Analysis

Longer

May 1, 2022

Contents

Chapter 1

Basics of Riemannian Geometry

1.1 Model Riemannian manifolds

1.1.1 Symmetries of Riemannian manifolds

The main feature of the Riemannian manifolds we are going to introduce is that they are all highly symmetric, meaning that they have large groups of isometries.

Let (M, g) be a Riemannian manifold. Recall that $\mathrm{Iso}(M, g)$ denotes the set of all isometries from M to itself, which is a group under composition. We say that (M, g) is a **homogeneous Riemannian manifold** if $\mathrm{Iso}(M, g)$ acts transitively on M, which is to say that for each pair of points $p, q \in M$, there is an isometry $\varphi : M \to M$ such that $\varphi(p) = q$.

The isometry group does more than just act onM itself. For every $\varphi \in \mathrm{Iso}(M, g)$, the global differential $d\varphi$ maps TM to itself and restricts to a linear isometry $d\varphi_p : T_pM \to T_pM$ for each $p \in M$. Given a point $p \in M$, let $\mathrm{Iso}_p(M, g)$ denote the **isotropy subgroup at p**, that is, the subgroup of $\mathrm{Iso}(M, g)$ consisting of isometries that fix p. For each $\varphi \in \mathrm{Iso}_p(M, g)$, the linear map $d\varphi_p$ takes T_pM to itself, and the map $I_p : \mathrm{Iso}_p(M, g) \to \mathrm{GL}(T_pM)$ given by $I_p(\varphi) = d\varphi_p$ is a representation of $\mathrm{Iso}_p(M, g)$, called the isotropy representation. We say that M is **isotropic at p** if the isotropy representation of $\mathrm{Iso}_p(M, g)$ acts transitively on the set of unit vectors in T_pM. If M is isotropic at every point, we say simply that M is **isotropic**.

There is an even stronger kind of symmetry than isotropy. Let $\mathrm{O}(M)$ denote the set of all orthonormal bases for all tangent spaces of M:

$$\mathrm{O}(M) = \coprod_{p \in M} \{\text{orthonormal bases for } T_pM\}.$$

There is an induced action of $\mathrm{Iso}(M, g)$ on $\mathrm{O}(M)$, defined by using the differential of an isometry φ to push an orthonormal basis at p forward to an orthonormal basis at $\varphi(p)$:

$$\varphi \cdot (e_1, \ldots, e_n) = (d\varphi_p(e_1), \ldots, d\varphi_p(e_n)).$$

We say that (M, g) is **frame-homogeneous** if this induced action is transitive on $\mathrm{O}(M)$, or in other words, if for all $p, q \in M$ and choices of orthonormal bases at p and q, there is an isometry taking p to q and the chosen basis at p to the one at q.

With these definitions, the following result is immediate.

Proposition 1.1.1. *Let (M, g) be a Riemannian manifold.*

(a) If M is isotropic at one point and it is homogeneous, then it is isotropic.

(b) If M is frame-homogeneous, then it is homogeneous and isotropic.

1.1.2 Euclidean spaces

The simplest and most important model Riemannian manifold is of course n-dimensional Euclidean space, which is just $\mathbb{R}^n$ with the Euclidean metric $\bar{g}$. Somewhat more generally, if V is any n-dimensional real vector space endowed with an inner product, we can set $g(v, w) = \langle v, w \rangle$ for any $p \in V$ and any $v, w \in T_pV$. Choosing an orthonormal basis $(e_1, \ldots, e_n)$ for V defines a basis isomorphism from $\mathbb{R}^n$ to V that sends $(x_1, \ldots, x_n)$ to $x^i e_i$, this is easily seen to be an isometry of (V, g) with $(\mathbb{R}^n, \bar{g})$, so all n-dimensional inner product spaces are isometric to each other as Riemannian manifolds.

Euclidean isometry group

In Exercise **??** we have showed that the isometry group of $(\mathbb{R}^n, \bar{g})$ is isomorphic to the Euclidean group $\mathrm{E}(n)$. It has a faithful representation given by the map $\rho : \mathrm{E}(n) \to \mathrm{GL}(n + 1, \mathbb{R})$ defined in block form by

$$\rho(b, A) \begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix}$$

where b is considered an $n - 1$ column matrix.

The Euclidean group acts on $\mathbb{R}^n$ via

$$(b, A) \cdot x = b + Ax.$$

It is clear that when $\mathbb{R}^n$ is endowed with the Euclidean metric, this action is isometric and the induced action on $\mathrm{O}(\mathbb{R}^n)$ is transitive. Thus each Euclidean space is frame-homogeneous.

1.1.3 Spheres

Our second class of model Riemannian manifolds comes in a family, with one for each positive real number. Given $R > 0$, let $S^n(R)$ denote the sphere of radius R centered at the origin in $\mathbb{R}^{n+1}$, endowed with the metric $\mathring{g}_R$ (called the round metric of radius R) induced from the Euclidean metric on $\mathbb{R}^{n+1}$. When $R = 1$, it is the round metric on S^n and we use the notation $\mathrm{g} = \mathrm{g}_1$.

One of the first things one notices about the spheres is that like Euclidean spaces, they are highly symmetric. We can immediately write down a large group of isometries of $S^n(R)$ by observing that the linear action of the orthogonal group $\mathrm{O}(n + 1)$ on $\mathbb{R}^{n+1}$ preserves $S^n(R)$ and the Euclidean metric, so its restriction to $S^n(R)$ acts isometrically on the sphere.

e homogeneous

Proposition 1.1.2. *The group $\mathrm{O}(n + 1)$ acts transitively on $\mathrm{O}(S^n(R))$, and thus each round sphere is frame-homogeneous.*

Proof. It suffices to show that given any $p \in S^n(R)$ and any orthonormal basis (e_i) for $T_pS^n(R)$, there is an orthogonal map that takes the "north pole" $N = (0, \ldots, 0, R)$ to p and the basis $\partial_1, \ldots, \partial_n$ for $T_N S^n(R)$ to (e_i).

To do so, think of p as a vector of length R in $\mathbb{R}^{n+1}$, and let $\widehat{p} = p/|p|$ denote the unit vector in the same direction. Since the basis vectors (e_i) are tangent to the sphere, they are orthogonal to $\widehat{p}$, so $(e_1, \ldots, e_n, \widehat{p})$ is an orthonormal basis for $\mathbb{R}^{n+1}$. Let A be the matrix whose columns are these basis vectors. Then $A \in \mathrm{O}(n+1)$, and by elementary linear algebra, A takes the standard basis vectors $\partial_1, \ldots, \partial_n$ to $(e_1, \ldots, e_n, \widehat{p})$. It follows that $A(N) = p$. Moreover, since A acts linearly on $\mathbb{R}^{n+1}$, its differential is represented in standard coordinates by the same matrix as A itself, so $dA_N : T_N(\partial_i) = e_i$ for $i = 1, \ldots, n$, and A is the desired orthogonal map. $\square$

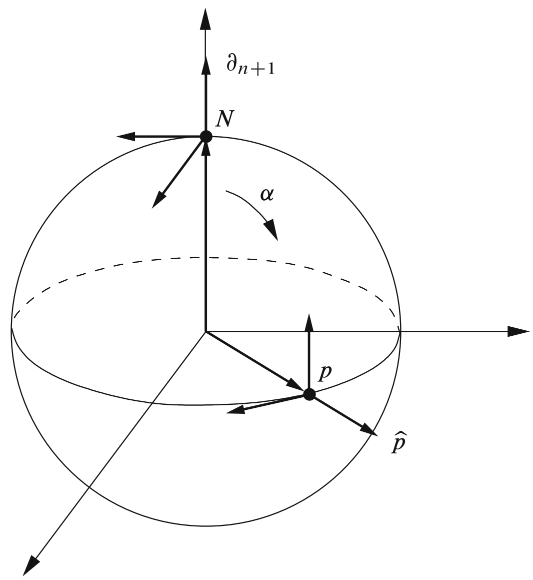

Figure 1.1: Transitivity of $\mathrm{O}(n+1)$ on $\mathrm{O}(S^n(R))$.

Another important feature of the round metrics—one that is much less evident than their symmetry—is that they bear a certain close relationship to the Euclidean metrics, which we now describe. Two metrics g_1 and g_2 on a manifold M are said to be **conformally related** (or **pointwise conformal** or just **conformal**) to each other if there is a positive function $f \in C^\infty(M)$ such that $g_2 = fg_1$. Given two Riemannian manifolds (M, g) and $(\widetilde{M}, \widetilde{g})$, a diffeomorphism $\varphi : M \to \widetilde{M}$ is called a **conformal diffeomorphism** (or a **conformal transformation**) if it pulls $\widetilde{g}$ back to a metric that is conformal to g:

$$\varphi^*\widetilde{g} = fg$$

for some $f \in C^\infty(M)$. The following proposition shows that conformal diffeomorphisms are the same as **angle-preserving diffeomorphisms**.

Proposition 1.1.3. *Let g_1 and g_2 be two Riemannian metrics on M. Then g_1 and g_2 are conformal if and only if they define the same angles.*

Therefore, a diffeomorphism is a conformal equivalence if and only if it preserves angles.

Proof. One direction is clear. Now assume that g_1 and g_2 define the same angle. Let (E_i) be a local orthonormal frame for g_1, and consider the g_2-angle between E_i and $(\cos\theta)E_i + (\sin\theta)E_j$. Since g_1 and g_2 define the same angle, we have $\langle E_i, E_j\rangle_{g_2} = 0$ for $i \neq j$, and thus

$$\begin{aligned}\cos\theta &= \frac{\langle E_i, (\cos\theta)E_i + (\sin\theta)E_j\rangle_{g_2}}{|E_i|_{g_2} \cdot |(\cos\theta)E_i + (\sin\theta)E_j|_{g_2}} \\ &= \cos\theta \cdot \frac{|E_i|_{g_2}}{\sqrt{\cos^2\theta |E_i|^2_{g_2} + \sin^2\theta |E_j|^2_{g_2}}}.\end{aligned}$$

This then implies that $|E_i|_{g_2} = |E_j|_{g_2}$. Since i, j are arbitrary, we conclude that $g_2 = fg_1$ for a function $f : M \to \mathbb{R}$. Since g_2 is smooth, it follows that f is smooth. Therefore g_1 and g_2 are conformal. $\square$

Two Riemannian manifolds are said to be **conformally equivalent** if there is a conformal diffeomorphism between them. A Riemannian manifold (M, g) is said to be **locally conformally flat** if every point of M has a neighborhood that is conformally equivalent to an open set in $(\mathbb{R}^n, \bar{g})$.

A conformal equivalence between $\mathbb{R}^n$ and $S^n(R)$ minus a point is provided by stereographic projection from the north pole. This is given by the following formula

$$\sigma(\xi, \tau) = u = \frac{R\xi}{R - \tau}$$

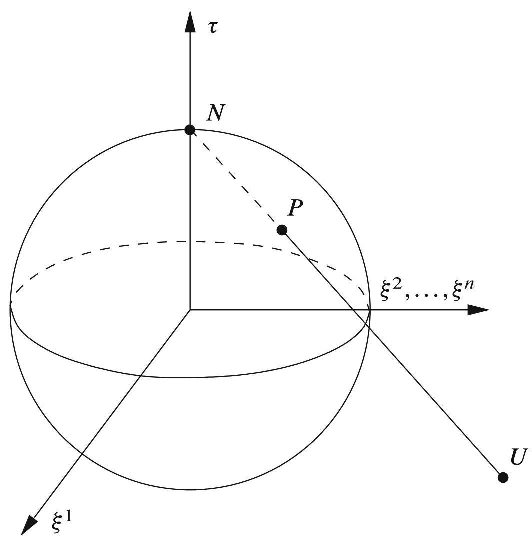

Figure 1.2: Stereographic projection.

It follows from this formula that σ is defined and smooth on all of $S^n(R) - \{N\}$. The easiest way to see that it is a diffeomorphism is to compute its inverse:

$$\sigma^{-1}(u) = (\xi, \tau) = \Big(\frac{2Ru}{|u|^2 + R^2}, R\frac{|u|^2 - R^2}{|u|^2 + R^2}\Big).$$

Proposition 1.1.4. *Stereographic projection is a conformal diffeomorphism between $S^n(R) - \{N\}$ and $\mathbb{R}^n$.*

Proof. The inverse map σ^{-1} is a smooth parametrization of $S^n(R) - \{N\}$, so we can use it to compute the pullback metric. Using the usual technique of substitution to compute pullbacks, we obtain the following coordinate representation of $\mathring{g}_R$ in stereographic coordinates:

$$(\sigma^{-1})^* \mathring{g}_R = (\sigma^{-1})^* \bar{g} = \sum_i \Big(d\Big(\frac{2R^2 u^i}{|u|^2 + R^2}\Big)\Big)^2 + \Big(d\Big(R\frac{|u|^2 - R^2}{|u|^2 + R^2}\Big)\Big)^2.$$

If we expand each of these terms individually, we get

$$d\Big(\frac{2R^2u^i}{|u|^2+R^2}\Big) = \frac{2R^2du^i}{|u|^2+R^2} - \frac{4R^2u^i\sum_j u^j du^j}{(|u|^2+R^2)^2};$$

$$d\Big(R\frac{|u|^2-R^2}{|u|^2+R^2}\Big) = \frac{2R\sum_j u^j du^j}{|u|^2+R^2} - \frac{2R(|u|^2-R^2)\sum_j u^j du^j}{(|u|^2+R^2)^2} = \frac{4R^3\sum_j u^j du^j}{(|u|^2+R^2)^2}.$$

Therefore,

$$\begin{aligned}(\sigma^{-1})^*\mathring{g}_R &= \frac{4R^2\sum_i (du^i)^2}{(|u|^2+R^2)^2} - \frac{16R^4(\sum_j u^j du^j)^2}{(|u|^2+R^2)^3} + \frac{16R^4|u|^2(\sum_j u^j du^j)}{(|u|^2+R^2)^4} + \frac{16R^6(\sum_j u^j du^j)^2}{(|u|^2+R^2)^4} \\ &= \frac{4R^2\sum_i (du^i)^2}{(|u|^2+R^2)^2}.\end{aligned}$$

In other words,

$$(\sigma^{-1})^*\mathring{g}_R = \frac{4R^2}{(|u|^2+R^2)^2}\bar{g},$$

where $\bar{g}$ now represents the Euclidean metric on $\mathbb{R}^n$, and so σ is a conformal diffeomorphism. □

Corollary 1.1.5. *Each sphere with a round metric is locally conformally flat.*

Proof. Stereographic projection gives a conformal equivalence between a neighborhood of any point except the north pole and Euclidean space; applying a suitable rotation and then stereographic projection (or stereographic projection from the south pole), we get such an equivalence for a neighborhood of the north pole as well. □

1.1.4 Hyperbolic spaces

Our third class of model Riemannian manifolds is perhaps less familiar than the other two. For each $n \geq 1$ and each $R > 0$ we will define a frame-homogeneous Riemannian manifold $\mathbb{H}^n(R)$, called **hyperbolic space of radius *R***. There are four equivalent models of the hyperbolic spaces, each of which is useful in certain contexts. In the next theorem, we introduce all of them and show that they are isometric.

erbolic space **Theorem 1.1.6.** *Let n be an integer greater than 1. For each fixed $R > 0$, the following Riemannian manifolds are all mutually isometric.*

(a) *(**Hyperboloid model**) $\mathbb{H}^n(R)$ is the submanifold of Minkowski space $\mathbb{R}^{n,1}$ defined in standard coordinates (ξ, τ) as the "upper sheet" $\{\tau > 0\}$ of the two-sheeted hyperboloid $|\xi|^2 - \tau^2 = -R^2$, with the induced metric $\breve{g}^1_R = \iota^*\bar{q}$ where $\iota : \mathbb{H}^n(R) \to \mathbb{R}^{n,1}$ is inclusion, and $\bar{q}$ is the Minkowski metric:*

$$\bar{q} = (d\xi^1)^2 + \cdots + (d\xi^n)^2 - (d\tau)^2. \tag{1.1}$$

Hyperbolic sp

(b) *(**Beltrami-Klein model**) $\mathbb{K}^n(R)$ is the ball of radius R centered at the origin in $\mathbb{R}^n$, with the metric given in coordinates $(w^1, \ldots, w^n)$ by*

$$\breve{g}^2_R = R^2\frac{(dw^1)^2 + \cdots + (dw^n)^2}{R^2 - |w|^2} + R^2\frac{(w^1dw^1 + \cdots + w^ndw^n)^2}{(R^2-|w|^2)^2}. \tag{1.2}$$

Hyperbolic sp

(c) ***(Poincaré ball model))*** *$\mathbb{B}^n(R)$ is the ball of radius R centered at the origin in $\mathbb{R}^n$, with the metric given in coordinates $(u^1, \dots, u^n)$ by*

$$\breve{g}^3_R = 4R^4 \frac{(du^1)^2 + \cdots + (du^n)^2}{(R^2 - |u|^2)^2}. \tag{1.3}$$

Hyperbolic sp

(d) ***(Poincaré half-space model)*** *$\mathbb{U}^n(R)$ is the upper half-space in $\mathbb{R}^n$ defined in coordinates $(x^1, \dots, x^{n-1}, y)$ by $\mathbb{U}^n(R) = \{(x, y) : y > 0\}$, endowed with the metric*

$$\breve{g}^4_R = R^2 \frac{(dx^1)^2 + \cdots + (dx^{n-1})^2 + (dy)^2}{y^2} \tag{1.4}$$

Hyperbolic sp

Proof. Let $R > 0$ be given. We need to verify that $\mathbb{H}^n(R)$ is actually a Riemannian submanifold of $\mathbb{R}^{n,1}$, or in other words that $\breve{g}^1_R$ is positive definite. One way to do this is to show, as we will below, that it is the pullback of $\breve{g}^2_R$ or $\breve{g}^3_R$ (both of which are manifestly positive definite) by a diffeomorphism. Alternatively, here is a direct proof using some of the theory of submanifolds of pseudo-Riemannian manifolds.

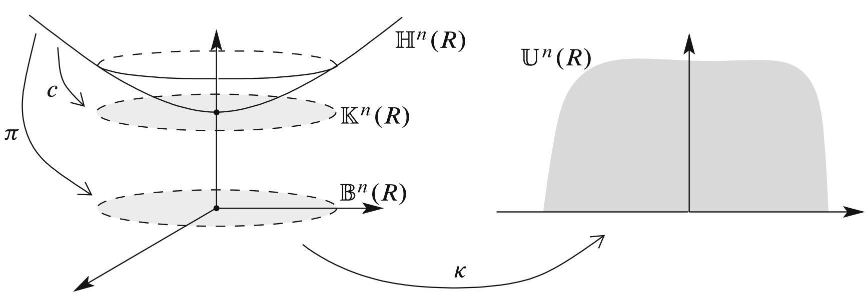

Figure 1.3: Isometries among the hyperbolic models.

Note that $\mathbb{H}^n(R)$ is an open subset of a level set of the smooth function $f : \mathbb{R}^{n,1} \to \mathbb{R}$ given by $f(\xi, \tau) = (\xi^1)^2 + \cdots + (\xi^n)^2 - \tau^2$. We have

$$df = 2\xi^1 d\xi^2 + \cdots + 2\xi^n d\xi^n - 2\tau d\tau,$$

and therefore the gradient of f with respect to $\bar{q}$ is given by

$$\operatorname{grad} f = 2\xi^1 \frac{\partial}{\partial \xi^1} + \cdots + 2\xi^n \frac{\partial}{\partial \xi^n} + 2\tau \frac{\partial}{\partial \tau}. \tag{1.5}$$

Hyperbolic sp

Direct computation gives $\bar{q}(\operatorname{grad} f, \operatorname{grad} f) = 4(\sum_i (\xi^i)^2 - \tau^2) = -4R^2 < 0$, therefore $\mathbb{H}^n(R)$ is a pseudo-Riemannian manifold of signature $(n, 0)$, that is, a Riemannian manifold.

We will show that all four Riemannian manifolds are mutually isometric by defining isometries $c : \mathbb{H}^n(R) \to \mathbb{K}^n(R)$, $\pi : \mathbb{H}^n(R) \to \mathbb{B}^n(R)$ and $\kappa : \mathbb{B}^n(R) \to \mathbb{U}^n(R)$.

We begin with a geometric construction of a diffeomorphism called **central projection** from the hyperboloid to the ball, $c : \mathbb{H}^n(R) \to \mathbb{K}^n(R)$, which turns out to be an isometry between the two metrics given in (a) and (b). For any $P = (\xi, \tau) \in \mathbb{R}^{n,1}$, set $c(P) = w \in \mathbb{K}^n(R)$, where $W = (w, 0) \in \mathbb{R}^{n,1}$ is the point where the line from the origin to P intersects the hyperplane $\{(\xi, \tau) : \tau = R\}$. Because W is characterized as the unique scalar multiple of P whose last coordinate is R, we have $W = RP/\tau$, and therefore $c(\xi, \tau) = R\xi/\tau$.

The relation $|\xi|^2 - \tau^2 = -\mathbb{R}^2$ guarantees that $|c(\xi, \tau)|^2 = R^2(1 - R^2/\tau^2) < R^2$, so c maps $\mathbb{H}^n(R)$ into $\mathbb{K}^n(R)$. To show that c is a diffeomorphism, we determine its inverse map. Let

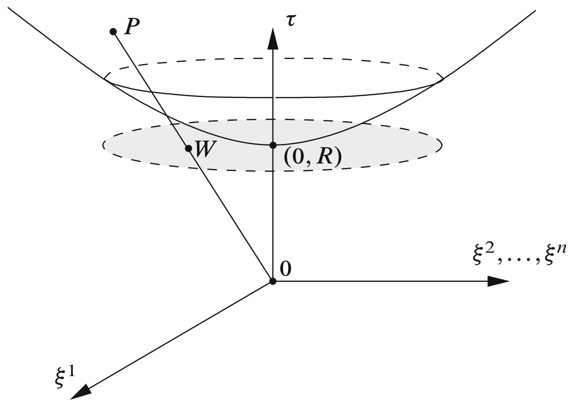

Figure 1.4: Central projection from the hyperboloid to the Beltrami-Klein model.

Hyperbolic ce

$w \in \mathbb{K}^n(R)$ be arbitrary. The unique positive scalar λ such that the point $(\xi, \tau) = \lambda(w, R)$ lies on $\mathbb{H}^n(R)$ is characterized by $\lambda^2|w|^2 - \lambda^2 R^2 = -R^2$, and therefore

$$\lambda = \frac{R}{\sqrt{R^2 - |w|^2}}.$$

It follows that the following smooth map is an inverse for c:

$$c^{-1}(w) = (\xi, \tau) = \Big(\frac{Rw}{\sqrt{R^2 - |w|^2}}, \frac{R^2}{\sqrt{R^2 - |w|^2}}\Big).$$

Thus c is a diffeomorphism. To show that it is an isometry between $\breve{g}^1_R$ and $\breve{g}^2_R$, we use the fact that $\breve{g}^1_R$ is the metric induced from $\bar{q}$, analogously to the computation we did for stereographic projection above. With (ξ, τ) defined above, we have

$$d\xi^i = \frac{R dw^i}{\sqrt{R^2 - |w|^2}} + \frac{Rw^i \sum_j w^j dw^j}{(R^2 - |w|^2)^{3/2}}, \quad d\tau = \frac{R^2 \sum_j w^j dw^j}{(R^2 - |w|^2)^{3/2}}.$$

It then follows that

$$\begin{aligned}
(c^{-1})^* \breve{g}^1_R &= \sum_i (d\xi^i)^2 - (d\tau)^2 \\
&= \frac{R^2 \sum_i (dw^i)^2}{R^2 - |w|^2} + \frac{R^2 |w|^2 (\sum_j w^j dw^j)^2}{(R^2 - |w|^2)^3} + \frac{2R^3 (\sum_j w^j dw^j)^2}{(R^2 - |w|^2)^2} - \frac{R^4 (\sum_j w^j dw^j)^2}{(R^2 - |w|^2)^3} \\
&= R^2 \frac{\sum_i (dw^i)^2}{R^2 - |w|^2} + R^2 \frac{(\sum_j w^j dw^j)^2}{(R^2 - |w|^2)^2}.
\end{aligned}$$

Next we describe a diffeomorphism $\pi : \mathbb{H}^n(R) \to \mathbb{B}^n(R)$ from the hyperboloid to the ball, called **hyperbolic stereographic projection**, which is an isometry between the metrics of (a) and (c). Let $S \in \mathbb{R}^{n,1}$ denote the point $S = (0, \dots, 0, -R)$. For any $P = (\xi, \tau) \in \mathbb{H}^n(R) \subseteq \mathbb{R}^{n,1}$, set $\pi(P) = u \in \mathbb{B}^n(R)$, where $U = (u, 0) \in \mathbb{R}^{n,1}$ is the point where the line through S and P intersects the hyperplane $\{(\xi, \tau) : \tau = 0\}$. The point U is characterized by $(U - S) = \lambda(P - S)$ for some nonzero scalar λ, so we get $\lambda = R/(R + \tau)$ and thus

$$\pi(\xi, \tau) = u = \frac{R\xi}{R + \tau},$$

which takes its values in $\mathbb{B}^n(R)$ because $|\pi(\xi, \tau)|^2 = R^2(\tau^2 - R^2)/(R + \tau)^2 < R^2$. A computation similar to the ones before shows that the inverse map is

$$\pi^{-1}(u) = (\xi, \tau) = \Big(\frac{2R^2 u}{R^2 - |u|^2}, R\frac{R^2 + |u|^2}{R^2 + |u|^2} \Big).$$

We will show that $(\pi^{-1})^* \breve{g}^1_R = \breve{g}^3_R$. The computation proceeds just as in the spherical case, with

$$d\xi^i = \frac{2R^2 du^i}{R^2 - |u|^2} + 4R^2 \frac{u^i \sum_j u^j du^j}{(R^2 - |u|^2)^2}, \quad d\tau = \frac{4R^2 \sum_j u^j du^j}{(R^2 - |u|^2)^2}$$

and therefore

$$\begin{aligned} (\pi^{-1})^* \breve{g}^1_R &= \frac{4R^2 \sum_i (du^i)^2}{(R^2 - |u|^2)^2} + \frac{16R^4 (\sum_j u^j du^j)^2}{(R^2 - |u|^2)^3} + \frac{16R^4 |u|^2 \sum_j u^j du^j}{(R^2 - |u|^2)^4} - \frac{16R^6 \sum_j u^j du^j}{(R^2 - |u|^2)^4} \\ &= 4R^2 \frac{\sum_i (du^i)^2}{(R^2 - |u|^2)^2} = \breve{g}^3_R. \end{aligned}$$

Now we consider the Poincaré half-space model, by constructing an explicit diffeomorphism

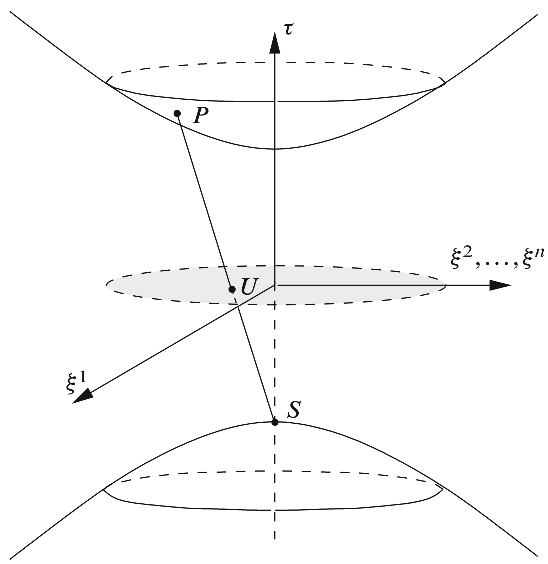

Figure 1.5: Hyperbolic stereographic projection. Hyperbolic st

$\kappa : \mathbb{U}^n(R) \to \mathbb{B}^n(R)$. In this case it is convenient to write the coordinates on the ball as $(u, v) = (u^1, \ldots, u^{n-1}, v)$. In the 2-dimensional case, it is easy to write down in complex notation $w = u + iv$ and $z = x + iy$. It is a Mobious transformation:

$$\kappa(z) = w = iR\frac{z - iR}{z + iR}. \tag{1.6}$$

Hyperbolic sp

Since $\kappa(iR) = 0$, $\kappa(0) = -iR$ and $\kappa(\infty) = iR$, it is a complex-analytic diffeomorphism taking $\mathbb{U}^2(R)$ onto $\mathbb{B}^2(R)$. Separating z into real and imaginary parts, we can also write this in real terms as

$$\kappa(x, y) = (u, v) = \Big(\frac{2R^2 x}{|x|^2 + (y + R)^2}, R\frac{|x|^2 + y^2 - R^2}{|x|^2 + (y + R)^2} \Big).$$

This same formula makes sense in any dimension n if we interpret x to mean $(x^1, \ldots, x^{n-1})$, and it is easy to check that it maps the upper half-space $\{y > 0\}$ into the ball of radius R.

