}$. If $K \neq I$ then $\langle e_I, e_K \rangle = 0$, while K^c and I have at least one element in common, so $e_I \wedge \star e_K = 0$. Hence both sides equal zero. So we must only check the equation for $I = K$, and without loss of generality we may assume (by relabeling the indices) that $I = \{1, 2, \dots, k\}$. Then the left hand side of (11.7) is

$$(-1)^{r(I^c)}v$$

while the right hand side is $(-1)^{q+r(I)}v$ by (11.2). Since $q = r(I) + r(I^c)$ the result follows.

One might think that (11.7) implies that \star acting on $\wedge^k V$, $k \neq 0, n$ determines the scalar product, but this is not quite true. Here is the simplest (and very important) counterexample. Take $V = \mathbf{R}^2$ with the standard positive definite scalar product and $k = 1$. So $\star : \wedge^1 V = V \rightarrow V$. In terms of an oriented orthonormal basis we have $\star e_1 = e_2$, $\star e_2 = -e_1$, thus \star is (counterclockwise) rotation through ninety degrees. Any (non-zero) multiple of the standard scalar product will determine the same notion of angle, and hence the same \star operator. Thus, in two dimensions, the \star operator only determines the metric up to scale.

The reason for the breakdown in the argument is that the v occurring on the right hand side of (11.7) depends on the choice of metric. It is clear from (11.7) that the star operator acting on $\wedge^k V$ determines the induced scalar product on $\wedge^k V$ up to scale. Indeed, let $\langle \cdot, \cdot \rangle'$ denote a second scalar product on V . Let v' denote the element of $\wedge^n V$ determined by the scalar product $\langle \cdot, \cdot \rangle'$, so

$$v' = av$$

for some non-zero constant, $a > 0$. Finally, for purposes of the present argument, let us use more precise notation and denote the scalar products induced on $\wedge^k V$ by $\langle \cdot, \cdot \rangle_k$ and $\langle \cdot, \cdot \rangle'_k$. Then (11.7) implies that

$$\langle \cdot, \cdot \rangle'_k = \frac{1}{a} \langle \cdot, \cdot \rangle_k. \quad (11.8)$$

For example, suppose that we know that the original scalar products on V differ by a positive scalar factor, say

$$\langle \cdot, \cdot \rangle' = c \langle \cdot, \cdot \rangle, \quad c > 0.$$

Then

$$\langle \cdot, \cdot \rangle'_k = c^k \langle \cdot, \cdot \rangle_k$$

while

$$v' = \frac{1}{c^{n/2}}v$$

since $\langle v, v \rangle'_n = c^n \langle v, v \rangle$. Hence the fact the the star operators are the same on $\wedge^k V$ implies that $c = 1$ for any k other than $k = \frac{n}{2}$. This was exactly the point of breakdown in our two dimensional example where $n = 2, k = 1$.

In general, if $\langle \cdot, \cdot \rangle$ is positive definite, and $\langle \cdot, \cdot \rangle'$ is any other non-degenerate scalar product, then the principal axis theorem (the diagonalization theorem for symmetric matrices) from linear algebra says that we can find a basis e_1, \dots, e_n which is orthonormal for $\langle \cdot, \cdot \rangle$ and orthogonal with respect to $\langle \cdot, \cdot \rangle'$ with

$$\langle e_i, e_i \rangle' = s_i, \quad s_i \neq 0.$$

Then

$$\langle e_I, e_I \rangle' = s_{i_1} \cdots s_{i_k} \langle e_I, e_I \rangle, \quad I = \{i_1, \dots, i_k\}.$$

The only way that (11.8) can hold for a given $0 < k < n$ is for all the s_i to be equal. Let s denote this common value of the s_i . Then $a = |s|^{-n/2}$ and we can conclude that $s = \pm 1$ if $k \neq n/2$ and in fact that $s = 1$ if, in addition, k is odd.

