Lectures on Differential Geometry
Math 240C

John Douglas Moore
Department of Mathematics
University of California
Santa Barbara, CA, USA 93106
e-mail: moore@math.ucsb.edu

June 6, 2011

Preface

This is a set of lecture notes for the course Math 240C given during the Spring of 2011. The notes will evolve as the course progresses.

The starred sections are less central to the course, and may be omitted by some readers.

Contents

Chapter 1

Riemannian geometry

1.1 Review of tangent and cotangent spaces

We will assume some familiarity with the theory of smooth manifolds, as presented, for example, in [3] or in [12].

Suppose that M is a smooth manifold and $p \in M$, and that $\mathcal{F}(p)$ denotes the space of pairs (U, f), where U is an open subset of M containing p and $f : U \to \mathbb{R}$ is a smooth function. If $\phi = (x^1, \ldots, x^n) : U \to \mathbb{R}^n$ is a smooth coordinate system on M with $p \in U$, and $(U, f) \in \mathcal{F}(p)$, we define

$$\left.\frac{\partial}{\partial x^i}\right|_p (f) = D_i(f \circ \phi^{-1})(\phi(p)) \in \mathbb{R},$$

where D_i denotes differentiation with respect to the i-th component. We thereby obtain an $\mathbb{R}$-linear map

$$\left.\frac{\partial}{\partial x^i}\right|_p : \mathcal{F}(p) \longrightarrow \mathbb{R},$$

called a *directional derivative operator*, which satisfies the Leibniz rule,

$$\left.\frac{\partial}{\partial x^i}\right|_p (fg) = \left(\left.\frac{\partial}{\partial x^i}\right|_p (f) \right) g(p) + f(p) \left(\left.\frac{\partial}{\partial x^i}\right|_p (g) \right),$$

and in addition depends only on the "germ" of f at p,

$$f \equiv g \quad \text{on some neighborhood of } p \quad \Rightarrow \left.\frac{\partial}{\partial x_i}\right|_p (f) = \left.\frac{\partial}{\partial x_i}\right|_p (g).$$

The set of all linear combinations

$$\sum_{i=1}^{n} a^i \left.\frac{\partial}{\partial x^i}\right|_p$$

of these basis vectors comprises the *tangent space* to M at p and is denoted by T_pM. Thus for any given smooth coordinate system $(x^1, \dots, x^n)$ on M, we have a corresponding basis

$$\left(\left.\frac{\partial}{\partial x^1}\right|_p, \dots, \left.\frac{\partial}{\partial x^n}\right|_p \right)$$

for the tangent space T_pM.

The notation we have adopted makes it easy to see how the components (a^i) of a tangent vector transform under change of coordinates. If $\psi = (y^1, \dots, y^n)$ is a second smooth coordinate system on M, the new basis vectors are related to the old by the chain rule,

$$\left.\frac{\partial}{\partial y^i}\right|_p = \sum_{j=1}^n \frac{\partial x^j}{\partial y^i}(p) \left.\frac{\partial}{\partial x^j}\right|_p, \quad \text{where} \quad \frac{\partial x^j}{\partial y^i}(p) = D_i(x^j \circ \psi^{-1})(\psi(p)).$$

The disjoint union of all of the tangent spaces forms the *tangent bundle*

$$TM = \bigcup \{T_pM : p \in M\},$$

which has a projection $\pi : TM \to M$ defined by $\pi(T_pM) = p$. If $\phi = (x^1, \dots, x^n)$ is a coordinate system on $U \subset M$, we can define a corresponding coordinate system

$$\tilde{\phi} = (x^1, \dots, x^n, \dot{x}^1, \dots, \dot{x}^n) \quad \text{on} \quad \pi^{-1}(U) \subset TM$$

by letting

$$x^i \left(\sum_{j=1}^n a^j \left.\frac{\partial}{\partial x^j}\right|_p \right) = x^i(p), \qquad \dot{x}^i \left(\sum_{j=1}^n a^j \left.\frac{\partial}{\partial x^j}\right|_p \right) = a^i. \tag{1.1}$$

For the various choices of charts (U, ϕ), the corresponding charts $(\pi^{-1}(U), \tilde{\phi})$ form an atlas making TM into a smooth manifold of dimension $2n$, as you saw in Math 240A.

The *cotangent space* to M at p is simply the dual space T_p^*M to T_pM. Thus an element of T_p^*M is simply a linear functional

$$\alpha : T_pM \longrightarrow \mathbb{R}.$$

Corresponding to the basis

$$\left(\left.\frac{\partial}{\partial x^1}\right|_p, \dots, \left.\frac{\partial}{\partial x^n}\right|_p \right)$$

of T_pM is the dual basis

$$\left(dx^1|_p, \dots, dx^n|_p \right), \quad \text{defined by} \quad dx^i|_p \left(\left.\frac{\partial}{\partial x^j}\right|_p \right) = \delta^i_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

The elements of T_p^*M, called cotangent vectors, are just the linear combinations of these basis vectors

$$\sum_{i=1}^n a_i dx^i|_p$$

Once again, under change of coordinates the basis elements transform by the chain rule,

$$dy^i|_p = \sum_{j=1}^n \frac{\partial y^i}{\partial x^j}(p) dx^j|_p.$$

An important example of cotangent vector is the differential of a function at a point. If $p \in U$ and $f : U \to \mathbb{R}$ is a smooth function, then the *differential* of f at p is the element $df|_p \in T_p^*M$ defined by $df|_p(v) = v(f)$. If $(x^1, \ldots, x^n)$ is a smooth coordinate system defined on U, then

$$df|_p = \sum_{i=1}^n \frac{\partial f}{\partial x^i}(p) dx^i|_p.$$

Just as we did for tangent spaces, we can take the disjoint union of all of the cotangent spaces forms the *cotangent bundle*

$$T^*M = \bigcup \{T_p^*M : p \in M\},$$

which has a projection $\pi : TM \to M$ defined by $\pi(T_pM) = p$. If $\phi = (x^1, \ldots, x^n)$ is a coordinate system on $U \subset M$, we can define a corresponding coordinate system

$$\tilde{\phi} = (x^1, \ldots, x^n, p_1, \ldots, p_n) \quad \text{on} \quad \pi^{-1}(U) \subset TM$$

by letting

$$x^i \left(\sum_{j=1}^n a^j \left. \frac{\partial}{\partial x^j} \right|_p \right) = x^i(p), \qquad p_i \left(\sum_{j=1}^n a_j dx^j|_p \right) = a_i.$$

(It is customary to use $(p_1, \ldots, p_n)$ to denote *momentum coordinates* on the cotangent space.) For the various choices of charts (U, ϕ) on M, the corresponding charts $(\pi^{-1}(U), \tilde{\phi})$ on T^*M form an atlas making T^*M into a smooth manifold of dimension $2n$.

We can generalize this construction and consider *tensor products* of tangent and cotangent spaces. For example, the tensor product of the cotangent space with itself, denoted by $\otimes^2 T_p^*M$, is the linear space of bilinear maps

$$g : T_pM \times T_pM \longrightarrow \mathbb{R}.$$

If $\phi = (x^1, \ldots, x^n) : U \to \mathbb{R}^n$ is a smooth coordinate system on M with $p \in U$, we can define

$$dx^i|_p \otimes dx^j|_p : T_pM \times T_pM \longrightarrow \mathbb{R}$$

$$\text{by} \quad dx^i|_p \otimes dx^j|_p \left(\left. \frac{\partial}{\partial x^k} \right|_p, \left. \frac{\partial}{\partial x^l} \right|_p \right) = \delta^i_k \delta^j_l.$$

Then

$$\{dx^i|_p \otimes dx^j|_p : 1 \leq i \leq n, 1 \leq j \leq n\}$$

is a basis for $\otimes^2 T_p^* M$, and a typical element of $\otimes^2 T_p^* M$ can be written as

$$\sum_{i,j=1}^{n} g_{ij}(p) dx^i|_p \otimes dx^j|_p,$$

where the $g_{ij}(p)$'s are elements of $\mathbb{R}$.

1.2 Riemannian metrics

Definition. Let M be a smooth manifold. A *Riemannian metric* on M is a function which assigns to each $p \in M$ a (positive-definite) inner product $\langle \cdot, \cdot \rangle_p$ on $T_p M$ which "varies smoothly" with $p \in M$. A *Riemannian manifold* is a pair $(M, \langle \cdot, \cdot \rangle)$ consisting of a smooth manifold M together with a Riemannian metric $\langle \cdot, \cdot \rangle$ on M.

Of course, we have to explain what we mean by "vary smoothly." This is most easily done in terms of local coordinates. If $\phi = (x^1, \ldots, x^n) : U \to \mathbb{R}^n$ is a smooth coordinate system on M, then for each choice of $p \in U$, we can write

$$\langle \cdot, \cdot \rangle_p = \sum_{i,j=1}^{n} g_{ij}(p) dx^i|_p \otimes dx^j|_p.$$

We thus obtain functions $g_{ij} : U \to \mathbb{R}$, and we say that $\langle \cdot, \cdot \rangle_p$ varies smoothly with p if the functions g_{ij} are smooth. We call the functions g_{ij} the *components* of the Riemannian metric with respect to the coordinate system $\phi = (x^1, \ldots, x^n)$.

Note that the functions g_{ij} satisfy the symmetry condition $g_{ij} = g_{ji}$ and the condition that the matrix (g_{ij}) be positive definite. We will sometimes write

$$\langle \cdot, \cdot \rangle|_U = \sum_{i,j=1}^{n} g_{ij} dx^i \otimes dx^j.$$

If $\psi = (y^1, \ldots, y^n)$ is a second smooth coordinate system on $V \subseteq M$, with

$$\langle \cdot, \cdot \rangle|_V = \sum_{i,j=1}^{n} h_{ij} dy^i \otimes dy^j,$$

it follows from the chain rule that, on $U \cap V$,

$$g_{ij} = \sum_{k.l=1}^{n} h_{kl} \frac{\partial y^k}{\partial x^i} \frac{\partial y^l}{\partial x^j}.$$

We will sometimes adopt the Einstein summation convention and leave out the summation sign:

$$g_{ij} = h_{kl}\frac{\partial y^k}{\partial x^i}\frac{\partial y^l}{\partial x^j}.$$

We remark in passing that this is how a "covariant tensor field of rank two" transforms under change of coordinates.

Using a Riemannian metric, one can "lower the index" of a tangent vector at p, producing a corresponding cotangent vector and vice versa. Indeed, if $v \in T_pM$, we can construct a corresponding cotangent vector α_v by the formula

$$\alpha_v(w) = \langle v, w\rangle_p.$$

In terms of components,

$$\text{if} \quad v = \sum_{i=1}^n a^i \left.\frac{\partial}{\partial x^i}\right|_p, \quad \text{then} \quad \alpha_v = \sum_{i,j=1}^n g_{ij}(p)a^j dx^i|_p.$$

Similarly, given a cotangent vector $\alpha \in T_p^*M$ we "raise the index" to obtain a corresponding tangent vector $v_\alpha \in T_pM$. In terms of components,

$$\text{if} \quad \alpha = \sum_{i=1}^n a_i dx^i|_p, \quad \text{then} \quad v_\alpha = \sum_{i,j=1}^n g^{ij}(p)a_j \left.\frac{\partial}{\partial x^i}\right|_p,$$

where $(g^{ij}(p))$ is the matrix inverse to $(g_{ij}(p))$. Thus a Riemannian metric transforms the differential $df|_p$ of a function to a tangent vector

$$\operatorname{grad}(f)(p) = \sum_{i,j=1}^n g^{ij}(p)\frac{\partial f}{\partial x^j}(p) \left.\frac{\partial}{\partial x^i}\right|_p,$$

called the *gradient* of f at p. Needless to say, in elementary several variable calculus this raising and lowering of indices is done all the time using the usual Euclidean dot product as Riemannian metric.

Example 1. The simplest example of a Riemannian manifold is n-dimensional *Euclidean space* $\mathbb{E}^n$, which is simply $\mathbb{R}^n$ together with its standard rectangular cartesian coordinate system $(x^1, \ldots, x^n)$, and the Euclidean metric

$$\langle\cdot,\cdot\rangle_E = dx^1 \otimes dx^1 + \cdots + dx^n \otimes dx^n.$$

In this case, the components of the metric are simply

$$g_{ij} = \delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

We will often think of the Euclidean metric as being defined by the dot product,

$$\left\langle \sum_{i=1}^n a^i \left.\frac{\partial}{\partial x^i}\right|_p, \sum_{j=1}^n b^j \left.\frac{\partial}{\partial x^j}\right|_p \right\rangle = \left(\sum_{i=1}^n a^i \left.\frac{\partial}{\partial x^i}\right|_p\right) \cdot \left(\sum_{j=1}^n b^j \left.\frac{\partial}{\partial x^j}\right|_p\right) = \sum_{i=1}^n a^i b^i.$$

Example 2. Suppose that M is an n-dimensional smooth manifold and that $F : M \to \mathbb{R}^N$ is a smooth imbedding. We can give $\mathbb{R}^N$ the Euclidean metric defined in the preceding example. For each choice of $p \in M$, we can then define an inner product $\langle \cdot, \cdot \rangle_p$ on T_pM by

$$\langle v, w \rangle_p = F_{*p}(v) \cdot F_{*p}(w), \quad \text{for } v, w \in T_pM.$$

Here F_{*p} is the differential of F at p defined in terms of a smooth coordinate system $\phi = (x^1, \ldots, x^n)$ by the explicit formula

$$F_{*p}\left(\left. \frac{\partial}{\partial x^i} \right|_p \right) = D_i(F \circ \phi^{-1})(\phi(p)) \in \mathbb{R}^N.$$

Clearly, $\langle v, w \rangle_p$ is symmetric, and it is positive definite because F is an immersion. Moreover,

$$\begin{aligned} g_{ij}(p) = \left\langle \left. \frac{\partial}{\partial x^i} \right|_p, \left. \frac{\partial}{\partial x^j} \right|_p \right\rangle &= F_{*p}\left(\left. \frac{\partial}{\partial x^i} \right|_p \right) \cdot F_{*p}\left(\left. \frac{\partial}{\partial x^j} \right|_p \right) \\ &= D_i(F \circ \phi^{-1})(\phi(p)) \cdot D_j(F \circ \phi^{-1})(\phi(p)), \end{aligned}$$

so $g_{ij}(p)$ depends smoothly on p. Thus the imbedding F induces a Riemannian metric $\langle \cdot, \cdot \rangle$ on M, and we write

$$\langle \cdot, \cdot \rangle = F^* \langle \cdot, \cdot \rangle_E.$$

It is an interesting fact that this construction includes *all* Riemannian manifolds.

Definition. Let $(M, \langle \cdot, \cdot \rangle)$ be a Riemannian manifold, and suppose that $\mathbb{E}^N$ denotes $\mathbb{R}^N$ with the Euclidean metric. An imbedding $F : M \to \mathbb{E}^N$ is said to be *isometric* if $\langle \cdot, \cdot \rangle = F^* \langle \cdot, \cdot \rangle_E$.

The local problem of finding an isometric imbedding is equivalent to solving the nonlinear system of partial differential equations

$$\frac{\partial F}{\partial x_i} \cdot \frac{\partial F}{\partial x_j} = g_{ij},$$

where the g_{ij}'s are known functions and the $\mathbb{E}^n$-valued function F is unknown. Unfortunately, this system does not belong to one of the standard types (elliptice, parabolic, hyperbolic) that are studied in PDE theory.

Nash's Imbedding Theorem. *If $(M, \langle \cdot, \cdot \rangle)$ is any smooth Riemannian manifold, there exists an isometric imbedding $F : M \to \mathbb{E}^N$ into Euclidean space of some dimension N.*

This was regarded as a landmark theorem when it first appeared [18]. The proof is difficult, involves subtle techniques from the theory of nonlinear partial differential equations, and is beyond the scope of this course.

A special case of Example 2 consists of two-dimensional smooth manifolds which are imbedded in $\mathbb{E}^3$. These are usually called *smooth surfaces* in $\mathbb{E}^3$ and are studied extensively in undergraduate courses in "curves and surfaces." This subject was extensively developed during the nineteenth century and was summarized in 1887-96 in a monumental four-volume work, *Leçons sur la théorie générale des surfaces et les applications géométriques du calcul infinitésimal*, by Jean Gaston Darboux. Indeed, the theory of smooth surfaces in $\mathbb{E}^3$ still provides much geometric intuition regarding Riemannian geometry of higher dimensions.

What kind of geometry does a Riemannian metric provide a smooth manifold M? Well, to begin with, we can use a Riemannian metric to define the lengths of tangent vectors. If $v \in T_pM$, we define the length of v by the formula

$$\|v\| = \sqrt{\langle v, v\rangle_p}.$$

Second, we can use the Riemannian metric to define angles between vectors: The angle θ between two nonzero vectors $v, w \in T_pM$ is the unique $\theta \in [0, \pi]$ such that

$$\langle v, w\rangle_p = \|v\|\|w\| \cos\theta.$$

Third, one can use the Riemannian metric to define lengths of curves. Suppose that $\gamma : [a, b] \to M$ is a smooth curve with velocity vector

$$\gamma'(t) = \sum_{i=1}^{n} \frac{dx^i}{dt} \left.\frac{\partial}{\partial x^i}\right|_{\gamma(t)} \in T_{\gamma(t)}M, \quad \text{for } t \in [a, b].$$

Then the *length* of γ is given by the integral

$$L(\gamma) = \int_a^b \sqrt{\langle \gamma'(t), \gamma'(t)\rangle_{\gamma(t)}}dt.$$

We can also write this in local coordinates as

$$L(\gamma) = \int_a^b \sqrt{\sum_{i,j=1}^{n} g_{ij}(\gamma(t))\frac{dx^i}{dt}\frac{dx^j}{dt}}dt.$$

Note that if $F : M \to \mathbb{E}^N$ is an isometric imbedding, then $L(\gamma) = L(F \circ \gamma)$. Thus the lengths of a curve on a smooth surface in $\mathbb{E}^3$ is just the length of the corresponding curve in $\mathbb{E}^3$. Since any Riemannian manifold can be isometrically imbedded in some $\mathbb{E}^N$, one might be tempted to try to study the Riemannian geometry of M via the Euclidean geometry of the ambient Euclidean space. However, this is not necessarily an efficient approach, since sometimes the isometric imbedding is quite difficult to construct.

Example 3. Suppose that $\mathbb{H}^2 = \{(x, y) \in \mathbb{R}^2 : y > 0\}$, with Riemannian metric

$$\langle \cdot, \cdot\rangle = \frac{1}{y^2}(dx \otimes dx + dy \otimes dy).$$

A celebrated theorem of David Hilbert states that $(\mathbb{H}^2, \langle\cdot,\cdot\rangle)$ has no isometric imbedding in $\mathbb{E}^3$ and although isometric imbeddings in Euclidean spaces of higher dimension can be constructed, none of them is easy to describe. The Riemannian manifold $(\mathbb{H}^2, \langle\cdot,\cdot\rangle)$ is called the *Poincaré upper half-plane*, and figures prominently in the theory of Riemann surfaces. It is the foundation for the non-Euclidean geometry discovered by Bolyai and Lobachevsky in the nineteenth century.

1.3 Geodesics

Our first goal is to generalize concepts from Euclidean geometry to Riemannian geometry. One of principal concepts in Euclidean geometry is the notion of straight line. What is the analog of this concept in Riemannian geometry? One candidate would be the curve between two points in a Riemannian manifold which has shortest length (if such a curve exists).

1.3.1 Smooth paths

Suppose that p and q are points in the Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$. If a and b are real numbers with $a < b$, we let

$$\Omega_{[a,b]}(M;p,q) = \{ \text{ smooth paths } \gamma : [a,b] \to M : \gamma(a) = p, \gamma(b) = q\}.$$

We can define two functions $L, J : \Omega_{[a,b]}(M;p,q) \to \mathbb{R}$ by

$$L(\gamma) = \int_a^b \sqrt{\langle \gamma'(t), \gamma'(t)\rangle_{\gamma(t)}}dt, \qquad J(\gamma) = \frac{1}{2}\int_a^b \langle \gamma'(t), \gamma'(t)\rangle_{\gamma(t)}dt.$$

Although our geometric goal is to understand the *length* L, it is often convenient to study this by means of the closely related *action* J. Notice that L is invariant under reparametrization of γ, so once we find a single curve which minimizes L we have an infinite-dimensional family. This, together with the fact that the formula for L contains a troublesome radical in the integrand, make J far easier to work with than L.

It is convenient to regard J as a smooth function on the "infinite-dimensional" manifold $\Omega_{[a,b]}(M;p,q)$. At first, we use the notion of infinite-dimensional manifold somewhat informally, in more advanced courses on global analysis it is important to make the notion precise.

Proposition 1. $[L(\gamma)]^2 \leq 2(b-a)J(\gamma)$. *Moreover, equality holds if and only if* $\langle \gamma'(t)\gamma'(t)\rangle$ *is constant if and only if* γ *has constant speed.*

Proof: We use the Cauchy-Schwarz inequality:

$$L(\gamma)]^2 = \left[\int_a^b \sqrt{\langle\gamma'(t),\gamma'(t)\rangle}dt\right]^2$$
$$\leq \left[\int_a^b dt\right]\left[\int_a^b \langle\gamma'(t),\gamma'(t)\rangle dt\right] = 2(b-a)J(\gamma). \quad (1.2)$$

Equality holds if and only if the functions $\langle\gamma'(t),\gamma'(t)\rangle$ and 1 are linearly dependent, that is, if and only if γ has constant speed.

Proposition 2. *Suppose that M has dimension at least two. An element $\gamma \in \Omega_{[a,b]}(M;p,q)$ minimizes J if and only if it minimizes L and has constant speed.*

Sketch of proof: One direction is easy. Suppose that γ has constant speed and minimizes L. Then, if $\lambda \in \Omega_{[a,b]}(M;p,q)$,

$$2(b-a)J(\gamma) = [L(\gamma)]^2 \leq [L(\lambda)]^2 \leq 2(b-a)J(\lambda),$$

and hence $J(\gamma) \leq J(\lambda)$.

We will only sketch the proof of the other direction for now; later a complete proof will be available. Suppose that γ minimizes J, but does not minimize L, so there is $\lambda \in \Omega$ such that $L(\lambda) < L(\gamma)$. Approximate λ by an immersion λ_1 such that $L(\lambda_1) < L(\gamma)$; this is possible by a special case of an approximation theorem due to Whitney (see [10], page 27). Since the derivative λ_1' is never zero, the function $s(t)$ defined by

$$s(t) = \int_a^t |\lambda_1'(t)|dt$$

is invertible and λ_1 can be reparametrized by arc length. It follows that we can find an element of $\lambda_2 : [a,b] \to M$ which is a reparametrization of λ_1 of constant speed. But then

$$2(b-a)J(\lambda_2) = [L(\lambda_2)]^2 = [L(\lambda_1)]^2 < [L(\gamma)]^2 \leq 2(b-a)J(\gamma),$$

a contradiction since γ was supposed to minimize J. Hence γ must in fact minimize L. By a similar argument, one shows that if γ minimizes J, it must have constant speed.

The preceding propositions motivate use of the function $J : \Omega_{[a,b]}(M;p,q) \to \mathbb{R}$ instead of L. We want to develop enough of the calculus on the "infinite-dimensional manifold" $\Omega_{[a,b]}(M;p,q)$ to enable us to find the critical points of J. This is exactly the idea that Marston Morse used in his celebrated "calculus of variations in the large," a subject presented beautifully in Milnor's classical book [15].

To start with, we need the notion of a smooth curve

$$\bar{\alpha} : (-\epsilon, \epsilon) \to \Omega_{[a,b]}(M; p, q) \quad \text{such that} \quad \bar{\alpha}(0) = \gamma,$$

where γ is a given element of Ω. (In fact, we would like to define smooth charts on the path space $\Omega_{[a,b]}(M; p, q)$, but for now a simpler approach will suffice.) We will say that a *variation* of γ is a map

$$\bar{\alpha} : (-\epsilon, \epsilon) \to \Omega_{[a,b]}(M; p, q)$$

such that $\bar{\alpha}(0) = \gamma$ and if

$$\alpha : (-\epsilon, \epsilon) \times [a, b] \to M \quad \text{is defined by} \quad \alpha(s, t) = \bar{\alpha}(s)(t),$$

then α is smooth.

Definition. An element $\gamma \in \Omega_{[a,b]}(M; p, q)$ is a *critical point* for J if

$$\frac{d}{ds} \left(J(\bar{\alpha}(s)) \right) \Big|_{s=0} = 0, \quad \text{for every variation } \bar{\alpha} \text{ of } \gamma. \tag{1.3}$$

Definition. An element $\gamma \in \Omega_{[a,b]}(M; p, q)$ is called a *geodesic* if it is a critical point for J.

Thus the geodesics are the candidates for curves of shortest length from p to q, that is candidates for the generalization of the notion of straight line from Euclidean to Riemannian geometry.

We would like to be able to determine the geodesics in Riemannian manifolds. It is easiest to do this for the case of a Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$ that has been provided with an isometric imbedding in $\mathbb{E}^N$. Thus we imagine that $M \subseteq \mathbb{E}^N$ and thus each tangent space $T_p M$ can be regarded as a linear subspace of $\mathbb{R}^N$. Moreover,

$$\langle v, w \rangle_p = v \cdot w, \quad \text{for} \quad v, w \in T_p M \subseteq \mathbb{E}^N,$$

where the dot on the right is the dot product in $\mathbb{E}^N$. If

$$\bar{\alpha} : (-\epsilon, \epsilon) \to \Omega_{[a,b]}(M; p, q)$$

is a variation of an element $\gamma \in \Omega_{[a,b]}(M; p, q)$, with corresponding map

$$\alpha : (-\epsilon, \epsilon) \times [a, b] \to M \subseteq \mathbb{E}^N,$$

then

$$\begin{aligned} \frac{d}{ds} \left(J(\bar{\alpha}(s)) \right) \Big|_{s=0} &= \frac{d}{ds} \left[\frac{1}{2} \int_a^b \frac{\partial \alpha}{\partial t}(s,t) \cdot \frac{\partial \alpha}{\partial t}(s,t) dt \right] \Bigg|_{s=0} \\ &= \int_a^b \frac{\partial^2 \alpha}{\partial s \partial t}(s,t) \cdot \frac{\partial \alpha}{\partial t}(s,t) dt \Bigg|_{s=0} = \int_a^b \frac{\partial^2 \alpha}{\partial s \partial t}(0,t) \cdot \frac{\partial \alpha}{\partial t}(0,t) dt, \end{aligned}$$

where α is regarded as an $\mathbb{E}^N$-valued function. If we integrate by parts, and use the fact that

$$\frac{\partial \alpha}{\partial s}(0,b) = 0 = \frac{\partial \alpha}{\partial s}(0,a),$$

we find that

$$\frac{d}{ds}\left(J(\bar{\alpha}(s))\right)\Big|_{s=0} = -\int_a^b \frac{\partial \alpha}{\partial s}(0,t) \cdot \frac{\partial^2 \alpha}{\partial t^2}(0,t)dt = -\int_a^b V(t) \cdot \gamma''(t)dt, \quad (1.4)$$

where $V(t) = (\partial\alpha/\partial s)(0,t)$ is called the *variation field* of the variation field $\bar{\alpha}$.

Note that $V(t)$ can be an arbitrary smooth $\mathbb{E}^N$-valued function such that

$$V(a) = 0, \quad V(b) = 0, \quad V(t) \in T_{\gamma(t)}M, \quad \text{for all } t \in [a,b],$$

that is, V can be an arbitrary element of the "tangent space"

$$T_\gamma\Omega = \{ \text{ smooth maps } V : [a,b] \to \mathbb{E}^N \\ \text{such that } V(a) = 0 = V(b), V(t) \in T_{\gamma(t)}M \text{ for } t \in [a,b] \}.$$

We can define a linear map $dJ(\gamma) : T_\gamma\Omega \to \mathbb{R}$ by

$$dJ(\gamma)(V) = -\int_a^b \langle V(t), \gamma''(t)\rangle \, dt = \frac{d}{ds}\left(J(\bar{\alpha}(s))\right)\Big|_{s=0},$$

for any variation $\bar{\alpha}$ with variation field V. We think of $dJ(\gamma)$ as the *differential* of J at γ.

If $dJ(\gamma)(V) = 0$ for all $V \in T_\gamma\Omega$, then $\gamma''(t)$ must be perpendicular to $T_{\gamma(t)}M$ for all $t \in [a,b]$. In other words, $\gamma : [a,b] \to M$ is a geodesic if and only if

$$\left(\gamma''(t)\right)^\top = 0, \quad \text{for all } t \in [a,b], \quad (1.5)$$

where $\left(\gamma''(t)\right)^\top$ denotes the orthogonal projection of $\gamma''(t)$ into $T_{\gamma(t)}M$. To see this rigorously, we choose a smooth function $\eta : [a,b] \to \mathbb{R}$ such that

$$\eta(a) = 0 = \eta(b), \quad \eta > 0 \quad \text{on} \quad (a,b),$$

and set

$$V(t) = \eta(t)\left(\gamma''(t)\right)^\top,$$

which is clearly an element of $T_\gamma\Omega$. Then $dJ(\gamma)(V) = 0$ implies that

$$\int_a^b \eta(t)\left(\gamma''(t)\right)^\top \cdot \gamma''(t)dt = \int_a^b \eta(t)\left\|\left(\gamma''(t)\right)^\top\right\|^2 dt = 0.$$

Since the integrand is nonnegative it must vanish identically, and (1.5) must indeed hold.

We have thus obtained a simple equation (1.5) which characterizes geodesics in a submanifold M of $\mathbb{E}^N$. The geodesic equation is a generalization of the simplest second-order linear ordinary differential equation, the equation of a particle moving with zero acceleration in Euclidean space, which asks for a vector-valued function

$$\gamma : (a, b) \longrightarrow \mathbb{E}^N \qquad \text{such that} \qquad \gamma''(t) = 0.$$

Its solutions are the constant speed straight lines. The simplest way to make this differential equation nonlinear is to consider an imbedded submanifold M of $\mathbb{E}^N$ with the induced Riemannian metric, and ask for a function

$$\gamma : (a, b) \longrightarrow M \subset \mathbb{E}^N \qquad \text{such that} \qquad (\gamma''(t))^\top = 0.$$

In simple terms, we are asking for the curves which are as straight as possible subject to the constraint that they lie within M.

Example. Suppose that

$$M = \mathbb{S}^n = \{(x^1, \ldots, x^{n+1}) \in \mathbb{E}^{n+1} : (x^1)^2 + \cdots + (x^{n+1})^2 = 1\}.$$

Let $\mathbf{e}_1$ and $\mathbf{e}_2$ be two unit-length vectors in $\mathbb{S}^n$ which are perpendicular to each other and define the unit-speed great circle $\gamma : [a, b] \to \mathbb{S}^n$ by

$$\gamma(t) = (\cos t)\mathbf{e}_1 + (\sin t)\mathbf{e}_2.$$

Then a direct calculation shows that

$$\gamma''(t) = -(\cos t)\mathbf{e}_1 - (\sin t)\mathbf{e}_2 = -\gamma(t).$$

Hence $(\gamma''(t))^\top = 0$ and γ is a geodesic on $\mathbb{S}^n$. We will see later that all unit speed geodesics on $\mathbb{S}^n$ are obtained in this manner.

1.3.2 Piecewise smooth paths

For technical reasons, it is often convenient to consider the more general space of piecewise smooth paths,

$$\hat{\Omega}_{[a,b]}(M; p, q) = \{ \text{ piecewise smooth paths } \gamma : [a, b] \to M : \gamma(a) = p, \gamma(b) = q\}.$$

an excellent exposition of which is given in Milnor [15], §11. By piecewise smooth, we mean γ is continuous and there exist $t_0 < t_1 < \cdots < t_N$ with $t_0 = a$ and $t_N = b$ such that $\gamma|[t_{i-1}, t_i]$ is smooth for $1 \leq i \leq N$. In this case, a *variation* of γ is a map

$$\bar{\alpha} : (-\epsilon, \epsilon) \to \hat{\Omega}_{[a,b]}(M; p, q)$$

such that $\bar{\alpha}(0) = \gamma$ and if

$$\alpha : (-\epsilon, \epsilon) \times [a, b] \to M \quad \text{is defined by} \quad \alpha(s, t) = \bar{\alpha}(s)(t),$$

then there exist $t_0 < t_1 < \cdots < t_N$ with $t_0 = a$ and $t_N = b$ such that

$$\alpha|(-\epsilon,\epsilon) \times [t_{i-1}, t_i]$$

is smooth for $1 \leq i \leq N$. As before, we find that

$$\frac{d}{ds}\left(J(\bar{\alpha}(s))\right)\Big|_{s=0} = \int_a^b \frac{\partial^2 \alpha}{\partial s \partial t}(0,t) \cdot \frac{\partial \alpha}{\partial t}(0,t) dt,$$

but now the integration by parts is more complicated because $\gamma'(t)$ is not continuous at $t_1, \ldots, t_{N-1}$. If we let

$$\gamma'(t_i-) = \lim_{t \to t_i -} \gamma'(t), \quad \gamma'(t_i+) = \lim_{t \to t_i +} \gamma'(t),$$

a short calculation shows that (1.6) yields the *first variation formula*

$$\frac{d}{ds}\left(J(\bar{\alpha}(s))\right)\Big|_{s=0} = -\int_a^b V(t) \cdot \gamma''(t) dt - \sum_{i=1}^{N-1} V(t_i) \cdot (\gamma'(t_i+) - \gamma'(t_i-)), \quad (1.6)$$

whenever $\bar{\alpha}$ is any variation of γ with variation field V. If

$$dJ(\gamma)(V) = \frac{d}{ds}\left(J(\bar{\alpha}(s))\right)\Big|_{s=0} = 0$$

for all variation fields V in the tangent space

$$T_\gamma \hat{\Omega} = \{ \text{ piecewise smooth maps } V : [a,b] \to \mathbb{E}^N \text{ such that } V(a) = 0 = V(b), V(t) \in T_{\gamma(t)} M \text{ for } t \in [a,b] \ \},$$

it follows that $\gamma'(t_i+) = \gamma'(t_i-)$ for every i and $(\gamma''(t))^\top = 0$. Thus critical points on the more general space of piecewise smooth paths are also smooth geodesics.

Exercise I. Due Friday, April 8. Suppose that M^2 is the right circular cylinder defined by the equation $x^2 + y^2 = 1$ in $\mathbb{E}^3$. Show that for each choice of real numbers a and b the curve

$$\gamma : \mathbb{R} \to M^2 \subseteq \mathbb{E}^3 \quad \text{defined by} \quad \gamma_{a,b}(t) = \begin{pmatrix} \cos(at) \\ \sin(at) \\ bt \end{pmatrix}$$

is a geodesic.

1.4 Hamilton's principle*

Of course, we would like a formula for geodesics that does not depend upon the existence of an isometric imbedding. To derive such a formula, it is convenient to regard the action J in a more general context, namely that of classical mechanics.

Definition. A *simple mechanical system* is a triple $(M, \langle\cdot,\cdot\rangle, \phi)$, where $(M, \langle\cdot,\cdot\rangle)$ is a Riemannian manifold and $\phi : M \to \mathbb{R}$ is a smooth function.

We call M the *configuration space* of the simple mechanical system. If $\gamma : [a, b] \to M$ represents the motion of the system,

$$\frac{1}{2}\langle\gamma'(t), \gamma'(t)\rangle = (\text{kinetic energy at time } t),$$
$$\phi(\gamma(t)) = (\text{potential energy at time } t).$$

Example 1: the Kepler problem. If a planet of mass m is moving around a star of mass M_0 with $M_0 >> m$, the star assumed to be stationary, we might take

$$M = \mathbb{R}^3 - \{(0,0,0)\},$$
$$\langle\cdot,\cdot\rangle = m(dx \otimes dx + dy \otimes dy + dz \otimes dz),$$
$$\phi(x, y, z) = \frac{-GmM_0}{\sqrt{x^2 + y^2 + z^2}}. \quad (1.7)$$

Here G is the gravitational constant. Sir Isaac Newton derived Kepler's three laws from this simple mechanical system, using his second law:

$$(\text{Force}) = (\text{mass})(\text{acceleration}), \quad \text{where} \quad (\text{Force}) = -\text{grad}(\phi). \quad (1.8)$$

The central symmetry of the problem suggests that we should use spherical coordinates. But that raises the question: How do we express Newton's second law in terms of general curvilinear coordinates? Needless to say, this system of equations can be solved exactly, thereby deriving Kepler's three laws of planetary motion. This was one of the major successes of Sir Isaac Newton's *Philosophiae naturalis principia mathematica* which appeared in 1687.

Example 2: the gravitational field of an oblate ellipsoid. The sun actually rotates once every 25 days, and some stars actually rotate much more rapidly. Thus the gravitational potential of a rapidly rotating star should differ somewhat from the simple formula given in Example 1. For the more general situation, we let $M = \mathbb{R}^3 - D$ where D is a closed compact domain with smooth boundary containing the origin,

$$\langle\cdot,\cdot\rangle = m(dx \otimes dx + dy \otimes dy + dz \otimes dz),$$

and $\phi : M \to \mathbb{R}$ is the "Newtonian gravitational potential" determined as follows: We imagine first that the mass (say ρ_0 were all concentrated at the point $(x_0, y_0, z_0) \in D$. Then just as in Example 1, we can set

$$\phi(x, y, z) = \frac{-Gm\rho_0}{\sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}}.$$

More generally, if $\rho(x,y,z)$ represents mass *density* at $(x,y,z) \in D$, the contribution to the potential $\phi(x,y,z)$ from a small coordinate cube located at $(\bar{x},\bar{y},\bar{z})$, with sides of length $d\bar{x}$, $d\bar{y}$, $d\bar{z}$ is given by

$$\frac{-Gm\rho(\bar{x},\bar{y},\bar{z})}{\sqrt{(x-\bar{x})^2+(y-\bar{y})^2+(z-\bar{z})^2}}d\bar{x}d\bar{y}d\bar{z}.$$

Summing over all the infinitesimal contributions to the potential, we finally obtain the desired formula for potential:

$$\phi(x,y,z) = \int\int\int_D \frac{-Gm\rho(\bar{x},\bar{y},\bar{z})}{\sqrt{(x-\bar{x})^2+(y-\bar{y})^2+(z-\bar{z})^2}}d\bar{x}d\bar{y}d\bar{z}. \tag{1.9}$$

Remarkably, if the mass distribution is spherically symmetric one gets exactly the same function as in (1.7) with M_0 being the integral of ρ over D. But in the case of an oblate ellipsoid, the potential is different. One can do the integration numerically and determine the numerical solutions to Newton's equations in this case, using software such as Mathematica, and one finds that the orbits are no longer exactly closed, but the perihelion of the planet precesses instead.

More generally, still, one could consider a nebula in space within the region D represented by a dust with mass density $\rho(x,y,z)$. It turns out that the "Newtonian gravitationl potential" $\phi(x,y,z)$ defined by (1.9) is the solution to Poisson's equation

$$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}+\frac{\partial^2\phi}{\partial z^2} = 4\pi Gm\rho \tag{1.10}$$

which goes to zero like $1/r$ as one approaches infinity, a fact which could be proven using the divergence theorem.

Example 3: the rigid body. To construct an interesting example in which the configuration space M is not Euclidean space, we take $M = SO(3)$, the group of real orthogonal 3×3 matrices of determinant one, regarded as the space of "configurations" of a rigid body B in $\mathbb{R}^3$ which has its center of mass located at the origin. We want to describe the motion of the rigid body as a path $\gamma : [a,b] \to M$. If p is a point in the rigid body with coordinates (x^1,x^2,x^3) at time $t=0$, we suppose that the coordinates of this point at time t will be

$$\gamma(t)\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix}, \qquad \text{where} \qquad \gamma(t) = \begin{pmatrix} a_{11}(t) & a_{12}(t) & a_{13}(t) \\ a_{21}(t) & a_{22}(t) & a_{23}(t) \\ a_{31}(t) & a_{32}(t) & a_{33}(t) \end{pmatrix} \in SO(3),$$

and $\gamma(0) = I$, the identity matrix. Then the velocity $v(t)$ of the particle p at time t will be

$$v(t) = \gamma'(t)\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^3 a'_{1i}(t)x^i \\ \sum_{i=1}^3 a'_{2i}(t)x^i \\ \sum_{i=1}^3 a'_{3i}(t)x^i \end{pmatrix},$$

and hence

$$v(t)\cdot v(t) = \sum_{i,j,k=1}^3 a'_{ki}(t)a'_{kj}(t)x^ix^j.$$

Suppose now that $\rho(x^1, x^2, x^3)$ is the density of matter at (x^1, x^2, x^3) within the rigid body. Then the total kinetic energy within the rigid body at time t will be given by the expression

$$K = \frac{1}{2} \sum_{i,j,k} \left(\int_B \rho(x^1, x^2, x^3) x^i x^j dx^1 dx^2 dx^3 \right) a'_{ki}(t) a'_{kj}(t).$$

We can rewrite this as

$$K = \frac{1}{2} \sum_{i,j,k} c_{ij} a'_{ki}(t) a'_{kj}(t), \qquad \text{where} \qquad c_{ij} = \int_B \rho(x^1, x^2, x^3) x^i x^j dx^1 dx^2 dx^3,$$

and define a Riemannian metric $\langle \cdot, \cdot \rangle$ on $M = SO(3)$ by

$$\langle \gamma'(t), \gamma'(t) \rangle = \sum_{i,j,k} c_{ij} a'_{ki}(t) a'_{kj}(t). \tag{1.11}$$

Then once again $(1/2)\langle \gamma'(t), \gamma'(t) \rangle$ represents the kinetic energy, this time of the rigid body B when its motion is represented by the curve $\gamma : (a, b) \to M$. We remark that the constants

$$I_{ij} = \text{Trace}(c_{ij})\delta_{ij} - c_{ij}$$

are called the *moments of inertia* of the rigid body.

A smooth function $\phi : SO(3) \to \mathbb{R}$ can represent the potential energy for the rigid body. In classical mechanics books, the motion of a top is described by means of a simple mechanical system which has configuration space $SO(3)$ with a suitable left-invariant metric and potential ϕ. Applied to the rotating earth, the same equations explain the precession of the equinoxes, according to which the axis of the earth traverses a circle in the celestial sphere once every 26,000 years. This means that astrologers will have to relearn their craft every few thousand years, because the sun will traverse a different path through the constellations, forcing a recalibration of astrological signs.

We want a formulation of Newton's second law (1.8) which helps to solve problems such as those we have just mentioned. In Lagrangian mechanics, the equations of motion for a simple mechanical system are derived from a variational principle. The key step is to define the *Lagrangian* to be the kinetic energy minus the potential energy. More precisely, for a simple mechanical system $(M, \langle \cdot, \cdot \rangle, \phi)$, we define the Lagrangian as a function on the tangent bundle of M,

$$\mathcal{L} : TM \to \mathbb{R} \quad \text{by} \quad \mathcal{L}(v) = \frac{1}{2}\langle v, v \rangle - \phi(\pi(v)),$$

where $\pi : TM \to M$ is the usual projection. We can then define the *action* $J : \Omega_{[a,b]}(M; p, q) \to \mathbb{R}$ by

$$J(\gamma) = \int_a^b \mathcal{L}(\gamma'(t)) dt.$$

As before, we say that $\gamma \in \Omega$ is a critical point for J if (1.3) holds. We can then give the Lagrangian formulation of Newton's law as follows:

Hamilton's principle. *If γ represents the motion of a simple mechanical system, then γ is a critical point for J.*

Thus the problem of finding curves in a Riemannian manifold from p to q of shortest length is put into the broader context of finding the trajectories for simple mechanical systems. Although we will focus on the geodesic problem, the earliest workers in the theory of geodesics must have been partially motivated by the fact that we were simply studying simple mechanical systems in which the potential function is zero.

We will soon see that if $\gamma \in \Omega_{[a,b]}(M; p, q)$ is a critical point for J and $[c, d] \subseteq [a, b]$, then the restriction of γ to $[c, d]$ is also a critical point for J, this time on the space $\Omega_{[c,d]}(M; r, s)$, where $r = \gamma(c)$ and $s = \gamma(d)$. Thus we can assume that $\gamma([a, b]) \subseteq U$, where $(U, x^1, \ldots, x^n)$ is a coordinate system on M, and we can express $\mathcal{L}$ in terms of the coordinates $(x^1, \ldots, x^n, \dot{x}^1, \ldots, \dot{x}^n)$ on $\pi^{-1}(U)$ described by (1.1). If

$$\gamma(t) = (x^1(t), \ldots, x^n(t)), \quad \text{and} \quad \gamma'(t) = (x^1(t), \ldots, x^n(t), \dot{x}^1(t), \ldots, \dot{x}^n(t)),$$

then

$$\mathcal{L}(\gamma'(t)) = \mathcal{L}(x^1(t), \ldots, x^n(t), \dot{x}^1(t), \ldots, \dot{x}^n(t)).$$

Euler-Lagrange Theorem. *A point $\gamma \in \Omega_{[a,b]}(M; p, q)$ is a critical point for the action $J \Leftrightarrow$ its coordinate functions satisfy the Euler-Lagrange equations*

$$\frac{\partial \mathcal{L}}{\partial x^i} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^i}\right) = 0. \tag{1.12}$$

Proof: We prove only the implication $\Rightarrow$, and leave the other half (which is quite a bit easier) as an exercise. We make the assumption that $\gamma([a, b]) \subseteq U$, where U is the domain of local coordinates as described above.

For $1 \leq i \leq n$, let $\psi^i : [a, b] \to \mathbb{R}$ be a smooth function such that $\psi^i(a) = 0 = \psi^i(b)$, and define a variation

$$\alpha : (-\epsilon, \epsilon) \times [a, b] \to U \quad \text{by} \quad \alpha(s, t) = (x^1(t) + s\psi^1(t), \ldots, x^n(t) + s\psi^n(t)).$$

Let $\dot{\psi}^i(t) = (d/dt)(\psi^i)(t)$. Then

$$J(\bar{\alpha}(s)) = \int_a^b \mathcal{L}(\cdots, x^i(t) + s\psi^i(t), \ldots, \dot{x}^i(t) + s\dot{\psi}^i(t), \ldots)dt,$$

so it follows from the chain rule that

$$\frac{d}{ds}\left(J(\bar{\alpha}(s))\right)\bigg|_{s=0} = \int_a^b \sum_{i=1}^n \left[\frac{\partial \mathcal{L}}{\partial x^i}(x^i(t), \dot{x}^i(t))\psi^i(t) + \frac{\partial \mathcal{L}}{\partial \dot{x}^i}(x^i(t), \dot{x}^i(t))\dot{\psi}^i(t)\right] dt.$$

Since $\psi^i(a) = 0 = \psi^i(b)$,

$$0 = \int_a^b \sum_{i=1}^n \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^i}\psi^i\right) dt = \int_a^b \sum_{i=1}^n \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^i}\right)\psi^i dt + \int_a^b \sum_{i=1}^n \frac{\partial \mathcal{L}}{\partial \dot{x}^i}\dot{\psi}^i dt,$$

and hence

$$\frac{d}{ds}\left(J(\bar{\alpha}(s))\right)\bigg|_{s=0} = \int_a^b \sum_{i=1}^n \left[\frac{\partial \mathcal{L}}{\partial x^i}\psi^i - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^i}\right)\psi^i\right] dt.$$

Thus if γ is a critical point for J, we must have

$$0 = \int_a^b \sum_{i=1}^n \left[\frac{\partial \mathcal{L}}{\partial x^i} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^i}\right)\right]\psi^i dt,$$

for every choice of smooth functions $\psi(t)$. In particular, if $\eta : [a,b] \to \mathbb{R}$ is a smooth function such that

$$\eta(a) = 0 = \eta(b), \quad \eta > 0 \quad \text{on} \quad (a,b),$$

and we set

$$\psi^i(t) = \eta(t)\left[\frac{\partial \mathcal{L}}{\partial x^i}(x^i(t), \dot{x}^i(t)) - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^i}(x^i(t), \dot{x}^i(t))\right)\right],$$

then

$$\int_a^b \eta(t)\left[\frac{\partial \mathcal{L}}{\partial x^i} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^i}\right)\right]^2 dt = 0.$$

Since the integrand is nonnegative, it must vanish identically and hence (1.12) must hold.

For a simple mechanical system, the Euler-Lagrange equations yield a derivation of Newton's second law of motion. Indeed, if

$$\langle \cdot, \cdot \rangle = \sum_{i,j=1}^n g_{ij} dx^i dx^j,$$

then in the standard coordinates $(x^1, \ldots, x^n, \dot{x}^1, \ldots, \dot{x}^n)$,

$$\mathcal{L}(\gamma'(t)) = \frac{1}{2}\sum_{i,j=1}^n g_{ij}(x^1, \ldots, x^n)\dot{x}^i\dot{x}^j - \phi(x^1, \ldots x^n).$$

Hence

$$\frac{\partial \mathcal{L}}{\partial x^i} = \frac{1}{2}\sum_{j,k=1}^n \frac{\partial g_{jk}}{\partial x^i}\dot{x}^j\dot{x}^k - \frac{\partial \phi}{\partial x^i},$$

$$\frac{\partial \mathcal{L}}{\partial \dot{x}^i} = \sum_{j=1}^n g_{ij}\dot{x}^j, \qquad \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^i}\right) = \sum_{j,k=1}^n \frac{\partial g_{ij}}{\partial x^k}\dot{x}^j\dot{x}^k + \sum_{j=1}^n g_{ij}\ddot{x}^j,$$

where $\ddot{x}^j = d^2x^j/dt^2$. Thus the Euler-Lagrange equations become

$$\sum_{j=1}^{n} g_{ij}\ddot{x}^j + \sum_{j,k=1}^{n} \frac{\partial g_{ij}}{\partial x^k}\dot{x}^j\dot{x}^k = \frac{1}{2}\sum_{j,k=1}^{n} \frac{\partial g_{jk}}{\partial x^i}\dot{x}^j\dot{x}^k - \frac{\partial \phi}{\partial x^i}$$

or

$$\sum_{j=1}^{n} g_{ij}\ddot{x}^j + \frac{1}{2}\sum_{j,k=1}^{n} \left(\frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^i}\right)\dot{x}^j\dot{x}^k = -\frac{\partial \phi}{\partial x^i}.$$

We multiply through by the matrix (g^{ij}) which is inverse to (g_{ij}) to obtain

$$\ddot{x}^l + \sum_{j,k=1}^{n} \Gamma^l_{jk}\dot{x}^j\dot{x}^k = -\sum_{i=1}^{n} g^{li}\frac{\partial \phi}{\partial x^i}, \tag{1.13}$$

where

$$\Gamma^l_{ij} = \frac{1}{2}\sum_{k=1}^{n} g^{lk}\left(\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}\right). \tag{1.14}$$

The expressions Γ^l_{ij} are called the *Christoffel symbols*. Note that if $(x^1, \ldots, x^n)$ are rectangular cartesian coordinates in Euclidean space, the Christoffel symbols vanish.

We can interpret the two sides of (1.13) as follows:

$$\ddot{x}^l + \sum_{j,k=1}^{n} \Gamma^l_{jk}\dot{x}^j\dot{x}^k = (\text{acceleration})^l,$$

$$-\sum_{i=1}^{n} g^{li}\frac{\partial \phi}{\partial x^i} = (\text{force per unit mass})^l.$$

Hence equation (1.13) is just the statement of Newton's second law, force equals mass times acceleration, for simple mechanical systems.