Hyperbolic space-1

From the formula (1.6) we can construct the inverse of κ:

$$\kappa^{-1}(u,v)=(x,y)=\Big(\frac{2R^2u}{|u|^2+(v-R)^2},R\frac{R^2-|u|^2-v^2}{|u|^2+(v-R)^2}\Big),$$

so κ is a diffeomorphism. We compute that

$$du^i=\frac{2R^2dx^i}{|x|^2+(y+R)^2}-\frac{4R^2x^i\sum_j x^jdx^j}{(|x|^2+(y+R)^2)^2}-\frac{4R^2x^i(y+R)dy}{(|x|^2+(y+R)^2)^2},$$

and

$$\begin{aligned}dv&=\frac{2Rx^i\sum_j x^jdx^j}{|x|^2+(y+R)^2}-\frac{2R(|x|^2+y^2-R^2)\sum_j x^jdx^j}{(|x|^2+(y+R)^2)^2}\\&\quad+\frac{2Ry(|x|^2+(y+R)^2)-2(y+R)(|x|^2+(y+R)^2)}{(|x|^2+(y+R)^2)^2}\\&=\frac{4R^2(y+R)\sum_j x^jdx^j}{(|x|^2+(y+R)^2)^2}+\frac{2R^2((y+R)^2-|x|^2)dy}{(|x|^2+(y+R)^2)^2}.\end{aligned}$$

With these, we have (for simplicity, we set $\Delta:=|x|^2+(y+R)^2$)

$$\begin{aligned}&(du^1)^2+\cdots+(du^{n-1})^2+(dv)^2\\&=\frac{4R^4\sum_i(dx^i)^2}{\Delta^2}+\frac{16R^4|x|^2(\sum_j x^jdx^j)^2}{\Delta^4}+\frac{16R^4|x|^2(y+R)^2(dy)^2}{\Delta^4}\\&\quad-\frac{16R^4(\sum_j x^jdx^j)^2}{\Delta^3}-\frac{16R^4(y+R)(\sum_j x^jdx^j)dy}{\Delta^3}+\frac{32R^4|x|^2(y+R)(\sum_j x^jdx^j)dy}{\Delta^4}\\&\quad+\frac{16R^4(y+R)^2(\sum_j x^jdx^j)^2}{\Delta^4}+\frac{4R^4[(y+R)^2-|x|^2]^2(dy)^2}{\Delta^4}\\&\quad+\frac{16R^4(y+R)[(y+R)^2-|x|^2](\sum_j x^jdx^j)dy}{\Delta^4}\\&=\frac{4R^2}{\Delta^2}\sum_i(dx^i)^2+(dy)^2\Big[\frac{16R^4|x|^2(y+R)^2}{\Delta^4}+\frac{4R^4[(y+R)^2-|x|^2]^2}{\Delta^4}\Big]\\&\quad+(\sum_j x^jdx^j)^2\Big[\frac{16R^4|x|^2}{\Delta^4}-\frac{16R^4}{\Delta^3}+\frac{16R^4(y+R)^2}{\Delta^4}\Big]\\&\quad+(\sum_j x^jdx^j)dy\Big[-\frac{16R^4(y+R)}{\Delta^3}+\frac{32R^4|x|^2(y+R)}{\Delta^4}+\frac{16R^4(y+R)[(y+R)^2-|x|^2]}{\Delta^4}\Big].\end{aligned}$$

Now we analyse the summands. First we have

$$\begin{aligned}&\frac{16R^4|x|^2(y+R)^2}{\Delta^4}+\frac{4R^4[(y+R)^2-|x|^2]^2}{\Delta^4}=\frac{4R^4}{\Delta^4}(4|x|^2(y+R)^2+[(y+R)^2-|x|^2]^2)\\&=\frac{4R^4}{\Delta^4}[(y+R)^2+|x|^2]^2=\frac{4R^4}{\Delta^2}.\end{aligned}$$

Second, note that

$$\frac{16R^4|x|^2}{\Delta^4}-\frac{16R^4}{\Delta^3}+\frac{16R^4(y+R)^2}{\Delta^4}=\frac{16R^4}{\Delta^4}(|x|^2-\Delta+(y+R)^2)=0,$$

and finally,

$$\begin{aligned}
&-\frac{16R^4(y+R)}{\Delta^3}+\frac{32R^4|x|^2(y+R)}{\Delta^4}+\frac{16R^4(y+R)[(y+R)^2-|x|^2]}{\Delta^4}\\
&=\frac{16R^4(y+R)}{\Delta^4}(\Delta-2|x|^2+(y+R)^2-|x|^2)=0.
\end{aligned}$$

Therefore we derive from our previous calculation that

$$(du^1)^2+\cdots+(du^{n-1})^2+(dv)^2=\frac{4R^4}{\Delta^2}((dx^1)^2+\cdots+(dx^{n-1})^2+(dy)^2).$$

To end the verification, we observe that

$$\begin{aligned}
|u|^2+v^2&=\frac{R^2}{\Delta^2}(4R^2|x|^2+(|x|^2+y^2-R^2)^2))\\
&=\frac{R^2}{\Delta^2}(|x|^4+y^2+R^2+2R^2|x|^2+2|x|^2y^2-2R^2y^2)\\
&=\frac{R^2}{\Delta^2}((|x|^2+y^2+R^2)^2-(2Ry)^2)\\
&=\frac{R^2(|x|^2+(y+R)^2)(|x|^2+(y-R)^2)}{\Delta^2}\\
&=\frac{R^2(|x|^2+(y-R)^2)}{\Delta}.
\end{aligned}$$

and therefore

$$\frac{1}{(R^2-|(u,v)|^2)^2}=\frac{\Delta^2}{R^2(\Delta-|x|^2-(y-R)^2)^2}=\frac{\Delta^2}{16R^6y^2}.$$

Once this is established, we then get

$$\begin{aligned}
\kappa^*\breve{g}^3_R&=\frac{4R^4}{(R^2-|(u,v)|^2)^2}((du^1)^2+\cdots+(du^{n-1})^2+(dv)^2)\\
&=\frac{4R^4\Delta^2}{16R^6y^2}\cdot\frac{4R^4}{\Delta^2}((dx^1)^2+\cdots+(dx^{n-1})^2+(dy)^2)\\
&=R^2\frac{(dx^1)^2+\cdots+(dx^{n-1})^2+(dy)^2}{y^2}\\
&=\breve{g}^4_R.
\end{aligned}$$

This finishes the proof. □

We often use the generic notation $\mathbb{H}^n(R)$ to refer to any one of the Riemannian manifolds of Theorem 1.1.6, and $\breve{g}_R$ to refer to the corresponding metric; the special case $R=1$ is denoted by $(\mathbb{H}^n,\breve{g})$ and is called simply **hyperbolic space**, or in the 2-dimensional case, the **hyperbolic plane**.

Because all of the models for a given value of R are isometric to each other, when analyzing them geometrically we can use whichever model is most convenient for the application we have in mind. The next corollary is an example in which the Poincaré ball and half-space models serve best.

Corollary 1.1.7. *Each hyperbolic space is locally conformally flat.*

Proof. In either the Poincaré ball model or the half-space model, the identity map gives a global conformal equivalence with an open subset of Euclidean space. □

The symmetries of $\mathbb{H}^n(R)$ are most easily seen in the hyperboloid model. Let $\mathrm{O}(n,1)$ denote the group of linear maps from $\mathbb{R}^{n,1}$ to itself that preserve the Minkowski metric, called the $(n+1)$-dimensional Lorentz group. Note that each element of $\mathrm{O}(n,1)$ preserves the hyperboloid $\{|\xi|^2 - \tau^2 = -R^2\}$, which has two components determined by $\tau > 0$ and $\tau < 0$. We let $\mathrm{O}^+(n,1)$ denote the subgroup of $\mathrm{O}(n,1)$ consisting of maps that take the $\tau > 0$ component of the hyperboloid to itself. (This is called the **orthochronous Lorentz group**, because physically it represents coordinate changes that preserve the forward time direction.) Then $\mathrm{O}^+(n,1)$ preserves $\mathbb{H}^n(R)$, and because it preserves $\bar{q}$ it acts isometrically on $\mathbb{H}^n(R)$.

Proposition 1.1.8. *The group $\mathrm{O}^+(n,1)$ acts transitively on $\mathrm{O}(\mathbb{H}^n(R))$, and therefore $\mathbb{H}^n(R)$ is frame-homogeneous.*

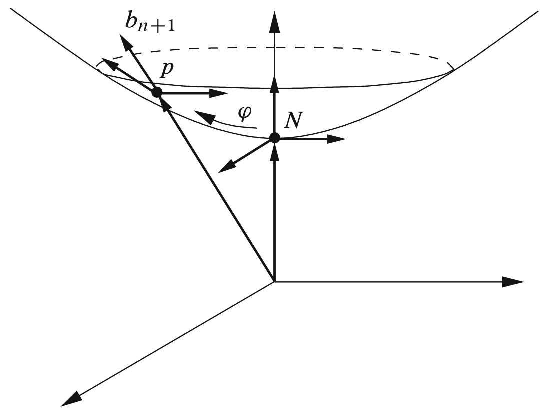

Figure 1.6: Frame homogeneity of $\mathbb{H}^n(R)$.

Proof. The argument is entirely analogous to the proof of Proposition 1.1.2, so we give only a sketch. Suppose $p \in \mathbb{H}^n(R)$ and (e_i) is an orthonormal basis for $T_p\mathbb{H}^n(R)$. Identifying $p \in \mathbb{R}^{n,1}$ with an element of $T_p\mathbb{R}^{n,1}$ in the usual way, we can regard $\widehat{p} = p/R$ as a $\bar{q}$-unit vector in $T_p\mathbb{R}^{n,1}$, and (1.5) shows that it is a scalar multiple of the $\bar{q}$-gradient of the defining function f and thus is orthogonal to $T_p\mathbb{H}^n(R)$ with respect to $\bar{q}$. Thus $(e_1, \dots, e_n, e_{n+1} = \widehat{p})$ is a $\bar{q}$-orthonormal basis for $\mathbb{R}^{n,1}$, and $\bar{q}$ has the following expression in terms of the dual basis (β^i):

$$\bar{q} = (\beta^1)^2 + \cdots + (\beta^n)^2 - (\beta^{n+1})^2.$$

Thus the matrix whose columns are $(e_1, \dots, e_{n+1})$ is an element of $\mathrm{O}^+(n,1)$ sending $N = (0, \dots, 0, R)$ to p and ∂_i to e_i. □

1.1.5 Invariant metrics on Lie groups

Lie groups provide us with another large class of homogeneous Riemannian manifolds. Let G be a Lie group. A Riemannian metric g on G is said to be **left-invariant** if it is invariant under all left translations: $L_\varphi^* g = g$ for all $\varphi \in G$. Similarly, g is **right-invariant** if it is invariant under all right translations, and **bi-invariant** if it is both left- and right-invariant. The next lemma shows that left-invariant metrics are easy to come by.

Lemma 1.1.9. *Let G be a Lie group and let $\mathfrak{g}$ be its Lie algebra of left-invariant vector fields.*

(a) *A Riemannian metric g on G is left-invariant if and only if for all $X, Y \in \mathfrak{g}$, the function $g(X,Y)$ is constant on G.*

(b) *The restriction map $g \mapsto g_e \in \Sigma^2(T_e^*G)$ together with the natural identification $T_eG \cong \mathfrak{g}$ gives a bijection between left-invariant Riemannian metrics on G and inner products on g.*

Proof. We only need to prove (a). Note that, since $\mathfrak{g}$ consists of left-invariant vector fields, for every $\varphi \in G$ we have

$$g(X,Y)_\varphi = g_\varphi(X_\varphi, Y_\varphi) = g_\varphi(d(L_\varphi)_e(X_e), d(L_\varphi)_e(Y_e)) = (L_\varphi^* g)_e(X_e, Y_e) = (L_\varphi^* g)(X,Y)_e.$$

Therefore $g(X,Y)$ is constant on G if g is left-invariant.

Conversely, if $g(X,Y)$ is constant for all $X, Y \in \mathfrak{g}$, we prove that $L_\varphi^* g = g$ for any $\varphi \in G$. In fact, fix φ and let $\psi \in G$ and $u, v \in T_\psi G$, then we can choose $X, Y \in \mathfrak{g}$ such that $X_\psi = u$, $Y_\psi = v$, so it follows that

$$\begin{aligned}
(L_\varphi^* g)_\psi(u,v) &= (L_\varphi^* g)_\psi(X,Y)_\psi \\
&= g_{\varphi\psi}(d(L_\varphi)_\psi(X_\psi), d(L_\varphi)_\psi(Y_\psi)) \\
&= g_{\varphi\psi}(X_{\varphi\psi}, Y_{\varphi\psi}) \\
&= g(X,Y)_{\varphi\psi} \\
&= g(X,Y)_\psi.
\end{aligned}$$

This then implies $L_\varphi^* g = g$, so g is left-invariant. □

Thus all we need to do to construct a left-invariant metric is choose any inner product on g, and define a metric on G by applying that inner product to left-invariant vector fields. Right-invariant metrics can be constructed in a similar way using right-invariant vector fields. Since a Lie group acts transitively on itself by either left or right translation, every left-invariant or right-invariant metric is homogeneous.

Much more interesting are the bi-invariant metrics. Fortunately, there is a complete answer to the question of which Lie groups admit bi-invariant metrics.

We begin with a proposition that shows how to determine whether a given leftinvariant metric is bi-invariant, based on properties of the adjoint representation of the group. Recall that this is the representation $\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g})$ given by $\mathrm{Ad}(\varphi) = (C_\varphi)_* : \mathfrak{g} \to \mathfrak{g}$, where C_φ is the automorphism defined by conjugation: $C_\varphi(\psi) = \varphi\psi\varphi^{-1}$.

ic bi-inv iff **Proposition 1.1.10.** *Let G be a Lie group and $\mathfrak{g}$ its Lie algebra. Suppose g is a left-invariant Riemannian metric on G, and let $\langle \cdot\,, \cdot \rangle$ denote the corresponding inner product on $\mathfrak{g}$ as in Lemma 1.1.9. Then* Riemann metric left-inv *g is bi-invariant if and only if $\langle \cdot\,, \cdot \rangle$ is invariant under the action of $\mathrm{Ad}(G) \subseteq \mathrm{GL}(\mathfrak{g})$, in the sense that $\langle \mathrm{Ad}(\varphi)X, \mathrm{Ad}(\varphi)Y \rangle = \langle X, Y \rangle$ for all $X, Y \in \mathfrak{g}$ and $\varphi \in G$.*

Proof. We begin the proof with some preliminary computations. Suppose g is left-invariant and $\langle \cdot\,, \cdot \rangle$ is the associated inner product on g. Let $\varphi \in G$ be arbitrary, and note that C_φ is the composition of left multiplication by φ followed by right multiplication by φ^{-1}. Thus for every $X \in \mathfrak{g}$, left-invariance implies $(R_{\varphi^{-1}})_* X = (R_{\varphi^{-1}})_*(L_\varphi)_* X = (C_\varphi)_* X = \mathrm{Ad}(\varphi)X$. Therefore, for all $\psi \in G$ and $X, Y \in \mathfrak{g}$, we have

$$\begin{aligned}
((R_{\varphi^{-1}})^* g)_\psi(X_\psi, Y_\psi) &= g_{\psi\varphi^{-1}}\big(((R_{\varphi^{-1}})_* X)_{\psi\varphi^{-1}}, ((R_{\varphi^{-1}})_* Y)_{\psi\varphi^{-1}}\big) \\
&= g_{\psi\varphi^{-1}}\big(\mathrm{Ad}(\varphi)X)_{\psi\varphi^{-1}}, \mathrm{Ad}(\varphi)Y)_{\psi\varphi^{-1}}\big) \\
&= \langle \mathrm{Ad}(\varphi)X, \mathrm{Ad}(\varphi)Y \rangle_\psi.
\end{aligned}$$

Now assume that $\langle\cdot\,,\cdot\rangle$ is invariant under $\mathrm{Ad}(G)$. Then the expression on the last line above is equal to $\langle X,Y\rangle_\psi = g_\psi(X_\psi,Y_\psi)$, which shows that $(R_{\varphi^{-1}})^*g = g$. Since this is true for all $\varphi \in G$, it follows that g is bi-invariant.

Conversely, assuming that g is bi-invariant, we have $(R_{\varphi^{-1}})^*g = g$ for each $\varphi \in G$, so the above computation yields

$$\langle X,Y\rangle_\psi = g_\psi(X_\psi,Y_\psi) = ((R_{\varphi^{-1}})^*g)_\psi(X_\psi,Y_\psi) = \langle \mathrm{Ad}(\varphi)X, \mathrm{Ad}(\varphi)Y\rangle_\psi,$$

which shows that $\langle\cdot\,,\cdot\rangle$ is $\mathrm{Ad}(G)$-invariant. □

In order to apply the preceding proposition, we need a lemma about finding invariant inner products on vector spaces. Recall that for every finite-dimensional real vector space V, $GL(V)$ denotes the Lie group of all invertible linear maps from V to itself. If H is a subgroup of $\mathrm{GL}(V)$, an inner product $\langle\cdot\,,\cdot\rangle$ on V is said to be ***H*-invariant** if $\langle hx,hy\rangle = \langle x,y\rangle$ for all $x,y \in V$ and $h \in H$.

prod inv iff **Lemma 1.1.11.** *Suppose V is a finite-dimensional real vector space and H is a subgroup of $\mathrm{GL}(V)$. There exists an H-invariant inner product on V if and only if H has compact closure in $\mathrm{GL}(V)$.*

Proof. Assume first that there exists an H-invariant inner product $\langle\cdot\,,\cdot\rangle$ on V. This implies that H is contained in the subgroup $\mathrm{O}(V) \subseteq \mathrm{GL}(V)$ consisting of linear isomorphisms of V that are orthogonal with respect to this inner product. Choosing an orthonormal basis of V yields a Lie group isomorphism between $\mathrm{O}(V)$ and $\mathrm{O}(n) \subseteq \mathrm{GL}_n(\mathbb{R})$ (where $n = \dim V$), so $\mathrm{O}(V)$ is compact; and the closure of H is a closed subset of this compact group, and thus is itself compact.

Conversely, suppose H has compact closure in $\mathrm{GL}(V)$, and let K denote the closure. Then by Proposition topo subgroup closure **??** K is also a subgroup, and thus it is a Lie group by the closed subgroup theorem. Let $\langle\cdot\,,\cdot\rangle$ be an arbitrary inner product on V, and let μ be a right-invariant volume form on K (for example, the volume form of some right-invariant metric on K). Then define a new inner product $(\cdot\,,\cdot)$ on V by

$$(x,y) = \int_K \langle kx,ky\rangle\mu.$$

It follows directly from the definition that $(\cdot\,,\cdot)$ is symmetric and bilinear over $\mathbb{R}$. For each nonzero $x \in V$, we have $\langle kx,kx\rangle > 0$ everywhere on K, so $(x,y) > 0$, showing that $(\cdot\,,\cdot)$ is indeed an inner product.

To see that it is invariant under K, let $k_0 \in K$ be arbitrary. Then for all $x,y \in V$ and $k \in K$, we have

$$\begin{aligned}(k_0x,k_0y) &= \int_K \langle kk_0x,kk_0y\rangle\mu \\ &= \int_K R_{k_0}^*(\langle kx,ky\rangle\mu) \\ &= \int_K \langle kx,ky\rangle\mu = (x,y).\end{aligned}$$

Thus $(\cdot\,,\cdot)$ is K-invariant, and it is also H-invariant because $H \subseteq K$. □

bi-inv exist **Theorem 1.1.12.** *Let G be a Lie group and $\mathfrak{g}$ its Lie algebra. Then G admits a bi-invariant metric if and only if $\mathrm{Ad}(G)$ has compact closure in $\mathrm{GL}(\mathfrak{g})$.*

The most important application of the preceding theorem is to compact groups.

compact group

Corollary 1.1.13. *Every compact Lie group admits a bi-invariant Riemannian metric.*

Proof. If G is compact, then $\mathrm{Ad}(G)$ is a compact subgroup of $\mathrm{GL}(\mathfrak{g})$ because $\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g})$ is continuous. □

Another application is to prove that certain Lie groups do not admit bi-invariant metrics. One way to do this is to note that if $\mathrm{Ad}(G)$ has compact closure in $\mathrm{GL}(\mathfrak{g})$, then every orbit of $\mathrm{Ad}(G)$ must be a bounded subset of $\mathfrak{g}$ with respect to any choice of norm, because it is contained in the image of the compact set $\overline{\mathrm{Ad(G)}}$ under a continuous map of the form $\varphi \mapsto \varphi(X_0)$ from $\mathrm{GL}(\mathfrak{g})$ to $\mathfrak{g}$. Thus if one can find an element $X_0 \in \mathfrak{g}$ and a subset $S \subseteq G$ such that the elements of the form $\mathrm{Ad}(\varphi)X_0$ are unbounded in $\mathfrak{g}$ for $\varphi \in S$, then there is no bi-invariant metric.

Here are some examples.

Example 1.1.14 (Invariant Metrics on Lie Groups).

(a) Every left-invariant metric on an abelian Lie group is bi-invariant, because the adjoint representation is trivial. Thus the Euclidean metric on $\mathbb{R}^n$ and the flat metric on T^n are both bi-invariant.

(b) If a metric g on a Lie group G is left-invariant, then the induced metric on every Lie subgroup $H \subseteq G$ is easily seen to be left-invariant. Similarly, if g is bi-invariant, then the induced metric on H is bi-invariant.

(c) The Lie group $\mathrm{SL}_2(\mathbb{R})$ admits many left-invariant metrics, but no bi-invariant ones. To see this, recall that the Lie algebra of $\mathrm{SL}_2(\mathbb{R})$ is isomorphic to the algebra $\mathfrak{sl}(2, \mathbb{R})$ of trace-free 2×2 matrices, and the adjoint representation is given by $\mathrm{Ad}(A)X = AXA^{-1}$. If we let $X_0 = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) \in \mathfrak{sl}(2, \mathbb{R})$ and $A_c = \begin{smallmatrix} c & 0 \\ 0 & 1/c \end{smallmatrix} \in \mathrm{SL}_2(\mathbb{R})$ for $c > 0$, then $\mathrm{Ad}(A_c)X_0 = \left(\begin{smallmatrix} 0 & c^2 \\ 0 & 0 \end{smallmatrix}\right)$, which is unbounded as $c \to +\infty$. Thus the orbit of X_0 is not contained in any compact subset, which implies that there is no bi-invariant metric on $\mathrm{SL}_2(\mathbb{R})$. A similar argument shows that $\mathrm{SL}_n(\mathbb{R})$ admits no bi-invariant metric for any $n \geq 2$. In view of (b) above, this shows also that $\mathrm{GL}_n(\mathbb{R})$ admits no bi-invariant metric for $n \geq 2$. (Of course, $\mathrm{GL}_1(\mathbb{R})$ does admit bi-invariant metrics because it is abelian.)

(d) With S^3 regarded as a submanifold of $\mathbb{C}^2$, the map

$$(w, z) \mapsto \begin{pmatrix} w & z \\ -\bar{z} & \bar{w} \end{pmatrix}$$

gives a diffeomorphism from S^3 to $\mathrm{SU}(2)$. Under the inverse of this map, the round metric on S^3 pulls back to a bi-invariant metric on $\mathrm{SU}(2)$.

(e) Let $\mathfrak{o}(n)$ denote the Lie algebra of $\mathrm{O}(n)$, identified with the algebra of skew symmetric $n \times n$ matrices, and define a bilinear form on $\mathfrak{o}(n)$ by

$$\langle A, B \rangle = \mathrm{tr}(A^T B).$$

This is an Ad-invariant inner product, and thus determines a bi-invariant Riemannian metric on $\mathrm{O}(n)$.

(f) Let $\mathbb{U}^n$ be the upper half-space as defined in Theorem 1.1.6. We can regard $\mathbb{U}^n$ as a Lie group by identifying each point $(x, y) = (x^1, \dots, x^{n-1}, y) \in \mathbb{U}^n$ with an invertible $n \times n$ matrix as follows:

$$(x, y) \longleftrightarrow \begin{pmatrix} I_{n-1} & 0 \\ x^T & y \end{pmatrix}$$

Then the hyperbolic metric $\breve{g}^4_R$ is left-invariant on $\mathbb{U}^n$ but not right-invariant.

(g) For $n \geq 1$, the $(2n+1)$-dimensional **Heisenberg group** is the Lie subgroup $H^n \subseteq \mathrm{GL}_{n+2}(\mathbb{R})$ defined by

$$H_n = \left\{ \begin{pmatrix} 1 & x^T & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y \in \mathbb{R}^n, z \in \mathbb{R} \right\},$$

where x and y are treated as column matrices. These are the simplest examples of **nilpotent Lie groups**. There are many left-invariant metrics on H_n, but no bi-invariant ones.

(h) Our last example is a group that plays an important role in the classification of 3-manifolds. Let Sol denote the following 3-dimensional Lie subgroup of $\mathrm{GL}_3(\mathbb{R})$:

$$\mathrm{Sol} = \left\{ \begin{pmatrix} e^z & 0 & x \\ 0 & e^{-z} & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{R} \right\}.$$

This group is the simplest non-nilpotent example of a **solvable Lie group**. Like the Heisenberg groups, Sol admits left-invariant metrics but not bi-invariant ones.

1.1.6 Other homogeneous Riemannian manifolds

There are many homogeneous Riemannian manifolds besides the frame-homogeneous ones and the Lie groups with invariant metrics. To identify other examples, it is natural to ask the following question: If M is a smooth manifold endowed with a smooth, transitive action by a Lie group G (called a homogeneous G-space or just a homogeneous space), is there a Riemannian metric on M that is invariant under the group action?

The next theorem gives a necessary and sufficient condition for existence of an invariant Riemannian metric that is usually easy to check.

Proposition 1.1.15 (Existence of Invariant Metrics on Homogeneous Spaces). *Suppose G is a Lie group and M is a homogeneous G-space. Let p_0 be a point in M, and let $I_{p_0} G_{p_0} \to \mathrm{GL}(T_{p_0} M)$ denote the isotropy representation at p_0. There exists a G-invariant Riemannian metric on M if and only if $I_{p_0}(G_{p_0})$ has compact closure in $\mathrm{GL}(T_{p_0} M)$.*

Proof. Assume first that g is a G-invariant metric on M. Then the inner product g_{p_0} on $T_{p_0} M$ is invariant under the isotropy representation, so it follows from Lemma 1.1.11 that $I_{p_0}(G_{p_0})$ has compact closure in $\mathrm{GL}(T_{p_0} M)$.

Conversely, assume that $I_{p_0}(G_{p_0})$ has compact closure in $\mathrm{GL}(T_{p_0} M)$. Lemma 1.1.11 shows that there is an inner product g_{p_0} on $T_{p_0} M$ that is invariant under the isotropy representation. For arbitrary $p \in M$, we define an inner product g_p on $T_p M$ by choosing an element $\varphi \in G$ such that $\varphi(p) = p_0$ and setting

$$g_p = (d\varphi_p)^* g_{p_0}.$$

If φ_1, φ_2 are any two such elements of G, then $\varphi_1 = h\varphi_2$ with $h = \varphi_1\varphi_2^{-1} \in G_{p_0}$, so

$$(d\varphi_1|_p)^* g_{p_0} = (d(h\varphi_2)|_p)^* g_{p_0} = (d\varphi_2|p)^*(dh_{p_0})^* g_{p_0} = (d\varphi_2|_p)^* g_{p_0},$$

showing that g is well defined as a rough tensor field on M. It is clear that g is G-invariant, so it remains only to show that it is smooth.

The map $\pi : G \to M$ given by $\pi(\psi) = \psi \cdot p_0$ is a smooth surjection because the action is smooth and transitive. Given $\varphi \in G$, if we let $\theta_\varphi : M \to M$ denote the map $p \mapsto \varphi \cdot p$ and $L_\varphi : G \to G$ the left translation by φ, then the map π satisfies

$$\pi \circ L_\varphi(\psi) = (\varphi\psi) \cdot p_0 = \varphi(\psi \cdot p_0) = \theta_\varphi \circ \pi(\psi) \tag{1.7}$$ Riemann G-inv

so it is equivariant with respect to these two actions. Thus it is a submersion by the equivariant rank theorem.

Define a rough 2-tensor field τ on G by $\tau = \pi^* g$. For all $\varphi \in G$, (Riemann G-inv metric-1 1.7) implies

$$L_\varphi^* \tau = L_\varphi^* \pi^* g = (\pi \circ L_\varphi)^* g = \pi^* \theta_\varphi^* g = \pi^* g = \tau,$$

where the next-to-last equality follows from the G-invariance of g. Thus τ is a left-invariant tensor field on G. Every basis $(X_1, \dots, X_n)$ for the Lie algebra of G forms a smooth global left-invariant frame for G, and with respect to such a frame the components $\tau(X_i, X_j)$ are constant; thus τ is a smooth tensor field on G (Corollary Lie rough left-inv ??).

For each $p \in M$, the fact that π is a surjective smooth submersion implies that there exist a neighborhood U of p and a smooth local section $\sigma : U \to G$. Then

$$g|_U = (\pi \circ \sigma)^* g = \sigma^* \pi^* g = \sigma^* \tau,$$

showing that g is smooth on U. Since this holds in a neighborhood of each point, g is smooth. □

Corollary 1.1.16. *Suppose G is a Lie group and M is a homogeneous G-space that admits at least one g-invariant metric. Show that for each $p \in M$, the map $g \mapsto g_p$ gives a bijection between G-invariant metrics on M and $I_p(G_p)$-invariant inner products on T_pM.*

The next corollary, which follows immediately from Theorem Riemann G-inv metric 1.1.15, addresses the most commonly encountered case.

nv if isotopy **Corollary 1.1.17.** *If a Lie group G acts smoothly and transitively on a smooth manifold M with compact isotropy groups, then there exists a G-invariant Riemannian metric on M.*

1.1.7 Model pseudo-Riemannian manifolds

The definitions of the Euclidean, spherical, and hyperbolic metrics can easily be adapted to give analogous classes of frame-homogeneous pseudo-Riemannian manifolds.

The first example is one we have already seen: the pseudo-Euclidean space of signature (r, s) is the pseudo-Riemannian manifold $(\mathbb{R}^{r,s}, \bar{q}^{(r,s)})$, where $\bar{q}^{(r,s)}$ is the pseudo-Riemannian metric defined by

$$\bar{q}^{(r,s)} = (d\xi^1)^2 + \cdots + (d\xi^r)^2 - (d\tau^1)^2 - \cdots - (d\tau^s)^2.$$

There are also pseudo-Riemannian analogues of the spherical and hyperbolic metrics. For nonnegative integers r and s and a positive real number $\mathbb{R}$, we define the pseudospher $(S^{r,s}(R), \mathring{q}_R^{(r,s)})$ and the pseudohyperbolic space $(\mathbb{H}^{r,s}(R), \breve{q}_R^{(r,s)})$ as follows. As manifolds,

$S^{r,s}(R) \subseteq \mathbb{R}^{r+1,s}$ and $\mathbb{H}^{r,s}(R) \subseteq \mathbb{R}^{r,s+1}$ are defined by

$$S^{r,s}(R) = \{(\xi, \tau) : (\xi^1)^2 + \cdots + (\xi^{r+1})^2 - (\tau^1)^2 - \cdots - (\tau^r)^2 = R^2\},$$
$$\mathbb{H}^{r,s}(R) = \{(\xi, \tau) : (\xi^1)^2 + \cdots + (\xi^r)^2 - (\tau^1)^2 - \cdots - (\tau^{r+1})^2 = -R^2\},$$

The metrics are the ones induced from the respective pseudo-Euclidean metrics.