I don't know how to deal the case of a general (non-definite) scalar product in so straightforward a manner. Perhaps you can work this out. But let me deal with the case of importance to us, a Lorentzian metric on a four dimensional space, so a metric of signature $(1, 3)$ or $(3, 1)$. For $k = 1$, we know from the above discussion that the star operator determines the metric completely. The case $k = 3$ reduces to the case $k = 1$ since $\star^2 = (-1)^{3 \cdot 1 + 1} \text{id} = \text{id}$ in this degree. The only remaining case is $k = 2$, where we know that \star only determines $\langle \cdot, \cdot \rangle_2$ up to a scalar. So the best we can hope for is that $\star : \wedge^2 V \rightarrow \wedge^2 V$ determines $\langle \cdot, \cdot \rangle$ up to a scalar multiple. The following proof (in the form of exercises) involves facts that will be useful to us later on when we study curvature properties of black holes, so we will need them anyway. What we are trying to prove is that (in our situation of a Minkowski metric in four dimensions) the equality $\langle \cdot, \cdot \rangle_2 = b \langle \cdot, \cdot \rangle_2$ for some $b \neq 0$ implies that $\langle \cdot, \cdot \rangle' = s \langle \cdot, \cdot \rangle$ for some $s \neq 0$:

1. Show that the metric $\langle \cdot, \cdot \rangle_2$ induced on $\wedge^2 V$ from the Minkowski metric $\langle \cdot, \cdot \rangle$ on V has signature $(3, 3)$. (It doesn't matter for this result whether we use signature $(1, 3)$ or or $(3, 1)$ for our Minkowski metric.)

2. Let $u, v \in V$. Show that $\langle u \wedge v, u \wedge v \rangle_2 = 0$ if and only if the plane, $P_{\{u, v\}}$ spanned by u and v is degenerate, i.e. the restriction of $\langle \cdot, \cdot \rangle$ to $P_{\{u, v\}}$ is singular. This means that $\{0\} \neq P_{\{u, v\}}^\perp \cap P_{\{u, v\}}$. Now $P_{\{u, v\}}^\perp \neq P_{\{u, v\}}$ since there are no totally null planes in V . So $P_{\{u, v\}}^\perp \cap P_{\{u, v\}}$ is a line consisting of null vectors, that is of vectors n satisfying $\langle n, n \rangle = 0$. Show that all other vectors in $P_{\{u, v\}}$ are spacelike. That is, if $w \in P_{\{u, v\}}, w \notin P_{\{u, v\}}^\perp$ then $\langle w, w \rangle < 0$ if we use the signature $(1, 3)$ or $\langle w, w \rangle > 0$ if we use the signature $(3, 1)$. Conversely, if n is any non-zero nullvector and w is any spacelike vector perpendicular to n then the plane spanned by n and w is a degenerate plane so that $\langle u \wedge v, u \wedge v \rangle_2 = 0$ for any pair of vectors spanning this plane.

Notice that if u', v' is some other pair of vectors spanning the plane $P_{\{u,v\}}$, then $u' \wedge v' = bu \wedge v$ for some scalar $b \neq 0$. Conversely, if $u' \wedge v' = bu \wedge v$, $b \neq 0$, then u', v' span the same plane, $P_{\{u,v\}}$ as do u and v . So every line of null decomposable bivectors (i.e. a line of the form $\{ru \wedge v\}$), $\langle u \wedge v, u \wedge v \rangle_2 = 0$ determines a line of null vectors, $\{cn\}$. Conversely, if we start with the line $\{cn\}$ of null vectors, let

$$Q_n := n^\perp$$

be the orthogonal complement of n . It is a three dimensional subspace of V containing n ; all elements of Q_n not lying on the line $\{cn\}$ being spacelike. The choice of any spacelike vector, w , in Q_n , say with $|\langle w, w \rangle| = 1$ then determines a degenerate plane containing n and lying in Q_n . We thus get a whole “circle” of null planes P with

$$\{0\} \subset \{cn\} \subset P \subset Q_n \subset V.$$

In general, a chain of increasing subspaces is called a “flag” in the mathematical literature. If the dimensions increase by one at each step it is called a “complete flag”. What we have here is that each $u \wedge v$ with $\langle u \wedge v, u \wedge v \rangle_2 = 0$ determines a special kind of complete flag, starting with a line of null vectors. (Penrose uses the following picturesque language: he calls $\{cn\}$ the flagpole about which the plane P rotates.) All this is overkill for our present purpose, but will be needed later on. What we do conclude for our current needs is that the cone of null bivectors, $\{\omega \in \wedge^2 V | \langle \omega, \omega \rangle_2 = 0\}$ determines the cone of null vectors, $N := \{w \in V | \langle w, w \rangle = 0\}$. So we can conclude the proof with the following:

3. Let W be any vector space with a non-degenerate scalar product $\langle \cdot, \cdot \rangle$ of type (p, q) with $p \neq 0, q \neq 0$ and let $N := \{w \in W | \langle w, w \rangle = 0\}$ be its null cone. If $\langle \cdot, \cdot \rangle'$ is any other (non degenerate) scalar product with the same null cone then $\langle \cdot, \cdot \rangle' = s \langle \cdot, \cdot \rangle$ for some non-zero scalar, s .