In the case where $\phi = 0$, we obtain differential equations for the geodesics on M,

$$\ddot{x}^i + \sum_{j,k=1}^{n} \Gamma^i_{jk}\dot{x}^j\dot{x}^k = 0, \tag{1.15}$$

where the Γ^i_{jk}'s are the Christoffel symbols.

1.5 The Levi-Civita connection

In modern differential geometry, the Christoffel symbols Γ^k_{ij} are regarded as the components of a connection. We now describe how that goes.

You may recall from Math 240A that a *smooth vector field* on the manifold M is a smooth map

$$X : M \to TM \quad \text{such that} \quad \pi \circ X = \text{id}_M,$$

where $\pi : TM \to M$ is the usual projection, or equivalently a smooth map

$$X : M \to TM \quad \text{such that} \quad X(p) \in T_pM.$$

The restriction of a vector field to the domain U of a smooth coordinate system $(x^1, \ldots, x^n)$ can be written as

$$X|U = \sum_{i=1}^{n} f^i \frac{\partial}{\partial x^i}, \quad \text{where} \quad f^i : U \to \mathbb{R}.$$

If we evaluate at a given point $p \in U$ this specializes to

$$X(p) = \sum_{i=1}^{n} f^i(p) \left. \frac{\partial}{\partial x^i} \right|_p .$$

A vector field X can be regarded as a first-order differential operator. Thus, if $g : M \to \mathbb{R}$ is a smooth function, we can operate on g by X, thereby obtaining a new smooth function $Xg : M \to \mathbb{R}$ by $(Xg)(p) = X(p)(g)$.

We let $\mathcal{X}(M)$ denote the space of all smooth vector fields on M. It can be regarded as a real vector space or as an $\mathcal{F}(M)$-module, where $\mathcal{F}(M)$ is the space of all smooth real-valued functions on M, where the multiplication $\mathcal{F}(M) \times \mathcal{X}(M) \to \mathcal{X}(M)$ is defined by $(fX)(p) = f(p)X(p)$.

Definition. A *connection* on the tangent bundle TM is an operator

$$\nabla : \mathcal{X}(M) \times \mathcal{X}(M) \longrightarrow \mathcal{X}(M)$$

that satisfies the following axioms (where we write $\nabla_X Y$ for $\nabla(X, Y)$):

$$\nabla_{fX+gY} Z = f \nabla_X Z + g \nabla_Y Z, \tag{1.16}$$

$$\nabla_Z (fX + gY) = (Zf)X + f \nabla_Z X + (Zg)Y + g \nabla_Z Y, \tag{1.17}$$

for $f, g \in \mathcal{F}(M)$ and $X, Y, Z \in \mathcal{X}(M)$.

Note that (1.17) is the usual "Leibniz rule" for differentiation. We often call $\nabla_X Y$ the *covariant derivative* of Y in the direction of X.

Lemma 1. *Any connection ∇ is local; that is, if U is an open subset of M,*

$$X|U \equiv 0 \quad \Rightarrow \quad (\nabla_X Y)|U \equiv 0 \quad \textit{and} \quad (\nabla_Y X)|U \equiv 0,$$

for any $Y \in \mathcal{X}(M)$.

Proof: Let p be a point of U and choose a smooth function $f : M \to \mathbb{R}$ such that $f \equiv 0$ on a neighborhood of p and $f \equiv 1$ outside U. Then

$$X|U \equiv 0 \Rightarrow fX \equiv X.$$

Hence

$$(\nabla_X Y)(p) = \nabla_{fX} Y(p) = f(p)(\nabla_X Y)(p) = 0,$$
$$(\nabla_Y X)(p) = \nabla_Y (fX)(p) = f(p)(\nabla_Y X)(p) + (Yf)(p)X(p) = 0.$$

Since p was an arbitrary point of U, we conclude that $(\nabla_X Y)|U \equiv 0$ and $(\nabla_Y X)|U \equiv 0$.

This lemma implies that if U is an arbitrary open subset of M a connection ∇ on TM will restrict to a unique well-defined connection ∇ on TU.

Thus we can restrict to the domain U of a local coordinate system $(x^1, \ldots, x^n)$ and define the components $\Gamma^k_{ij} : U \to \mathbb{R}$ of the connection by

$$\nabla_{\partial/\partial x^i} \frac{\partial}{\partial x^j} = \sum_{k=1}^n \Gamma^k_{ij} \frac{\partial}{\partial x^k}.$$

Then if X and Y are smooth vector fields on U, say

$$X = \sum_{i=1}^n f^i \frac{\partial}{\partial x^i}, \quad Y = \sum_{j=1}^n g^j \frac{\partial}{\partial x^j},$$

we can use the connection axioms and the components of the connection to calculate $\nabla_X Y$:

$$\nabla_X Y = \sum_{i=1}^n \left[\sum_{j=1}^n f^j \frac{\partial g^i}{\partial x^j} + \sum_{j,k=1}^n \Gamma^i_{jk} f^j g^k \right] \frac{\partial}{\partial x^i}. \tag{1.18}$$

Lemma 2. *$(\nabla_X Y)(p)$ depends only on $X(p)$ and on the values of Y along some curve tangent to $X(p)$.*

Proof: This follows immediately from (1.18).

Because of the previous lemma, we can $\nabla_v X \in T_p M$, whenever $v \in T_p M$ and X is a vector field defined along some curve tangent to v at p, by setting

$$\nabla_v X = (\nabla_{\tilde{V}} \tilde{X})(p),$$

for any choice of extensions $\tilde{V}$ of v and $\tilde{X}$ of X. In particular, if $\gamma : [a, b] \to M$ is a smooth curve, we can define the vector field $\nabla_{\gamma'} \gamma'$ along γ. Recall that we define the *Lie bracket* of two vector fields X and Y by $[X, Y](f) = X(Y(f)) - Y(X(f))$. If X and Y are smooth vector fields on the domain U of local corrdinates $(x^1, \ldots, x^n)$, say

$$X = \sum_{i=1}^n f^i \frac{\partial}{\partial x^i}, \quad Y = \sum_{j=1}^n g^j \frac{\partial}{\partial x^j},$$

then

$$[X,Y] = \sum_{i,j=1}^{n} f^i \frac{\partial g^j}{\partial x^i}\frac{\partial}{\partial x^j} - \sum_{i,j=1}^{n} g^j \frac{\partial f^i}{\partial x^j}\frac{\partial}{\partial x^i}.$$

Fundamental Theorem of Riemannian Geometry. *If $(M, \langle\cdot,\cdot\rangle)$ is a Riemannian manifold, there is a unique connection on TM such that*

1. *∇ is symmetric, that is, $\nabla_X Y - \nabla_Y X = [X,Y]$, for $X, Y \in \mathcal{X}(M)$,*
2. *∇ is metric, that is, $X\langle Y,Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$, for $X, Y, Z \in \mathcal{X}(M)$.*

This connection is called the *Levi-Civita connection* of the Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$.

To prove the theorem we express the two conditions in terms of local coordinates $(x^1, \ldots, x^n)$ defined on an open subset U of M. In terms of the components of ∇, defined by the formula

$$\nabla_{\partial/\partial x^i}\frac{\partial}{\partial x^j} = \sum_{k=1}^{n} \Gamma^k_{ij}\frac{\partial}{\partial x^k}, \tag{1.19}$$

the first condition becomes

$$\Gamma^k_{ij} = \Gamma^k_{ji}, \quad \text{since} \quad \left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] = 0.$$

Thus the Γ^k_{ij}'s are symmetric in the lower pair of indices. If we write

$$\langle\cdot,\cdot\rangle = \sum_{i,j=1}^{n} g_{ij}dx^i \otimes dx^j,$$

then the second condition yields

$$\begin{aligned}\frac{\partial g_{ij}}{\partial x^k} &= \frac{\partial}{\partial x^k}\left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right\rangle = \left\langle \nabla_{\partial/\partial x^k}\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right\rangle + \left\langle \frac{\partial}{\partial x^i}, \nabla_{\partial/\partial x^k}\frac{\partial}{\partial x^j}\right\rangle \\ &= \left\langle \sum_{l=1}^{n}\Gamma^l_{ki}\frac{\partial}{\partial x^l}, \frac{\partial}{\partial x^j}\right\rangle + \left\langle \frac{\partial}{\partial x^i}, \sum_{l=1}^{n}\Gamma^l_{kj}\frac{\partial}{\partial x^l}\right\rangle = \sum_{l=1}^{n} g_{lj}\Gamma^l_{ki} + \sum_{l=1}^{n} g_{il}\Gamma^l_{kj}.\end{aligned}$$

In fact, the second condition is equivalent to

$$\frac{\partial g_{ij}}{\partial x^k} = \sum_{l=1}^{n} g_{lj}\Gamma^l_{ki} + \sum_{l=1}^{n} g_{il}\Gamma^l_{kj}. \tag{1.20}$$

We can permute the indices in (1.20), obtaining

$$\frac{\partial g_{jk}}{\partial x^i} = \sum_{l=1}^{n} g_{lk}\Gamma^l_{ij} + \sum_{l=1}^{n} g_{jl}\Gamma^l_{ik}. \tag{1.21}$$

and

$$\frac{\partial g_{ki}}{\partial x^j} = \sum_{l=1}^{n} g_{li}\Gamma^l_{jk} + \sum_{l=1}^{n} g_{kl}\Gamma^l_{ji}. \tag{1.22}$$

Subtracting (1.20) from the sum of (1.21) and (1.22), and using the symmetry of Γ^l_{ij} in the lower indices, yields

$$\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k} = 2\sum_{l=1}^{n} g_{lk}\Gamma^l_{ij}.$$

Thus if we let (g^{ij}) denote the matrix inverse to $g_{ij})$, we find that

$$\Gamma^l_{ij} = \frac{1}{2}\sum_{k=1}^{n} g^{lk}\left(\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}\right), \tag{1.23}$$

which is exactly the formula (1.14) we obtained before by means of Hamilton's principle.

This proves uniqueness of the connection which is both symmetric and metric. For existence, we define the connection locally by (1.19), where the Γ^l_{ij}'s are defined by (1.23) and check that the resulting connection is both symmetric and metric. (Note that by uniqueness, the locally defined connections fit together on overlaps.)

In the special case where the Riemannian manifolds is Euclidean space $\mathbb{E}^N$ the Levi-Civita connection is easy to describe. In this case, we have global rectangular cartesian coordinates $(x^1, \ldots, x^N)$ on $\mathbb{E}^N$ and any vector field Y on $\mathbb{E}^N$ can be written as

$$Y = \sum_{i=1}^{N} f^i \frac{\partial}{\partial x^i}, \quad \text{where} \quad f^i : \mathbb{E}^N \to \mathbb{R}.$$

In this case, the Levi-Civita connection ∇^E has components $\Gamma^k_{ij} = 0$, and therefore the operator ∇^E satisfies the formula

$$\nabla^E_X Y = \sum_{i=1}^{N} (Xf^i)\frac{\partial}{\partial x^i}.$$

It is easy to check that this connection which is symmetric and metric for the Euclidean metric.

If M is an imbedded submanifold of $\mathbb{E}^N$ with the induced metric, then one can define a connection $\nabla : \mathcal{X}(M) \times \mathcal{X}(M) \to \mathcal{X}(M)$ by

$$(\nabla_X Y)(p) = (\nabla^E_X Y(p))^\top,$$

where $(\cdot)^\top$ is the orthogonal projection into the tangent space. (Use Lemma 5.2 to justify this formula.) It is a straightforward exercise to show that ∇ is symmetric and metric for the induced connection, and hence ∇ is the Levi-Civita connection for M. Note that if $\gamma : [a,b] \to M$ is a smooth curve, then

$$(\nabla_{\gamma'}\gamma')(t) = (\gamma''(t))^{\top},$$

so a smooth curve in $M \subseteq \mathbb{E}^N$ is a geodesic if and only if $\nabla_{\gamma'}\gamma' \equiv 0$. If we want to develop the subject independent of Nash's imbedding theorem, we can make the

Definition. If $(M, \langle\cdot,\cdot\rangle)$ is a Riemannian manifold, a smooth path $\gamma : [a,b] \to M$ is a *geodesic* if it satisfies the equation $\nabla_{\gamma'}\gamma' \equiv 0$, where ∇ is the Levi-Civita connection.

In terms of local coordinates, if

$$\gamma' = \sum_{i=1}^{n} \frac{d(x^i \circ \gamma)}{dt} \frac{\partial}{\partial x^i},$$

then a straightforward calculation yields

$$\nabla_{\gamma'}\gamma' = \sum_{i=1}^{n} \left[\frac{d^2(x^i \circ \gamma)}{dt^2} + \sum_{j,k=1}^{n} \Gamma^i_{jk} \frac{d(x^j \circ \gamma)}{dt} \frac{d(x^k \circ \gamma)}{dt} \right] \frac{\partial}{\partial x^i}. \tag{1.24}$$

This reduces to the equation (1.15) we obtained before from Hamilton's principle. Note that

$$\frac{d}{dt}\langle\gamma',\gamma'\rangle = \gamma'\langle\gamma',\gamma'\rangle = 2\langle\nabla_{\gamma'}\gamma',\gamma'\rangle = 0,$$

so geodesics automatically have constant speed.

More generally, if $\gamma : [a,b] \to M$ is a smooth curve, we call $\nabla_{\gamma'}\gamma'$ the *acceleration* of γ. Thus if $(M, \langle\cdot,\cdot\rangle, \phi)$ is a simple mechanical system, its equations of motion can be written as

$$\nabla_{\gamma'}\gamma' = -\mathrm{grad}(\phi), \tag{1.25}$$

where in terms of local coordinates $(x^1, \ldots, x^n)$ on $U \subseteq M$,

$$\mathrm{grad}(\phi) = \sum g^{ji}(\partial V/\partial x^i)(\partial/\partial x^j).$$

Note that these equations of motion can be written as follows:

$$\begin{cases} \frac{dx^l}{dt} = \dot{x}^l, \\ \frac{d\dot{x}^l}{dt} = -\sum_{j,k=1}^{n} \Gamma^l_{jk}\dot{x}^j\dot{x}^k - \sum_{i=1}^{n} g^{li}\frac{\partial\phi}{\partial x^i}. \end{cases} \tag{1.26}$$

In terms of the local coordinates $(x^1, \ldots, x^n, \dot{x}^1, \ldots, \dot{x}^n)$ on $\pi^{-1}(U) \subseteq TM$, this system of differential equations correspond to the vector field

$$X = \sum_{i=1}^{n} \dot{x}^i \frac{\partial}{\partial x^i} - \sum_{i=1}^{n} \left(\sum_{j,k=1}^{n} \Gamma^i_{jk}\dot{x}^j\dot{x}^k + \sum_{i=1}^{n} g^{ij}\frac{\partial\phi}{\partial x^j} \right) \frac{\partial}{\partial \dot{x}^i}.$$

It follows from the fundamental existence and uniqueness theorem from the theory of ordinary differential equations ([3], Chapter IV, §4) that given an element $v \in T_pM$, there is a unique solution to this system, defined for $t \in (-\epsilon, \epsilon)$ for some $\epsilon > 0$, which satisfies the initial conditions

$$x^i(0) = x^i(p), \qquad \dot{x}^i(0) = \dot{x}^i(v).$$

In the special case where $\phi = 0$, we can restate this as:

Existence and Uniqueness Theorem for Geodesics. *Given $p \in M$ and $v \in T_pM$, there is a unique geodesic $\gamma : (-\epsilon, \epsilon) \to M$ for some $\epsilon > 0$ such that $\gamma(0) = p$ and $\gamma'(0) = v$.*

Simplest Example. In the case of Euclidean space $\mathbb{E}^n$ with the standard Euclidean metric, $g_{ij} = \delta_{ij}$, the Christoffel symbols vanish, $\Gamma^l_{jk} = 0$, and the equations for geodesics become

$$\frac{d^2x^i}{dt^2} = 0.$$

The solutions are

$$x^i = a^i t + b^i,$$

the straight lines parametrized with constant speed.

Exercise II. Due Friday, April 15. Consider the upper half-plane $\mathbb{H}^2 = \{(x, y) \in \mathbb{R}^2 : y > 0\}$, with Riemannian metric

$$\langle \cdot, \cdot \rangle = \frac{1}{y^2}(dx \otimes dx + dy \otimes dy),$$

the so-called Poincaré upper half plane.

a. Calculate the Christoffel symbols Γ^k_{ij}.

b. Write down the equations for the geodesics, obtaining two equations

$$\frac{d^2x}{dt^2} = \cdots, \quad \frac{d^2y}{dt^2} = \cdots.$$

c. Show that the vertical half-lines $x = c$ are the images of geodesics and find their constant speed parametrizations.

1.6 First variation of J: intrinsic version

Now that we have the notion of connection available, it might be helpful to review the argument that the function

$$J : \Omega_{[a,b]}(M; p, q), \to \mathbb{R} \quad \text{defined by} \quad J(\gamma) = \frac{1}{2}\int_a^b \langle \gamma'(t), \gamma'(t) \rangle dt,$$

has geodesics as its critical points, and recast the argument in a form that is independent of choice of isometric imbedding.

In fact, the argument we gave before goes through with only one minor change, namely given a variation

$$\bar{\alpha} : (-\epsilon, \epsilon) \to \Omega \quad \text{with corresponding} \quad \alpha : (-\epsilon, \epsilon) \times [a, b] \to M,$$

we must make sense of the partial derivatives

$$\frac{\partial \alpha}{\partial s}, \quad \frac{\partial \alpha}{\partial t}, \quad \ldots,$$

since we can no longer regard α as a vector valued function.

But these is a simple remedy. We look at the first partial derivatives as maps

$$\frac{\partial \alpha}{\partial s}, \frac{\partial \alpha}{\partial t} : (-\epsilon, \epsilon) \times [a, b] \to TM \qquad \text{such that} \quad \pi \circ \left(\frac{\partial \alpha}{\partial s} \right) = \alpha, \quad \pi \circ \left(\frac{\partial \alpha}{\partial t} \right) = \alpha.$$

In terms of local coordinates, these maps are defined by

$$\left(\frac{\partial \alpha}{\partial s} \right)(s, t) = \sum_{i=1}^{n} \frac{\partial (x^i \circ \alpha)}{\partial s}(s, t) \left. \frac{\partial}{\partial x^i} \right|_{\alpha(s,t)},$$

$$\left(\frac{\partial \alpha}{\partial t} \right)(s, t) = \sum_{i=1}^{n} \frac{\partial (x^i \circ \alpha)}{\partial t}(s, t) \left. \frac{\partial}{\partial x^i} \right|_{\alpha(s,t)}.$$

We define higher order derivatives via the Levi-Civita connection. Thus for example, in terms of local coordinates, we set

$$\nabla_{\partial/\partial s} \left(\frac{\partial \alpha}{\partial t} \right) = \sum_{k=1}^{n} \left[\frac{\partial^2 (x^k \circ \alpha)}{\partial s \partial t} + \sum_{i,j=1}^{n} (\Gamma^k_{ij} \circ \alpha) \frac{\partial (x^i \circ \alpha)}{\partial s} \frac{\partial (x^i \circ \alpha)}{\partial t} \right] \left. \frac{\partial}{\partial x^k} \right|_{\alpha},$$

thereby obtaining a map

$$\nabla_{\partial/\partial s} \left(\frac{\partial \alpha}{\partial t} \right) : (-\epsilon, \epsilon) \times [a, b] \to TM \quad \text{such that} \quad \pi \circ \nabla_{\partial/\partial s} \left(\frac{\partial \alpha}{\partial t} \right) = \alpha.$$

Similarly, we define

$$\nabla_{\partial/\partial t} \left(\frac{\partial \alpha}{\partial t} \right), \qquad \nabla_{\partial/\partial t} \left(\frac{\partial \alpha}{\partial s} \right),$$

and so forth. In short, we replace higher order derivatives by covariant derivatives using the Levi-Civita connection for the Riemmannian metric.

The properties of the Levi-Civita connection imply that

$$\nabla_{\partial/\partial s} \left(\frac{\partial \alpha}{\partial t} \right) = \nabla_{\partial/\partial t} \left(\frac{\partial \alpha}{\partial s} \right)$$

and

$$\frac{\partial}{\partial t}\left\langle \frac{\partial \alpha}{\partial s}, \frac{\partial \alpha}{\partial t}\right\rangle = \left\langle \nabla_{\partial/\partial t}\left(\frac{\partial \alpha}{\partial s}\right), \frac{\partial \alpha}{\partial t}\right\rangle + \left\langle \frac{\partial \alpha}{\partial s}, \nabla_{\partial/\partial t}\left(\frac{\partial \alpha}{\partial t}\right)\right\rangle.$$

With these preparations out of the way, we can now proceed as before and let

$$\bar{\alpha} : (-\epsilon, \epsilon) \to \Omega_{[a,b]}(M; p, q)$$

be a smooth path with $\bar{\alpha}(0) = \gamma$ and

$$\frac{\partial \alpha}{\partial s}(0, t) = V(t),$$

where V is an element of the tangent space

$$T_\gamma \Omega = \{ \text{ smooth maps } V : [a, b] \to TM \\ \text{such that } \pi \circ V(t) = \gamma(t) \text{ for } t \in [a, b], \text{ and } V(a) = 0 = V(b) \ \}.$$

Then just as before,

$$\begin{aligned} \frac{d}{ds}\left(J(\bar{\alpha}(s))\right)\Big|_{s=0} &= \frac{d}{ds}\left[\frac{1}{2}\int_a^b \left\langle \frac{\partial \alpha}{\partial t}(s,t), \frac{\partial \alpha}{\partial t}(s,t) dt\right\rangle\right]\Big|_{s=0} \\ &= \int_a^b \left\langle \nabla_{\partial/\partial s}\left(\frac{\partial \alpha}{\partial t}\right)(0,t), \frac{\partial \alpha}{\partial t}(0,t)\right\rangle dt \\ &= \int_a^b \left\langle \nabla_{\partial/\partial t}\left(\frac{\partial \alpha}{\partial s}\right)(0,t), \frac{\partial \alpha}{\partial t}(0,t)\right\rangle dt \\ = \int_a^b \left[\frac{\partial}{\partial t}\left\langle \frac{\partial \alpha}{\partial s}(0,t), \frac{\partial \alpha}{\partial t}(0,t)\right\rangle \right. &- \left.\left\langle \frac{\partial \alpha}{\partial s}(0,t), \nabla_{\partial/\partial t}\frac{\partial \alpha}{\partial t}(0,t)\right\rangle\right] dt. \end{aligned}$$

Since

$$\frac{\partial \alpha}{\partial s}(0, b) = 0 = \frac{\partial \alpha}{\partial s}(0, a),$$

we obtain

$$\frac{d}{ds}\left(J(\bar{\alpha}(s))\right)\Big|_{s=0} = -\int_a^b \langle V(t), (\nabla_{\gamma'}\gamma')(t)\rangle dt.$$

We call this the *first variation* of J in the direction of V, and write

$$dJ(\gamma)(V) = -\int_a^b \langle V(t), (\nabla_{\gamma'}\gamma')(t)\rangle dt. \tag{1.27}$$

A critical point for J is a point $\gamma \in \Omega_{[a,b]}(M; p, q)$ at which $dJ(\gamma) = 0$, and the above argument shows that the critical points for J are exactly the geodesics for the Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$.

Of course, we could modify the above derivation to determine the first variation of the action

$$J(\gamma) = \frac{1}{2}\int_a^b \langle \gamma'(t), \gamma'(t)\rangle dt - \int_a^b \phi(\gamma(t)) dt$$

for a simple mechanical system $(M, \langle\cdot,\cdot\rangle, \phi)$ such as those considered in §**??**. We would find after a short calculation that

$$dJ(\gamma)(V) = -\int_a^b \langle V(t), (\nabla_{\gamma'}\gamma')(t)\rangle dt - \int_a^b d\phi(V)(\gamma(t)dt$$
$$= -\int_a^b \langle V(t), (\nabla_{\gamma'}\gamma')(t) - \text{grad}(\phi)(\gamma(t))\rangle \, dt.$$

Just as in §1.4, the critical points are solutions to Newton's equation (1.25).

1.7 Lorentz manifolds

The notion of Riemmannian manifold has a generalization which is extremely useful in Einstein's theory of relativity, both special and general, as described for example is the standard texts [17] or [22].

Definition. Let M be a smooth manifold. A *pseudo-Riemannian metric* on M is a function which assigns to each $p \in M$ a nondegenerate symmetric bilinear map

$$\langle\cdot,\cdot\rangle_p : T_pM \times T_pM \longrightarrow \mathbb{R}$$

which which varies smoothly with $p \in M$. As before , varying smoothly with $p \in M$ means that if $\phi = (x^1, \ldots, x^n) : U \to \mathbb{R}^n$ is a smooth coordinate system on M, then for $p \in U$,

$$\langle\cdot,\cdot\rangle_p = \sum_{i,j=1}^n g_{ij}(p) dx^i|_p \otimes dx^j|_p,$$

where the functions $g_{ij} : U \to \mathbb{R}$ are smooth. The conditions that $\langle\cdot,\cdot\rangle_p$ be symmetric and nondegenerate are expressed in terms of the matrix (g_{ij}) by saying that (g_{ij}) is a symmetric matrix and has nonzero determinant.

It follows from linear algebra that for any choice of $p \in M$, local coordinates $(x^1, \ldots, x^n)$ can be chosen so that

$$(g_{ij}(p)) = \begin{pmatrix} -I_{p\times p} & 0 \\ 0 & I_{q\times q} \end{pmatrix},$$

where $I_{p\times p}$ and $I_{q\times q}$ are $p \times p$ and $q \times q$ identity matrices with $p + q = n$. The pair (p, q) is called the *signature* of the pseudo-Riemannian metric.

Note that a pseudo-Riemannian metric of signature $(0, n)$ is just a Riemannian metric. A pseudo-Riemannian metric of signature $(1, n-1)$ is called a *Lorentz metric*.

A *pseudo-Riemannian manifold* is a pair $(M, \langle\cdot,\cdot\rangle)$ where M is a smooth manifold and $\langle\cdot,\cdot\rangle$ is a pseudo-Riemannian metric on M. Similarly, a *Lorentz manifold* is a pair $(M, \langle\cdot,\cdot\rangle)$ where M is a smooth manifold and $\langle\cdot,\cdot\rangle$ is a Lorentz metric on M.

Example. Let $\mathbb{R}^{n+1}$ be given coordinates $(t, x^1, \ldots, x^n)$, with t being regarded as time and $(x^1, \ldots, x^n)$ being regarded as Euclidean coordinates in space, and consider the Lorentz metric

$$\langle \cdot, \cdot \rangle = -c^2 dt \otimes dt + \sum_{i=1}^{n} dx^i \otimes dx^i, \tag{1.28}$$

where the constant c is regarded as the speed of light. When endowed with this metric, $\mathbb{R}^{n+1}$ is called *Minkowski space-time* and is denoted by $\mathbb{L}^{n+1}$. Four-dimensional Minkowski space-time is the arena for *special relativity*.

The arena for *general relativity* is a more general four-dimensional Lorentz manifold $(M, \langle \cdot, \cdot \rangle)$, also called space-time. In the case of general relativity, the components g_{ij} of the metric are regarded as potentials for the gravitational forces.

In either case, points of space-time can be thought of as *events* that happen at a given place in space and at a given time. The trajectory of a moving particle can be regarded as curve of events, called its *world line*.

If p is an event in a Lorentz manifold $(M, \langle \cdot, \cdot \rangle)$, the tangent space T_pM inherits a Lorentz inner product

$$\langle \cdot, \cdot \rangle_p : T_pM \times T_pM \longrightarrow \mathbb{R}.$$

We say that an element $v \in T_pM$ is

1. *timelike* if $\langle v, v \rangle < 0$,
2. *spacelike* if $\langle v, v \rangle > 0$, and
3. *lightlike* if $\langle v, v \rangle = 0$.

A parametrized curve $\gamma : [a, b] \to M$ into a Lorentz manifold $(M, \langle \cdot, \cdot \rangle)$ is said to be timelike if $\gamma'(u)$ is timelike for all $u \in [a, b]$. If a parametrized curve $\gamma : [a, b] \to M$ represents the world line of a massive object, it is timelike and the integral

$$L(\gamma) = \frac{1}{c} \int_a^b \sqrt{-\langle \gamma'(u) \gamma'(u) \rangle} du \tag{1.29}$$

is the *elapsed time* measured by a clock moving along the world line γ. We call $L(\gamma)$ the *proper time* of γ.

The Twin Paradox. The fact that elapsed time is measured by the integral (1.29) has counterintuitive consequences. Suppose that $\gamma : [a, b] \to \mathbb{L}^4$ is a timelike curve in four-dimensional Minkowski space-time, parametrized so that

$$\gamma(t) = (t, x^1(t), x^2(t), x^3(t)).$$

Then

$$\gamma'(t) = \frac{\partial}{\partial t} + \sum_{i=1}^{3} \frac{dx^i}{dt} \frac{\partial}{\partial x^i}, \quad \text{so} \quad \langle \gamma'(t), \gamma'(t) \rangle = -c^2 + \sum_{i=1}^{3} \left(\frac{dx^i}{dt} \right)^2,$$

and hence

$$L(\gamma) = \int_a^b \frac{1}{c}\sqrt{c^2 - \sum_{i=1}^3 \left(\frac{dx^i}{dt}\right)^2} dt = \int_a^b \sqrt{1 - \frac{1}{c^2}\sum_{i=1}^3 \left(\frac{dx^i}{dt}\right)^2} dt. \qquad (1.30)$$

Thus if a clock is at rest with respect to the coordinates, that is $dx^i/dt \equiv 0$, it will measure the time interval $b - a$, while if it is in motion it will measure a somewhat shorter time interval. This failure of clocks to synchronize is what is called the twin paradox.

Equation (1.30) states that in Minkowski space-time, straight lines maximize L among all timelike world lines from an event p to an event q. When given an affine parametrization such curves have zero acceleration.

In general relativity, Minkowki spce-time is replaced by a more general Lorentz manifold. The world line of a massive body, not subject to any forces other than gravity, will also maximize L, and if it is appropriately parametrized, it will have zero acceleration in terms of the Lorentz metric $\langle \cdot, \cdot \rangle$. Just as in the Riemannian case, it is easier to describe the critical point behavior of the closely related *action*

$$J : \Omega_{[a,b]}(M : p, q) \to \mathbb{R}, \quad \text{defined by} \quad J(\gamma) = \frac{1}{2}\int_a^b \langle \gamma'(t), \gamma'(t) \rangle dt.$$

The critical points of J for a Lorentz manifold $(M, \langle \cdot, \cdot \rangle)$ are called its *geodesics*.

How does one determine the geodesics in a Lorentz manifold? Fortunately, the fundamental theorem of Riemannian geometry generalizes immediately to pseudo-Riemannian metrics;

Fundamental Theorem of pseudo-Riemannian Geometry. *If $\langle \cdot, \cdot \rangle$ is a pesudo-Riemannian metric on a smooth manifold M, there is a unique connection on TM such that*

1. *∇ is symmetric, that is, $\nabla_X Y - \nabla_Y X = [X, Y]$, for $X, Y \in \mathcal{X}(M)$,*
2. *∇ is metric, that is, $X\langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle$, for $X, Y, Z \in \mathcal{X}(M)$.*

The proof is identical to the proof we gave before. Moreover, just as before, we can define the Christoffel symbols for local coordinates, and they are given by exactly the same formula (1.23). Finally, by the first variation formula, one shows that a smooth parametrized curve $\gamma : [a, b] \to M$ is a geodesic if and only if it satisfies the equation $\nabla_{\gamma'}\gamma' \equiv 0$.

We can now summarize the main ideas of general relativity, that are treated in much more detail in [17] or [22]. General relativity is a theory of the gravitational force, and there are two main components:

1. The matter and energy in space-time tells space-time how to curve in accordance with the Einstein field equations. These Einstein field equations are described in terms of the curvature of the Lorentz manifold and determine the Lorentz metric. (We will describe curvature in the following sections.)

2. Timelike geodesics are exactly the world lines of massive objects which are subjected to no forces other than gravity, while lightlike geodesics are the trajectories of light rays.

Riemannian geometry, generalized to the case of Lorentz metrics, was exactly the tool that Einstein needed to develop his theory.

1.8 The Riemann-Christoffel curvature tensor

Let $(M, \langle \cdot, \cdot \rangle)$ be a Riemannian manifold (or more generally a pseudo-Riemannian manifold) with Levi-Civita connection ∇. If $\mathcal{X}(M)$ denotes the space of smooth vector fields on M, we define

$$R : \mathcal{X}(M) \times \mathcal{X}(M) \times \mathcal{X}(M) \longrightarrow \mathcal{X}(M)$$

by

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

We call R the *Riemann-Christoffel curvature tensor* of $(M, \langle \cdot, \cdot \rangle)$.

Proposition 1. *The operator R is multilinear over functions, that is,*

$$R(fX,Y)Z = R(X,fY)Z = R(X,Y)fZ = fR(X,Y)Z.$$

Proof: We prove only the equality $R(X,Y)fZ = fR(X,Y)Z$, leaving the others as easy exercises:

$$\begin{aligned} R(X,Y)fZ &= \nabla_X \nabla_Y (fZ) - \nabla_Y \nabla_X (fZ) - \nabla_{[X,Y]}(fZ) \\ &= \nabla_X((Yf)Z + f\nabla_Y Z) - \nabla_Y((Xf)Z + f\nabla_X Z) - [X,Y](f)Z - f\nabla_{[X,Y]}(Z) \\ &= XY(f)Z + (Yf)\nabla_X Z + (Xf)\nabla_Y Z + f\nabla_X \nabla_Y Z \\ &\quad - YX(f)Z - (Xf)\nabla_Y Z - (Yf)\nabla_X Z - f\nabla_Y \nabla_Y Z \\ &\quad - [X,Y](f)Z - f\nabla_{[X,Y]}(Z) \\ &= f(\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z) = fR(X,Y)Z. \end{aligned}$$

Since the connection ∇ can be localized by Lemma 1 of §5, so can the curvature; that is, if U is an open subset of M, $(R(X,Y)Z)|U$ depends only $X|U$, $Y|U$ and $Z|U$. Thus Proposition 1 allows us to consider the curvature tensor as defining a multilinear map

$$R_p : T_pM \times T_pM \times T - pM \to T_pM.$$

Moreover, the curvature tensor can be determined in local coordinates from its component functions $R^l_{ijk} : U \to \mathbb{R}$, defined by the equations

$$R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\frac{\partial}{\partial x^k} = \sum_{l=1}^{n} R^l_{kij} \frac{\partial}{\partial x^l}.$$

Proposition 2. *The components R^l_{ijk} of the Riemann-Christoffel curvature tensor are determined from the Christoffel symbols Γ^l_{jk} by the equations*

$$R^l_{kij} = \frac{\partial}{\partial x^i}(\Gamma^l_{jk}) - \frac{\partial}{\partial x^j}(\Gamma^l_{ik}) + \sum_{m=1}^{n} \Gamma^l_{mi}\Gamma^m_{jk} - \sum_{m=1}^{n} \Gamma^l_{mj}\Gamma^m_{ik}. \qquad (1.31)$$

The proof is a straightforward computation.

From the Riemann-Christoffel tensor one can in turn construct other local invariants of Riemannian geometry. For example, the *Ricci curvature* of a pseudo-Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$ is the bilinear form

$$\mathrm{Ric}_p : T_pM \times T_pM \to \mathbb{R} \quad \text{defined by} \quad \mathrm{Ric}_p(x,y) = (\text{Trace of } v \mapsto R_p(v,x)y).$$

Of course, the Ricci curvature also determines a bilinear map

$$\mathrm{Ric} : \mathcal{X}(M) \times \mathcal{X}(M) \longrightarrow \mathcal{F}(M).$$

In terms of coordinates we can write

$$\mathrm{Ric} = \sum_{i,j=1}^{n} R_{ij} dx^i \otimes dx^j, \quad \text{where} \quad R_{ij} = \sum_{k=1}^{n} R^k_{ikj}.$$

Finally, the *scalar curvature* of $(M, \langle\cdot,\cdot\rangle)$ is the function

$$s : M \to \mathbb{R} \quad \text{defined by} \quad s = \sum_{i,j=1}^{n} g^{ij} R_{ij},$$

where $(g^{ij}) = (g_{ij})^{-1}$. It is easily verified that s is independent of choice of local coordinates.

The simplest example of course is Euclidean space $\mathbb{E}^N$. In this case, the metric coefficients g_{ij} are constant, and hence it follows from (1.23) that the Christoffel symbols $\Gamma^k_{ij} = 0$. Thus it follows from Proposition 8.2 that the curvature tensor R is identically zero. Recall that in this case, the Levi-Civita connection ∇^E on $\mathbb{E}^N$ is given by the simple formula

$$\nabla^E_X \left(\sum_{i=1}^{N} f^i \frac{\partial}{\partial x^i} \right) = \sum_{i=1}^{N} X(f^i) \frac{\partial}{\partial x^i}.$$

Similarly, Minkowski space-time $\mathbb{L}^{n+1}$, which consists of the manifold $\mathbb{R}^{n+1}$ with coordinates $(t, x^1, \ldots, x^n)$ and Lorentz metric

$$\langle\cdot,\cdot\rangle = -c^2 dt \otimes dt + \sum_{i=1}^{n} dx^i \otimes dx^i$$

has vanishing Christoffel symbols $\Gamma^k_{ij} = 0$ and hence vanishing curvature.

The next class of examples are the submanifolds of $\mathbb{E}^N$ (or submanifolds of Minkowski space-time) with the induced Riemannian metric. It is often relatively easy to calculate the curvature of these submanifolds by means of the so-called Gauss equation, as we now explain. Thus suppose that $\iota : M \to \mathbb{E}^N$ is an imbedding and agree to identify $p \in M$ with $\iota(p) \in \mathbb{E}^N$ and $v \in T_pM$ with its image $\iota_*(v) \in T_p\mathbb{E}^N$. If $p \in M$ and $v \in T_p\mathbb{E}^N$, we let

$$v = v^\top + v^\perp, \quad \text{where} \quad v^\top \in T_pM \quad \text{and} \quad v^\perp \perp T_pM.$$

Thus $(\cdot)^\top$ is the orthogonal projection into the tangent space and $(\cdot)^\perp$ is the orthogonal projection into the normal space, the orthogonal complement to the tangent space. We have already noted we can define the Levi-Civita connection $\nabla : \mathcal{X}(M) \times \mathcal{X}(M) \to \mathcal{X}(M)$ by the formula

$$(\nabla_X Y)(p) = (\nabla^E_X Y(p))^\top.$$

If we let $\mathcal{X}^\perp(M)$ denote the space of vector fields in $\mathbb{E}^N$ which are defined at points of M and are perpendicular to M, then we can define

$$\alpha : \mathcal{X}(M) \times \mathcal{X}(M) \to \mathcal{X}^\perp(M) \quad \text{by} \quad \alpha(X,Y) = (\nabla^E_X Y(p))^\perp.$$

We call α the *second fundamental form* of M in $\mathbb{E}^N$.

Proposition 3. *The second fundamental form satisfies the identities:*

$$\alpha(fX,Y) = \alpha(X,fY) = f\alpha(X,Y), \quad \alpha(X,Y) = \alpha(Y,X).$$

Indeed,

$$\alpha(fX,Y) = (\nabla^E_{fX}Y)^\perp = f(\nabla^E_X Y)^\perp = f\alpha(X,Y),$$

$$\alpha(X,fY) = (\nabla^E_X)(fY))^\perp = ((Xf)Y + f\nabla^E_X Y)^\perp = f\alpha(X,Y),$$

so α is bilinear over functions. It therefore suffices to establish $\alpha(X,Y) = \alpha(Y,X)$ in the case where $[X,Y] = 0$, but in this case

$$\alpha(X,Y) - \alpha(Y,X) = (\nabla^E_X Y - \nabla^E_Y X)^\perp = 0.$$

There is some special terminology that is used in the case where $\gamma : (a,b) \to M \subseteq \mathbb{E}^N$ is a unit speed curve. In this case, we say that the acceleration $\gamma''(t) \in T_{\gamma(t)}\mathbb{E}^N$ is the *curvature* of γ, while

$$(\gamma''(t))^\top = (\nabla_{\gamma'}\gamma')(t) = (\textit{geodesic curvature} \text{ of } \gamma \text{ at } t),$$

$$(\gamma''(t))^\perp = \alpha(\gamma'(t), \gamma'(t)) = (\textit{normal curvature} \text{ of } \gamma \text{ at } t).$$

Thus if $x \in T_pM$ is a unit length vector, $\alpha(x,x)$ can be interpreted as the normal curvature of some curve tangent to x at p.

Gauss Theorem. The curvature tensor R of a submanifold $M \subseteq \mathbb{E}^N$ is given by the *Gauss equation*

$$\langle R(X,Y)W,Z\rangle = \alpha(X,Z)\cdot\alpha(Y,W) - \alpha(X,W)\cdot\alpha(Y,Z), \tag{1.32}$$

where X, Y, Z and W are elements of $\mathcal{X}(M)$, and the dot on the right denotes the Euclidean metric in the ambient space $\mathbb{E}^N$.

Proof: Since Euclidean space has zero curvature,

$$\nabla^E_X\nabla^E_Y W - \nabla^E_Y\nabla^E_X W - \nabla^E_{[X,Y]}W = 0, \tag{1.33}$$

and hence if the dot denotes the Euclidean dot product,

$$\begin{aligned}
0 &= (\nabla^E_X\nabla^E_Y W)\cdot Z - (\nabla^E_Y\nabla^E_X W)\cdot Z - \nabla^E_{[X,Y]}W\cdot Z \\
&= X(\nabla^E_Y W\cdot Z) - \nabla^E_Y W\cdot\nabla^E_X Z - Y(\nabla^E_X W\cdot Z) + \nabla^E_X W\cdot\nabla^E_Y Z - \nabla^E_{[X,Y]}W\cdot Z \\
&= X\langle\nabla_Y W, Z\rangle - \langle\nabla_Y W,\nabla_X Z\rangle - \alpha(Y,W)\cdot\alpha(X,Z) \\
&\quad - Y\langle\nabla_X W, Z\rangle - \langle\nabla_X W,\nabla_Y Z\rangle - \alpha(X,W)\cdot\alpha(Y,Z) - \langle\nabla_{[X,Y]}W, Z\rangle.
\end{aligned}$$

Thus we find that

$$\begin{aligned}
0 &= \langle\nabla_X\nabla_Y W, Z\rangle - \alpha(Y,W)\cdot\alpha(X,W) \\
&\qquad - \langle\nabla_Y\nabla_X W, Z\rangle + \alpha(X,W)\cdot\alpha(Y,W) - \langle\nabla_{[X,Y]}W, Z\rangle.
\end{aligned}$$

This yields

$$\langle\nabla_X\nabla_Y W - \nabla_Y\nabla_X W - \nabla_{[X,Y]}W, Z\rangle = \alpha(Y,W)\cdot\alpha(X,W) - \alpha(X,W)\cdot\alpha(Y,W),$$

which is exactly (1.32).

Example 1. We can consider the sphere of radius a about the origin in $\mathbb{E}^{n+1}$:

$$\mathbb{S}^n(a) = \{(x^1,\ldots,x^{n+1}) \in \mathbb{E}^{n+1} : (x^1)^2 + \cdots + (x^{n+1})^2 = a^2\}.$$

If $\gamma : (-\epsilon,\epsilon) \to \mathbb{S}^n(a) \subseteq \mathbb{E}^{n+1}$ is a unit speed great circle, say

$$\gamma(t) = a\cos((1/a)t)\mathbf{e}_1 + a\sin((1/a)t)\mathbf{e}_2,$$

where $(\mathbf{e}_1,\mathbf{e}_2)$ are orthonormal vectors located at the origin in $\mathbb{E}^{n+1}$, then a direct calculation shows that

$$\gamma''(t) = -\frac{1}{a}\mathbf{N}(\gamma(t)),$$

where $\mathbf{N}(p)$ is the outward pointing unit normal to $\mathbb{S}^n(a)$ at the point $p \in \mathbb{S}^n(a)$. In particular,

$$(\nabla_{\gamma'}\gamma')(t) = (\gamma''(t))^\top = 0,$$

so γ is a geodesic, and in accordance with the Existence and Uniqueness Theorem for Geodesics from §1.5, we see that the geodesics in $\mathbb{S}^n(a)$ are just the constant speed great circles. Moreover, the second fundamental form of $\mathbb{S}^n(a)$ in $\mathbb{E}^{n+1}$ satisfies

$$\alpha(x,x) = -\frac{1}{a}\mathbf{N}(p), \quad \text{for all unit length } x \in T_p\mathbb{S}^n(a).$$

If x does not have unit length, then

$$\alpha\left(\frac{x}{\|x\|}, \frac{x}{\|x\|}\right) = -\frac{1}{a}\mathbf{N}(p) \quad \Rightarrow \quad \alpha(x,x) = -\frac{1}{a}\langle x, x\rangle \mathbf{N}(p).$$

By polarization, we obtain

$$\alpha(x,y) = \frac{-1}{a}\langle x, y\rangle \mathbf{N}(p), \quad \text{for all } x, y \in T_p\mathbb{S}^n(a).$$

Thus substitution into the Gauss equation yields

$$\begin{aligned}\langle R(x,y)w, z\rangle = &\left(\frac{-1}{a}\langle x, z\rangle \mathbf{N}(p)\right) \cdot \left(\frac{-1}{a}\langle y, w\rangle \mathbf{N}(p)\right) \\ &- \left(\frac{-1}{a}\langle x, w\rangle \mathbf{N}(p)\right) \cdot \left(\frac{-1}{a}\langle y, x\rangle \mathbf{N}(p)\right).\end{aligned}$$

Thus we finally obtain a formula for the curvature of $\mathbb{S}^n(a)$:

$$\langle R(x,y)w, z\rangle = \frac{1}{a^2}(\langle x, z\rangle\langle y, w\rangle - \langle x, w\rangle\langle y, z\rangle).$$

Definition. If $(M_1, \langle\cdot,\cdot\rangle_1)$ and $(M_2, \langle\cdot,\cdot\rangle_2)$ are pseudo-Riemannian manifolds, a diffeomorphism $\phi : M_1 \to M_2$ is said to be an *isometry* if

$$\langle(\phi_*)_p(v), (\phi_*)_p(w)\rangle_2 = \langle v, w\rangle_1, \quad \text{for all } v, w \in T_pM_1 \text{ and all } p \in M_1. \quad (1.34)$$

Of course, we can rewrite (1.34) as $\phi^*\langle\cdot,\cdot\rangle_2 = \langle\cdot,\cdot\rangle_1$, where

$$\phi^*\langle v, w\rangle_2 = \langle(\phi_*)_p(v), (\phi_*)_p(w)\rangle_2, \quad \text{for } v, w \in T_pM_1.$$

Note that the isometries from a pseudo-Riemannian manifold to itself form a group under composition.

Thus, for example, the orthogonal group $O(n+1)$ acts as a group of isometries on $\mathbb{S}^n(a)$, a group of dimension $(1/2)n(n+1)$.

Just like we considered hypersurfaces in $\mathbb{E}^{n+1}$, we can calculate the curvature of "spacelike hypersurfaces" in Minkowski space-time $\mathbb{L}^{n+1}$. In this case, the Christoffel symbols Γ^k_{ij} are zero, so the Levi-Civita connection ∇^L of $\mathbb{L}^{n+1}$ is defined by

$$\nabla^L_X\left(f^0\frac{\partial}{\partial t} + \sum_{i=1}^N f^i\frac{\partial}{\partial x^i}\right) = X(f^0)\frac{\partial}{\partial t} + \sum_{i=1}^N X(f^i)\frac{\partial}{\partial x^i}.$$

Suppose that M is an n-dimensional manifold and $\iota : M \to \mathbb{L}^{n+1}$ is an imbedding. We say that $\iota(M)$ is a *spacelike hypersurface* if the standard Lorentz metric on $\mathbb{L}^{n+1}$ induces a positive-definite Riemannian metric on M. For simplicity, let us set the speed of light $c = 1$ so that the Lorentz metric on $\mathbb{L}^{n+1}$ is simply

$$\langle \cdot, \cdot \rangle_L = -dt \otimes dt + \sum_{i=1}^{n} dx^i \otimes dx^i.$$

Just as in the case where the ambient space is Euclidean space, we find that the Levi-Civita connection $\nabla : \mathcal{X}(M) \times \mathcal{X}(M) \to \mathcal{X}(M)$ on TM is given by the formula

$$(\nabla_X Y)(p) = (\nabla^L_X Y(p))^\top,$$

where $(\cdot)^\top$ is the orthogonal projection into the tangent space. If we let $\mathcal{X}^\perp(M)$ denote the vector field in $\mathbb{L}^N$ which are defined at points of M and are perpendicular to M, we can define the second fundamental form of M in $\mathbb{L}^{n+1}$ by

$$\alpha : \mathcal{X}(M) \times \mathcal{X}(M) \to \mathcal{X}^\perp(M) \quad \text{by} \quad \alpha(X,Y) = (\nabla^E_X Y(p))^\perp,$$

where $(\cdot)^\perp$ is the orthogonal projection to the orthogonal complement to the tangent space. Moreover, the curvature of the spacelike hypersurface is given by the Gauss equation

$$\langle R(X,Y)W, Z \rangle = \langle \alpha(X,Z), \alpha(Y,W) \rangle_L - \langle \alpha(X,W), \alpha(Y,Z) \rangle_L, \tag{1.35}$$

where X, Y, Z and W are elements of $\mathcal{X}(M)$.

Example 2. We can now construct a second important example of an n-dimensional Riemannian manifold for which we can easily calculate geodesics and curvature. Namely, we can set

$$\mathbb{H}^n(a) = \{(t, x^1 \dots, x^n) \in \mathbb{L}^{n+1} : t^2 - (x^1)^2 - \cdots - (x^n)^2 = a^2,\ t > 0\},$$

the set of future-pointed timelike vectors $\mathbf{v}$ situated at the origin in $\mathbb{L}^{n+1}$ such that $\mathbf{v} \cdot \mathbf{v} = -a^2$, where the dot now denotes the Lorentz metric on $\mathbb{L}^{n+1}$. Of course, $\mathbb{H}^n(a)$ is nothing other than the upper sheet of a hyperboloid of two sheets. Clearly $\mathbb{H}^n(a)$ is an imbedded submanifold of $\mathbb{L}^{n+1}$ and we claim that the induced metric on $\mathbb{H}^n(a)$ is positive-definite.

To prove this, we could consider $(x^1, \dots, x^n)$ as global coordinates on $\mathbb{H}^n(a)$, so that

$$t = \sqrt{a^2 + (x^1)^2 + \cdots + (x^n)^2}.$$

Then

$$dt = \frac{\sum_{i=1}^n x^i dx^i}{\sqrt{a^2 + (x^1)^2 + \cdots + (x^n)^2}},$$

and the induced metric on $\mathbb{H}^n(a)$ is

$$\langle \cdot, \cdot \rangle = -\frac{\sum x^i x^j dx^i \otimes dx^j}{a^2 + (x^1)^2 + \cdots + (x^n)^2} + dx^1 \otimes dx^1 + \cdots + dx^n \otimes dx^n.$$

Thus

$$g_{ij} = \delta_{ij} - \frac{\sum x^i x^j}{a^2 + (x^1)^2 + \cdots + (x^n)^2},$$

and from this expression we immediately see that the induced metric on $H^n(a)$ is indeed positive-definite.