Theorem 1.1.18. *For all r, s, and R as above, $S^{r,s}(R)$ and $\mathbb{H}^{r,s}(R)$ are pseudo-Riemannian manifolds of signature (r, s).*

Proof. The defining function for $S^{r,s}$ and $\mathbb{H}^{r,s}$ are

$$f(\xi, \tau) = (\xi^1)^2 + \cdots + (\xi^{r+1})^2 - (\tau^1)^2 - \cdots - (\tau^s)^2,$$
$$g(\xi, \tau) = (\xi^1)^2 + \cdots + (\xi^r)^2 - (\tau^1)^2 - \cdots - (\tau^{s+1})^2.$$

Note that we have

$$df = 2\xi^1 d\xi^1 + \cdots + 2\xi^{r+1} d\xi^{r+1} - 2\tau^1 d\tau^1 - \cdots - 2\tau^s,$$
$$dg = 2\xi^1 d\xi^1 + \cdots + 2\xi^r d\xi^r - 2\tau^1 d\tau^1 - \cdots - 2\tau^{s+1}.$$

Therefore from the representation of the metrics, we have

$$\operatorname{grad} f = 2\xi^1 \frac{\partial}{\partial \xi^1} + \cdots + 2\xi^{r+1} \frac{\partial}{\partial \xi^{r+1}} + 2\tau^1 \frac{\partial}{\partial \tau^1} + \cdots + 2\tau^s \frac{\partial}{\partial \tau^s},$$
$$\operatorname{grad} g = 2\xi^1 \frac{\partial}{\partial \xi^1} + \cdots + 2\xi^r \frac{\partial}{\partial \xi^r} + 2\tau^1 \frac{\partial}{\partial \tau^1} + \cdots + 2\tau^{s+1} \frac{\partial}{\partial \tau^{s+1}},$$

With this, we can compute that

$$\bar{q}^{r+1,s}(\operatorname{grad} f, \operatorname{grad} f) = 4[(\xi^1)^2 + \cdots + (\xi^{r+1})^2 - (\tau^1)^2 - \cdots - (\tau^s)^2] = 4R^2,$$
$$\bar{q}^{r,s+1}(\operatorname{grad} g, \operatorname{grad} g) = 4[(\xi^1)^2 + \cdots + (\xi^r)^2 - (\tau^1)^2 - \cdots - (\tau^{s+1})^2] = -4R^2,$$

pseudo Riemann submani

Therefore by Proposition **??** we know that $S^{r,s}(R)$ and $\mathbb{H}^{r,s}(R)$ both have signature (r, s). □

For pseudo-Riemannian manifolds, though, it is necessary to modify the definition of frame homogeneity slightly. If (M, g) is a pseudo-Riemannian manifold of signature (r, s), let us say that an orthonormal basis for some tangent space T_pM is in **standard order** if the expression for g_p in terms of the dual basis (ε^i) is $(\varepsilon^1)^2 + \cdots + (\varepsilon^r)^2 - (\varepsilon^{r+1})^2 - \cdots - (\varepsilon^{r+s})^2$, with all positive terms coming before the negative terms. With this understanding, we define $\mathrm{O}(M)$ to be the set of all standard-ordered orthonormal bases for all tangent spaces to M, and we say that (M, g) is frame-homogeneous if the isometry group acts transitively on $\mathrm{O}(M)$.

Proposition 1.1.19. *All pseudo-Euclidean spaces, pseudospheres, and pseudohyperbolic spaces are frame-homogeneous.*

hyperbolic frame homogeneous

Proof. This is similar to Proposition 1.1.8. □

In the case of signature $(n, 1)$, the Lorentz manifolds $(S^{n,1}(R), \mathring{q}^{(n,1)})$ and $(\mathbb{H}^{n,1}, \breve{q}_R^{(n,1)})$ are called **de Sitter space of radius R** and **anti-de Sitter space of radius R**, respectively.

1.2 Connections

1.2.1 Differentiating Vector Fields in $\mathbb{R}^n$

To see why we need a new kind of differentiation operator, let us begin by thinking informally about curves in $\mathbb{R}^n$. Let $I \subseteq \mathbb{R}$ be an interval and $\gamma : I \to \mathbb{R}^n$ a smooth curve, written in standard coordinates as $\gamma(t) = (\gamma^1(t), \dots, \gamma^n(t))$. Such a curve has a well-defined velocity $\gamma'(t)$ and acceleration $\gamma''(t)$ at each $t \in I$, computed by differentiating the components:

$$\gamma'(t) = \dot{\gamma}^1(t)\frac{\partial}{\partial x^1}\Big|_{\gamma(t)} + \cdots + \dot{\gamma}^n(t)\frac{\partial}{\partial x^n}\Big|_{\gamma(t)}, \tag{2.1}$$ Euclidean cur

$$\gamma''(t) = \dot{\gamma}^1(t)\frac{\partial}{\partial x^1}\Big|_{\gamma(t)} + \cdots + \dot{\gamma}^n(t)\frac{\partial}{\partial x^n}\Big|_{\gamma(t)}. \tag{2.2}$$ Euclidean cur

A curve γ in $\mathbb{R}^n$ is a straight line if and only if it has a parametrization for which $\gamma''(t) = 0$.

We can also make sense of directional derivatives of vector fields on $\mathbb{R}^n$, just by computing ordinary directional derivatives of the component functions in standard coordinates: Let $Y \in \mathfrak{X}(\mathbb{R}^n)$ be a vector field and $v = v^i\partial_i \in T_p\mathbb{R}^n$ be any vector. We can define the **Euclidean directional derivative of Y in the direction v** by the formula

$$\overline{\nabla}_v Y = v^i \frac{\partial Y^j}{\partial x^i}(p)\frac{\partial}{\partial x^j}\Big|_p = v(Y^j)\frac{\partial}{\partial x^j}\Big|_p. \tag{2.3}$$ Euclidean dir

If $X \in \mathfrak{X}(\mathbb{R}^n)$ is another vector field, we then obtain a new vector field $\overline{\nabla}_X Y$ by evaluating $\overline{\nabla}_{X_p} Y$ at each point

$$\overline{\nabla}_X Y = X(Y^j)\frac{\partial}{\partial x^j}.$$

More generally, we can play the same game on submanifolds of $\mathbb{R}^n$. Suppose $M \subseteq \mathbb{R}^n$ is an embedded submanifold, and consider a smooth curve $\gamma : I \to M$. We want to think of a **geodesic** in M as a curve in M that is "as straight as possible". Of course, if M itself is curved, then $\gamma'(t)$ (thought of as a vector in $\mathbb{R}^n$) will probably have to vary, or else the curve will leave M. But we can try to insist that the velocity not change any more than necessary for the curve to stay in M. One way to do this is to compute the Euclidean acceleration $\gamma''(t)$ as above, and then apply the tangential projection $\pi^\top : T_{\gamma(t)}\mathbb{R}^n \to T_{\gamma(t)}M$ (see Prop. **??** Riemann normal bundle). This yields a vector $\gamma''(t)^\top = \pi^\top(\gamma''(t))$ tangent to M, which we call the **tangential acceleration of** γ. It is reasonable to say that γ is as straight as it is possible for a curve in M to be if its tangential acceleration is zero.

Similarly, suppose Y is a smooth vector field on (an open subset of) M, and we wish to ask how much Y is varying in M in the direction of a vector $v \in T_pM$. Just as in the case of velocity vectors, if we look at it from the point of view of $\mathbb{R}^n$, the vector field Y might be forced to vary just so that it can remain tangent to M. But one plausible way to answer the question is to extend Y to a smooth vector field $\widetilde{Y}$ on an open subset of $\mathbb{R}^n$, compute the Euclidean directional derivative of $\widetilde{Y}$ in the direction v, and then project orthogonally onto T_pM. Let us define the tangential directional derivative of Y in the direction v to be

$$\nabla_v^\top Y = \pi^\top(\overline{\nabla}_v \widetilde{Y}). \tag{2.4}$$ Euclidean tan

If $\widetilde{Y}_1$ and $\widetilde{Y}_2$ are two extensions of Y on some open neighborhood of M, then by Proposition **??** tangent space subm we have $v(\widetilde{Y}_1 - \widetilde{Y}_2) = 0$, therefore $\nabla_v^\top Y$ is well-defined. Moreover, it can be easily seen that

$\nabla_v^\top$ is preserved by rigid motions of $\mathbb{R}^n$, in the sense that if $F \in \mathrm{E}(n)$ then

$$dF_p(\nabla_v^\top Y) = \nabla_{dF_p(v)}(F_* Y).$$

However, at this point there is little reason to believe that the tangential directional derivative is an intrinsic invariant of M (one that depends only on the Riemannian geometry of M with its induced metric).

On an abstract Riemannian manifold, for which there is no ambient Euclidean space in which to differentiate, this technique is not available. Thus we have to find some way to make sense of the acceleration of a smooth curve in an abstract manifold. Let $\gamma : I \to M$ be such a curve. As you know from your study of smooth manifold theory, at each time $t \in I$, the velocity of γ is a well-defined vector $\gamma'(t) \in T_{\gamma(t)}M$ whose representation in any coordinates is given by (2.1), just as in Euclidean space.

However, unlike velocity, acceleration has no such coordinate-independent interpretation. For example, consider the parametrized circle in the plane given in Cartesian coordinates by $\gamma(t) = (x(t), y(t)) = (\cos t, \sin t)$. As a smooth curve in $\mathbb{R}^2$, it has an acceleration vector at time t given by

$$\gamma''(t) = x''(t)\frac{\partial}{\partial x}\Big|_{\gamma(t)} + y''(t)\frac{\partial}{\partial y}\Big|_{\gamma(t)} = -\cos t\frac{\partial}{\partial x}\Big|_{\gamma(t)} - \sin t\frac{\partial}{\partial y}\Big|_{\gamma(t)}.$$

But in polar coordinates, the same curve is described by $(r(t), \theta(t)) = (1, t)$. In these coordinates, if we try to compute the acceleration vector by the analogous formula, we get

$$\gamma''(t) = r''(t)\frac{\partial}{\partial x}\Big|_{\gamma(t)} + \theta''(t)\frac{\partial}{\partial y}\Big|_{\gamma(t)} = 0.$$

The problem is this: to define $\gamma''(t)$ by differentiating $\gamma'(t)$ with respect to t, we have to take a limit of a difference quotient involving the vectors $\gamma'(t+h)$ and $\gamma'(t)$ but these live in different vector spaces ($T_{\gamma(t+h)}M$ and $T_{\gamma(t)}M$ respectively), so it does not make sense to subtract them. The definition of acceleration works in the special case of smooth curves in $\mathbb{R}^n$ expressed in standard coordinates (or more generally, curves in any finite-dimensional vector space expressed in linear coordinates) because each tangent space can be naturally identified with the vector space itself. On a general smooth manifold, there is no such natural identification (The different result we get above comes from two ways to identify the tangent space).

The velocity vector $\gamma'(t)$ is an example of a vector field along a curve. To interpret the acceleration of a curve in a manifold, what we need is some coordinate-independent way to differentiate vector fields along curves. To do so, we need a way to compare values of the vector field at different points, or intuitively, to "connect" nearby tangent spaces. This is where a connection comes in: it will be an additional piece of data on a manifold, a rule for computing directional derivatives of vector fields.

1.2.2 Connections

It turns out to be easiest to define a connection first as a way of differentiating sections of vector bundles. The definition is meant to capture the essential properties of the Euclidean and tangential directional derivative operators ($\overline{\nabla}$ and $\nabla^\top$) that we defined above.

Let $\pi : E \to M$ be a smooth vector bundle over a smooth manifold M with or without boundary, and let $\pi(E)$ denote the space of smooth sections of E. A **connection in** E is a map

$$\nabla : \mathfrak{X}(M) \times \Gamma(E) \to \Gamma(E)$$

written $(X, Y) \mapsto \nabla_X Y$, satisfying the following properties:

(i) $\nabla_X Y$ is linear over $C^\infty(M)$ in X: for $f_1, f_2 \in C^\infty(M)$ and $X_1, X_2 \in \mathfrak{X}(M)$,

$$\nabla_{f_1 X_1 + f_2 X_2} Y = f_1 \nabla_{X_1} Y + f_2 \nabla_{X_2} Y.$$

(ii) $\nabla_X Y$ is linear over $\mathbb{R}$ in Y: for $a_1, a_2 \in \mathbb{R}$ and $Y_1, Y_2 \in \Gamma(E)$,

$$\nabla_X (a_1 Y_1 + a_2 Y_2) = a_1 \nabla_X Y_1 + a_2 \nabla_X Y_2.$$

(iii) ∇ satisfies the following product rule: for $f \in C^\infty(M)$,

$$\nabla_X (fY) = (Xf)Y + f \nabla_X Y.$$

The operator $\nabla_X Y$ is called the **covariant derivative** of Y in the direction X.

Although a connection is defined by its action on global sections, it follows from the definitions that it is actually a local operator, as the next lemma shows.

nection local **Lemma 1.2.1.** *Suppose ∇ is a connection in a smooth vector bundle $E \to M$. For every $X \in \mathfrak{X}(M)$, $Y \in \Gamma(E)$, and $p \in M$, the covariant derivative $\nabla_X Y|_p$ depends only on the values of X and Y in an arbitrarily small neighborhood of p. More precisely, if $X = \widetilde{X}$ and $Y = \widetilde{Y}$ on a neighborhood of p, then $\nabla_X Y|_p = \nabla_{\widetilde{X}} \widetilde{Y}$.*

Proof. First consider Y. Replacing Y by $Y - \widetilde{Y}$ shows that it suffices to prove $\nabla_X Y|_p = 0$ if Y vanishes on a neighborhood of p.

Thus suppose Y is a smooth section of E that is identically zero on a neighborhood U of p. Choose a bump function $\varphi \in C^\infty(M)$ with support in U such that $\varphi(p) = 1$. The hypothesis that Y vanishes on U implies that $\varphi Y \equiv 0$ on all of M, so for every $X \in \mathfrak{X}(M)$, we have $\nabla_X(\varphi Y) = \nabla_X(0 \cdot \varphi Y) = 0$. Thus the product rule gives

$$0 = \nabla_X(\varphi Y) = (X\varphi)Y + \varphi(\nabla_X Y). \tag{2.5}$$ connection lo

Now $Y \equiv 0$ on the support of φ, so the first term on the right is identically zero. Evaluating connection local-1 (2.5) at p shows that $\nabla_X Y|_p = 0$. The argument for X is similar but easier, using $\nabla_{\varphi X} Y = \varphi \nabla_X Y$. □

n restriction **Proposition 1.2.2 (Restriction of a Connection).** *Suppose ∇ is a connection in a smooth vector bundle $E \to M$. For every open subset $U \subseteq M$, there is a unique connection ∇^U on the restricted bundle $E|_U$ that satisfies the following relation for every $X \in \mathfrak{X}(M)$ and $Y \in \Gamma(E)$:*

$$\nabla^U_{X|_U}(Y|_U) = (\nabla_X Y)|_U. \tag{2.6}$$ connection re

Proof. First we prove uniqueness. Suppose ∇^U is any such connection and $X \in \mathfrak{X}(U)$ and $Y \in \Gamma(E|_U)$ are arbitrary. Given $p \in U$, we can use a bump function to construct a smooth vector field $\widetilde{X} \in \mathfrak{X}(M)$ and a smooth section $\widetilde{Y}$ such that $\widetilde{X}|_U$ agrees with X and $\widetilde{Y}|_U$ with Y on some neighborhood of p, and then Lemma connection local 1.2.1 together with connection restriction-1 (2.6) implies

$$\nabla^U_X Y|_p = \nabla^U_{\widetilde{X}|_U}(\widetilde{Y}|_U)|_p = \nabla^U_{\widetilde{X}}(\widetilde{Y})|_p. \tag{2.7}$$ connection re

Since the right-hand side is completely determined by ∇, this shows that ∇^U is uniquely defined if it exists.

To prove existence, given $X \in \mathfrak{X}(U)$ and $Y \in \Gamma(E|_U)$, for every $p \in U$ we just construct $\widetilde{X}$ and $\widetilde{Y}$ as above, and define $\nabla^U_X Y|_p$ by (2.7). This is independent of the choices of $\widetilde{X}$ and $\widetilde{Y}$ by Lemma 1.2.1, and it is smooth because the same formula holds on some neighborhood of p. The fact that it satisfies the properties of a connection follows from that ∇ is a connection. □

In the situation of this proposition, we typically just refer to the restricted connection as ∇ instead of ∇^U; the proposition guarantees that there is no ambiguity in doing so.

Lemma 1.2.1 tells us that we can compute the value of $\nabla_X Y$ at p knowing only the values of X and Y in a neighborhood of p. In fact, as the next proposition shows, we need only know the value of X at p itself.

Proposition 1.2.3. *Under the hypotheses of Lemma 1.2.1, $\nabla_X Y|_p$ depends only on the values of Y in a neighborhood of p and the value of X at p.*

Proof. The claim about Y was proved in Lemma 1.2.1. To prove the claim about X, it suffices by linearity to assume that $X_p = 0$ and show that $\nabla_X Y = 0$. Choose a coordinate neighborhood U of p, and write $X = X^i \partial_i$ in coordinates on U, with $X_i(p) = 0$. Thanks to Proposition 1.2.2, it suffices to work with the restricted connection on U, which we also denote by ∇. For every $Y \in \Gamma(E|_U)$, we have

$$\nabla_X Y|_p = \nabla_{X^i \partial_i} Y|_p = X^i(p) \nabla_{\partial_i} Y|_p = 0.$$

This gives the claim. □

Thanks to Propositions 1.2.2 and 1.2.3, we can make sense of the expression $\nabla_v Y$ when v is some element of T_pM and Y is a smooth local section of E defined only on some neighborhood of p. To evaluate it, let X be a vector field on a neighborhood of p whose value at p is v, and set $\nabla_v Y = \nabla_X Y|_p$. Proposition 1.2.3 shows that the result does not depend on the extension chosen. Henceforth, we will interpret covariant derivatives of local sections of bundles in this way without further comment.

1.2.3 Connections in the tangent bundle

For Riemannian or pseudo-Riemannian geometry, our primary concern is with connections in the tangent bundle, so for the rest of the section we focus primarily on that case. A connection in the tangent bundle is often called simply a **connection on M**. (The terms **affine connection** and **linear connection** are also sometimes used in this context, but there is little agreement on the precise definitions of these terms, so we avoid them.)

Suppose M is a smooth manifold with or without boundary. By the definition we just gave, a connection in TM is a map

$$\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$$

satisfying properties (i)(iii) above. For computations, we need to examine how a connection appears in terms of a local frame. Let (E_i) be a smooth local frame for TM on an open subset $U \subseteq M$. For every choice of the indices i and j, we can expand the vector field $\nabla_{E_i} E_j$ in terms of this same frame:

$$\nabla_{E_i} E_j = \Gamma^k_{ij} E_k.$$

As i, j, and k range from 1 to $n = \dim M$, this defines n^3 smooth functions $\Gamma^k_{ij} : U \to \mathbb{R}$, called the **connection coefficients** of ∇ with respect to the given frame. The following proposition shows that the connection is completely determined in U by its connection coefficients.

Proposition 1.2.4. *Let M be a smooth manifold with or without boundary, and let ∇ be a connection in TM. Suppose (E_i) is a smooth local frame over an open subset $U \subseteq M$, and let $\{\Gamma^k_{ij}\}$ be the connection coefficients of ∇ with respect to this frame. For smooth vector fields $X, Y \in \mathfrak{X}(U)$, written in terms of the frame as $X = X^i E_i$, $Y = Y^j E_j$, one has*

$$\nabla_X Y = \big(X(Y^k) + X^i Y^j \Gamma^k_{ij}\big) E_k. \tag{2.8}$$

Proof. Just use the defining properties of a connection and compute:

$$\nabla_X(Y^j E_j) = X(Y^j)E_j + Y^j \nabla_{X^i E_i} E_j = X(Y^j)E_j + X^i Y^j \nabla_{E_i} E_j = X(Y^j)E_j + X^i Y^j \Gamma^k_{ij} E_k.$$

Renaming the dummy index in the first term yields (2.8). □

Once the connection coefficients (and thus the connection) have been determined in some local frame, they can be determined in any other local frame on the same open set by the result of the following proposition.

Proposition 1.2.5 (Transformation Law for Connection Coefficients). *Let M be a smooth manifold with or without boundary, and let ∇ be a connection in TM. Suppose we are given two smooth local frames (E_i) and $(\widetilde{E}_j)$ for TM on an open subset $U \subseteq M$, related by $\widetilde{E}_i = A^j_i E_j$ for some matrix of functions A^j_i. Let $\{\Gamma^k_{ij}\}$ and $\{\widetilde{\Gamma}^k_{ij}\}$ denote the connection coefficients of ∇ with respect to these two frames. Then*

$$\widetilde{\Gamma}^k_{ij} = A^p_i E_p(A^q_j)(A^{-1})^k_q + A^p_i A^q_j \Gamma^r_{pq}(A^{-1})^k_r. \tag{2.9}$$

Proof. Write $\widetilde{E}_i = A^p_i E_p$ and $\widetilde{E}_j = A^q_j E_q$, then by (2.8),

$$\nabla_{\widetilde{E}_i} \widetilde{E}_j = A^p_i E_p(A^q_j) E_q + A^p_i A^q_j \Gamma^r_{pq} E_r = A^p_i E_p(A^q_j)(A^{-1})^k_q \widetilde{E}_k + A^p_i A^q_j \Gamma^r_{pq}(A^{-1})^k_r \widetilde{E}_k.$$

Therefore

$$\widetilde{\Gamma}^k_{ij} = A^p_i E_p(A^q_j)(A^{-1})^k_q + A^p_i A^q_j \Gamma^r_{pq}(A^{-1})^k_r,$$

as claimed. □

Observe that the second term above is exactly what the transformation law would be if Γ^k_{ij} were the components of a $(1,2)$-tensor field; but the first term is of a different character, because it involves derivatives of the transition matrix.

1.2.4 Existence of connections

So far, we have studied properties of connections but have not produced any, so you might be wondering whether they are plentiful or rare. In fact, they are quite plentiful, as we will show shortly. Let us begin with the simplest example.

Example 1.2.6 (The Euclidean Connection). In $T\mathbb{R}^n$, define the Euclidean connection $\overline{\nabla}$ by formula (2.3). It is easy to check that this satisfies the required properties for a connection, and that its connection coefficients in the standard coordinate frame are all zero.

Here is a way to construct a large class of examples.

Example 1.2.7 (**The Tangential Connection on a Submanifold of $\mathbb{R}^n$**)**.** Let $M \subseteq \mathbb{R}^n$ be an embedded submanifold. Define a connection $\nabla^\top$ on TM, called the **tangential connection**, by setting

$$\nabla_X Y = \pi^\top(\overline{\nabla}_{\widetilde{X}} \widetilde{Y}|_M),$$

where $\pi^\top$ is the orthogonal projection onto TM, $\overline{\nabla}$ is the Euclidean connection, and $\widetilde{X}$, $\widetilde{Y}$ extensions of X and Y to an open set in $\mathbb{R}^n$ (Such extensions exist by Proposition **??**). Since the value of $\overline{\nabla}_{\widetilde{X}} \widetilde{Y}$ at a point $p \in M$ depends only on $\widetilde{X}_p = X_p$, this just boils down to defining $(\nabla^\top_X Y)|_p$ to be equal to the tangential directional derivative $\nabla^\top_{X_p} Y$ that we defined in (2.4) above. By Proposition **??** this value is independent of the choice of extension $\widetilde{Y}$, so $\nabla^\top$ is well defined. Smoothness is easily verified by expressing $\nabla_{\widetilde{X}} \widetilde{Y}$ in terms of an adapted orthonormal frame.

It is immediate from the definition that $\nabla^\top$ is linear over $C^\infty(M)$ in X and over $\mathbb{R}$ in Y, so to show that $\nabla^\top$ is a connection, only the product rule needs to be checked. Let $f \in C^\infty(M)$, and let $\widetilde{f}$ be an extension of f to a neighborhood of M in $\mathbb{R}^n$. Then $\widetilde{f}\widetilde{Y}$ is a smooth extension of fY to a neighborhood of M, so

$$\begin{aligned}
\nabla^\top_X(fY) &= \pi^\top(\overline{\nabla}_{\widetilde{X}}(\widetilde{f}\widetilde{Y})|_M) \\
&= \pi^\top((\widetilde{X}\widetilde{f})\widetilde{Y}|_M) + \pi^\top(\widetilde{f}\overline{\nabla}_{\widetilde{X}}\widetilde{Y}|_M) \\
&= (Xf)\nabla^\top_X Y + f\nabla^\top_X Y.
\end{aligned}$$

Thus $\nabla^\top$ is a connection.

In fact, there are many connections on $\mathbb{R}^n$, or indeed on every smooth manifold that admits a global frame (for example, every manifold covered by a single smooth coordinate chart). The following lemma shows how to construct all of them explicitly.

Lemma 1.2.8. *Suppose M is a smooth n-manifold with or without boundary, and M admits a global frame (E_i). Formula (1.2.4) gives a one-to-one correspondence between connections in TM and choices of n^3 smooth real-valued functions $\{\Gamma^k_{ij}\}$ on M.*

Proof. Every connection determines functions $\{\Gamma^k_{ij}\}$ by $\nabla_{E_i} E_j = \Gamma^k_{ij} E_k$, and Proposition 1.2.4 shows that those functions satisfy (2.8). On the other hand, given $\{\Gamma^k_{ij}\}$, we can define $\nabla_X Y$ by (2.8). It is easy to see that the resulting expression is smooth if X and Y are smooth, linear over $\mathbb{R}$ in Y, and linear over $C^\infty(M)$ in X. To prove that it is a connection, only the product rule requires checking; this is a straightforward computation. □

Proposition 1.2.9. *The tangent bundle of every smooth manifold with or without boundary admits a connection.*

Proof. Let M be a smooth manifold with or without boundary, and cover M with coordinate charts $\{U_\alpha\}$ the preceding lemma guarantees the existence of a connection ∇^α on each U_α. Choose a partition of unity $\{\psi_\alpha\}$ subordinate to $\{U_\alpha\}$. We would like to patch the various ∇^α's together by the formula

$$\nabla_X Y = \sum_\alpha \psi_\alpha \nabla^\alpha_X Y.$$

Because the set of supports of the ψ_α's is locally finite, the sum on the right-hand side has only finitely many nonzero terms in a neighborhood of each point, so it defines a smooth vector field on M. It is immediate from this definition that $\nabla_X Y$ is linear over $\mathbb{R}$ in Y and linear over $C^\infty(M)$ in X. We have to be a bit careful with the product rule, though, since a linear combination of connections is not necessarily a connection (In fact, only a convex linear combination of connections is again a connection). By direct computation,

$$\begin{aligned}\nabla_X(fY) &= \sum_\alpha \psi_\alpha \nabla^\alpha_X(fY) = \sum_\alpha \psi_\alpha((Xf)Y + f\nabla^\alpha_X Y)\\ &= (Xf)Y\sum_\alpha \psi_\alpha + f\sum_\alpha \psi_\alpha \nabla^\alpha_X Y\\ &= (Xf)Y + f\nabla_X Y.\end{aligned}$$

This finishes the proof. □

Although a connection is not a tensor field, the next proposition shows that the difference between two connections is.

erence tensor **Proposition 1.2.10 (The Difference Tensor).** *Let M be a smooth manifold with or without boundary. For any two connections ∇^0 and ∇^1 in TM, define a map $D(X,Y) : \mathfrak{X}(M)\times\mathfrak{X}(M)\to\mathfrak{X}(M)$ by*

$$D(X,Y) = \nabla^1_X Y - \nabla^0_X Y.$$

*Then D is bilinear over $C^\infty(M)$, and thus defines a $(1,2)$-tensor field called the **difference tensor** between ∇^0 and ∇^1.*

Proof. It is immediate from the definition that D is linear over $C^\infty(M)$ in its first argument, because both ∇^0 and ∇^1 are. To show that it is linear over $C^\infty(M)$ in the second argument, expand $D(X,Y)$ using the product rule, and note that the two terms in which f is differentiated cancel each other. □

Now that we know there is always one connection in TM, we can use the result of the preceding proposition to say exactly how many there are.

affine space **Proposition 1.2.11.** *Let M be a smooth manifold with or without boundary, and let ∇^0 be any connection in TM. Then the set $\mathcal{A}(TM)$ of all connections in TM is equal to the following affine space:*

$$\mathcal{A}(TM) = \{\nabla^0 + D : D \in \Gamma(T^{(1,2)}TM)\}.$$

where $D \in \Gamma(T^{(1,2)}TM)$ is interpreted as a map from $\mathfrak{X}(M)\times\mathfrak{X}(M)$ to $\mathfrak{X}(M)$ as in Proposition **??**, *and $\nabla^0 + D$ is defined by* tensor end iso

$$(\nabla^0 + D)_X Y = \nabla^0_X Y + D(X,Y).$$

1.2.5 Covariant derivatives of tensor fields

By definition, a connection in TM is a rule for computing covariant derivatives of vector fields. We now show that every connection in TM automatically induces connections in all tensor bundles over M, and thus gives us a way to compute covariant derivatives of tensor fields of any type.

def on tensor **Proposition 1.2.12.** *Let M be a smooth manifold with or without boundary, and let ∇ be a connection in TM. Then ∇ uniquely determines a connection in each tensor bundle $T^{(k,l)}TM$, also denoted by ∇, such that the following four conditions are satisfied.*

(i) *In* $T^{(1,0)}TM = TM$, ∇ *agrees with the given connection.*

(ii) *In* $T^{(0,0)}TM = C^\infty(M)$, ∇ *is given by ordinary differentiation of functions:*

$$\nabla_X f = Xf.$$

(iii) ∇ *obeys the following product rule with respect to tensor products:*

$$\nabla_X(F \otimes G) = (\nabla_X F) \otimes G + F \otimes (\nabla_X G).$$

(iv) ∇ *commutes with all contractions: if* tr *denotes a trace on any pair of indices, one covariant and one contravariant, then*

$$\nabla_X(\mathrm{tr}F) = \mathrm{tr}(\nabla_X F).$$

This connection also satisfies the following additional properties:

(a) ∇ *obeys the following product rule with respect to the natural pairing between a covector field* ω *and a vector field* Y*:*

$$\nabla_X\langle\omega, Y\rangle = \langle\nabla_X\omega, Y\rangle + \langle\omega, \nabla_X Y\rangle.$$

(b) *For all* $F \in \Gamma(T^{(k,l)}TM)$*, smooth* 1*-forms* $\omega^1, \ldots, \omega^k$ *and smooth vector fields* $Y_1, \ldots, Y_l$*,*

$$\begin{aligned}(\nabla_X F)(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l) &= X\big(F(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l)\big) \\ - \sum_{i=1}^{k} F(\omega^1, \ldots, \nabla_X\omega^i, \ldots, \omega^k, Y_1, \ldots, Y_l) &- \sum_{j=1}^{l} F(\omega^1, \ldots, \omega^k, Y_1, \ldots, \nabla_X Y_j, \ldots, Y_l).\end{aligned} \tag{2.10}$$

connection de

Proof. First we showthat every family of connections on all tensor bundles satisfying (i)–(iv) also satisfies (a) and (b). Suppose we are given such a family of connections, all denoted by ∇. To prove (a), note that $\langle\omega, Y\rangle = \mathrm{tr}(\omega \otimes Y)$, as can be seen by evaluating both sides in coordinates, where they both reduce to $\omega_i Y^i$. Therefore, (i)–(iv) imply

$$\begin{aligned}\nabla_X\langle\omega, Y\rangle &= \nabla_X\big(\mathrm{tr}(\omega \otimes Y)\big) = \mathrm{tr}\big(\nabla_X(\omega \otimes Y)\big) \\ &= \mathrm{tr}(\nabla_X\omega \otimes Y + \omega \otimes \nabla_X Y) \\ &= \langle\nabla_X\omega, Y\rangle + \langle\omega, \nabla_X Y\rangle.\end{aligned}$$

Then (b) is proved by induction using a similar computation applied to

$$F(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l) = \mathrm{tr}^{k+l}(F \otimes \omega^1 \cdots \otimes \omega^k \otimes Y_1 \otimes \cdots \otimes Y_l).$$

where each trace operator acts on an upper index of F and the lower index of the corresponding 1-form, or a lower index of F and the upper index of the corresponding vector field.