So we now know that in our four dimensional Minkowski space, a knowledge of $\star : \wedge^2 V \rightarrow \wedge^2 V$ determines the metric up to scale. Here are some more special facts we will need later.

4. Show that $\star : \wedge^2 V \rightarrow \wedge^2 V$ is self adjoint relative to $\langle \cdot, \cdot \rangle_2$, i.e.

$$\langle \star \lambda, \omega \rangle = \langle \lambda, \star \omega \rangle \quad \forall \lambda, \omega \in \wedge^2 V.$$

The next three problems relate to the discussion in Chapter IX. It follows

from our general formula that $\star : \wedge^2 V \rightarrow \wedge^2 V$ satisfies $\star^2 = -\text{id}$. This means that $\star : \wedge^2 V \rightarrow \wedge^2 V$ has eigenvalues i and $-i$. In order to have actual eigenvectors, we must complexify. So we introduce the space

$$\wedge^2 V_{\mathbf{C}} := \wedge^2 V \otimes \mathbf{C}.$$

An element of $\wedge^2 V_{\mathbf{C}}$ is an expression of the form $\lambda + i\omega$, $\lambda, \omega \in \wedge^2 V$. Any linear operator on $\wedge^2 V$ automatically extends to become a complex linear operator on $\wedge^2 V_{\mathbf{C}}$. For example $\star(\lambda + i\omega) := \star\lambda + i\star\omega$. Similarly, every real bilinear form on $\wedge^2 V$ extends to a complex bilinear form on $\wedge^2 V_{\mathbf{C}}$. For example, $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_2$ (we will now revert to the imprecise notation and drop the subscript 2) extends as

$$\langle \lambda + i\omega, \tau + i\rho \rangle := \langle \lambda, \tau \rangle + i\langle \omega, \tau \rangle + i\langle \lambda, \rho \rangle - \langle \omega, \rho \rangle.$$

The subspaces

$$\wedge^2 V_{\mathbf{C}}^+ := \{\lambda - i\star\lambda\} \quad \wedge^2 V_{\mathbf{C}}^- := \{\lambda + i\star\lambda\} \quad \lambda \in \wedge^2 V$$

are complex linear subspaces which are the $+i$ and $-i$ eigenspaces of \star on $\wedge^2 V_{\mathbf{C}}$. They are each of three complex dimensions and

$$\wedge^2 V_{\mathbf{C}} = \wedge^2 V_{\mathbf{C}}^+ \oplus \wedge^2 V_{\mathbf{C}}^-.$$

In the physics literature they have the unfortunate names of the space of “self dual” and “anti-self dual” bivectors.

5. Show that these two subspaces are orthogonal under (the complex extension of) $\langle \cdot, \cdot \rangle$.

The (real) vector space $\wedge^2 V$ has dimension 6. Hence the space of symmetric two tensors over $\wedge^2 V$, the space $S^2(\wedge^2 V)$ has dimension $6 \cdot 7/2 = 21$. The operator $\star : \wedge^2 V \rightarrow \wedge^2 V$ induces an operator (shall we also denote it by \star ?) of $S^2(\wedge^2 V) \rightarrow S^2(\wedge^2 V)$. The eigenvalues of this induced operator will be all possible products of two factors of either i or $-i$, so the eigenvalues of the induced operator $\star : S^2(\wedge^2 V) \rightarrow S^2(\wedge^2 V)$ are ± 1 . The corresponding eigenspaces are now real.

6. Show that the dimension of the -1 eigenspace is 12 and the dimension of the $+1$ eigenspace is 9. (Hint: The dimensions of real eigenspaces do not change if we complexify and then consider dimensions over the complex numbers with the same real eigenvalues of the complexified operator. Describe the space $S^2(W_1 \oplus W_2)$, the symmetric two tensors over a direct sum of two vector spaces.)