Just as $\mathbb{S}^n(a)$ is invariant under a large group of isometries, so is $H^n(a)$. Indeed, we can set

$$I_{1,n} = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

and define a Lie subgroup $O(1,n)$ of the general linear group by

$$O(1,n) = \{A \in GL(n+1,\mathbb{R}) : A^T I_{1,n} A = I_{1,n}\}.$$

Elements of $O(1,n)$ are called *Lorentz transformations*, and it is easily checked that they act as isometries on $\mathbb{L}^{n+1}$. The index two subgroup

$$O^+(1,n) = \{A \in O(1,n) : (\mathbf{v} \text{ future-pointing}) \;\Rightarrow\; (A\mathbf{v} \text{ future-pointing})\}$$

preserves the upper sheet $H^n(a)$ of the hyperboloid of two sheets, and hence acts as isometries on $H^n(a)$. Of course, just like $O(n+1)$, the group $O^+(1,n)$ of *orthochronous Lorentz transformations* has dimension $(1/2)n(n+1)$.

Suppose that $p \in \mathbb{H}^n(a)$, that $\mathbf{e}_0$ is a future-pointing unit length timelike vector such that $p = a\mathbf{e}_0$ and Π is a two-dimensional plane that passes through the origin and contains $\mathbf{e}_0$. Using elementary linear algebra, Π must also contain a unit length spacelike vector $\mathbf{e}_1$ such that $\langle \mathbf{e}_0, \mathbf{e}_1 \rangle_L = 0$. (In fact, after an orthochronous Lorentz transformation, we can arrange that $\mathbf{e}_)$ points along the t-axis in $\mathbb{L}^{n+1}$. Then the smooth curve

$$\gamma : (-\epsilon, \epsilon) \to \mathbb{L}^{n+1} \quad \text{defined by} \quad \gamma(t) = a\cosh(t/a)\mathbf{e}_0 + a\sinh(t/a)\mathbf{e}_1$$

lies in $H^n(a)$ because

$$a^2(\cosh(t/a))^2 - a^2(\sinh(t/a))^2 = a^2.$$

Note that γ is spacelike and direct calculation shows that

$$\langle \gamma'(t), \gamma'(t) \rangle_L = -(\sinh(t/a))^2 + (\cosh(t/a))^2 = 1.$$

Moreover,

$$\gamma''(t) = \frac{1}{a}(\cosh(t/a)\mathbf{e}_0 + \sinh(t/a)\mathbf{e}_1) = \frac{1}{a}\mathbf{N}(\gamma(t)),$$

where $\mathbf{N}(p)$ is the unit normal to $\mathbb{H}^n(a)$ at p. Thus

$$(\nabla_{\gamma'}\gamma')(t) = (\gamma''(t))^\top = 0,$$

so γ is a geodesic and

$$\alpha(\gamma'(t),\gamma'(t)) = (\gamma''(t))^\perp = \frac{1}{a}\mathbf{N}(\gamma(t)).$$

Since we can construct a unit speed geodesic γ in M as above with $\gamma(0) = p$ for any $p \in H^n(a)$ and $\gamma'(0) = \mathbf{e}_1$ for any unit length $\mathbf{e}_1 \in T_p\mathbb{H}^n(a)$, we have constructed all the unit-speed geodesics in $\mathbb{H}^n(a)$.

Thus just as in the case of the sphere, we can use the Gauss equation (1.35) to determine the curvature of $\mathbb{H}^n(a)$. Thus the second fundamental form of $\mathbb{H}^n(a)$ in $\mathbb{L}^{n+1}$ satisfies

$$\alpha(x,x) = -\frac{1}{a}\mathbf{N}(p), \quad \text{for all unit length } x \in T_p\mathbb{H}^n(a),$$

where $\mathbf{N}(p)$ is the future-pointing unit normal to M. If x does not have unit length, then

$$\alpha\left(\frac{x}{\|x\|},\frac{x}{\|x\|}\right) = \frac{1}{a}\mathbf{N}(p) \quad \Rightarrow \quad \alpha(x,x) = \frac{1}{a}\langle x,x\rangle\mathbf{N}(p).$$

By polarization, we obtain

$$\alpha(x,y) = \frac{1}{a}\langle x,y\rangle\mathbf{N}(p), \quad \text{for all } x,y \in T_p\mathbb{N}^n(a).$$

Thus substitution into the Gauss equation yields

$$\begin{aligned}\langle R(x,y)w,z\rangle = \left(\frac{1}{a}\langle x,z\rangle\mathbf{N}(p)\right)\cdot\left(\frac{1}{a}\langle y,w\rangle\mathbf{N}(p)\right)\\ -\left(\frac{1}{a}\langle x,w\rangle\mathbf{N}(p)\right)\cdot\left(\frac{1}{a}\langle y,x\rangle\mathbf{N}(p)\right).\end{aligned}$$

Since $\mathbf{N}(p)$ is timelike and hence $\langle\mathbf{N}(p),\mathbf{N}(p)\rangle_L = -1$, we finally obtain a formula for the curvature of $\mathbb{S}^n(a)$:

$$\langle R(x,y)w,z\rangle = \frac{-1}{a^2}(\langle x,z\rangle\langle y,w\rangle - \langle x,w\rangle\langle y,z\rangle).$$

The Riemannian manifold $\mathbb{H}^n(a)$ is called *hyperbolic space*, and its geometry is called *hyperbolic geometry*. We have constructed a model for hyperbolic geometry, the upper sheet of the hyperboloid of two sheets in $\mathbb{L}^{n+1}$, and have seen that the geodesics in this model are just the intersections with two-planes passing through the origin in $\mathbb{L}^{n+1}$.

When M is either $\mathbb{S}^n(a)$ and $\mathbb{H}^n(a)$, there is an isometry ϕ which takes any point p of M to any other point q and any orthonormal basis of T_pM to any orthonormal basis of T_qM. This allows us to construct non-Euclidean geometries for $\mathbb{S}^n(a)$ and $\mathbb{H}^n(a)$ which are quite similar to Euclidean geometry. In the case of $\mathbb{H}^n(a)$ all the postulates of Euclidean geometry are satisfied except for the parallel postulate.

1.9 Curvature symmetries; sectional curvature

The Riemann-Christoffel curvature tensor is the basic local invariant of a pseudo-Riemannian manifold. If M has dimension n, one would expect R to have n^4 independent components R^l_{ijk}, but the number of independent components is cut down considerably because of the curvature symmetries. Indeed, the following proposition shows that a two-dimensional Riemannian manifold has only one independent curvature component, a three-dimensional manifold only six, a four-dimensional manifold only twenty:

Proposition 1. *The curvature tensor R of a pseudo-Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$ satisfies the identities:*

1. $R(X,Y) = -R(Y,X)$,
2. $R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0$,
3. $\langle R(X,Y)W, Z\rangle = -\langle R(X,Y)Z, W\rangle$, *and*
4. $\langle R(X,Y)W, Z\rangle = -\langle R(W,Z)X, Y\rangle$.

Remark 1. If we assumed the Nash imbedding theorem (in the Riemannian case), we could derive these identities immediately from the Gauss equation (1.32).

Remark 2. We can write the above curvature symmetries in terms of the components R^l_{ijk} of the curvature tensor. Actually, it is easier to express these symmetries if we "lower the index" and write

$$R_{lkij} = \sum_{p=1}^{n} g_{lp} R^p_{kij}. \qquad (1.36)$$

In terms of these components, the curvature symmetries are

$$R_{lkij} = -R_{lkji}, \quad R_{lkij} + R_{ljki} + R_{lijk} = 0,$$
$$R_{lkij} = -R_{klij}, \quad R_{lkij} = R_{ijlk}.$$

In view of the last symmetry the lowering of the index into the third position in (1.36) is consistent with regarding the R_{ijlk}'s as the components of the map

$$R : T_pM \times T_pM \times T_pM \times T_pM \to \mathbb{R} \quad \text{by} \quad R(X,Y,Z,W) = \langle R(X,Y)W, Z\rangle.$$

Proof of proposition: Note first that since R is a tensor, we can assume without loss of generality that all brackets of X, Y, Z and W are zero. Then

$$R(X,Y) = \nabla_X\nabla_Y - \nabla_Y\nabla_X = -(\nabla_Y\nabla_X - \nabla_X\nabla_Y) = -R(Y,X),$$

establishing the first identity. Next,

$$\begin{aligned}R(X,Y)Z + R(Y,Z)X + R(Z,X)Y &= \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z \\ &\quad + \nabla_Y\nabla_Z X - \nabla_Z\nabla_Y X + \nabla_Z\nabla_X Y - \nabla_X\nabla_Z Y \\ &= \nabla_X(\nabla_Y Z - \nabla_Z Y) + \nabla_Y(\nabla_Z X - \nabla_X Z) + \nabla_Z(\nabla_X Y - \nabla_Y X) = 0,\end{aligned}$$

the last equality holding because ∇ is symmetric. For the third identity, we calculate

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

Hence the symmetric part of the bilinear form

$$(W, Z) \mapsto \langle R(X,Y)W, Z\rangle$$

is zero, from which the third identity follows. Finally, it follows from the first and second identities that

$$\langle R(X,Y)W, Z\rangle = -\langle R(Y,X)W, Z\rangle = \langle R(X,W)Y, Z\rangle + \langle R(W,Y)X, Z\rangle,$$

and from the third and second that

$$\langle R(X,Y)W, Z\rangle = -\langle R(X,Y)Z, W\rangle = \langle R(Y,Z)X, W\rangle + \langle R(Z,X)Y, W\rangle.$$

Adding the last two expressions yields

$$\begin{aligned}2\langle R(X,Y)W, Z\rangle &= \langle R(X,W)Y, Z\rangle \\ &\quad + \langle R(W,Y)X, Z\rangle + \langle R(Y,Z)X, W\rangle + \langle R(Z,X)Y, W\rangle. \qquad (1.37)\end{aligned}$$

Exchanging the pair (X, Y) with (W, Z) yields

$$\begin{aligned}2\langle R(W,Z)X, Y\rangle &= \langle R(W,X)Z, Y\rangle \\ &\quad + \langle R(X,Z)W, Y\rangle + \langle R(Z,Y)W, X\rangle + \langle R(Y,W)Z, X\rangle. \qquad (1.38)\end{aligned}$$

Each term on the right of (1.37) equals one of the terms on the right of (1.38), so

$$\langle R(X,Y)W, Z\rangle = \langle R(W,Z)X, Y\rangle,$$

finishing the proof of the proposition.

Using the first and third of the symmetries we can define a linear map called the *curvature operator*,

$$\mathcal{R} : \Lambda^2 T_p M \to \Lambda^2 T_p M \quad \text{by} \quad \langle \mathcal{R}(x \wedge y), z \wedge w\rangle = \langle R(x,y)w, z\rangle.$$

It follows from the fourth symmetry that $\mathcal{R}$ is symmetric,

$$\langle \mathcal{R}(x \wedge y), z \wedge w \rangle = \langle \mathcal{R}(z \wedge w), x \wedge y \rangle,$$

so all the eigenvalues of $\mathcal{R}$ are real and $\Lambda^2 T_p M$ has an orthonormal basis consisting of eigenvectors.

Proposition 2. *Let*

$$R, S : T_p M \times T_p M \times T_p M \times T_p M \to \mathbb{R}$$

be two quadrilinear functions which satisfy the curvature symmetries. If

$$R(x, y, x, y) = S(x, y, x, y), \quad \text{for all } x, y \in T_p M,$$

then $R = S$.

Proof: Let $T = R - S$. Then T satisfies the curvature symmetries and

$$T(x, y, x, y) = 0, \quad \text{for all } x, y \in T_p M.$$

Hence

$$\begin{aligned} 0 &= T(x, y + z, x, y + z) \\ &= T(x, y, x, y) + T(x, y, x, z) + T(x, z, x, y) + T(x, z, x, z) \\ &= 2T(x, y, x, z), \end{aligned}$$

so $T(x, y, x, z) = 0$. Similarly,

$$0 = T(x + z, y, x + z, w) = T(x, y, z, w) + T(z, y, x, w),$$

$$0 = T(x + w, y, z, x + w) = T(x, y, z, w) + T(w, y, z, x).$$

Finally,

$$\begin{aligned} 0 &= 2T(x, y, z, w) + T(z, y, x, w) + T(w, y, z, x) \\ &= 2T(x, y, z, w) - T(y, z, x, w) - T(z, x, y, w) = 3T(x, y, z, w). \end{aligned}$$

So $T = 0$ and $R = S$.

This proposition shows that the curvature is completely determined by the sectional curvatures, defined as follows:

Definition. Suppose that σ is a two-dimensional subspace of $T_p M$ such that the restriction of $\langle \cdot, \cdot \rangle$ to σ is nondegenerate. Then the *sectional curvature* of σ is

$$K(\sigma) = \frac{\langle R(x, y)y, x \rangle}{\langle x, x \rangle \langle y, y \rangle - \langle x, y \rangle^2},$$

whenever (x, y) is a basis for σ. The curvature symmetries imply that $K(\sigma)$ is independent of the choice of basis.

Recall our key three examples, the so-called *spaces of constant curvature*:

$$\begin{cases} \text{If } M = \mathbb{E}^n, \text{ then } K(\sigma) \equiv 0 \text{ for all two-planes } \sigma \subseteq T_pM. \\ \text{If } M = \mathbb{S}^n(a), \text{ then } K(\sigma) \equiv 1/a^2 \text{ for all two-planes } \sigma \subseteq T_pM. \\ \text{If } M = \mathbb{H}^n(a), \text{ then } K(\sigma) \equiv -1/a^2 \text{ for all two-planes } \sigma \subseteq T_pM. \end{cases}$$

Along with the projective space $\mathbb{P}^n(a)$, which is obtained from $\mathbb{S}^n(a)$ by identifying antipodal points, these are the most symmetric Riemannian manifolds possible; it can be shown that they are the only n-dimensional Riemannian manifolds which have an isometry group of maximal possible dimension $(1/2)n(n+1)$.

1.10 Gaussian curvature of surfaces

We now make contact with the theory of surfaces in $\mathbb{E}^3$ as described in undergraduate texts such as [19]. If $(M, \langle \cdot, \cdot \rangle)$ is a two-dimensional Riemannian manifold, then there is only one two-plane at each point p, namely T_pM. In this case, we can define a smooth function $K : M \to \mathbb{R}$ by

$$K(p) = K(T_pM) = (\text{sectional curvature of } T_pM).$$

The function K is called the *Gaussian curvature* of M, and is easily checked to also equal $s/2$, where s is the scalar curvature of M. As we saw in the previous section, the Gaussian curvature of a two-dimensional Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$ determines the entire Riemann-Christoffel curvature tensor.

An important special case is that of a two-dimensional smooth surface M^2 imbedded in $\mathbb{E}^3$, with M^2 given the induced Riemannian metric. We say that M is *orientable* if it is possible to choose a smooth unit normal $\mathbf{N}$ to M,

$$\mathbf{N} : M^2 \to \mathbb{S}^2(1) \subseteq \mathbb{E}^3, \quad \text{with } \mathbf{N}(p) \text{ perpendicular to } T_pM.$$

Such a choice of unit normal is said to determine an *orientation* of M^2.

If N_pM is the orthogonal complement to T_pM in Euclidean space, then the second fundamental form $\alpha : T_pM \times T_pM \to N_pM$ determines a symmetric bilinear form $h : T_pM \times T_pM \to \mathbb{R}$ by the formula

$$h(x, y) = \alpha(x, y) \cdot \mathbf{N}(p), \quad \text{for } x, y \in T_pM,$$

and this $\mathbb{R}$-valued symmetric bilinear is also often called the *second fundamental form* of the surface M. Note that if we reverse orientation, h changes sign.

Recall that if (x^1, x^2) is a smooth coordinate system on M, we can define the components of the induced Riemannian metric on M^2 by the formulae

$$g_{ij} = \left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle, \quad \text{for } i, j = 1, 2.$$

If $F : M^2 \to \mathbb{E}^3$ is the imbedding than the components of the induced Riemannian metric (also called the first fundamental form) are given by the formula

$$g_{ij} = F_* \left(\frac{\partial}{\partial x^i} \right) \cdot F_* \left(\frac{\partial}{\partial x^j} \right) = \frac{\partial F}{\partial x^i} \cdot \frac{\partial F}{\partial x^j}.$$

Similarly, we can define the components of the second fundamental form by

$$h_{ij} = h\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right), \quad \text{for } i,j = 1,2.$$

These components can be found by the explicit formula

$$h_{ij} = \left(\nabla^E_{\partial/\partial x^i} \frac{\partial}{\partial x^j}\right) \cdot \mathbf{N} = \frac{\partial^2 F}{\partial x^i \partial x^j} \cdot \mathbf{N}.$$

If we let

$$X = \frac{\partial}{\partial x^1}, \quad Y = \frac{\partial}{\partial x^2},$$

then it follows from the definition of Gaussian curvature and the Gauss equation that

$$\begin{aligned} K &= \frac{\langle R(X,Y)Y,X\rangle}{\langle X,X\rangle\langle Y,Y\rangle - \langle X,Y\rangle^2} \\ &= \frac{\alpha(X,X)\cdot\alpha(Y,Y) - \alpha(X,Y)\cdot\alpha(X,Y)}{\langle X,X\rangle\langle Y,Y\rangle - \langle X,Y\rangle^2} \\ &= \frac{h_{11}h_{22} - h_{12}^2}{g_{11}g_{22} - g_{12}^2} = \frac{\begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix}}{\begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix}}. \end{aligned} \tag{1.39}$$

In his *General investigations regarding curved surfaces* of 1827, Karl Friedrich Gauss defined the Gaussian curvature by (1.39). His Theorema Egregium stated that the Gaussian curvature depends only on the Riemannian metric (g_{ij}). From our viewpoint, this follows from the fact that the Riemann-Christoffel curvature is determined by the Levi-Civita connection, which in turn is determined by the Riemannian metric.

One says that the *intrinsic geometry* of a surface $M^2 \subseteq \mathbb{E}^3$ is the geometry of its first fundamental form, that is, its Riemannian metric. Everything that can be defined in terms of the Riemannian metric, such as the geodesics, also belongs to the intrinsic geometry of the surface. The second fundamental form $\alpha = h\mathbf{N}$ determines more—it determines also the *extrinisic geometry* of the surface. Thus, for example, a short calculation shows that the plane

$$F_1 : \mathbb{R}^2 \to \mathbb{E}^3 \quad \text{defined by} \quad F_1(u,v) = (u,0,v)$$

and the cylinder over the catenary

$$F_2 : \mathbb{R}^2 \to \mathbb{E}^3 \quad \text{defined by} \quad F_2(u,v) = \left(\log\left(u + \sqrt{u^2+1}\right), \sqrt{u^2+1}, v\right)$$

both induce the same Riemannian metric

$$\langle \cdot, \cdot \rangle = du \otimes du + dv \otimes dv,$$

so they have the same intrinsic geometry, yet their second fundamental forms are different, so they have different extrinsic geometry.

If $\gamma : (-\epsilon, \epsilon) \to S$ has unit speed, then $\alpha(\gamma'(0), \gamma'(0))$ is simply the normal curvature of γ at $t = 0$. Thus the normal curvatures of curves passing through p at time 0 are just the values of the function

$$\kappa : T_p' M \longrightarrow R, \quad \text{defined by} \quad \kappa(v) = \alpha(\mathbf{v}, \mathbf{v}),$$

where $T_p' M$ is the unit circle in the tangent space:

$$T_p' M = \{\mathbf{v} \in T_p M : \mathbf{v} \cdot \mathbf{v} = 1\}.$$

It is a well known fact from real analysis that a continuous function on a circle must achieve its maximum and minimum values. These values are called the *principal curvatures* and are denoted by $\kappa_1(p)$ and $\kappa_2(p)$. The values of the principal curvatures can be found via the method of Lagrange multipliers from second-year calculus: one seeks the maximum and minimum values of the function

$$\kappa(v) = \sum_{i,j=1}^{2} h_{ij} v^i v^j \quad \text{subject to the constraint} \quad \langle v, v \rangle = \sum_{i,j=1}^{2} g_{ij} v^i v^j = 1.$$

One thus finds that the principal curvatures are just the roots of the equation for λ:

$$\begin{vmatrix} h_{11} - \lambda g_{11} & h_{12} - \lambda g_{12} \\ h_{12} - \lambda g_{21} & h_{22} - \lambda g_{22} \end{vmatrix} = 0.$$

One easily verifies that the Gaussian curvature is just the product of the principal curvatures, $K = \kappa_1 \kappa_2$, but we can also construct an important extrinsic quantity,

$$(\text{the mean curvature}) = H = \frac{1}{2}(\kappa_1 + \kappa_2).$$

Surfaces which locally minimize area can be shown to have mean curvature zero. Such surfaces are called minimal surfaces and a vast literature is devoted to their study.

Example 1. Let us consider the *catenoid*, the submanifold of $\mathbb{R}^3$ defined by the equation

$$r = \sqrt{x^2 + y^2} = \cosh z,$$

where (r, θ, z) are cylindrical coordinates. This is obtained by rotating the catenary around the z-axis in (x, y, z)-space. As parametrization, we can take $M^2 = \mathbb{R} \times S^1$ and

$$F : \mathbb{R} \times S^1 \to \mathbb{E}^3 \quad \text{by} \quad F(u, v) = \begin{pmatrix} \cosh u \cos v \\ \cosh u \sin v \\ u \end{pmatrix}.$$

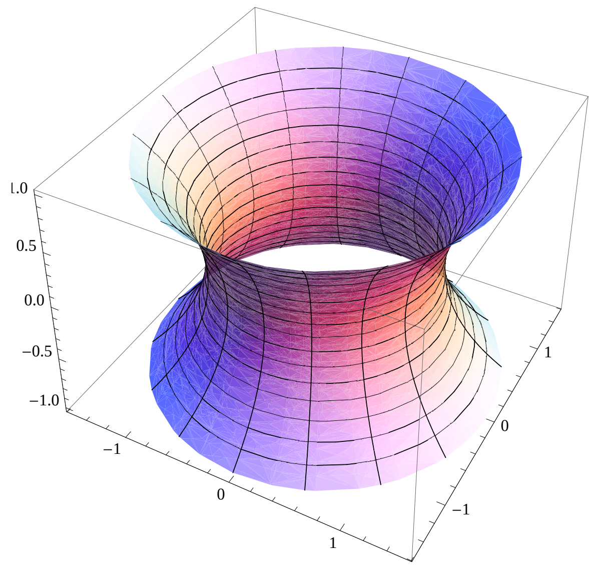

Figure 1.1: The catenoid is the unique complete minimal surface of revolution in $\mathbb{E}^3$.

Here u is the coordinate on $\mathbb{R}$ and v is the coordinate on S^1 which is just the quotient group $\mathbb{R}/\mathbb{Z}$, where $\mathbb{Z}$ is the cyclic group generated by 2π. Then

$$\frac{\partial F}{\partial u} = \begin{pmatrix} \sinh u \cos v \\ \sinh u \sin v \\ 1 \end{pmatrix} \quad \text{and} \quad \frac{\partial F}{\partial v} = \begin{pmatrix} -\cosh u \sin v \\ \cosh u \cos v \\ 0 \end{pmatrix},$$

and hence the coefficients of the first fundamental form in this case are

$$g_{11} = 1 + \sinh^2 u = \cosh^2 u, \quad g_{12} = 0 \quad \text{and} \quad g_{22} = \cosh^2 u.$$

The induced Riemannian metric (or first fundamental form) in this case is

$$\langle \cdot, \cdot \rangle = \cosh^2 u (du \otimes du + dv \otimes dv).$$

To find a unit normal, we calculate

$$\frac{\partial F}{\partial u} \times \frac{\partial F}{\partial v} = \begin{vmatrix} \mathbf{i} & \sinh u \cos v & -\cosh u \sin v \\ \mathbf{j} & \sinh u \sin v & \cosh u \cos v \\ \mathbf{k} & 1 & 0 \end{vmatrix} = \begin{pmatrix} -\cosh u \cos v \\ -\cosh u \sin v \\ \cosh u \sinh u \end{pmatrix},$$

and then a unit normal to the surface can be given by the formula

$$\mathbf{N} = \frac{\frac{\partial F}{\partial u} \times \frac{\partial F}{\partial v}}{\left| \frac{\partial F}{\partial u} \times \frac{\partial F}{\partial v} \right|} = \frac{1}{\cosh u} \begin{pmatrix} -\cos v \\ -\sin v \\ \sinh u \end{pmatrix}.$$

To calculate the second fundamental form, we need the second order partial derivatives,

$$\frac{\partial^2 F}{\partial u^2} = \begin{pmatrix} \cosh u \cos v \\ \cosh u \sin v \\ 0 \end{pmatrix}, \quad \frac{\partial^2 F}{\partial u \partial v} = \begin{pmatrix} -\sinh u \sin v \\ \sinh u \cos v \\ 0 \end{pmatrix},$$

and

$$\frac{\partial^2 F}{\partial v^2} = \begin{pmatrix} -\cosh u \cos v \\ -\cosh u \sin v \\ 0 \end{pmatrix}.$$

These give the coefficients of the second fundamental form

$$h_{11} = \frac{\partial^2 F}{\partial u^2} \cdot \mathbf{N} = -1, \qquad h_{12} = h_{21} = \frac{\partial^2 F}{\partial u \partial v} \cdot \mathbf{N} = 0,$$

and

$$h_{22} = \frac{\partial^2 F}{\partial v^2} \cdot \mathbf{N} = 1.$$

From these components we can easily calculate the Gaussian curvature of the catenoid

$$K = \frac{-1}{(\cosh u)^4}.$$

Moreover, one can check that the catenoid has mean curvature zero, so it is a minimal surface. In fact, it is not difficult to show that the catenoid is the only complete minimal surface of revolution in $\mathbb{E}^3$.

Example 2. Another important example is the *pseudosphere*, the submanifold of $\mathbb{R}^3$ parametrized by $\mathbf{x} : M^2 \to R^3$, where $M^2 = (0, \infty) \times S^1$ and

$$F : (0, \infty) \times S^1 \to \mathbb{E}^3 \quad \text{by} \quad F(u, v) = \begin{pmatrix} e^{-u} \cos v \\ e^{-u} \sin v \\ \int_1^u \sqrt{1 - e^{-2w}} dw \end{pmatrix},$$

where u is the coordinate in $(0, \infty)$ and v is the coordinate on S^1 once again. This surface is obtained by rotating a curve called the tractrix around the z-axis in (x, y, z)-space. Then

$$\frac{\partial F}{\partial u} = \begin{pmatrix} -e^{-u} \cos v \\ -e^{-u} \sin v \\ \sqrt{1 - e^{-2u}} \end{pmatrix} \quad \text{and} \quad \frac{\partial F}{\partial v} = \begin{pmatrix} -e^{-u} \sin v \\ e^{-u} \cos v \\ 0 \end{pmatrix},$$

and hence the coefficients of the first fundamental form in this case are

$$g_{11} = 1, \quad g_{12} = 0 \quad \text{and} \quad g_{22} = e * -2u.$$

To find a unit normal, we once again calculate

$$\frac{\partial F}{\partial u} \times \frac{\partial F}{\partial v} = \begin{vmatrix} \mathbf{i} & -e^{-u} \cos v & -e^{-u} \sin v \\ \mathbf{j} & -e^{-u} \sin v & e^{-u} \cos v \\ \mathbf{k} & \sqrt{1 - e^{-2u}} & 0 \end{vmatrix} = \begin{pmatrix} -e^{-u} \sqrt{1 - e^{-2u}} \cos v \\ -e^{-u} \sqrt{1 - e^{-2u}} \sin v \\ -e^{-2u} \end{pmatrix},$$

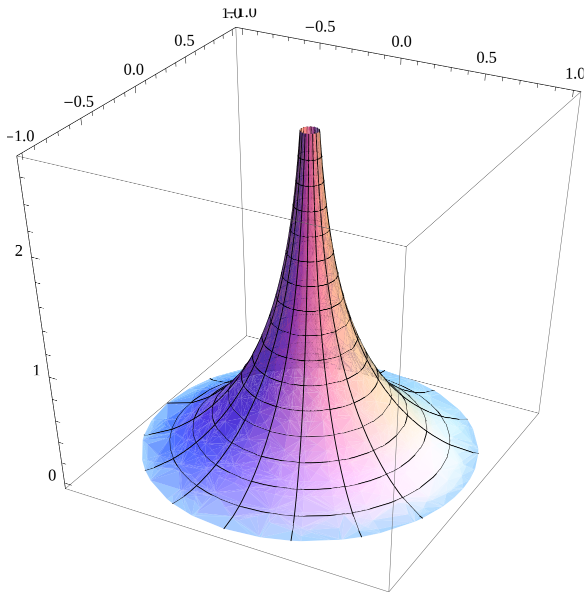

Figure 1.2: The pseudosphere is a surface of revolution in $\mathbb{E}^3$ with $K = -1$.

and then the unit normal to the surface is given by

$$\mathbf{N} = \frac{\frac{\partial F}{\partial u} \times \frac{\partial F}{\partial v}}{\left|\frac{\partial F}{\partial u} \times \frac{\partial F}{\partial v}\right|} = \begin{pmatrix} -\sqrt{1-e^{-2u}}\cos v \\ -\sqrt{1-e^{-2u}}\sin v \\ -e^{-2u} \end{pmatrix}.$$

To calculate the second fundamental form, we need the second order partial derivatives,

$$\frac{\partial^2 F}{\partial u^2} = \begin{pmatrix} e^{-u}\cos v \\ e^{-u}\sin v \\ e^{-u}(1-e^{-2u})^{-1/2} \end{pmatrix}, \quad \frac{\partial^2 F}{\partial u \partial v} = \begin{pmatrix} e^{-u}\sin v \\ -e^{-u}\cos v \\ 0 \end{pmatrix},$$

and

$$\frac{\partial^2 F}{\partial v^2} = \begin{pmatrix} -e^{-u}\cos v \\ -e^{-u}\sin v \\ 0 \end{pmatrix}.$$

These give the coefficients of the second fundamental form

$$h_{11} = \frac{\partial^2 F}{\partial u^2} \cdot \mathbf{N} = \frac{e^{-u}}{\sqrt{1-e^{-2u}}}, \qquad h_{12} = h_{21} = \frac{\partial^2 F}{\partial u \partial v} \cdot \mathbf{N} = 0,$$

and

$$h_{22} = \frac{\partial^2 F}{\partial v^2} \cdot \mathbf{N} = e^{-u}\sqrt{1-e^{-2u}}.$$

From these components we see that the Gaussian curvature of the pseudosphere is $K = -1$.

Exercise III. Due Friday, April 22. Consider the torus $T^2 = S^1 \times S^1$ with imbedding

$$F : T^2 \to S \quad \text{by} \quad F(u, v) = \begin{pmatrix} (2 + \cos u)\cos v \\ (2 + \cos u)\sin v \\ \sin u \end{pmatrix},$$

where u and v are the angular coordinates on the two S^1 factors, with $u+2\pi = u$, $v + 2\pi = v$.

a. Calculate the components g_{ij} of the induced Riemannian metric on M^2.

b. Calculate a continuously varying unit normal $\mathbf{N}$ and the components h_{ij} of the second fundamental form of M^2.

c. Determine the Gaussian curvature K.

1.11 Matrix Lie groups

In addition to the spaces of constant curvature, there is another class of manifolds for which the geodesics and curvature can be computed with relative ease, the compact Lie groups with biinvariant Riemannian metrics. Before discussing these examples, we give a brief review of Lie groups and Lie algebras. For a more detailed discussion, one could refer to Chapter 20 of [12].

Suppose that G is a Lie group and $\sigma \in G$. We can define the *left translation* by σ,

$$L_\sigma : G \to G \quad \text{by} \quad L_\sigma(\tau) = \sigma\tau,$$

a map which is clearly a diffeomorphism. Similarly, we can define *right translation*

$$R_\sigma : G \to G \quad \text{by} \quad R_\sigma(\tau) = \tau\sigma.$$

A vector field X on G is said to be *left invariant* if $(L_\sigma)_*(X) = X$ for all $\sigma \in G$, where

$$(L_\sigma)_*(X)(f) = X(f \circ L_\sigma) \circ L_\sigma^{-1}.$$

A straightforward calculation shows that if X and Y are left invariant vector fields on G, then so is their bracket $[X, Y]$. Thus the space

$$\mathfrak{g} = \{X \in \mathcal{X}(G) : (L_\sigma)_*(X) = X \text{ for all } \sigma \in G \}$$

is closed under Lie bracket, and the real bilinear map

$$[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$$

is skew-symmetric (that is, $[X, Y] = -[Y, X]$), and satisfies the *Jacobi identity*

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

Thus $\mathfrak{g}$ is a Lie algebra and we call it the *Lie algebra* of G. If e is the identity of the Lie group, restriction to T_eG yields an isomorphism $\alpha : \mathfrak{g} \to T_eG$. The inverse $\beta : T_eG \to \mathfrak{g}$ is defined by $\beta(v)(\sigma) = (L_\sigma)_*(v)$.

The most important examples of Lie groups are the *general linear group*

$$GL(n,\mathbb{R}) = \{\ n \times n \text{ matrices } A \text{ with real entries } : \det A \neq 0\},$$

and its subgroups, which are called *matrix Lie groups*. For $1 \le i, j \le n$, we can define coordinates

$$x^i_j : GL(n,\mathbb{R}) \to \mathbb{R} \quad \text{by} \quad x^i_j((a^i_j)) = a^i_j.$$

Of course, these are just the rectangular cartesian coordinates on an ambient Euclidean space in which $GL(n,\mathbb{R})$ sits as an open subset. If $X = (x^i_j) \in GL(n,\mathbb{R})$, left translation by X is a linear map, so is its own differential. Thus

$$(L_X)_* \left(\sum_{i,j=1}^n a^i_j \frac{\partial}{\partial x^i_j} \right) = \sum_{i,j,k=1}^n x^i_k a^k_j \frac{\partial}{\partial x^i_j}.$$

If we allow X to vary over $GL(n,\mathbb{R})$ we obtain a left invariant vector field

$$X_A = \sum_{i,j,k=1}^n a^i_j x^k_i \frac{\partial}{\partial x^k_j}$$

which is defined on $GL(n,\mathbb{R})$. It is the unique left invariant vector field on $GL(n,\mathbb{R})$ which satisfies the condition

$$X_A(I) = \sum_{i,j=1}^n a^i_j \left. \frac{\partial}{\partial x^i_j} \right|_I ,$$

where I is the identity matrix, the identity of the Lie group $GL(n,\mathbb{R})$. Every left invariant vector field on $GL(n,\mathbb{R})$ is obtained in this way, for some choice of $n \times n$ matrix $A = (a^i_j)$. A direct calculation yields

$$[X_A, X_B] = X_{[A,B]}, \quad \text{where} \quad [A,B] = AB - BA, \tag{1.40}$$

which gives an alternate proof that left invariant vector fields are closed under Lie brackets in this case. Thus the Lie algebra of $GL(n,\mathbb{R})$ is isomorphic to

$$\mathfrak{g}l(n,\mathbb{R}) \cong T_IG = \{\ n \times n \text{ matrices } A \text{ with real entries } \},$$

with the usual bracket of matrices as Lie bracket.

For a general Lie group G, if $X \in \mathfrak{g}$, the integral curve θ_X for X such that $\theta_X(0) = e$ satisfies the identity $\theta_X(s+t) = \theta_X(s) \cdot \theta_X(t)$ for sufficiently small s and t. Indeed,

$$t \mapsto \theta_X(s+t) \quad \text{and} \quad t \mapsto \theta_X(s) \cdot \theta_X(t)$$

are two integral curves for X which agree when $t = 0$, and hence must agree for all t. From this one can easily argue that $\theta_X(t)$ is defined for all $t \in \mathbb{R}$, and thus θ_X provides a Lie group homomorphism

$$\theta_X : \mathbb{R} \longrightarrow G.$$

We call θ_X the *one-parameter group* which corresponds to $X \in \mathfrak{g}$. Since the vector field X is left invariant, the curve

$$t \mapsto L_\sigma(\theta_X(t)) = \sigma\theta_X(t) = R_{\theta_X(t)}(\sigma)$$

is the integral curve for X which passes through σ at $t = 0$, and therefore the one-parameter group $\{\phi_t : t \in \mathbb{R}\}$ of diffeomorphisms on G which corresponds to $X \in \mathfrak{g}$ is given by

$$\phi_t = R_{\theta_X(t)}, \quad \text{for } t \in \mathbb{R}.$$

In the case where $G = GL(n, \mathbb{R})$ the one-parameter groups are easy to describe. In this case, if $A \in \mathfrak{g}l(n, \mathbb{R})$, we claim that the corresponding one-parameter group is

$$\theta_A(t) = e^{tA} = I + tA + \frac{1}{2!}t^2A^2 + \frac{1}{2!}t^3A^3 + \cdots .$$

Indeed, it is easy to prove directly that the power series converges for all $t \in \mathbb{R}$, and termwise differentiation shows that it defines a smooth map. The usual proof that $e^{t+s} = e^t e^s$ extends to a proof that $e^{(t+s)A} = e^{tA}e^{sA}$, so $t \mapsto e^{tA}$ is a one-parameter group. Finally, since

$$\frac{d}{dt}(e^{tA}) = Ae^{tA} = e^{tA}A \quad \text{the curve} \quad t \mapsto e^{tA}$$

is tangent to A at the identity.

If G is a Lie subgroup of $GL(n, \mathbb{R})$, then its left invariant vector fields are defined by taking elements of $T_I G \subseteq T_I GL(n, \mathbb{R})$ and spreading them out over G by left translations of G. Thus the left invariant vector fields on G are just the restrictions of the elements of $\mathfrak{g}l(n, \mathbb{R})$ which are tangent to G.

We can use the one-parameter groups to determine which elements of $\mathfrak{g}l(n, \mathbb{R})$ are tangent to G at I. Consider, for example, the *orthogonal group*,

$$O(n) = \{A \in GL(n, \mathbb{R}) : A^T A = I\},$$

where $(\cdot)^T$ denotes transpose, and its identity component, the *special orthogonal group*,

$$SO(n) = \{A \in O(n) : \det A = 1\}.$$

In either case, the corresponding Lie algebra is

$$\mathfrak{o}(n) = \{A \in \mathfrak{g}l(n, \mathbb{R}) : e^{tA} \in O(n) \text{ for all } t \in \mathbb{R} \ \}.$$

Differentiating the equation

$$(e^{tA})^T e^{tA} = I \quad \text{yields} \quad (e^{tA})^T A^T e^{tA} + (e^{tA})^T A e^{tA} = 0,$$

and evaluating at $t = 0$ yields a formula for the Lie algebra of the orthogonal group,

$$\mathfrak{o}(n) = \{A \in \mathfrak{g}l(n, \mathbb{R}) : A^T + A = 0\},$$

the Lie algebra of skew-symmetric matrices.

The complex general linear group,

$$GL(n, \mathbb{C}) = \{\ n \times n \text{ matrices } A \text{ with complex entries } : \det A \neq 0\},$$

is also frequently encountered, and its Lie algebra is

$$\mathfrak{g}l(n, \mathbb{C}) \cong T_e G = \{\ n \times n \text{ matrices } A \text{ with complex entries }\},$$

with the usual bracket of matrices as Lie bracket. It can be regarded as a Lie subgroup of $GL(2n, \mathbb{R})$. The *unitary group* is

$$U(n) = \{A \in GL(n, \mathbb{C}) : \bar{A}^T A = I\},$$

and its Lie algebra is

$$\mathfrak{u}(n) = \{A \in \mathfrak{g}l(n, \mathbb{C}) : \bar{A}^T + A = 0\},$$

the Lie algebra of skew-Hermitian matrices, while the *special unitary group*

$$SU(n) = \{A \in U(n) : \det A = 1\}$$

has Lie algebra

$$\mathfrak{s}u(n) = \{A \in \mathfrak{u}(n) : \text{Trace}(A) = 0\}.$$

We can also develop a general linear group based upon the quaternions. The space $\mathbb{H}$ of quaternions can be regarded as the space of complex 2×2 matrices of the form

$$Q = \begin{pmatrix} t + iz & x + iy \\ -x + iy & t - iz \end{pmatrix}, \tag{1.41}$$

where $(t, x, y, z) \in \mathbb{R}^4$ and $i = \sqrt{-1}$. As a real vector space, $\mathbb{H}$ is generated by the four matrices

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{i} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{j} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \mathbf{k} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

the matrix product restricting to the cross product on the subspace spanned by $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$. Thus, for example, $\mathbf{ij} = \mathbf{k}$ in agreement with the cross product. The *conjugate* of a quaternion Q defined by (1.41) is

$$\bar{Q} = \begin{pmatrix} t - iz & -x - iy \\ x - iy & t + iz \end{pmatrix},$$

and

$$\bar{Q}^T Q = (t^2 + x^2 + y^2 + z^2) I = \langle Q, Q \rangle I,$$

where $\langle \;,\; \rangle$ denotes the Euclidean dot product on $\mathbb{H}$.

We can now define

$$GL(n, \mathbb{H}) = \{ \; n \times n \text{ matrices } A \text{ with quaternion entries } \; : \det A \neq 0\};$$

the representation (1.41) showing how this can be regarded as a Lie subgroup of $GL(2n, \mathbb{C})$. Finally, we can define the *compact symplectic group*

$$Sp(n) = \{A \in GL(n, \mathbb{F}) : \bar{A}^T A = I\},$$

where the bar is now defined to be quaternion conjugation of each quaternion entry of A. Note that $Sp(n)$ is a compact Lie subgroup of $U(2n)$, and its Lie algebra is

$$\mathfrak{s}p(n) = \{A \in \mathfrak{g}l(n, \mathbb{C}) : \bar{A}^T + A = 0\},$$

where once again conjugation of A is understood to be quaternion conjugation of each matrix entry.

Lie Group-Lie algebra correspondence: If G and H are Lie groups and $h : G \to H$ is a Lie group homomorphism, we can define a map

$$h_* : \mathfrak{g} \to \mathfrak{h} \quad \text{by} \quad h_*(X) = \beta[(h_*)_e(X(e))],$$

and one can check that this is yields a Lie algebra homomorphism. This gives rise to a "covariant functor" from the category of Lie groups and Lie group homomorphisms to the category of Lie algebras and Lie algebra homomorphisms. A somewhat deeper theorem shows that for any Lie algebra $\mathfrak{g}$ there is a unique simply connected Lie group G with Lie algebra $\mathfrak{g}$. (For example, there is a unique simply connected Lie group corresponding to $\mathfrak{o}(n)$, and this turns out to be a double cover of $SO(n)$ called $\mathrm{Spin}(n)$.) This correspondence between Lie groups and Lie algebras often reduces problems regarding Lie groups to Lie algebras, which are much simpler objects that can be studied via techniques of linear algebra.

A Lie algebra is said to be *simple* if it is nonabelian and has no nontrivial ideals. A compact Lie group is said to be *simple* if its Lie algebra is simple. The compact simply connected Lie groups were classified by Wilhelm Killing and Élie Cartan in the late nineteenth century. In addition to $\mathrm{Spin}(n)$, $SU(n)$ and $Sp(n)$, there are exactly five exceptional Lie groups. The classification of these groups is one of the primary goals of a basic course in Lie group theory.

1.12 Lie groups with biinvariant metrics

It is easiest to compute curvature and geodesics in Riemannian manifolds which have a large group of isometries, such as the space forms that we described before in §1.8. Compact Lie groups also have Riemannian metrics which have large isometry groups. Indeed, the Riemannian metric (1.11) used in classical mechanics for determining the motion of a rigid body is a left-invariant metric on $SO(3)$, that is the diffeomorphism $L_\sigma : SO(3) \to SO(3)$ is an isometry for each $\sigma \in SO(3)$, as one checks by an easy calculation. Even more symmetric are the biinvariant metrics which we study in this section.

Definition. Suppose that G is a Lie group. A pseudo-Riemannian metric on G is *biinvariant* if the diffeomorphisms L_σ and R_σ are isometries for every $\sigma \in G$.

Proposition. *Every compact Lie group has a biinvariant Riemannian metric.*

Proof: We first note that it is easy to construct left-invariant Riemannian metrics on any Lie group. Such a metric is defined by a symmetric bilinear form on the Lie algebra,

$$\langle \cdot, \cdot \rangle : \mathfrak{g} \times \mathfrak{g} \longrightarrow \mathbb{R}.$$

If $n = \dim G$, we can use a basis of left invariant one-forms $(\theta^1, \dots, \theta^n)$ for G to construct a nonzero left invariant n-form $\Theta = \theta^1 \wedge \cdots \wedge \theta^n$. This nowhere zero n-form defines an orientation for G, so if G is compact we can define the *Haar integral* of any smooth function $f : G \to \mathbb{R}$ by

$$(\text{Haar integral of } f) = \int_G f(\sigma) d\sigma = \frac{\int_G f\Theta}{\int_G \Theta}.$$

We can then average a given left invariant metric over right translations defining

$$\langle\langle \cdot, \cdot \rangle\rangle : \mathfrak{g} \times \mathfrak{g} \longrightarrow \mathbb{R} \quad \text{by} \quad \langle\langle X, Y \rangle\rangle = \int_G \langle R_\sigma^* X, R_\sigma^* Y \rangle d\sigma.$$

The resulting averaged metric $\langle\langle \cdot, \cdot \rangle\rangle$ is the sought-after biinvariant metric.

Example 1. We can define a Riemannian metric on $GL(n, \mathbb{R})$ by

$$\langle \cdot, \cdot \rangle = \sum_{i,j=1}^{n} dx_j^i \otimes dx_j^i. \tag{1.42}$$

This is just the Euclidean metric that $GL(n, \mathbb{R})$ inherits as an open subset of $\mathbb{E}^{n^2}$. Although this metric on $GL(n, \mathbb{R})$ is not biinvariant, we claim that the metric it induces on the subgroup $O(n)$ is biinvariant.

To prove this, it suffices to show that the metric (1.42) is invariant under L_A and R_A, when $A \in O(n)$. If $A = (a_j^i) \in O(n)$ and $B = (b_j^i) \in GL(n, \mathbb{R})$, then

$$(x_j^i \circ L_A)(B) = x_j^i(AB) = \sum_{k=1}^{n} x_k^i(A) x_j^k(B) = \sum_{k=1}^{n} a_k^i x_j^k(B),$$

so that

$$L_A^*(x_j^i) = x_j^i \circ L_A = \sum_{k=1}^{n} a_k^i x_j^k.$$

It follows that

$$L_A^*(dx_j^i) = \sum_{k=1}^{n} a_k^i dx_j^k,$$

and hence

$$L_A^*\langle\cdot,\cdot\rangle = \sum_{i,j=1}^n L_A^*(dx_j^i)\otimes L_A^*(dx_j^i)$$

$$= \sum_{i,j,k,l=1}^n a_k^i dx_j^k \otimes a_l^i dx_j^l = \sum_{i,j,k,l=1}^n (a_k^i a_l^i) dx_j^k \otimes dx_j^l.$$

Since $A^T A = I$, $\sum_{i=1}^n a_k^i a_l^i = \delta_{il}$, and hence

$$L_A^*\langle\cdot,\cdot\rangle = \sum_{j,k,l=1}^n \delta_{kl} dx_j^k \otimes dx_j^l = \langle\cdot,\cdot\rangle.$$

By a quite similar computation, one shows that

$$R_A^*\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle, \quad \text{for } A \in O(n).$$

Hence the Riemannian metric defined by (1.42) is indeed invariant under right and left translations by elements of the compact group $O(n)$. Thus (1.42) induces a biinvariant Riemannian metric on $O(n)$, as claimed. Note that if we identify $T_I O(n)$ with the Lie algebra $\mathfrak{o}(n)$ of skew-symmetric matrices, this Riemannian metric is given by

$$\langle X, Y\rangle = \text{Trace}(X^T Y), \quad \text{for} \quad X, Y \in \mathfrak{o}(n).$$

Example 2. The unitary group $U(n)$ is an imbedded subgroup of $GL(2n,\mathbb{R})$ which lies inside $O(2n)$, and hence if $\langle\cdot,\cdot\rangle_E$ is the Euclidean metric induced on $GL(2n,\mathbb{R})$,

$$L_A^*\langle\cdot,\cdot\rangle_E = \langle\cdot,\cdot\rangle_E = R_A^*\langle\cdot,\cdot\rangle_E, \quad \text{for } A \in U(n).$$

Thus the Euclidean metric on $GL(2n,\mathbb{R})$ induces a biinvariant Riemannian metric on $U(n)$. If we identify $T_I U(n)$ with the Lie algebra $\mathfrak{u}(n)$ of skew-Hermitian matrices, one can check that this Riemannian metric is given by

$$\langle X, Y\rangle = 2\text{Re}\left(\text{Trace}(X^T \bar{Y})\right), \quad \text{for} \quad X, Y \in \mathfrak{u}(n). \tag{1.43}$$

Example 3. The compact symplectic group $Sp(n)$ is an imbedded subgroup of $GL(4n,\mathbb{C})$ which lies inside $O(4n)$, and hence if $\langle\cdot,\cdot\rangle_E$ is the Euclidean metric induced on $GL(4n,\mathbb{R})$,

$$L_A^*\langle\cdot,\cdot\rangle_E = \langle\cdot,\cdot\rangle_E = R_A^*\langle\cdot,\cdot\rangle_E, \quad \text{for } A \in Sp(n).$$

Thus the Euclidean metric on $GL(4n,\mathbb{R})$ induces a biinvariant Riemannian metric on $Sp(n)$.

Proposition. *Suppose that G is a Lie group with a biinvariant pseudo-Riemannian metric $\langle\cdot,\cdot\rangle$. Then*

1. *geodesics passing through the identity $e \in G$ are just the one-parameter subgroups of G,*

2. *the Levi-Civita connection on TG is defined by*

$$\nabla_X Y = \frac{1}{2}[X, Y], \quad \text{for } X, Y \in \mathfrak{g},$$

3. *the curvature tensor is given by*

$$\langle R(X, Y)W, Z\rangle = \frac{1}{4}\langle [X, Y], [Z, W]\rangle, \quad \text{for } X, Y, Z, W \in \mathfrak{g}. \tag{1.44}$$

Before proving this, we need to some facts about the Lie bracket that are proven in [12]. Recall that if X is a vector field on a smooth manifold M with one-parameter group of local diffeomorphisms $\{\phi_t : t \in \mathbb{R}\}$ and Y is a second smooth vector field on M, then the Lie bracket $[X, Y]$ is determined by the formula

$$[X, Y](p) = -\left.\frac{d}{dt}\left((\phi_t)_*(Y)(p)\right)\right|_{t=0}. \tag{1.45}$$

Definition. A vector field X on a pseudo-Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$ is said to be *Killing* if its one-parameter group of local diffeomorphisms $\{\phi_t : t \in \mathbb{R}\}$ consists of isometries.

The formula (1.45) for the Lie bracket has the following consequence needed in the proof of the theorem:

Lemma. *If X is a Killing field, then*

$$\langle \nabla_Y X, Z\rangle + \langle Y, \nabla_Z X\rangle = 0, \quad \text{for } Y, Z \in \mathcal{X}(M).$$

Proof: Note first that if X is fixed

$$\langle \nabla_Y X, Z\rangle(p) \quad \text{and} \quad \langle Y, \nabla_Z X\rangle(p)$$

depend only on $Y(p)$ and $Z(p)$. Thus we can assume without loss of generality that $\langle Y, Z\rangle$ is constant. Then, since X is Killing, $\langle (\phi_t)_*(Y), (\phi_t)_*(Z)\rangle$ is constant, and

$$0 = \left\langle \left.\frac{d}{dt}\left((\phi_t)_*(Y)\right)\right|_{t=0}, Z\right\rangle + \left\langle Y, \left.\frac{d}{dt}\left((\phi_t)_*(Z)\right)\right|_{t=0}\right\rangle$$
$$= -\langle [X, Y], Z\rangle - \langle Y, [X, Z]\rangle.$$

On the other hand, since ∇ is the Levi-Civita connection,

$$0 = X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle.$$

Adding the last two equations yields the statement of the lemma.