Next we address uniqueness. Assume again that ∇ represents a family of connections satisfying (i)(iv), and hence also (a) and (b). Observe that (ii) and (a) imply that the covariant derivative of every 1-form ω can be computed by

$$(\nabla_X\omega)(Y) = X(\omega(Y)) - \omega(\nabla_X Y). \tag{2.11}$$

connection de

It follows that the connection on 1-forms is uniquely determined by the original connection in TM. Similarly, (b) gives a formula that determines the covariant derivative of every tensor field F in terms of covariant derivatives of vector fields and 1-forms, so the connection in every tensor bundle is uniquely determined.

Now to prove existence, we first define covariant derivatives of 1-forms by (2.11), and then we use (2.10) to define ∇ on all other tensor bundles. The first thing that needs to be checked is that the resulting expression is multilinear over $C^\infty(M)$ in each ω^i and Y_j, and therefore defines a smooth tensor field. This is done by inserting $f\omega^i$ in place of ω^i, or fY_j in place of Y_j, and expanding the right-hand side, noting that the two terms in which f is differentiated cancel each other out. Once we know that $\nabla_X F$ is a smooth tensor field, we need to check that it satisfies the defining properties of a connection. Linearity over $C^\infty(M)$ in X and linearity over $\mathbb{R}$ in F are both evident from (2.10) and (2.11), and the product rule in F follows easily from the fact that differentiation of functions by X satisfies the product rule. It is then a straightforward computation to show that the resulting connection satisfies conditions (i)–(v). □

Proposition 1.2.13. *Let M be a smooth manifold with or without boundary, and let ∇ be a connection in TM. Suppose (E_i) is a local frame for M, (ε^j) is its dual coframe, and $\{\Gamma_{ij}^k\}$ are the connection coefficients of ∇ with respect to this frame. Let X be a smooth vector field, and let X^iE_i be its local expression in terms of this frame.*

(a) The covariant derivative of a 1-form $\omega = \omega_i\varepsilon^i$ is given locally by

$$\nabla_X\omega = \big(X(\omega_k) - X^j\Gamma_{jk}^i\omega_i\big)\varepsilon^k.$$

(b) If $F \in \Gamma(T^{(k,l)}TM)$ is a smooth mixed tensor field of any rank, expressed locally as

$$F = F_{j_1\dots j_l}^{i_1\dots i_k} E_{i_1}\otimes\cdots\otimes E_{i_k}\otimes\varepsilon^{j_1}\otimes\cdots\otimes\varepsilon^{j_l},$$

then the covariant derivative of F is given locally by

$$(\nabla_X F)_{j_1\dots j_l}^{i_1\dots i_k} = X(F_{j_1\dots j_l}^{i_1\dots i_k}) + \sum_{s=1}^{k} X^p\Gamma_{pq}^{i_s}F_{j_1\dots j_l}^{i_1\dots q\dots i_k} - \sum_{r=1}^{l} X^p\Gamma_{pj_r}^{q}F_{j_1\dots q\dots j_l}^{i_1\dots i_k}.$$

Proof. First we compute that

$$\begin{aligned}(\nabla_X\omega)(E_k) &= X(\omega(E_k)) - \omega(\nabla_X E_k) = X(\omega_k) - \omega(X^j\nabla_{E_j}E_k)\\ &= X(\omega_k) - \omega(X^j\Gamma_{jk}^iE_i) = X(\omega_k) - \omega_iX^j\Gamma_{jk}^i.\end{aligned}$$

Therefore the first equatin follows. For (b), first we use (a) to obtain $\nabla_X\varepsilon^{i_s} = -X^p\Gamma_{pq}^{i_s}\varepsilon^q$, and so we have

$$\begin{aligned}&(\nabla_X F)_{j_1\dots j_l}^{i_1\dots i_k} = (\nabla_X F)(\varepsilon^{i_1},\dots,\varepsilon^{i_k},E_{j_1},\dots,E_{j_l})\\ &= X\big(F(\varepsilon^{i_1},\dots,\varepsilon^{i_k},E_{j_1},\dots,E_{j_l})\big)\\ &-\sum_{s=1}^{k}F(\varepsilon^{i_1},\dots,\nabla_X\varepsilon^{i_s},\dots,\varepsilon^{i_k},E_{j_1},\dots,E_{j_l}) - \sum_{r=1}^{l}F(\varepsilon^{i_1},\dots,\varepsilon^{i_k},E_{j_1},\dots,\nabla_XE_{j_r},\dots,E_{j_l})\\ &= X(F_{j_1\dots j_l}^{i_1\dots i_k}) + \sum_{s=1}^{k}X^p\Gamma_{pq}^{i_s}F(\varepsilon^{i_1},\dots,\varepsilon^q,\dots,\varepsilon^{i_k},E_{j_1},\dots,E_{j_l})\\ &-\sum_{r=1}^{l}X^p\Gamma_{pj_r}^{q}F(\varepsilon^{i_1},\dots,\varepsilon^{i_k},E_{j_1},\dots,E_q,\dots,E_{j_l})\\ &= X(F_{j_1\dots j_l}^{i_1\dots i_k}) + \sum_{s=1}^{k}X^p\Gamma_{pq}^{i_s}F_{j_1\dots j_l}^{i_1\dots q\dots i_k} - \sum_{r=1}^{l}X^p\Gamma_{pj_r}^{q}F_{j_1\dots q\dots j_l}^{i_1\dots i_k}.\end{aligned}$$

Therefore the claim follows. □

Because the covariant derivative $\nabla_X F$ of a tensor field (or, as a special case, a vector field) is linear over $C^\infty(M)$ in X, the covariant derivatives of F in all directions can be handily encoded in a single tensor field whose rank is one more than the rank of F, as follows.

Proposition 1.2.14 (The Total Covariant Derivative). *Let M be a smooth manifold with or without boundary and let ∇ be a connection in TM. For every $F \in \Gamma(T^{(k,l)}TM)$, the map*

$$\nabla : \underbrace{\Omega^1(M) \times \cdots \times \Omega^1(M)}_{k \text{ folds}} \times \underbrace{\mathfrak{X}(M) \times \cdots \times \mathfrak{X}(M)}_{l+1 \text{ folds}} \to C^\infty(M)$$

given by

$$(\nabla F)(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l, X) = (\nabla_X F)(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l)$$

*defines a smooth $(k, l+1)$-tensor field on M called the **total covariant derivative** of F.*

tensor char

Proof. This follows immediately from the tensor characterization lemma (Lemma **??**): $\nabla_X F$ is a tensor field, so it is multilinear over $C^\infty(M)$ in its $k+l$ arguments; and it is linear over $C^\infty(M)$ in X by definition of a connection. □

When we write the components of a total covariant derivative in terms of a local frame, it is standard practice to use a semicolon to separate indices resulting from differentiation from the preceding indices. Thus, for example, if Y is a vector field written in coordinates as $Y = Y^i E_i$, the components of the $(1,1)$-tensor field ∇Y are written $Y^i_{;j}$, so that

$$\nabla Y = Y^i_{;j} E_i \otimes \varepsilon^j,$$

with

$$Y^i_{;j} = E_j Y^i + \Gamma^i_{jk} Y^k.$$

For a 1-form ω, the formulas read

$$\nabla \omega = \omega_{i;j} \varepsilon^i \otimes \varepsilon^j,$$

with

$$\omega_{i;j} = E_j \omega_i - \Gamma^k_{ij} \omega_k.$$

More generally, the next lemma gives a formula for the components of total covariant derivatives of arbitrary tensor fields.

ve expression

Proposition 1.2.15. *Let M be a smooth manifold with or without boundary and let ∇ be a connection in TM, and let (E_i) be a smooth local frame for TM and $\{\Gamma^k_{ij}\}$ the corresponding connection coefficients. The components of the total covariant derivative of a (k,l)-tensor field F with respect to this frame are given by*

$$F^{i_1 \ldots i_k}_{j_1 \ldots j_l; p} = E_p(F^{i_1 \ldots i_k}_{j_1 \ldots j_l}) + \sum_{s=1}^{k} \Gamma^{i_s}_{pq} F^{i_1 \ldots q \ldots i_k}_{j_1 \ldots j_l} - \sum_{t=1}^{l} \Gamma^q_{p j_t} F^{i_1 \ldots i_k}_{j_1 \ldots q \ldots j_l}.$$

connection tensor frame expression

Proof. By Proposition 1.2.13, the coefficients of $\nabla_{E_p} F$ is given by

$$(\nabla_{E_p} F)^{i_1 \ldots i_k}_{j_1 \ldots j_l} = E_p(F^{i_1 \ldots i_k}_{j_1 \ldots j_l}) + \sum_{s=1}^{k} \Gamma^{i_s}_{pq} F^{i_1 \ldots q \ldots i_k}_{j_1 \ldots j_l} - \sum_{t=1}^{l} \Gamma^q_{p j_t} F^{i_1 \ldots i_k}_{j_1 \ldots q \ldots j_l}.$$

This, by definition, is equal to $F^{i_1 \ldots i_k}_{j_1 \ldots j_l; p}$, thus the claim follows. □

Corollary 1.2.16. *Let M be a smooth manifold with or without boundary and let ∇ be a connection in TM, then for every $F \in \Gamma(T^{(k,l)}TM)$ and $X \in \mathfrak{X}(M)$ we have*

$$\nabla_X F = \operatorname{tr}(\nabla F \otimes X).$$

Proof. Choose a smooth local frame (E_i) with dual frame (ε^j), then by Proposition 1.2.13 and Proposition 1.2.15 we have

$$\begin{aligned}
(\nabla_X F)^{i_1\dots i_k}_{j_1\dots j_l} &= X(F^{i_1\dots i_k}_{j_1\dots j_l}) + \sum_{s=1}^{k} X^p \Gamma^{i_s}_{pq} F^{i_1\dots q\dots i_k}_{j_1\dots j_l} - \sum_{t=1}^{l} X^p \Gamma^{q}_{pj_t} F^{i_1\dots i_k}_{j_1\dots q\dots j_l} \\
&= X^p E_p(F^{i_1\dots i_k}_{j_1\dots j_l}) + \sum_{s=1}^{k} X^p \Gamma^{i_s}_{pq} F^{i_1\dots q\dots i_k}_{j_1\dots j_l} - \sum_{t=1}^{l} X^p \Gamma^{q}_{pj_t} F^{i_1\dots i_k}_{j_1\dots q\dots j_l} \\
&= X^p \Big(E_p(F^{i_1\dots i_k}_{j_1\dots j_l}) + \sum_{s=1}^{k} \Gamma^{i_s}_{pq} F^{i_1\dots q\dots i_k}_{j_1\dots j_l} - \sum_{t=1}^{l} \Gamma^{q}_{pj_t} F^{i_1\dots i_k}_{j_1\dots q\dots j_l} \Big) \\
&= X^p F^{i_1\dots i_k}_{j_1\dots j_l;p} \\
&= \operatorname{tr}(\nabla F \otimes X)^{i_1\dots i_k}_{j_1\dots j_l}.
\end{aligned}$$

Therefore the claim. □

1.2.6 Second covariant derivatives

Having defined the tensor field ∇F for a (k,l)-tensor field F, we can in turn take its total covariant derivative and obtain a $(k,l{+}2)$-tensor field $\nabla^2 F$. Given vector fields $X, Y \in \mathfrak{X}(M)$, let us introduce the notation $\nabla^2_{X,Y} F$ for the (k,l)-tensor field obtained by inserting X, Y in the last two slots of $\nabla^2 F$:

$$\nabla^2_{X,Y} F(\dots) = \nabla^2 F(\dots, Y, X)$$

Note the reversal of order of X and Y: this is necessitated by our convention that the last index position in ∇F is the one resulting from differentiation, while it is conventional to let $\nabla^2_{X,Y}$ stand for differentiating first in the Y direction, then in the X direction.

It is important to be aware that $\nabla^2_{X,Y} F$ is not the same as $\nabla_Y(\nabla_X F)$. The main reason is that the former is linear over $C^\infty(M)$ in Y, while the latter is not. The relationship between the two expressions is given in the following proposition.

Proposition 1.2.17. *Let M be a smooth manifold with or without boundary and let ∇ be a connection in TM. For every smooth vector field or tensor field F,*

$$\nabla^2_{X,Y} F = \nabla_X(\nabla_Y F) - \nabla_{\nabla_X Y} F.$$

Proof. By Corollary 1.2.16 $\nabla^2_{X,Y} F$ can be expressed as an iterated trace:

$$\nabla^2_{X,Y} F = \operatorname{tr}(\operatorname{tr}(\nabla^2 F \otimes X) \otimes Y)).$$

Therefore, since ∇_X commutes with contraction and satisfies the product rule with respect to tensor products (Prop. 1.2.12), we have

$$\begin{aligned}\nabla_X(\nabla_Y F) &= \nabla_X(\operatorname{tr}(\nabla F \otimes Y)) = \operatorname{tr}(\nabla_X(\nabla F \otimes Y)) \\ &= \operatorname{tr}(\nabla_X(\nabla F) \otimes Y) + \operatorname{tr}(\nabla F \otimes \nabla_X Y) \\ &= \operatorname{tr}(\operatorname{tr}(\nabla^2 F \otimes X) \otimes Y) + \nabla_{\nabla_X Y} F \\ &= \nabla^2_{X,Y} F + \nabla_{\nabla_X Y} F.\end{aligned}$$

Therefore the claim follows. □

Example 1.2.18 (The Covariant Hessian). Let f be a smooth function on M. Then $\nabla f \in \Gamma(T^{(0,1)}TM) = \Omega^1(M)$ is just the 1-form df, because both tensors have the same action on vectors: $\nabla f(X) = \nabla_X f = Xf = df(X)$. The 2-tensor $\nabla^2 f = \nabla(df)$ is called the covariant Hessian of f. Proposition 1.2.17 shows that its action on smooth vector fields X, Y can be computed by the following formula:

$$\nabla^2 f(Y, X) = \nabla^2_{X,Y} f = \nabla_X(\nabla_Y f) - \nabla_{\nabla_X Y} f = X(Yf) - (\nabla_X Y)f.$$

In any local coordinates, it is

$$\nabla^2 f = f_{;ij} dx^i \otimes dx^j, \quad f_{;ij} = \partial_j \partial_i - \Gamma^k_{ji} \partial_k.$$

1.2.7 Covariant derivatives along curves

Now we can address the question that originally motivated the definition of connections: How can we make sense of the derivative of a vector field along a curve? Let M be a smooth manifold with or without boundary. Given a smooth curve $\gamma : I \to M$, a **vector field along γ** is a continuous map $V : I \to TM$ such that $V(t) \in T_{\gamma(t)}M$ for every $t \in I$; it is a **smooth vector field along γ** if it is smooth as a map from I to TM. We let $\mathfrak{X}(\gamma)$ denote the set of all smooth vector fields along γ. It is a real vector space under pointwise vector addition and multiplication by constants, and it is a module over $C^\infty(M)$ with multiplication defined pointwise.

The most obvious example of a vector field along a smooth curve γ is the curve's velocity: $\gamma'(t) \in T_{\gamma(t)}M$ for each t, and its coordinate expression (2.1) shows that it is smooth. Here is another example: if γ is a curve in $\mathbb{R}^2$, let $N(t) = R\gamma'(t)$, where R is counterclockwise rotation by $-\pi/2$, so $N(t)$ is normal to $\gamma'(t)$. In standard coordinates, $N(t) = (-\dot\gamma^2(t), \dot\gamma^1(t))$, so N is a smooth vector field along γ.

A large supply of examples is provided by the following construction: suppose $\gamma : I \to M$ is a smooth curve and $\widetilde{V}$ is a smooth vector field on an open subset of M containing the image of γ. Define $V : I \to TM$ by setting $V(t) = \widetilde{V}_{\gamma(t)}$ for each $t \in I$. Since V is equal to the composition $\widetilde{V} \circ \gamma$, it is smooth. A smooth vector field along γ is said to be **extendible** if there exists a smooth vector field $\widetilde{V}$ on a neighborhood of the image of γ that is related to V in this way. Not every vector field along a curve need be extendible; for example, if $\gamma(t_1) = \gamma(t_2)$ but $\gamma'(t_1) \neq \gamma'(t_2)$, then γ' is not extendible. Even if γ is injective, its velocity need not be extendible, as the next example shows.

Example 1.2.19. Consider the figure eight curve $\gamma : (-\pi, \pi) \to \mathbb{R}^2$ defined by

$$\gamma(t) = (\sin 2t, \sin t).$$

Its image is a set that looks like a figure eight in the plane. It is easy to see that Γ is an injective smooth immersion, but its velocity vector field is not extendible.

More generally, a **tensor field along γ** is a continuous map σ from I to some tensor bundle $T^{(k,l)}TM$ such that $\sigma(t) \in T^{(k,l)}(T_{\gamma(t)}M)$ for each $t \in I$. It is a **smooth tensor field along γ** if it is smooth as a map from I to $T^{(k,l)}TM$, and it is **extendible** if there is a smooth tensor field $\widetilde{\sigma}$ on a neighborhood of $\gamma(I)$ such that $\sigma = \widetilde{\sigma} \circ \gamma$. Here is the promised interpretation of a connection as a way to take derivatives of vector fields along curves.

Theorem 1.2.20 (Covariant Derivative Along a Curve). *Let M be a smooth manifold with or without boundary and let ∇ be a connection in TM. For each smooth curve $\gamma : I \to M$, the connection determines a unique operator*

$$D_t : \mathfrak{X}(\gamma) \to \mathfrak{X}(\gamma)$$

*called the **covariant derivative along γ**, satisfying the following properties:*

(i) Linearity over $\mathbb{R}$: for $a, b \in \mathbb{R}$,

$$D_t(aV + bW) = aD_tV + bD_tW.$$

(ii) Product rule: for $f \in C^\infty(I)$,

$$D_t(fV) = f'V + fD_tV.$$

(iii) If $V \in \mathfrak{X}(\gamma)$ is extendible, then for every extension $\widetilde{V}$ of V,

$$D_t(V) = \nabla_{\gamma'(t)}\widetilde{V}.$$

There is an analogous operator on the space of smooth tensor fields of any type along γ.

Proof. For or simplicity, we prove the theorem for the case of vector fields along γ; the proof for arbitrary tensor fields is essentially identical except for notation.

First we show uniqueness. Suppose D_t is such an operator, and let $t_0 \in I$ be arbitrary. An argument similar to that of Lemma 1.2.1 shows that the value of D_tV at t_0 depends only on the values of V in any interval $(t_0 - \varepsilon, t_0 + \varepsilon)$ containing t_0. (If t_0 is an endpoint of I, extend a coordinate representation of γ to a slightly bigger open interval, prove the lemma there, and then restrict back to I.)

Choose smooth coordinates (x^i) for M in a neighborhood of $\gamma(t_0)$, and write

$$V(t) = V^j(t)\frac{\partial}{\partial x^j}\Big|_{\gamma(t)}$$

for t near t_0, where $V^1, \dots, V^n$ are smooth real-valued functions defined on some neighborhood of t_0 in I. By the properties of D_t, since each ∂_j is extendible,

$$\begin{aligned} D_tV(t) &= \dot{V}^j(t)\partial_j|_{\gamma(t)} + V^j(t)\nabla_{\gamma'(t)}\partial_j|_{\gamma(t)} \\ &= \big(\dot{V}^k(t) + \dot{\gamma}^i(t)V^j(t)\Gamma^k_{ij}(\gamma(t))\big)\partial_k|_{\gamma(t)}. \end{aligned} \tag{2.12}$$

This shows that such an operator is unique if it exists.

For existence, if $\gamma(I)$ is contained in a single chart, we can define D_tV by (2.12); it is easy verification that it satisfies the requisite properties. In the general case, we can cover $\gamma(I)$ with coordinate charts and define D_tV by this formula in each chart, and uniqueness implies that the various definitions agree whenever two or more charts overlap. $\square$

(It is worth noting that in the physics literature, the covariant derivative along a curve is sometimes called the **absolute derivative**.)

Now we can improve Proposition 1.2.3 by showing that $\nabla_v Y$ actually depends only on the values of Y along any curve through p whose velocity is v.

Proposition 1.2.21. *Let M be a smooth manifold with or without boundary, let ∇ be a connection in TM, and let $p \in M$ and $v \in T_pM$. Suppose Y and $\widetilde{Y}$ are two smooth vector fields that agree at points in the image of some smooth curve $\gamma : I \to M$ such that $\gamma(0) = p$ and $\gamma'(0) = v$. Then $\nabla_v Y = \nabla_v \widetilde{Y}$.*

Proof. We can define a smooth vector field Z along γ by $Z(t) = Y_{\gamma(t)} = \widetilde{Y}_{\gamma(t)}$. Since both Y and $\widetilde{Y}$ are extensions of Z, it follows from condition (iii) in Theorem 1.2.20 that both $\nabla_v Y$ and $\nabla_v \widetilde{Y}$ are equal to $D_t Z(t_0)$. □

1.2.8 Geodesics

Armed with the notion of covariant differentiation along curves, we can now define acceleration and geodesics.

Let M be a smooth manifold with or without boundary and let ∇ be a connection in TM. For every smooth curve $\gamma : I \to M$, we define the acceleration of γ to be the vector field $D_t\gamma'$ along γ. A smooth curve γ is called a geodesic (with respect to ∇) if its acceleration is zero: $D_t\gamma' = 0$. In terms of smooth coordinates (x^i), if we write the component functions of γ as $\gamma(t) = (x^1(t), \dots, x^n(t))$, then it follows from (2.12) that γ is a geodesic if and only if its component functions satisfy the following geodesic equation:

$$\ddot{x}^k(t) + \dot{x}^i(t)\dot{x}^j(t)\Gamma^k_{ij}(x(t)) = 0, \tag{2.13}$$

where we use $x(t)$ as an abbreviation for the n-tuple of component functions $(x^1(t), \dots, x^n(t))$. This is a system of second-order ordinary differential equations (ODEs) for the real-valued functions $x^1(t), ;x^n(t)$. The next theorem uses ODE theory to prove existence and uniqueness of geodesics with suitable initial conditions.

Theorem 1.2.22 (Existence and Uniqueness of Geodesics). *Let M be a smooth manifold and ∇ a connection in TM. For every $p \in M$, $w \in T_pM$, and $t_0 \in \mathbb{R}$, there exist an open interval $I \to \mathbb{R}$ containing t_0 and a geodesic $\gamma : I \to M$ satisfying $\gamma(t_0) = p$ and $\gamma'(t_0) = w$. Any two such geodesics agree on their common domain.*

Proof. Let (x^i) be smooth coordinates on some neighborhood U of p. A smooth curve in U, written as $\gamma(t) = (x^1(t), \dots, x^n(t))$, is a geodesic if and only if its component functions satisfy (2.13). The standard trick for proving existence and uniqueness for such a second-order system is to introduce auxiliary variables $v^i = \dot{x}^i$ to convert it to the following equivalent first-order system in twice the number of variables:

$$\begin{aligned} \dot{x}^k(t) &= v^k(t) \\ \dot{v}^k(t) &= -v^i(t)v^j(t)\Gamma^k_{ij}(x(t)). \end{aligned} \tag{2.14}$$

Treating $(x^1, \dots, x^n, v^1, \dots, v^n)$ as coordinates on $U \times \mathbb{R}^n$, we can recognize (2.14) as the equations for the flow of the vector field $G \in \mathfrak{X}(U \times \mathbb{R}^n)$ given by

$$G_{(x,v)} = v^k \frac{\partial}{\partial x^k}\bigg|_{(x,v)} - v^i v^j \Gamma^k_{ij}(x) \frac{\partial}{\partial v^k}\bigg|_{(x,v)}. \tag{2.15}$$

By the fundamental theorem on flows, for each $p \in U \times \mathbb{R}^n$ and $t_0 \in \mathbb{R}$, there exist an open interval I_0 containing t_0 and a unique smooth solution $\zeta : I \to U \times \mathbb{R}^n$ to this system satisfying the initial condition $\zeta(t_0) = (p, w)$. If we write the component functions of ζ as $\zeta(t) = (x^i(t), v^i(t))$, then we can easily check that the curve $\gamma(t) = (x^i(t))$ in U satisfies the existence claim of the theorem.

To prove the uniqueness claim, suppose $\gamma, \widetilde{\gamma} : I \to M$ are both geodesics defined on some open interval with $\gamma(t_0) = \widetilde{\gamma}(t_0)$ and $\gamma'(t_0) = \gamma'(t_0)$. In any local coordinates around $\gamma(t_0)$, we can define smooth curves $\zeta, \widetilde{\zeta} : (t_0 - \varepsilon, t_0 + \varepsilon) \to U \times \mathbb{R}^n$ as above. These curves both satisfy the same initial value problem for the system (4.17), so by the uniqueness of ODE solutions, they agree on $(t_0 - \varepsilon, t_0 + \varepsilon)$ for some $\varepsilon > 0$. Suppose for the sake of contradiction that $\gamma(b) \neq \widetilde{\gamma}(b)$ for some $b \in I$. First suppose $b > t_0$, and let β be the infimum of numbers $b \in I$ such that $b > t_0$ and $\gamma(b) \neq \widetilde{\gamma}(b)$. Then $\beta \in I$, and by continuity, $\gamma(\beta) = \widetilde{\gamma}(\beta)$ and $\gamma'(\beta) = \widetilde{\gamma}'(\beta)$. Applying local uniqueness in a neighborhood of β, we conclude that γ and $\widetilde{\gamma}$ agree on a neighborhood of β, which contradicts our choice of β. Arguing similarly to the left of t_0, we conclude that $\gamma \equiv \widetilde{\gamma}$ on all of I. $\square$

A geodesic $\gamma : I \to M$ is said to be maximal if it cannot be extended to a geodesic on a larger interval, that is, if there does not exist a geodesic $\widetilde{\gamma} : \widetilde{I} \to M$ defined on an interval $\widetilde{I}$ properly containing I and satisfying $\widetilde{\gamma}|_I = \gamma$. A **geodesic segment** is a geodesic whose domain is a compact interval.

Corollary 1.2.23. *Let M be a smooth manifold and let ∇ be a connection in TM. For each $p \in M$ and $v \in T_pM$, there is a unique maximal geodesic $\gamma : I \to M$ with $\gamma(0) = p$ and $\gamma'(0) = v$, defined on some open interval I containing 0.*

Proof. Given $p \in M$ and $v \in T_pM$, let I be the union of all open intervals containing 0 on which there is a geodesic with the given initial conditions. By Theorem 1.2.22, all such geodesics agree where they overlap, so they define a geodesic $\gamma : I \to M$, which is obviously the unique maximal geodesic with the given initial conditions. $\square$

The unique maximal geodesic γ with $\gamma(0) = p$ and $\gamma'(0) = v$ is often called simply the **geodesic with initial point p and initial velocity v**, and is denoted by γ_v. (For simplicity, we do not specify the initial point p in the notation; it can implicitly be recovered from v by $p = \pi(v)$, where $\pi : TM \to M$ is the natural projection.)

1.2.9 Parallel transport

Another construction involving covariant differentiation along curves that will be useful later is called **parallel transport**. As we did with geodesics, we restrict attention here to manifolds without boundary.

Let M be a smooth manifold and let ∇ be a connection in TM. A smooth vector or tensor field V along a smooth curve γ is said to be **parallel** along γ (with respect to ∇) if $D_tV \equiv 0$. Thus a geodesic can be characterized as a curve whose velocity vector field is parallel along the curve.

The fundamental fact about parallel vector and tensor fields along curves is that every tangent vector or tensor at any point on a curve can be uniquely extended to a parallel field along the entire curve. Before we prove this claim, let us examine what the equation of parallelism looks like in coordinates. Given a smooth curve γ with a local coordinate representation $\gamma(t) = (\gamma^1(t), \ldots, \gamma^n(t))$, formula (2.12) shows that a vector field V is parallel along

γ if and only if

$$\dot{V}^k(t) = -\dot{\gamma}^i(t)V^j\Gamma_{ij}^k(\gamma(t)) \tag{2.16}$$

with analogous expressions based on Proposition 1.2.15 for tensor fields of other types. In each case, this is a system of first-order linear ordinary differential equations for the unknown coefficients of the vector or tensor field—in the vector case, the functions $(V^1(t), \ldots, V^n(t))$. The usual ODE theorem guarantees the existence and uniqueness of a solution for a short time, given any initial values at $t = t_0$; but since the equation is linear, we can actually show much more: there exists a unique solution on the entire parameter interval.

Theorem 1.2.24 (Existence, Uniqueness, and Smoothness for Linear ODEs). *Let $I \subseteq \mathbb{R}$ be an open interval, and for $1 \le i, j \le n$, let $A_j^k : I \to \mathbb{R}$ be smooth functions. For all $t_0 \in I$ and every initial vector $(c^1, \ldots, c^n) \in \mathbb{R}^n$, the linear initial value problem*

$$\begin{aligned} \dot{V}(t) &= A(t)V(t) \\ V(t_0) &= c \end{aligned} \tag{2.17}$$

has a unique smooth solution on all of I, and the solution depends smoothly on $(t, c) \in \mathbb{R}^n$.

Theorem 1.2.25 (Existence and Uniqueness of Parallel Transport). *Suppose M is a smooth manifold with or without boundary, and ∇ is a connection in TM. Given a smooth curve $\gamma : I \to M$, $t_0 \in I$, and a vector $v \in T_{\gamma(t_0)}M$ or tensor $v \in T^{(k,l)}(T_{\gamma(t_0)}M)$, there exists a unique parallel vector or tensor field V along γ such that $V(t_0) = v$.*

Proof. As in the proof of Theorem 1.2.22, we carry out the proof for vector fields. The case of tensor fields differs only in notation.

First suppose $\gamma(I)$ is contained in a single coordinate chart. Then V is parallel along γ if and only if its components satisfy the linear system of ODEs (2.16). Theorem 1.2.24 guarantees the existence and uniqueness of a solution on all of I with any initial condition $V(t_0) = v$.

Now suppose $\gamma(I)$ is not covered by a single chart. Let β denote the supremum of all $b > t_0$ for which a unique parallel transport exists on $[t_0, b]$. (The argument for $t < t_0$ is similar.) We know that $\beta > t_0$, since for b close enough to t_0, $\gamma([t_0, b])$ is contained in a single chart and the above argument applies. Then a unique parallel transport V exists on $[t_0, \beta)$. If β is equal to $\sup I$, we are done. If not, choose smooth coordinates on an open set containing $\gamma(\beta - \delta, \beta + \delta)$ for some positive δ. Then there exists a unique parallel vector field $\widetilde{V}$ on $(\beta - \delta, \beta + \delta)$ satisfying the initial condition $\widetilde{V}(\beta - \delta/2) = V(\beta - \delta/2)$. By uniqueness, $V = \widetilde{V}$ on their common domain, and therefore $\widetilde{V}$ is a parallel extension of V past β, which is a contradiction. □

The vector or tensor field whose existence and uniqueness are proved in Theorem 1.2.25 is called the **parallel transport of v along γ**. For each $t_0, t_1 \in I$, we define a map

$$P_{t_0 t_1}^{\gamma} : T_{\gamma(t_0)}M \to T_{\gamma(t_1)}M,$$

called the **parallel transport map**, by setting $P_{t_0 t_1}^{\gamma}(v) = V(t_1)$ for each $v \in T_{\gamma(t_0)}M$, where V is the parallel transport of v along γ. This map is linear, because the equation of parallelism is linear. It is in fact an isomorphism, because $P_{t_1 t_0}^{\gamma}$ is an inverse for it.