The reason that the preceding problem will be of importance to us is that the curvature tensor R at any point of a Lorentzian manifold can be thought of as lying in $S^2(\wedge^2 V)$ where $V = TM_x^*$, the cotangent space at a point. Actually, one of the Bianchi identities (the cyclic sum condition) imposes one additional algebraic constraint on the curvature tensor so that R lies in a 20 dimensional subspace of the 21 dimensional space $S^2(\wedge^2 V)$. The Einstein condition in free space will turn out to further restrict R to lie in an eleven dimensional subspace of the twelve dimensional space of -1 eigenvectors, and the more stringent condition of being Ricci flat will restrict R to lie in a ten dimensional subspace of this eleven dimensional space. We will spend a good bit of time studying this ten dimensional space.

11.3 The star operator on forms.

If M is an oriented semi-Rimannian manifold, we can consider the star operator associated to each cotangent space. Thus, operating pointwise, we get a star operator mapping k -forms into $(n - k)$ forms, where $n = \dim M$:

$$\star : \Omega^k(M) \rightarrow \Omega^{n-k}(M).$$

Many of the important equations of physics have simple expressions in terms of the star operator on forms. the purpose of the rest of these exercises is to describe some of them. In fact, all of the equations we shall write down will be for various star operators of flat space of two, three and four dimensions. But the general formulation goes over unchanged for curved spaces or spacetimes.

11.3.1 For \mathbf{R}^2 .

We take as our orthonormal frame of forms to be dx, dy and the orientation two form to be $v := dx \wedge dy$. Then

$$\star dx = dy, \quad \star dy = -dx$$

as we have already seen.

7. For any pair of smooth real valued functions u and v , let

$$\omega := udx - vdy.$$

Write out the pair of equations

$$d\star\omega = 0, \quad d\omega = 0 \tag{11.9}$$

as a system of two partial differential equations for u and v . (We will find later on that Maxwell's equations in the absence of sources has exactly this same

expression, except that for Maxwell's equations ω is a two form on Minkowski space instead of being a one form on the plane.) If we allow complex valued forms, write $f = u + iv$ and $dz = dx + idy$ then the above pair of equations can be written as

$$d[f dz] = 0.$$

It then follows from Stokes' theorem that the integral of $f dz$ around the boundary of any region where f is defined (and smooth) must be zero. This is known as the Cauchy integral theorem. Notice that

$$f dz = \omega + i \star \omega$$

is the anti-self dual form corresponding to ω in the terminology of the preceding section.

11.3.2 For \mathbf{R}^3 .

We have the orthonormal coframe field dx, dy, dz , with $v = dx \wedge dy \wedge dz$, so $\star 1 = v$,

$$\begin{aligned} \star dx &= dy \wedge dz \\ \star dy &= -dx \wedge dz \\ \star dz &= dx \wedge dy \end{aligned}$$

with

$$\star^2 = 1$$

in all degrees. Let

$$\Delta := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Let

$$\begin{aligned} \theta &= adx + bdy + cdz \\ \Omega &= Adx \wedge dy + Bdx \wedge dz + Cdy \wedge dz. \end{aligned}$$

8. Show that

$$\star d \star d \theta - d \star d \star \theta = -(\Delta a)dx - (\Delta b)dy - (\Delta c)dz \quad (11.10)$$

and

$$-\star d \star d \Omega + d \star d \star \Omega = -(\Delta A)dx \wedge dy - (\Delta B)dx \wedge dz - (\Delta C)dy \wedge dz. \quad (11.11)$$

11.3.3 For $\mathbf{R}^{1,3}$.

We will choose the metric to be of type (1, 3) so that we have the “orthonormal” coframe field cdt, dx, dy, dz with

$$\langle cdt, cdt \rangle = 1$$

and

$$\langle dx, dx \rangle = \langle dy, dy \rangle = \langle dz, dz \rangle = -1.$$

We will choose

$$v = cdt \wedge dx \wedge dy \wedge dz.$$

This fixes the star operator. But I am faced with an awkward notational problem in the next section when we will discuss the Maxwell equations and the relativistic London equations: We will want to deal with the star operator on \mathbf{R}^3 and $\mathbf{R}^{1,3}$ simultaneously, in fact in the same equation. I could use a subscript, say \star_3 to denote the three dimensional star operator and \star_4 to denote the four dimensional star operator. This would clutter up the equations. So I have opted to keep the symbol \star for the star operator in three dimensions and for the purpose of the rest of this section only, use a different symbol, \clubsuit , for the star operator in four dimensions. So

$$\clubsuit(cdt \wedge dx \wedge dy \wedge dz) = 1, \quad \clubsuit 1 = -cdt \wedge dx \wedge dy \wedge dz$$