Application. If X is a Killing field on the pseudo-Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$ and $\gamma : (a, b) \to M$ is a geodesic, then since $\langle \nabla_Y X, Y \rangle = 0$,

$$\frac{d}{dt}\langle \gamma', X \rangle = \gamma' \langle \gamma', X \rangle = \langle \nabla_{\gamma'} \gamma', X \rangle + \langle \gamma', \nabla_{\gamma'} X \rangle = 0,$$

where we think of γ' as a vector field defined along γ. Thus $\langle \gamma', X \rangle$ is constant along the geodesic. This often gives very useful constraints on the qualitative behavior of geodesic flow.

We now turn to the proof of the Proposition: First note that since the metric $\langle \cdot, \cdot \rangle$ is left invariant,

$$X, Y \in \mathfrak{g} \quad \Rightarrow \quad \langle X, Y \rangle \quad \text{is constant.}$$

Since the metric is right invariant, each $R_{\theta_X(t)}$ is an isometry, and hence X is a Killing field. Thus

$$\langle \nabla_Y X, Z \rangle + \langle \nabla_Z X, Y \rangle = 0, \quad \text{for } X, Y, Z \in \mathfrak{g}.$$

In particular,

$$\langle \nabla_X X, Y \rangle = -\langle \nabla_Y X, X \rangle = -\frac{1}{2} Y \langle X, X \rangle = 0.$$

Thus $\nabla_X X = 0$ for $X \in \mathfrak{g}$ and the integral curves of X must be geodesics.

Next note that

$$0 = \nabla_{X+Y}(X + Y) = \nabla_X X + \nabla_X Y + \nabla_Y X + \nabla_Y Y = \nabla_X Y + \nabla_Y X.$$

Averaging the equations

$$\nabla_X Y + \nabla_Y X = 0, \qquad \nabla_X Y - \nabla_Y X = [X, Y]$$

yields the second assertion of the proposition.

Finally, if $X, Y, Z \in \mathfrak{g}$, use of the Jacobi identity yields

$$\begin{aligned} R(X,Y)Z &= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \\ &= \frac{1}{4}[X,[Y,Z]] - \frac{1}{4}[Y,[X,Z]] - \frac{1}{2}[[X,Y],Z] = -\frac{1}{4}[[X,Y],Z]. \end{aligned}$$

On the other hand, if $X, Y, Z \in \mathfrak{g}$,

$$0 = 2X\langle Y, Z \rangle = 2\langle \nabla_X Y, Z \rangle + 2\langle Y, \nabla_X Z \rangle = \langle [X,Y], Z \rangle + \langle Y, [X,Z] \rangle.$$

Thus we conclude that

$$\langle R(X,Y)W, Z \rangle = -\frac{1}{4}\langle [[X,Y],W], Z \rangle = \frac{1}{4}\langle [X,Y],[Z,W] \rangle,$$

finishing the proof of the third assertion.

Remark. If G is a Lie group with a biinvariant pseudo-Riemannian metric, the map

$$\nu : G \to G \quad \text{defined by} \quad \nu(\sigma) = \sigma^{-1},$$

is an isometry. Indeed, it is immediate that $(\nu_*)_e = -\mathrm{id}$ is an isometry, and the identity

$$\nu = R_{\sigma^{-1}} \circ \nu \circ L_{\sigma^{-1}}$$

shows that $(\nu_*)_\sigma$ is an isometry for each $\sigma \in G$. Thus ν is an isometry of G which reverses geodesics through the identity e. More generally, the map $I_\sigma = L_{\sigma^{-1}} \circ \nu \circ L_\sigma$ is an isometry which reverses geodesics through σ

Definition. A *Riemannian symmetric space* is a Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$ such that for each $p \in M$ there is an isometry $I_p : M \to M$ which fixes p and reverses geodesics through p.

Examples include not just the Lie groups with biinvariant Riemannian metrics and the spaces of constant curvature, but also complex projective space with the metric described in the next section.

1.13 Projective spaces; Grassmann manifolds

We have jsut seen that if G is a compact Lie group with a biinvariant Riemannian metric $\langle \cdot, \cdot \rangle$, it is easy to compute the geodesics in G and the curvature of G. The Riemannian symmetric spaces defined at the end of the previous section provide a more general class of Riemannian manifolds fin which one can easily calculate geodesics and curvature. In 1926-27, Élie Cartan completely classified the Riemannian symmetric spaces (the classification is presented in [9]), and these provide a treasure box of examples on which one can test possible conjectures. We give only the briefest introduction to this theory, and consider a few symmetric spaces that can be realized as submanifolds $M \subseteq G$ with the metric induced from a biinvariant Riemannian metric on G. An important case is the complex projective space with its "Fubini-Study" metric, a space which plays a central role in algebraic geometry.

We assume as known the basic theory of homogeneous spaces, as described in Chapter 9 of [12]. If G is a Lie group and H is a compact subgroup, the *homogeneous space* of left cosets G/H is a smooth manifold and the projection $\pi : G \to G/H$ is a smooth submersion. Moreover, the map $G \times G/H \to G/H$, defined by $(\sigma, \tau H) \to \sigma\tau H$, is smooth.

Suppose therefore that G is a compact Lie group with a biinvariant Riemannian metric, that $s : G \to G$ is a group homomorphism such that $s^2 = \mathrm{id}$, and that

$$H = \{\sigma \in G : s(\sigma) = \sigma\},$$

a compact Lie subgroup of G. In this case, the group homomorphism s induces a Lie algebra homomorphism $s_* : \mathfrak{g} \to \mathfrak{g}$ such that $s_*^2 = \mathrm{id}$. We let

$$\mathfrak{h} = \{X \in \mathfrak{g} : s_*(X) = X\}, \quad \mathfrak{p} = \{X \in \mathfrak{g} : s_*(X) = -X\}.$$

Then $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$ is a direct sum decomposition, and the fact that s_* is a Lie algebra homomorphism implies that

$$[\mathfrak{h}, \mathfrak{h}] \subseteq \mathfrak{h}, \quad [\mathfrak{h}, \mathfrak{p}] \subseteq \mathfrak{p}, \quad [\mathfrak{p}, \mathfrak{p}] \subseteq \mathfrak{h}.$$

Finally, note that $\mathfrak{h}$ is the Lie algebra of H and hence is isomorphic to the tangent space to H at the identity e, while $\mathfrak{p}$ is isomorphic to the tangent space to G/H at eH.

Under these conditions we can define a map

$$F : G/H \to G \quad \text{by} \quad F(\sigma H) = \sigma s(\sigma^{-1}). \tag{1.46}$$

Note that if $h \in H$ then $F(\sigma h) = \sigma h s(h^{-1}\sigma^{-1}) = \sigma s(\sigma^{-1})$, so F is a well-defined map on the homogeneous space G/H.

Lemma. *The map F defined by (1.46) is an imbedding. Moreover, the geodesics for the induced Riemannian metric*

$$\langle \cdot, \cdot \rangle_{G/H} = F^* \langle \cdot, \cdot \rangle_G \tag{1.47}$$

on G/H are just the left translates of one-parameter subgroups of G which are tangent to $F(G/H)$ at some point of $F(G/H)$.

Since G/H is compact, we need only show that F is a one-to-one immersion. But

$$\sigma s(\sigma^{-1}) = \tau s(\tau^{-1}) \quad \Leftrightarrow \quad \tau^{-1}\sigma = s(\tau^{-1}\sigma) \quad \Leftrightarrow \quad \tau^{-1}\sigma \in H,$$

so F is one-to-one. To see that F is an immersion, one first checks that

$$X \in \mathfrak{p} \quad \Rightarrow \quad s\left(e^{-tX}e^{-uX}\right) = s\left(e^{-tX}\right) s\left(e^{-uX}\right) \quad \Rightarrow \quad t \mapsto s(e^{-tX})$$

is a one-parameter group and checking the derivative at $t = 0$ shows that $s(e^{-tX}) = e^{tX}$ and hence $F(e^{tX}H) = e^{2tX}$. Thus

$$(F_*)_{eH} : T_{eH}(G/H) \cong \mathfrak{p} \longrightarrow T_e G \cong \mathfrak{g}$$

is injective. Moreover,

$$F(\sigma e^{tX}H) = \sigma e^{tX} s(e^{-tX}) s(\sigma^{-1}) = L_\sigma \circ R_{\sigma^{-1}}(e^{2tX}), \quad \text{for } X \in \mathfrak{p}, \tag{1.48}$$

and since L_σ and $R_{\sigma^{-1}}$ are diffeomorphisms, this quickly implies that $(F_*)_{\sigma H}$ is injective for each $\sigma H \in G/H$, so F is indeed an immersion.

Thus $F(G/H)$ can be thought of as an imbedded submanifold of G. The curves

$$\gamma(t) = F(\sigma e^{tX}H) \quad \text{for } \sigma \in G \text{ and } X \in \mathfrak{p} \tag{1.49}$$

are images under isometries of one-parameter subgroups of G by (1.48), and hence geodesics in G which lie within G/H. For simplicity of notation, we now identify G/H with $F(G/H)$. If ∇^G and $\nabla^{G/H}$ are the Levi-Civita connections on G and G/H respectively, and $(\cdot)^\top$ is the projection from TG to $T(G/H)$, then each curve (1.49) satisfies

$$\nabla^{G/H}_{\gamma'(t)}\gamma'(t) = \left(\nabla^G_{\gamma'(t)}\gamma'(t)\right)^\top = 0,$$

and hence is also a geodesic within G/H, verifying the last statement of the Lemma.

The submanifolds $H \subseteq G$ and G/H intersect transversely at $e \in G$. We will see later that these submanifolds are in fact generated by the one-parameter subgroups θ_X emanating from e within G. When X ranges over $\mathfrak{p}$, the one-parameter subgroups θ_X cover H, while when X ranges over $\mathfrak{p}$ these one-parameter subgroups cover G/H.

Finally, note that the Lie group G acts as a group of isometries on G/H. Moreover, the isometry s of G takes G/H to itself, and restricts to

$$s_{eH} : G/H \to G/H \quad \text{defined by} \quad s_{eH}(\sigma H) = s(\sigma)H,$$

an isometry which reverses geodesics through the point eH. Using the transitive group G of isometries one gets an isometry reversing geodesics through any point σH in G/H, demonstrating that G/H is indeed a Riemannian symmetric space.

Example 1. Suppose $G = O(n)$ and s is conjugation with the element

$$I_{p,q} = \begin{pmatrix} -I_{p\times p} & 0 \\ 0 & I_{q\times q} \end{pmatrix}, \quad \text{where} \quad p + q = n.$$

Thus

$$s(A) = I_{p,q} A I_{p,q}, \quad \text{for } A \in O(n),$$

and it is easily verified that s preserves the biinvariant metric and is a group homomorphism. In this case $H = O(p) \times O(q)$, and the quotient $O(n)/O(p) \times O(q)$ is the *Grassmann manifold* of real p-planes in n-space. The special case $O(n)/O(1) \times O(n-1)$ of one-dimensional subspaces of $\mathbb{R}^n$ is also known as real projective space $\mathbb{R}P^{n-1}$.

Example 2. Suppose $G = U(n)$ and s is conjugation with the element

$$I_{p,q} = \begin{pmatrix} -I_{p\times p} & 0 \\ 0 & I_{q\times q} \end{pmatrix}, \quad \text{where} \quad p + q = n.$$

In this case $H = U(p) \times U(q)$ and the quotient $U(n)/U(p) \times U(q)$ is the *Grassmann manifold* of complex p-planes in n-space. The special case $U(n)/U(1) \times U(n-1)$ of complex one-dimensional subspaces of $\mathbb{C}^n$ is also known as complex projective space $\mathbb{C}P^{n-1}$.

Example 3. Finally, suppose $G = Sp(n)$ and s is conjugation with the element

$$I_{p,q} = \begin{pmatrix} -I_{p\times p} & 0 \\ 0 & I_{q\times q} \end{pmatrix}, \quad \text{where} \quad p + q = n.$$

In this case $H = Sp(p) \times Sp(q)$ and the quotient $Sp(n)/Sp(p) \times Sp(q)$ is the *Grassmann manifold* of quaternionic p-planes in n-space. The special case $Sp(n)/Sp(1) \times Sp(n-1)$ of complex one-dimensional subspaces of $\mathbb{H}^n$ is also known as quaternionic projective space $\mathbb{H}P^{n-1}$.

Curvature Theorem. *Suppose that G is a compact Lie group with a biinvariant Riemannian metric, $s : G \to G$ a Lie automorphism such that $s^2 = id$ and that $H = \sigma \in G : s(\sigma) = \sigma\}$. Then the curvature of the Riemannian metric on G/H defined by (1.47) is given by*

$$\langle R(X,Y)W, Z\rangle = \langle [X,Y], [Z,W]\rangle, \quad \textit{for } X, Y, Z, W \in T_{eH}(G/H) \cong \mathfrak{p}. \quad (1.50)$$

Indeed, this curvature formula follows from the Gauss equation for a submanifold M of a Riemannian manifold $(N, \langle\cdot,\cdot\rangle)$, when M is given the induced submanifold metric. To prove this extended Gauss equation, one follows the discussion already given in §1.8, except that we replace the ambient Euclidean space $\mathbb{E}^N$ with a general Riemannian manifold $(N, \langle\cdot,\cdot\rangle)$.

Thus if $p \in M \subseteq N$ and $v \in T_pN$, we let

$$v = v^\top + v^\perp, \quad \text{where} \quad v^\top \in T_pM \quad \text{and} \quad v^\perp \perp T_pM,$$

$(\cdot)^\top$ and $(\cdot)^\perp$ being the orthogonal projection into the tangent space and normal space. The Levi-Civita connection ∇^M on M is then defined by the formula

$$(\nabla^M_X Y)(p) = (\nabla^N_X Y(p))^\top,$$

where ∇^N is the Levi-Civita connection on N. If we let $\mathcal{X}^\perp(M)$ denote the vector fields in N which are defined at points of M and are perpendicular to M, then we can define the second fundamental form

$$\alpha : \mathcal{X}(M) \times \mathcal{X}(M) \to \mathcal{X}^\perp(M) \quad \text{by} \quad \alpha(X,Y) = (\nabla^N_X Y(p))^\perp.$$

As before, it satisfied the identities:

$$\alpha(fX,Y) = \alpha(X,fY) = f\alpha(X,Y), \quad \alpha(X,Y) = \alpha(Y,X).$$

If $\gamma : (a,b) \to M \subseteq \mathbb{E}^N$ is a unit speed curve, we call $(\nabla^N_{\gamma'}\gamma')$ the geodesic curvature of γ in N, while

$$\left(\nabla^N_{\gamma'}\gamma'\right)^\top = \nabla^M_{\gamma'}\gamma' = (\text{geodesic curvature of } \gamma \text{ in } M),$$

$$\left(\nabla^N_{\gamma'}\gamma'\right)^\perp = \alpha(\gamma,\gamma') = (\text{normal curvature of } \gamma).$$

With these preparations, we can now state:

Extended Gauss Theorem. The curvature tensor R of a submanifold M of a Riemannian manifold N with the induced Riemannian metric is given by the *Gauss equation*

$$\langle R^M(X,Y)W,Z\rangle = \langle R^N(X,Y)W,Z\rangle \\ + \langle \alpha(X,Z),\alpha(Y,W)\rangle - \langle \alpha(X,W),\alpha(Y,Z)\rangle, \quad (1.51)$$

where X, Y, Z and W are elements of $\mathcal{X}(M)$.

The proof of the extended Gauss equation (1.51) is just like that of the previous Gauss equation, except that we replace the equation (1.33) with

$$\nabla^N_X\nabla^N_Y W - \nabla^N_Y\nabla^N_X W - \nabla^N_{[X,Y]}W = \langle R^N(X,Y)W,Z\rangle,$$

and then follow exactly the same steps as before.

Proof of the Curvature Theorem: We simply note that since geodesics in G/H are also geodesics in the ambient manifold G, the second fundamental form α vanishes, so (1.50) follows immediately from (1.51).

Exercise IV. Due Wednesday, May 4. We consider the special case of the above construction in which $G = U(n)$ and s is conjugation with

$$I_{1,n-1} = \begin{pmatrix} -1 & 0 \\ 0 & I_{(n-1)\times(n-1)} \end{pmatrix},$$

so that the fixed point set of the automorphism s is $H = U(1) \times U(n-1)$ and $G/H = \mathbb{C}P^{n-1}$.

a. Recall that the Lie algebra $\mathfrak{u}(n)$ divides into a direct sum $\mathfrak{u}(n) = \mathfrak{h} \oplus \mathfrak{p}$, where

$$\mathfrak{h} = \{X \in \mathfrak{g} : s_*(X) = X\}, \quad \mathfrak{p} = \{X \in \mathfrak{g} : s_*(X) = -X\},$$

and $\mathfrak{h}$ is just the Lie algebra of $U(1) \times U(n-1)$. Consider two elements

$$X = \begin{pmatrix} 0 & -\bar{\xi}_2 & \cdots & -\bar{\xi}_n \\ \xi_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdot \\ \xi_n & 0 & \cdots & 0 \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} 0 & -\bar{\eta}_2 & \cdots & -\bar{\eta}_n \\ \eta_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdot \\ \eta_n & 0 & \cdots & 0 \end{pmatrix}$$

of $\mathfrak{p}$, and determine their Lie bracket $[X,Y] \in \mathfrak{h}$.

SOLUTION: To simplify notation, we write

$$X = \begin{pmatrix} 0 & -\bar{\xi}^T \\ \xi & 0 \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} 0 & -\bar{\eta}^T \\ \eta & 0 \end{pmatrix}, \quad (1.52)$$

where ξ and η are column vectors in $\mathbb{C}^{n-1}$. Then ordinary matrix multiplication shows that

$$XY = \begin{pmatrix} -\bar{\xi}^T\eta & 0 \\ 0 & -\xi\bar{\eta}^T \end{pmatrix}, \quad YX = \begin{pmatrix} -\bar{\eta}^T\xi & 0 \\ 0 & -\eta\bar{\xi}^T \end{pmatrix},$$

so that

$$[X,Y] = \begin{pmatrix} -\bar{\xi}^T\eta + \bar{\eta}^T\xi & 0 \\ 0 & -\xi\bar{\eta}^T + \eta\bar{\xi}^T \end{pmatrix}.$$

b. Use the formula for curvature of G/H to show that the sectional curvatures $K(\sigma)$ for $\mathbb{C}P^{n-1}$ satisfy the inequalities $a^2 \leq K(\sigma) \leq 4a^2$ for some $a^2 > 0$.

SOLUTION: As inner product on $T_I U(n)$, we use

$$\langle A, B\rangle = \frac{1}{2}\mathrm{Re}\left(\mathrm{Trace}(A^T\bar{B})\right), \quad \text{for} \quad A, B \in \mathfrak{u}(n).$$

This differs by a constant factor from the Riemannian metric induced by the natural imbedding into $\mathbb{E}^{(2n)^2}$, but with the rescaled metric

$$\langle X, Y\rangle = \mathrm{Re}(\xi^T\bar{\eta}),$$

when X and Y are given by (1.52). To simplify the calculations, assume that

$$\langle X, X\rangle = \langle Y, Y\rangle = 1, \quad \text{and} \quad \langle X, Y\rangle = 0.$$

Then

$$|\xi|^2 = |\eta|^2 = 1 \quad \text{and} \quad \xi^T\bar{\eta} = -\eta^T\bar{\xi} \in \sqrt{-1}\mathbb{R} \subseteq \mathbb{C};$$

in other words, $\xi^T\bar{\eta}$ is purely imaginary. Then

$$\begin{aligned} &\langle [X,Y],[X,Y]\rangle \\ &\quad = \frac{1}{2}\mathrm{Trace}\begin{pmatrix} -\eta^T\bar{\xi} + \xi^T\bar{\eta} & 0 \\ 0 & -\bar{\eta}\xi^T + \bar{\xi}\eta^T \end{pmatrix}\begin{pmatrix} -\xi^T\bar{\eta} + \eta^T\bar{\xi} & 0 \\ 0 & -\bar{\xi}\eta^T + \bar{\eta}\xi^T \end{pmatrix} \\ &\qquad = \frac{1}{2}\mathrm{Trace}\begin{pmatrix} 4\left|\mathrm{Im}(\xi^T\bar{\eta})\right|^2 & 0 \\ 0 & (-\bar{\eta}\xi^T + \bar{\xi}\eta^T)(-\bar{\xi}\eta^T + \bar{\eta}\xi^T) \end{pmatrix} \\ &\qquad\qquad = 2\left|\mathrm{Im}(\xi^T\bar{\eta})\right|^2 + |\xi|^2|\eta|^2 + \left|\mathrm{Im}(\xi^T\bar{\eta})\right|^2 \\ &\qquad\qquad\qquad = |\xi|^2|\eta|^2 + 3\left|\mathrm{Im}(\xi^T\bar{\eta})\right|^2. \end{aligned}$$

The last expression ranges between 1 and 4, and it follows from the Cauchy-Schwarz inequality that it achieves its maximum when $\eta = i\xi$. Thus if σ is the two-plane spanned by X and Y,

$$K(\sigma) = \frac{\langle [X,Y],[X,Y]\rangle}{\langle X,X\rangle\langle Y,Y\rangle - \langle X,Y\rangle^2} = \left[|\xi|^2|\eta|^2 + 3\left|\mathrm{Im}(\xi^T\bar{\eta})\right|^2\right]$$

lies in the interval $[1,4]$, achieving both extreme values when $n-1 \geq 2$.

Remark. Once one has complex projective space $\mathbb{C}P^{n-1}$ with its Fubini-Study metric (the unique Riemannian metric up to scale factor which is invariant under the action of $U(n)$) one can construct a host of new examples of Riemannian manifolds, namely the complex analytic submanifolds of $\mathbb{C}P^{n-1}$. A famous theorem of Chow states these are algebraic varieties without singularities, that is, each such submanifold can be represented as the zero locus of a finite number of homogeneous polynomials with complex coefficients. This brings us into contact with two major subjects within contemporary mathematics, Kähler geometry and algebraic geometry over the complex field, both of which are treated in the beautiful text by Griffiths and Harris [6].

Chapter 2

Normal coordinates

2.1 Definition of normal coordinates

Our next goal is to develop a system of local coordinates centered at a given point p in a Riemannian manifold which are as Euclidean as possible. Such coordinates can also be constructed for pseudo-Riemannian manifolds, and their construction is based upon the following series of propositions.

Proposition 1. *Suppose that $(M, \langle \cdot, \cdot \rangle)$ is a pseudo-Riemannian manifold and $p \in M$. Then there is an open neighborhood V of 0 in T_pM such that if $v \in T_pM$ the unique geodesic γ_v in M which satisfies the initial conditions $\gamma_v(0) = p$ and $\gamma_v'(0) = v$ is defined on the interval $[0, 1]$.*

Proof: According to ODE theory applied to the second-order system of differential equations

$$\frac{d^2x^i}{dt^2} + \sum_{j,k=1}^{n} \Gamma^i_{jk} \frac{dx^j}{dt} \frac{dx^k}{dt} = 0,$$

there is a neighborhood W of 0 in T_pM and an $\epsilon > 0$ such that the geodesic γ_w is defined on $[0, \epsilon]$ for all $w \in W$. Let $V = \epsilon W$. Then if $v \in V$, $v = \epsilon w$ for some $w \in W$, and since $\gamma_v(t) = \gamma_w(\epsilon t)$, γ_v is defined on $[0, 1]$, proving the proposition.

Definition. The *exponential map*

$$\exp_p : V \to M \quad \text{is defined by} \quad \exp_p(v) = \gamma_v(1).$$

Remark. Note that if G is a compact subgroup of $GL(n, \mathbb{R})$ with a biinvariant Riemannian metric $\langle \cdot, \cdot \rangle$ as constructed in §1.12,

$$\exp_I A = e^{tA}, \quad \text{for } A \in T_I G \cong \mathfrak{g} \subseteq \mathfrak{gl}(n, \mathbb{R}).$$

Similarly, if G is a compact subgroup of $GL(n,\mathbb{R})$ with a biinvariant Riemannian metric, $s : G \to G$ is a group homomorphism such that $s^2 = \text{id}$, and that

$$H = \{\sigma \in G : s(\sigma) = \sigma\},$$

we can divide the Lie algebra $\mathfrak{g}$ of G into a direct sum $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$, where $\mathfrak{h}$ is the Lie algebra of H and $\mathfrak{p} \cong T_{eH}(G/H)$. In this case, the imbedding

$$F : G/H \to G, \quad F(\sigma H) = \sigma s(\sigma^{-1})$$

realizes G/H as a submanifold of G, and

$$\exp_{eH} A = e^{tA}, \quad \text{for } A \in \mathfrak{p} \cong T_{eH}(G/H).$$

These facts explain the origin of the term "exponential map."

Note that if $v \in V$, $t \mapsto \exp_p(tv)$ is a geodesic (because $\exp_p(tv) = \gamma_{tv}(1) = \gamma_v(t)$), and hence $\exp_p$ takes straight line segments through the origin in T_pM to geodesic segments through p in M.

Proposition 2. *There is an open neighborhood $\tilde{U}$ of 0 in T_pM which $\exp_p$ maps diffeomorphically onto an open neighborhood U of p in M.*

Proof: By the inverse function theorem, it will suffice to show that

$$((\exp_p)_*)_0 : T_0(T_pM) \longrightarrow T_pM$$

is an isomorphism. We identify $T_0(T_pM)$ with T_pM. If $v \in T_pM$, define

$$\lambda_v : \mathbb{R} \to T_pM \quad \text{by} \quad \lambda_v(t) = tv.$$

Then $\lambda_v'(0) = v$ and

$$\begin{aligned} ((\exp_p)_*)_0(v) = ((\exp_p)_*)_0(\lambda_v'(0)) = (\exp_p) \circ \lambda_v)'(0) \\ = \frac{d}{dt}(\exp_p(tv))\bigg|_{t=0} = \frac{d}{dt}(\gamma_v(t))\bigg|_{t=0} = v, \end{aligned}$$

so $((\exp_p)_*)_0$ is indeed an isomorphism.

It will sometimes be useful to have a stronger version of the above proposition, proven by the same method, but making use of the map

$$\exp : (\text{neighborhood of 0-section in } TM) \longrightarrow M \times M,$$

defined by

$$\exp(v) = (p, \exp_p(v)), \quad \text{for } v \in T_pM.$$

Proposition 3. *Given a point $p_0 \in M$ there is an open neighborhood $\tilde{W}$ of the zero vector 0 of $T_{p_0}M$ within TM which exp maps diffeomorphically onto an open neighborhood W of (p_0, p_0) in $M \times M$.*

Proof: If 0 denotes the zero vector in $T_{p_0}M$, it suffices to show that

$$(\exp_*)_0 : T_0(TM) \longrightarrow T_{(p_0,p_0)}(M \times M)$$

is an isomorphism. Since both vector spaces have the same dimension it suffices to show that $(\exp_*)_0$ is an epimorphism. Let

$$\pi_1 : M \times M \to M, \qquad \pi_2 : M \times M \to M$$

denote the projections on the first and second factors, respectively. Then $\pi_i \circ \exp : TM \to M$ is the bundle projection $TM \to M$ and hence $((\pi_1 \circ \exp)_*)_0$ is an epimorphism. On the other hand, the composition

$$T_{p_0}M \subseteq TM \xrightarrow{\exp} M \times M \xrightarrow{\pi_2} M$$

is just $\exp_{p_0}$ and hence $((\pi_2 \circ \exp)_*)_0$ is an epimorphism by the previous proposition. Hence $(\exp_*)_0$ is indeed an epimorphism as claimed.

Corollary 4. *Suppose that $(M, \langle\cdot,\cdot\rangle)$ is a Riemannian manifold and $p_0 \in M$. Then there is an open neighborhood U of p_0 and an $\epsilon > 0$ such that $\exp_p$ maps*

$$\{v \in T_pM : \langle v, v\rangle < \epsilon^2\}$$

diffeomorphically onto an open subset of M for all $p \in U$.

If $(M, \langle\cdot,\cdot\rangle)$ is a Riemannian manifold and $p \in M$. If we choose a basis $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ for T_pM, orthonormal with respect to the inner product $\langle\cdot,\cdot\rangle_p$, we can define "flat" coordinates $(\dot{x}^1, \dots, \dot{x}^n)$ on T_pM by

$$\dot{x}^i(v) = a^i \quad \Leftrightarrow \quad v = \sum_{i+1}^{n} a^i \mathbf{e}_i.$$

If U is an open neighborhood of $p \in M$ such that $\exp_p$ maps an open neighborhood $\tilde{U}$ of $0 \in T_pM$ diffeomorphically onto U, we can define coordinates

$$(x^1, \dots, x^n) : U \to \mathbb{R}^n \quad \text{by} \quad x^i \circ \exp_p = \dot{x}^i, .$$

which we call *Riemannian normal coordinates* centered at p, or sometimes simply *normal coordinates*.

If $(M, \langle\cdot,\cdot\rangle)$ is a Lorentz manifold as described in §1.7, which we study in units for which the speed of light is one, we can choose linear coordinates $(\dot{x}^0, \dot{x}^1, \dots \dot{x}^n)$ on T_pM so that the Lorentz metric on T_pM is represented by

$$\langle\cdot,\cdot\rangle = -d\dot{x}^0 \otimes d\dot{x}^0 + d\dot{x}^1 \otimes d\dot{x}^1 + \cdots + d\dot{x}^n \otimes d\dot{x}^n = \sum_{i,j=0}^{n} \eta_{ij} d\dot{x}^i \otimes d\dot{x}^j,$$

where

$$(\eta_{ij}) = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{2.1}$$

These coordinates on T_pM then project to coordinates $(x^0, x^1, \ldots x^n)$ defined on an open neighborhood $U = \exp_p(V)$ of $p \in M$ which we call *Lorentz normal coordinates*.

More generally, we could define normal coordinates for a pseudo-Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$ of arbitrary signature.

Let us focus first on the Riemannian case, and suppose that in terms of Riemannian normal coordinates centered at p, we have

$$\langle\cdot,\cdot\rangle = \sum_{i,j=1}^{n} g_{ij} dx^i \otimes dx^j.$$

It is interesting to determine the Taylor series expansion of the g_{ij}'s about p.

Of course, we have $g_{ij}(p) = \delta_{ij}$. To evaluate the first order derivatives, we note that whenever a^1, ..., a^n are constants, the curve γ defined by

$$x^i \circ \gamma(t) = a^i t$$

is a geodesic in M by definition of the exponential map. Thus the functions $x^i = x^i \circ \gamma$ must satisfy the geodesic equation

$$\ddot{x}^k + \sum_{i,j=1}^{n} \Gamma^k_{ij} \dot{x}^i \dot{x}^j = 0.$$

Substitution into this equation yields

$$\sum_{i,j=1}^{n} \Gamma^k_{ij}(p) a^i a^j = 0.$$

Since this holds for all choices of the constants $(a^1, \ldots, a^n)$ we conclude that $\Gamma^k_{ij}(p) = 0$. It then follows from (1.20) that

$$\frac{\partial g_{ij}}{\partial x^k}(p) = 0.$$

Later we will see that the Taylor series for the Riemannian metric in normal coordinates centered at p is given by

$$g_{ij} = \delta_{ij} - \frac{1}{3} \sum_{k,l=1}^{n} R_{ikjl}(p) x^k x^l + (\text{higher order terms}).$$

This formula gives a very explicit formula for how much a Riemannian metric differs from the Euclidean metric near a given point p. Note that when we look at a neighborhood of p under higher and higher magnification, it looks more and more like flat Euclidean space, the curvature measuring the deviation from flatness.

There is a similar Taylor series for Lorentz manifolds,

$$g_{ij} = \eta_{ij} - \frac{1}{3} \sum_{k,l=1}^{n} R_{ikjl}(p) x^k x^l + (\text{higher order terms}),$$

where the η_{ij}'s are the components of the flat Lorentz metric in Minkowski space-time defined by (2.1). In this case, when we look at a neighborhood of an event p in a curved space-time under higher and higher magnification, it looks more and more like flat Minkowski space-time in what is called an "inertial frame." These inertial frames represent coordinate systems in which gravitational forces vanish up to second order. Thus a "freely falling" space station in outer space is in an innertial frame and for experiments inside the space station the gravitational force represented by the Christoffel symbols vanishes. The curvature tensor represents tidal forces that act over large distances. They would become evident to an observer falling into the singularity in the middle of a black hole, the tidal forces pulling an arm in one direction, a leg in another, until all classical forms of matter are destroyed.

Before establishing these Taylor series expansions, we will need the so-called Gauss lemma.

2.2 The Gauss Lemma

Riemannian case. Suppose that $(x^1, \ldots, x^n)$ are normal coordinates centered at a point p in a Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$, and defined on an open neighborhood U of p. We can then define a radial function

$$r : U \to \mathbb{R} \quad \text{by} \quad r = \sqrt{(x^1)^2 + \cdots + (x^n)^2},$$

and a radial vector field S on $U - \{p\}$ by

$$S = \sum_{i=1}^{n} \frac{x^i}{r} \frac{\partial}{\partial x^i}.$$

For $1 \leq i, j \leq n$, let E_{ij} be the rotation vector field on U defined by

$$E_{ij} = x^i \frac{\partial}{\partial x^j} - x^j \frac{\partial}{\partial x^i}. \tag{2.2}$$

Lemma 1. $[E_{ij}, S] = 0$.

Proof: This can be verified by direct calculation. For a more conceptual argument, one can note that the one-parameter group of local diffeomorphisms $\{\phi_t : t \in \mathbb{R}\}$ on U induced by E_{ij} consists of rotations in terms of the normal coordinates, so $(\phi_t)_*(S) = S$, so

$$[E_{ij}, S] = -\frac{d}{dt}\left((\phi_t)_*(S)\right)\Big|_{t=0} = 0.$$

Lemma 2. *If ∇ is the Levi-Civita connection on M, then $\nabla_S S = 0$.*

Proof: If $(a^q, \ldots, a^n)$ are real numbers such that $\sum (a^i)^2 = 1$, then the curve γ defined by

$$x^i(\gamma(t)) = a^i t$$

is an integral curve for S. On the other hand,

$$\gamma(t) = \exp_p \left(\sum_{i=1}^n a^i t \left. \frac{\partial}{\partial x^i} \right|_p \right),$$

and hence γ is a geodesic. We conclude that all integral curves for S are geodesics and hence $\nabla_S S = 0$.

Lemma 3. $\langle S, S \rangle \equiv 1$.

Proof: If γ is as in the preceding lemma,

$$\frac{d}{dt} \langle \gamma'(t), \gamma'(t) \rangle = 2 \langle \nabla_{\gamma'(t)} \gamma'(t), \gamma'(t) \rangle = 0,$$

so $\gamma'(t)$ must have constant length. But

$$\langle \gamma'(0), \gamma'(0) \rangle = \sum_{i=1}^n (a^i)^2 = 1,$$

so we conclude that $\langle S, S \rangle \equiv 1$.

Gauss Lemma I. $\langle S, E_{ij} \rangle \equiv 0$.

Proof: We calculate the derivative of $\langle S, E_{ij} \rangle$ in the radial direction:

$$\begin{aligned} S \langle S, E_{ij} \rangle = \langle \nabla_S S, E_{ij} \rangle + \langle S, \nabla_S E_{ij} \rangle &= \langle S, \nabla_S E_{ij} \rangle \\ &= \langle S, \nabla_{E_{ij}} S \rangle = \frac{1}{2} E_{ij} \langle S, S \rangle = 0. \end{aligned}$$

Thus $\langle S, E_{ij} \rangle$ is constant along the geodesic rays emanating from p. let $\|X\| = \sqrt{\langle X, X \rangle}$. Then as $(x^1, \ldots, x^n) \to (0, \ldots 0)$,

$$|\langle S, E_{ij} \rangle| \leq \|S\| \|E_{ij}\| = \|E_{ij}\| \to 0.$$

If follows that the constant $\langle S, E_{ij} \rangle$ must be zero.

Before proving the next lemma, we observe that

$$S(r) = 1, \qquad E_{ij}(r) = 0.$$

These fact can be verified by direct computation.

Gauss Lemma II. *$dr = \langle S, \cdot \rangle$; in other words, $dr(X) = \langle S, X \rangle$, whenever X is a smooth vector field on $U - \{p\}$.*

Proof: It clearly suffices to prove this when either $X = S$ or $X = E_{ij}$. In the first case,

$$dr(S) = S(r) = 1 = \langle S, S \rangle,$$

while in the second,

$$dr(E_{ij}) = E_{ij}(r) = 0 = \langle S, E_{ij} \rangle.$$

Lorentz case: Suppose now that $(M, \langle \cdot, \cdot \rangle)$ is a Lorentz manifold with Lorentz normal coordinates $(x^0, x^1, \dots x^n)$ defined on an open neighborhood $U = \exp_p(V)$ of $p \in M$. We let

$$U_- = \exp_p(\{v \in v : \langle v, v \rangle < 0\}), \quad U_+ = \exp_p(\{v \in v : \langle v, v \rangle > 0\}),$$

the images of the timelike and spacelike vectors in V respectively.

In this case, we can define two functions

$$t : U_- \to \mathbb{R} \quad \text{by} \quad t = \sqrt{(x^0)^2 - (x^1)^2 - \cdots - (x^n)^2},$$
$$r : U_+ \to \mathbb{R} \quad \text{by} \quad r = \sqrt{(x^1)^2 + \cdots + (x^n)^2 - (x^0)^2}.$$

We now have two radial vector fields T on U_- and S on U_+ defined by

$$T = \frac{x^0}{t}\frac{\partial}{\partial x^0} + \sum_{i=1}^{n} \frac{x^i}{t}\frac{\partial}{\partial x^i}, \quad S = \frac{x^0}{r}\frac{\partial}{\partial x^0} + \sum_{i=1}^{n} \frac{x^i}{r}\frac{\partial}{\partial x^i}.$$

For $1 \leq i, j \leq n$, let E_{ij} be the rotation vector field on U defined by (2.2) and for $1 \leq i \leq n$, define the infinitesimal Lorentz transformation E_{0i} by

$$E_{0i} = x^i \frac{\partial}{\partial x^0} + x^0 \frac{\partial}{\partial x^i}.$$

We can then carry out exactly the same steps for proving the Gauss Lemma:

Lemma 1. $[E_{ij}, T] = [E_{ij}, S] = 0$ *and* $[E_{0i}, T] = [E_{0i}, S] = 0$.

Lemma 2. *If* ∇ *is the Levi-Civita connection on* M*, then* $\nabla_T T = \nabla_S S = 0$.

Lemma 3. $\langle T, T \rangle \equiv -1$ *and* $\langle S, S \rangle \equiv 1$.

Gauss Lemma III. *For Lorentz manifolds,*

$$\langle T, E_{ij} \rangle \equiv 0 \equiv \langle T, E_{0i} \rangle \quad \text{and} \quad \langle S, E_{ij} \rangle \equiv 0 \equiv \langle S, E_{0i} \rangle.$$

The proofs are straightforward modifications of the Riemannian case.

2.3 Curvature in normal coordinates

The following theorem explains how the curvature of a Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$ measures deviation from the Euclidean metric.

Taylor Series Theorem. *The Taylor series for the Riemannian metric* (g_{ij}) *in terms of normal coordinates centered at a point* $p \in M$ *is given by*

$$g_{ij} = \delta_{ij} - \frac{1}{3} \sum_{k,l=1}^{n} R_{ikjl}(p) x^k x^l + (\text{higher order terms}).$$

The Taylor series for a Lorentz metric (g_{ij}) *in terms of normal coordinates centered at an event* $p \in M$ *is given by*

$$g_{ij} = \eta_{ij} - \frac{1}{3} \sum_{k,l=1}^{n} R_{ikjl}(p) x^k x^l + (\text{higher order terms}).$$

To prove this, we make use of "constant extensions" of vectors in T_pM, relative to the normal coordinates $(x^1, \dots, x^n)$. Suppose that $w \in T_pM$ and

$$w = \sum_{i=1}^{n} a^i \left. \frac{\partial}{\partial x^i} \right|_p .$$

Then the *constant extension* of w is the vector field

$$W = \sum_{i=1}^{n} a^i \frac{\partial}{\partial x^i}.$$

Since there is a genuine constant vector field in T_pM which is $\exp_p$-related to W, W depends only on w, not on the choice of normal coordinates.

We define a quadrilinear map

$$G : T_pM \times T_pM \times T_pM \times T_pM \longrightarrow \mathbb{R}$$

as follows:

$$G(x, y, z, w) = XY\langle Z, W\rangle(p),$$

where X, Y, Z and W and the constant extensions of x, y, z and w. Thus the components of G will be the second order derivatives of the metric tensor.

Lemma. *The quadralinear form G satisfies the following symmetries:*

1. $G(x, y, z, w) = G(y, x, z, w)$,
2. $G(x, y, z, w) = G(x, y, w, z)$,
3. $G(x, x, x, x) = 0$,
4. $G(x, x, x, y) = 0$,
5. $G(x, y, z, w) = G(z, w, x, y)$, *and*
6. $G(x, y, z, w) + G(x, z, w, y) + G(x, w, y, z) = 0$.

Proof: The second of these identities is immediate and the first follows from equality of mixed partials. The other identities require more work.

For the identity $G(w,w,w,w)=0$, we let $W=\sum a^i(\partial/\partial x^i)$; then the curve $\gamma:(-\epsilon,\epsilon)\to M$ defined by

$$x^i(\gamma(t))=a^i t$$

is an integral curve for W such that $\gamma(0)=p$. It is also a constant speed geodesic and hence

$$WW\langle W,W\rangle(p)=0.$$

We next check that $G(w,w,w,z)=0$, focusing first on the Riemannian case. It clearly suffices to prove this when z is unit length and perpendicular to a unit length w. We can choose our normal coordinates so that

$$W=\frac{\partial}{\partial x^1},\qquad Z=\frac{\partial}{\partial x^2}.$$

We consider the curve γ in M defined by

$$x^1\circ\gamma(t)=t,\qquad x^i\circ\gamma(t)=0,\quad \text{for } i>1.$$

Along γ we have $W=S$ and $Z=(1/x^1)E_{12}$, so it follows from Gauss Lemma I that $\langle W,Z\rangle\equiv 0$ along γ, and hence

$$WW\langle W,Z\rangle(p)=0. \tag{2.3}$$

The Lorentz case is similar: We choose γ to be spacelike or timelike, so that $\gamma'=S$ or $\gamma'=T$ along γ, and to achieve (2.3) we need to apply Gauss Lemma III with two types of infinitesimal Lorentz transformations, E_{ij} and E_{0i}.

It follows from the first four symmetries that whenever $u,v\in T_pM$ and $t\in\mathbb{R}$,

$$\begin{aligned}0&=G(u+tv,u+tv,u+tv,u-tv)\\&=t^3(\text{something})+t^2[G(v,v,u,u)-G(u,u,v,v)]+t(\text{something}),\end{aligned}$$

where we have used symmetries 1 and 2 to eliminate some terms in the sum. Since this identity must hold for all t, the coefficient of t^2 must be zero, so

$$G(u,u,v,v)=G(v,v,u,u),$$

which yields the fifth symmetry.

To obtain the final identity, we let

$$v_1,v_2,v_3,v_4\in T_pM\quad\text{and}\quad t_1,t_2,t_3,t_4\in\mathbb{R},$$

and note that

$$G\left(\sum t_iv_i,\sum t_jv_j,\sum t_kv_k,\sum t_lv_l\right)=0.$$

The coefficient of $t_1t_2t_3t_4$ must vanish, and hence

$$\sum_{\sigma\in S_4} G\left(v_{\sigma(1)}, v_{\sigma(2)}, v_{\sigma(3)}, v_{\sigma(4)}\right) = 0.$$

This, together with the earlier symmetries, yields the last symmetry.

Now we let

$$g_{ij,kl} = G\left(\left.\frac{\partial}{\partial x^k}\right|_p, \left.\frac{\partial}{\partial x^l}\right|_p, \left.\frac{\partial}{\partial x^i}\right|_p, \left.\frac{\partial}{\partial x^j}\right|_p\right).$$

Lemma. $R_{iljk}(p) = g_{ik,jl}(p) - g_{ij,lk}(p)$.

Proof: Since the Christoffel symbols Γ^k_{ij} vanish at p, it follows that

$$\frac{\partial}{\partial x^i}(\Gamma^l_{jk})(p) = \frac{1}{2}\frac{\partial}{\partial x^i}\left[\frac{\partial g_{lj}}{\partial x^k} + \frac{\partial g_{lk}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^l}\right](p),$$

$$\frac{\partial}{\partial x^j}(\Gamma^l_{ik})(p) = \frac{1}{2}\frac{\partial}{\partial x^j}\left[\frac{\partial g_{li}}{\partial x^k} + \frac{\partial g_{lk}}{\partial x^i} - \frac{\partial g_{ik}}{\partial x^l}\right](p),$$

and hence we conclude from Proposition 2 from §1.8 that

$$\begin{aligned} R_{ijlk}(p) = R_{lkij}(p) &= \frac{1}{2}\left[\frac{\partial^2 g_{jl}}{\partial x^i \partial x^k} - \frac{\partial^2 g_{jk}}{\partial x^i \partial x^l} + \frac{\partial^2 g_{ik}}{\partial x^j \partial x^l} - \frac{\partial^2 g_{il}}{\partial x^j \partial x^k}\right](p) \\ &= \frac{1}{2}\left[g_{jl,ik} + g_{ik,jl} - g_{jk,il} - g_{il,jk}\right](p) = g_{ik,jl}(p) - g_{il,jk}(p), \end{aligned}$$

where the comma denotes differentiation and we have used the fifth symmetry of G.

From the last lemma and the sixth symmetry, we now conclude that

$$R_{ikjl}(p) + R_{iljk}(p) = g_{il,jk}(p) - g_{ij,lk}(p) + g_{ik,lj}(p) - g_{ij,lk}(p) = -3g_{ij,kl}(p).$$

We therefore conclude that

$$\frac{\partial^2 g_{ij}}{\partial x^k \partial x^l}(p) = -\frac{1}{3}[R_{ikjl}(p) + R_{iljk}(p)].$$

Substitution into the Taylor expansion

$$g_{ij} = \delta_{ij} + \frac{1}{2}\sum_{k,l=1}^{n} \frac{\partial^2 g_{ij}}{\partial x^k \partial x^l}(p) x^k x^l + (\text{higher order terms})$$

or

$$g_{ij} = \eta_{ij} + \frac{1}{2}\sum_{k,l=1}^{n} \frac{\partial^2 g_{ij}}{\partial x^k \partial x^l}(p) x^k x^l + (\text{higher order terms})$$

now yields the Taylor Series Theorem.

Exercise V. Due Wednesday, May 11. Suppose that $(M, \langle \cdot,\cdot \rangle)$ is a two-dimensional Riemannian manifold.

a. Show that if (x^1, x^2) are Riemannian normal coordinates centered at $p \in M$, then

$$\langle \cdot,\cdot \rangle = (dx^1)^2 + (dx^2)^2 - \frac{1}{3}K(p)(x^1 dx^2 - x^2 dx^1)^2 + (\text{higher order terms}), \quad (2.4)$$

where $K(p)$ is the Gaussian curvature of M at p.

b. We can introduce *geodesic polar coordinates*,

$$x^1 = r\cos\theta, \qquad x^2 = r\sin\theta.$$

Show that the formula (2.4) can be rewritten as

$$\langle \cdot,\cdot \rangle = dr \otimes dr + \left[r^2 - \frac{r^4}{3}K(p) \right] d\theta \otimes d\theta + (\text{higher order terms}).$$

c. If $r > 0$, let C_r be the geodesic circle defined by the radial geodesic polar coordinate equal to r. In the plane, the length of this circle would be given by $L(C_r) = 2\pi r$. Show that in the curved surface, on the other hand, we have

$$K(p) = \frac{3}{\pi} \lim_{r \to 0} \frac{2\pi r - L(C_r)}{r^3}.$$

Thus in a surface with positive Gaussian curvature, the length of the geodesic circle grows more slowly than in the Euclidean plane, while when the Gaussian curvature is negative, the length of the geodesic circle grows more rapidly.

SOLUTION TO PART c: In the curved surface, we have

$$\begin{aligned} L(C_r) = \int_0^{2\pi} \sqrt{g_{22}} d\theta &= \int_0^{2\pi} r\sqrt{1 - (1/3)K(p)r^2 + (\text{h.o.t})}\ d\theta \\ &= \int_0^{2\pi} r\left(1 - (1/6)K(p)r^2 + (\text{h.o.t})\right) d\theta \\ &= 2\pi r - \frac{\pi}{3}K(\mathbf{p})r^3 + (\text{h.o.t}), \end{aligned}$$

where (h.o.t) stands for higher order terms.Hence

$$\lim_{r \to 0} \frac{2\pi r - L(C_r)}{r^3} = \frac{\pi}{3}K(\mathbf{p}),$$

or after rearrangement,

$$K(\mathbf{p}) = \frac{3}{\pi} \lim_{r \to 0} \frac{2\pi r - L(C_r)}{r^3}.$$

2.4 Tensor analysis

Perhaps it is time to describe the classical component notation for tensor fields which is used alot by physicists. Indeed, we will describe the tensor notation as used by Einstein in his earliest expositions of general relativity [5]. We begin by describing the tensor bundles, then tensor fields and their covariant derivatives and differentials.

Recall that T_p^*M is the vector space of linear maps $\alpha : T_pM \to \mathbb{R}$. We define the k-fold *tensor product* $\otimes^k T_p^*M$ to be the vector space of $\mathbb{R}$-multilinear maps

$$\phi : \overbrace{T_pM \times T_pM \times \cdots \times T_pM}^{k} \longrightarrow \mathbb{R}.$$

Thus $\otimes^1 T_p^*M$ is just the space of linear functionals on T_pM which is T_p^*M itself, while by convention $\otimes^0 T_p^*M = \mathbb{R}$.

We can define a product on it as follows. If $\phi \in \otimes^k T_p^*M$ and $\psi \in \otimes^l T_p^*M$, we define $\phi \otimes \psi \in \otimes^{k+l} T_p^*M$ by

$$(\phi \otimes \psi)(v_1, \ldots, v_{k+l}) = \phi(v_1, \ldots, v_k)\psi(v_{k+1}, \ldots, v_{k+l}).$$

This multiplication is called the *tensor product* and is bilinear,

$$(a\phi + \tilde{\phi}) \otimes \psi = a\phi \otimes \psi + \tilde{\phi} \otimes \psi, \qquad \phi \otimes (a\psi + \tilde{\psi}) = a\phi \otimes \psi + \phi \otimes \tilde{\psi},$$

as well as associative,

$$(\phi \otimes \psi) \otimes \omega = \phi \otimes (\psi \otimes \omega).$$

Hence we can write $\phi \otimes \psi \otimes \omega$ with no danger of confusion. The tensor product makes the direct sum

$$\otimes^* T_p^*M = \sum_{i=0}^{\infty} \otimes^k T_p^*M$$

into a graded algebra over $\mathbb{R}$, called the *tensor algebra* of T_p^*M.