It is also useful to extend the parallel transport operation to curves that are merely piecewise smooth. Given an admissible curve $\gamma : [a, b] \to M$, a map $V : [a, b] \to TM$ such that

$V(t) \in T_{\gamma(t)}M$ for each t is called a **piecewise smooth vector field along γ** if V is continuous and there is an admissible partition $(a_0, \dots, a_k)$ for γ such that V is smooth on each subinterval $[a_{i-1}, a_i]$. We will call any such partition an **admissible partition** for V. A piecewise smooth vector field V along γ is said to be **parallel along γ** if $D_tV = 0$ wherever V is smooth.

Corollary 1.2.26 (Parallel Transport Along Piecewise Smooth Curves). *Suppose M is a smooth manifold with or without boundary, and ∇ is a connection in TM. Given an admissible curve $\gamma : [a, b] \to M$ and a vector $v \in T_{\gamma(a)}M$ or tensor $v \in T^{(k,l)}(T_{\gamma(a)}M)$, there exists a unique piecewise smooth parallel vector or tensor field V along γ such that $V(a) = v$, and V is smooth wherever γ is.*

Proof. Let $(a_0, \dots, a_k)$ be an admissible partition for γ. First define $V|_{[a_0,a_1]}$ to be the parallel transport of v along the first smooth segment $\gamma([a_0, a_1])$; then define $V|_{[a_1,a_2]}$ to be the parallel transport of $V(a_1)$ along the next smooth segment; and continue by induction. □

Here is an extremely useful tool for working with parallel transport. Given any basis $(e_1, \dots, e_n)$ for $T_{\gamma(t_0)}M$, we can parallel transport the vectors e_i along γ, thus obtaining an n-tuple of parallel vector fields $(E_1, \dots, E_n)$ along γ. Because each parallel transport map is an isomorphism, the vectors $(E_i(t))$ form a basis for $T_{\gamma(t)}M$ at each point $\gamma(t)$. Such an n-tuple of vector fields along γ is called a **parallel frame along γ**. Every smooth (or piecewise smooth) vector field along γ can be expressed in terms of such a frame as $V(t) = V^i(t)E_i(t)$ and then the properties of covariant derivatives along curves, together with the fact that the E_i's are parallel, imply

$$D_tV(t) = \dot{V}^i(t)E_i(t). \tag{2.18}$$

wherever V and γ are smooth. This means that a vector field is parallel along γ if and only if its component functions with respect to the frame $(E_i(t))$ are constants. The parallel transport map is the means by which a connection "connects" nearby tangent spaces. The next theorem and its corollary show that parallel transport determines covariant differentiation along curves, and thereby the connection itself.

Theorem 1.2.27 (Parallel Transport Determines Covariant Differentiation). *Let M be a smooth manifold with or without boundary, and let ∇ be a connection in TM. Suppose $\gamma : I \to M$ is a smooth curve and V is a smooth vector field along γ. For each $t_0 \in I$,*

$$D_tV(t_0) = \lim_{t_1 \to t_0} \frac{P^{\gamma}_{t_1 t_0}V(t_1) - V(t_0)}{t_1 - t_0}. \tag{2.19}$$

Proof. Let (E_i) be a parallel frame along γ, and write $V(t) = V^i(t)E_i(t)$ for $t \in I$. On the one hand, (2.18) shows that $D_tV(t_0) = \dot{V}^i(t_0)E_i(t_0)$. On the other hand, for every fixed $t_1 \in I$, the parallel transport of the vector $V(t_1)$ along γ is the constant-coefficient vector field $W(t) = V(t_1)E_i(t)$ along γ, so $P^{\gamma}_{t_1 t_0}V(t_1) = V^i(t_1)E_{t_0}$. Inserting these formulas into (2.19) and taking the limit as $t_1 \to t_0$, we conclude that the right-hand side is also equal to $\dot{V}^i(t_0)E_i(t_0)$. □

Corollary 1.2.28 (Parallel Transport Determines the Connection). *Let M be a smooth manifold with or without boundary, and let ∇ be a connection in TM. Suppose X and Y are smooth vector fields on M. For every $p \in M$,*

$$\nabla_X Y|_p = \lim_{h \to 0} \frac{P^{\gamma}_{h0}Y_{\gamma(h)} - Y_p}{h} \tag{2.20}$$

where $\gamma : I \to M$ is any smooth curve such that $\gamma(0) = p$ and $\gamma'(0) = X_p$.

Proof. Given $p \in M$ and a smooth curve γ such that $\gamma(0) = p$ and $\gamma'(0) = X_p$, let $V(t)$ denote the vector field along γ determined by Y, so $V(t) = Y_{\gamma(t)}$. By property (iii) of Theorem 1.2.20, $\nabla_X Y|_p$ is equal to $D_t V(0)$, so the result follows from Theorem 1.2.27. □

A smooth vector or tensor field on M is said to be **parallel** (with respect to ∇) if it is parallel along every smooth curve in M. For example, every constant-coefficient vector field on $\mathbb{R}^n$ is parallel.

Proposition 1.2.29. *Suppose M is a smooth manifold with or without boundary, ∇ is a connection in TM, and A is a smooth vector or tensor field on M. Then A is parallel on M if and only if $\nabla A \equiv 0$.*

Proof. Since $D_t A(t) = \nabla_{\gamma'(t)} A$, it is clear that if $\nabla A \equiv 0$, then A is parallel. Now assume that A is parallel, then it follows that $\nabla_v A = 0$ for any $p \in M$ and $v \in T_p M$. This then means $\nabla_X A = 0$ for any vector field $X \in \mathfrak{X}(M)$, so $\nabla A \equiv 0$. □

Although Theorem 1.2.25 showed that it is always possible to extend a vector at a point to a parallel vector field along any given curve, it may not be possible in general to extend it to a parallel vector field on an open subset of the manifold. The impossibility of finding such extensions is intimately connected with the phenomenon of curvature, which will occupy a major portion of our attention.

1.2.10 Pullback connections

Like vector fields, connections in the tangent bundle cannot be either pushed forward or pulled back by arbitrary smooth maps. However, there is a natural way to pull back such connections by means of a diffeomorphism.

Suppose M and $\widetilde{M}$ are smooth manifolds and $\varphi : M \to \widetilde{M}$ is a diffeomorphism. For a smooth vector field $X \in \mathfrak{X}(M)$, recall that the pushforward of X is the unique vector field $\varphi_* X \in \mathfrak{X}(\widetilde{M})$ that satisfies $d\varphi_p(X_p) = (\varphi_* X)_{\varphi(p)}$ for all $p \in M$.

Lemma 1.2.30 (Pullback Connections). *Suppose M and $\widetilde{M}$ are smooth manifolds with or without boundary. If $\widetilde{\nabla}$ is a connection in $T\widetilde{M}$ and $\varphi : M \to \widetilde{M}$ is a diffeomorphism, then the map $\varphi^*\widetilde{\nabla} : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ defined by*

$$(\varphi^*\widetilde{\nabla})_X Y = \varphi^*\big(\widetilde{\nabla}_{\varphi_* X}(\varphi_* Y)\big). \tag{2.21}$$

is a connection in TM, called the ***pullback of $\widetilde{\nabla}$ by φ****.*

Proof. It is immediate from the definition that $\varphi^*\widetilde{\nabla}$ is linear over $\mathbb{R}$ in Y. To see that it is linear over $C^\infty(M)$ in X, let $f \in C^\infty(M)$, and let $\widetilde{f} = \varphi_* f = f \circ \varphi^{-1}$, so $\varphi_*(fX) = \widetilde{f}\varphi_* X$. Then

$$\begin{aligned}
(\varphi^*\widetilde{\nabla})_{fX} Y &= \varphi^*\big(\widetilde{\nabla}_{\varphi_*(fX)}(\varphi_* Y)\big) \\
&= \varphi^*\big(\widetilde{\nabla}_{\widetilde{f}\varphi_* X}(\varphi_* Y)\big) \\
&= \varphi^*\big(\widetilde{f}\widetilde{\nabla}_{\varphi_* X}(\varphi_* Y)\big) \\
&= f\varphi^*\big(\widetilde{\nabla}_{\varphi_* X}(\varphi_* Y)\big) \\
&= f(\varphi^*\widetilde{\nabla})_X Y.
\end{aligned}$$

Finally, to prove the product rule in Y, let f and $\widetilde{f}$ be as above, and note that by Corollary ?? vector field push we have $\varphi^*\big((\varphi_* X)f\big) = X(\varphi^* f)$. Thus

$$\begin{aligned}(\varphi^*\widetilde{\nabla})_X(fY) &= \varphi^*\big(\widetilde{\nabla}_{\varphi_* X}(\widetilde{f}\varphi_* Y)\big)\\ &= \varphi^*\big(\widetilde{f}\widetilde{\nabla}_{\varphi_* X}(\varphi_* Y) + (\varphi_* X)(\widetilde{f})\varphi_* Y\big)\\ &= f(\varphi^*\widetilde{\nabla})_X Y + (Xf)Y.\end{aligned}$$

This completes the proof. □

Now we show the various concepts defined in terms of connections–covariant derivatives along curves, parallel transport, and geodesics–all behave as expected with respect to pullback connections.

ull back prop

Proposition 1.2.31 (Properties of Pullback Connections). *Suppose M and $\widetilde{M}$ are smooth manifolds with or without boundary, and $\varphi : M \to \widetilde{M}$ is a diffeomorphism. Let $\widetilde{\nabla}$ be a connection in $T\widetilde{M}$ and let $\nabla = \varphi^*\widetilde{\nabla}$ be the pullback connection in TM. Suppose $\gamma : I \to M$ is a smooth curve*

(a) *φ takes covariant derivatives along curves to covariant derivatives along curves: if V is a smooth vector field along γ, then*

$$d\varphi \circ D_t V = \widetilde{D}_t(d\varphi \circ V),$$

where D_t is covariant differentiation along γ with respect to ∇, and $\widetilde{D}_t$ is covariant differentiation along $\varphi \circ \gamma$ with respect to $\widetilde{\nabla}$.

(b) *φ takes geodesics to geodesics: if φ is a ∇-geodesic in M, then $\varphi \circ \gamma$ is a $\widetilde{\nabla}$-geodesic in $\widetilde{M}$.*

(c) *φ takes parallel transport to parallel transport: for every $t_0, t_1 \in I$*

$$d\varphi_{\gamma(t_1)} \circ P^{\gamma}_{t_0 t_1} = P^{\varphi\circ\gamma}_{t_0 t_1} \circ d\varphi_{\gamma(t_0)}.$$

connection pull back-1

Proof. From (2.21) we have, for extendible vector fields V, that

$$\begin{aligned}d\varphi \circ D_t V &= \varphi_*(D_t V) = \varphi^*(\nabla_{\gamma'} V)\\ &= \varphi^*\varphi_*(\widetilde{\nabla}_{\varphi_*\gamma'}\varphi_* V)\\ &= \widetilde{\nabla}_{(\varphi\circ\gamma)'}\varphi_* V\\ &= \widetilde{D}_t(d\varphi \circ V).\end{aligned}$$

For nonextendible vector fields along γ, we can imitate the proof in Proposition 1.2.20 connection derivative curv. This proves (a), and $(b), (c)$ are easy consequences of (a). □

1.2.11 Exercise

n torsion def

Exercise 1.2.1. Let M be a smooth manifold and let ∇ be a connection in TM. Define a map $\tau : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ by

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

(a) Show that τ is a $(1, 2)$-tensor field, called the **torsion tensor of ∇**.

(b) We say that ∇ is symmetric if its torsion vanishes identically. Show that ∇ is symmetric if and only if its connection coefficients with respect to every coordinate frame are symmetric: $\Gamma^k_{ij} = \Gamma^k_{ji}$.

(c) Show that ∇ is symmetric if and only if the covariant Hessian $\nabla^2 f$ of every smooth function $f \in C^\infty(M)$ is a symmetric 2-tensor field.

(d) Show that the Euclidean connection $\overline{\nabla}$ on $\mathbb{R}^n$ is symmetric.

Proof. Recall that in local frames we have

$$\nabla_X Y = \big(X(Y^k) + X^i Y^j \Gamma_{ij}^k\big) E_k.$$

Therefore

$$\begin{aligned}\tau(X,Y) &= \big(X(Y^k) + X^i Y^j \Gamma_{ij}^k\big)E_k - \big(Y(X^k) + Y^i X^j \Gamma_{ij}^k\big)E_k - (X(Y^k) - Y(X^k))E_k \\ &= X^i Y^j (\Gamma_{ij}^k - \Gamma_{ji}^k) E_k.\end{aligned}$$

Therefore τ is a $(1,2)$-tensor, and this also proves (b). Finally, in Example 1.2.18 we computed that

$$\begin{aligned}\nabla^2 f(Y,X) - \nabla^2 f(X,Y) &= X(Yf) - Y(Xf) - (\nabla_X Y - \nabla_Y X)f \\ &= ([X,Y] - \nabla_X Y + \nabla_Y X)f \\ &= -\tau(X,Y)f.\end{aligned}$$

Therefore if $\tau = 0$ if and only if $\nabla^2 f$ is symmetric for every f. $\square$

Exercise 1.2.2. Suppose M is a smooth manifold (without boundary), $I \subseteq \mathbb{R}$ is an interval (bounded or not, with or without endpoints), and $\gamma : I \to M$ is a smooth curve.

(a) Show that for every $t_0 \in I$ such that $\gamma'(t_0) \neq 0$, there is a connected neighborhood J of t_0 in I such that every smooth vector field along $\gamma|_J$ is extendible.

(b) Show that if I is an open interval or a compact interval and γ is a smooth embedding, then every smooth vector field along γ is extendible.

Proof. If $\gamma'(t_0) \neq 0$, then there is a compact neighborhood of t_0 on which γ is an injective immersion. It follows that γ is a local embedding near t_0. Then by choosing a slice coordinate near t_0, we see every vector field along γ can be extended near t_0.

It γ is an embedding, then every point of $\gamma(I)$ has a slice chart, so we can extend any vector field along γ near this point. By a partition of unity we can extend it on the whole curve. $\square$

Exercise 1.2.3. Let M be a smooth manifold, and let ∇^0 and ∇^1 be two connections on TM.

(a) Show that ∇^0 and ∇^1 have the same torsion if and only if their difference tensor is symmetric, i.e., $D(X,Y) = D(X,Y)$ for all X and Y.

(b) Show that ∇^0 and ∇^1 determine the same geodesics if and only if their difference tensor is antisymmetric, i.e., $D(X,Y) = -D(X,Y)$ for all X and Y.

Proof. Let τ^0, τ^1 be the torsion of ∇^0, ∇^1, respectively. Then note that

$$\begin{aligned}D(X,Y) - D(Y,X) &= (\nabla_X^0 Y - \nabla_X^1 Y) - (\nabla_Y^0 X - \nabla_Y^1 X) \\ &= \nabla_X^0 Y - \nabla_Y^0 X - (\nabla_X^1 Y - \nabla^1 Y X) \\ &= \tau^0(X,Y) - \tau^1(X,Y).\end{aligned}$$

Therefore $D(X,Y)$ is symmetric if and only if $\tau^0 = \tau^1$.

For (b), first we calculate:

$$\begin{aligned} D(X,Y)+D(Y,X) &= (\nabla^0_X Y - \nabla^1_X Y) + (\nabla^0_Y X - \nabla^1_Y X) \\ &= (\nabla^0_X Y + \nabla^0_Y X) - (\nabla^1_X Y - \nabla^1 Y X) \end{aligned}$$

Therefore $D(X,Y)$ is antisymmetric if and only if $\nabla^0_X Y + \nabla^0_Y X = \nabla^1_X Y + \nabla^1_Y X$. If this holds, then $\nabla^0_{\gamma'}\gamma' = \nabla^1_{\gamma'}\gamma'$, so ∇^0 and ∇^1 define the same geodesics. Conversely, if they defines the same geodesics, then by Proposition 1.2.21 and Proposition 1.2.3 we get $\nabla^0_X X = \nabla^1_X X$ for every $X \in \mathfrak{X}(M)$. Note that

$$\begin{aligned} &\nabla_{X+Y}(X+Y) - \nabla_{X-Y}(X-Y) \\ &= \nabla_X X + \nabla^0_Y Y + \nabla_X Y + \nabla_Y X - (\nabla_X X + \nabla_Y Y - \nabla_X Y - \nabla_Y X) \\ &= 2(\nabla_X Y + \nabla_Y X). \end{aligned}$$

Therefore we conclude that $\nabla^0_X Y + \nabla^0_Y X = \nabla^1_X Y + \nabla^1_Y X$, and thus $D(X,Y)$ is antisymmetric. □

Exercise 1.2.4. Suppose M is a smooth manifold endowed with a connection, $\gamma : I \to M$ is a smooth curve, and $Y \in \mathfrak{X}(M)$. Prove that if Y is parallel along γ, then it is parallel along every reparametrization of γ.

Proof. Let $\varphi : J \to I$ be a reparametrization of γ. Set $V = Y \circ \gamma \circ \varphi$, then we only need to show V is parallel. Choose smooth coordinates (x^i) for M in a neighborhood of $\gamma(t_0)$, and write $Y = (Y^1(t), \ldots, Y^n(t)), V = (V^1(t), \ldots, V^n(t))$. Then we have $V^i = Y^i \circ \varphi$. Moreover, since Y is parallel, we have

$$\dot{Y}^k(t) + \dot{\gamma}^i(t) Y^j(t) \Gamma^k_{ij}(\gamma(t)) = 0$$

for all k. Let $\widetilde{\gamma} = \gamma \circ \varphi$, then the derivative of $V(t)$ along $\gamma \circ \varphi$ has components

$$\begin{aligned} &\dot{V}^k(t) + \dot{\widetilde{\gamma}}^i(t) V^j(t) \Gamma^k_{ij}(\widetilde{\gamma}(t)) \\ &= \dot{Y}^k(\varphi(t))\varphi'(t) + \dot{\gamma}(\varphi(t))\varphi'(t) Y^j(\varphi(t)) \Gamma^k_{ij}(\gamma \circ \varphi(t)) \\ &= \big(\dot{Y}^k(\varphi(t)) + \dot{\gamma}(\varphi(t)) Y^j(\varphi(t)) \Gamma^k_{ij}(\gamma \circ \varphi(t))\big)\varphi'(t) \\ &= 0. \end{aligned}$$

Therefore $V(t)$ is parallel along $\gamma \circ \varphi$. □

Exercise 1.2.5. Suppose G is a Lie group.

(a) Show that there is a unique connection ∇ in TG with the property that every left-invariant vector field is parallel.

(b) Show that the torsion tensor of ∇ is zero if and only if the identity component of G is abelian.

Proof. Recall that for each $d(L_\varphi)$ is an isomorphism between the tangent spaces of G for each $\varphi \in G$, so we define $P^\gamma_{t_0 t_1} = d(L_{\gamma(t_0)\gamma(t_1)^{-1}})_{\gamma(t_1)}$, and so

$$\nabla_X Y|_\varphi = \lim_{h \to 0} \frac{d(L_{\varphi\gamma(h)^{-1}})_{\gamma(h)}(Y_{\gamma(h)}) - Y_\varphi}{h}.$$

where $\gamma : I \to M$ is a smooth curve such that $\gamma(0) = \varphi$ and $\gamma'(0) = X_\varphi$. By Corollary 1.2.28 this defines a connection on TG. Now for any left-invariant vector field $X \in \mathfrak{g}$, we have

$$d(L_{\varphi\gamma(h)^{-1}})_{\gamma(h)}(X_{\gamma(h)}) = X_\varphi.$$

Therefore $\nabla_{\gamma'(t)} X = 0$ for any smooth curve $\gamma : I \to G$, and so X is parallel.

Now assume that $\tau(X, Y) = 0$, then $\nabla_X Y = 0$ for any $X, Y \in \mathfrak{g}$ by (a), and so $[X, Y] = 0$. This means the Lie algebra $\mathfrak{g}$ is abelian, and so is the identity component since it is generated by the image of $\exp$. □

1.3 The Levi-Civita connection

Except where noted otherwise, the results and proofs of this section do not use positivity of the metric, so they apply equally well to Riemannian and pseudo-Riemannian manifolds.

1.3.1 The tangential connection revisited

We are eventually going to show that on each Riemannian manifold there is a natural connection that is particularly well suited to computations in Riemannian geometry. Since we get most of our intuition about Riemannian manifolds from studying submanifolds of $\mathbb{R}^n$ with the induced metric, let us start by examining that case.

Let $M \subseteq \mathbb{R}^n$ be an embedded submanifold. As a guiding principle, a geodesic in M should be "as straight as possible". A reasonable way to make this rigorous is to require that the geodesic have no acceleration in directions tangent to the manifold, or in other words that its acceleration vector have zero orthogonal projection onto TM. The tangential connection defined in Example 1.2.7 is perfectly suited to this task, because it computes covariant derivatives on M by taking ordinary derivatives in $\mathbb{R}^n$ and projecting them orthogonally to TM.

It is easy to compute covariant derivatives along curves in M with respect to the tangential connection. Suppose $\gamma : I \to M$ is a smooth curve. Then γ can be regarded as either a smooth curve in M or a smooth curve in $\mathbb{R}^n$, and a smooth vector field V along γ that takes its values in TM can be regarded as either a vector field along γ in M or a vector field along γ in $\mathbb{R}^n$. Let $\overline{D}_t V$ denote the covariant derivative of V along γ (as a curve in $\mathbb{R}^n$) with respect to the Euclidean connection $\overline{\nabla}$, and let $D_t^\top V$ denote its covariant derivative along γ (as a curve in M) with respect to the tangential connection $\nabla^\top$. The next proposition shows that the two covariant derivatives along γ have a simple relationship to each other.

Proposition 1.3.1. *Let $M \subseteq \mathbb{R}^n$ be an embedded submanifold, $\gamma : I \to M$ a smooth curve in M, and V a smooth vector field along γ that takes its values in TM. Then for each $t \in I$,*

$$D_t^\top V(t) = \pi^\top(\overline{D}_t V(t)).$$

Proof. Let $t_0 \in I$ be arbitrary. By Proposition **??**, on some neighborhood U of $\gamma(t_0)$ in $\mathbb{R}^n$ there is an adapted orthonormal frame for TM, that is, a local orthonormal frame $(E_1, \ldots, E_n)$ for $T\mathbb{R}^n$ such that $(E_1, \ldots, E_k)$ restricts to an orthonormal frame for TM at points of $M \cap U$ (where $k = \dim M$). If $\varepsilon > 0$ is small enough that $\gamma(t_0 - \varepsilon, t_0 + \varepsilon) \subseteq U$, then for $t \in (t_0 - \varepsilon, t_0 + \varepsilon)$ we can write

$$V(t) = V^1(t) E_1|_{\gamma(t)} + \cdots + V^k(t) E_k|_{\gamma(t)},$$

for some smooth functions $V^1, \dots, V^k : (t_0 - \varepsilon, t_0 + \varepsilon) \to \mathbb{R}$. Formula (2.12) yields

$$\begin{aligned}
\pi^\top(\overline{D}_t V(t)) &= \pi^\top \Big(\sum_{i=1}^k \big(\dot V^i(t) E_i|_{\gamma(t)} + V^i(t) \overline{\nabla}_{\gamma'(t)} E_i|_{\gamma(t)} \big) \Big) \\
&= \sum_{i=1}^k \Big(\dot V^i(t) E_i|_{\gamma(t)} + V^i(t) \pi^\top (\overline{\nabla}_{\gamma'(t)} E_i|_{\gamma(t)}) \Big) \\
&= \sum_{i=1}^k \Big(\dot V^i(t) E_i|_{\gamma(t)} + V^i(t) \nabla^\top_{\gamma'(t)} E_i|_{\gamma(t)}) \Big) \\
&= D_t^\top V(t).
\end{aligned}$$

This completes the proof. □

Corollary 1.3.2. *Suppose $M \subseteq \mathbb{R}^n$ is an embedded submanifold. A smooth curve $\gamma : I \to M$ is a geodesic with respect to the tangential connection on M if and only if its ordinary acceleration $\gamma''(t)$ is orthogonal to $T_{\gamma(t)}M$ for all $t \in I$.*

Proof. As noted in Example 1.2.6, the connection coefficients of the Euclidean connection on $\mathbb{R}^n$ are all zero. Thus it follows from (2.12) that the Euclidean covariant derivative of γ' along γ is just its ordinary acceleration: $\overline{D}_t \gamma'(t) = \gamma''(t)$. The corollary then follows from Proposition 1.3.1. □

These considerations can be extended to pseudo-Riemannian manifolds as well. Let $(\mathbb{R}^{r,s}, \bar q^{(r,s)})$ be the pseudo-Euclidean space of signature (r, s). If $M \subseteq \mathbb{R}^{r,s}$ is an embedded Riemannian or pseudo-Riemannian submanifold, then for each $p \in M$, the tangent space $T_p\mathbb{R}^{r,s}$ decomposes as a direct sum $T_pM \oplus N_pM$, where $N_pM = (T_pM)^\perp$ is the orthogonal complement of T_pM with respect to $\bar q^{(r,s)}$. We let $\pi^\top : T_p\mathbb{R}^{r,s} \to T_pM$ be the $\bar q^{(r,s)}$-orthogonal projection, and define the **tangential connection** $\nabla^\top$ on M by

$$\nabla^\top_X Y = \pi^\top(\nabla_{\widetilde X} \widetilde Y),$$

where $\widetilde X$ and $\widetilde Y$ are smooth extensions of X and Y to a neighborhood of M, and ∇ is the ordinary Euclidean connection on $\mathbb{R}^{r,s}$. This is a well-defined connection on M by the same argument as in the Euclidean case, and the next proposition is proved in exactly the same way as Corollary 1.3.2.

Proposition 1.3.3. *Suppose M is an embedded Riemannian or pseudo-Riemannian submanifold of the pseudo-Euclidean space $\mathbb{R}^{r,s}$. A smooth curve $\gamma : I \to M$ is a geodesic with respect to $\nabla^\top$ if and only if $\gamma''(t)$ is $\bar q^{(r,s)}$-orthogonal to $T_{\gamma(t)}M$ for all $t \in I$.*

1.3.2 Connections on abstract Riemannian manifolds

Metric connections

The Euclidean connection on $\mathbb{R}^n$ has one very nice property with respect to the Euclidean metric: it satisfies the product rule

$$\overline{\nabla}_X \langle Y, Z \rangle = \langle \overline{\nabla}_X Y, Z \rangle + \langle Y, \overline{\nabla}_X Z \rangle.$$

The Euclidean connection has the same property with respect to the pseudo-Euclidean metric on $\mathbb{R}^{r,s}$. It is almost immediate that the tangential connection on a Riemannian or pseudo-Riemannian submanifold satisfies the same product rule, if we now interpret all the vector fields as being tangent to M and interpret the inner products as being taken with respect to the induced metric on M.

This property makes sense on an abstract Riemannian or pseudo-Riemannian manifold. Let g be a Riemannian or pseudo-Riemannian metric on a smooth manifold M (with or without boundary). A connection ∇ on TM is said to be **compatible with g**, or to be a **metric connection**, if it satisfies the following product rule for all $X, Y, Z \in \mathfrak{X}(M)$:

$$\nabla_X \langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle. \tag{3.1}$$

The next proposition gives several alternative characterizations of compatibility with a metric, any one of which could be used as the definition.

Proposition 1.3.4 (Characterizations of Metric Connections). *Let (M, g) be a Riemannian or pseudo-Riemannian manifold (with or without boundary), and let ∇ be a connection on TM. The following conditions are equivalent:*

(a) ∇ is compatible with g: $\nabla_X \langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$.

(b) g is parallel with respect to ∇: $\nabla g \equiv 0$.

(c) In terms of any smooth local frame (E_i), the connection coefficients of ∇ satisfy

$$\Gamma^l_{ki} g_{lj} + \Gamma^l_{kj} g_{il} = E_k(g_{ij}). \tag{3.2}$$

(d) If V, W are smooth vector fields along any smooth curve γ, then

$$\frac{d}{dt}\langle V, W\rangle = \langle D_t Y, Z\rangle + \langle Y, D_t Z\rangle. \tag{3.3}$$

(e) If V, W are parallel vector fields along a smooth curve γ in M, then $\langle V, W\rangle$ is constant along γ.

(f) Given any smooth curve γ in M, every parallel transport map along γ is a linear isometry.

(g) Given any smooth curve γ in M, every orthonormal basis at a point of γ can be extended to a parallel orthonormal frame along γ.

Proof. First we prove $(a) \Leftrightarrow (b)$. By Proposition 1.2.12, the total covariant derivative of the symmetric 2-tensor g is given by

$$(\nabla g)(Y, Z, X) = (\nabla_X g)(Y, Z) = X(g(Y, Z)) - g(\nabla_X Y, Z) - g(Y, \nabla_Z).$$

This is zero for all X, Y, Z if and only if (3.1) is satisfied for all X, Y, Z.

To prove $(b) \Leftrightarrow (c)$, note that Proposition 1.2.15 shows that the components of ∇g in terms of a smooth local frame (E_i) are

$$g_{ij;k} = E_k(g_{ij}) - \Gamma^l_{ki} g_{lj} - \Gamma^l_{kj} g_{il}.$$

These are all zero if and only if (3.2) is satisfied.

Next we prove $(a) \Leftrightarrow (b)$. Assume (a), and let V, W be smooth vector fields along a smooth curve $\gamma : I \to M$. Given $t_0 \in I$, in a neighborhood of $\gamma(t_0)$ we may choose coordinates (x^i) and write $V = V^i \partial_i$ and $W = W^j \partial_j$ for some smooth functions $V^i, W^i : I \to \mathbb{R}$.

Applying (3.1) to the extendible vector fields ∂_i, ∂_j, we obtain

$$\begin{aligned}\frac{d}{dt}\langle V, W\rangle &= \frac{d}{dt}\big(V^i, W^j\langle \partial_i, \partial_j\rangle\big)\\ &= (\dot V^i W^j + V^i \dot W^j)\langle \partial_i, \partial_j\rangle + V^i W^i\big(\langle \nabla_{\gamma'(t)}\partial_i, \partial_j\rangle + \langle \partial_i, \nabla_{\gamma'(t)}\partial_j\rangle\big)\\ &= \langle D_t V, W\rangle + \langle V, D_t W\rangle.\end{aligned}$$

which proves (d). Conversely, if (d) holds, then in particular it holds for extendible vector fields along γ, and then (a) follows from part (iii) of Theorem 1.2.28.

Now we will prove $(d) \Rightarrow (e) \Rightarrow (f) \Rightarrow (g) \Rightarrow (d)$. Assume first that (d) holds. If V and W are parallel along γ, then (3.3) shows that $\langle V, W\rangle$ has zero derivative with respect to t, so it is constant along γ.