while

$$\begin{aligned} \clubsuit cdt &= -dx \wedge dy \wedge dz \\ \clubsuit dx &= -cdt \wedge dy \wedge dz \\ \clubsuit dy &= cdt \wedge dx \wedge dz \\ \clubsuit dz &= -cdt \wedge dx \wedge dy \end{aligned}$$

which we can summarize as

$$\begin{aligned} \clubsuit cdt &= -\star 1 \\ \clubsuit \theta &= -cdt \wedge \star \theta \quad \text{for} \\ \theta &= adx + bdy + cdz \end{aligned}$$

and

$$\begin{aligned} \clubsuit cdt \wedge dx &= dy \wedge dz \\ \clubsuit cdt \wedge dy &= -dx \wedge dz \\ \clubsuit cdt \wedge dz &= dx \wedge dy \\ \clubsuit dx \wedge dy &= -cdt \wedge dz \\ \clubsuit dx \wedge dz &= cdt \wedge dy \\ \clubsuit dy \wedge dz &= -cdt \wedge dx. \end{aligned}$$

Notice that the last three equations follow from the preceding three because $\clubsuit^2 = -\text{id}$ as a map on two forms in $\mathbf{R}^{1,3}$. We can summarize these last six equations as

$$\clubsuit(cdt \wedge \theta) = \star\theta, \quad \clubsuit\Omega = -cdt \wedge \star\Omega.$$

I want to make it clear that in these equations $\theta = adx + bdy + cdz$ where the functions a, b , and c can depend on all four variables, t, x, y and z . Similarly Ω is a linear combination of $dx \wedge dy, dx \wedge dz$ and $dy \wedge dz$ whose coefficients can depend on all four variables. So we may think of θ and Ω as forms on three space which depend on time.

We have $\clubsuit^2 = \text{id}$ on one forms and on three forms which checks with

$$\clubsuit(cdt \wedge dx \wedge dy) = -dz$$

or, more generally,

$$\clubsuit(cdt \wedge \Omega) = -\star\Omega.$$

11.4 Electromagnetism.

We begin with two regimes in which we solely use the star operator on \mathbf{R}^3 . Then we will pass to the full relativistic theory.

11.4.1 Electrostatics.

The objects of the theory are:

a linear differential form, E , called the *electric field strength*. A point charge e experiences the force eE . The integral of E along any path gives the voltage drop along that path. The units of E are

$$\frac{\text{voltage}}{\text{length}} = \frac{\text{energy}}{\text{charge} \cdot \text{length}}.$$

The *dielectric displacement*, D , which is a two form. In principle, we could measure $D(v_1, v_2)$ where $v_1, v_2 \in T\mathbf{R}_x^3 \sim \mathbf{R}^3$ are a pair of vectors as follows: construct a parallel-plate capacitor whose plates are metal parallelograms determined by hv_1, hv_2 where h is a small positive number. Place these plates with the corner at x touch them together, then separate them. They acquire charges $\pm Q$. The orientation of \mathbf{R}^3 picks out one of these two plates which we call the top plate. Then

$$D(v_1, v_2) = \lim_{h \rightarrow 0} \frac{\text{charge on top plate}}{h^2}.$$

The units of D are

$$\frac{\text{charge}}{\text{area}}.$$

The *charge density* which is a three form, ρ . (We identify densities with three forms since we have an orientation.)

The key equations in the theory are:

$$dE = 0$$

which, in a simply connected region implies that that $E = -du$ for some function, u called the potential.

The integral of D over the boundary surface of some three dimensional region is the total charge in the region *Gauss' law*:

$$\int_{\partial U} D = \int_U \rho$$

which, by Stokes, can be written differentially as

$$dD = \rho.$$

(I will use units which absorb the traditional 4π into ρ .)

Finally there is a constitutive equation relating E and D . In an isotropic medium it is given by

$$D = \epsilon \star E$$

where ϵ is called the dielectric factor. In a homogeneous medium it is a constant, called the dielectric constant. In particular, the dielectric constant of the vacuum is denoted by ϵ_0 . The units of ϵ_0 are

$$\frac{\text{charge}}{\text{area}} \times \frac{\text{charge} \cdot \text{length}}{\text{energy}} = \frac{(\text{charge})^2}{\text{energy} \cdot \text{length}}.$$

The laws of electrostatics, since they involve the star operator, determine the three dimensional Euclidean geometry of space.