Proposition. *If $(x^1, \ldots, x^n)$ are smooth coordinates defined on an open neighborhood of $p \in M$, then*

$$\{dx^{i_1}|_p \otimes \cdots \otimes dx^{i_k}|_p : 1 \leq i_1 \leq n, \ldots, 1 \leq i_k \leq n\}$$

is a basis for $\otimes^k T_p^*M$. Thus $\otimes^k T_p^*M$ has dimension n^k.

Sketch of proof: For linear independence, suppose that

$$\sum a_{i_1 \cdots i_k} dx^{i_1}|_p \otimes \cdots \otimes dx^{i_k}|_p = 0.$$

Then

$$0 = \sum a_{i_1 \cdots i_k} dx^{i_1}|_p \otimes \cdots \otimes dx^{i_k}|_p \left(\left.\frac{\partial}{\partial x^{j_1}}\right|_p, \ldots, \left.\frac{\partial}{\partial x^{j_k}}\right|_p \right)$$

$$= \sum a_{i_1 \cdots i_k} dx^{i_1}|_p \left(\left.\frac{\partial}{\partial x^{j_1}}\right|_p \right) dx^{i_k}|_p \left(\left.\frac{\partial}{\partial x^{j_k}}\right|_p \right) = a_{j_1 \cdots j_k}.$$

To show that the elements span, suppose that $\phi \in \otimes^k T_p^* M$, and show that

$$\phi = \sum a_{i_1 \cdots i_k} dx^{i_1}|_p \otimes \cdots \otimes dx^{i_k}|_p,$$
$$\text{where} \quad \sum a_{i_1 \cdots i_k} = \phi \left(\frac{\partial}{\partial x^{i_1}}\bigg|_p, \ldots, \frac{\partial}{\partial x^{i_k}}\bigg|_p \right).$$

We now let $\otimes^k T^* M = \bigcup \{\otimes^k T_p^* M\}$, a disjoint union. Just as in the case of the tangent and cotangent bundles, $\otimes^k T^* M$ has a smooth manifold structure, together with a projection $\pi : \otimes^k T^* M \to M$ such that $\pi(\otimes^k T_p^* M) = p$. We can describe the coordinates for the smooth structure on $\otimes^k T^* M$ as follows: If $(x^1, \ldots, x^n)$ are smooth coordinates on an open set $U \subseteq M$, the corresponding smooth coordinates on $\pi^{-1}(U)$ are are the pullbacks of $(x^1, \ldots, x^n)$ to $\pi^{-1}(U)$, together with the additional coordinates $p_{i_1 \cdots i_k} : \pi^{-1}(U) \to \mathbb{R}$ defined by

$$p_{i_1 \cdots i_k} \left(\sum_{j_1, \ldots, j_k} a_{j_1 \cdots j_k} dx^{j_1}|_p \otimes \cdots \otimes dx^{j_k}|_p \right) = a_{i_1 \cdots i_k}.$$

We can regard $\otimes^k T^* M$ as the total space of a vector bundle of rank n^k over M, as described in Chapter 5 of [12].

If U is an open subset of M, a *covariant tensor field* of rank k on U is a smooth map

$$T : U \to \otimes^k T_p^* M \quad \text{such that} \quad \pi \circ T = \mathrm{id}_U.$$

Informally, we can say that a covariant tensor field of rank k on U is a function T which assigns to each point $p \in U$ an element $T(p) \in \otimes^k T_p^* M$ in such a way that $T(p)$ varies smoothly with p.

Let $\Gamma(\otimes^k T^* M)$ denote the real vector space of covariant tensor fields of rank k on M, $\Gamma(\otimes^k T^* M|U)$ the space of covariant tensor fields of rank k on U. If $(U, (x^1, \ldots, x^n))$ is a smooth coordinate system on M, we can define

$$dx^{i_1} \otimes \cdots \otimes dx^{i_k} \in \Gamma(\otimes^k T^* M|U)$$
$$\text{by} \quad (dx^{i_1} \otimes \cdots \otimes dx^{i_k})(p) = dx^{i_1}|_p \otimes \cdots \otimes dx^{i_k}|_p.$$

Then any element $T \in \Gamma(\otimes^k T^* M|U)$ can be written uniquely as a sum

$$T = \sum_{i_1, \ldots, i_k} T_{i_1 \cdots i_k} dx^{i_1} \otimes \cdots \otimes dx^{i_k},$$

where each $T_{i_1 \cdots i_k} : U \to \mathbb{R}$ is a smooth function, called a component of T. If

$$S \in \Gamma(\otimes^k T^* M|U) \quad \text{and} \quad T \in \Gamma(\otimes^l T^* M|U),$$

then we can define the *tensor product*

$$S \otimes T \in \Gamma(\otimes^{k+l} T^* M|U) \quad \text{by} \quad (S \otimes T)(p) = S(p) \otimes T(p).$$

In terms of components with respect to local coordinates,

$$(S \otimes T)_{i_1 \cdots i_k j_1 \cdots j_l} = S_{i_1 \cdots i_k} T_{j_1 \dots j_l}.$$

Note that if $f \in \Gamma(\otimes^0 T^* M) = \mathcal{F}(M)$ and $T \in \Gamma(\otimes^k T^* M)$, then the tensor product reduces to the usual product $f \otimes T = fT$.

An important example of a covariant tensor field of rank four is the Riemann-Christoffel curvature tensor R of a pseudo-Riemannian manifold, which in terms of local coordinates $(U, (x^1, \dots, x^n))$ can be written

$$R = \sum_{i,j,k,l} R_{ijkl} dx^i \otimes dx^j \otimes dx^k \otimes dx^l.$$

More generally, we can define the *tensor product* $(\otimes^r T_p M) \otimes (\otimes^s T_p^* M)$ to be the vector space of $\mathbb{R}$-multilinear maps

$$\phi : \left(\overbrace{T_p^* M \times \cdots \times T_p^* M}^{r} \right) \otimes \left(\overbrace{T_p M \times \cdots \times T_p M}^{s} \right) \longrightarrow \mathbb{R}.$$

In this case, any element of $(\otimes^r T_p M) \otimes (\otimes^s T_p^* M)$ can be written uniquely as a sum

$$\sum a^{i_1 \cdots i_r}_{j_1, \cdots j_s} \left. \frac{\partial}{\partial x^{i_1}} \right|_p \otimes \cdots \otimes \left. \frac{\partial}{\partial x^{i_r}} \right|_p \otimes dx^{j_1}|_p \otimes \cdots \otimes dx^{j_s}|_p.$$

By generalizing the preceding construction in the obvious way we can construct a smooth manifold structure on

$$T^{r,s} M = (\otimes^r TM) \otimes (\otimes^s T^* M) = \bigcup \{ (\otimes^r T_p M) \otimes (\otimes^s T_p^* M) : p \in M \},$$

making this into the total space of a vector bundle of rank n^{r+s} over M. If U is an open subset of M, a *tensor field* which has *contravariant rank* r and *covariant rank* s over U is a smooth map

$$T : U \longrightarrow T^{r,s} M \quad \text{such that} \quad \pi \circ T = \mathrm{id}_U.$$

We let $\Gamma(\otimes^k T^{r,s} M)$ denote the real vector space of tensor fields of contravariant rank r and covariant rank s on M.

In terms of local coordinates $(U, (x^1, \dots, x^n))$, a tensor field $T \in \Gamma(T^{r,s} M)$ can be written as

$$T = \sum T^{i_1 \cdots i_r}_{j_1, \cdots j_s} \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_s},$$

where the functions $T^{i_1 \cdots i_r}_{j_1, \cdots j_s} : U \to \mathbb{R}$ are called the *components* of T. Under change of coordinates to $(\bar{U}, (\bar{x}^1, \dots, \bar{x}^n))$, one finds that

$$T = \sum \bar{T}^{k_1 \cdots k_r}_{l_1, \cdots l_s} \frac{\partial}{\partial \bar{x}^{k_1}} \otimes \cdots \otimes \frac{\partial}{\partial \bar{x}^{k_r}} \otimes d\bar{x}^{l_1} \otimes \cdots \otimes d\bar{x}^{l_s},$$

where

$$\bar{T}^{k_1\cdots k_r}_{l_1,\cdots l_s} = \sum T^{i_1\cdots i_r}_{j_1,\cdots j_s} \frac{\partial \bar{x}^{k_1}}{\partial x^{i_1}} \cdots \frac{\partial \bar{x}^{k_r}}{\partial x^{i_r}} \frac{\partial x^{j_1}}{\partial \bar{x}^{l_1}} \cdots \frac{\partial x^{j_s}}{\partial \bar{x}^{l_s}}. \tag{2.5}$$

In classical tensor analysis, one thinks of a tensor field as being defined by the collection of component functions $T^{i_1\cdots i_r}_{j_1,\cdots j_s}$, one collection for each local coordinate system, subject to the requirement that under change of coordinates the components transform according to (2.5). Although this notation uses lots of indices, it turns out to be relatively efficient when doing computations in terms of local coordinates.

Given a tensor field T of contravariant rank r and covariant rank s, with components $T^{i_1\cdots i_r}_{j_1,\cdots j_s}$, one can form a new tensor of contravariant rank $r-1$ and covariant rank $s-1$ which has components

$$S^{i_2\cdots i_r}_{j_2,\cdots j_s} = \sum_{i=1}^{n} T^{i_1 i_2\cdots i_r}_{i j_2,\cdots j_s}.$$

We say that S is obtained from T by contraction on the first contravariant and first covariant indices. Using a pseudo-Riemannian metric

$$\langle\cdot,\cdot\rangle = \sum_{i,j=1}^{n} g_{ij} dx^1 \otimes dx^j \quad \text{with inverse matrix} \quad (g^{ij}) = (g_{ij})^{-1}$$

we can raise and lower indices by the formulae

$$T^{i_1\cdots i_r j_1}{}_{j_2,\cdots j_s} = \sum g^{j_1 k} T^{i_1\cdots i_r}{}_{k j_2,\cdots j_s}, \quad T^{i_1\cdots i_{r-1}}{}_{i_r j_1,\cdots j_s} = \sum g_{i_r k} T^{i_1\cdots i_{r-1} k}{}_{j_1,\cdots j_s}.$$

For example, we can raise the first index of the Riemann-Christoffel curvature tensor obtaining $R^i{}_{jkl}$ and then contract on the first and third indices to obtain the component form of the Ricci tensor

$$R_{jl} = \sum_{i=1}^{n} R^i{}_{jil}.$$

Raising an index of the Ricci tensor and contracting once again then yields the scalar curvature s. Note that when we have pseudo-Riemannian metric present, lowering indices allows us to reduce all tensor fields to covariant tensor fields; this is sometimes advantageous because covariant tensor fields pull back under smooth maps from one manifold to another, while arbitrary tensor fields do not have that useful property.

Covariant derivatives: Suppose that X is a smooth vector field on M. Given a vector field Y on M, the Levi-Civita connection defines a new vector field $\nabla_X Y$. If $A \in \Gamma(T^*M)$ is a covariant tensor field of rank one, which is the same thing as a one-form on M we can take its *covariant derivative* $\nabla_X(A) \in \Gamma(T^*M)$ by forcing the "Leibniz rule":

$$X(A(Y)) = \nabla_X A(Y) + A(\nabla_X Y) \quad \text{or} \quad \nabla_X A(Y) = X(A(Y)) - A(\nabla_X Y).$$

We can then inductively define the *covariant derivative* of any covariant tensor field by enforcing the "Leibniz rule": thus if $S \in \Gamma(\otimes^k T^*M)$ and $T \in \Gamma(\otimes^l T^*M)$, then we require

$$\nabla_X(S \otimes T) = \nabla_X S \otimes T + S \otimes \nabla_X T.$$

One of the defining properties of the Levi-Civita connection then implies that for every vector field X on M,

$$\nabla_X \langle \cdot, \cdot \rangle = 0. \tag{2.6}$$

Using this fact, one can show that the covariant derivative ∇_X extends to arbitrary tensor fields of contravariant rank r and covariant rank s in such a way that it commutes with the raising and lowering of indices.

Covariant differentials: If Y is a vector field on M, the *covariant differential* of Y is the tensor field

$$\nabla . Y \in \Gamma(T^{1,1}M).$$

In terms of local coordinates $(U, (x^1, \ldots, x^n))$, we can write

$$Y = \sum_{i=1}^n Y^i \frac{\partial}{\partial x^i} \quad \text{and then} \quad \nabla . Y = \sum_{i,j=1}^n Y^i_{;j} \frac{\partial}{\partial x_i} dx^j,$$

where

$$Y^i_{;j} = dx^i(\nabla_{\partial/\partial x^j} Y) = \cdots = \frac{\partial Y^i}{\partial x^j} + \sum_{k=1}^n \Gamma^i_{jk} Y^k.$$

One can conveniently define the *divergence* of a vector field Y by taking the contraction of the covariant differential,

$$\operatorname{div}(Y) = \sum_{i=1}^n Y^i_{;i}.$$

Similarly, if ω is a covariant tensor field on M of rank one, that is a differential one-form, the *covariant differential* of ω is the tensor field

$$\nabla . \omega \in \Gamma(\otimes^2 T^*M).$$

n terms of local coordinates $(U, (x^1, \ldots, x^n))$, we can write

$$\omega = \sum_{i=1}^n \omega_i dx^i \quad \text{and then} \quad \nabla . \omega = \sum_{i,j=1}^n \omega_{i;j} dx^i \otimes dx^j,$$

where

$$\omega_{i;j} = \nabla_{\partial/\partial x^j} \omega \left(\frac{\partial}{\partial x^j} \right) = \cdots = \frac{\partial \omega^i}{\partial x^j} - \sum_{k=1}^n \Gamma^k_{ij} \omega_k.$$

It is an easy exercise to show that

$$(d\omega)_{ij} = \omega_{i;j} - \omega_{j;i},$$

the Christoffel symbols magically vanishing when one skew-symmetrizes. (Ordinary derivatives are sufficient for defining exterior derivatives!)

More generally, one can take covariant differentials of tensor products, Thus if Y is a vector field on M and ω is a differential one-form on M, then

$$\nabla.(Y \otimes \omega) = \nabla.Y \otimes \omega + Y \otimes \nabla.\omega,$$

and if Y and ω have components Y^i and ω_j as in the preceding paragraphs, the formulae with Christoffel symbols are

$$(Y \otimes \omega)^i_{j;k} = \frac{\partial}{\partial x^k}(Y^i \omega_j) + \sum_{l=1}^{n} \Gamma^i_{kl} Y^l \omega_j - \sum_{l=1}^{n} \Gamma^l_{kj} Y^i \omega_l.$$

More generally still, the covariant differential of a tensor field

$$T \in \Gamma(T^{r,s}M) \quad \text{is the tensor field} \quad \nabla.T \in \Gamma(T^{r,s+1}M).$$

If in terms of local coordinates $(U, (x^1, \ldots, x^n))$, the tensor field is expressed as

$$T = \sum T^{i_1 \cdots i_r}_{j_1, \cdots j_s} \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_s},$$

then

$$\nabla.T = \sum T^{i_1 \cdots i_r}_{j_1, \cdots j_s;k} \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_s} \otimes dx^k,$$

where

$$T^{i_1 \cdots i_r}_{j_1, \cdots j_s;k} = \frac{\partial T^{i_1 \cdots i_r}_{j_1, \cdots j_s}}{\partial x^k} + \sum_{k=1}^{n} \Gamma^{i_1}_{lk} T^{l i_2 \cdots i_r}_{j_1, \cdots j_s} + \cdots + \sum_{k=1}^{n} \Gamma^{i_r}_{lk} T^{i_1 \cdots i_{r-1} l}_{j_1, \cdots j_s}$$
$$- \sum_{k=1}^{n} \Gamma^l_{j_1 k} T^{i_1 \cdots i_r}_{l j_2, \cdots j_s} - \cdots - \sum_{k=1}^{n} \Gamma^l_{j_s k} T^{i_1 \cdots i_r}_{j_1, \cdots j_{s-1} l}. \quad (2.7)$$

As a special case of this notation, (2.6) takes the simple form

$$g_{ij;k} = 0. \quad (2.8)$$

The covariant differential commutes with contractions and satisfies the Leibniz rule for tensor products, for example,

$$(S_{ij} T_{kl})_{;r} = S_{ij;r} T_{kl} + S_{ij} T_{kl;r},$$

The fact (2.8) that the covariant differential of the metric is zero implies that covariant differential commutes with raising and lowering of indices.

Comma goes to semicolon: The tensor analysis operations we have just described are supplemented by a simple rule for transforming equations formulated within special relativity into the curved space-time of general relativity. This rule is called by Misner, Thorne and Wheeler [17], the "comma goes to semicolon" rule. One starts with a system of equations which might describe the way matter behaves within Newtonian physics, for example, one could take Maxwell's equations from electricity and magnetism. One them in special relativistic form with partial derivatives denoted by commas. Then to express the equations in appropriate form consistent with general relativity, one replaces the ordinary derivatives (commas) by covariant derivatives (semicolons). This idea is treated in much more detail in courses on general relativity, so some readers may want to skim the following more detailed example of this procedure.

Digression on fluid mechanics: As an example of the comma goes to semicolon procedure, one might consider a perfect fluid, a fluid in three-space which is described by a vector field $V = (V^1, V^2, V^3)$ representing the velocity of the fluid and a scalar function ρ which represents the density of the fluid. In Newtonian physics the motion of the fluid is described by the so-called Euler equations

$$\frac{\partial \rho}{\partial t} = \sum_{j=1}^{3} \frac{\partial}{\partial x^j}(\rho V^j) = 0, \qquad \frac{\partial V^i}{\partial t} + \sum_{j=1}^{3} V^j \frac{\partial V^i}{\partial x^j} = -\frac{\partial p}{\partial x^i}, \tag{2.9}$$

where p is an additional function called the pressure, which is usually related to the density ρ by an "equation of state." For simplicity, let us take the pressure to be identically zero, which gives a rather boring fluid in which the fluid particles just travel along straight lines, but this simple model of a fluid is often used in cosmological models.

One then seeks to find a relativistic version of the fluid equations, which is invariant under Lorentz transformations in Minkowski space-time $\mathbb{L}^4$, which we recall is just $\mathbb{R}^4$ with the Lorentz metric

$$\langle \cdot, \cdot \rangle = \sum_{i,j=0}^{3} \eta_{ij} dx^i \otimes dx^j, \quad \text{where} \quad (\eta_{ij}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In special relativity, we need to describe the fluid by density function ρ and a four-component vector field $V = (V^0, V^1, V^2, V^3)$ on Minkowski space-time tangent to the world-lines of the fluid particles and satisfying the condition that

$$\sum_{i,j=0}^{3} \eta_{ij} V^i V^j = -1. \tag{2.10}$$

One finds that equations which reduce to the Euler equations with zero pressure when fluid velocities are small compared to the speed of light (which make

(V^1, V^2, V^3) very small compared to $V^0 \doteq 1$) are

$$\sum_{j=0}^{3} \frac{\partial T^{ij}}{\partial x^j} = 0, \quad \text{where} \quad T^{ij} = \rho V^i V^j. \tag{2.11}$$

If we agree to let a comma denote partial differentiation, we can write this as

$$\sum_{j=0}^{3} T^{ij}{}_{,j} = 0.$$

The comma-goes-to-semicolon rule now says that the equations for a perfect fluid with zero pressure in a curved space-time (with the metric representing the gravitational field) are just

$$\sum_{j=0}^{3} T^{ij}{}_{;j} = 0,$$

where now the semicolon signifies taking a covariant differential rather than simply differentiating components. Of course, the semicolon equation reduces to the former in the special case of Lorentz space-time, in which all the Christoffel symbols vanish. If we substitute from (2.11), the equations become

$$\sum_{j=0}^{3} \rho V^i{}_{;j} V^j + \sum_{j=0}^{3} V^i (\rho V^j)_{;j} = 0 \quad \text{which yield} \quad \sum_{j=0}^{3} (\rho V^j)_{;j} = 0$$

upon multiplication with V_i and summing over i. This approximates the equation of continuity when velocities are small, which is the first of the Euler equations. Once we know it holds, it follows that

$$\sum_{j=0}^{3} V^i{}_{;j} V^j = 0,$$

which one can verify is just the condition that the flow lines of the fluid be geodesics.

More generally, when pressure is nonzero, the Euler equations (2.9) are the low velocity limit of the special relativistic equations

$$\sum_{j=0}^{3} T^{ij}{}_{,j} = 0, \quad \text{where} \quad T^{ij} = (\rho + p) V^i V^j + p\eta^{ij}$$

and (η^{ij}) just denotes the inverse to the matrix (η_{ij}) and the components V^i once again satisfy (2.10), as fully explained in §5.10 of [17]. According to the comma goes to semicolon rule the versions of the Euler equations compatible with general relativity is just

$$\sum_{j=0}^{3} T^{ij}{}_{;j} = 0, \quad \text{where} \quad T^{ij} = (\rho + p) V^i V^j + p g^{ij}$$

and (g^{ij}) is the matrix inverse to (g_{ij}).

Semicolon goes to comma: Normal coordinates allow us to stand this rule on its head. Suppose that we want to verify a tensor equation which involves covariant derivatives. Since we only need to verify this equation at a point, we can choose the coordinates we work in to be normal coordinates. But then the Christoffel symbols vanish, and we can replace the covariant differential by ordinary derivatives of components. The following exercise illustrates this principle.

Exercise VI. Due Wednesday, May 25. a. Use normal coordinates to prove the *Bianchi identity*

$$(\nabla_X R)(Y,Z)W + (\nabla_Y R)(Z,X)W + (\nabla_Z R)(X,Y)W = 0. \tag{2.12}$$

Since (2.12) is a tensor equation, it suffices to prove this at any given point $p \in M$. But we can choose normal coordinates at p and let X, Y, Z and W be coordinate fields for the normal coordinates. And then

$$\nabla_Z W(p) = \nabla_X W(p) = \nabla_Y W(p) = 0,$$

making the identity easy to prove. This illustrates the power of using normal coordinates in calculations.

b. Show that the identity (2.12) can be written in component form as

$$R_{ijrs;k} + R_{jkrs;i} + R_{kirs;j} = 0.$$

c. One can construct a new curvature tensor G_{ij}, called the *Einstein curvature*, from the Ricci and scalar curvatures by setting

$$G_{ij} = R_{ij} - \frac{1}{2} s\, g_{ij}.$$

Show that

$$G^{ij}{}_{;j} = 0. \tag{2.13}$$

Remark. The fact that the Einstein tensor has zero divergence was the clue that led Einstein to his field equations, $G^{ij} = 8\pi T^{ij}$, where T^{ij} is the stress-energy tensor, which is given by (2.11) for a perfect dust.

SOLUTION: a. Note first that since $[X,Y] = [X,Z] = [Z,X] = 0$,

$$\begin{aligned}\nabla_X (R(Y,Z)W) + \nabla_Y (R(Z,X)W) + \nabla_Z (R(X,Y)W) &= \nabla_X(\nabla_Y\nabla_Z - \nabla_Z\nabla_Y)W \\ + \nabla_Y(\nabla_Z\nabla_X - \nabla_X\nabla_Z)W &+ \nabla_Z(\nabla_X\nabla_Y - \nabla_Y\nabla_X)W \\ &= R(X,Y)(\nabla_Z X) + R(Y,Z)(\nabla_X W) + R(Z,X)(\nabla_Y W).\end{aligned}$$

Then using the fact that

$$\nabla_Z W(p) = \nabla_X W(p) = \nabla_Y W(p) = 0 \tag{2.14}$$

we see that

$$\nabla_X (R(Y,Z)W)(p) + \nabla_Y (R(Z,X)W)(p) + \nabla_Z (R(X,Y)W)(p) = 0.$$

Now using the Leibniz rule, we see that at the point p

$$\begin{aligned} 0 = \nabla_X (R(Y,Z)W) + \nabla_Y (R(Z,X)W) + \nabla_Z (R(X,Y)W) \\ = (\nabla_X R)(Y,Z)W + (\nabla_Y R)(Z,X)W + (\nabla_Z R)(X,Y)W \\ + R(Y,Z)(\nabla_X W) + R(Z,X)(\nabla_Y W) + R(X,Y)(\nabla_Z W). \end{aligned}$$

thus it follows once again from (2.15) that

$$(\nabla_X R)(Y,Z)W(p) + (\nabla_Y R)(Z,X)W(p) + (\nabla_Z R)(X,Y)W(p) = 0.$$

Since p was an arbitrary point, we have verified the tensor equation

$$(\nabla_X R)(Y,Z)W + (\nabla_Y R)(Z,X)W + (\nabla_Z R)(X,Y)W = 0.$$

b. From here, it is more convenient to use index notation. In index notation, the last equation reads

$$R^r_{sij;k} + R^r_{sjk;i} + R^r_{ski;j} = 0.$$

Lowering the index then yields

$$R_{rsij;k} + R_{rsjk;i} + R_{rski;j} = 0 \quad \text{or} \quad R_{ijrs;k} + R_{jkrs;i} + R_{kirs;j} = 0.$$

c. Finally, we write

$$\sum g^{ir} g^{js} R_{ijrs;k} + \sum g^{ir} g^{js} R_{jkrs;i} + \sum g^{ir} g^{js} R_{kirs;j} = 0,$$

which simplifies to

$$s_{;k} - \sum g^{ir} R_{kr;i} - \sum g^{js} R_{ks;j} = 0 \quad \text{or} \quad \left(R^i{}_{k;i} - \frac{1}{2} s_{;k} \right) = 0,$$

equivalent to $G^i{}_{j;i} = 0$, which is the same as (2.13).

2.5 Riemannian manifolds as metric spaces

We can use the Riemannian normal coordinates constructed in the previous sections to establish the following important result. Before stating it, we note that a smooth curve $\lambda : [0,1] \to M$ is called *regular* if $\lambda'(t)$ is never zero; it is easy to prove that regular curve can always be reparametrized to have nonzero constant speed.

Local Minimization Theorem. *Suppose that $(M^n, \langle \cdot, \cdot \rangle)$ is a Riemannian manifold and that V is an open ball of radius $\epsilon > 0$ centered at $0 \in T_pM$ which*

$\exp_p$ maps diffeomorphically onto an open neighborhood U of p in M. Suppose that $v \in V$ and that $\gamma : [0,1] \to M$ is the geodesic defined by $\gamma(t) = \exp_p(tv)$. Let $q = \exp_p(v)$. If $\lambda : [0,1] \to M$ is any smooth curve with $\lambda(0) = p$ and $\lambda(1) = q$, then

1. *$L(\lambda) \geq L(\gamma)$, and if equality holds and λ is regular, then λ is a reparametrization of γ.*
2. *$J(\lambda) \geq J(\gamma)$, with equality holding only if $\lambda = \gamma$.*

To prove the first of these assertions, we use normal coordinates $(x^1, \ldots, x^n)$ defined on U. Note that $L(\gamma) = r(q)$. Suppose that $\lambda : [0,1] \to M$ is any smooth curve with $\lambda(0) = p$ and $\lambda(1) = q$.

Case I. Suppose that λ does not leave U. Then it follows from Gauss Lemma II that

$$\begin{aligned} L(\lambda) = \int_0^1 \sqrt{\langle \lambda'(t), \lambda'(t) \rangle} dt = \int_0^1 \|\lambda'(t)\| dt \geq \int_0^1 \langle \lambda'(t), R(\lambda(t) \rangle dt \\ \geq \int_0^1 dr(\lambda'(t)) dt = (r \circ \lambda)(1) - (r \circ \lambda)(0) = L(\gamma). \end{aligned}$$

Moreover, equality holds only if $\lambda'(t)$ is a nonnegative multiple of $R(\lambda(t))$ which holds only if λ is a reparametrization of γ.

Case II. Suppose that λ leaves U at some first time $t_0 \in (0,1)$. Then

$$\begin{aligned} L(\lambda) = \int_0^1 \sqrt{\langle \lambda'(t), \lambda'(t) \rangle} dt > \int_0^{t_0} \|\lambda'(t)\| dt \geq \int_0^{t_0} \langle \lambda'(t), R(\lambda(t) \rangle dt \\ \geq \int_0^{t_0} dr(\lambda'(t)) dt = (r \circ \lambda)(t_0) - (r \circ \lambda)(0) = \epsilon > L(\gamma). \end{aligned}$$

The second assertion is proven in a similar fashion.

If $(M, \langle \cdot, \cdot \rangle)$ is a Riemannian manifold, we can define a distance function

$$d : M \times M \longrightarrow \mathbb{R}$$

by setting

$$\begin{aligned} d(p,q) = \inf\{\ L(\gamma) \text{ such that } \gamma : [0,1] \to M \text{ is a smooth path} \\ \text{with } \gamma(0) = p \text{ and } \gamma(1) = q\ \}. \end{aligned}$$

Then the previous theorem shos that $d(p,q) = 0$ implies that $p = q$. Hence

1. $d(p,q) \geq 0$, with equality holding if and only if $p = q$,
2. $d(p,q) = d(q,p)$, and

3. $d(p,r) \leq d(p,q) + d(q,r)$.

Thus (M, d) is a metric space. It is relatively straightforward to show that the metric topology on M agrees with the usual topology of M. In particular, the distance function

$$d : M \times M \longrightarrow \mathbb{R}$$

is continuous.

Definition. If p and q are points in a Riemannian manifold M, a *minimal geodesic* from p to q is a geodesic $\gamma : [a,b] \to M$ such that

$$\gamma(a) = p, \quad \gamma(b) = q \quad \text{and} \quad L(\gamma) = d(p,q).$$

An open set $U \subset M$ is said to be *geodesically convex* if whenever p and q are elements of U, there is a unique minimal geodesic from p to q and moreover, that minimal geodesic lies entirely within U.

Geodesic Convexity Theorem. *Suppose that $(M^n, \langle \cdot, \cdot \rangle)$ is a Riemannian manifold. Then M has an open cover by geodesically convex open sets.*

A proof could be constructed based upon the preceding arguments, but we omit the details. (One proof is outlined in Problem 6.4 from [13].)

2.6 Completeness

We return now to a variational problem that we considered earlier. Given two points p and q in a Riemannian manifold M, does there exist a minimal geodesic from p to q? For this variational problem to have a solution we need an hypothesis on the Riemannian metric.

Definition. A pseudo-Riemannian manifold $(M, \langle \cdot, \cdot)$ is said to be *geodesically complete* if geodesics in M can be extended indefinitely without running off the manifold. Equivalently, $(M, \langle \cdot, \cdot)$ is geodesically complete if $\exp_p$ is globally defined for all $p \in M$.

Examples: The spaces of constant curvature $\mathbb{E}^n$, $\mathbb{S}^n(a)$ and $\mathbb{H}^n(a)$ are all geodesically complete, as are the compact Lie groups with biinvariant metrics and the Grassmann manifolds. On the other hand, nonempty proper open subsets of any of these spaces are not geodesically complete.

Minimal Geodesic Theorem I. *Suppose that $(M^n, \langle \cdot, \cdot \rangle)$ is a connected and geodesically complete Riemannian manifold. Then any two points p and q of M can be connected by a minimal geodesic.*

The idea behind the proof is extremely simple. Given $p \in M$, the geodesic completeness assumption implies that $\exp_p$ is globally defined. Let $a = d(p,q)$, then we should have $q = \exp_p(av)$, where v is a unit length vector in T_pM which "points in the direction" of q.

More precisely, let $\tilde{B}_\epsilon$ be a closed ball of radius ϵ centered at 0 in T_pM, and suppose that $\tilde{B}_\epsilon$ is contained in a an open set which is mapped diffeomorphically by $\exp_p$ onto an open neighborhood of p in M. Let $\tilde{S}_\epsilon$ be the boundary of $\tilde{B}_\epsilon$ and let S be the image of $\tilde{S}$ under $\exp_p$. Since S is a compact subset of M there is a point $m \in S$ of minimal distance from q. We can write $m = \exp_p(\epsilon v)$ for some unit length $v \in T_pM$. Finally, we define

$$\gamma : [0, a] \to M \quad \text{by} \quad \gamma(t) = \exp_p(tv).$$

Then γ is a candidate for the minimal geodesic from p to q.

To finish the proof, we need to show that $\gamma(a) = q$. It will suffice to show that

$$d(\gamma(t), q) = a - t, \tag{2.15}$$

for all $t \in [0, a]$. Note that $d(\gamma(t), q) \geq a - t$, because if $d(\gamma(t), q) < a - t$, then

$$d(p, q) \leq d(p, \gamma(t)) + d(\gamma(t), q) < t + (a - t) = a.$$

Moreover, if (2.15) holds for $t_0 \in [0, a]$, it also holds for all $t \in [0, t_0]$, because if $t \in [0, t_0]$, then

$$d(\gamma(t), q) \leq d(\gamma(t), \gamma(t_0)) + d(\gamma(t_0), q) \leq (t_0 - t) + (a - t_0) = a - t.$$

We let

$$t_0 = \sup\{t \in [0, a] : d(\gamma(t), q) = a - t\},$$

and note that $d(\gamma(t_0), q) = a - t_0$ by continuity. We will show that:

1. $t_0 \geq \epsilon$, and
2. $0 < t_0 < a$ leads to a contradiction.

To establish the first of these assertions, we note that by the Local Minimization Theorem from §2.5,

$$d(p, q) = \inf\{d(p, r) + d(r, q) : r \in S\} = \epsilon + \inf\{d(r, q) : r \in S\} = \epsilon + d(m, q),$$

and hence $a = \epsilon + d(m, q) = \epsilon + d(\gamma(\epsilon), q)$.

To prove the second assertion, we construct a sphere S about $\gamma(t_0)$ as we did for p, and let m be the point on S of minimal distance from q. Then

$$d(\gamma(t_0), q) = \inf\{d(\gamma(t_0), r) + d(r, q) : r \in S\} = \epsilon + d(m, q),$$

and hence

$$a - t_0 = \epsilon + d(m, q), \quad \text{so} \quad a - (t_0 + \epsilon) = d(m, q).$$

Note that $d(p, m) \geq t_0 + \epsilon$ because otherwise

$$d(p, q) \leq d(p, m) + d(m, q) < t_0 + \epsilon + a - (t_0 + \epsilon) = a,$$

so the broken geodesic from p to $\gamma(t_0)$ to m has length $t_0 + \epsilon = d(p, m)$. If the broken geodesic had a corner it could be shortened by rounding off the corner, a fact which follows from the first variation formula (1.6) for piecewise smooth paths. Hence m must lie on the image of γ, so $\gamma(t_0 + \epsilon) = m$, contradicting the maximality of t_0.

It follows that $t_0 = a$, $d(\gamma(a), q) = 0$ and $\gamma(a) = q$, finishing the proof of the theorem.

Basic idea used in the preceding proof: If you have a regular piecewise smooth curve (that is, $\gamma'(t)$ is never zero and at corners, the left and right-hand limits of $\gamma'(t)$ exist and are nonzero), then the curve can be shortened by "rounding corners." Needless to say, this is a useful technique.

For a Riemannian manifold, we also have a notion of completeness in terms of metric spaces. Fortunately, the two notions of completeness coincide:

Hopf-Rinow Theorem. *Suppose that $(M^n, \langle \cdot, \cdot \rangle)$ is a connected Riemannian manifold. Then (M, d) is complete as a metric space if and only if $(M, \langle \cdot, \cdot \rangle)$ is geodesically complete.*

To prove this theorem, suppose first that $(M^n, \langle \cdot, \cdot \rangle)$ is geodesically complete. Let p be a fixed point in M and (q_i) a Cauchy sequence in (M, d). We need to show that (q_i) converges to a point $q \in M$. We can assume without loss of generality that $d(q_i, q_j) < \epsilon$ for some $\epsilon > 0$, and let $K = d(p, q_1)$, so that $d(p, q_i) \leq K + \epsilon$ for all i. It follows from the Minimal Geodesic Theorem that $q_i = \exp_p(v_i)$ for some $v_i \in T_pM$, and $\|v_i\| \leq K + \epsilon$. Completeness of $\mathbb{R}^n$ with its usual Euclidean metric implies that (v_i) has a convergent subsequence, which converges to some point $v \in T_pM$. Then $q = \exp_p(v)$ is a limit of the Cauchy sequence (q_i), and (M, d) is indeed a complete metric space.

To prove the converse, we suppose that (M, d) is a complete metric space, but $(M^n, \langle \cdot, \cdot \rangle)$ is not geodesically complete. Then there is some unit speed geodesic $\gamma : [0, b) \to M$ which extends to no interval $[0, b+\delta)$ for $\delta > 0$. Let (t_i) be a sequence from $[0, b)$ such that $t_i \to b$. If $p_i = \gamma(t_i)$, then $d(p_i, p_j) \leq |t_i - t_j|$, so the sequence (p_i) is a Cauchy sequence within the metric space (M, d). Let p_0 be the limit of (p_i). Then by Corollary 4 from §2.1, we see that there is some fixed $\epsilon > 0$ such that $\exp_{p_i}(v)$ is defined for all $|v| < \epsilon$ when i is sufficiently large. This implies γ can be extended a distance ϵ beyond p_i when i is sufficiently large, yielding a contradiction.

2.7 Smooth closed geodesics

If we are willing to strengthen completeness to compactness, we can give another proof of the Minimal Geodesic Theorem, which is quite intuitive and illustrates techniques that are commonly used for calculus of variations problems. Moreover, this approach is easily modified to give a proof that a compact Riemannian manifold which is not simply connected must possess a nonconstant smooth closed geodesic. Of course, this can be thought of as a special case of periodic motion within classical mechanics.

Simplifying notation a little, we let

$$\Omega(M;p,q) = \{ \text{ smooth maps } \gamma : [0,1] \to M \text{ such that } \gamma(0) = p \text{ and } \gamma(1) = q \ \}$$

$$\text{and let} \quad \Omega(M;p,q)^a = \{\gamma \in \Omega(M;p,q) : J(\gamma) < a\}.$$

Assuming that M is compact, we can conclude from Proposition 3 of §2.1 that there is a $\delta > 0$ such that any p and q in M with $d(p,q) < \delta$ are connected by a unique minimal geodesic

$$\gamma_{p,q} : [0,1] \to M \quad \text{with} \quad L(\gamma_{p,q}) = d(p,q).$$

Moreover, if $\delta > 0$ is sufficiently small, the ball of radius δ about any point is geodesically convex and $\gamma_{p,q}$ depends smoothly on p and q. If $\gamma : [b, b+\epsilon] \to M$ is a smooth path and

$$\epsilon < \frac{\delta^2}{2a}, \quad \text{then} \quad J(\gamma) \leq a \quad \Rightarrow L(\gamma) \leq \sqrt{2a\epsilon} < \delta.$$

as we see from (1.2).

Choose $N \in \mathbb{N}$ such that $1/N < \epsilon$, and if $\gamma \in \Omega(M;p,q)^a$, let $p_i = \gamma(i/N)$, for $0 \leq i \leq N$. Then γ is approximated by the map $\tilde{\gamma} : [0,1] \to M$ such that

$$\tilde{\gamma}(t) = \gamma_{p_{i-1}p_i}\left(\frac{(i-1)+t}{N}\right), \quad \text{for} \quad t \in \left[\frac{i-1}{N}, \frac{i}{N}\right].$$

Thus $\tilde{\gamma}$ lies in the space of "broken geodesics,"

$$BG_N(M;p,q) = \{ \text{ maps } \gamma : [0,1] \to M \text{ such that } \gamma| \left[\frac{i-1}{N}, \frac{i}{N}\right] \text{ is a constant speed geodesic } \},$$

and $\Omega(M;p,q)^a$ is approximated by

$$BG_N(M;p,q)^a = \{\gamma \in BG_N(M;p,q) : J(\gamma) < a\}.$$

Suppose that γ is an element of $BG_N(M;p,q)^a$. Then if

$$p_i = \gamma\left(\frac{i}{N}\right), \quad \text{then} \quad d(p_{i-1}, p_i) \leq \sqrt{\frac{2a}{N}} < \sqrt{2a\epsilon} < \delta,$$

so γ is completely determined by

$$(p_0, p_1, \ldots, p_i, \ldots, p_N), \quad \text{where} \quad p_0 = p, \quad p_N = q.$$

Thus we have an injection

$$j : BG_N(M;p,q)^a \to \overbrace{M \times M \times \cdots \times M}^{N-1},$$

$$j(\gamma) = \left(\gamma\left(\frac{1}{N}\right), \ldots, \gamma\left(\frac{N-1}{N}\right)\right).$$

We also have a map $r : \Omega(M;p,q)^a \to BG_N(M;p,q)^a$ defined as follows: If $\gamma \in \Omega(M;p,q)^a$, let $r(\gamma)$ be the broken geodesic from

$$p = p_0 \text{ to } p_1 = \gamma\left(\frac{1}{N}\right) \text{ to } \cdots \text{ to } p_{N-1} = \gamma\left(\frac{N-1}{N}\right) \text{ to } q.$$

We can regard $r(\gamma)$ as the closest approximation to γ in the space of broken geodesics.

Minimal Geodesic Theorem II. *Suppose that $(M^n, \langle\cdot,\cdot\rangle)$ is a compact connected Riemannian manifold. Then any two points p and q of M can be connected by a minimal geodesic.*

To prove this, let

$$\mu = \inf\{J(\gamma) : \gamma \in \Omega(M;p,q)\}.$$

Choose $a > \mu$, so that $\Omega(M;p,q)^a$ is nonempty, and let (γ_j) be a sequence in $\Omega(M;p,q)^a$ such that $J(\gamma_j) \to \mu$. Let $\tilde{\gamma}_j = r(\gamma_j)$, the corresponding broken geodesic from

$$p = p_{0j} \text{ to } p_{1j} = \gamma\left(\frac{1}{N}\right) \text{ to } \cdots \text{ to } p_{(N-1)j} = \gamma\left(\frac{N-1}{N}\right) \text{ to } q,$$

and note that $J(\tilde{\gamma}_j) \leq J(\gamma_j)$.

Since M is compact, we can choose a subsequence (j_k) such that (p_{ij_k}) converges to some point $p_i \in M$ for each i. Hence a subsequence of $(\tilde{\gamma}_j)$ converges to an element $\tilde{\gamma} \in BG_N(M;p,q)^a$. Moreover,

$$J(\tilde{\gamma}_j) \leq \lim_{j\to\infty} J(\tilde{\gamma}_j) \leq \lim_{j\to\infty} J(\gamma_j) = \mu.$$

The curve $\tilde{\gamma}$ must be of constant speed, because otherwise we could decrease J be reparametrizing $\tilde{\gamma}$. Hence $\tilde{\gamma}$ must also minimize length L on $BG_N(M;p,q)^a$.

Finally, $\tilde{\gamma}$ cannot have any corners, because if it did, we could decrease length by rounding corners. (This follows from the first variation formula for piecewise smooth curves given in §1.3.2.) We conclude that $\tilde{\gamma} : [0,1] \to M$ is a smooth geodesic with $L(\tilde{\gamma}) = d(p,q)$, that is, $\tilde{\gamma}$ is a minimal geodesic from p to q, finishing the proof of the theorem.

Remark. Note that $BG_N(M;p,q)^a$ can be regarded as a finite-dimensional manifold which approximates the infinite-dimensional space $\Omega(M;p,q)^a$. This is a powerful idea which Marston Morse used in his critical point theory for geodesics. (See [15] for a thorough working out of this approach.)

Although the preceding theorem is weaker than the one presented in the previous section, the technique of proof can be extended to other contexts. We say that two smooth curves

$$\gamma_1 : S^1 \to M \quad \text{and} \quad \gamma_2 : S^1 \to M$$

are *freely homotopic* if there is a continuous path

$$\Gamma : [0,1] \times S^1 \to M \quad \text{such that} \quad \Gamma(0,t) = \gamma_1(t) \quad \text{and} \quad \Gamma(1,t) = \gamma_2(t).$$

We say that a connected manifold M is *simply connected* if any smooth path $\gamma : S^1 \to M$ is freely homotopic to a constant path. Thus M is simply connected if and only if its fundamental group, as defined in [8], is zero.

As before, we can approximate the space $\mathrm{Map}(S^1, M)$ of smooth maps $\gamma : S^1 \to M$ by a finite-dimensional space, where S^1 is regarded as the interval $[0, 1]$ with the points 0 and 1 identified. This time the finite-dimensional space is the space of "broken geodesics,"

$$BG_N(S^1, M) = \{ \text{ maps } \gamma : [0,1] \to M \text{ such that}$$
$$\gamma| \left[\frac{i-1}{N}, \frac{i}{N} \right] \text{is a constant speed geodesic and } \gamma(0) = \gamma(1) \ \}.$$

Just as before, when a is sufficiently small, then

$$\mathrm{Map}(S^1, M)^a = \{\gamma \in \mathrm{Map}(S^1, M) : J(\gamma) < a\}$$

is approximated by

$$BG_N(S^1, M)^a = \{\gamma \in BG_N(S^1, M) : J(\gamma) < a\}.$$

Moreover, if $p_i = \gamma(i/N)$, then γ is completely determined by

$$(p_1, p_2, \ldots, p_i, \ldots, p_N).$$

Thus we have an injection

$$j : BG_N(S^1, M)^a \to \overbrace{M \times M \times \cdots \times M}^{N},$$
$$j(\gamma) = \left(\gamma \left(\frac{1}{N} \right), \ldots, \gamma \left(\frac{N-1}{N} \right), \gamma(1) \right).$$

We also have a map $r : \mathrm{Map}(S^1, M)^a \to BG_N(S^1, M)^a$ defined as follows: If $\gamma \in \mathrm{Map}(S^1, M)^a$, let $r(\gamma)$ be the broken geodesic from

$$\gamma(0) \text{ to } p_1 = \gamma \left(\frac{1}{N} \right) \text{ to } \cdots \text{ to } \ p_{N-1} = \gamma \left(\frac{N-1}{N} \right) \text{ to } p_N = \gamma(1).$$

Closed Geodesic Theorem. *Suppose that $(M^n, \langle \cdot, \cdot \rangle)$ is a compact connected Riemannian manifold which is not simply connected. Then there is a nonconstant smooth closed geodesic in M which minimizes length among all nonconstant smooth closed curves in M^n.*

The proof is virtually identical to that for the Minimal Geodesic Theorem II except for a minor change in notation. We note that since M is not simply connected, the space

$$\mathcal{F} = \{\gamma \in \mathrm{Map}(S^1, M) : \ \gamma \text{ is not freely homotopic to a constant } \}$$

is nonempty, and we let

$$\mu = \inf\{J(\gamma) : \gamma \in \mathcal{F}\}.$$

Choose $a > \mu$, so that $\mathcal{F}^a = \{\gamma \in \mathcal{F} : J(\gamma) < a\}$ is nonempty, and let (γ_j) be a sequence in $\mathcal{F}^a$ such that $J(\gamma_j) \to \mu$. Let $\tilde{\gamma}_j = r(\gamma_j)$, the corresponding broken geodesic and from

$$p_{Nj} = \gamma(0) \text{ to } p_{1j} = \gamma\left(\frac{1}{N}\right) \text{ to } \cdots \text{ to } p_{(N-1)j} = \gamma\left(\frac{N-1}{N}\right) \text{ to } p_{Nj} = \gamma(1),$$

and note that $J(\tilde{\gamma}_j) \leq J(\gamma_j)$.

Since M is compact, we can choose a subsequence (j_k) such that (p_{ij_k}) converges to some point $p_i \in M$ for each i. Hence a subsequence of $(\tilde{\gamma}_j)$ converges to an element $\tilde{\gamma} \in BG_N(S^1, M)^a$. Moreover,

$$J(\tilde{\gamma}_j) \leq \lim_{j\to\infty} J(\tilde{\gamma}_j) \leq \lim_{j\to\infty} J(\gamma_j) = \mu.$$

The curve $\tilde{\gamma}$ must be of constant speed, because otherwise we could decrease J be reparametrizing $\tilde{\gamma}$. Hence $\tilde{\gamma}$ must also minimize length L on $BG_N(S^1, M)^a$.

Finally, $\tilde{\gamma}$ cannot have any corners, because if it did, we could decrease length by rounding corners. (This follows again from the first variation formula for piecewise smooth curves given in §1.3.2.) We conclude that $\tilde{\gamma} : S^1 \to M$ is a smooth geodesic which is not constant since it cannot even be freely homotopic to a constant.

Remarks. It was proven by Fet and Liusternik that any compact connected Riemannian manifold has at least one smooth closed geodesic. As in the proof of the preceding theorem, one needs a constraint to pull against and such constraints are provided by the standard topological invariants of the free loop space $\mathrm{Map}(S^1, M)$, namely

$$\pi_k(\mathrm{Map}(S^1, M)) \quad \text{and} \quad H^k(\mathrm{Map}(S^1, M); \mathbb{Z}).$$

A nonconstant geodesic $\gamma : S^1 \to M$ is said to be *prime* if it is not of the form $\gamma \circ \pi$, where $\pi : S^1 \to S^1$ is a covering of degree $k \geq 2$. Wilhelm Klingenberg raised the question: Is it true that any compact simply connected Riemannian manifold has infinitely many prime smooth closed geodesics? Gromoll and Meyer [7] made a significant advance on this question by showing that if M had only finitely many prime geodesics, then the ranks of $H^k(\mathrm{Map}(S^1, M); \mathbb{R})$ must be bounded. Unfortunately, it is quite difficult to calculate the integer cohomology ring of $\mathrm{Map}(S^1, M)$ using the usual techniques of algebraic topology. However, Quillen and Sullivan were able to simplify the calculations of rational or real cohomology sufficiently (via Sullivan's theory of minimal models) to enable Vigué and Sullivan to show that if M has only finitely many prime geodesics, then the real cohomology algebra $H^*(\mathrm{Map}(S^1, M); \mathbb{R})$ must be generated as an algebra by a single element.

Nevertheless, the basic question raised by Klingenberg still appears to be open for compact simply connected Riemannian manifolds such as S^n, when $n \geq 3$.

The partial solution of this question by Gromoll, Meyer and others represents an impressive application of algebraic topology to a problem very much at the center of geometry and mechanics. One need only recall that proving existence of periodic solutions to problems in celestial mechanics was one of the prime motivations for Henri Poincaré's research which led to the development of qualitative methods for solving differential equations.

Chapter 3

Curvature and topology

3.1 Overview

Curvature is the most important local invariant of a Riemannian or pseudo-Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$. It measures the deviation from flatness in the coordinates (normal coordinates) which are as flat as possible near a given point p. It is natural to ask what is the relationship between the curvature of a pseudo-Riemannian manifold and its topology.

There are many ways of looking at the curvature. Recall from §1.9 that we can organize the curvature into a curvature operator

$$\mathcal{R} : \Lambda^2 T_pM \to \Lambda^2 T_pM \quad \text{defined by} \quad \langle \mathcal{R}(x \wedge y), z \wedge w\rangle = \langle R(x,y)w, z\rangle,$$

and one of the curvature symmetries states that this curvature operator is symmetric,

$$\langle \mathcal{R}(x \wedge y), z \wedge w\rangle = \langle x \wedge y, \mathcal{R}(z \wedge w)\rangle .$$

Hence by theorems from linear algebra, $\Lambda^2 T_pM$ has a basis consisting of eigenvectors for $\mathcal{R}$ and all of the eigenvalues of $\mathcal{R}$ are real.

Definition. We say that $(M, \langle\cdot,\cdot\rangle)$ has *positive curvature operators* if for every $p \in M$, the eigenvalues of $\mathcal{R}$ are > 0 and that it has *nonpositive curvature operators* if for every $p \in M$, the eigenvalues of $\mathcal{R}$ are ≤ 0.