Now assume (e). Let v_0, w_0 be arbitrary vectors in $T_{\gamma(t_0)}M$, and let V, W be their parallel transports along γ, so that $V(t_0) = v_0$, $W(t_0) = w_0$, $P^\gamma_{t_0t_1}(v_0) = V(t_1)$ and $P^\gamma_{t_0t_1}(w_0) = W(t_1)$. Because $\langle V, W\rangle$ is constant along γ, it follows that

$$\langle P^\gamma_{t_0t_1}(v_0), P^\gamma_{t_0t_1}(w_0)\rangle = \langle V(t_1), W(t_1)\rangle = \langle V(t_0), W(t_0)\rangle = \langle v_0, w_0\rangle,$$

so $P^\gamma_{t_0t_1}$ is a linear isometry.

Next, assuming (f), we suppose $\gamma : I \to M$ is a smooth curve and (e_i) is an orthonormal basis for $T_{\gamma(t_0)}M$, for some $t_0 \in I$. We can extend each e_i by parallel transport to obtain a smooth parallel vector field E_i along γ, and the assumption that parallel transport is a linear isometry guarantees that the resulting n-tuple (E_i) is an orthonormal frame at all points of γ.

Finally, assume that (g) holds, and let (E_i) be a parallel orthonormal frame along γ. Given smooth vector fields V and W along γ, we can express them in terms of this frame as $V = V^i E_i$ and $W = W^j E_j$. The fact that the frame is orthonormal means that the metric coefficients $g_{ij} = \langle E_i, E_j\rangle$ are constants along γ (± 1 or 0), and the fact that it is parallel means that $D_t = \dot V^i E_i$ and $D_t = \dot W^i E_i$. Thus both sides of (3.3) reduce to the following expression:

$$g_{ij}(\dot V^i W^j + V^i \dot W^j).$$

This proves (d). $\square$

Corollary 1.3.5. *Suppose (M, g) is a Riemannian or pseudo-Riemannian manifold with or without boundary, ∇ is a metric connection on M, and $\gamma : I \to M$ is a smooth curve.*

(a) $|\gamma'(t)|$ is constant if and only if $D_t\gamma'(t)$ is orthogonal to $\gamma'(t)$ for all $t \in I$.

(b) If γ is a geodesic, then $|\gamma'(t)|$ is constant.

Proof. If ∇ is a metric connection on M, then for any curve $\gamma : I \to M$,

$$\frac{d}{dt}\langle \gamma'(t), \gamma'(t)\rangle = 2\langle D_t\gamma'(t), \gamma'(t)\rangle.$$

Therefore $|\gamma'(t)|$ is constant if and only if $\langle D_t\gamma'(t), \gamma'(t)\rangle = 0$, that is, if and only if $D_t\gamma'(t)$ is orthogonal to $\gamma'(t)$. In particular, if γ is a geodesic, then $D_t\gamma'(t) = 0$, so this holds. $\square$

Proposition 1.3.6. *If M is an embedded Riemannian or pseudo-Riemannian submanifold of $\mathbb{R}^n$ or $\mathbb{R}^{r,s}$, the tangential connection on M is compatible with the induced Riemannian or pseudo-Riemannian metric.*

Proof. We will show that ∇ satisfies (3.1). Suppose $X, Y, Z \in \mathfrak{X}(M)$ and let $\widetilde{X}, \widetilde{Y}, \widetilde{Z}$ be smooth extensions of them to an open subset of $\mathbb{R}^n$ or $\mathbb{R}^{r,s}$. At points of M, we have

$$\begin{aligned}
\nabla^\top_X \langle Y, Z\rangle &= X\langle Y, Z\rangle = \widetilde{X}\langle \widetilde{Y}, \widetilde{Z}\rangle \\
&= \overline{\nabla}_{\widetilde{X}}\langle \widetilde{Y}, \widetilde{Z}\rangle \\
&= \langle \overline{\nabla}_{\widetilde{X}}\widetilde{Y}, \widetilde{Z}\rangle + \langle \widetilde{Y}, \overline{\nabla}_{\widetilde{X}}\widetilde{Z}\rangle \\
&= \langle \pi^\top(\overline{\nabla}_{\widetilde{X}}\widetilde{Y}), \widetilde{Z}\rangle + \langle \widetilde{Y}, \pi^\top(\overline{\nabla}_{\widetilde{X}}\widetilde{Z})\rangle \\
&= \langle \nabla^\top_X Y, Z\rangle + \langle Y, \nabla^\top_X Z\rangle
\end{aligned}$$

where the next-to-last equality follows from the fact that $\widetilde{Y}$ and $\widetilde{Z}$ are tangent to M. □

Symmetric connections

It turns out that every abstract Riemannian or pseudo-Riemannian manifold admits many different metric connection, so requiring compatibility with the metric is not sufficient to pin down a unique connection on such a manifold. To do so, we turn to another key property of the tangential connection. Recall the definition (2.1 of the Euclidean connection. The expression on the right-hand side of that definition is reminiscent of part of the coordinate expression for the Lie bracket:

$$[X, Y] = X(Y^i)\frac{\partial}{\partial x^i} - Y(X^i)\frac{\partial}{\partial x^i}.$$

In fact, the two terms in the Lie bracket formula are exactly the coordinate expressions for $\overline{\nabla}_X Y$ and $\overline{\nabla}_Y X$. Therefore, the Euclidean connection satisfies the following identity for all smooth vector fields X, Y:

$$\overline{\nabla}_X Y - \overline{\nabla}_X Y = [X, Y].$$

This expression has the virtue that it is coordinate-independent and makes sense for every connection on the tangent bundle. We say that a connection ∇ on the tangent bundle of a smooth manifold M is **symmetric** if

$$\nabla_X Y - \nabla_X Y = [X, Y]$$

for all $X, Y \in \mathfrak{X}(M)$. The symmetry condition can also be expressed in terms of the torsion tensor of the connection; this is the smooth $(1,2)$-tensor field $\tau : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ defined by

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

Thus a connection ∇ is symmetric if and only if its torsion vanishes identically. It follows from the result of Exercise 1.2.1 that a connection is symmetric if and only if its connection coefficients in every coordinate frame satisfy $\Gamma^k_{ij} = \Gamma^k_{ji}$; this is the origin of the term "symmetric".

Proposition 1.3.7. *If M is an embedded (pseudo-)Riemannian submanifold of a (pseudo-)Euclidean space, then the tangential connection on M is symmetric.*

Proof. Let M be an embedded Riemannian or pseudo-Riemannian submanifold of $\mathbb{R}^n$, where $\mathbb{R}^n$ is endowed either with the Euclidean metric or with a pseudo-Euclidean metric $\bar{q}^{(r,s)}$, $r + s = n$. Let $X, Y \in \mathfrak{X}(M)$, and let $\widetilde{X}, \widetilde{Y}$ be smooth extensions of them to an open subset of the ambient space. If $\iota : M \hookrightarrow \mathbb{R}^n$ represents the inclusion map, it follows that X and Y

are ι-related to $\widetilde{X}$ and $\widetilde{Y}$, respectively, and thus by the naturality of the Lie bracket, $[X,Y]$ is ι-related to $[\widetilde{X},\widetilde{Y}]$. In particular, $[\widetilde{X},\widetilde{Y}]$ is tangent to M, and its restriction to M is equal to $[X,Y]$. Therefore,

$$\begin{aligned}\nabla^\top_X Y - \nabla^\top_Y X &= \pi^\top(\overline{\nabla}_{\widetilde{X}}\widetilde{Y} - \overline{\nabla}_{\widetilde{Y}}\widetilde{X})\\ &= \pi^\top([\widetilde{X},\widetilde{Y}])\\ &= [\widetilde{X},\widetilde{Y}]|_M\\ &= [X,Y],\end{aligned}$$

and so $\nabla^\top$ is symmetric. □

The last two propositions show that if we wish to single out a connection on each Riemannian or pseudo-Riemannian manifold in such a way that it matches the tangential connection when the manifold is presented as an embedded submanifold of $\mathbb{R}^n$ or $\mathbb{R}^{r,s}$ with the induced metric, then we must require at least that the connection be compatible with the metric and symmetric. It is a pleasant fact that these two conditions are enough to determine a unique connection.

Theorem 1.3.8 (Fundamental Theorem of Riemannian Geometry). *Let (M,g) be a Riemannian or pseudo-Riemannian manifold (with or without boundary). There exists a unique connection ∇ on TM that is compatible with g and symmetric. It is called the **Levi-Civita connection** of g (or also, when g is positive definite, the **Riemannian connection**).*

Proof. We prove uniqueness first, by deriving a formula for ∇. Suppose, therefore, that ∇ is such a connection, and let $X,Y,Z \in \mathfrak{X}(M)$. Writing the compatibility equation three times with X,Y,Z cyclically permuted, we obtain

$$\begin{aligned}X\langle Y,Z\rangle &= \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle,\\ Y\langle Z,X\rangle &= \langle \nabla_Y Z, X\rangle + \langle Z, \nabla_Y X\rangle,\\ Z\langle X,Y\rangle &= \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle.\end{aligned}$$

Using the symmetry condition on the last term in each line, this can be rewritten as

$$\begin{aligned}X\langle Y,Z\rangle &= \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_Z X\rangle + \langle Y, [X,Z]\rangle,\\ Y\langle Z,X\rangle &= \langle \nabla_Y Z, X\rangle + \langle Z, \nabla_X Y\rangle + \langle Z, [Y,X]\rangle,\\ Z\langle X,Y\rangle &= \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Y Z\rangle + \langle X, [Z,Y]\rangle.\end{aligned}$$

Adding the first two of these equations and subtracting the third, we obtain

$$\begin{aligned}X\langle Y,Z\rangle + Y\langle Z,X\rangle - Z\langle X,Y\rangle &= 2\langle \nabla_X Y, Z\rangle + \langle Y,[X,Z]\rangle + \langle Z,[Y,X]\rangle - \langle X,[Z,Y]\rangle\\ &= 2\langle \nabla_X Y, Z\rangle - \langle Y,[Z,X]\rangle - \langle Z,[X,Y]\rangle + \langle X,[Y,Z]\rangle.\end{aligned}$$

Finally, solving for $\langle \nabla_X Y, Z\rangle$, we get

$$\langle \nabla_X Y, Z\rangle = \frac{1}{2}\big(X\langle Y,Z\rangle + Y\langle Z,X\rangle - Z\langle X,Y\rangle - \langle X,[Y,Z]\rangle + \langle Y,[Z,X]\rangle + \langle Z,[X,Y]\rangle\big) \tag{3.4}$$

Levi-Civita c

Now suppose ∇_1 and ∇_2 are two connections on TM that are symmetric and compatible with g. Since the right-hand side of (3.4) does not depend on the connection, it follows that $\langle \nabla^1_X Y - \nabla^2_X Y, Z\rangle = 0$ for all X,Y,Z. This can happen only if $\nabla^1_X Y = \nabla^2_X Y$ for all X and Y, so $\nabla^1 = \nabla^2$.

To prove existence, we use (3.4), or rather a coordinate version of it. It suffices to prove that such a connection exists in each coordinate chart, for then uniqueness ensures that the connections in different charts agree where they overlap.

Let $(U, (x^i))$ be any smooth local coordinate chart. Applying (3.4) to the coordinate vector fields, whose Lie brackets are zero, we obtain

$$\langle \nabla_{\partial_i} \partial_j, \partial_l \rangle = \frac{1}{2}\big(\partial_i \langle \partial_j, \partial_l\rangle + \partial_j \langle \partial_l, \partial_i\rangle - \partial_l \langle \partial_i, \partial_j \rangle\big) \tag{3.5}$$

Recall the definitions of the metric coefficients and the connection coefficients: $g_{ij} = \langle \partial_i, \partial_j \rangle$, $\nabla_{\partial_i}\partial_j = \Gamma^k_{ij}\partial_k$. Inserting these into (3.5) yields

$$\Gamma^k_{ij} g_{kl} = \frac{1}{2}(\partial_i g_{jl} + \partial_j g_{li} - \partial_l g_{ij}). \tag{3.6}$$

Finally, multiplying both sides by the inverse matrix g^{kl}, we get

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl}(\partial_i g_{jl} + \partial_j g_{li} - \partial_l g_{ij}). \tag{3.7}$$

This formula certainly defines a connection in each chart, and it is evident from the formula that $\Gamma^k_{ij} = \Gamma^k_{ji}$, so the connection is symmetric. Thus only compatibility with the metric needs to be checked. Using (3.6) twice, we get

$$\Gamma^l_{ki} g_{lj} + \Gamma^l_{kj} g_{il} = \frac{1}{2}(\partial_k g_{ij} + \partial_i g_{jk} - \partial_j g_{ki}) + \frac{1}{2}(\partial_k g_{ji} + \partial_j g_{ik} - \partial_i g_{kj}) = \partial_k g_{ij}$$

By Proposition 1.3.4(c), this shows that ∇ is compatible with g. □

A bonus of this proof is that it gives us explicit formulas that can be used for computing the Levi-Civita connection in various circumstances.

Corollary 1.3.9 (Formulas for the Levi-Civita Connection). *Let (M, g) be a Riemannian or pseudo-Riemannian manifold (with or without boundary), and let ∇ be its Levi-Civita connection.*

(a) In terms of vector fields: If X, Y, Z are smooth vector fields on M, then

$$\langle \nabla_X Y, Z\rangle = \frac{1}{2}\big(X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle - \langle X, [Y,Z]\rangle + \langle Y, [Z,X]\rangle + \langle Z, [X,Y]\rangle\big) \tag{3.8}$$

*This is known as **Koszul's formula**.*

(b) In coordinates: In any smooth coordinate chart for M, the coefficients of the Levi-Civita connection are given by

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl}(\partial_i g_{jl} + \partial_j g_{li} - \partial_l g_{ij}). \tag{3.9}$$

(c) In a local frame: Let (E_i) be a smooth local frame on an open subset $U \subseteq M$, and let $c^k_{ij} : U \to \mathbb{R}$ be the n^3 smooth functions defined by

$$[E_i, E_j] = c^k_{ij} E_k. \tag{3.10}$$

Then the coefficients of the Levi-Civita connection in this frame are

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl}(E_i g_{jl} + E_j g_{li} - E_l g_{ij} - g_{im} c^m_{jl} + g_{jm} c^m_{li} + g_{lm} c^m_{ij}) \tag{3.11}$$

(d) *In a local orthonormal frame: If g is Riemannian, (E_i) is a smooth local orthonormal frame, and the functions c_{ij}^k are defined by (3.10), then*

$$\Gamma_{ij}^k = \frac{1}{2}(-c_{jk}^i + c_{ki}^j + c_{ij}^k). \tag{3.12}$$

On every Riemannian or pseudo-Riemannian manifold, we will always use the Levi-Civita connection fromnowonwithout further comment. Geodesics with respect to this connection are called **Riemannian** (or **pseudo-Riemannian**) **geodesics**, or simply "geodesics" as long as there is no risk of confusion. The connection coefficients Γ_{ij}^k of the Levi-Civita connection in coordinates, given by (3.9), are called the **Christoffel symbols** of g.

Now we give an intrinsic expression of the Levi-Civita connection.

Proposition 1.3.10. *Let (M, g) be a Riemannian or pseudo-Riemannian manifold (with or without boundary), then the Levi-Civita connection is given by*

$$2g(\nabla_X Y, Z) = (\mathfrak{L}_Y g)(X, Z) + (dY^\flat)(X, Z).$$

Proof. We have, by the Koszul formula, that

$$\begin{aligned}
&(\mathfrak{L}_Y g)(X, Z) + (dY^\flat)(X, Z)\\
&= Y(g(X, Z)) - g([Y, X], Z) - g(X, [Y, Z]) + X(Y^\flat(Z)) - Z(Y^\flat(X)) - Y^\flat([X, Z])\\
&= Y(g(X, Z)) - g([Y, X], Z) - g(X, [Y, Z]) + X(g(Y, Z)) - Z(g(X, Y)) - g(Y, [X, Z])\\
&= X(g(Y, Z)) + Y(g(Z, X)) - Z(g(X, Y)) - g(X, [Y, Z]) + g(Y, [Z, X]) + g(Z, [X, Y])\\
&= X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle - \langle X, [Y, Z]\rangle + \langle Y, [Z, X]\rangle + \langle Z, [X, Y]\rangle\\
&= 2g(\nabla_X Y, Z),
\end{aligned}$$

This gives the claim. □

Example 1.3.11. We consider the Levi-Civita connection of a Lie group with a bi-invariant metric. Let G be a Lie group and $\mathfrak{g}$ its Lie algebra. Suppose g is a bi-invariant Riemannian metric on G and ∇ is the Levi-Civita connection of g. Since g is left-invariant, by Proposition 1.1.9 we know that $\langle X, Y\rangle$ is constant for any $X, Y \in \mathfrak{g}$. Therefore the Koszul's formula (3.8) becomes

$$\langle \nabla_X Y, Z\rangle = \frac{1}{2}\big(-\langle X, [Y, Z]\rangle + \langle Y, [Z, X]\rangle + \langle Z, [X, Y]\rangle\big). \tag{3.13}$$

Also, since g is bi-invariant, by Proposition 1.1.10 we know that $\langle \mathrm{Ad}(\varphi)X, \mathrm{Ad}(\varphi)Y\rangle = \langle X, Y\rangle$ is constant on G for every $\varphi \in G$. In particular, the function

$$\langle \mathrm{Ad}(\exp(tZ))X, \mathrm{Ad}(\exp(tZ)Y\rangle$$

is constant on G. Differentiating this equation with respect to t and use the property of metric connection and that $\mathrm{Ad}_* = \mathrm{ad}$, we get

$$\begin{aligned}
0 &= \frac{d}{dt}\Big|_{t=0}\langle \mathrm{Ad}(\exp(tZ))X, \mathrm{Ad}(\exp(tZ)Y\rangle\\
&= \langle \mathrm{ad}_Z X, Y\rangle + \langle X, \mathrm{ad}_Z Y\rangle\\
&= \langle [Z, X], Y\rangle + \langle X, [Z, Y]\rangle.
\end{aligned}$$

Together with (3.13), we then get

$$\langle \nabla_X Y, Z\rangle = \frac{1}{2}\langle [X, Y], Z\rangle$$

for all $X, Y \in \mathfrak{g}$ and $Z \in \mathfrak{X}(G)$. This then implies $\nabla_X Y = [X, Y]/2$.

The next proposition shows that these connections are familiar ones in the case of embedded submanifolds of Euclidean or pseudo-Euclidean spaces.

Proposition 1.3.12 (The Levi-Civita connection in the Euclidean case).

(a) *The Levi-Civita connection on a (pseudo-)Euclidean space is equal to the Euclidean connection* $\overline{\nabla}$.

(b) *Suppose M is an embedded (pseudo-)Riemannian submanifold of a (pseudo-)Euclidean space. Then the Levi-Civita connection on M is equal to the tangential connection $\nabla^\top$.*

Proof. We observed that the Euclidean connection is symmetric and compatible with both the Euclidean metric $\bar{g}$ and the pseudo-Euclidean metrics $\bar{q}^{(r,s)}$, which implies (a). Part (b) then follows from Propositions 1.3.6 and 1.3.7. □

An important consequence of the definition is that because Levi-Civita connections are defined in coordinate-independent terms, they behave well with respect to isometries.

Proposition 1.3.13 (Naturality of the Levi-Civita Connection). *Suppose (M, g) and $(\widetilde{M}, \widetilde{g})$ are Riemannian or pseudo-Riemannian manifolds with or without boundary, and let ∇ denote the Levi-Civita connection of g and $\widetilde{\nabla}$ that of $\widetilde{g}$. If $\varphi : M \to \widetilde{M}$ is an isometry, then $\varphi^*\widetilde{\nabla} = \nabla$.*

Proof. By uniqueness of the Levi-Civita connection, it suffices to show that the pullback connection $\varphi^*\widetilde{\nabla}$ is symmetric and compatible with g. The fact that φ is an isometry means that for any $X, Y \in \mathfrak{X}(M)$ and $p \in M$,

$$\langle Y_p, Z_p\rangle = \langle d\varphi_p(Y_p), d\varphi_p(Z_p)\rangle = \langle (\varphi_* Y)_{\varphi(p)}, (\varphi_* Z)_{\varphi(p)}\rangle,$$

or in other words, $\langle Y, Z\rangle = \langle \varphi_* Y, \varphi_* Z\rangle \circ \varphi$. Therefore

$$\begin{aligned}
X\langle Y, Z\rangle &= X(\langle \varphi_* Y, \varphi_* Z\rangle \circ \varphi)\\
&= \varphi^*\big((\varphi_* X)\langle \varphi_* Y, \varphi_* Z\rangle\big)\\
&= \varphi^*\big(\langle \widetilde{\nabla}_{\varphi_* X}\varphi_* Y, \varphi_* Z\rangle + \langle \varphi_* Y, \widetilde{\nabla}_{\varphi_* X}\varphi_* Z\rangle\big)\\
&= \langle \varphi^*\widetilde{\nabla}_{\varphi_* X}\varphi_* Y, \varphi^*\varphi_* Z\rangle + \langle \varphi^*\varphi_* Y, \varphi^*\widetilde{\nabla}_{\varphi_* X}\varphi_* Z\rangle\\
&= \langle (\varphi^*\widetilde{\nabla})_X Y, Z\rangle + \langle Y, (\varphi^*\widetilde{\nabla})_X Z\rangle.
\end{aligned}$$

which shows that the pullback connection is compatible with g. Symmetry is proved as follows:

$$\begin{aligned}
(\varphi^*\widetilde{\nabla})_X Y - (\varphi^*\widetilde{\nabla})_Y X &= \varphi^*\big(\widetilde{\nabla}_{\varphi_* X}(\varphi_* Y) - \widetilde{\nabla}_{\varphi_* X}(\varphi_* Y)\big)\\
&= \varphi^*([\varphi_* X, \varphi_* Y])\\
&= [X, Y].
\end{aligned}$$

This completes the proof. □

Corollary 1.3.14 (Naturality of Geodesics). *Suppose that (M, g) and $(\widetilde{M}, \widetilde{g})$ are Riemannian or pseudo-Riemannian manifolds with or without boundary, and $\varphi : M \to \widetilde{M}$ is a local isometry. If γ is a geodesic in M, then $\varphi \circ \gamma$ is a geodesic in $\widetilde{M}$.*

Proof. This is an immediate consequence of Proposition 1.2.31, together with the fact that being a geodesic is a local property. □

Like every connection on the tangent bundle, the Levi-Civita connection induces connections on all tensor bundles.

form parallel

Proposition 1.3.15. *Suppose (M,g) is a Riemannian or pseudo-Riemannian manifold. The connection induced on each tensor bundle by the Levi-Civita connection is compatible with the induced inner product on tensors, in the sense that $X\langle F,G\rangle = \langle \nabla_X F, G\rangle + \langle F, \nabla_G\rangle$ for every vector field X and every pair of smooth tensor fields $F,G \in \Gamma(T^{(r,s)}TM)$.*

Proof. Since every tensor field can be written as a sum of tensor products of vector and/or covector fields, it suffices to consider the case in which $F = \alpha \otimes \cdots \otimes \alpha_{r+s}$ and $G = \beta_1 \otimes \cdots \otimes \beta_{r+s}$, where α_i and β_i are covariant or contravariant 1-tensor fields, as appropriate. In this case, the formula follows from (**??**) by a routine computation. Riemann tensor inner prod-2 $\square$

form parallel

Proposition 1.3.16. *Let (M,g) be an oriented Riemannian manifold. The Riemannian volume form of g is parallel with respect to the Levi-Civita connection.*

Proof. Let $p \in M$ and $v \in T_pM$ be arbitrary, and let $\gamma : (-\varepsilon,\varepsilon) \to M$ be a smooth curve satisfying $\gamma(0) = p$ and $\gamma'(0) = v$. Let $(E_1,\ldots,E_n)$ be a parallel oriented orthonormal frame along γ. Since $dV_g(E_1,\ldots,E_n) = 1$ and $D_tE_i = 0$ along γ, formula (2.10) shows that $\nabla_v(dV_g) = D_t(dV_g)|_{t=0} = 0$. connection def on tensor-1 $\square$

Recall that the Hessian of a smooth function is symmetric if the connection is symmetric. In the case of the Levi-Civita connection, we can say more.

ie derivative

Proposition 1.3.17. *Let (M,g) be an oriented Riemannian manifold and ∇ be the Levi-Civita connection. Let $f \in C^\infty(M)$, then*

$$\nabla^2 f(X,Y) = \langle \nabla_X \operatorname{grad} f, Y\rangle = \frac{1}{2}(\mathfrak{L}_{\operatorname{grad} f}\, g)(X,Y).$$

Proof. Recall that

$$\nabla^2 f(Y,X) = \nabla_X\nabla_Y f - \nabla_{\nabla_X Y} f,$$

therefore $\nabla^2 f$ is symmetric if ∇ is symmetric. Then we observe that

$$\begin{aligned}
\nabla^2 f(X,Y) = \nabla^2 f(Y,X) &= (\nabla_X df)(Y) \\
&= X(df(Y)) - df(\nabla_X Y) \\
&= X\langle \operatorname{grad} f, Y\rangle - \langle \operatorname{grad} f, \nabla_X Y\rangle \\
&= \langle \nabla_X \operatorname{grad} f, Y\rangle
\end{aligned}$$

Now since $\nabla^2 f$ is symmetric, we also have

$$\nabla^2 f(X,Y) = \langle \nabla_Y \operatorname{grad} f, X\rangle.$$

Combine this two equations, we find that

$$\begin{aligned}
(\mathfrak{L}_{\operatorname{grad} f}\, g)(X,Y) &= (\operatorname{grad} f)(g(X,Y)) - g([\operatorname{grad} f, X], Y) - g(X, [\operatorname{grad} f, Y]) \\
&= g(\nabla_{\operatorname{grad} f} X, Y) + g(X, \nabla_{\operatorname{grad} f} Y) - g([\operatorname{grad} f, X], Y) - g(X, [\operatorname{grad} f, Y]) \\
&= g(\nabla_X \operatorname{grad} f, Y) + g(X, \nabla_Y \operatorname{grad} f) \\
&= 2\nabla^2 f(X,Y).
\end{aligned}$$

Therefore the claim follows. $\square$

Proposition 1.3.18. *The musical isomorphisms commute with the total covariant derivative operator: if F is any smooth tensor field with a contravariant i-th index position, and $\flat$ represents the operation of lowering the i th index, then*

$$\nabla(F^\flat) = (\nabla F)^\flat. \tag{3.14}$$

Similarly, if G has a covariant i-th position and $\sharp$ denotes raising the i-th index, then

$$\nabla(G^\sharp) = (\nabla G)^\sharp. \tag{3.15}$$

Proof. Recall that $F^\flat = \operatorname{tr}(F \otimes g)$, where the trace is taken on the i-th and last indices of $F \otimes g$. Because g is parallel, for every vector field X we have $\nabla_X(F \otimes g) = (\nabla_X F) \otimes g$. Because ∇_X commutes with traces, therefore,

$$\nabla_X(F^\flat) = \nabla_X(\operatorname{tr}(F \otimes g)) = \operatorname{tr}((\nabla_X F) \otimes g) = (\nabla_X F)^\flat.$$

This shows that when X is inserted into the last index position on both sides of (3.15), the results are equal. Since X is arbitrary, this proves (3.15).

Because the sharp and flat operators are inverses of each other when applied to the same index position, (3.15) follows by substituting $F = G^\sharp$ into (3.15) and applying $\sharp$ to both sides. □

1.3.3 The exponential map

In what follows, we let (M, g) be a Riemannian or pseudo-Riemannian n-manifold, endowed with its Levi-Civita connection. Corollary 1.2.23 showed that each initial point $p \in M$ and each initial velocity vector $v \in T_pM$ determine a unique maximal geodesic γ_v. To deepen our understanding of geodesics, we need to study their collective behavior, and in particular, to address the following question: How do geodesics change if we vary the initial point or the initial velocity? The dependence of geodesics on the initial data is encoded in a map from the tangent bundle into the manifold, called the exponential map, whose properties are fundamental to the further study of Riemannian geometry.

(It is worth noting that the existence of the exponential map and the basic properties expressed below hold for every connection in TM, not just for the Levi-Civita connection. For simplicity, we restrict attention here to the latter case, because that is all we need. We also restrict to manifolds without boundary, in order to avoid complications with geodesics running into a boundary.)

The next lemma shows that geodesics with proportional initial velocities are related in a simple way.

Lemma 1.3.19 (Rescaling Lemma). *For every $p \in M$, $v \in T_pM$, and $c, t \in \mathbb{R}$,*

$$\gamma_{cv}(t) = \gamma_v(ct), \tag{3.16}$$

whenever either side is defined.

Proof. If $c = 0$, then both sides of (3.16) are equal to p for all $t \in \mathbb{R}$, so we may assume that $c \neq 0$. It suffices to show that $\gamma_{cv}(t)$ exists and (3.16) holds whenever the right-hand side is defined. (The same argument with the substitutions $v = c'v'$, $t = c't'$, and $c = 1/c'$ then implies that the conclusion holds when only the left-hand side is known to be defined.)

Suppose the maximal domain of γ_v is the open interval $I \subseteq \mathbb{R}$. For simplicity, write $\gamma = \gamma_v$, and define a new curve $\widetilde{\gamma} : c^{-1}I \to M$ by $\widetilde{\gamma}(t) = \gamma(ct)$, where $c^{-1}I = \{c^{-1}t : t \in I\}$.

We will show that $\widetilde{\gamma}$ is a geodesic with initial point p and initial velocity cv; it then follows by uniqueness and maximality that it must be equal to γ_{cv}.

It is immediate from the definition that $\widetilde{\gamma}(0) = \gamma(0) = p$. Choose any smooth local coordinates on M and write the coordinate representation of γ as $\gamma(t) = (\gamma^1, \dots, \gamma^n)$, then the chain rule gives

$$\dot{\widetilde{\gamma}}^i(t) = \frac{d}{dt}\gamma^i(ct) = c\dot{\gamma}^i(ct).$$

In particular, it follows that $\widetilde{\gamma}'(0) = c\gamma'(0) = cv$.

Now let D_t and $\widetilde{D}_t$ denote the covariant differentiation operators along γ and $\widetilde{\gamma}$, respectively. Using the chain rule again in coordinates yields

$$\begin{aligned}
\widetilde{D}_t\widetilde{\gamma}'(t) &= \Big(\frac{d}{dt}\dot{\widetilde{\gamma}}^k(t) + \Gamma^k_{ij}(\widetilde{\gamma}(t))\dot{\widetilde{\gamma}}^i(t)\dot{\widetilde{\gamma}}^j(t)\Big)\partial_k \\
&= \Big(c^2\ddot{\gamma}^k(ct) + c^2\Gamma^k_{ij}(\gamma(ct))\dot{\gamma}^i(ct)\dot{\gamma}^j(ct)\Big)\partial_k \\
&= c^2 D_t\gamma'(ct) = 0.
\end{aligned}$$

Thus $\widetilde{\gamma}$ is a geodesic, so $\widetilde{\gamma} = \gamma_{cv}$, as claimed. □

The assignment $v \mapsto \gamma_v$ defines a map from TM to the set of geodesics in M. More importantly, by virtue of the rescaling lemma, it allows us to define a map from (a subset of) the tangent bundle to M itself, which sends each line through the origin in T_pM to a geodesic.