11.4.2 Magnetoquasistatics.

In this regime, it is assumed that there are no static charges, so $\rho = 0$, and that Maxwell's term $\partial D/\partial t$ can be ignored; energy is stored in the magnetic field rather than in capacitors.

The fundamental objects are:

a one form E giving the electric force field. The force on a charge e is eE , as before.

a two form B giving the magnetic induction or the magnetic flux density. The force on a current element I (which is a vector) is $i(I)B$ where i denotes interior product.

The current flux, J which is two form [measured in (amps)/(area)].

a one form, H called the magnetic excitation or the magnetic field. The integral of H over the boundary, C of a surface S is equal to the flux of current through the surface. This is *Ampere's law*.

$$\int_C H = \int_S J \quad (11.12)$$

Faraday's law of induction says that

$$-\frac{d}{dt} \int_S B = \int_C E. \quad (11.13)$$

By Stokes' theorem, the differential form of Ampere's law is

$$dH = J, \quad (11.14)$$

and of Faraday's law is

$$\frac{\partial B}{\partial t} = -dE. \quad (11.15)$$

Faraday's law implies that the time derivative of dB vanishes. But in fact we have the stronger assertion (Hertz's law)

$$dB = 0. \quad (11.16)$$

Equations (11.14), (11.15), and (11.16) are the structural laws of electrodynamics in the magnetoquasistatic approximation. We must supplement them by constitutive equations. One of these is

$$B = \mu \star H, \quad (11.17)$$

where \star denotes the star operator in three dimensions.

According to Ampere's law, H has units

$$\frac{\text{charge}}{\text{time} \cdot \text{length}}$$

while according to Faraday's law B has units

$$\frac{\text{energy} \cdot \text{time}}{\text{charge} \cdot (\text{length})^2}$$

so that μ has units

$$\frac{\text{energy} \cdot (\text{time})^2}{(\text{charge})^2 \cdot \text{length}}.$$

Thus $\epsilon \cdot \mu$ has units

$$\frac{(\text{time})^2}{(\text{length})^2} = (\text{velocity})^{-2}$$

and it was Maxwell's great discovery, the foundation stone of all that has happened since in physics, that

$$\frac{1}{\epsilon_0 \mu_0} = c^2$$

where c is the speed of light. (This discussion is a bit premature in our present regime of quasimagnetostatics where D plays no role.)

We need one more constitutive equation, to relate the current to the electromagnetic field. In ordinary conductivity, one mimics the equation

$$V = RI$$

for a resistor in a network by *Ohm's law*:

$$J = \sigma \star E. \tag{11.18}$$

According to the Drude theory (as modified by Sommerfeld) the charge carriers are free electrons and σ can be determined semi-empirically from a model involving the mean free time between collisions as a parameter. Notice that in ordinary conductivity the charge carrier is something external to the electromagnetic field, and σ is not regarded as a fundamental constant of nature (like c , say) but is an empirical parameter to be derived from another theory, say statistical mechanics. In fact, Drude proposed the theory of the free electron gas in 1900, some three years after the discovery of the electron, by J.J. Thompson, and it had a major success in explaining the law of Wiedemann and Franz, relating thermal conductivity to electrical conductivity. However, if you look at the lengthy article on conductivity in the 1911 edition of the Encyclopedia Britannica, written by J.J. Thompson himself, you will find no mention of electrons in the section on conductivity in solids. The reason is that Drude's theory gave absolutely the wrong answer for the specific heat of metals, and this was only rectified in 1925 in the brilliant paper by Sommerfeld where he replaces Maxwell Boltzmann statistics by the Fermi-Dirac statistics. All this is explained in a solid state physics course. I repeat my main point - σ is not a fundamental constant and the source of J is external to the electromagnetic fields.

11.4.3 The London equations.

In the superconducting domain, it is natural to mimic a network inductor which satisfies the equation

$$V = L \frac{dI}{dt}.$$

So the Londons (1933) introduced the equation

$$E = \Lambda \star \frac{\partial J}{\partial t}, \tag{11.19}$$

where Λ is an empirical parameter similar to the conductance, but the analogue of inductance of a circuit element. Equation (11.19) is known as the *first London equation*. If we assume that Λ is a constant, we have

$$\frac{\partial}{\partial t} \star dH = \star \frac{\partial J}{\partial t} = \frac{1}{\Lambda} E.$$

Setting $H = \mu^{-1} \star B$, applying d , and using (11.15) we get

$$\frac{\partial}{\partial t} \left(d \star d \star B + \frac{\mu}{\Lambda} B \right) = 0. \quad (11.20)$$