One of the first theorems relating curvature to topology was a theorem first proven by von Mangoldt and Hadamard in the classical theory of surfaces and then extended to Riemannian manifolds of arbitrary dimension by Élie Cartan in 1928. Recall from §2.7 that a connected manifold M is *simply connected* if any smooth path $\gamma : S^1 \to M$ is freely homotopic to a constant path. The Hadamard-Cartan Theorem implies that a simply connected complete Riemannian manifold with nonpositive curvature operators must be diffeomorphic to Euclidean space.

The Hadamard-Cartan theorem contrasts with a 2008 theorem of Böhm and Wilking [1] which states that compact simply connected Riemannian manifolds with positive curvature operators are diffeomorphic to spheres. The reason the Hadamard-Cartan theorem was proven so much earlier is that it used only the theory of geodesics, while the Böhm-Wilking result used the Ricci flow of Hamilton that was developed to settle the Poincaré conjecture for three-manifolds.

Actually, however, the Hadamard-Cartan Theorem is somewhat stronger than what we stated. This is because it is sectional curvatures which govern the behavior of geodesics. Positive sectional curvatures cause geodesics emanating from a point $p \in M$ to converge, while nonpositive sectional curvatures cause them to diverge, and it is the latter fact which underlies the proof of the Hadamard-Cartan Theorem. Thus nonpositive curvature operators can be replaced by a weaker hypothesis in the Hadamarad-Cartan Theorem: ffor every $p \in M$,

$$\langle \mathcal{R}(x \wedge y), x \wedge y \rangle \leq 0, \quad \text{for all decomposable } x \wedge y \in \Lambda^2 T_p^* M,$$

the assumption that $(M, \langle \cdot, \cdot \rangle)$ has *nonpositive sectional curvatures*.

Asserting that a manifold has positive sectional curvatures is weaker than saying that it has positive curvature operators. Indeed, the complex and quaternionic projective spaces have positive sectional curvatures and are not homeomorphic to spheres. However, one can use convergence properties of geodesics to prove key theorems regarding Ricci curvature. One of the oldest of these is the theorem of Myers (1941) which asserts that a complete Riemannian manifold whose Ricci curvature satisfies the condition

$$\mathrm{Ric}(v, v) \geq \frac{n-1}{a^2} \langle v, v \rangle, \quad \text{for all } v \in TM, \tag{3.1}$$

where a is a nonzero real number, must be compact and its distance function must satisfy the condition $d(p, q) \leq \pi a$, for all $p, q \in M$. Like the Hadamard-Cartan Theorem the argument is based upon the influence of curvature on geodesics, this time the fact that positive curvature causes geodesics to focus. It has an analog in Lorentz geometry that led to the celebrated singularity theorems of Hawking and Penrose (see [17] for a discussion of these theorems).

Although Myers' Theorem puts a major restriction on the topology of compact manifolds of positive Ricci curvature, it is known that any manifold of dimension at least three has a complete Riemannian metric with negative Ricci curvature and finite volume [14]. Thus there are no significant topological restrictions on manifolds of negative Ricci or scalar curvature. The question of which compact manifolds admit metrics of positive scalar curvature, on the other hand, has generated numerous important theorems, many using techniques from the theory of linear elliptic equations on Riemannian manifolds, as well as spin geometry [11]. Whether a compact manifold of dimension four has a metric of positive scalar curvature often depends on invariants which go beyond the usual topological invariants of algebraic topology. Indeed, existence of positive scalar curvature metrics on compact four-manifolds is related to the more refined invariants of Donaldson and Seiberg and Witten, which depend upon the smooth structure of the manifold, not just its topological type.

We next build up the machinery needed to prove the theorems of Hadamard-Cartan and Myers, the basic theorems relating curvature to topology. Along the way we will prove uniqueness of the standard simply connected Riemannian manifolds of constant sectional curvature.

3.2 Parallel transport along curves

Let $(M, \langle \cdot, \cdot \rangle)$ be a pseudo-Riemannian manifold with Levi-Civita connection ∇. Suppose that $\gamma : [a, b] \to M$ is a smooth curve. A *smooth vector field in M along γ* is a smooth function

$$X : [a, b] \to TM \quad \text{such that} \quad X(t) \in T_{\gamma(t)}M \quad \text{for all } t \in [a, b].$$

We can define the *covariant derivative* of such a vector field along γ,

$$\nabla_{\gamma'} X : [a, b] \to TM, \quad \text{so that} \quad (\nabla_{\gamma'} X)(t) \in T_{\gamma(t)}M.$$

If $(x^1, \ldots, x^n)$ are local coordinates in terms of which

$$X(t) = \sum_{i=1}^{n} f^i(t) \frac{\partial}{\partial x^i}\bigg|_{\gamma(t)}, \qquad \gamma'(t) = \sum_{i=1}^{n} \frac{d(x^i \circ \gamma)}{dt}(t) \frac{\partial}{\partial x^i}\bigg|_{\gamma(t)},$$

then a short calculation shows that

$$\nabla_{\gamma'} X(t) = \sum_{i=1}^{n} \left[\frac{df^i}{dt}(t) + \sum_{j,k=1}^{n} \Gamma^i_{jk}(\gamma(t)) \frac{d(x^j \circ \gamma)}{dt}(t) f^k(t) \right] \frac{\partial}{\partial x^i}\bigg|_{\gamma(t)}. \tag{3.2}$$

We would write this last equation in tensor notation as

$$\sum_{j=1}^{n} f^i{}_{;j} \frac{d(x^j \circ \gamma)}{dt} = 0 \quad \text{along } \gamma.$$

Definition. We say that a vector field X along γ is *parallel* if $\nabla_{\gamma'} X \equiv 0$.

Proposition. *If $\gamma : [a, b] \to M$ is a smooth curve, $t_0 \in [a, b]$ and $v \in T_{\gamma(t_0)}M$, then there is a unique vector field X along γ which is parallel along γ and takes the value v at t_0:*

$$\nabla_{\gamma'} X \equiv 0 \quad \textit{and} \quad X(t_0) = v. \tag{3.3}$$

Proof: Suppose that in terms of local coordinates,

$$v = \sum_{i=1}^{n} a^i \frac{\partial}{\partial x^i}\bigg|_{\gamma(t_0)}.$$

Then it follows from (3.2) that (3.3) is equivalent to the linear initial value problem

$$\frac{df^i}{dt} + \sum_{j,k=1}^{n} \Gamma^i_{jk} \frac{d(x^j \circ \gamma)}{dt} f^k = 0, \quad f^i(t_0) = a^i.$$

It follows from the theory of linear ordinary differential equations (which is must simpler than the general theory of differential equations) that this initial value problem has a unique solution defined on the interval $[a, b]$.

If $\gamma : [a, b] \to M$ is a smooth path we can define a vector space isomorphism

$$\tau_\gamma : T_{\gamma(a)}M \to T_{\gamma(b)}M \quad \text{by} \quad \tau_\gamma(v) = X(b),$$

where X is the unique vector field along γ which is parallel and satisfies the initial condition $X(a) = v$. Similarly, we can define such an isomorphism τ_γ if γ is only piecewise smooth. We call τ_γ the *parallel transport* along γ.

Note that if X and Y are parallel along γ, then since the Levi-Civita connection ∇ is metric, $\langle X, Y \rangle$ is constant along γ; indeed,

$$\gamma' \langle X, Y \rangle = \langle \nabla_{\gamma'} X, Y \rangle + \langle X, \nabla_{\gamma'} Y \rangle = 0.$$

It follows that τ_γ is an isometry from $T_{\gamma(a)}M$ to $T_{\gamma(b)}M$.

Parallel transport depends very much on the path γ. For example, we could imagine parallel transport on the unit two-sphere $\mathbb{S}^2 \subseteq \mathbb{E}^3$ along the following piecewise smooth geodesic triangle γ: We start at the north pole $n \in \mathbb{S}^2$ and follow the prime meridian to the equator, then follow the equator through θ radians of longitude, and finally follow a meridian of constant longitude back up to the north pole. The resulting isometry from $T_n\mathbb{S}^2$ to itself is then just a rotation through the angle θ.

3.3 Geodesics and curvature

We now consider the differential equation that is generated when we have a "deformation through geodesics."

Suppose that $\gamma : [a, b] \to M$ is a smooth curve and that $\alpha : (-\epsilon, \epsilon) \times [a, b] \to M$ is a smooth map such that $\alpha(0, t) = \gamma(t)$. We can consider the map α as defining a family of smooth curves

$$\bar{\alpha}(s) : [a, b] \to M, \quad \text{for } s \in (-\epsilon, \epsilon), \quad \text{such that} \quad \bar{\alpha}(0) = \gamma,$$

if we set $\bar{\alpha}(s)(t) = \alpha(s, t)$. A *smooth vector field in M along α* is a smooth function

$$X : (-\epsilon, \epsilon) \times [a, b] \to TM$$
$$\text{such that} \quad X(s, t) \in T_{\alpha(s,t)}M \quad \text{for all } (s, t) \in (-\epsilon, \epsilon) \times [a, b].$$

We can take the covariant derivatives $\nabla_{\partial/\partial s}X$ and $\nabla_{\partial/\partial t}X$ of such a vector field along α just as we did for vector fields along curves. (In fact, we already carried out this construction in a special case in §1.6.) If $(x^1, \dots, x^n)$ are local coordinates in terms of which

$$X(s,t) = \sum_{i=1}^{n} f^i(s,t) \frac{\partial}{\partial x^i}\bigg|_{\alpha(s,t)}$$

$$\text{and we write} \quad \frac{\partial}{\partial s}(s,t) = \frac{\partial \alpha}{\partial s}(s,t) = \sum_{i=1}^{n} \frac{\partial (x^i \circ \alpha)}{\partial s}(s,t) \frac{\partial}{\partial x^i}\bigg|_{\alpha(s,t)},$$

then a short calculation yields

$$(\nabla_{\partial/\partial s} X)(s,t) = \sum_{i=1}^{n} \left[\frac{\partial f^i}{\partial s} + \sum_{j,k=1}^{n} (\Gamma^i_{jk} \circ \alpha) \frac{\partial (x^j \circ \alpha)}{\partial s} f^k \right](s,t) \, \frac{\partial}{\partial x^i}\bigg|_{\alpha(s,t)}.$$

In tensor notation we would write this as

$$\sum_{j=1}^{n} f^i{}_{:j} \frac{d(x^j \circ \alpha)}{dt} = 0 \quad \text{along } \alpha.$$

Of course, a similar local coordinate formula can be given for $\nabla_{\partial/\partial t} X$.

Important examples of vector fields along α include

$$\frac{\partial \alpha}{\partial s} \quad \text{and} \quad \frac{\partial \alpha}{\partial t},$$

and it follows quickly from the symmetry $\Gamma^k_{ij} = \Gamma^k_{ji}$ and the local coordinate formulae that

$$\nabla_{\partial/\partial s}\left(\frac{\partial \alpha}{\partial t}\right) = \nabla_{\partial/\partial t}\left(\frac{\partial \alpha}{\partial s}\right).$$

Just as in §1.8, the covariant derivatives do not commute, and this failure is described by the curvature: Thus if X is a smooth vector field along α,

$$\nabla_{\partial/\partial s} \circ \nabla_{\partial/\partial t} X - \nabla_{\partial/\partial t} \circ \nabla_{\partial/\partial s} X = R\left(\frac{\partial \alpha}{\partial s}, \frac{\partial \alpha}{\partial t}\right) X.$$

We say that α a *deformation* of γ and call

$$X(t) = \frac{\partial \alpha}{\partial s}(0,t) \in T_{\gamma(t)} M$$

the corresponding *deformation field*.

Proposition 1. *If α is a deformation such that each $\bar{\alpha}(s)$ is a geodesic, then the deformation field X must satisfy Jacobi's equation:*

$$\nabla_{\gamma'} \nabla_{\gamma'} X + R(X, \gamma')\gamma' = 0. \tag{3.4}$$

Proof: Since $\bar{\alpha}(s)$ is a geodesic for every s,

$$\nabla_{\partial/\partial t}\left(\frac{\partial\alpha}{\partial t}\right) \equiv 0 \quad \text{and hence} \quad \nabla_{\partial/\partial s}\nabla_{\partial/\partial t}\left(\frac{\partial\alpha}{\partial t}\right) = 0.$$

By the definition of curvature (see §1.8)

$$\nabla_{\partial/\partial t}\nabla_{\partial/\partial s}\left(\frac{\partial\alpha}{\partial t}\right) + R\left(\frac{\partial\alpha}{\partial s}, \frac{\partial\alpha}{\partial t}\right)\frac{\partial\alpha}{\partial t} = 0$$
$$\text{or} \quad \nabla_{\partial/\partial t}\nabla_{\partial/\partial t}\left(\frac{\partial\alpha}{\partial s}\right) + R\left(\frac{\partial\alpha}{\partial s}, \frac{\partial\alpha}{\partial t}\right)\frac{\partial\alpha}{\partial t} = 0.$$

Evaluation at $s = 0$ now yields (3.4), finishing the proof.

Remark. The Jacobi equation can be regarded as the linearization of the geodesic equation near a given geodesic γ.

Definition. A vector field X along a geodesic γ which satisfies the Jacobi equation (3.4) is called a *Jacobi field*.

Suppose that γ is a unit speed geodesic and that $(E_1, \ldots E_n)$ are parallel orthonormal vector fields along γ such that $E_1 = \gamma'$. We can then define the component functions of the curvature with respect to $(E_1, \ldots E_n)$ by

$$R(E_k, E_l)E_j = \sum_{i=1}^{n} R^i{}_{jkl}E_i.$$

where our convention is that the upper index i gets lowered to the first position. If $X = \sum f^i E_i$, then the Jacobi equation becomes

$$\frac{d^2 f^i}{dt^2} + \sum_{j=1}^{n} R^i_{1j1} f^j = 0. \tag{3.5}$$

This second order linear system of ordinary differential equations will possess a $2n$-dimensional vector space of solutions along γ. The Jacobi fields which vanish at a given point will form a linear subspace of dimension n.

Example. Suppose that $(M, \langle\cdot,\cdot\rangle)$ is a complete Riemannian manifold of constant sectional curvature k, in other words,

$$K(\sigma) \equiv k, \quad \text{whenever } \sigma \subset T_pM \text{ is a two-dimensional subspace, } p \in M.$$

Then the Riemann-Christoffel curvature tensor is given by

$$R(X,Y)W = k[\langle Y, W\rangle X - \langle X, W\rangle Y].$$

In this case, X will be a Jacobi field if and only if

$$\nabla_{\gamma'}\nabla_{\gamma'}X = -R(X,\gamma')\gamma' = k[\langle X, \gamma'\rangle\gamma' - \langle\gamma', \gamma'\rangle X].$$

Equivalently, if we assume that $\gamma : [0, b] \to M$ is unit speed and write $X = \sum f^i E_i$, where $(E_1, \ldots, E_n)$ is a parallel orthonormal frame along γ such that $E_1 = \gamma'$, then $R^1{}_{1j1} = 0$ for all j, and for $2 \le i \le n$,

$$R^i{}_{1j1} = \begin{cases} k, & \text{for } j = i, \\ 0, & \text{for } j \neq i. \end{cases}$$

Thus writing out (3.5) for the constant curvature case yields

$$\begin{cases} \frac{d^2 f^1}{dt^2} = 0, \\ \frac{d^2 f^i}{dt^2} = -k f^i, & \text{for } 2 \le i \le n. \end{cases}$$

The solutions are

$$f^1(t) = a^1 + b^1 t,$$

and for $2 \le i \le n$,

$$f^i(t) = \begin{cases} a^i \cos(\sqrt{k}t) + b^i \sin(\sqrt{k}t), & \text{for } k > 0, \\ a^i + b^i t, & \text{for } k = 0, \\ a^i \cosh(\sqrt{-k}t) + b^i \sinh(\sqrt{-k}t), & \text{for } k < 0, \end{cases} \tag{3.6}$$

Here $a^1, b^1, \ldots, a^n, b^n$ are constants of integration to be that are determined by the initial conditions.

Definition. Suppose that $\gamma : [a, b] \to M$ is a geodesic in a pseudo-Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$ with $\gamma(a) = p$ and $\gamma(b) = q$. We say that p and q are *conjugate* along γ if $p \neq q$ and there is a nonzero Jacobi field X along γ such that $X(a) = 0 = X(b)$.

For example, antipodal points on the sphere $\mathbb{S}^n(1)$ of radius one are conjugate along the great circle geodesics which join them, while $\mathbb{E}^n$ and $\mathbb{H}^n(1)$ do not have any conjugate points.

Suppose that p is a point in a geodesically complete pseudo-Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$ and $v \in T_p M$. We can then define a geodesic $\gamma_v : [0, 1] \to M$ by $\gamma_v(t) = \exp_p(tv)$. We say that v belongs to the *conjugate locus* in $T_p M$ if $\gamma_v(0)$ and $\gamma_v(1)$ are conjugate along γ_v.

Proposition 2. *A vector $v \in T_p M$ belongs to the conjugate locus if and only if $(\exp_p)_*$ is singular at v, that is, there is a nonzero vector $w \in T_v(T_p M)$ such that $(\exp_p)_{*v}(w) = 0$.*

Proof: We use the following construction: If $w \in T_v(T_p M)$, we define

$$\alpha_w : (\epsilon, \epsilon) \times [0, 1] \to M \quad \text{by} \quad \alpha_w(s, t) = \exp_p(t(v + sw)).$$

We set

$$X_w(t) = \frac{\partial \alpha_w}{\partial s}(0, t),$$

a Jacobi field along γ_v which vanishes at $\gamma_v(0)$. As w ranges throughout T_pM, X_w rnges throughout the n-dimensional space of Jacobi fields along γ_v which vanish at $\gamma_v(0)$.

$\Leftarrow$: If $(\exp_p)_{*v}(w) = 0$ where $w \neq 0$, then X_w is a nonzero Jacobi field along γ_v which vanishes at $\gamma_v(0)$ and $\gamma_v(1)$, so v belongs to the conjugate locus.

$\Rightarrow$: If v belongs to the conjugate locus, there is a nonzero Jacobi field along γ_v which vanishes at $\gamma_v(0)$ and $\gamma_v(1)$, and this vector field must be of the form X_w for some nonzero $w \in T_v(T_pM)$. But then $(\exp_p)_{*v}(w) = X_w(1) = 0$, and hence $(\exp_p)_*$ is singular at v.

Example. Let us consider the n-sphere $\mathbb{S}^n$ of constant curvature one. If p is the north pole in $\mathbb{S}^n$, it follows from (3.6) that the conjugate locus in $T_p\mathbb{S}^n$ is a family of concentric spheres of radius $k\pi$, where $k \in \mathbb{N}$.

3.4 The Hadamard-Cartan Theorem

We now turn to the proof of the Hadamard-Cartan Theorem, which states that the exponential map at any point is a smooth covering, in accordance with the following definition: A smooth map $\pi : \tilde{M} \to M$ is a *smooth covering* if π is onto, and each $q \in M$ possesses an open neighborhood U such that $\pi^{-1}(U)$ is a disjoint union of open sets each of which is mapped diffeomorphically by π onto U. Such an open set $U \subset M$ is said to be *evenly covered*.

Hadamard-Cartan Theorem I. *Let $(M, \langle \cdot, \cdot \rangle)$ be a complete connected Riemannian manifold with nonpositive sectional curvatures. Then the exponential map*

$$\exp_p : T_pM \longrightarrow M$$

is a smooth covering.

It is a theorem from basic topology as we will see in §3.5 (or see Chapter 1 of [8]) that a smooth covering of a simply connected space must be a diffeomorphism. Thus the Hadamard-Cartan Theorem will imply that a simply connected complete Riemannian manifold with nonpositive sectional curvatures must be diffeomorphic to Euclidean space.

Suppose that $(M, \langle \cdot, \cdot \rangle)$ is a complete Riemannian manifold. A point $p \in M$ is said to be a *pole* if the conjugate locus in T_pM is empty. For example, it follows from the explicit formulae we derived for Jacobi fields that any point in Euclidean space $\mathbb{E}^n$ or hyperbolic space $\mathbb{H}^n$ is a pole. The first step in establishing the Hadamard-Cartan Theorem consists of proving the following assertion:

Lemma 1. *If $(M, \langle \cdot, \cdot \rangle)$ is a complete connected Riemannian manifold whose curvature R satisfies the condition*

$$\langle R(x, y)y, x \rangle \leq 0, \quad \text{for all } x, y \in T_qM \text{ and all } q \in M,$$

then any point $p \in M$ is a pole.

Proof: Suppose that $p \in M$, $v \in T_pM$ and $\gamma(t) = \exp_p(tv)$. We need to show that $p = \gamma(0)$ and $q = \gamma(1)$ are not conjugate along γ.

Suppose, on the contrary, that X is a nonzero Jacobi field along γ which vanishes at $\gamma(0)$ and $\gamma(1)$. Thus

$$\nabla_{\gamma'}\nabla_{\gamma'}X + R(X,\gamma')\gamma' = 0, \quad \langle \nabla_{\gamma'}\nabla_{\gamma'}X, X\rangle = -\langle R(X,\gamma')\gamma', X\rangle \geq 0.$$

Hence

$$\int_0^1 \langle \nabla_{\gamma'}\nabla_{\gamma'}X, X\rangle dt \geq 0,$$

and integrating by parts yields

$$\int_0^1 \left[\frac{d}{dt}\langle \nabla_{\gamma'}X, X\rangle - \langle \nabla_{\gamma'}X, \nabla_{\gamma'}X\rangle \right] dt \geq 0,$$

and since the first term integrates to zero, we obtain

$$\int_0^1 -\langle \nabla_{\gamma'}X, \nabla_{\gamma'}X\rangle dt \geq 0.$$

It follows that $\nabla_{\gamma'}X \equiv 0$ and hence X is identically zero, a contradiction.

Thus to finish the proof of the Hadamard-Cartan Theorem, we need only prove the

Pole Theorem. *If $(M, \langle \cdot, \cdot \rangle)$ is a complete connected Riemannian manifold and $p \in M$ is a pole, then $\exp_p : T_pM \to M$ is a smooth covering.*

To prove this, we need to show that π is onto and each $p \in M$ has an open neighborhood U such that $\pi^{-1}(U)$ is the disjoint union of open sets, each of which is mapped diffeomorphically by π onto U.

Since $\exp_p$ is nonsingular at every $v \in T_pM$, we can define a Riemannian metric $\langle\langle \cdot, \cdot \rangle\rangle$ on T_pM by

$$\langle\langle x, y\rangle\rangle = \langle (\exp_p)_*(x), (\exp_p)_*(y)\rangle, \quad \text{for all } x, y \in T_v(T_pM).$$

Locally, $\exp_p$ is an isometry from $(T_pM, \langle\langle \cdot, \cdot \rangle\rangle)$ to $(M, \langle \cdot, \cdot \rangle)$ and it takes lines through the origin in T_pM to geodesics through $p \in M$. Hence lines through the origin must be geodesics in the Riemannian manifold $(T_pM, \langle\langle \cdot, \cdot \rangle\rangle)$. It therefore follows from the Hopf-Rinow Theorem from §2.6 that $(T_pM, \langle\langle \cdot, \cdot \rangle\rangle)$ is complete. Thus the theorem will follow from the following lemma:

Lemma 2. *If $\pi : \tilde{M} \to M$ is a local isometry of connected Riemannian manifolds with $\tilde{M}$ complete, then π is a smooth covering.*

Proof of lemma: Let $q \in M$. We need to show that q lies in an open set $U \subseteq M$ which is evenly covered, i.e. that $\pi^{-1}(U)$ is a disjoint union of open sets each of which is mapped diffeomorphically onto U.

There exists $\epsilon > 0$ such that $\exp_q$ maps the open ball of radius 2ϵ in T_qM diffeomorphically onto $\{r \in M : d(q,r) < 2\epsilon\}$. Let $\{\tilde{q}_\alpha : \alpha \in A\}$ be the set of points in $\tilde{M}$ which are mapped by π to q, and let

$$U = \{r \in M : d(r,q) < \epsilon\}, \qquad \tilde{U}_\alpha = \{\tilde{r} \in \tilde{M} : d(\tilde{r}, \tilde{q}_\alpha) < \epsilon\}.$$

Choose a point $\tilde{q}_\alpha \in \pi^{-1}(q)$, and let

$$B_\epsilon = \{v \in T_qM : \|v\| < \epsilon\}, \quad \tilde{B}_\epsilon = \{\tilde{v} \in T_{\tilde{q}_\alpha}\tilde{M} : \|\tilde{v}\| < \epsilon\}.$$

Since π takes geodesics to geodesics, we have a commutative diagram

$$\begin{array}{ccc} \tilde{B}_\epsilon & \xrightarrow{\pi_*} & B_\epsilon \\ \exp_{\tilde{q}_\alpha}\Big\downarrow & & \exp_q\Big\downarrow \\ \tilde{U}_\alpha & \xrightarrow{\pi} & U \end{array}$$

Note that $\exp_{\tilde{q}_\alpha}$ is globally defined and maps onto $\tilde{U}_\alpha$ because $\tilde{M}$ is complete, and π_* and $\exp_q$ are diffeomorphisms. It follows that π must map $\tilde{U}_\alpha$ diffeomorphically onto U.

If $\tilde{r} \in \tilde{U}_\alpha \cap \tilde{U}_\beta$, we would have geodesics $\tilde{\gamma}_\alpha$ and $\tilde{\gamma}_\beta$ of length $< \epsilon$ from $\tilde{q}_\alpha$ and $\tilde{q}_\beta$ to $\tilde{r}$. These would project to geodesics γ_α and γ_β of length $< \epsilon$ from q to $r = \pi(\tilde{r})$. By uniqueness of geodesics in normal coordinate charts, we would have $\gamma_\alpha = \gamma_\beta$. Since π is a local isometry, $\tilde{\gamma}_\alpha$ and $\tilde{\gamma}_\beta$ would satisfy the same initial conditions at $\tilde{r}$. Thus $\tilde{\gamma}_\alpha = \tilde{\gamma}_\beta$, so $\tilde{q}_\alpha = \tilde{q}_\beta$ and $\alpha = \beta$. We have shown that $\tilde{U}_\alpha \cap \tilde{U}_\beta \neq \emptyset$ only if $\alpha = \beta$.

Suppose now that $\tilde{r} \in \pi^{-1}(U)$, with $r = \pi(\tilde{r}) \in U$. Then there is a unit-speed geodesic γ from r to q of length $< \epsilon$. There is a unit-speed geodesic $\tilde{\gamma}$ in $\tilde{M}$ starting from $\tilde{r}$ whose initial conditions project to those of γ. Then $\pi \circ \tilde{\gamma} = \gamma$ and hence $\tilde{\gamma}$ proceeds from $\tilde{r}$ to $\tilde{q}_\alpha$ in time $< \epsilon$ for some $\alpha \in A$. Thus $\tilde{r} \in \tilde{U}_\alpha$ for some $\alpha \in A$, and

$$\pi^{-1}(U) = \bigcup \{\tilde{U}_\alpha : \alpha \in A\}.$$

Thus every point in M lies in an open set which is evenly covered. One easily checks that $\pi(\tilde{M})$ is both open and closed in M. Since M is connected, π is surjective and the lemma is proven. This in turn proves the Pole Theorem and the Hadamard-Cartan Theorem.

In the next section, we will describe the properties of the fundamental group of a smooth manifold and its relationship to smooth coverings. Using those properties, one can restate the Hadamard-Cartan Theorem as:

Hadamard-Cartan Theorem II. *Let $(M, \langle \cdot, \cdot \rangle)$ be a complete simply connected Riemannian manifold with nonpositive sectional curvatures. Then the exponential map*

$$\exp_p : T_pM \longrightarrow M$$

is a diffeomorphism.

3.5 The fundamental group*

For the benefit of readers who may not have taken the basic sequence in algebraic topology, this section gives a brief treatment of the notion of fundamental group. If you have not seen the fundamental group before, focus first on the definitions of fundamental group, simply connected and universal covers, take the main theorems on faith, and gradually return to the proofs after you see how these concepts are used. A more detailed treatement of the fundamental group can be found in Chapter 1 of [8], which is available on the internet.

3.5.1 Definition of the fundamental group*

Suppose that X is a metrizable topological space and that x_0 and x_1 are points of X. We let

$$P(X; x_0, x_1) = \{ \text{ continuous paths } \gamma : [0,1] \to X : \gamma(0) = x_0, \gamma(1) = x_1\}.$$

If $\gamma, \lambda \in P(X, x_0, x_1)$ we say that γ and λ are *homotopic* relative to the endpoints $\{0,1\}$ and write $\gamma \simeq \lambda$ if there is a continuous map $\alpha : [0,1] \times [0,1] \to X$ such that

$$\alpha(x,0) = x_0, \quad \alpha(s,1) = x_1, \quad \alpha(0,t) = \gamma(t), \quad \alpha(1,t) = \lambda(t).$$

We let $\pi_1(X, x_0, x_1)$ denote the quotient space of $P(X, x_0, x_1)$ by the equivalence relation defined by $\simeq$, and if $\gamma \in P(X, x_0, x_1)$, we let $[\gamma] \in \pi_1(X; x_0, x_1)$ denote the corresponding equivalence class. If d is a metric defining the topology on X, we define a metric on $P(X; x_0, x_1)$ by

$$d(\gamma, \lambda) = \sup\{d(\gamma(t), \lambda(t)) : t \in [0,1]\},$$

then $P(X; x_0, x_1)$ becomes a metric space itself and it has a resulting topology. In this case, $\pi_1(X, x_0, x_1)$ can be regarded as the collection of path components of $P(X; x_0, x_1)$.

Suppose that $\gamma \in P(X; x_0, x_1)$ and $\lambda \in P(X; x_1, x_2)$, and define $\gamma \cdot \lambda \in P(X; x_0, x_2)$ by

$$(\gamma \cdot \lambda)(t) = \begin{cases} \gamma(2t), & \text{for } t \in [0, 1/2], \\ \lambda(2t-1), & \text{for } t \in [1/2, 1]. \end{cases}$$

Finally, if $[\gamma] \in \pi_1(X; x_0, x_1)$ and $[\lambda] \in \pi_1(X; x_1, x_2)$, we claim that we can define a product $[\gamma][\lambda] = [\gamma \cdot \lambda] \in \pi_1(X; x_0, x_2)$. We need to show that this product

$$\pi_1(X; x_0, x_1) \times \pi_1(X; x_1, x_2) \longrightarrow \pi_1(X; x_0, x_2)$$

is well-defined; in other words, if $\gamma \simeq \tilde{\gamma}$ and $\lambda \simeq \tilde{\lambda}$, then

$$\gamma \cdot \lambda \simeq \tilde{\gamma} \cdot \tilde{\lambda}.$$

We show that $\gamma \simeq \tilde{\gamma}$ implies that $\gamma \cdot \lambda \simeq \tilde{\gamma} \cdot \lambda$. If α is the homotopy from γ to $\tilde{\gamma}$, we define

$$\beta : [0,1] \times [0,1] \to X \quad \text{by} \quad \beta(s.t) = \begin{cases} \alpha(s, 2t), & \text{for } t \in [0, 1/2], \\ \lambda(2t-1), & \text{for } t \in [1/2, 1]. \end{cases}$$

This gives the required homotopy from $\gamma \cdot \lambda$ to $\tilde{\gamma} \cdot \lambda$. The fact that $\lambda \simeq \tilde{\lambda}$ implies that $\gamma \cdot \lambda \simeq \gamma \cdot \tilde{\lambda}$ is quite similar.

The case where $x_0 = x_1$ is particularly important. We denote $\pi_1(X, x_0, x_0)$ by $\pi_1(X, x_0)$, and call it the *fundamental group* of X at x_0.

Theorem. *The multiplication operation defined above makes $\pi_1(X, x_0)$ into a group.*

To prove this we must first show that $\pi_1(X, x_0)$ has an identity. We let ε be the constant path, $\varepsilon(t) = x_0$ for all $t \in [0,1]$, and claim that

$$[\gamma \cdot \varepsilon] = [\gamma] = [\varepsilon \cdot \gamma], \quad \text{for all } [\gamma] \in \pi_1(X, x_0).$$

We prove the first equality, the other being similar; to do this, we need to construct a homotopy from $\gamma \cdot \varepsilon$ to γ. We simply define $\alpha : [0,1] \times [0,1] \to X$ by

$$\alpha(s,t) = \begin{cases} \gamma\left(\frac{2t}{s+1}\right), & \text{for } t \leq (1/2)(s+1), \\ x_0, & \text{for } t \geq (1/2)(s+1). \end{cases}$$

Then $\alpha(0,t) = (\gamma \cdot \varepsilon)(t)$ and $\alpha(1,t) = \gamma(t)$.

To prove associativity of multiplication, we need to show that if γ, λ and μ are elements of $P(X, x_0)$, the

$$(\gamma \cdot \lambda) \cdot \mu \simeq \gamma \cdot (\lambda \cdot \mu).$$

To do this, we define $\alpha : [0,1] \times [0,1] \to X$ by

$$\alpha(s,t) = \begin{cases} \gamma\left(\frac{4t}{s+1}\right), & \text{for } t \leq (1/4)(s+1), \\ \lambda(4t - s - 1), & \text{for } (1/4)(s+1) \leq t \leq (1/4)(s+2), \\ \mu\left(\frac{4t-s-2}{2-t}\right), & \text{for } t \geq (1/4)(s+2). \end{cases}$$

Then $\alpha(0,t) = (\gamma \cdot \lambda) \cdot \mu$ while $\alpha(1,t) = \gamma \cdot (\lambda \cdot \mu)$, so multiplication is indeed associative.

Finally, given $\gamma \in P(X, x_0)$, we define $\gamma^{-1}(t) = \gamma(1-t)$. To prove that $[\gamma^{-1}]$ is the inverse to $[\gamma]$, we must show that

$$\gamma \cdot \gamma^{-1} \simeq \varepsilon \quad \text{and} \quad \gamma^{-1} \cdot \gamma \simeq \varepsilon.$$

For the first of these, we define a homotopy $\alpha : [0,1] \times [0,1] \to X$ by

$$\alpha(s,t) = \begin{cases} \gamma(2t), & \text{for } t \leq (1/2)(1-s), \\ \gamma(1-s), & \text{for } (1/2)(1-s) \leq t \leq (1/2)(1+s), \\ \gamma(2-2t), & \text{for } t \geq (1/2)(1+s). \end{cases}$$

Then $\alpha(0,t) = (\gamma \cdot \gamma^{-1})(t)$ while $\alpha(s,1) = x_0$. The homotopy for $\gamma^{-1} \cdot \gamma \simeq \varepsilon$ is constructed in a similar fashion.

A continuous map $F : X \to Y$ with $F(x_0) = y_0$ induces a map

$$F_\sharp : P(X, x_0) \to P(Y, y_0) \quad \text{by} \quad F_\sharp(\gamma) = F \circ \gamma,$$

and it is easily checked that

$$\gamma \simeq \lambda \quad \Rightarrow \quad F \circ \gamma \simeq F \circ \lambda,$$

so that $F_\sharp$ induces a set-theoretic map

$$F_\sharp : \pi_1(X, x_0) \to \pi_1(Y, y_0).$$

Moreover, it is immediate that $F_\sharp$ is in fact a group homomorphism. We thus obtain a covariant function from the catergory of pointed metrizable topological spaces (X, x_0) and continuous maps $F : (X, x_0) \to (Y, y_0)$ preserving base points to the category of groups and group homomorphisms.

Remark. If X is pathwise connected, the fundamental groups based at different points are isomorphic. This is proven by techniques similar to those utilized in the proof of the preceding theorem. Indeed, if $\gamma \in P(X, x_0, x_1)$, we can define a map $h_\gamma : P(X, x_0) \to P(X, x_1)$ by

$$h_\gamma(\lambda) = \begin{cases} \gamma(1-3t), & \text{for } t \in [0, 1/3], \\ \lambda(3t-1), & \text{for } t \in [1/3, 2/3], \\ \gamma(3t-2), & \text{for } t \in [2/3, 1]. \end{cases}$$

By arguments similar to those used in the proof of the preceding theorem, one checks that this yields a well-defined group homomorphism

$$h_{[\gamma]} : \pi_1(X, x_0) \to \pi_1(X, x_1) \quad \text{by} \quad h_{[\gamma]}([\lambda]) = [h_\gamma \lambda].$$

Finally, if $\gamma^{-1} \in P(X, x_1, x_0)$ is defined by $\gamma^{-1}(t) = \gamma(1-t)$, one checks that $h_{[\gamma^{-1}]}$ is an inverse to $h_{[\gamma]}$.

Definition. We say that a metrizable topological space X is *simply connected* if it is pathwise connected and $\pi_1(X, x_0) = 0$. (The above remark shows that this condition does not depend on the choice of base point x_0.

3.5.2 Homotopy lifting*

To calculate the fundamental groups of spaces, one often uses the notion of covering space. A continuous map $\pi : \tilde{X} \to X$ is a *covering* if it is onto and every $x \in X$ lies in an open neighborhood U such that $\pi^{-1}(U)$ is a disjoint union of open sets each of which is mapped homeomorphically by π onto U. Such an open set is said to be *evenly covered*. Coverings have two important useful properties:

Homotopy Lifting Theorem. *Suppose that* $\pi : \tilde{X} \to X$ *is a covering. If*

$$\tilde{\gamma} : [0,1] \to \tilde{X}, \qquad \alpha : [0,1] \times [0,1] \to X$$

are continuous maps such that $\pi(\tilde{\gamma})((t) = \alpha(0,t)$*, then there exists a continuous map*

$$\tilde{\alpha} : [0,1] \times [0,1] \to \tilde{X}$$

such that $\tilde{\alpha}(0,t) = \tilde{\gamma}(t)$ *and* $\pi \circ \tilde{\alpha} = \alpha$*.*

In words, the homotopy α can be lifted to $\tilde{\alpha}$ taking values in $\tilde{X}$.

To prove this, we let $\mathcal{U}$ be an open cover of $[01,] \times [0,1] \subseteq \mathbb{R}^2$ consisting of open sets of the form $(a,b) \times (c,d)$ where a, b, c, d are rational and

$$\alpha((a,b) \times (c,d) \cap [0,1] \times [0,1])$$

lies in an evenly covered open subset of M. Since $[0,1] \times [0,1]$ is compact, a finite subcollection

$$\{(a_1, b_1 \times (c_1, d_1), \ldots, (a_k, b_k) \times (c_k, d_k)\}$$

of $\mathcal{U}$ covers $[0,1] \times [0,1]$. Choose a positive integer m so that ma_1, ..., md_k are all integers and let $n = 2m$. For $1 \leq i, j \leq n$, let

$$D_{ij} = \left[\frac{i-1}{n}, \frac{i}{n}\right] \times \left[\frac{j-1}{n}, \frac{j}{n}\right].$$

Then $\alpha(D_{ij})$ is contained in an evenly covered open subset U_{ij} of X.

The idea now is to define $\tilde{\alpha}$ inductively on D_{11}, D_{12}, ...,D_{1n}, D_{21}, ..., D_{2n}, ..., D_{n1}, ..., D_{nn}.

When we get to the (i,j)-stage, $\tilde{\alpha}$ is already defined on a connected part of the bounary of D_{ij} and the image lies in some $\tilde{U}_{ij}$ which is mapped homeomorphically onto an evenly covered open subset U_{ij} of X. We are forced to define $\tilde{\alpha}|D_{ij}$ by

$$\tilde{\alpha}|D_{ij} = (\pi|\tilde{U}_{ij})^{-1} \circ H|\alpha_{ij}.$$

This gives the unique extension of $\tilde{\alpha}$ to D_{ij} and an induction on i and j then finishes the proof of the Unique Path Lifting Theorem.

Remark. In the Homotopy Lifting Theorem, we could consider the case of a degenerate path $\tilde{\gamma}(t) \equiv \tilde{p}$ and a degenerate homotopy $\alpha(s,t) = \lambda(s)$. In this case, the Homotopy Lifting Theorem gives rise to a existence of a path $\tilde{\lambda}$ covering a given path λ in X. The following theorem shows that this lifted path is unique:

Unique Path Lifting Theorem. *Suppose that* $\pi : \tilde{X} \to X$ *is a covering. If* $\gamma, \lambda : [0,1] \to \tilde{X}$ *are two continuous maps such that* $\gamma(0) = \lambda(0)$ *and* $\pi \circ \gamma = \pi \circ \lambda$*, then* $\gamma = \lambda$*.*

To prove this, we let $J = \{t \in [0,1] : \gamma(t) = \lambda(t)\}$. We claim that J is both open and closed. Indeed, if $t \in J$, $\gamma(t) = \lambda(t)$ and $\gamma(t) = \lambda(t)$ lies in some open set $\tilde{U}$ which is mapped homeomorphically onto an open set U in X. Clearly $t \subseteq \gamma^{-1}(\tilde{U}) \cap \lambda^{-1}(\tilde{U}) \subseteq J$. Hence J is open.

On the other hand, if $t \in J$, there exist open sets $\tilde{U}_1$ and $\tilde{U}_2$ such that $\gamma(t) \in \tilde{U}_1$ and $\lambda(t) \in \tilde{U}_2$, where the two sets $\tilde{U}_1$ and $\tilde{U}_2$ are disjoint open sets mapped homemorphically by π onto an open subset U of X. Thus

$$t \in \gamma^{-1}(\tilde{U}_1) \cap \lambda^{-1}(\tilde{U}_2) \subseteq [0,1] - J,$$

and J is closed.

Example. Suppose that

$$\pi : \mathbb{R} \to \mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\} \quad \text{by} \quad \pi(t) = e^{2\pi i t}. \tag{3.7}$$

One checks that π is a smooth covering. We can use the previous theorems to calculate the fundamental group $\pi_1(\mathbb{S}^2, 1)$.

Indeed, suppose that $\gamma \in P(S^1, 1)$. Then the Unique Path Lifting Theorem implies that there is a unique $\tilde{\gamma} : [0,1] \to \mathbb{R}$ such that $\tilde{\gamma}(0) = 0$ and $\pi \circ \tilde{\gamma} = \gamma$. Since $\gamma(1) = 1$, there exists an element $k \in \mathbb{Z}$ such that $\tilde{\gamma}(1) = k$. If $\gamma \simeq \lambda \in P(\mathbb{S}^2, 1)$ by means of a homotopy $\alpha : [0,1] \times [0,1] \to \mathbb{S}^1$, we can use the Homotopy Lifting Theorem to construct $\tilde{\alpha} : [0,1] \times [0,1] \to \mathbb{R}$ such that $\alpha(0,t) = \tilde{\gamma}(t)$ and $\pi \circ \tilde{\alpha} = \alpha$. Unique path lifting implies that $\tilde{\alpha}(s,0) = 0$, $\tilde{\alpha}(s,1) = k$. and $\tilde{\alpha}(1,t) = \tilde{\lambda}(t)$. Thus $\tilde{\gamma}(1) = \tilde{\lambda}(1)$, and we obtain a well-defined map

$$h : \pi_1(\mathbb{S}^1, 1) \to \mathbb{Z} \quad \text{such that} \quad h([\gamma]) = \tilde{\gamma}(1).$$

It is easily checked that h is a homomorphism and since $h([e^{2\pi k i t}]) = k$ for $k \in \mathbb{Z}$, we see that h is surjection. Finally, if $h([\gamma]) = 0$, then $\tilde{\gamma}(1) = 0$ and hence $\tilde{\gamma}$ is homotopic to a constant, and γ itself must be homotopic to a constant. Thus we conclude that $\pi_1(\mathbb{S}^2, 1) \cong \mathbb{Z}$.

Degree of maps from S^1 to S^1: Suppose that $F : S^1 \to S^1$ is a continuous map. Then $\gamma : S^1 \to S^1$ determines a homomorphism of fundamental groups

$$\gamma_\sharp : \pi_1(S^1) \to \pi_1(S^1),$$

and since $\pi_1(S^1) \cong \mathbb{Z}$, this group homomorphism must be multiplication by some integer $n \in \mathbb{Z}$. We set $\deg(\gamma) = n$ and call it the *degree* of γ.

Regarding S^1 as

$$S^1 = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}.$$

we note that the differential form $xdy - ydx$ is closed but not exact. However, if π is the covering (3.7), then $\pi^*(ydx - xdy) = d\theta$ for some globallly defined real-valued function θ on $\mathbb{R}$. If $\gamma : S^1 \to S^1$ is smooth, then γ lifts to a smooth map $\tilde{\gamma} : S^1 \to \mathbb{R}$ and

$$\deg(\gamma) = \frac{1}{2\pi} \int_{\tilde{\gamma}} d\theta = \frac{1}{2\pi} \int_{\gamma} (xdy - ydx).$$

By very similar arguments, one could calculate the fundamental groups of many other spaces. For example, if $T^n = \mathbb{E}^n/\mathbb{Z}^n$, the usual n-torus, then

$$\pi_1(T^n, x_0) = \overbrace{\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}}^{n},$$

while if $\mathbb{R}P^n$ is the real projective space obtained by identifying antipodal points on $\mathbb{S}^n(1)$, then $\pi_1(\mathbb{R}P^n, x_0) = \mathbb{Z}_2$.

Finally, the Homotopy Lifting Theorem allows us to finish the argument that a complete simply connected Riemannian manifold which has nonpositive sectional curvatures must be diffeomorphic to $\mathbb{R}^n$:

Covering Theorem. *Suppose that $\pi : \tilde{M} \to M$ is a smooth covering, were $\tilde{M}$ and M are pathwise connected. If M is simply connected, then π is a diffeomorphism.*

To prove this we need only show that π is one-to-one. Suppose that $\tilde{p}$ and $\tilde{q}$ are points in $\tilde{M}$ such that $\pi(\tilde{p}) = \pi(\tilde{q})$. Since $\tilde{M}$ is pathwise connected, there is a continuous path $\tilde{\gamma} : [0,1] \to \tilde{M}$ such that $\tilde{\gamma}(1) = \tilde{p}$ and $\tilde{\gamma}(1) = \tilde{q}$. Let $\gamma = \pi \circ \tilde{\gamma}$. If $p = \pi(\tilde{p}) = \pi(\tilde{q})$, then $\gamma \in P(M, p)$. Since M is simply connected there is a continuous map $\alpha : [0,1] \times [0,1] \to M$ such that

$$\alpha(s,0) = p = \alpha(s,1), \quad \alpha(0,t) = \gamma(t), \quad \alpha(1,t) = p.$$

By the Homotopy Lifting Theorem, there is a continuous map $\tilde{\alpha} : [0,1] \times [0,1] \to \tilde{M}$ such that

$$\tilde{\alpha}(0,t) = \tilde{\gamma}(t), \qquad \pi \circ \tilde{\alpha} = \alpha.$$

The Unique Path Lifting Theorem implies that

$$\tilde{\alpha}(s,0) = \tilde{\alpha}(0,0) = \tilde{\gamma}(0) = \tilde{p} \quad \text{and} \quad \tilde{\alpha}(s,1) = \tilde{\alpha}(0,1) = \tilde{\gamma}(1) = \tilde{q}.$$

On the other hand, the Unique Path Lifting Theorem also implies that $\tilde{\alpha}(1,t)$ is constant. Hence $\tilde{p} = \tilde{q}$ and π is indeed one-to-one, exactly what we wanted to prove.

3.5.3 Universal covers*

The final fact often needed regarding the fundamental group and covering spaces is the existence of a universal cover.

Universal Cover Theorem. *If M is a connected smooth manifold, there exists a simply connected smooth manifold $\tilde{M}$ together with a smooth covering $\pi : \tilde{M} \to M$. Moreover, if $\tilde{M}_1$ is another simply connected smooth manifold with with smooth covering $\pi_1 : \tilde{M}_1 \to M$, there exists a smooth diffeomorphism $T : \tilde{M} \to \tilde{M}_1$ such that $\pi = \pi_1 \circ T$.*

We sketch the argument. A complete proof can be found in [8].

We start with a point $p_0 \in M$ and let

$$\tilde{M} = \{(p, [\gamma]) : p \in M, [\gamma] \in \pi_1(M; p_0, p)\}.$$

We can then define $\pi : \tilde{M} \to M$ by $\pi(p, [\gamma]) = p$. We need to define a metrizable topology on $\tilde{M}$, check that π is a covering and show that $\tilde{M}$ is simply connected. Then $\tilde{M}$ inherits a unique smooth manifold structure such that π is a local diffeomorphism.

Here is the idea for constructing the topology: Suppose that $\tilde{p} = (p, [\gamma]) \in \tilde{M}$ and let U be a contractible neighborhood of p within M. We then let

$$\tilde{U}_{(p,[\gamma])} = \{(q, [\lambda]) \in \tilde{M} : q \in U, [\lambda] = [\gamma \cdot \alpha], \text{ where } \alpha \text{ lies entirely within } U \}.$$

Then π maps $\tilde{U}_{(p,[\gamma])}$ homeomorphically onto U. From this one concludes that π is a covering.

To show that $\tilde{M}$ is simply connected, we suppose that $\tilde{\lambda} : [0,1] \to \tilde{M}$ is a continuous path with

$$\tilde{\lambda}(0) = \tilde{\lambda}(1) = (p_0, [\varepsilon]),$$

where ε is the constant path at p_0. Then $\lambda = \pi \circ \tilde{\lambda}$ is a closed curve from p_0 to p_0. For $t \in [0,1]$, we define

$$\lambda_t : [0,1] \to M \quad \text{by} \quad \lambda_t(s) = \lambda(st),$$

and define

$$\hat{\lambda} : [0,1] \to \tilde{M} \quad \text{by} \quad \hat{\lambda}(t) = (\lambda(t), [\lambda_t]).$$

Then $\hat{\lambda}(0) = (p_0, [\varepsilon])$ and $\pi \circ \hat{\lambda} = \lambda$. By the Unique Path Lifting Theorem, $\hat{\lambda} = \tilde{\lambda}$. But

$$\hat{\lambda}(1) = \tilde{\lambda}(1) \quad \Rightarrow \quad (p_0, [\lambda]) = (p_0, [\varepsilon]) \quad \Rightarrow \quad [\lambda] = [\varepsilon].$$

Thus λ is homotopic to a constant in M and by the Homotopy Lifting Theorem, $\tilde{\lambda}$ is homotopic to a constant in $\tilde{M}$.

If $\tilde{M}_1$ is another simply connected smooth manifold with with smooth covering $\pi_1 : \tilde{M}_1 \to M$, we choose $\tilde{p}_1 \in \tilde{M}_1$ such that $\pi_1(\tilde{p}_1) = p_0$. We then define

$$T : \tilde{M} \to \tilde{M}_1 \quad \text{by} \quad T(p, [\gamma]) = \tilde{\gamma}(1),$$

where $\tilde{\gamma} : [0,1] \to \tilde{M}_1$ is the unique lift of γ such that $\tilde{\gamma}(0) = \tilde{p}_1$. It is then relatively straightforward to check that T is a diffeomorphism such that $\pi = \pi_1 \circ T$.

Definition. If M is a connected smooth manifold and $\pi : \tilde{M} \to M$ is a smooth covering with $\tilde{M}$ simply connected, we say that $\pi : \tilde{M} \to M$ (or sometime $\tilde{M}$ itself) is the *universal cover* of M.