Define a subset $\mathcal{E} \subseteq TM$, the **domain of the exponential map**, by

$$\mathcal{E} = \{v \in TM : \gamma_v \text{ is defined on an interval containing } [0,1]\},$$

and then define the exponential map $\exp : \mathcal{E} \to M$ by

$$\exp(v) = \gamma_v(1).$$

For each $p \in M$, the **restricted exponential map at p**, denoted by $\exp_p$, is the restriction of exp to the set $\mathcal{E}_p = \mathcal{E} \cap T_pM$.

The next proposition describes some essential features of the exponential map. Recall that a subset of a vector space V is said to be star-shaped with respect to a point $x \in S$ if for every $y \in S$, the line segment from x to y is contained in S.

mann exp prop

Proposition 1.3.20 (Properties of the Exponential Map). *Let (M, g) be a Riemannian or pseudo-Riemannian manifold, and let $\exp : \mathcal{E} \to M$ be its exponential map.*

(a) $\mathcal{E}$ is an open subset of TM containing the image of the zero section, and each set $\mathcal{E}_p \subseteq T_pM$ is star-shaped with respect to 0.

(b) For each $v \in TM$, the geodesic γ_v is given by

$$\gamma_v(t) = \exp(tv). \tag{3.17}$$

Riemann exp p

for all t such that either side is defined.

(c) The exponential map is smooth.

(d) For each point $p \in M$, the differential $d(\exp_p)_0 : T_0(T_pM) \cong T_pM \to T_pM$ is the identity map of T_pM, under the usual identification of $T_0(T_pM)$ with T_pM.

Proof. Write $n = \dim M$. The rescaling lemma with $t = 1$ says precisely that $\exp(cv) = \gamma_{cv}(1) = \gamma_v(c)$ whenever either side is defined; this is (b). Moreover, if $v \in \mathcal{E}_p$, then by definition γ_v is defined at least on $[0,1]$. Thus for $0 \le t \le 1$, the rescaling lemma says that

$$\exp_p(tv) = \gamma_{tv}(1) = \gamma_v(t)$$

is defined. This shows that $\mathcal{E}_p$ is star-shaped with respect to 0.

Next we will show that $\mathcal{E}$ is open and $\exp$ is smooth. To do this, recall that in Theorem 1.2.22 we defined a vector field G and showed the geodesics are integral cuvers of G. The importance of G stems from the fact that it actually defines a global vector field on the total space of TM, called the **geodesic vector field**. The key observation, to be proved below, is that G acts on a function $f \in C^\infty(TM)$ by

$$Gf(p,v) = \frac{d}{dt}\Big|_{t=0} f(\gamma_v(t), \gamma_v'(t)). \tag{3.18}$$

To prove that G satisfies (3.18) we choose a chart $(U,(x^i))$ and write the components of the geodesic $\gamma_v(t)$ as $x^i(t)$ and those of its velocity as $v^i(t) = \dot{x}^i(t)$. Using the chain rule and the geodesic equation in the form (2.14), we can write the right-hand side of (3.18) as

$$\begin{aligned}\frac{d}{dt}\Big|_{t=0} f(\gamma_v(t), \gamma_v'(t)) &= \Big(\frac{\partial f}{\partial x^k}(x(t), v(t))\dot{x}^k(t) + \frac{\partial f}{\partial v^k}(x(t), v(t))\dot{v}^k(t)\Big)\Big|_{t=0} \\ &= \frac{\partial f}{\partial x^k}(p,v)v^k - \frac{\partial f}{\partial v^k}(p,v)v^i v^j \Gamma_{ij}^k(p) = Gf(p,v).\end{aligned}$$

The fundamental theorem on flows (Theorem **??**) shows that there exist an open set $\mathcal{D} \subseteq \mathbb{R} \times TM$ containing $\{0\} \times TM$ and a smooth map $\theta : \mathcal{D} \to TM$, such that each curve $\theta^{(p,v)} = \theta(t,(p,v))$ is the unique maximal integral curve of G starting at (p,v), defined on an open interval containing 0. Now suppose $(p,v) \in \mathcal{E}$. This means that the geodesic γ_v is defined at least on the interval $[0,1]$, and therefore so is the integral curve of G starting at $(p,v) \in TM$. Since $(1,(p,v)) \in \mathcal{D}$, there is a neighborhood of $(1,(p,v))$ in $\mathbb{R} \times TM$ on which the flow of G is defined. In particular, this means that there is a neighborhood of (p,v) on which the flow exists for $t \in [0,1]$ (recall the definition of a flow domain) and therefore on which the exponential map is defined. This shows that $\mathcal{E}$ is open.

Since geodesics are projections of integral curves of G under the map $\pi : TM \to M$, it follows that the exponential map can be expressed as

$$\exp_p(v) = \gamma_v(1) = \pi \circ \theta(1,(p,v))$$

wherever it is defined, and therefore $\exp_p(v)$ is a smooth function of (p,v).

To compute $d(\exp_p)_0(v)$ for an arbitrary vector $v \in T_pM$, we just need to choose a curve γ in T_pM starting at 0 whose initial velocity is v, and compute the initial velocity of $\exp_p \circ \tau$. A convenient curve is $\tau(t) = tv$, which yields

$$d(\exp_p)_0(v) = \frac{d}{dt}\Big|_{t=0} (\exp_p \circ \tau)(0) = \frac{d}{dt}\Big|_{t=0} \exp_p(tv) = \frac{d}{dt}\Big|_{t=0} \gamma_v(t) = v.$$

Thus $d(\exp_p)_0$ is the identity map. $\square$

Corollary 1.3.14 on the naturality of geodesics translates into the following important property of the exponential map.

Proposition 1.3.21. *Suppose (M,g) and $(\widetilde{M},\widetilde{g})$ are Riemannian or pseudo-Riemannian manifolds* *and $\varphi : M \to \widetilde{M}$ is a local isometry. Then for every $p \in M$, the following diagram commutes:*

$$\begin{tikzcd}
\mathcal{E}_p \arrow[r, "d\varphi_p"] \arrow[d, "\exp_p"'] & \widetilde{\mathcal{E}}_{\varphi(p)} \arrow[d, "\exp_{\varphi(p)}"] \\
M \arrow[r, "\varphi"] & \widetilde{M}
\end{tikzcd}$$

where $\mathcal{E}_p \subseteq T_pM$ and $\widetilde{\mathcal{E}}_{\varphi(p)} \subseteq T_{\varphi(p)}\widetilde{M}$ are the domains of the restricted exponential maps $\exp_p$ (with respect to g) and $\exp_{\varphi(p)}$ (with respect to $\widetilde{g}$), respectively.

Proof. By Corollary geodesic naturality 1.3.14 if γ is a geodesic of M, then $\varphi \circ \gamma$ is a geodesic of $\widetilde{M}$. With this, everything is clear. □

An important consequence of the naturality of the exponential map is the following proposition, which says that local isometries of connected manifolds are completely determined by their values and differentials at a single point.

Proposition 1.3.22. *Let (M, g) and $(\widetilde{M}, \widetilde{g})$ be Riemannian or pseudo-Riemannian manifolds, with M connected. Suppose $\varphi : M \to \widetilde{M}$ are local isometries such that for some point $p \in M$, we have $\varphi(p) = \psi(p)$ and $d\varphi_p = d\psi_p$. Then $\varphi \equiv \psi$.*

Proof. Define a set

$$S = \{p \in M : \varphi(p) = \psi(p) \text{ and } d\varphi_p = d\psi_p\}.$$

It is clear that S is closed by continuity, and is nonempty by our assumption. Now let $p \in S$, by Proposition Riemann exp prop 1.3.20 we know that there exists $\varepsilon > 0$ such that $\exp_p$ and $\exp_{\varphi(p)}$ are diffeomorphisms on $B_\varepsilon(0)$. Since $d\varphi_p$ is an isometry, we have $d\varphi_p(B_\varepsilon(0)) = B_\varepsilon(0)$, and thus by the naturality of the exponential map, we know that

$$\begin{aligned}
\varphi|_{\exp_p(B_\varepsilon(0))} &= \left((\exp_{\varphi(p)})|_{B_\varepsilon(0)}\right) \circ d\varphi_p \circ \left((\exp_p)|_{B_\varepsilon(0)}\right)^{-1} \\
&= \left((\exp_{\varphi(p)})|_{B_\varepsilon(0)}\right) \circ d\psi_p \circ \left((\exp_p)|_{B_\varepsilon(0)}\right)^{-1} \\
&= \psi|_{\exp_p(B_\varepsilon(0))}
\end{aligned}$$

By differentiating this equation we get $d\varphi_q = d\psi_q$ for $q \in \exp_p(B_\varepsilon(0))$, so $\exp_p(B_\varepsilon(0)) \subseteq S$. This proves S is open, and by connectivity we then get $S = M$. □

A Riemannian or pseudo-Riemannian manifold (M, g) is said to be **geodesically complete** if every maximal geodesic is defined for all $t \in \mathbb{R}$, or equivalently if the domain of the exponential map is all of TM. It is easy to construct examples of manifolds that are not geodesically complete; for example, in every proper open subset of $\mathbb{R}^n$ with its Euclidean metric or with a pseudo-Euclidean metric, there are geodesics that reach the boundary in finite time. Similarly, on $\mathbb{R}^n$ with the metric $(\sigma^{-1})^*\mathring{g}$ obtained from the sphere by stereographic projection, there are geodesics that escape to infinity in finite time. Geodesically complete manifolds are the natural setting for global questions in Riemannian or pseudo-Riemannian geometry, and most of our attention will be focused on them later.

1.3.4 Normal neighborhoods and normal coordinates

We continue to let (M, g) be a Riemannian or pseudo-Riemannian manifold of dimension n (without boundary). Recall that for every $p \in M$, the restricted exponential map $\exp_p$

maps the open subset $\mathcal{E}_p \subseteq T_pM$ smoothly into M. Because $d(\exp_p)_0$ is invertible, the inverse function theorem guarantees that there exist a neighborhood V of the origin in T_pM and a neighborhood U of p in M such that $\exp_p : V \to U$ is a diffeomorphism. A neighborhood U of $p \in M$ that is the diffeomorphic image under $\exp_p$ of a star-shaped neighborhood of $0 \in T_pM$ is called a **normal neighborhood** of p.

Every orthonormal basis (e_i) for T_pM determines a basis isomorphism $B : \mathbb{R}^n \to T_pM$ by $B(x^1, \ldots, x^n) = x^i e_i$. If $U = \exp_p(V)$ is a normal neighborhood of p, we can combine this isomorphism with the exponential map to get a smooth coordinate map $\varphi = B^{-1} \circ (\exp_p|_V)^{-1} : U \to \mathbb{R}^n$

$$\begin{tikzcd}
T_pM \arrow[r, "B^{-1}"] & \mathbb{R}^n \\
U \arrow[u, "(\exp_p|_V)^{-1}"] \arrow[ur, "\varphi"'] &
\end{tikzcd}$$

Such coordinates are called (**Riemannian** or **pseudo-Riemannian**) **normal coordinates centered at** $\boldsymbol{p}$.

dinate unique **Proposition 1.3.23 (Uniqueness of Normal Coordinates).** *Let (M, g) be a Riemannian or pseudo-Riemannian n-manifold, p a point of M, and U a normal neighborhood of p. For every normal coordinate chart on U centered at p, the coordinate basis is orthonormal at p; and for every orthonormal basis (e_i) for T_pM, there is a unique normal coordinate chart (x^i) on U such that $\partial_i|_p = e_i$ for $i = 1, \ldots, n$. In the Riemannian case, any two normal coordinate charts (x^i) and $(\widetilde{x}^j)$ are related by*

$$\widetilde{x}^j = A_i^j x^i \tag{3.19}$$ Riemann norma

for some (constant) matrix $A_i^j \in \mathrm{O}(n)$.

Proof. Let φ be a normal coordinate chart on U centered at p, with coordinate functions (X^i). By definition, this means that $\varphi = B^{-1} \circ (\exp_p)^{-1}$, where $B : \mathbb{R}^n \to T_pM$ is the basis isomorphism determined by some orthonormal basis (e_i) for T_pM. Note that $(d\varphi_p)^{-1} = d(\exp_p)_0 \circ dB_0 = B$ because $d(\exp_p)_0$ is the identity and B is linear. Thus $\partial_i|_p = (d\varphi_p)^{-1}(\partial_i|_0) = B(\partial_i|_0) = e_i$, which shows that the coordinate basis is orthonormal at p. Conversely, every orthonormal basis (e_i) for T_pM yields a basis isomorphism B and thus a normal coordinate chart $\varphi = B^{-1} \circ (\exp_p)^{-1}$, which satisfies $\partial_i|_p = e_i$ by the computation above.

If $\widetilde{\varphi} = \widetilde{B}^{-1} \circ (\exp_p)^{-1}$ is another such chart, then

$$\widetilde{\varphi} \circ \varphi^{-1} = \widetilde{B}^{-1} \circ (\exp_p)^{-1} \circ (\exp_p) \circ B = \widetilde{B}^{-1} \circ B,$$

Riemann normal coordinate unique-1

which is a linear isometry of $\mathbb{R}^n$ and therefore has the form (3.19) in terms of standard coordinates on $\mathbb{R}^n$. Since $(\widetilde{x}^j)$ and (x^i) are the same coordinates if and only if (A_i^j) is the identity matrix, this shows that the normal coordinate chart associated with a given orthonormal basis is unique. $\square$

ordinate prop **Proposition 1.3.24 (Properties of Normal Coordinates).** *Let (M, g) be a Riemannian or pseudo-Riemannian n-manifold, and let $(U, (x^i))$ be any normal coordinate chart centered at $p \in M$.*

(a) *The coordinates of p are $(0, \ldots, 0)$.*

(b) *The components of the metric at p are $g_{ij} = \delta_{ij}$ if g is Riemannian, and $g_{ij} = \pm\delta_{ij}$ otherwise.*

(c) For every $v = v^i \partial_i \in T_pM$, the geodesic γ_v starting at p with initial velocity v is represented in normal coordinates by the line

$$\gamma_v(t) = (tv^1, \ldots, tv^n) \tag{3.20}$$

as long as t is in some interval I containing 0 such that $\gamma_v(I) \subseteq U$.

(d) The Christoffel symbols in these coordinates vanish at p.

(e) All of the first partial derivatives of g_{ij} in these coordinates vanish at p.

Proof. Part (a) follows directly from the definition of normal coordinates, and parts (b) and (c) follow from Propositions 1.3.23 and 1.3.20(b), respectively.

To prove (d), let $v = v^i \partial_i|_p \in T_pM$ be arbitrary. The geodesic equation (2.13) for $\gamma_v(t) = (tv^1, \ldots, tv^n)$ simplifies to

$$\Gamma^k_{ij}(tv)v^i v^j = 0.$$

Evaluating this expression at $t = 0$ shows that $\Gamma^k_{ij}(0)v^i v^j = 0$ for every index k and every vector v. In particular, with $v = \partial_a$ for some fixed a, this shows that $\Gamma^k_{aa} = 0$ for each a and k (no summation). Substituting $v = \partial_a + \partial_b$ and $v = \partial_a - \partial_b$ for any fixed pair of indices a and b and subtracting, we conclude also that Γ^k_{ab} at p for all a, b, k. Finally, (e) follows from (d) together with (3.2) in the case $E_k = \partial_k$. $\square$

Because they are given by the simple formula (3.20), the geodesics starting at p and lying in a normal neighborhood of p are called **radial geodesics**.

1.3.5 Tubular neighborhoods and Fermi coordinates

The exponential map and normal coordinates give us a good understanding of the behavior of geodesics starting a point. Now we generalize those constructions to geodesics starting on any embedded submanifold. We restrict attention to the Riemannian case, because we will be using the Riemannian distance function.

Suppose (M, g) is a Riemannian manifold, $P \subseteq M$ is an embedded submanifold, and $\pi : NP \to P$ is the normal bundle of P in M. Let $\mathcal{E} \subseteq TM$ denote the domain of the exponential map of M, and let $\mathcal{E}_P = \mathcal{E} \cap NP$. Let $E : \mathcal{E}_P \to M$ denote the restriction of $\exp$ (the exponential map of M) to $\mathcal{E}_P$. We call E the **normal exponential map** of P in M.

A **normal neighborhood of P in M** is an open subset $U \subseteq M$ that is the diffeomorphic image under E of an open subset $V \subseteq \mathcal{E}_P$ whose intersection with each fiber N_xP is star-shaped with respect to 0. We will be primarily interested in normal neighborhoods of the following type: a normal neighborhood of P in M is called a **tubular neighborhood** if it is the diffeomorphic image under E of a subset $V \subseteq \mathcal{E}_P$ of the form

$$V = \{(x, v) \in NP : |v|_g < \delta(x)\}, \tag{3.21}$$

for some positive continuous function $\delta : P \to \mathbb{R}$. If U is the diffeomorphic image of such a set V for a constant function $\delta \equiv \varepsilon$, then it is called a **uniform tubular neighborhood of radius ε**, or an **ε-tubular neighborhood**.

Theorem 1.3.25 (Tubular Neighborhood Theorem). *Let (M, g) be a Riemannian manifold. Every embedded submanifold of M has a tubular neighborhood in M, and every compact submanifold has a uniform tubular neighborhood.*

Proof. Let $P \subseteq M$ be an embedded submanifold, and $P_0 \subseteq NP$ be the zero section. We begin by showing that E is a local diffeomorphism on a neighborhood of P_0. By the inverse function theorem, it suffices to show that the differential $dE_{(x,0)}$ is bijective at each point $(x,0) \in P_0$. This follows easily from the following two facts:

- The restriction of E to P_0 is the obvious diffeomorphism $P_0 \to M$, so $dE_{(x,0)}$ maps the subspace $T_{(x,0)}P_0 \subseteq T_{(x,0)}NP$ isomorphically onto T_xP.

- The restriction of E to the fiber N_xP is the restricted map $\exp_x$, which is a diffeomorphism at 0, so $dE_{(x,0)}$ maps $T_{(x,0)}(N_xP)$ isomorphically onto N_xP.

Since $T_xM = T_xP \oplus N_xP$, these together show that $dE_{(x,0)}$ is surjective, and hence is bijective for dimensional reasons. Thus, E is a diffeomorphism on a neighborhood of $(x,0)$ in NP, which we can take to be of the form $V_\delta(x) = \{(x',v') \in NM : d_g(x,x') < \delta, |v'|_g < \delta\}$ for some $\delta > 0$.

To complete the proof, we need to show that there is an open subset V of the form (Riemann tubular neig 3.21) on which E is a global diffeomorphism. For each point $x \in P$, let

$$\rho(x) = \sup\{\delta < 1 : E \text{ is a diffeomorphism from } V_\delta(x) \text{ to its image}\}.$$

The argument in the preceding paragraph implies that $\rho : P \to \mathbb{R}$ is positive.

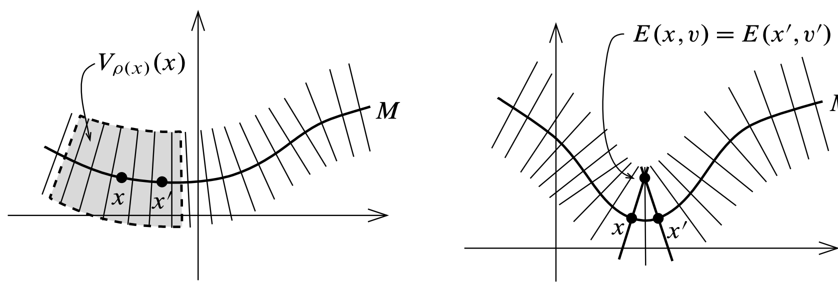

Figure 1.1: Continuity of ρ and Injectivity of E.

To show it is continuous, let $x, x' \in P$ be arbitrary, and suppose first that $d_g(x,x') < \rho(x)$. Then by the triangle inequality, $V_\delta(x')$ is contained in $V_{\rho(x)}(x)$ for $\delta = \rho(x) - d_g(x,x')$, which implies that $\rho(x') \geq \rho(x) - d_g(x,x')$, or

$$\rho(x) - \rho(x') \leq d_g(x,x').$$

On the other hand, if $d_g(x,x') \geq \rho(x)$, then the inequality above holds for trivial reasons. Reversing the roles of x and x' yields an analogous inequality, which shows that $|\rho(x) - \rho(x')| \leq d_g(x,x')$, so ρ is continuous.

Now let $V = \{(x,v) \in NP : |v|_g < \rho(x)/2\}$. We will show that E is injective on V. Suppose that (x,v) and (x',v') are points in V such that $E(x,v) = E(x',v')$. Assume without loss of generality that $\rho(x) \leq \rho(x')$. Because $\exp_x(v) = \exp_{x'}(v')$, there is an admissible curve from x to x' of length $|v|_g + |v'|_g$ (recall that the velocity of a geodesic is constant (Corollary metric connection geodesic velocity constant 1.3.5)), and thus

$$d_g(x,x') \leq |v|_g + |v'|_g \leq \frac{1}{2}\rho(x) + \frac{1}{2}\rho(x') \leq \rho(x).$$

Therefore, both (x,v) and (x',v') are in $V_{\rho(x)}(x)$. Since (x,v) and (x',v') are then in $V_\delta(x)$ for some $\delta < \rho(x)$ and E is injective on $V_\delta(x)$, this implies $(x,v) = (x',v')$.

The set $U = E(V)$ is open in M because $E|_V$ is a local diffeomorphism and thus an open map. It follows that $E : V \to U$ is a smooth bijection and a local diffeomorphism, hence a diffeomorphism by Proposition local diff prop **??**. Therefore, U is a tubular neighborhood of P.

Finally, if P is compact, then the continuous function $1/2\rho$ achieves a minimum value $\varepsilon > 0$ on P, so U contains a uniform tubular neighborhood of radius ε. □

Fermi coordinates

Now we will construct coordinates on a tubular neighborhood that are analogous to Riemannian normal coordinates around a point. Let P be an embedded k-dimensional submanifold of a Riemannian n-manifold (M, g), and let $U \subseteq M$ be a normal neighborhood of P, with $U = E(V)$ for some appropriate open subset $V \subseteq NP$.

Let (W_0, ψ) be a smooth coordinate chart for P, and let $(E_1, \dots, E_{n-k})$ be a local orthonormal frame for the normal bundle NP; by shrinking W_0 if necessary, we can assume that the coordinates and the local frame are defined on the same open subset $W_0 \subseteq P$. Let $\widehat{W}_0 = \psi(W_0) \subseteq \mathbb{R}^k$, and let $NP|_{W_0}$ be the portion of the normal bundle over W_0. The coordinate map and frame (E_j) yield a diffeomorphism $B : \widehat{W}_0 \times \mathbb{R}^{n-k} \to NP|_{W_0}$ defined by

$$B(x^1, \dots, x^k, v^1, \dots, v^{n-k}) = (p, v^1 E_1|_p + \cdots + v^{n-k} E_{n-k}|_p),$$

where $p = \psi^{-1}(x^1, \dots, x^k)$. Let $V_0 = V \cap NP|_{W_0} \subseteq NP$ and $U_0 = E(V_0) \subseteq M$, and define a smooth coordinate map $\varphi : U_0 \to \mathbb{R}^n$ by $\varphi = B^{-1} \circ (E|_{V_0})^{-1}$:

$$\begin{tikzcd} NP|_{W_0} \arrow[r, "B^{-1}"] & \widehat{W}_0 \times \mathbb{R}^{n-k} \subseteq \mathbb{R}^n \\ U_0 \arrow[u, "(E|_{V_0})^{-1}"] \arrow[ur, "\varphi"'] & \end{tikzcd}$$

The coordinate map can also be written

$$\varphi : (p, v^1 E_1|_p + \cdots + v^{n-k} E_{n-k}|_p) \mapsto (x^1(p), \dots, x^k(p), v^1, \dots, v^{n-k}).$$

Coordinates of this form are called **Fermi coordinates**. Here is the analogue of Proposition Riemann normal coordinate prop 1.3.24 for Fermi coordinates.

ordinate prop **Proposition 1.3.26 (Properties of Fermi Coordinates).** *Let P be an embedded k-dimensional submanifold of a Riemannian n-manifold (M, g), let U be a normal neighborhood of P in M, and let $(x^1, \dots, x^k, v^1, \dots, v^{n-k})$ be Fermi coordinates on an open subset $U_0 \subseteq U$. For convenience, we also write $x^{k+j} = v^j$ for $j = 1, \dots, n-k$.*

(a) $P \cap U_0$ is the set of points where $x^{k+1} = \cdots = x^n = 0$.

(b) At each point $p \in P \cap U_0$, the metric components satisfy the following:

$$g_{ij} = g_{ji} = \begin{cases} 0, & 1 \le i \le k \text{ and } k+1 \le j \le n; \\ \delta_{ij}, & k+1 \le i, j \le n. \end{cases}$$

(c) For every $p \in P \cap U_0$ and $v = v^1 E_1|_p + \cdots + v^{n-k} E_{n-k}|_p \in N_pP$, the geodesic γ_v starting at p with initial velocity v is the curve with coordinate expression

$$\gamma_v(t) = (x^1(p), \dots, x^k(p), tv^1, \dots, tv^{n-k}).$$

(d) *At each $p \in P \cap U_0$, the Christoffel symbols in these coordinates satisfy $\Gamma^k_{ij} = 0$, provided $k+1 \leq i, j \leq n$.*

(e) *At each $p \in P \cap U_0$, the partial derivatives $\partial_i g_{jk}(p)$ vanish for $k+1 \leq i, j, k \leq n$.*

Proof. Part (a) and (b) follow directly from the definition of Fermi coordinates, and (c) follows from Proposition 1.3.20(b).

To prove (d), let $v = v^i E_i|_p \in N_p P$ be arbitrary. The geodesic equation (2.13) for $\gamma_v(t) = (tv^1, \ldots, tv^n)$ simplifies to

$$\Gamma^k_{ij}(x^1(p), \ldots, x^k(p), tv^1, \ldots, tv^{n-k})v^i v^j = 0.$$

Evaluating this expression at $t = 0$ shows that $\Gamma^k_{ij}(p)v^i v^j = 0$ for every index k and every vector v. In particular, with $v = \partial_a$ for some fixed a, this shows that $\Gamma^k_{aa} = 0$ for each a and k (no summation). Substituting $v = \partial_a + \partial_b$ and $v = \partial_a - \partial_b$ for any fixed pair of indices a and b and subtracting, we conclude also that Γ^k_{ab} at p for all a, b, k. Finally, (e) follows from (d) together with (3.2) in the case $E_k = \partial_k$. □

1.3.6 Geodesics of the model spaces

Euclidean Space

On $\mathbb{R}^n$ with the Euclidean metric, Proposition 1.3.7 shows that the Levi-Civita connection is the Euclidean connection. Therefore, as one would expect, constant coefficient vector fields are parallel, and the Euclidean geodesics are straight lines with constant-speed parametrizations. Every Euclidean space is geodesically complete.

Spheres

Because the round metric on the sphere $S^n(R)$ is induced by the Euclidean metric on $\mathbb{R}^{n+1}$, it is easy to determine the geodesics on a sphere using Corollary 1.3.2. Define a great circle on $S^n(R)$ to be any subset of the form $S^n(R) \cap \Pi$, where $\Pi \subseteq \mathbb{R}^{n+1}$ is a 2-dimensional linear subspace.

Proposition 1.3.27. *A nonconstant curve on $S^n(R)$ is a maximal geodesic if and only if it is a periodic constant-speed curve whose image is a great circle. Thus every sphere is geodesically complete.*

Proof. Let $p \in S^n(R)$ be arbitrary. Because $f(x) = |x|^2$ is a defining function for $S^n(R)$, a vector $v \in T_p\mathbb{R}^{n+1}$ is tangent to $S^n(R)$ if and only if $df_p(v) = 2\langle v, p\rangle = 0$, where we think of v as a vector by means of the usual identification of $\mathbb{R}^{n+1}$ with $T_p\mathbb{R}^{n+1}$. Thus $T_pS^n(R)$ is exactly the set of vectors orthogonal to p.

Suppose v is an arbitrary nonzero vector in $T_pS^n(R)$. Let $a = |v|/R$ and $\hat{v} = v/a$ (so $\hat{v} = R$), and consider the smooth curve $\gamma : \mathbb{R} \to \mathbb{R}^{n+1}$ given by

$$\gamma(t) = (\cos at)p + (\sin at)\hat{v}.$$

By direct computation, $|\gamma(t)|^2 = R^2$, so $\gamma(t) \in S^n(R)$ for all t. Moreover

$$\gamma'(t) = -a(\sin at)p + a(\cos at)\hat{v}, \quad \gamma''(t) = -a^2(\cos at)p - a^2(\sin at)\hat{v}.$$

Because $\gamma''(t)$ is proportional to $\gamma(t)$ (thinking of both as vectors in $\mathbb{R}^{n+1}$), it follows that $\gamma''(t)$ is $\bar{g}$-orthogonal to $T_pS^n(R)$, so γ is a geodesic in $S^n(R)$ by Corollary 1.3.2. Since $\gamma(0) = p$ and $\gamma(0) = a\hat{v} = v$, it follows that $\gamma = \gamma_v$.

Each γ_v is periodic of period $2\pi/a$, and has constant speed by Corollary 1.3.5 (or by direct computation). The image of γ_v is the great circle formed by the intersection of $S^n(R)$ with the linear subspace spanned by $\{p, \hat{v}\}$.

Conversely, suppose that C is a great circle formed by intersecting $S^n(R)$ with a two dimensional subspace Π, and let $\{v, w\}$ be an orthonormal basis for Π. Then C is the image of the geodesic with initial point $p = Rw$ and initial velocity v. □

Hyperbolic spaces

The geodesics of hyperbolic spaces can be determined by an analogous procedure using the hyperboloid model.

Proposition 1.3.28. *A nonconstant curve in a hyperbolic space is a maximal geodesic if and only if it is a constant-speed embedding of $\mathbb{R}$ whose image is one of the following:*

*(a) Hyperboloid Model: The intersection of $\mathbb{H}^n(R)$ with a 2-dimensional linear subspace of $\mathbb{R}^{n,1}$, called a **great hyperbola**.*

(b) Beltrami-Klein: The interior of a line segment whose endpoints both lie on $\partial\mathbb{K}^n(R)$.

(c) Ball Model: The interior of a diameter of $\mathbb{B}^n(R)$, or the intersection of $\mathbb{B}^n(R)$ with a Euclidean circle that intersects $\partial\mathbb{B}^n(R)$ orthogonally.

(d) Half-Space Model: The intersection of $\mathbb{U}^n(R)$ with one of the following: a line parallel to the y-axis or a Euclidean circle with center on $\partial\mathbb{U}^n(R)$.

Every hyperbolic space is geodesically complete.