From this one can deduce that an applied external field will not penetrate, but not the full Meissner effect expelling all magnetic fields in any superconducting region. Here is a sample argument about the non-penetration of imposed magnetic fields into a superconducting domain: Since $dB = 0$ we can write

$$d \star d \star B = (d \star d \star + \star d \star d) B = -\Delta B,$$

where Δ is the usual three dimensional Laplacian applied to the coefficients of B (using(11.11)). Suppose we have a situation which is invariant under translation in the x and z direction. For example an infinite slab of width $2a$ with sides at $y = \pm a$ parallel to the $y = 0$ plane. Then assuming the solution also invariant, (11.20) becomes

$$\left(\frac{\mu}{\Lambda} - \frac{\partial^2}{\partial y^2} \right) \frac{\partial B}{\partial t} = 0.$$

If we assume symmetry with respect to $y = 0$ in the problem, we get

$$\frac{\partial B}{\partial t} = C(t) \cosh \frac{y}{\lambda},$$

where

$$\lambda = \sqrt{\frac{\Lambda}{\mu}}$$

is called the *penetration depth* of the superconducting material. It is typically of order $.1\mu\text{m}$. Suppose we impose some time dependent external field which takes on the the value

$$b(t)dx \wedge dy,$$

for example, on the surface of the slab. Continuity then gives

$$\frac{\partial B}{\partial t} = b'(t) \frac{\cosh y/\lambda}{\cosh a/\lambda}.$$

The quotient on the right decays exponentially with penetration y/λ . So externally applied magnetic fields do not penetrate, in the sense that the time derivative of the magnetic flux vanishes exponentially within a few multiples of the penetration depth. But the full Meissner effect says that all magnetic fields in the interior are expelled.

So the Londons proposed strengthening (11.20) by requiring that the expression in parenthesis in (11.20) be actually zero, instead of merely assuming that it is a constant. Since $d \star B = \mu dH = \mu J$ (assuming that μ is a constant) we get

$$d \star J = -\frac{1}{\Lambda} B. \quad (11.21)$$

Equation (11.21) is known as *the second London equation*.

11.4.4 The London equations in relativistic form.

We can write the two London equations in relativistic form, by letting

$$j = -J \wedge dt$$

be the three form representing the current in space time. In general, we write

$$j := \rho - J \wedge dt \quad (11.22)$$

as the three form in space time giving the relativistic “current”, but in the quasistatic regime $\rho = 0$.

We have

$$\clubsuit j = \frac{1}{c} \star J,$$

a one form on space time with no dt component (under our assumption of the absence of static charge in our space time splitting). So

$$cd(\clubsuit j) = d_{space} \star J - \frac{\partial \star J}{\partial t} \wedge dt,$$

where the d on the left is the full d operator on space time. (From now on, until the end of this handout, we will be in space-time, and so use d to denote the full d operator in four dimensions, and use d_{space} to denote the three dimensional d operator.)

We recall that in the relativistic treatment of Maxwell’s equations, the electric field and the magnetic induction are combined to give the electromagnetic field

$$F = B + E \wedge dt$$

so that Faraday’s law, (11.15), and Hertz’s law, (11.16) are combined into the single equation,

$$dF = 0, \quad (11.23)$$

known as the *first Maxwell equation*. We see that the two London equations can also be combined to give

$$d\clubsuit c\Lambda j = -F, \quad (11.24)$$

which implies (11.23). This suggests that superconductivity involves modifying Maxwell’s equations, in contrast to ordinary conductivity which is supplementary to Maxwell’s equations.

11.4.5 Maxwell's equations.

To see the nature of this modification, we recall the second Maxwell equation which involves the two form

$$G = D - H \wedge dt$$

where D is the “dielectric displacement”, as above. Recall that

$$d_{space}D$$

gives the density of charge according to Gauss' law. The *second Maxwell equation* combines Gauss' law and Maxwell's modification of Ampere's law into the single equation

$$dG = j, \tag{11.25}$$

where the three current, j is given by (11.22). The product $(\epsilon\mu)^{-1/2}$ has the units of velocity, as we have seen, and let us assume that we are in the vacuum or in a medium for which that this velocity, is c , the same value as the vacuum. So using the corresponding Lorentz metric on space time to define our \clubsuit operator the combined constitutive relations can be written as

$$G = -\frac{1}{c\mu} \clubsuit F,$$

or using units where $c = 1$ more simply as

$$G = -\frac{1}{\mu} \clubsuit F. \tag{11.26}$$

From now on, we will use “natural” units in which $c = 1$ and in which energy and mass have units $(\text{length})^{-1}$.