Note that the above construction of the universal cover shows that the elements of the fundamental group of M correspond in a one-to-one fashion with the elements of $\pi^{-1}(p_0)$. One can also show that there is a one-to-one correspondence between elements of the fundamental group of M and the group of deck transformations of the universal cover $\pi : \tilde{M} \to M$, where a *deck transformation* is a diffeomorphism

$$T : \tilde{M} \to \tilde{M} \quad \text{such that} \quad \pi \circ T = \pi.$$

3.6 Uniqueness of simply connected space forms

The Hadamard-Cartan Theorem has an important consequence regarding *space forms*, that is, complete Riemannian manifolds whose sectional curvatures are constant:

Space Form Theorem. *Let k be a given real number. If $(M, \langle\cdot,\cdot\rangle)$ and $(\tilde{M}, \widetilde{\langle\cdot,\cdot\rangle})$ are complete simply connected Riemannian manifolds of constant curvature k, then $(M, \langle\cdot,\cdot\rangle)$ and $(\tilde{M}, \widetilde{\langle\cdot,\cdot\rangle})$ are isometric.*

The proof divides into two cases, the case where $k \leq 0$ and the case where $k > 0$. It is actually the first case to which the Hadamard-Cartan Theorem directly applies.

Case I. Suppose that $k \leq 0$. In this case, the idea for the proof is really simple. Let $p \in M$ and $\tilde{p} \in \tilde{M}$. Then

$$\exp_p : T_pM \to M \quad \text{and} \quad \exp_{\tilde{p}} : T_{\tilde{p}}\tilde{M} \to \tilde{M}$$

are both diffeomorphisms by the Hadamard-Cartan Theorem. Let $\tilde{F} : T_pM \to T_{\tilde{p}}\tilde{M}$ be a linear isometry and let

$$F = \exp_{\tilde{p}} \circ \tilde{F} \circ \exp_p^{-1}.$$

Clearly F is a diffeomorphism, and it suffices to show that F is an isometry from M onto $\tilde{M}$.

Suppose that $q \in M$, $v \in T_qM$. Let $\tilde{q}$ be the corresponding point in $\tilde{M}$ and let $\tilde{v}$ be the corresponding vector in $T_{\tilde{q}}\tilde{M}$. It suffices to show that $\|v\| = \|\tilde{v}\|$.

Since M is complete, there is a geodesic $\gamma : [0,1] \to M_1$ such that $\gamma(0) = p$ and $\gamma(1) = q$. The geodesic γ is the image under $\exp_p$ of a line segment in T_pM. The commutativity of the diagram

$$\begin{array}{ccc} T_pM & \xrightarrow{\tilde{F}} & T_{\tilde{p}}\tilde{M} \\ {\scriptstyle \exp_p}\downarrow & & {\scriptstyle \exp_{\tilde{p}}}\downarrow \\ M & \xrightarrow{F} & \tilde{M} \end{array}$$

shows that F will take γ to a geodesic $\tilde{\gamma}$ from $\tilde{p}$ to $\tilde{q}$. Moreover, F takes Jacobi fields along γ to Jacobi fields along $\tilde{\gamma}$. Let V be the unique Jacobi field along

γ which vanishes at p and is equal to v at q. Then $\tilde{V} = F_* V$ is the Jacobi field along $\tilde{\gamma}$ which vanishes at $\tilde{p}$ and is equal to $\tilde{v}$ at $\tilde{q}$. Since $\tilde{F}$ is an isometry,

$$\text{the length of } \nabla_{\gamma'} V(0) = \text{ the length of } \nabla_{\tilde{\gamma}'} \tilde{V}(0).$$

It follows from the explicit formula (3.6) for Jacobi fields that the lengths of V and $\tilde{V}$ are equal at corresponding points, and hence $\|v\| = \|\tilde{v}\|$. This completes the proof when $k \leq 0$.

Case II. Suppose now that $k > 0$ and let $a = 1/\sqrt{k}$. It suffices to show that if $(\tilde{M}, \widetilde{\langle \cdot, \cdot \rangle})$ is an n-dimensional complete simply connected Riemannian manifold of constant curvature k, then it is globally isometric to $(\mathbb{S}^n(a), \langle \cdot, \cdot \rangle)$, where $\langle \cdot, \cdot \rangle$ is the standard metric on $\mathbb{S}^n(a)$. This case is a little more involved than the previous one, because $\mathbb{S}^n(a)$ is not diffeomorphic to its tangent space.

Lemma 1. *Suppose that $p \in \mathbb{S}^n(a)$, $\tilde{p} \in \tilde{M}$ and $\tilde{F} : T_p \mathbb{S}^n(a) \to T_{\tilde{p}} \tilde{M}$ is a linear isometry. If q is the antipodal point to p in $\mathbb{S}^n(a)$, then there is a unique smooth map*

$$F : \mathbb{S}^n(a) - \{q\} \to \tilde{M} \quad \textit{such that} \quad (F_*)_p = \tilde{F}.$$

The proof is similar to the construction given for Case I. Note first that $\exp_p$ maps

$$\{v \in T_p \mathbb{S}^n(a) : \sqrt{\langle v, v \rangle} < \pi\} \quad \text{diffeomorphically onto} \quad \mathbb{S}^n(a) - \{q\}.$$

Since we need $(F_*)_p = \tilde{F}$ and F must take geodesics to geodesics, we are forced to define $F : \mathbb{S}^n(a) - \{q\} \to \tilde{M}$ by

$$F = \exp_{\tilde{p}} \circ \tilde{F} \circ \exp_p^{-1},$$

just as in the previous case, establishing uniqueness. The argument given in Case I then shows that F is indeed an isometric mapping:

$$\widetilde{\langle F_*(v), F_*(w) \rangle} = \langle v, w \rangle, \quad \text{for } v, w \in T_q \mathbb{S}^n(a),$$

establishing existence.

Returning to the proof of the theorem, we choose a point $p \in M$ and apply Lemma 1 to obtain an isometric map $F : \mathbb{S}^n(a) - \{q\} \to \tilde{M}$, where q is the antipodal point to p. Let r be a point in $\mathbb{S}^n(a) - \{p, q\}$. Then $(F_*)_r : T_r \mathbb{S}^n(a) \to \tilde{M}$ is a linear isometry. We obtain Lemma 1 once again to obtain an isometric map $\hat{F} : \mathbb{S}^n(a) - \{s\} \to \tilde{M}$, where s is the antipodal point to r. By uniqueness, $F = \hat{F}$ on overlaps. Hence F extends to a map $\bar{F} : \mathbb{S}^n(a) \to \tilde{M}$. In particular, $\bar{F}$ takes the antipodal point q to p to a single point of $\tilde{M}$.

Clearly, $\bar{F}$ is an immersion and a local isometry. Since $\bar{F}$ maps M onto $\tilde{M}$ by commutativity of the diagram

$$\begin{array}{ccc} T_p\mathbb{S}^n(a) & \xrightarrow{\tilde{F}} & T_{\tilde{p}}\tilde{M} \\ \exp_p \big\downarrow & & \exp_{\tilde{p}} \big\downarrow \\ \mathbb{S}^n(a) & \xrightarrow{\bar{F}} & \tilde{M}, \end{array}$$

we see that $\tilde{M}$ is compact. Thus the rest of the proof will follow from:

Lemma 2. *Suppose that M and $\tilde{M}$ are compact smooth n-dimensional smooth manifolds and $\bar{F} : M \to \tilde{M}$ is an immersion. Then $\bar{F}$ is a smooth covering map.*

The proof is a straightforward exercise in the theory of covering spaces, and much simpler than the proof of Lemma 2 in §3.4 because any point in $\tilde{M}$ can have only finitely many preimages.

In the application to our theorem, $\tilde{M}$ is simply connected, so $\bar{F}$ is a diffeomorphism. Thus we obtain the required diffeomorphism from $(\mathbb{S}^n(a), \langle \cdot, \cdot \rangle)$ to $(\tilde{M}, \widetilde{\langle \cdot, \cdot \rangle})$.

3.7 Non simply connected space forms

There are many space forms which are not simply connected. Since these are covered by the simply connected space forms $\mathbb{E}^n$, $\mathbb{S}^n(a)$ and $\mathbb{H}^n(a)$, they provide many examples of smooth coverings, the theory of which is described in §3.5.

In the case of $\mathbb{E}^n$, we can take a basis $(v_1, \ldots, v_n)$ for $\mathbb{R}^n$ and consider the free abelian subgroup $\mathbb{Z}^n$ of $\mathbb{R}^n$ which is generated by the elements of the basis; thus

$$\mathbb{Z}^n = \{m_1 v_1 + \cdots m_n v_n : m_1, \ldots m_n \in \mathbb{Z}\}.$$

As usual, we let $\mathbb{E}^n$ denote $\mathbb{R}^n$ with the flat Euclidean metric. Then the quotient group $T^n = \mathbb{E}^n / \mathbb{Z}^n$ inherits a flat Riemannian metric; the resulting Riemannian manifold is called a flat n-torus. Note that $\pi_1(T^n) \cong \mathbb{Z}^n$.

In the positive curvature case, we can take identify antipodal points in $\mathbb{S}^n(1)$ obtaining the n-dimensional real projective space $\mathbb{R}P^n(1)$. The obvious projection $\pi : \mathbb{S}^n(1) \to \mathbb{R}P^n(1)$ is an important example of a smooth covering. Since the antipodal map is an isometry, there is a unique Riemannian metric $\langle \cdot, \cdot \rangle$ on $\mathbb{R}P^n(1)$ such that $\pi^* \langle \cdot, \cdot \rangle$ is the metric of constant curvature one on $\mathbb{S}^n(1)$. Thus $\mathbb{R}P^n(1)$ has a metric of constant curvature one and $\pi_1(\mathbb{R}P^n(1)) \cong \mathbb{Z}_2$.

There are many other Riemannian manifolds which have metrics of constant curvature one. To construct further examples with constant positive curvature in the case where $n = 3$, we make use of Hamilton's quaternions.

A quatenion Q can be regarded as a 2×2 matrix with complex entries,

$$Q = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}.$$

The set $\mathbb{H}$ of quaternions can be regarded as a four-dimensional real vector space with basis

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad E_x = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad E_y = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad E_z = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

Thus if

$$Q = \begin{pmatrix} t+iz & x+iy \\ -x+iy & t-iz \end{pmatrix}, \quad \text{we can write} \quad Q = tI + xE_x + yE_y + zE_z,$$

for unique choice of $(t, x, y, z) \in \mathbb{E}^4$. The determinant of the quaternion,

$$\det Q = (t+iz)(t-iz) - (-x+iy)(x+iy) = t^2 + x^2 + y^2 + z^2,$$

can be taken to be the Euclidean length of the quaternion.

Matrix multiplication makes $\mathbb{H} - \{0\}$ into a noncommutative Lie group. If

$$Q = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}, \quad \text{then} \quad Q^{-1} = \frac{1}{|Q|^2} \begin{pmatrix} \bar{a} & -b \\ \bar{b} & a \end{pmatrix}, \tag{3.8}$$

where $|Q|^2 = |a|^2 + |b|^2 = \det Q$. The determinant map

$$\det : \mathbb{H} - \{0\} \longrightarrow \mathbb{R}^+ = \text{(positive real numbers)}$$

is a group homomorphism when the group operation on $\mathbb{R}^+$ is ordinary multiplication, and we can identify $\mathbb{S}^3(1)$ with the group of unit-length quaternions

$$\{Q \in \mathbb{H} : \det Q = 1\} = \{tI + xE_x + yE_y + zE_z : t^2 + x^2 + y^2 + z^2 = 1\}.$$

Since $\mathbb{S}^3(1)$ is the kernel of the determinant map, it is a Lie subgroup of $\mathbb{H} - \{0\}$. It follows directly from (3.8) that

$$\mathbb{S}^3(1) \cong SU(2) = \{A \in GL(2, \mathbb{C} : A^{-1} = \bar{A}^T, \ \det A = 1\},$$

which is just the special unitary group.

If $A \in \mathbb{S}^3(1)$ and $Q \in \mathbb{H}$, then $\det(AQ) = \det(Q) = \det(QA)$ so the induced metric on $\mathbb{S}^3(1)$ is a biinvariant metric.

Moreover, if $A \in \mathbb{S}^3(1)$, we can define a linear isometry

$$\pi(A) : \mathbb{H} \to \mathbb{H} \quad \text{by} \quad \pi(A)(Q) = AQA^{-1}.$$

Since $\pi(A)$ preserves the t-axis, it can induces a map from the (x, y, z)-hyperplane to itself, and can be regarded as an element of

$$SO(3) = \{B \in GL(3, \mathbb{R}) : B^T B = I \text{ and } \det B = 1\}.$$

Moreover, the map $\pi : \mathbb{S}^3(1) \to SO(3)$ is a group homomorphism, and it is an easy exercise to check that the kernel of π is $\{\pm I\}$. It follows that $SO(3)$ is in fact diffeomorphic to $\mathbb{R}P^3$, and we can consider the group of unit-length quaternions as the universal cover of $SO(3)$. Note in particular, that $SO(3)$ is not simply connected, and in fact $\pi_1(SO(3)) \cong \mathbb{Z}_2$.

The group $SO(3)$ has many interesting finite subgroups. For example, the group of symmetries of a polygon of n sides is a group of order $2n$ called the dihedral group and denoted by $\mathbf{D}_n$. It is generated by a rotation through an angle $2\pi/n$ in the plane and be a reflection, which can be regarded as a rotation in an ambient $\mathbb{E}^3$. Thus the dihedral group can be regarded as a subgroup of $SO(3)$.

One also has groups of rotations of the five platonic solids, the tetrahedron, the cube, the octahedron, the dodecahedron and the icosahedron. The group of rotations of the tetrahedron $\mathbf{T}$ is just the alternating group on four letters and has order 12. The group of rotations $\mathbf{O}$ of the octahedron is isomorphic to the group of rotations of the cube and has order 24. Finally, the group of rotations $\mathbf{I}$ of the icosahedron is isomorphic to the group of rotations of the dodecahdron and has order 60. It is proven in §2.6 of Wolf [23] that the only finite subgroups of $SO(3)$ are cyclic and those isomorphic to $\mathbf{D}_n$, $\mathbf{T}$, $\mathbf{O}$ and $\mathbf{I}$.

One can take the preimage of these groups under the projection $\pi : \mathbb{S}^3(1) \to SO(3)$ obtaining the binary dihedral groups $\mathbf{D}_n^*$, the binary tetrahedral group $\mathbf{T}^*$, the binary octahedral group $\mathbf{O}^*$ and the binary icosahedral group $\mathbf{I}^*$. Thus one gets many examples of finite subgroups $G \subseteq \mathbb{S}^3(1)$. For each of these, one can construct the universal cover

$$\pi : \mathbb{S}^3(1) \to \mathbb{S}^3(1)/G,$$

left translations by elements of G being the deck transformations. Since these left translations are isometries, the quotient space $\mathbb{S}^3(1)/G$ inherits a Riemannian metric of constant curvature, the quotient space now having fundamental group G.

We can produce yet more examples by constructing finite subgroups of $SO(4)$ which act on $\mathbb{S}^3(1)$ without fixed points. For constructing such examples, it is helpful to know that $SU(2) \times SU(2)$ is a double cover of $SO(4)$. Indeed, if $(A_+, A_-) \in SU(2) \times SU(2)$, we can define

$$\pi(A_+, A_-) : \mathbb{H} \to \mathbb{H} \quad \text{by} \quad \pi(A_+, A_-)(Q) = A_+ Q A_-^{-1}.$$

This provides a surjective Lie group homomorphism $\pi : SU(2) \times SU(2) \to SO(4)$ with kernel $\{(I, I), (-I, -I)\}$. Once again, we find that $\pi_1(SO(4)) \cong \mathbb{Z}_2$.

Finally, one can show that any compact oriented connected surface of genus $g \geq 2$ possesses a Riemannian metric of constant negative curvature. In higher dimensions, there is an immense variety of nonsimply connected manifolds of constant negative curvature; such manifolds are called *hypberbolic manifolds*, and they possess a rich theory (see [21]).

3.8 Second variation of action

Curvature also affects the topology of M indirectly, through its effect on the stability of geodesics. We recall from §1.3 that geodesics are critical points of the action function $J : \Omega(M; p, q) \to \mathbb{R}$, where

$$\Omega(M; p, q) = \{ \text{ smooth paths } \gamma : [0, 1] \to M : \gamma(0) = p, \gamma(1) = q\},$$

and the action J is defined by

$$J(\gamma) = \frac{1}{2} \int_0^1 \langle \gamma'(t), \gamma'(t) \rangle_{\gamma(t)} dt.$$

Recall that a *variation* of γ is a map

$$\bar{\alpha} : (-\epsilon, \epsilon) \to \Omega(M; p, q)$$

such that $\bar{\alpha}(0) = \gamma$ and the map

$$\alpha : (-\epsilon, \epsilon) \times [a, b] \to M \quad \text{defined by} \quad \alpha(s, t) = \bar{\alpha}(s)(t),$$

is smooth. Our next goal is to calculate the second derivative of $J(\bar{\alpha}(s))$ at $s = 0$ when $\bar{\alpha}(0)$ is a geodesic, which gives a test for stability because at a local minimum the second derivative must be nonnegative. This second derivative is called the *second variation* of J at γ. We will see that the sectional curvature of M plays a crucial role in the formula for second variation.

The first step in deriving the second variation formula is to differentiate under the integral sign which yields

$$\begin{aligned} \frac{d^2}{ds^2} \left(J(\bar{\alpha}(s))\right)\Big|_{s=0} &= \frac{d^2}{ds^2} \left[\frac{1}{2} \int_0^1 \left\langle \frac{\partial \alpha}{\partial t}(s, t), \frac{\partial \alpha}{\partial t}(s, t) \right\rangle dt \right]\Big|_{s=0} \\ &= \left[\frac{1}{2} \int_0^1 \frac{\partial^2}{\partial s^2} \left\langle \frac{\partial \alpha}{\partial t}(s, t), \frac{\partial \alpha}{\partial t}(s, t) \right\rangle dt \right]\Big|_{s=0} \\ = \int_0^1 &\left[\left\langle \nabla_{\frac{\partial}{\partial s}} \left(\frac{\partial \alpha}{\partial t} \right), \nabla_{\frac{\partial}{\partial s}} \left(\frac{\partial \alpha}{\partial t} \right) \right\rangle + \left\langle \nabla_{\frac{\partial}{\partial s}} \nabla_{\frac{\partial}{\partial s}} \left(\frac{\partial \alpha}{\partial t} \right), \left(\frac{\partial \alpha}{\partial t} \right) \right\rangle \right] dt \Big|_{s=0} \\ = \int_0^1 &\left[\left\langle \nabla_{\frac{\partial}{\partial t}} \left(\frac{\partial \alpha}{\partial s} \right), \nabla_{\frac{\partial}{\partial t}} \left(\frac{\partial \alpha}{\partial s} \right) \right\rangle + \left\langle \nabla_{\frac{\partial}{\partial s}} \nabla_{\frac{\partial}{\partial t}} \left(\frac{\partial \alpha}{\partial s} \right), \left(\frac{\partial \alpha}{\partial t} \right) \right\rangle \right] dt \Big|_{s=0}. \end{aligned}$$

Using the curvature tensor, we can interchange the order of differentiation to obtain

$$\begin{aligned} \frac{d^2}{ds^2} \left(J(\bar{\alpha}(s))\right)\Big|_{s=0} &= \int_0^1 \left[\left\langle \nabla_{\frac{\partial}{\partial t}} \left(\frac{\partial \alpha}{\partial s} \right), \nabla_{\frac{\partial}{\partial t}} \left(\frac{\partial \alpha}{\partial s} \right) \right\rangle \right. \\ + \left\langle \nabla_{\frac{\partial}{\partial t}} \nabla_{\frac{\partial}{\partial s}} \left(\frac{\partial \alpha}{\partial s} \right), \left(\frac{\partial \alpha}{\partial t} \right) \right\rangle &- \left. \left\langle R \left(\frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial s} \right) \left(\frac{\partial \alpha}{\partial s} \right), \left(\frac{\partial \alpha}{\partial t} \right) \right\rangle \right] dt \Big|_{s=0}. \end{aligned} \tag{3.9}$$

Now comes an integration by parts, using the formula

$$\begin{aligned} &\frac{\partial}{\partial t} \left\langle \nabla_{\frac{\partial}{\partial s}} \left(\frac{\partial \alpha}{\partial s} \right), \left(\frac{\partial \alpha}{\partial t} \right) \right\rangle \\ &\qquad = \left\langle \nabla_{\frac{\partial}{\partial t}} \nabla_{\frac{\partial}{\partial s}} \left(\frac{\partial \alpha}{\partial s} \right), \left(\frac{\partial \alpha}{\partial t} \right) \right\rangle + \left\langle \nabla_{\frac{\partial}{\partial s}} \left(\frac{\partial \alpha}{\partial s} \right), \nabla_{\frac{\partial}{\partial t}} \left(\frac{\partial \alpha}{\partial t} \right) \right\rangle. \end{aligned}$$

Note that

$$\int_0^1 \frac{\partial}{\partial t}\left\langle \nabla_{\frac{\partial}{\partial s}}\left(\frac{\partial \alpha}{\partial s}\right), \left(\frac{\partial \alpha}{\partial t}\right)\right\rangle = 0$$

because $\alpha(s,0)$ and $\alpha(s,1)$ are both constant. Hence (3.9) becomes

$$\frac{d^2}{ds^2}\left(J(\bar{\alpha}(s))\right)\Big|_{s=0} = \int_0^1 \Big[\left\langle \nabla_{\frac{\partial}{\partial t}}\left(\frac{\partial \alpha}{\partial s}\right), \nabla_{\frac{\partial}{\partial t}}\left(\frac{\partial \alpha}{\partial s}\right)\right\rangle$$
$$-\left\langle R\left(\frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial s}\right)\left(\frac{\partial \alpha}{\partial s}\right), \left(\frac{\partial \alpha}{\partial t}\right)\right\rangle - \left\langle \nabla_{\frac{\partial}{\partial s}}\left(\frac{\partial \alpha}{\partial s}\right), \nabla_{\frac{\partial}{\partial t}}\left(\frac{\partial \alpha}{\partial t}\right)\right\rangle\Big] dt\Big|_{s=0}.$$

Finally, we evaluate at $s=0$ and use the fact that $\bar{\alpha}(0)=\gamma$ is a geodesic to obtain

$$\frac{d^2}{ds^2}\left(J(\bar{\alpha}(s))\right)\Big|_{s=0} = \int_0^1 \left[\langle \nabla_{\gamma'}X, \nabla_{\gamma'}X\rangle - \langle R(X,\gamma')\gamma', X\rangle\right] dt,$$

where X is the variation field defined by $X(t) = (\partial\alpha/\partial s)(0,t)$.

If $\gamma \in \Omega(M;p,q)$, we define the "tangent space" to the "infinite-dimensional manifold" $\Omega(M;p,q)$ at the point γ to be

$$T_\gamma\Omega(M;p,q) = \{\text{ smooth vector fields } X \text{ along } \gamma \ : X(0) = 0 = X(1)\}.$$

Definition. If $\gamma \in \Omega(M;p,q)$ is a geodesic, the *index form* of J at γ is the symmetric bilinear form

$$I : T_\gamma\Omega(M;p,q) \times T_\gamma\Omega(M;p,q) \to \mathbb{R}$$

defined by

$$I(X,Y) = \int_0^1 \left[\langle \nabla_{\gamma'}X, \nabla_{\gamma'}Y\rangle - \langle R(X,\gamma')\gamma', Y\rangle\right] dt, \tag{3.10}$$

for $X, Y \in T_\gamma\Omega(M;p,q)$.

By the polarization identity, the index form at a geodesic γ is the unique real-valued symmetric bilinear form I on $T_\gamma\Omega(M;p,q)$ such that

$$I(X,X) = \frac{d^2}{ds^2}\left(J(\bar{\alpha}(s))\right)\Big|_{s=0},$$

whenever $\bar{\alpha} : (-\epsilon,\epsilon) \to \Omega(M;p,q)$ is a smooth variation of γ with variation field X.

We can integrate by parts in (3.10) to obtain

$$I(X,Y) = -\int_0^1 \langle \nabla_{\gamma'}\nabla_{\gamma'}X + R(X,\gamma')\gamma', Y\rangle dt = \int_0^1 \langle L(X), Y\rangle dt,$$

where L is the *Jacobi operator*, defined by

$$L(X) = -\nabla_{\gamma'}\nabla_{\gamma'}X - R(X,\gamma')\gamma'.$$

Thus $I(X,Y) = 0$ for all $Y \in T_\gamma\Omega(M;p,q)$ if and only if X is a Jacobi field in $T_\gamma\Omega(M;p,q)$.

Note that the second variation argument we have given shows that if γ is a minimizing geodesic from p to q, the index form I at γ must be positive semi-definite.

3.9 Myers' Theorem

Recall that the *Ricci curvature* of a Riemannian manifold $(M,\langle\cdot,\cdot\rangle)$ is the bilinear form

$$\mathrm{Ric} : T_pM \times T_pM \to \mathbb{R} \quad \text{defined by} \quad \mathrm{Ric}(x,y) = (\text{Trace of } v \mapsto R(v,x)y).$$

Myers' Theorem (1941). *If $(M,\langle\cdot,\cdot\rangle)$ is a complete connected n-dimensional Riemannian manifold such that*

$$Ric(v,v) \geq \frac{n-1}{a^2}\langle v,v\rangle, \quad \text{for all } v \in TM, \tag{3.11}$$

where a is a nonzero real number, then M is compact and $d(p,q) \leq \pi a$, for all $p,q \in M$. Moreover, the fundamental group of M is finite.

Proof of Myers' Theorem: It suffices to show that $d(p,q) \leq \pi a$, for all $p,q \in M$, because closed bounded subsets of a complete Riemannian manifold are compact.

Suppose that p and q are points of M with $d(p,q) > \pi a$. Let $\gamma : [0,1] \to M$ be a minimal geodesic with $\gamma(0) = p$ and $\gamma(1) = q$. Let $(E_1,\ldots,E_n)$ be a parallel orthonormal frame along γ with $\gamma' = d(p,q)E_1$. Finally, let

$$X_i(t) = \sin(\pi t)E_i(t), \quad \text{for } t \in [0,1] \text{ and } 2 \leq i \leq n.$$

Then for each i, $2 \leq i \leq n$,

$$\nabla_{\gamma'}X_i = \pi\cos(\pi t)E_i, \qquad \nabla_{\gamma'}\nabla_{\gamma'}X_i = -\pi^2\sin(\pi t)E_i,$$

and hence

$$\langle\nabla_{\gamma'}\nabla_{\gamma'}X_i, X_i\rangle = -\pi^2\sin^2(\pi t).$$

On the other hand,

$$\langle R(X_i,\gamma')\gamma', X_i\rangle = \sin^2(\pi t)d(p,q)^2\langle R(E_i,E_1)E_1,E_i\rangle,$$

so

$$\langle\nabla_{\gamma'}\nabla_{\gamma'}X_i + R(X_i,\gamma')\gamma', X_i\rangle = \sin^2(\pi t)[d(p,q)^2\langle R(E_i,E_1)E_1,E_i\rangle - \pi^2].$$

Hence

$$\sum_{i=2}^{n} I(X_i, X_i) = -\sum_{i=2}^{n} \int_0^1 \langle \nabla_{\gamma'} \nabla_{\gamma'} X_i + R(X_i, \gamma')\gamma', X_i \rangle dt$$
$$= \int_0^1 \sin^2(\pi t) \left[(n-1)\pi^2 - d(p,q)^2 \sum_{i=2}^{n} \langle R(E_i, E_1)E_1, E_i \rangle \right] dt$$
$$= \int_0^1 \sin^2(\pi t) \left[(n-1)\pi^2 - d(p,q)^2 \mathrm{Ric}(E_1, E_1) \right] dt.$$

Since $\mathrm{Ric}(E_1, E_1) \geq (n-1)/a^2$, we conclude that

$$\sum_{i=2}^{n} I(X_i, X_i) < \int_0^1 \sin^2(\pi t) \left[(n-1)\pi^2 - (n-1)\frac{d(p,q)^2}{a^2} \right] dt,$$

the expression in brackets being negative because $d(p,q) > \pi a$. This contradicts the assumption that γ is a minimal geodesic, by the second variation argument given in the preceding section.

To show that the fundamental group of M is finite, we let $\tilde{M}$ be the universal cover of M, and give $\tilde{M}$ the Riemannian metric $\pi^*\langle \cdot, \cdot \rangle$, where $\pi : \tilde{M} \to M$ is the covering map. The Ricci curvature of $\tilde{M}$ satisfies the same inequality (3.11) as the Ricci curvature of M; moreover $\tilde{M}$ is complete. Thus by the above argument $\tilde{M}$ must also be compact. Hence if $p \in M$, $\pi^{-1}(p)$ is a finite set of points. But by the arguments presented in §3.5.3, the number of points in $\pi^{-1}(p)$ is the order of the fundamental group of M. Thus the fundamental group of M must be finite.

For example, one can apply Myers' Theorem to show that $S^1 \times S^2$ cannot admit a Riemannian metric of positive Ricci curvature, because $\pi_1(S^1 \times S^2, x_0) \cong \mathbb{Z}$, and is therefore not finite. A famous open question posed by Hopf asks whether $S^2 \times S^2$ admits a Riemannian metric with positive sectional curvatures.

Another important application is to Lie groups with biinvariant Riemannian metrics. If G is a Lie group with Lie algebra $\mathfrak{g}$, then the *center* of the Lie algebra is

$$\mathfrak{z} = \{ X \in \mathfrak{g} : [X, Y] = 0 \text{ for all } Y \in \mathfrak{g} \}.$$

Recall that any compact Lie group possesses a biinvariant Riemannian metric. This fact has a partial converse:

Corollary. *Suppose that G is a Lie group which has a biinvariant Riemannian metric. If the Lie algebra of G has trivial center, then G is compact.*

Proof: We use the explicit formula for curvature of biinvariant Riemannian metrics presented in §1.12. If E_1 is a unit-length element of $\mathfrak{g}$, we can extend E_1 to an orthonormal basis $(E_1, \ldots, E_n)$ for $\mathfrak{g}$, and conclude that

$$\mathrm{Ric}(E_1, E_1) = \sum_{i=2}^{n} \langle R(E_1, E_i)E_i, E_1 \rangle = \sum_{i=2}^{n} \frac{1}{4} \langle [E_1, E_i], E_1, E_i] \rangle > 0.$$

As E_1 ranges over the unit sphere in $\mathfrak{g}$, the continuous function $E_1 \mapsto \text{Ric}(E_1, E_1)$ must assume its minimum value. Hence $\text{Ric}(E_1, E_1)$ is bounded below, and it follows from Myers' Theorem that G is compact.

3.10 Synge's Theorem

Recall from §2.7, smooth closed geodesics can be regarded as critical points for the action function $J : \text{Map}(S^1, M) \to \mathbb{R}$, where $\text{Map}(S^1, M)$ is the space of smooth closed curves and

$$J(\gamma) = \frac{1}{2}\int_0^1 \langle \gamma'(t), \gamma'(t) \rangle_{\gamma(t)} dt.$$

Here S^1 is regarded as being obtained from $[0, 1]$ by identifying endpoints of the interval. It is interesting to consider conditions under which such critical points are stable.

In this case a *variation* of a point $\gamma \in \text{Map}(S^1, M)$ is a map

$$\bar{\alpha} : (-\epsilon, \epsilon) \to \text{Map}(S^1, M)$$

such that $\bar{\alpha}(0) = \gamma$ and the map

$$\alpha : (-\epsilon, \epsilon) \times S^1 \to M \quad \text{defined by} \quad \alpha(s, t) = \bar{\alpha}(s)(t),$$

is smooth. We can calculate the second derivative of $J(\bar{\alpha}(s))$ at $s = 0$ when $\bar{\alpha}(0)$ is a smooth closed geodesic, just as we did in §3.8, and in fact the derivation is a little simpler because we do not have to worry about contributions from the boundary of $[0, 1]$. Thus we obtain the analogous result

$$\frac{d^2}{ds^2}\left(J(\bar{\alpha}(s))\right)\bigg|_{s=0} = \int_0^1 \left[\langle \nabla_{\gamma'} X, \nabla_{\gamma'} X \rangle - \langle R(X, \gamma')\gamma', X \rangle\right] dt, \tag{3.12}$$

where now the variation field X is an element of

$$T_\gamma \text{Map}(S^1, M) = \{ \text{ smooth vector fields } X \text{ along } \gamma : S^1 \to M \ \},$$

and by polarization we have an index form

$$I : T_\gamma \text{Map}(S^1, M) \times T_\gamma \text{Map}(S^1, M) \to \mathbb{R}$$
$$\text{defined by} \quad I(X, Y) = \int_0^1 \left[\langle \nabla_{\gamma'} X, \nabla_{\gamma'} Y \rangle - \langle R(X, \gamma')\gamma', Y \rangle\right] dt,$$

which must be positive semi-definite if the smooth closed geodesic is stable. The fact that the sectional curvature appears in the second variation formula (3.12) implies that there is a relationship between sectional curvatures and the stability of geodesics. This fact can be exploited to yield relationships between curvature and topology, as the following theorem demonstrates.

Synge's Theorem (1936). *Suppose that $(M, \langle\cdot,\cdot\rangle)$ is a compact Riemannian manifold with positive sectional curvatures. If M is even-dimensional and orientable then M is simply connected.*

To prove this theorem, we use the Closed Geodesic Theorem from §2.7. Indeed, the nonconstant geodesic γconstructed in the proof is stable, and hence if $\bar{\alpha}(s)$ is any variation of γ,

$$\frac{d^2}{ds^2}\left(J(\bar{\alpha}(s))\right)\bigg|_{s=0} = \int_0^1 \left[\langle \nabla_{\gamma'} X, \nabla_{\gamma'} X\rangle - \langle R(X,\gamma')\gamma', X\rangle\right] dt \geq 0.$$

To construct an explicit variation that decreases action, we $p = \gamma(0)$ and make use of the parallel transport around γ:

$$\tau_\gamma : T_pM \longrightarrow T_pM.$$

If M is orientable, we know that this is an orientation-preserving isometry of T_pM. We can set

$$e_1 = \frac{1}{L(\gamma)}\gamma'(0),$$

and extend to a positively oriented orthonormal frame $(e_1, \ldots, e_n)$ for T_pM. From the canonical form theorem for special orthogonal transformations, if M is even-dimensional, say of dimension $2m$, we can choose the orthonormal basis so that τ_γ is represented by the matrix

$$\begin{pmatrix} \cos\theta_1 & \sin\theta_1 & & & \cdot & & \cdot & & \cdot \\ -\sin\theta_1 & \cos\theta_1 & & & & & & & \\ & \cdot & & \cos\theta_2 & \sin\theta_2 & \cdot & & \cdot & \\ & & & -\sin\theta_2 & \cos\theta_2 & & & & \\ & \cdot & & & \cdot & \cdot & & \cdot & \\ & \cdot & & & \cdot & \cdot & \cos\theta_m & \sin\theta_m & \\ & & & & & & -\sin\theta_m & \cos\theta_m & \end{pmatrix}. \tag{3.13}$$

Since $\tau_\gamma(e_1) = e_1$, we can arrange that $\theta_1 = 0$ in the first block. But then it follows that $\tau_\gamma(e_2) = e_2$, so there is a unit-length vector e_2 perpendicular to e_1 which is preserved by parallel transport around γ. We let X be the smooth vector field along γ obtained by parallel transport of e_2 around γ. Then since M has positive sectional curvatures,

$$I(X,X) = \int_0^1 \left[-\langle R(X,\gamma')\gamma', X\rangle\right] dt < 0.$$

Thus if $\bar{\alpha} : (-\epsilon, \epsilon) \to \mathrm{Map}(S^1, M)$ is a deformation of γ with deformation field X,

$$\frac{d}{ds}\left(J(\bar{\alpha}(s))\right)\bigg|_{s=0} = 0, \qquad \frac{d^2}{ds^2}\left(J(\bar{\alpha}(s))\right)\bigg|_{s=0} < 0.$$

This contradicts the stability of the minimal geodesic γ, finishing the proof of the theorem.

Remark. It follows from Synge's Theorem, that the only even-dimensional complete Riemannian manifolds of constant curvature one are the spheres $\mathbb{S}^{2m}(1)$ and the projective spaces $\mathbb{R}P^{2m}(1)$ of constant curvature one, the latter being nonorientable.

Exercise VII. Do not hand in. Show that an odd-dimensional compact manifold with positive sectional curvatures is automatically orientable by an argument similar to that provided for Synge's Theorem. You can follow the outline:

a. Sketch how you would modify the proof of the Closed Geodesic Theorem from §2.7 to show that if M were not orientable, one could construct a smallest smooth closed geodesic γ among curves around which parallel transport is orientation-reversing.

b. Show that in this case, the orthogonal matrix representing the parallel transport must have determinant -1, and its standard form is like (3.13) except for an additional 1×1 block containing -1.

c. As before, since the tangent vector to γ gets transported to itself, there is an additional unit vector e_2 perpendicular to γ which is transported to itself, and this implies that there is a nonzero parallel vector field X perpendicular to γ. Use the second variation formula to show that one can deform in the direction of X to obtain an orientation-reversing curve which is shorter, thereby obtaining a contradiction just as before.

Chapter 4

Cartan's method of moving frames

Differential geometry can be formulated in many related notations. For example, there is the invariant notation for a connection ∇ as well the Christoffel symbols used in the classical tensor of §2.4 developed by Ricci and Levi-Civita. But there are also the connection forms utilized by Élie Joseph Cartan (1869-1951) who dominated differential geometry during the first half of the twentieth century.

Cartan's approach to differential geometry emphasized the usefulness of differential forms, as opposed to more general tensor fields. Differential forms have the following advantages:

1. Differential forms pull back under smooth maps, while arbitrary tensor fields do not.
2. The exterior derivative is defined in terms of ordinary derivatives, not covariant derivatives, and in particular, the exterior derivative does not depend upon a choice of Riemannian metric.

Indeed, Cartan argued that complicated equations in index notation, such as (2.7 led to a "debauchery of indices" which often hid the simple underlying geometric concepts.

An additional advantage to differential forms and the exterior derivative is that they lead directly to de Rham cohomology, named after Cartan's student Georges de Rham, who showed that the de Rham cohomology of a smooth manifold was a topological invariant in 1931. It is de Rham cohomology which is the form of cohomology most useful for application in physics; for example, Maxwell's equations from electricity and magnetism are expressed most simply in terms of differential forms, as described in §4.5 of [17]. Bott and Tu [2] give a treatment of algebraic topology from the differential form point of view.

Cartan liked to use differential forms in connection with the theory of moving frames, which can be thought of as an extension of the Frenet trihedral used in studying curves in Euclidean space. His ideas ultimately led to the simplest proofs of the Gauss-Bonnet theorem for surfaces, as well as its extensions in higher dimensions.

4.1 An easy method for calculating curvature

Suppose that $(M, \langle \cdot, \cdot \rangle)$ is an n-dimensional Riemannian manifold, U an open subset of M. By a *moving orthonormal frame* on U, we mean an ordered n-tuple $(E_1, \ldots, E_n)$ of vector fields on U such that for each $p \in U$, $(E_1(p), \ldots, E_n(p))$ is an orthonormal basis for T_pM, for each $p \in M$. Suppose that M is oriented. Then we say that a moving orthonormal frame $(E_1, \ldots, E_n)$ is *positively oriented* if $(E_1(p), \ldots, E_n(p))$ is a positively basis for T_pM, for each $p \in M$.

Given a moving orthonormal frame on U, we can construct a corresponding *moving orthonormal coframe* $(\theta_1, \ldots, \theta_n)$ by requiring that each θ_i be the smooth one-form on U such that

$$\theta_i(E_j) = \delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{4.1}$$

We can then write the restriction of the Riemannian metric to U as

$$\langle \cdot, \cdot \rangle | U = \sum_{i=1}^{n} \theta_i \otimes \theta_i.$$

Indeed, if

$$v = \sum_{i=1}^{n} a^i E_i(p) \quad \text{and} \quad w = \sum_{i=1}^{n} b^i E_i(p),$$

then

$$\langle v, w \rangle = \sum_{i,j=1}^{n} a^i b^j \langle E_i(p), E_j(p) \rangle = \sum_{i=1}^{n} a^i b^i = \sum_{i=1}^{n} \theta_i \otimes \theta_i(v, w).$$

Conversely, if we can write the Riemannian metric $\langle \cdot, \cdot \rangle$ in the form

$$\langle \cdot, \cdot \rangle | U = \sum_{i=1}^{n} \theta_i \otimes \theta_i,$$

where $\theta_1, \ldots \theta_n$ are smooth one-forms on U, then $(\theta_1, \ldots, \theta_n)$ is a moving orthonormal coframe on U, and one can use (4.1) to define a moving orthonormal frame $(E_1, \ldots, E_n)$ on U.

Corresponding to a given orthonormal frame $(E_1, \ldots, E_n)$, we can define *connection one-forms* ω_{ij} for $1 \leq i, j \leq n$ by

$$\nabla_X E_j = \sum_{i=1}^{n} \omega_{ij}(X) E_i \quad \text{or} \quad \omega_{ij}(X) = \langle E_i, \nabla_X E_j \rangle, \tag{4.2}$$

as well as *curvature two-forms* Ω_{ij} by

$$R(X,Y)E_j = \sum_{i=1}^{n} \Omega_{ij}(X,Y)E_i \quad \text{or} \quad \Omega_{ij}(X,Y) = \langle E_i, R(X,Y)E_j \rangle. \tag{4.3}$$

Since $\langle E_i, E_j \rangle = \delta^i_j$ and the Levi-Civita connection preserves the metric,

$$0 = X\langle E_i, E_j \rangle = \langle \nabla_X E_i, E_j \rangle + \langle E_i, \nabla_X E_j \rangle = \omega_{ji}(X) + \omega_{ij}(X),$$

and hence the matrix $\omega = (\omega_{ij})$ of connection one-forms is skew-symmetric. It follows from the curvature symmetries that the matrix $\Omega = (\Omega_{ij})$ of curvature two-forms is also skew-symmetric.

Theorem. *If $(M, \langle \cdot, \cdot \rangle)$ is a Riemannian manifold and $(E_1, \ldots, E_n)$ is a moving orthonormal coframe defined on an open subset U of M, with dual coframe $(\theta_1, \ldots, \theta_n)$, then the connection and curvature forms satisfy the structure equations of Cartan:*

$$d\theta_i = -\sum_{j=1}^{n} \omega_{ij} \wedge \theta_j, \tag{4.4}$$

$$d\omega_{ij} = -\sum_{j=1}^{n} \omega_{ik} \wedge \omega_{kj} + \Omega_{ij}. \tag{4.5}$$

Moreover, the ω_{ij}'s are the unique collection of one-forms which satisfy (4.4) together with the skew-symmetry condition

$$\omega_{ij} + \omega_{ji} = 0. \tag{4.6}$$

The proof of the two structure equations is based upon the familiar formula for the exterior derivative of a one-form:

$$d\theta(X,Y) = X(\theta(Y)) - Y(\theta(X)) - \theta([X,Y]). \tag{4.7}$$

Indeed, to establish (4.4), we need to verify that

$$d\theta_i(E_k, E_l) = -\sum_{j=1}^{n} (\omega_{ij} \wedge \theta_j)(E_k, E_l).$$

and a straightforward calculation shows that

$$\begin{aligned}
&d\theta_i(E_k, E_l) + \sum_{j=1}^{n} (\omega_{ij} \wedge \theta_j)(E_k, E_l) \\
&\qquad = E_k(\theta_i(E_l)) - E_l(\theta_i(E_k)) - \theta_i([E_k, E_l]) + \sum_{j=1}^{n} (\omega_{ij} \wedge \theta_j)(E_k, E_l) \\
&\qquad = -\theta_i([E_k, E_l]) + \sum_{j=1}^{n} \omega_{ij}(E_k)\theta_j(E_l) - \sum_{j=1}^{n} \omega_{ij}(E_l)\theta_j(E_k) \\
&\qquad = \omega_{il}(E_k) - \omega_{ik}(E_l) - \theta_i([E_k, E_l]) \\
&\qquad = \langle E_i, \nabla_{E_k} E_l - \nabla_{E_l} E_k - [E_k, E_l] \rangle = 0,
\end{aligned}$$

where we have used the fact that the Levi-Civita connection is symmetric in the last line of the calculation. Similarly,

$$\begin{aligned}
&d\omega_{ij}(E_k, E_l) + \sum_{r=1}^{n} (\omega_{ir} \wedge \omega_{rj})(E_k, E_l) \\
&\qquad = E_k(\omega_{ij}(E_l)) - E_l(\omega_{ij}(E_k)) - \omega_{ij}([E_k, E_l]) \\
&\qquad\quad + \sum_{r=1}^{n} \omega_{ir}(E_k)\omega_{rj}(E_l) - \sum_{r=1}^{n} \omega_{ir}(E_l)\omega_{rj}(E_k) \\
&\qquad = E_k(\omega_{ij}(E_l)) - E_l(\omega_{ij}(E_k)) - \omega_{ij}([E_k, E_l]) \\
&\qquad\qquad - \sum_{r=1}^{n} \omega_{ri}(E_k)\omega_{rj}(E_l) + \sum_{r=1}^{n} \omega_{ri}(E_l)\omega_{rj}(E_k).
\end{aligned}$$

But

$$\begin{aligned}
E_k(\omega_{ij}(E_l)) = E_k\langle E_i, \nabla_{E_l} E_j \rangle &= \langle \nabla_{E_k} E_i, \nabla_{E_l} E_j \rangle + \langle E_i, \nabla_{E_k} \nabla_{E_l} E_j \rangle \\
&= \sum_{r=1}^{n} \omega_{ri}(E_k)\omega_{rj}(E_l) + \langle E_i, \nabla_{E_k} \nabla_{E_l} E_j \rangle,
\end{aligned}$$

while

$$E_l(\omega_{ij}(E_k)) = \sum_{r=1}^{n} \omega_{ri}(E_l)\omega_{rj}(E_k) + \langle E_i, \nabla_{E_l} \nabla_{E_k} E_j \rangle.$$

Hence

$$\begin{aligned}
&d\omega_{ij}(E_k, E_l) + \sum_{r=1}^{n} (\omega_{ir} \wedge \omega_{rj})(E_k, E_l) \\
&\qquad = \langle E_i, \nabla_{E_k} \nabla_{E_l} E_j \rangle - \langle E_i, \nabla_{E_l} \nabla_{E_k} E_j \rangle - \langle E_i, \nabla_{[E_k, E_l]} E_j \rangle \\
&\qquad = \langle E_i, R(E_k, E_l) E_j \rangle,
\end{aligned}$$

and the second structure equation is established.

Finally, to prove the uniqueness of the ω_{ij}'s, we suppose that we have two matrices of one-forms $\omega = (\omega_{ij})$ and $\tilde{\omega} = (\tilde{\omega}_{ij})$ which satisfy the first structure equation (4.4) and the skew-symmetry condition (4.6). Then the one-forms $\phi_{ij} = \omega_{ij} - \tilde{\omega}_{ij}$ must satisfy

$$\sum_{j=1}^{n} \phi_{ij} \wedge \theta_j = 0, \qquad \phi_{ij} + \phi_{ji} = 0.$$

We can write

$$\phi_{ij} = \sum_{j,k=1}^{n} f_{ijk} \theta_j \wedge \theta_k,$$

where each f_{ijk} is a smooth real-valued function on U. Note that

$$\sum_{j=1}^{n} \phi_{ij} \wedge \theta_j = 0 \quad \Rightarrow \quad f_{ijk} = f_{ikj},$$

while

$$\phi_{ij} + \phi_{ji} = 0 \quad \Rightarrow \quad f_{ijk} = -f_{jik}.$$

Hence

$$f_{ijk} = -f_{jik} = -f_{jki} = f_{kji} = f_{kij} = -f_{ikj} = -f_{ijk}.$$

Thus the functions f_{ijk} must vanish, and uniqueness is established.

The Cartan structure equations often provide a relatively painless procedure for calculating curvature:

Example. Suppose that we let

$$\mathbb{H}^n = \{(x^1, \ldots, x^{n-1}, y) \in \mathbb{R}^n : y > 0\},$$

and give it the Riemannian metric

$$\langle \cdot, \cdot \rangle = \frac{1}{y^2} \left(dx^1 \otimes dx^1 + \cdots + dx^{n-1} \otimes dx^{n-1} + dy \otimes dy \right).$$

In this case, we can set

$$\theta_1 = \frac{1}{y} dx^1, \quad \ldots, \quad \theta_{n-1} = \frac{1}{y} dx^{n-1}, \quad \theta_n = \frac{1}{y} dy, \tag{4.8}$$

thereby obtaining an orthonormal coframe $(\theta^1, \ldots, \theta^n)$ on $\mathbb{H}^n$.

Differentiating (4.8) yields

$$d\theta_1 = \frac{1}{y^2} dx^1 \wedge dy = \theta_1 \wedge \theta_n, \quad \cdots$$

$$d\theta_{n-1} = \frac{1}{y^2} dx^{n-1} \wedge dy = \theta_{n-1} \wedge \theta_n, \quad d\theta_n = 0.$$

In other words, if $1 \leq a \leq n-1$,

$$d\theta_a = \theta_a \wedge \theta_n, \quad \text{while} \quad d\theta_n = 0.$$

The previous theorem says that there is a unique collection of one-forms ω_{ij} which satisfy the first structure equation and the skew-symmetry condition. We can solve for these connection forms, obtaining

$$\omega_{ab} = 0, \quad \text{for } 1 \leq a, b \leq n-1, \qquad \omega_{an} = -\theta_a, \quad \text{for } 1 \leq a \leq n-1.$$

From the explicit form of the ω_{ij}'s, it is now quite easy to show that the curvature two-forms are given by

$$\Omega_{ij} = -\theta_i \wedge \theta_j, \quad \text{for } 1 \leq i, j \leq n.$$

It now follows from (4.3) that

$$\begin{aligned}\Omega_{ij}(X,Y) = \langle E_i, R(X,Y)E_j \rangle &= -\theta^i \wedge \theta^j(X,Y) \\ &= -[\langle E_i, X\rangle\langle E_j, Y\rangle - \langle E_j, X\rangle\langle E_i, Y\rangle],\end{aligned}$$

so that

$$\langle R(X,Y)W, Z\rangle = -[\langle X, Z\rangle\langle Y, W\rangle - \langle X, W\rangle\langle Y, Z\rangle].$$

In other words the Riemannian manifold $(\mathbb{H}^n, \langle \cdot, \cdot \rangle)$ has constant sectional curvatures.

4.2 The curvature of a surface

The preceding theory simplifies considerably when applied to two-dimensional Riemannian manifolds, and yields a particularly efficient method of calculating Gaussian curvature of surfaces (compare §1.10).