11.4.6 Comparing Maxwell and London.

The material in this subsection, especially the comments at the end, might be acceptable in the mathematics department. You should be warned that they do not reflect the currently accepted physical theories of superconductivity, and hence might encounter some trouble in the physics department.

In classical electromagnetic theory, j is regarded as a source term in the sense that one introduces a one form, A , the four potential, with

$$F = -dA$$

and Maxwell's equations become the variational equations for the Lagrangian with Lagrange density

$$\mathcal{L}_M(A, j) = \frac{1}{2}dA \wedge \clubsuit dA - \mu A \wedge j. \tag{11.27}$$

This means the following: $\mathcal{L}_M(A, j)$ is a four form on $\mathbf{R}^{1,3}$ and we can imagine the “function”

$$\mathbf{L}_M(A, j) \text{ “} := \text{” } \int_{\mathbf{R}^{1,3}} \mathcal{L}_M(A, j).$$

It is of course not defined because the integral need not converge. But if C is any smooth one form with compact support, the variation

$$d(\mathbf{L}_M)_{(A,j)}[C] := \frac{d}{ds} \mathbf{L}_M(A + sC, j)|_{s=0}$$

is well defined, and the condition that this variation vanish for all such C gives Maxwell’s equations.

9. Show that these variational equations do indeed give Maxwell’s equations. Use $d(C \wedge \clubsuit A) = dC \wedge \clubsuit A - C \wedge d\clubsuit A$ and the fact that $\tau \wedge \clubsuit \omega = \omega \wedge \clubsuit \tau$ for two forms.

In particular, one has gauge invariance: A is only determined up to the addition of a closed one form, and the Maxwell equations become

$$d\clubsuit dA = \mu j. \quad (11.28)$$

For the London equations, if we apply \clubsuit to (11.24) and use (11.26) we get

$$\clubsuit d\clubsuit j = \frac{\mu}{\Lambda} G,$$

and so by the second Maxwell equation, (11.25) we have

$$d\clubsuit d\clubsuit j = \frac{1}{\lambda^2} j. \quad (11.29)$$

We no longer restrict j by requiring the absence of stationary charge, but do observe that “conservation of charge”. i.e. $dj = 0$ is a consequence of (11.29).

If we set

$$\clubsuit \Lambda j = A, \quad (11.30)$$

we see that the Maxwell Lagrange density (11.27) is modified to become the “Proca” Lagrange density

$$\mathcal{L}_L(A) = \frac{1}{2} \left(dA \wedge \clubsuit dA - \frac{1}{\lambda^2} A \wedge \clubsuit A \right). \quad (11.31)$$

10. Verify this.

A number of remarks are in order:

1. The London equations have no gauge freedom.
2. The Maxwell equations in free space (that is with $j = 0$) are conformally invariant. This is a general property of the star operator on middle degrees, in our case from \wedge^2 to \wedge^2 , as we have seen. But the London equations involve the star operator from \wedge^1 to \wedge^3 and hence depend on, and determine, the actual metric and not just on the conformal class. This is to be expected in that the Meissner effect involves the penetration depth, λ .
3. Since the units of λ are length, the units of $1/\lambda^2$ are $(\text{mass})^2$ as is to be expected. So the London modification of Maxwell's equations can be expressed as the addition of a masslike term to the massless photons. In fact, substituting a plane wave with four momentum \mathbf{k} directly into (11.29) shows that \mathbf{k} must lie on the mass shell $\mathbf{k}^2 = 1/\lambda^2$.
4. Since the Maxwell equations are the mass zero limit of the Proca equations, one might say that the London equations represent the more generic situation from the mathematical point of view. Perhaps the "true world" is always superconducting and we exist in some limiting case where the photon can be considered to have mass zero.
5. On the other hand, if one starts from a firm belief in gauge theories, then one would regard the mass acquisition as the result of spontaneous symmetry breaking via the Higgs mechanism. In the standard treatment one gets the Higgs field as the spin zero field given by a Cooper pair. But since the electrons are not needed for charge transport, as no external source term occurs in (11.29), one might imagine an entirely different origin for the Higgs field. Do we need electrons for superconductivity? We don't use them to give mass to quarks or leptons in the standard model.