Indeed, if $(M, \langle \cdot, \cdot \rangle)$ is an oriented two-dimensional Riemannian manifold, U an open subset of M, then a moving orthonormal frame (E_1, E_2) is uniquely determined up to a rotation: If (E_1, E_2) and $(\tilde{E}_1, \tilde{E}_2)$ are two positively-oriented moving orthonormal frames on a contractible open subset $U \subseteq M$, then

$$\begin{pmatrix} E_1 & E_2 \end{pmatrix} = \begin{pmatrix} \tilde{E}_1 & \tilde{E}_2 \end{pmatrix} \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix},$$

for some smooth function $\alpha : U \to \mathbb{R}$. It is no surprise that the corresponding moving orthonormal coframes (θ_1, θ_2) and $(\tilde{\theta}_1, \tilde{\theta}_2)$ are related by a similar formula:

$$\begin{pmatrix} \theta_1 & \theta_2 \end{pmatrix} = \begin{pmatrix} \tilde{\theta}_1 & \tilde{\theta}_2 \end{pmatrix} \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}.$$

It follows that the volume form is invariant under change of positively oriented moving orthonormal frame:

$$\theta_1 \wedge \theta_2 = \tilde{\theta}_1 \wedge \tilde{\theta}_2.$$

We claim that the corresponding skew-symmetric matrix of connection forms

$$\omega = \begin{pmatrix} 0 & \omega_{12} \\ -\omega_{12} & 0 \end{pmatrix},$$

transforms by the rule

$$\omega_{12} = \tilde{\omega}_{12} - d\alpha.$$

To see this, recall that ω_{12} is defined by the formula

$$\nabla_X E_2 = \omega_{12}(X) E_1,$$

and hence

$$\nabla_X (-\sin\alpha \tilde{E}_1 + \cos\alpha \tilde{E}_2 = \omega_{12}(X)(\cos\alpha \tilde{E}_1 + \sin\alpha \tilde{E}_2),$$

which expands to yield

$$\begin{aligned} -\cos\alpha d\alpha(X)\tilde{E}_1 + \sin\alpha d\alpha(X)\tilde{E}_2 - \sin\alpha \nabla_X \tilde{E}_1 + \cos\alpha \nabla_X \tilde{E}_2 \\ = \omega_{12}(X)(\cos\alpha \tilde{E}_1 + \sin\alpha \tilde{E}_2). \end{aligned}$$

Taking the inner product with $\tilde{E}_1$ yields

$$-\cos\alpha d\alpha(X) + \cos\alpha \langle \tilde{E}_1, \nabla_X \tilde{E}_2 \rangle = \omega_{12}(X) \cos\alpha,$$

and dividing by $\cos\alpha$ yields the desired formula

$$-d\alpha + \tilde{\omega}_{12} = \omega_{12}.$$

The skew-symmetric matrix of curvature forms

$$\Omega = \begin{pmatrix} 0 & \Omega_{12} \\ -\Omega_{12} & 0 \end{pmatrix}$$

is now determined by the Cartan's second structure equation

$$\Omega_{12} = d\omega_{12} = d\tilde{\omega}_{12} = \tilde{\Omega}_{12}.$$

Note that the curvature form Ω_{12} is independent of the choice of positively oriented moving orthonormal frame. Indeed, it follows from (4.3) that

$$\Omega_{12}(E_1, E_2) = \langle E_1, R(E_1, E_2) E_2 \rangle = K,$$

where K is the Gaussian curvature of $(M, \langle \cdot, \cdot \rangle)$, and hence

$$\Omega_{12} = K\theta_1 \wedge \theta_2.$$

This formula makes it easy to calculate the curvature of a surface using differential forms.

Definition. Suppose that $(M, \langle \cdot, \cdot \rangle)$ is an oriented two-dimensional Riemannian manifold. A positively-oriented coordinate system $(U, (x, y))$ is said to be *isothermal* if on U

$$\langle \cdot, \cdot \rangle = e^{2\lambda}(dx \otimes dx + dy \otimes dy), \tag{4.9}$$

where $\lambda : U \to \mathbb{R}$ is a smooth function.

Here is a deep theorem whose proof lies beyond the scope of the course:

Theorem. *Any oriented two-dimensional Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$ has an atlas consisting of isothermal coordinate systems.*

A proof (using regularity theory of elliptic operators) can be found on page 378 of [20]. Assuming the theorem, we can ask: What is the relationship between positively oriented isothermal coordinate systems?

Suppose that (x^1, x^2) and (u^1, u^2) are two positively oriented coordinate systems on U with

$$\langle \cdot, \cdot \rangle = \sum_{i,j=1}^{n} g_{ij} dx^i \otimes dx^j = \sum_{i,j=1}^{n} \tilde{g}_{ij} du^i \otimes du^j,$$

where

$$g_{ij} = e^{2\lambda}\delta_{ij}, \qquad \tilde{g}_{ij} = e^{2\mu}\delta_{ij}.$$

Then since the g_{ij}'s transform as the components of a covariant tensor of rank two,

$$g_{ij} = \sum_{k,l=1}^{n} \tilde{g}_{kl} \frac{\partial u^k}{\partial x^i} \frac{\partial u^l}{\partial x^j}$$

or

$$\begin{aligned}\begin{pmatrix} e^{2\lambda} & 0 \\ 0 & e^{2\lambda} \end{pmatrix} &= \begin{pmatrix} \partial u^1/\partial x^1 & \partial u^2/\partial x^1 \\ \partial u^1/\partial x^2 & \partial u^2/\partial x^2 \end{pmatrix} \begin{pmatrix} e^{2\mu} & 0 \\ 0 & e^{2\mu} \end{pmatrix} \begin{pmatrix} \partial u^1/\partial x^1 & \partial u^1/\partial x^2 \\ \partial u^2/\partial x^1 & \partial u^2/\partial x^2 \end{pmatrix} \\ &= J^T \begin{pmatrix} e^{2\mu} & 0 \\ 0 & e^{2\mu} \end{pmatrix} J, \quad \text{where} \quad J = \begin{pmatrix} \partial u^1/\partial x^1 & \partial u^1/\partial x^2 \\ \partial u^2/\partial x^1 & \partial u^2/\partial x^2 \end{pmatrix}.\end{aligned}$$

Hence

$$J^T J = e^{2\lambda - 2\mu} I, \quad \text{or} \quad B^T B = I, \quad \text{where} \quad B = e^{\mu - \lambda} J.$$

Since both coordinates are positively oriented, $B \in SO(2)$, and hence if U is contractible, we can write

$$B = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix},$$

for some function $\alpha : U \to \mathbb{R}$. Thus we see that

$$J = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \text{where} \quad a = d \quad \text{and} \quad b = -c.$$

This implies that

$$\frac{\partial u^1}{\partial x^1} = \frac{\partial u^2}{\partial x^2}, \qquad \frac{\partial u^1}{\partial x^2} = -\frac{\partial u^2}{\partial x^1}.$$

These are just the Cauchy-Riemann equations, which express the fact that the complex-valued function $w = u^1 + iu^2$ is a holomorphic function of $z = x^1 + ix^2$.

Thus isothermal coordinates make an oriented two-dimensional Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$ into a one-dimensional complex manifold, in accordance with the following definition:

Definition. An *n-dimensional complex manifold* is a second-countable Hausdorff space M, together with a collection $\mathcal{A} = \{(U_\alpha, \phi_\alpha) : \alpha \in A\}$ of "charts," where each U_α is an open subset of M and each ϕ_α is a homeomorphism from U_α onto an open subset of $\mathbb{C}^n$, such that $\phi_\alpha \circ \phi_\beta^{-1}$ is holomorphic where defined, for all $\alpha, \beta \in A$. A one-dimensional complex manifold is also called a *Riemann surface*.

We say that $\mathcal{A}$ is the atlas of holomorphic charts. If $(M, \mathcal{A})$ and $(N, \mathcal{B})$ are two complex manifolds, we say that a map $F : M \to N$ is *holomorphic* if $\psi_\beta \circ F \circ \phi_\alpha^{-1}$ is holomorphic where defined, for all charts $(U_\alpha, \phi_\alpha) \in \mathcal{A}$ and $(V_\beta, \psi_\beta) \in \mathcal{B}$. Two complex manifolds M and N are *holomorphically equivalent* if there is a holomorphic map $F : M \to N$ which has a holomorphic inverse $G : N \to M$.

In particular, we can speak of holomorphically equivalent Riemann surfaces. Two Riemannian metrics $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$ on an oriented surface M are said to be *conformally equivalent* if there is a smooth function $\lambda : M \to \mathbb{R}$ such that

$$\langle \cdot, \cdot \rangle_1 = e^{2\lambda} \langle \cdot, \cdot \rangle_2.$$

This defines an equivalence relation, and given the existence of isothermal coordinates, it is clear that conformal equivalence classes of Riemannian metrics are in one-to-one correspondence with Riemann surface structures on a given oriented surface M.

Exercise VIII. (Do not hand in.) a. Suppose that $(M, \langle \cdot, \cdot \rangle)$ is an oriented two-dimensional Riemannian manifold with isothermal coordinate system $(U, (x, y))$ so that the Riemannian is given by (4.9). Use the method of moving frames to show that the Gaussian curvature of M is given by the formula

$$K = -\frac{1}{e^{2\lambda}} \left(\frac{\partial^2 \lambda}{\partial x^2} + \frac{\partial^2 \lambda}{\partial y^2} \right).$$

Hint: To start with, let $\theta_1 = e^\lambda dx$ and $\theta_2 = e^\lambda dy$. Then calculate ω_{12} and Ω_{12}.

b. Consider the Poincaré disk, the open disk $D = \{(x, y) \in \mathbb{R}^2\}$ with the Poincaré metric

$$\langle \cdot, \cdot \rangle = \frac{4}{[1 - (x^2 + y^2)]^2} (dx \otimes dx + dy \otimes dy).$$

Show that the Gaussian curvature of $(D, \langle \cdot, \cdot \rangle)$ is given by $K = -1$.

c. Show that reflections through lines passing through the origin are isometries and hence that lines passing through the origin in D are geodesics for the Poincaré metric. Show that the boundary of D is infinitely far away along any of these lines, and hence the geodesics through the origin can be extended indefinitely. Conclude from the Hopf-Rinow theorem that $(D, \langle \cdot, \cdot \rangle)$ is a complete Riemannian manifold, hence isometric to the model of the hyperbolic plane we constructed in §1.8.

4.3 The Gauss-Bonnet formula for surfaces

We now sketch the proof of the Gauss-Bonnet formula for surfaces in a version that suggests how it might be extended to n-dimensional oriented Riemannian manifolds. (See [19] for a more leisurely treatment.)

We start with an oriented two-dimensional Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$ without boundary and a smooth vector field $X : M \to TM$ with finitely many zeros at the points $p_1, p_2, \ldots, p_k$ of M. Let $V = M - \{p_1, \ldots, p_k\}$ and define a unit length vector field $Y : V \to TM$ by $Y = X/\|X\|$.

The *covariant differential* $DY = \nabla.Y$ of Y is the endomorphism of TM defined by $v \mapsto \nabla_v Y$. We will find it convenient to regard $\nabla.Y$ as a one-form with values in TM:

$$DY = \nabla.Y \in \Omega^1(TM).$$

If (E_1, E_2) is a positively oriented orthonormal moving frame defined on an open subset $U \subseteq M$, we can write

$$Y = f_1 E_1 + f_2 E_2 \quad \text{on} \quad U \cap V,$$

and

$$\nabla.Y = (df_1 + \omega_{12} f_2)E_1 + (df_2 - \omega_{12} f_1)E_2.$$

We let J denote counterclockwise rotation through 90 degrees in the tangent bundle, so that

$$JE_1 = E_2, \quad JE_2 = -E_1,$$

and

$$JY = -f_2 E_1 + f_1 E_2 \quad \text{on} \quad U \cap V.$$

Then

$$\psi = \langle JY, DY \rangle = f_1(df_2 - \omega_{12} f_1) - f_2(df_1 + \omega_{12} f_2) = f_1 df_2 - f_2 df_1 - \omega_{12}$$

is a globally defined one-form on $V = M - \{p_1, \ldots, p_k\}$ which depends upon X, and since $d(f_1 df_2 - f_2 df_1) = 0$,

$$d\psi = -\Omega_{12} = -K \theta_1 \wedge \theta_2, \tag{4.10}$$

where $\theta_1 \wedge \theta_2$ is the area form on M. The idea behind the proof of the Gauss-Bonnet formula is to apply Stokes's Theorem to (4.10).

Let ϵ be a small positive number. For each zero p_i of X, we let $C_\epsilon(p_i) = \{q \in M : d(p_i, q) = \epsilon\}$, a circle which inherits an orientation by regarding it as the boundary of

$$D_\epsilon(p_i) = \{q \in M : d(p_i, q) \leq \epsilon\}.$$

Definition. The *rotation index* of X about p_i is

$$\omega(X, p_i) = \frac{1}{2\pi} \lim_{\epsilon \to 0} \int_{C_\epsilon(p_i)} \psi,$$

if this limit exists.

Lemma. *If $(M, \langle \cdot, \cdot \rangle)$ is a two-dimensional compact oriented Riemannian manifold and X is a vector field on M with finitely many isolated zeros, then at each zero p_i the rotation index $\omega(X, p_i)$ exists and depends only on X, not on the choice of Riemannian metric on M.*

To prove this, we make use of the notion of the degree $\deg(F)$ of a continuous map F from S^1 to itself, as described in §3.5.2. Recall that such a map $F : S^1 \to S^1$ determines a homomorphism of fundamental groups

$$F_\sharp : \pi_1(S^1) \to \pi_1(S^1),$$

and since $\pi_1(S^1) \cong \mathbb{Z}$, this group homomorphism must be multiplication by some integer $n \in \mathbb{Z}$. We set $\deg(F) = n$.

Note that for $\epsilon > 0$ sufficiently small, we can define a map

$$F_\epsilon : C_\epsilon(p_i) \to S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} \quad \text{by} \quad F_\epsilon(q) = (f_1(q), f_2(q)).$$

Then

$$\deg(F_\epsilon) = \frac{1}{2\pi} \int_{C_\epsilon(p_i)} F_\epsilon^*(x dy - y dx)$$
$$= \frac{1}{2\pi} \int_{C_\epsilon(p_i)} f_1 df_2 - f_2 df_1 = \frac{1}{2\pi} \lim_{\epsilon \to 0} \int_{C_\epsilon(p_i)} \psi.$$

Thus $\omega(X, p_i)$ does indeed exist and is an integer.

To see that this integer is independent of the choice of Riemannian metric, note that any two Riemannian metrics $\langle \cdot, \cdot \rangle_0$ and $\langle \cdot, \cdot \rangle_1$ can be connected by a smooth one-parameter family

$$t \mapsto \langle \cdot, \cdot \rangle_t = (1 - t)\langle \cdot, \cdot \rangle_0 + t \langle \cdot, \cdot \rangle_1.$$

We can let $\omega_t(X, p_i)$ be the degree of X at p_i with respect to $\langle \cdot, \cdot \rangle_t$. Then $\omega_t(X, p_i)$ is a continuously varying integer and must therefore be constant.

It follows from transversality theory (as presented for example in Hirsch [10]) that any compact oriented surface possesses a vector field which has finitely many nondegenerate zeros. If X is any such vector field, we can apply Stokes's Theorem to $W = M - \bigcup_{i=1}^k D_\epsilon(p_i)$:

$$\int_W K\theta_1 \wedge \theta_2 = \int_W -d\psi = \sum_{i=1}^k \int_{C_\epsilon(p_i)} \psi, \tag{4.11}$$

the extra minus sign coming from the fact that the orientation $C_\epsilon(p_i)$ inherits from W is opposite to the orientation it receives as boundary of $D_\epsilon(p_i)$. In the limit as $\epsilon \to 0$, we obtain

$$\int_M K\theta_1 \wedge \theta_2 = 2\pi \sum_{i=1}^k \omega(X, p_i). \tag{4.12}$$

Since the left-hand side of (4.12) does not depend on the vector field while the right-hand side does not depend on the metric, neither side can depend on either the vector field or the metric, so both sides must equal an integer-valued topological invariant of compact oriented smooth surfaces $\chi(M)$, which is called the *Euler characteristic* of M. Thus we obtain two theorems:

Poincaré Index Theorem. *Suppose that M be a two-dimensional compact oriented smooth manifold and that X is a vector field on M with finitely many isolated zeros at the points p_1, p_2, $\dots$, p_k. Then*

$$\sum_{i=1}^k \omega(X, p_i) = \chi(M).$$

Gauss-Bonnet Theorem. *Let $(M, \langle \cdot, \cdot \rangle)$ be a two-dimensional compact oriented Riemannian manifold with Gaussian curvature K and area form $\theta_1 \wedge \theta_2$. Then*

$$\frac{1}{2\pi} \int_M K\theta_1 \wedge \theta_2 = \chi(M).$$

Recall that a compact connected oriented surface is diffeomorphic to a sphere with h handles M_h. We can imbed M_h into $\mathbb{E}^3$ in such a way that the standard height function has exactly one nondegenerate maximum and one nodegenerate minimum, and $2h$ nondegenerate saddle points. The gradient X of the height function is then a vector field with nondegenerate zeros at the critical points of the height function. The maximum and minimum are zeros with rotation index one while each saddle point is a zero with rotation index -1. Thus $\chi(M_h) = 2 - 2h$.

The previous theorems can be extended to manifolds with boundary. In this case we consider a vector field X which has finitely many zeros at the points p_1, p_2, $\dots$, p_k in the interior of M, is perpendicular to ∂M along ∂M and points outward along ∂M. As before, we let $V = M - \{p_1, \dots, p_k\}$, define $Y : V \to TM$ by $Y = X/\|X\|$ and set $\psi = \langle JY, DY \rangle$. Then just as before

$$d\psi = -\Omega_{12} = -K\theta_1 \wedge \theta_2.$$

But this time, when we apply Stokes's Theorem to $W = M - \bigcup_{i=1}^k D_\epsilon(p_i)$ we obtain

$$\int_W K\theta_1 \wedge \theta_2 = \int_W -d\psi = -\int_{\partial M} \psi + \sum_{i=1}^k \int_{C_\epsilon(p_i)} \psi.$$

Thus when we let $\epsilon \to 0$, we obtain

$$\int_M K\theta_1 \wedge \theta_2 + \int_\partial M\psi = 2\pi \sum_{i=1}^k \omega(X, p_i). \tag{4.13}$$

Along ∂M, one can show that

$$\langle JY, DY \rangle = \kappa_g ds,$$

where κ_g is known as the geodesic curvature. Note that $\kappa_g = 0$ when ∂M consists of geodesics. As before, the left-hand side of (4.13) does not depend on the vector field while the right-hand side does not depend on the metric, so neither side can depend on either the vector field or the metric. The two sides must therefore equal a topological invariant which we call the *Euler characteristic* of M once again, thereby obtaining two theorems:

Poincaré Index Theorem for Surfaces with Boundary. *Suppose that M be a two-dimensional compact oriented smooth manifold with boundary ∂M and that X is a vector field on M with finitely many isolated zeros at the points $p_1, p_2, \ldots, p_k$ in the interior of M which is perpendicular to ∂M and points out along ∂M. Then*

$$\sum_{i=1}^k \omega(X, p_i) = \chi(M).$$

Gauss-Bonnet Theorem for Surfaces with Boundary. *Let M be a compact oriented smooth surface in with boundary ∂M. Then*

$$\int_M K dA + \int_{\partial S} \kappa_g ds = 2\pi\chi(M),$$

where f is the number of faces, e is the number of edges and v is the number of vertices in $\mathcal{T}$.

The celebrated uniformization theorem for Riemann surfaces shows that any Riemann surface has a complete Riemannian metric in its conformal equivalence class that has constant Gaussian curvature. For compact oriented surfaces, see that

1. the sphere has a Riemannian metric of constant curvature $K = 1$,
2. the torus T^2 has a metric of constant curvature $K = 0$,
3. and we will show in the next section that a sphere with h handles, where $h \geq 2$, has a Riemannian metric with constant curvature $K = -1$.

Of course, one could not expect such simple results for Riemannian manifolds of dimension ≥ 3, but as a first step, one might try to construct analogs of the Gauss-Bonnet formula for Riemannian manifolds of higher dimensions. Such an analog was discovered by Allendoerfer, Weil and Chern and is now called the generalized Gauss-Bonnet Theorem. This formula expresses the Euler characteristic of a compact oriented n-dimensional manifold as an integral of a curvature polynomial. It turns out that there are also several other topological invariants that can be expressed as integrals of curvature polynomials. This leads to an important topic within topology, the theory of characteristic classes, as developed by Chern, Pontrjagin and others [16].

4.4 Application to hyperbolic geometry*

The *hyperbolic plane* (also called the Poincaré upper half plane) is the open set $\mathbb{H}^2 = \{(x, y) \in R^2 : y > 0\}$ together with the "Riemannian metric"

$$g = ds^2 = \frac{1}{y^2}[dx \otimes dx + dy \otimes dy] = \left(\frac{dx}{y}\right)^2 + \left(\frac{dy}{y}\right)^2.$$

The geometry of the "Riemannian manifold" $(\mathbb{H}^2, g)$ has many striking similarlities to the geometry of the ordinary Euclidean plane. In fact, the geometry of this Riemannian manifold is exactly the non-Euclidean geometry, which had been studied by Bolyai and Lobachevsky towards the beginning of the nineteenth century. It would be nice indeed if this non-Euclidean geometry could be realized as the geometry on some surface in $\mathbb{R}^3$, but this is not the case because of a famous theorem of David Hilbert (1901): The hyperbolic plane $\mathbb{H}^2$ cannot be realized on a surface in $\mathbb{R}^3$. In fact, a part of the hyperbolic plane can be realized as the geometry of the pseudo-sphere, but according to Hilbert's Theorem, the entire hyperbolic plane cannot be realized as the geometry of a smooth surface in $\mathbb{R}^3$. Thus abstract Riemannian geometry is absolutely essential for putting non-Euclidean geometry into its proper context as an important special case of the differential geometry of surfaces.

To study the Riemannian geometry of the hyperbolic plane in more detail, we can utilize the Darboux-Cartan method of moving frames to calculate the curvature of this metric. We did this before and found that the Gaussian curvature is given by the formula $K \equiv -1$. In particular the Gaussian curvature of the hyperbolic plane is the same at every point, just like in the case of the sphere. (Of course, there is nothing special about the constant -1; any other negative constant could be achieved by rescaling the metric.)

Another quite useful fact is that angles measured via the hyperbolic metric are exactly the same as those measured via the standard Euclidean metric $dx^2 + dy^2$. Indeed, the form of the metric shows that the coordinates (x, y) are isothermal.

Amazingly, the geodesics in the hyperbolic plane are very simple. Indeed, the straight line $x = c$ is the fixed point set of the reflection

$$\phi : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2c - x \\ y \end{pmatrix},$$

which is easily seen to be an isometry of the metric:

$$\frac{1}{y^2}[d(2c - x)^2 + dy^2] = \frac{1}{y^2}[dx^2 + dy^2].$$

Thus if γ is the geodesic with initial conditions $\gamma(0) = (c, 1)$, $\gamma'(0) = (0, 1)$ then $\phi \circ \gamma$ is also a geodesic with the same initial conditions. By uniqueness of geodesics satisfying given initial conditions, $\phi \circ \gamma = \gamma$ and γ must lie in the vertical line $x = c$. It therefore follows that each vertical line $x = c$ is a geodesic. We can therefore ask if we can find a function $\alpha : \mathbb{R} \to [0, \infty)$ such that the curve

$$\gamma(t) = (0, \alpha(t))$$

is a unit-speed geodesic.

To solve this problem, note that

$$\frac{\alpha'(t)^2}{\alpha(t)^2} = 1 \quad \Rightarrow \quad \frac{\alpha'(t)}{\alpha(t)} = \pm 1 \quad \Rightarrow \quad \alpha(t) = ce^t \quad \text{or} \quad \alpha(t) = ce^{-t}.$$

We conclude that the x-axis is infinitely far away in terms of the Poincaré metric.

Other geodesics can be found by rewriting the metric in polar coordinates

$$ds^2 = \frac{1}{r^2 \sin^2 \theta}[dr^2 + r^2 d\theta^2] = \frac{1}{\sin^2 \theta}\left[\left(\frac{dr}{r}\right)^2 + d\theta^2\right].$$

The map ϕ which sends $r \mapsto 1/r$ and leaves θ alone is also an isometry:

$$\frac{r^2}{\sin^2 \theta}[(d(1/r))^2 + (1/r)^2 d\theta^2] = \frac{1}{r^2 \sin^2 \theta}[dr^2 + r^2 d\theta^2].$$

From this representation, it is easily seen that the map which sends $r \mapsto R^2/r$ and leaves θ alone is an isometry which fixes the semicircle

$$x^2 + y^2 = R^2, \qquad y > 0,$$

so this semicircle is also a geodesic. Since translation to the right or to the left are isometries of the hyperbolic metric, we see that all circles centered on the x-axis intersect the hyperbolic plane in geodesics. (Of course, we have not found their constant speed parametrizations.)

Thus semicircles perpendicular to the x-axis and vertical rays are geodesics. Since there is one of these passing through any point and in any direction, we have described all the geodesics on the hyperbolic plane.

There is another property which the hyperbolic plane shares with Euclidean space and the sphere but with no other surfaces. That is, the hyperbolic plane has a large group of isometries, namely enough isometries to rotate through an arbitrary angle about any point and translate any point to any other point.

To study isometries in general, it is useful to utilize complex notation $z = x + iy$. (This is beneficial because the coordinates are isothermal.) Then $\mathbb{H}^2$ is simply the set of complex numbers with positive imaginary part.

Theorem. *The map*

$$z \mapsto \phi(z) = \frac{az+b}{cz+d} \tag{4.14}$$

is an isometry whenever a, b, c and d are real numbers such that $ad - bc > 0$.

Proof: Note that the transformation (4.14) is unchanged if we make the replacements

$$a \mapsto \lambda a, \quad b \mapsto \lambda b, \quad c \mapsto \lambda c, \quad d \mapsto \lambda d,$$

where $\lambda > 0$. So we can assume without loss of generality that $ad - bc = 1$. We can then factor the map

$$\phi(z) = z' = \frac{az+b}{cz+d}$$

into a composition of four transformations

$$z_1 = z + \frac{d}{c}, \quad z_2 = c^2 z_1, \quad z_3 = -\frac{1}{z_2}, \quad z' = z_3 + \frac{a}{c}. \tag{4.15}$$

To see this, note that

$$z_2 = c(cz+d), \quad z_3 = -\frac{1}{c(cz+d)},$$
$$z' = \frac{a(cz+d)}{c(cz+d)} - \frac{1}{c(cz+d)} = \frac{acz+ad-1}{c(cz+d)} = \frac{az+b}{cz+d}.$$

The first and fourth transformations from (4.15) are translations of the hyperbolic plane which are easily seen to be isometries. It is straightforward to check (using polar coordinates) that the other two are also isometries. Since the composition of isometries is an isometry, ϕ itself is an isometry.

We will call the transformations

$$\phi(z) = \frac{az+b}{cz+d}, \qquad ad - bc = 1,$$

the *linear fractional transformations* and denote the space of linear fractional transformations by $PSL(2,\mathbb{R})$. The reflection $R : \mathbb{H}^2 \to \mathbb{H}^2$ in the line $x = 0$ is also an isometry, but it cannot be written as a special linear fractional transformation. It can be proven that any isometry of $\mathbb{H}^2$ can be written as a special linear fractional transformation, or as the composition $R \circ \phi$, where ϕ is a linear fractional transformation.

Suppose that ϕ is a linear fractional transformation of $\mathbb{H}^2$. If we set $c = 0$, we see that

$$\phi(z) = \frac{az+b}{d} = \frac{a}{d}z + \frac{b}{d} = a^2 z + ab,$$

since $d = 1/a$. This is a radial expansion or contraction about the origin, followed by a translation. Thus we can move any point in $\mathbb{H}^2$ to any other point by means of an isometry. This gives a rigorous proof that the hyperbolic plane is homogeneous; that is, it has the same geometric properties at every point!

Moreover, we can do an arbitrary rotation about a point. Suppose, for example, that we want to fix the point $i = (0, 1)$. We impose this condition on the linear fractional transformation

$$\phi(z) = w = \frac{az + b}{cz + d},$$

and do a short calculation

$$i = \frac{ai + b}{ci + d} \qquad \Rightarrow \qquad -c + id = ai + b \qquad \Rightarrow \qquad a = d,\ b = -c.$$

Since

$$ad - bc = 1 \Rightarrow a^2 + b^2 = 1,$$

we can reexpress the linear fractional transformation in terms of sines and cosines:

$$\phi(z) = w = \frac{\cos\theta z + \sin\theta}{-\sin\theta z + \cos\theta}.$$

This isometry fixes the point i but there are clearly not the identity for arbitrary values of θ. An orientation-preserving isometry which fixes a point must act as a rotation on the tangent space at that point. This can be proven in general, but there is also a direct way to see it in our very specific context.

Simply note that

$$dw = \frac{adz(cz + d) - cdz(az + b)}{(cz + d)^2} = \cdots = \frac{dz}{(cz + d)^2}.$$

If we evaluate at our chosen point, we see that if everything is evaluated at the point i,

$$dw = \frac{dz}{(ci + d)^2} = \frac{dz}{(-\sin\theta i + \cos\theta)^2} = \frac{dz}{(e^{-i\theta})^2} = e^{2i\theta} dz.$$

Writing this out in terms of real and imaginary parts by setting

$$dw = du + idv, \qquad dz = dx + idy,$$

we see that the linear fractional transformation rotates the tangent space through an angle of 2θ.

Since $\mathbb{H}^2$ has the same geometric properties at every point, we see that we can realize rotations about any point by means of isometries.

Thus we have a group of hyperbolic motions which just as rich as the group of Euclidean motions in the plane. We can therefore try to develop the geometry of hyperbolic space in exactly the same way as Euclidean geometry of the plane.

In Euclidean geometry, any line can be extended indefinitely. In hyperbolic geometry the geodesic from the point $i = (0, 1)$ straight down to the x-axis has infinite length as we have seen above. Since any geodesic ray can be taken to any other geodesic ray by means of an isometry, we see that *any* geodesic ray from a point in the hyperbolic plane to the x-axis must have infinite length. In other words, geodesics can be extended indefinitely in the hyperbolic plane.

In Euclidean geometry there is a unique straight line between any two points. Here is the hyperbolic analogue:

Proposition. *In hyperbolic geometry, there is a unique geodesic connecting any two points.*

Proof: Existence is easy. Just take a circle perpendicular to the x-axis (or vertical line) which connects the two points.

If two geodesics intersect in more than one point, they would form a geodesic biangle, which is shown to be impossible by the Gauss-Bonnet Theorem:

Exercise IX. (Do not hand in.) a. Construct a geodesic biangle in S^2 with its standard metric using two geodesics from the north to the south poles, and use the Gauss-Bonnet Theorem to show that area of the geodesic biangle is 2α, where α is the angle between the geodesics.

b. Use the Gauss-Bonnet theorem to show that there is no geodesic biangle in the hyperbolic plane.

Proposition. *In hyperbolic geometry, any isosceles triangle, with angles α, β and β once again, can be constructed, so long as $\alpha + 2\beta < \pi$.*

Proof: Starting from the point i, construct two downward pointing geodesics which approach the x-axis and are on opposite sides of the y-axis. We can arrange that each of these geodesics makes an angle of $\alpha/2$ with the y-axis. Move the same distance d along each geodesic until we reach the points p and q. Connect p and q by a geodesic, thereby forming a geodesic triangle.

The interior angles at p and q must be equal because the triangle is invariant under the reflection R in the y-axis. When $d \to 0$, the Gauss-Bonnet formula shows that the sum of the interior angles of the geodesic approaches π. On the other hand, as the vertices of the triangle approach the x axis, the interior angles β at p and q approach zero.

By the intermediate value theorem from analysis, β can assume any value such that

$$0 < \beta < \frac{1}{2}(\pi - 2\alpha),$$

and the Proposition is proven.

Once we have this Proposition, we can piece together eight congruent isosceles geodesic triangles with angles

$$\alpha = \frac{\pi}{4}, \qquad \beta = \frac{\pi}{8}$$

to form a geodesic octagon. All the sides have the same length so we can identify them in pairs. This prescription identifies all of the vertices on the geodesic octagon. Since the angles at the vertices add up to 2π, a neighborhood of the identified point can be made diffeomorphic to an open subset of $\mathbb{R}^2$.

We need to do some cut and paste geometry to see what kind of surface results. The answer is a sphere with two handles, the compact oriented surface of genus two. Actually, it is easiest to see this by working in reverse, and cutting the sphere with two handles along four circles which emanate from a given point. From this we can see that the sphere with two handles can be cut into an octagon.

This then leads to the following remarkable fact:

Theorem. *A sphere with two handles can be given an abstract Riemannian metric with Gaussian curvature $K \equiv -1$.*

Indeed, a similar construction enables give a metric of constant curvature one on a sphere with g handles, where g is any integer such that $g \geq 2$.

Exercise X. (Do not hand in.) Show that any geodesic triangle in the Poincaré upper half plane must have area $\leq \pi$.

Exercise XI. (Do not hand in.) We now return to consider the Riemannian manifold $(\mathbb{H}^n, \langle \cdot, \cdot \rangle)$ described at the end of §4.1. We claim that it is also *homogeneous*, that is, given any two point $p, q \in \mathbb{H}^n$, there is an isometry

$$\phi : (\mathbb{H}^n, \langle \cdot, \cdot \rangle) \longrightarrow (\mathbb{H}^n, \langle \cdot, \cdot \rangle)$$

such that $\phi(p) = q$. Indeed, we have shown this above in the case where $n = 2$.

a. Show that $(\mathbb{H}^n, \langle \cdot, \cdot \rangle)$ is also homogeneous.

b. Use this fact to show that $(\mathbb{H}^n, \langle \cdot, \cdot \rangle)$ is complete. (Hint: If the exponential map $\exp_p$ at p is defined on a ball of radius $\epsilon > 0$ then so is $\phi \circ \exp_p$, which can be taken to be the exponential map at q. Thus no matter how far any geodesic has been extended, it can be extended for a distance at least $\epsilon > 0$ for a fixed choice of ϵ, and this implies that geodesics can be extended indefinitely.

Remark. Thus it follows from the uniqueness of simply connected complete space forms, that $(\mathbb{H}^n, \langle \cdot, \cdot \rangle)$ is isometric to the model of hyperbolic space we constructed in §1.8.

Chapter 5

Appendix: General relativity*

We give a quick introduction to the key ideas of general relativity, to illustrate how the notions of curvature are used. All of this is treated in much more detail in [17] and [22]. We will see that the Bianchi identity was crucial in enabling Einstein to find the right form of the field equations for general relativity.

The key ideas of classical Newtonian mechanics are easily described. One imagines mass density $\rho(x^1, x^2, x^3)$ spread throughout Euclidean space $\mathbb{E}^3$ with Euclidean coordinates (x^1, x^2, x^3). One then solves for the Newtonian potential $\phi(x^1, x^2, x^3) = m\Phi(x^1, x^2, x^3)$ in the Poisson equation (1.10), which we can rewrite in terms of potential per unit mass Φ as

$$\frac{\partial^2 \Phi}{\partial (x^1)^2} + \frac{\partial^2 \Phi}{\partial (x^2)^2} + \frac{\partial^2 \Phi}{\partial (x^3)^2} = 4\pi G \rho(x^1, x^2, x^3), \tag{5.1}$$

where G is the Newtonian gravitational constant. Then one uses Hamilton's principle as described in §1.4) to derive the orbits of a planet which is subject to the a gravitational force which is minus the gradient of the Newtonian potential, which is just

$$m\frac{d^2(x^i)}{dt^2} = -m\frac{\partial \Phi}{\partial x^i}.$$

with m being the mass of the planet.

Einstein's idea was that gravitational forces should be manifested by a Lorentz metric on four-dimensional space time. Thus one might start with with the flat Lorentz metric (1.28) of special relativity,

$$\langle \cdot, \cdot \rangle = -c^2 dt \otimes dt + dx^1 \otimes dx^1 + dx^2 \otimes dx^2 + dx^3 \otimes dx^3,$$

a flat Lorentz metric on $\mathbb{R}^4$, and modify it slightly. The simplest such modification would be

$$\langle \cdot, \cdot \rangle = -(c^2 + 2\Phi) dt \otimes dt + dx^1 \otimes dx^1 + dx^2 \otimes dx^2 + dx^3 \otimes dx^3, \tag{5.2}$$

where Φ is the solution to the Poisson equation (5.1). If we required that the path

$$\tilde{\gamma} : (a,b) \longrightarrow \mathbb{R}^4, \quad \gamma(t) = \begin{pmatrix} t \\ \gamma(t) \end{pmatrix} = \begin{pmatrix} t \\ x^1(t) \\ x^2(t) \\ x^3(t) \end{pmatrix}$$

be a critical point of the action integral

$$J_0(\gamma) = \int_a^b \left[\left(-\frac{c^2}{2} - \Phi \right) + \frac{1}{2} \left(\left(\frac{dx^1}{dt} \right)^2 + \left(\frac{dx^2}{dt} \right)^2 + \left(\frac{dx^3}{dt} \right)^2 \right) \right] dt,$$

then in accordance with Hamilton's principle, we would recover the equations of motion for Newtonian mechanics, since the integrand differs by a constant from the Lagrangian we obtained before. But this action $J_0(\gamma)$ closely approximates the action integral

$$J(\tilde{\gamma}) = \int_a^b \left[\left(-\frac{c^2}{2} - \Phi \right) \left(\frac{dt}{d\tau} \right)^2 + \frac{1}{2} \left(\left(\frac{dx^1}{d\tau} \right)^2 + \left(\frac{dx^2}{d\tau} \right)^2 + \left(\frac{dx^3}{d\tau} \right)^2 \right) \right] d\tau$$

given by the Lorentz metric (5.2), for paths

$$\tilde{\gamma} : [a,b] \to \mathbb{R}^4, \qquad \tilde{\gamma}(\tau) = \begin{pmatrix} t(\tau) \\ x^1(\tau) \\ x^2(\tau) \\ x^3(\tau) \end{pmatrix},$$

and one might expect that J would give corrections to Newtonian mechanics that bring it into closer accord with special relativity. Indeed, in 1911, Einstein used essentially this metric to argue that light would bend when passing close to a massive object, but the deflection he predicted turned out to be off by a factor of two.

The problem is that our construction of the Lorentz metric was too ad hoc, we have not constructed a tensor equation which constructs the metric in a way consistent with local equivalence principles.

Nevertheless, we might imagine proceeding with the provisional metric (5.2) and see what form the Poisson equation takes. To this end, we calculate the Christoffel symbols for the Lorentz metric (5.2) with index conventions $1 \leq i,j,k \leq 3$. The result is that for the metric which should approximate classical Newtonian mechanics, all the Christoffel symbols vanish except for

$$\Gamma^i_{00} = \frac{\partial \Phi}{\partial x^i}, \quad \Gamma^0_{i0} = \Gamma^0_{0i} = -\frac{1}{c^2 + 2\Phi} \frac{\partial \Phi}{\partial x^i},$$

leading in turn to the geodesic equations

$$\frac{d^2 x^i}{d\tau^2} = -\frac{\partial \Phi}{\partial x^i} \left(\frac{dt}{d\tau} \right)^2,$$

which gives a close approximation to the orbits in Newtonian theory. We can use (1.31) to calculate the Ricci curvature, and find that

$$R_{00} = \sum_{j=0}^{3} R^j_{0j0} = \sum_{j=1}^{3} \frac{\partial}{\partial x^j}(\Gamma^j_{00}) - \sum_{m=1}^{3} \Gamma^0_{m0}\Gamma^m_{00} = \sum_{j=1}^{3} \frac{\partial^2 \Phi}{\partial (x^1)^2} - \sum_{m=1}^{3} \Gamma^0_{m0}\Gamma^m_{00}.$$

Thus one might be led to conjecture that the relativistic version of Poisson's equation should be

$$R_{00} = 4\pi G\rho,$$

when ρ is the density for a dust which is not moving with respect to the coordinate system (t, x^1, x^2, x^3).

But of course, we need a field equation for the metric which does not depend upon choice of local coordinates. What the above argument suggests is that the appropriate equation should involve the Ricci tensor. Moreover, it should involve the stress energy tensor T described in §2.4 and given by (2.11), satisfying the Euler equations for a perfect fluid:

$$T^{\mu\nu}{}_{;\nu} = 0.$$

(Throughout this chapter, we will use the index conventions $0 \leq \mu, \nu, \sigma, \tau \leq 3$ and $1 \leq i, j, k \leq 3$.)

We have provided some of the motivation leading to Einstein's choice of field equations. Indeed, here is how Einstein expressed himself in [5], pages 83-84:

> If there is an analogue of Poisson's equation in the general theory of relativity, then this equation must be a tensor equation for the tensor $g_{\mu\nu}$ of the gravitational potential; the energy tensor of matter must appear on the right-hand side of this equation. On the left-hand side of the equation there must be a differential tensor in the $g_{\mu\nu}$. We have to find this differential tensor. It is completely determined by the following three conditions:
>
> 1. It may contain no differential coefficients of the $g_{\mu\nu}$ higher than the second.
> 2. It must be linear in these second differential coefficients.
> 3. Its divergence must vanish identically.

From Einstein's criteria, we see that the Ricci tensor $R_{\mu\nu}$ would be a candidate for the left-hand side of the field equation, except for the fact that it does not have zero divergence. However, it follows from Exercise VI (see (2.13) that

$$G^{\mu\nu}{}_{;\nu} = 0, \quad \text{where} \quad G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} s \, g_{\mu\nu},$$

where s is the scalar curvature of the Lorentz metric. Thus the field equations which Einstein adopted were

$$R_{\mu\nu} - \frac{1}{2} s \, g_{\mu\nu} = \kappa T_{\mu\nu}, \tag{5.3}$$

where κ is a constant and $T_{\mu\nu}$ is the stress-energy tensor of matter within space-time. We can of course solve (5.3) for $R_{\mu\nu}$, the result being

$$R_{\mu\nu} = \kappa \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right), \quad \text{where} \quad T = g^{\mu\nu} T_{\mu\nu}. \tag{5.4}$$

Later we will see that the constant should be $\kappa = 8\pi G$.

We might now try to find approximate solutions to the field equations by linearizing them near the known solution for the flat space-time which contains no matter. For simplicity, we choose units so that the speed of light c is one, so that the metric coefficients for this flat space-time are

$$(\eta_{\mu\nu}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then the coefficients of the unknown metric should be of the form

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad \text{where} \quad |h_{\mu\nu}| << 1.$$

and in carrying out calculations with the linearization, we ignore all products of any derivatives of $h_{\mu\nu}$ with other derivatives of $h_{\mu\nu}$. This simplifies the calculation of the Ricci curvature immensely because we can ignore the products in the Christoffel symbols when calculating curvature. Indeed, it follows directly from (1.31) and the expressions for the Christoffel symbols that

$$\begin{aligned} R_{\mu\nu} = \sum R^{\sigma}_{\mu\sigma\nu} &= \sum \frac{\partial}{\partial x^{\sigma}} (\Gamma^{\sigma}_{\mu\nu}) - \sum \frac{\partial}{\partial x^{\nu}} (\Gamma^{\sigma}_{\mu\sigma}) \\ &= \frac{1}{2} \sum \eta^{\sigma\tau} (h_{\mu\sigma,\nu\tau} + h_{\nu\sigma,\mu\tau} - h_{\sigma\tau,\mu\nu}) - \frac{1}{2} \sum \eta^{\sigma\tau} h_{\mu\nu,\sigma\tau}, \end{aligned} \tag{5.5}$$

where the commas denote differentiation.

The equations (5.5) appear complicated, but they can be simplified immensely by a "gauge transformation," that is by a change of coordinates of the form

$$x^{\alpha} \mapsto \tilde{x}^{\alpha} = x^{\alpha} + \xi^{\alpha}, \quad \text{where} \quad |\xi^{\alpha}| << 1.$$

Indeed, let $\tilde{g}_{\mu\nu}$ denote the components of the metric with respect to the new coordinates $\tilde{x}^{\alpha}$ and write

$$\tilde{g}_{\mu\nu} = \eta_{\mu\nu} + \tilde{h}_{\mu\nu}, \quad \text{where} \quad |\tilde{h}_{\mu\nu}| << 1.$$

Then

$$\begin{aligned} \eta_{\mu\nu} + h_{\mu\nu} = g_{\mu\nu} &= \sum \frac{\partial \tilde{x}^{\sigma}}{\partial x^{\mu}} \frac{\partial \tilde{x}^{\tau}}{\partial x^{\nu}} \tilde{g}_{\sigma\tau} = \sum (\delta^{\sigma}_{\mu} + \xi^{\sigma}{}_{,\mu})(\delta^{\tau}_{\nu} + \xi^{\tau}{}_{,\nu})(\eta_{\sigma\tau} + \tilde{h}_{\sigma\tau}) \\ &= \eta_{\mu\nu} + \xi_{\mu,\nu} + \xi_{\nu,\mu} + \tilde{h}_{\mu\nu} + (\text{negligible terms}), \end{aligned}$$

so in the linear approximation we can write

$$h_{\mu\nu} = \tilde{h}_{\mu\nu} + \xi_{\mu,\nu} + \xi_{\nu,\mu}.$$

Lemma. *We can choose* ξ_α *so that*

$$\sum \eta^{\sigma\tau} \left(\tilde{h}_{\mu\sigma,\tau} - \frac{1}{2} \tilde{h}_{\sigma\tau,\mu} \right) = 0. \tag{5.6}$$

To prove this, note that

$$\sum \eta^{\sigma\tau} \left(h_{\mu\sigma,\tau} - \frac{1}{2} h_{\sigma\tau,\mu} \right) = \sum \eta^{\sigma\tau} \left(\tilde{h}_{\mu\sigma,\tau} - \frac{1}{2} \tilde{h}_{\sigma\tau,\mu} \right) + \sum \eta^{\sigma\tau} \xi_{mu,\sigma\tau}.$$

Thus we need only solve the equation

$$\Box\, \xi_\mu = \sum \eta^{\sigma\tau} \xi_{mu,\sigma\tau} = \left(h_{\mu\sigma,\tau} - \frac{1}{2} h_{\sigma\tau,\mu} \right),$$

where $\Box$ denotes the usual wave operator. We can solve this equation thereby achieving (5.6).

Thus after a gauge transformation, we can assume that

$$\sum \eta^{\sigma\tau} \left(h_{\mu\sigma,\tau} - \frac{1}{2} h_{\sigma\tau,\mu} \right) = 0.$$

Substitution into (5.5) eliminates the first three terms on the right-hand side, leaving the equation

$$R_{\mu\nu} = -\frac{1}{2} \Box\, h_{\mu\nu} = \frac{1}{2} \left[\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial (x^1)^2} - \frac{\partial^2}{\partial (x^2)^2} - \frac{\partial^2}{\partial (x^3)^2} \right] h_{\mu\nu}. \tag{5.7}$$

Once again $\Box$ denotes the wave operator. In the case where the stress-energy is zero, the Einstein field equations become

$$\Box\, h_{\mu\nu} = 0,$$

which is just the equation for gravitational waves propagating through Minkowski space-time.

In the time independent case, (5.7) reduces to

$$R_{\mu\nu} = -\frac{1}{2} \Delta h_{\mu\nu} = -\frac{1}{2} \left[\frac{\partial^2}{\partial (x^1)^2} + \frac{\partial^2}{\partial (x^2)^2} + \frac{\partial^2}{\partial (x^3)^2} \right] h_{\mu\nu}. \tag{5.8}$$

If we imagine that matter with density ρ is at rest with respect to the coordinates (t, x^1, x^2, x^3), then with respect to these coordinates

$$(T^{\mu\nu}) = (T_{\mu\nu}) = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \text{so} \quad T = \sum \eta^{\mu\nu} T_{\mu\nu} = -\rho$$

and

$$\left(T_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}T\right) = \begin{pmatrix} \rho/2 & 0 & 0 & 0 \\ 0 & \rho/2 & 0 & 0 \\ 0 & 0 & \rho/2 & 0 \\ 0 & 0 & 0 & \rho/2 \end{pmatrix}. \tag{5.9}$$

Thus in the time-independent case it follows from (5.4) and (5.7) that

$$\Delta h_{\mu\nu} = -\kappa\rho.$$

For this to agree with Poisson's equation, we have to take $\kappa = 8\pi G$, where G is Newton's constant of gravitation. We can then conclude that the solution to the linearized equations for a field of dust particles which are at rest with respect to the coordinates (t, x^1, x^2, x^3) will be

$$\langle\cdot,\cdot\rangle = -(1+2\Phi)dt \otimes dt + (1-2\Phi)\left(\sum_{i=1}^{3} dx^i \otimes dx^i\right),$$

where Φ is the Newtonian potential. If we had not used coordinates in which the speed of light is one, the result would have been

$$\langle\cdot,\cdot\rangle = -(c^2+2\Phi)dt \otimes dt + \left(1 - \frac{2\Phi}{c^2}\right)\left(\sum_{i=1}^{3} dx^i \otimes dx^i\right).$$

In the case of a gravitational field produced by a spherically summetric star, $\Phi = -(GM)/r$ outside the star, where M is the mass of the star and r is the radial coordinate. In this case,

$$\langle\cdot,\cdot\rangle = -\left(c^2 - \frac{2GM}{r}\right) dt \otimes dt + \left(1 + \frac{2GM}{c^2 r}\right)\left(\sum_{i=1}^{3} dx^i \otimes dx^i\right).$$

In contrast to (5.2) this metric does give the right result for deflection of light passing close to the star.

Einstein first published his field equations in 1915 and the bending of light by the sun was observed by Sir Arthur Eddington in the solar eclipse of May 29, 1919. This observation made Einstein famous. At present, it is commonplace to observe the effects of the bending of light by a "gravitational lens" in pictures of distant galaxies taken by space telescopes such as the Hubble and Herschel telescopes. This provides compelling experimental evidence that Einstein's field equations are indeed correct.

Bibliography

[1] C. Böhm and B. Wilking, *Manifolds with positive curvature operators are space forms*, Annals of Math. **187** (2008), 1079-1097.

[2] R. Bott and L. Tu, *Differential forms in algebraic topology*, Springer Verlag, New York, 1982.

[3] William M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, Academic Press, New York, 2003.

[4] S. S. Chern, *A simple intrinsic proof of the Gauss-Bonnet formula for closed Riemannian manifolds*, Annals of Math. **45** (1944), 747-752.

[5] A. Einstein, *The meaning of relativity*, Princeton Univ. Press, Princeton, NJ, 1922, 1955.

[6] P. Griffiths and J. Harris, *principles of algebraic geometry*, Wiley, New York, 1978.

[7] D. Gromoll and W. Meyer, *Periodic geodesics on compact Riemannian manifolds*, J. Differential Geometry **3** (1969), 493-510.

[8] A. Hatcher, *Algebraic topology*, Cambridge Univ. Press, Cambridge, UK, 2002.

[9] S. Helgason, *Differential geometry, Lie groups and symmetric spaces*, Academic Press, New York, 1978.

[10] M. Hirsch, *Differential topology*, Springer, New York, 1997.

[11] H. B. Lawson and M. L. Michelsohn, *Spin geometry*, Princeton Univ. Press, Princeton, NJ, 1989.

[12] J. M. Lee, *Introduction to smooth manifolds*, Springer, New York, 2002.

[13] J. M. Lee, *Riemannian manifolds: an introduction to curvature* , Springer, New York, 1997.

[14] J. Lohkamp, *Metrics of negative Ricci curvature*, Annals of Math. **140** (1994), 655-683.

[15] J. Milnor, *Morse theory*, Annals of Math. Studies **51**, Princeton Univ. Press, Princeton, NJ, 1963.

[16] J. Milnor and J. Stasheff, *Characteristic classes*, Princeton Univ. Press, Princeton NJ, 1974.

[17] C.W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation*, W. H. Freeman, New York, 1973.

[18] J. Nash, *The imbedding problem for Riemannian manifolds*. Annals of Math. **63** (1956), 20-63.

[19] A. Pressley, *Elementary differential geometry*, Springer, New York, 2001.

[20] M. Taylor, *Partial differential equations; basic theory*, Springer, New York, 1996.

[21] W. Thurston, *Three-dimensional geometry and topology*, Princeton Univ. Press, Princeton, NJ, 1997.

[22] R. M. Wald, *General relativity*, University of Chicago Press, Chicago, 1984.

[23] J. A. Wolf, *Spaces of constant curvature*, McGraw-Hill, New York, 1967.