Proof. We begin with the hyperboloid model, for which the proof is formally quite similar to what we just did for the sphere. Since the Riemannian connection on $\mathbb{H}^n(R)$ is equal to the tangential connection by Proposition 1.3.12, it follows from Corollary 1.3.2 that a smooth curve $\gamma : I \to \mathbb{H}^n(R)$ is a geodesic if and only if its acceleration $\gamma''(t)$ is everywhere $\bar{q}$-orthogonal to $T_{\gamma(t)}\mathbb{H}^n(R)$ (where $\bar{q} = \bar{q}^{n,1}$ is the Minkowski metric)

Let $p \in \mathbb{H}^n(R)$ be arbitrary. Note that $f(x) = \bar{q}(x,x)$ is a defining function for $\mathbb{H}^n(R)$, and (1.1) shows that the gradient of f at p is equal to $2p$ (where we regard p as a vector in $T_p\mathbb{R}^{n,1}$ as before). It follows that a vector $v \in T_p\mathbb{R}^{n,1}$ is tangent to $\mathbb{H}^n(R)$ if and only if $\bar{q}(p,v) = 0$. Let $v \in T_p\mathbb{H}^n(R)$ be an arbitrary nonzero vector. Put $a = |v|_{\bar{q}}/R = \bar{q}(v,v)^{1/2}/R$ and $\hat{v} = v/a$, and define $\gamma : \mathbb{R} \to \mathbb{R}^{n,1}$ by

$$\gamma(t) = (\cosh at)p + (\sinh at)\hat{v}.$$

Direct computation shows that γ takes its values in $\mathbb{H}^n(R)$ and that its acceleration vector is everywhere proportional to $\gamma(t)$. Thus $\gamma''(t)$ is $\bar{q}$-orthogonal to $T_{\gamma(t)}\mathbb{H}^n(R)$, so γ is a geodesic in $\mathbb{H}^n(R)$ and therefore has constant speed. Because it satisfies the initial conditions $\gamma(0) = p$ and $\gamma'(t) = a\hat{v} = v$, it is equal to γ_v. Note that γ_v is a smooth embedding of $\mathbb{R}$ into $H^n(R)$ whose image is the great hyperbola formed by the intersection between $\mathbb{H}^n(R)$ and the plane spanned by $\{p, \hat{v}\}$.

Conversely, suppose Π is any 2-dimensional linear subspace of $\mathbb{R}^{n,1}$ that has nontrivial intersection with $\mathbb{H}^n(R)$. Choose $p \in \Pi \cap \mathbb{H}^n(R)$, and let v be another nonzero vector in Π that is $\bar{q}$-orthogonal to p, which implies $v \in T_p\mathbb{H}^n(R)$. Using the computation above, we see that the image of the geodesic γ_v is the great hyperbola formed by the intersection of Π with $\mathbb{H}^n(R)$.

Before considering the other three models, note that since maximal geodesics in $\mathbb{H}^n(R)$ are constant-speed embeddings of R, it follows from naturality that maximal geodesics in each of the other models are also constant-speed embeddings of R. Thus each model is geodesically complete, and to determine the geodesics in the other models we need only determine their images.

Consider the BeltramiKlein model. Recall the isometry $c : \mathbb{H}^n(R) \to \mathbb{K}^n(R)$ given by $c(\xi, \tau) = R\xi/\tau$ (see (1.4)). The image of a maximal geodesic in $\mathbb{H}^n(R)$ is a great hyperbola, which is the set of points $(\xi, \tau) \in \mathbb{H}^n(R)$ that solve a system of $n - 1$ independent linear equations. Simple algebra shows that (ξ, τ) satisfies a linear equation $\alpha_i \xi^i + \beta\tau = 0$ if and only if $w = c(\xi, \tau) = R\xi/\tau$ satisfies the affine equation $\alpha_i \xi^i = -\beta R$. Thus c maps each great hyperbola onto the intersection of $\mathbb{K}^n(R)$ with an affine subspace of $\mathbb{R}^n$, and since it is the image of a smooth curve, it must be the intersection of $\mathbb{K}^n(R)$ with a straight line.

Next consider the Poincaré ball model. First consider the 2-dimensional case, and recall the inverse hyperbolic stereographic projection $\pi^{-1} : \mathbb{B}^2(R) \to \mathbb{H}^2(R)$ constructed in Theorem 1.1.6:

$$\pi^{-1}(u) = (\xi, \tau) = \Big(\frac{2R^2 u}{R^2 - |u|^2}, R\frac{R^2 + |u|^2}{R^2 - |u|^2}\Big).$$

In this case, a great hyperbola is the set of points on $\mathbb{H}^2(R)$ that satisfy a single linear equation $\alpha_i \xi^i + \beta\tau = 0$. In the special case $\beta = 0$, this hyperbola is mapped by π to a straight line segment through the origin, as can easily be seen from the geometric definition of π. If $\beta \neq 0$, we can assume (after multiplying through by a constant if necessary) that $\beta = -1$, and write the linear equation as $\tau = \alpha_i \xi^i = \alpha \cdot \xi$ (where the dot represents the Euclidean dot product between elements of $\mathbb{R}^2$). Under π^{-1}, this pulls back to the equation

$$R\frac{R^2 + |u|^2}{R^2 - |u|^2} = \frac{2R^2 \alpha \cdot u}{R^2 - |u|^2}$$

on the disk, which simplifies to

$$|u|^2 - 2R\alpha \cdot u + R^2 = 0.$$

Completing the square, we can write this as

$$|u - R\alpha|^2 = R^2(|\alpha|^2 - 1) \tag{3.22}$$

If $|\alpha|^2 \leq 1$, this locus is either empty or a point on $\partial\mathbb{B}^2(R)$, so it contains no points in $\mathbb{B}^2(R)$. Since we are assuming that it is the image of a maximal geodesic, we must therefore have $|\alpha|^2 > 1$. In that case, (3.22) is the equation of a circle with center R. and radius $R\sqrt{|\alpha|^2 - 1}$. At a point u_0 where the circle intersects $\partial\mathbb{B}^2(R)$, the three points 0, u_0, and $R\alpha$ form a triangle with sides $|u_0| = R$, $|R\alpha|$, and $|u_0 - R\alpha|$, which satisfy the Pythagorean identity by (3.22); therefore the circle meets $\partial\mathbb{B}^2(R)$ in a right angle.

In the higher-dimensional case, a geodesic on $\mathbb{H}^n(R)$ is determined by a 2-plane. If the 2-plane contains the point $(0, \ldots, 0, R)$, then the corresponding geodesic on $\mathbb{B}^n(R)$ is a line through the origin as before. Otherwise, we can use an orthogonal transformation in the $(\xi^1, \ldots, \xi^n)$ variables (which preserves $\breve{g}_R$) to move this 2-plane so that it lies in the (ξ^1, ξ^2, τ) subspace, and then we are in the same situation as in the 2-dimensional case.

Finally, consider the upper half-space model. The 2-dimensional case is easiest to analyze using complex notation. Recall the complex formula for the Möbious transform $\kappa : \mathbb{U}^2(\mathbb{R}) \to \mathbb{B}^2(R)$ given in Theorem 1.1.6:

$$\kappa(z) = w = iR\frac{z - iR}{z + iR}.$$

Substituting this into equation (5.26) and writing $w = u + iv$ and $\alpha = a + ib$ in place of $u = (u^1, u^2)$, $\alpha = (\alpha^1, \alpha^2)$ we get

$$R^2\frac{|z - iR|^2}{|z + iR|^2} - iR^2\overline{\alpha}\frac{z - iR}{z + iR} + iR^2\alpha\frac{\overline{z} + iR}{\overline{z} - iR} + R^2|\alpha|^2 = R^2(|\alpha|^2 - 1).$$

Multiplying through by $(z + iR)(\overline{z} - iR)/2R^2$ and simplifying yields

$$(1 - b)|z|^2 - 2aRx + (b + 1)R^2 = 0.$$

This is the equation of a circle with center on the x-axis, unless $b = 1$, in which case the condition $|\alpha|^2 > 1$ forces $a \neq 0$, and then it is a straight line $x =$ constant. The other class of geodesics on the ball, line segments through the origin, can be handled similarly.

In the higher-dimensional case, suppose first that $\gamma : \mathbb{R} \to \mathbb{U}^n(R)$ is a maximal geodesic such that $\gamma(0)$ lies on the y-axis and $\gamma'(0)$ is in the span of $\{\partial/\partial x^1, \partial/\partial y\}$. From the explicit formula (1.6) for κ, it follows that $\kappa \circ \gamma(0)$ lies on the v-axis in the ball, and $(\kappa \circ \gamma)'(0)$ is in the span of $\{\partial/\partial u^1, \partial/\partial v\}$. The image of the geodesic $\kappa \circ \gamma$ is either part of a line through the origin or an arc of a circle perpendicular to $\partial\mathbb{B}^n(R)$, both of which are contained in the (u^1, v)-plane. By the argument in the preceding paragraph, it then follows that the image of γ is contained in the (x^1, y)-plane and is either a vertical half-line or a semicircle centered on the $y = 0$ hyperplane. For the general case, note that translations and orthogonal transformations in the x-variables preserve vertical half-lines and circles centered on the $y = 0$ hyperplane in $\mathbb{U}^n(R)$, and they also preserve the metric $\breve{g}_R$. Given an arbitrary maximal geodesic $\gamma : \mathbb{R} \to \mathbb{U}^n(R)$, after applying an x-translation we may assume that $\gamma(0)$ lies on the y-axis, and after an orthogonal transformation in the x variables, we may assume that $\gamma'(0)$ is in the span of $\{\partial/\partial x^1, \partial/\partial y\}$; then the argument above shows that the image of γ is either a vertical half-line or a semicircle centered on the $y = 0$ hyperplane. □

1.3.7 Exercise

Exercise 1.3.1. Suppose (M, g) is a Riemannian manifold, and let div and ∇ be the divergence and Laplace operators.

(a) Show that for every vector field $X \in \mathfrak{X}(M)$, $\mathrm{div}X$ can be written in terms of the total covariant derivative as $\mathrm{div}X = \mathrm{tr}(\nabla X)$, and that if $X = X_i E_i$ in terms of some local frame, then $\mathrm{div}X = X^i_{;i}$.

(b) Show that the Laplace operator acting on a smooth function f can be expressed as

$$\Delta f = \mathrm{tr}_g(\nabla^2 f) \tag{3.23}$$

and in terms of any local frame,

$$\Delta f = g^{ij} f_{;ij} = f_{;i}{}^{i}. \tag{3.24}$$

Proof. Let $E_1, \ldots, E_n$ be a local orthonormal frame and dV_g the Riemannian volume form, then by definition and Proposition **??**:

$$(\mathrm{div}X)dV_g = (d \circ i_X)(dV_g) = (d \circ i_X + i_X \circ d)(dV_g) = \mathfrak{L}_X(dV_g).$$

This then implies

$$\begin{aligned}
\operatorname{div} X &= (\operatorname{div} X\, dV_g)(E_1,\dots,E_n) = (\mathfrak{L}_X dV_g)(E_1,\dots,E_n)\\
&= X(dV_g(E_1,\dots,E_n)) - \sum_{i=1}^n dV_g(E_1,\dots,\nabla_X E_i,\dots,E_n)\\
&= -\sum_{i=1}^n \langle \mathfrak{L}_X E_i, E_i\rangle = \frac{1}{2}\sum_{i=1}^n (\mathfrak{L}_X g)(E_i,E_i).
\end{aligned}$$

Now we turn to $\operatorname{tr}(\nabla X)$. We have

$$\begin{aligned}
\operatorname{tr}(\nabla X) &= \sum_{i=1}^n \langle (\nabla X)(E_i), E_i\rangle = \sum_{i=1}^n \langle \nabla_{E_i} X, E_i\rangle\\
&= \frac{1}{2}\sum_{i=1}^n \mathfrak{L}_X g(E_i,E_i) + \frac{1}{2}\sum_{i=1}^n dX^\flat(E_i,E_i) = \frac{1}{2}\sum_{i=1}^n \mathfrak{L}_X g(E_i,E_i),
\end{aligned}$$

Riemann Levi-Civita Lie exterior

where we use Proposition 1.3.10 and that $dX^\flat$ is antisymmetric. This gives (a).

Now for part (b), we have

$$\Delta f = \operatorname{div}(\operatorname{grad} f) = \operatorname{tr}(\nabla(\operatorname{grad} f)) = \operatorname{tr}(\nabla(df)^\sharp) = \operatorname{tr}((\nabla df)^\sharp) = \operatorname{tr}_g(\nabla df) = \operatorname{tr}_g(\nabla^2 f).$$

The coordinate expression then follows. □

1.4 Geodesics and distance

Most of the results of this section do not apply to general pseudo-Riemannian metrics, at least not without substantial modification. For this reason, we restrict our focus here to the Riemannian case. Also, the theory of minimizing curves becomes considerably more complicated in the presence of a nonempty boundary; thus, unless otherwise stated, throughout this chapter we assume that (M,g) is a Riemannian manifold without boundary. And because we will be using the Riemannian distance function, we assume for most results that M is connected.

1.4.1 Geodesics and minimizing curves

Let (M,g) be a Riemannian manifold. An admissible curve γ in M is said to be a minimizing curve if $L_g(\gamma) \le L_g(\widetilde{\gamma})$ for every admissible curve $\widetilde{\gamma}$ with the same endpoints. When M is connected, it follows from the definition of the Riemannian distance that γ is minimizing if and only if $L_g(\gamma)$ is equal to the distance between its endpoints.

Our first goal is to show that all minimizing curves are geodesics. To do so, we will think of the length function L_g as a functional on the set of all admissible curves in M with fixed starting and ending points. Our project is to search for minima of this functional. It suffices to note that if γ is a minimizing curve, and $\{\Gamma_s : s \in (-\varepsilon,\varepsilon)\}$ is a one-parameter family of admissible curves with the same endpoints such that $L_g(\Gamma_s)$ is a differentiable function of s and $\Gamma_0 = \gamma$, then by elementary calculus, the s-derivative of $L_g(\Gamma_s)$ must vanish at $s = 0$ because $L_g(\gamma_s)$ attains a minimum there.

Families of curves

To make this rigorous, we introduce some more definitions. Let (M, g) be a Riemannian manifold. Given intervals $I, J \subseteq \mathbb{R}$, a continuous map $\Gamma : J \times I \to M$ is called a **one parameter family of curves**. Such a family defines two collections of curves in M: the **main curves** $\Gamma_s(t) = \Gamma(s,t)$ defined for $t \in I$ by holding s constant, and the **transverse curves** $\Gamma^t(s) = \Gamma(s,t)$ defined for $s \in I$ by holding t constant. If such a family Γ is smooth (or at least continuously differentiable), we denote the velocity vectors of the main and transverse curves by

$$\partial_t \Gamma(s,t) = (\Gamma_s)'(t) \in T_{\Gamma(s,t)}M, \quad \partial_s \Gamma(s,t) = (\Gamma^t)'(s) \in T_{\Gamma(s,t)}M.$$

Each of these is an example of a **vector field along $\boldsymbol{\Gamma}$**, which is a continuous map $V : I \times J \to TM$ such that $V(s,t) \in T_{\Gamma(s,t)}M$ for each (s,t). The families of curves that will interest us most are of the following type. A one-parameter family Γ is called an admissible family of curves if

(i) its domain is of the form $J \times [a,b]$ for some open interval J.

(ii) there is a partition $(a_0, \ldots, a_k)$ of $[a,b]$ such that Γ is smooth on each rectangle of the form $J \times [a_{i-1}, a_i]$.

(iii) $\Gamma_s(t) = \Gamma(s,t)$ is an admissible curve for each $s \in J$.

Every such partition is called an **admissible partition for the family**.

If $\gamma : [a,b] \to M$ is a given admissible curve, a **variation of $\boldsymbol{\gamma}$** is an admissible family of curves $\Gamma : J \times [a,b] \to M$ such that J is an open interval containing 0 and $\Gamma_0 = \gamma$. It is called a proper variation if in addition, all of the main curves have the same starting and ending points: $\Gamma_s(a) = \gamma(a)$ and $\Gamma_s(b) = \gamma(b)$ for all $s \in J$.

In the case of an admissible family, the transverse curves are smooth on J for each t, but the main curves are in general only piecewise regular. Thus the velocity vector fields $\partial_s \Gamma$ and $\partial_t \Gamma$ are smooth on each rectangle $J \times [a_{i-1}, a_i]$, but not generally on the whole domain.

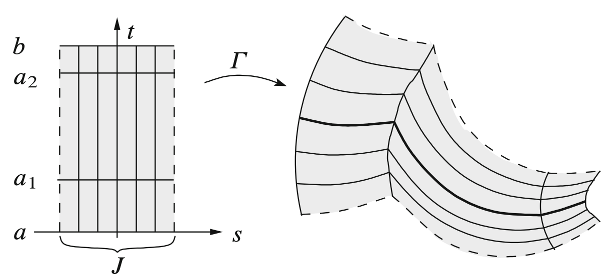

Figure 1.1: An admissible family.

We can say a bit more about $\partial_s \Gamma$, though. If Γ is an admissible family, a **piecewise smooth vector field along $\boldsymbol{\Gamma}$** is a (continuous) vector field along γ whose restriction to each rectangle $J \times [a_{i-1}, a_i]$ is smooth for some admissible partition $(a_0, \ldots, a_k)$ for Γ. In fact, $\partial_s \Gamma$ is always such a vector field. To see that it is continuous on the whole domain $J \times [a,b]$, note on the one hand that for each $i = 1, \ldots, k-1$, the values of $\partial_s \Gamma$ along the set $J \times \{a_i\}$ depend only on the values of Γ on that set, since the derivative is taken only with respect to the s variable;

on the other hand,$\partial_s\Gamma$ is continuous (in fact smooth) on each subrectangle $J \times [a_{i-1}, a_i]$ and $J \times [a_i, a_{i+1}]$, so the right-hand and left-hand limits at $t = a_i$ must be equal. Therefore $\partial_s\Gamma$ is always a piecewise smooth vector field along Γ. (However, $\partial_t\Gamma$ is typically not continuous at $t = a_i$.)

If Γ is a variation of γ, the **variation field of Γ** is the piecewise smooth vector field $V(t) = \partial_s\Gamma(0, t)$ along γ. We say that a vector field V along γ is **proper** if $V(a) = V(b) = 0$; it follows easily from the definitions that the variation field of every proper variation is itself proper.

Proposition 1.4.1. *If γ is an admissible curve and V is a piecewise smooth vector field along γ, then V is the variation field of some variation of γ. If V is proper, the variation can be taken to be proper as well.*

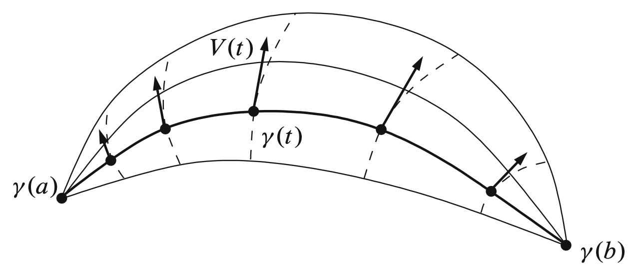

Figure 1.2: Every vector field along γ is a variation field.

Proof. Suppose γ and V satisfy the hypotheses, and set $\Gamma(s, t) = \exp_{\gamma(t)}(sV(t))$. By compactness of $[a, b]$, there is some positive ε such that Γ is defined on $(-\varepsilon, \varepsilon) \times [a, b]$. By composition, Γ is smooth on $(-\varepsilon, \varepsilon) \times [a_{i-1}, a_i]$ for each subinterval $[a_{i-1}, a_i]$ on which V is smooth, and it is continuous on its whole domain. By the properties of the exponential map, the variation field of Γ is V. Moreover, if $V(a) = 0$ and $V(b) = 0$, the definition gives $\Gamma(0, a) \equiv \gamma(a)$ and $\Gamma(0, b) \equiv \gamma(b)$, so Γ is proper. $\square$

If V is a piecewise smooth vector field along Γ, we can compute the covariant derivative of V either along the main curves (at points where V is smooth) or along the transverse curves; the resulting vector fields along Γ are denoted by D_tV and D_sV respectively. A key ingredient in the proof that minimizing curves are geodesics is the symmetry of the Levi-Civita connection. It enters into our proofs in the form of the following lemma.

Lemma 1.4.2 (Symmetry Lemma). *Let $\Gamma : J \times [a, b] \to M$ be an admissible family of curves in a Riemannian manifold. On every rectangle $J \times [a_{i-1}, a_i]$ where Γ is smooth,*

$$D_s\partial_t\Gamma = D_t\partial_s\Gamma.$$

Proof. This is a local question, so we may compute in local coordinates (x^i) around a point $\Gamma(s_0, t_0)$. Writing the components of Γ as $\Gamma(s, t) = (x^1(s, t), \ldots, x^n(s, t))$, we have

$$\partial_t\Gamma = \frac{\partial x^k}{\partial t}\partial_k, \quad \partial_s\Gamma = \frac{\partial x^k}{\partial s}\partial_k.$$

Then, using the coordinate formula (2.12) for covariant derivatives along curves, we obtain

$$D_s\partial_t\Gamma = \Big(\frac{\partial^2 x^k}{\partial s\partial t} + \frac{\partial x^i}{\partial s}\frac{\partial x^j}{\partial t}\Gamma^k_{ij}\Big)\partial_k, \quad D_t\partial_s\Gamma = \Big(\frac{\partial^2 x^k}{\partial t\partial s} + \frac{\partial x^i}{\partial t}\frac{\partial x^j}{\partial s}\Gamma^k_{ij}\Big)\partial_k.$$

Reversing the roles of i and j in the second line above and using the symmetry condition $\Gamma_{ij}^k = \Gamma_{ji}^k$, we conclude that these two expressions are equal. □

Minimizing curves are geodesics

We can now compute an expression for the derivative of the length functional along a variation of a curve. Traditionally, the derivative of a functional on a space of maps is called its **first variation**.

rst variation

Theorem 1.4.3 (First Variation Formula). *Let (M, g) be a Riemannian manifold. Suppose $\gamma : [a, b] \to M$ is a unit-speed admissible curve, $\Gamma : J \times [a, b] \to M$ is a variation of γ, and V is its variation field. Then $L_g(\Gamma_s)$ is a smooth function of s, and*

$$\frac{d}{ds}\Big|_{s=0} L_g(\Gamma_s) = -\int_a^b \langle V, D_t\gamma'\rangle dt - \sum_{i=1}^{k-1} \langle V(a_i), \Delta_i\gamma'\rangle + \langle V, \gamma'\rangle\Big|_a^b. \tag{4.1}$$

geodesic firs

where $(a_0, \dots, a_k)$ is an admissible partition for V, and for each $i = 1, \dots, k-1$, $\Delta_i\gamma' = \gamma'(a_i^+) - \gamma'(a_i^-)$ is the "jump" in the velocity vector field γ' at a_i. In particular, if Γ is a proper variation, then

$$\frac{d}{ds}\Big|_{s=0} L_g(\Gamma_s) = -\int_a^b \langle V, D_t\gamma'\rangle dt - \sum_{i=1}^{k-1} \langle V(a_i), \Delta_i\gamma'\rangle. \tag{4.2}$$

geodesic firs

Proof. On every rectangle $J \times [a_{i-1}, a_i]$ where Γ is smooth, since the integrand in $L_g(\Gamma_s)$ is smooth and the domain of integration is compact, we can differentiate under the integral sign as many times as we wish. Because $L_g(\Gamma_s)$ is a finite sum of such integrals, it follows that it is a smooth function of s.

Differentiating on the interval $[a_{i-1}, a_i]$ yields

$$\begin{aligned}
\frac{d}{ds} L_g(\Gamma_s|_{[a_{i-1},a_i]}) &= \int_{a_{i-1}}^{a_i} \frac{\partial}{\partial s} |\partial_t\Gamma_s| dt \\
&= \int_{a_{i-1}}^{a_i} \frac{1}{2|\partial_t\Gamma|} (\langle D_s\partial_t\Gamma, \partial_t\Gamma\rangle + \langle \partial_t\Gamma, D_s\partial_t\Gamma\rangle) dt \\
&= \int_{a_{i-1}}^{a_i} \frac{\langle D_s\partial_t\Gamma, \partial_t\Gamma\rangle}{|\partial_t\Gamma|} dt \\
&= \int_{a_{i-1}}^{a_i} \frac{\langle D_t\partial_s\Gamma, \partial_t\Gamma\rangle}{|\partial_t\Gamma|} dt
\end{aligned}$$

where we have used the symmetry lemma in the last line. Setting $s = 0$ and noting that $\partial_s\Gamma(0, t) = V(t)$ and $\partial_t\Gamma(0, t) = \gamma'(t)$ (which has length 1), we get

$$\begin{aligned}
\frac{d}{ds}\Big|_{s=0} L_g(\Gamma_s|_{[a_{i-1},a_i]}) &= \int_{a_{i-1}}^{a_i} \langle D_tV, \gamma'\rangle dt = \int_{a_{i-1}}^{a_i} \left(\frac{d}{dt}\langle V, \gamma'\rangle - \langle V, D_t\gamma'\rangle\right) dt \\
&= \langle V(a_i), \gamma'(a_i^-)\rangle - \langle V(a_{i-1}), \gamma'(a_{i-1}^+)\rangle - \int_{a_{i-1}}^{a_i} \langle V, D_t\gamma'\rangle dt.
\end{aligned}$$

geodesic first variation-1

Finally, summing over i, we obtain (4.1). □

Because every admissible curve has a unit-speed parametrization and length is independent of parametrization, the requirement in the above proposition that γ be of unit speed is not a real restriction, but rather just a computational convenience.

Theorem 1.4.4. *In a Riemannian manifold, every minimizing curve is a geodesic when it is given a unit-speed parametrization.*

Proof. Suppose $\gamma : [a,b] \to M$ is minimizing and of unit speed, and $(a_0, \dots, a_k)$ is an admissible partition for γ. If Γ is any proper variation of γ, then $L_g(\Gamma_s)$ is a smooth function of s that achieves its minimum at $s = 0$, so it follows from elementary calculus that $d(L_g(\Gamma_s))/ds = 0$ when $s = 0$. Since every proper vector field along γ is the variation field of some proper variation, the right-hand side of (4.2) must vanish for every such V.

First we show that $D_t\gamma' = 0$ on each subinterval $[a_{i-1}, a_i]$, so γ is a "broken geodesic". Choose one such interval, and let $\varphi \in C^\infty(\mathbb{R})$ be a bump function such that $\varphi > 0$ on (a_{i-1}, a_i) and $\varphi = 0$ elsewhere. Then (4.2) with $V = \varphi D_t\gamma'$ becomes

$$0 = -\int_{a_{i-1}}^{a_i} \varphi |D_t\gamma'|^2 dt.$$

Since the integrand is nonnegative and $\varphi > 0$ on (a_{i-1}, a_i), this shows that $D_t\gamma' = 0$ on each such subinterval.

Next we need to show that $\Delta_i\gamma' = 0$ for each i between 0 and k, which is to say that γ has no corners. For each such i, we can use a bump function in a coordinate chart to construct a piecewise smooth vector field V along γ such that $V(a_i) = \Delta_i\gamma'$ and $V(a_j) = 0$ for $j \neq i$. Then (4.2) reduces to $-|\Delta_i\gamma'|^2 = 0$, so $\Delta_i\gamma' = 0$ for each i.

Finally, since the two one-sided velocity vectors of γ match up at each a_i, it follows from uniqueness of geodesics that $\gamma|_{[a_i,a_{i+1}]}$ is the continuation of the geodesic $\gamma|_{[a_{i-1},a_i]}$, and therefore γ is smooth. □

The first variation formula actually tells us a bit more than is claimed in Theorem 1.4.4. In proving that γ is a geodesic, we did not use the full strength of the assumption that the length of Γ_s takes a minimum when $s = 0$; we only used the fact that its derivative is zero. We say that an admissible curve γ is a critical point of L_g if for every proper variation Γ_s of γ, the derivative of $L_g(\Gamma_s)$ with respect to s is zero at $s = 0$. Therefore we can strengthen Theorem 1.4.4 in the following way.

Corollary 1.4.5. *A unit-speed admissible curve γ is a critical point for L_g if and only if it is a geodesic.*

The geodesic equation $D_t\gamma' = 0$ thus characterizes the critical points of the length functional. In general, the equation that characterizes critical points of a functional on a space of maps is called the variational equation or the **Euler-Lagrange equation** of the functional.

Geodesics are locally minimizing

Next we turn to the converse of Theorem 1.4.4. It is easy to see that the literal converse is not true, because not every geodesic segment is minimizing. For example, every geodesic segment on S^n that goes more than halfway around the sphere is not minimizing, because the other portion of the same great circle is a shorter curve segment between the same two points. For that reason, we concentrate initially on local minimization properties of geodesics.

As usual, let (M, g) be a Riemannian manifold. A regular (or piecewise regular) curve $\gamma : I \to M$ is said to be **locally minimizing** if every $t_0 \in I$ has a neighborhood $I_0 \subseteq I$ such that whenever $a, b \in I_0$ with $a < b$, the restriction of γ to $[a, b]$ is minimizing.

al minimizing **Lemma 1.4.6.** *Every minimizing admissible curve segment is locally minimizing.*

Proof. Let $\gamma : I \to M$ be a minimizing admissible curve, then for any $a, b \in I$ the restriction $\gamma|_{[a,b]}$ is also minimizing, since otherwise we can replace this segment by a minimizing curve, which contradictc the fact that γ is a minimizing curve. $\square$

Our goal is to show that geodesics are locally minimizing. The proof will be based on a careful analysis of the geodesic equation in Riemannian normal coordinates. If ε is a positive number such that $\exp_p$ is a diffeomorphism from the ball $B_\varepsilon(0) \subseteq T_pM$ to its image (where the radius of the ball is measured with respect to the norm defined by g_p), then the image set $\exp_p(B_\varepsilon(0))$ is a normal neighborhood of p, called a **geodesic ball in M**, or sometimes an **open geodesic ball** for clarity. Also, if the closed ball $\overline{B_\varepsilon(0)}$ is contained in an open set $V \subseteq T_pM$ on which $\exp_p$ is a diffeomorphism onto its image, then $\exp_p(\overline{B_\varepsilon(0)})$ is called a **closed geodesic ball**, and $\exp_p(\partial B_\varepsilon(0))$ is called a **geodesic sphere**. Given such a V, by compactness there exists $\varepsilon' > \varepsilon$ such that $B_{\varepsilon'}(0) \subseteq V$, so every closed geodesic ball is contained in an open geodesic ball of larger radius. In Riemannian normal coordinates centered at p, the open and closed geodesic balls and geodesic spheres centered at p are just the coordinate balls and spheres.

Suppose U is a normal neighborhood of $p \in M$. Given any normal coordinates (x^i) on U centered at p, define the **radial distance function** $r : U \to \mathbb{R}$ by

$$r(x) = \sqrt{(x^1)^2 + \cdots + (x^n)^2}. \tag{4.3}$$

Riemann radia

and the **radial vector field** on $U \setminus \{p\}$, denoted by ∂_r, by

$$\frac{\partial}{\partial r} = \frac{x^i}{r}\frac{\partial}{\partial x^i}. \tag{4.4}$$

Riemann radia

In Euclidean space, $r(x)$ is the distance to the origin, and ∂_r is the unit vector field pointing radially outward from the origin. (The notation is suggested by the fact that ∂_r is a coordinate derivative in polar or spherical coordinates.)

stance smooth **Lemma 1.4.7.** *In every normal neighborhood U of $p \in M$, the radial distance function and the radial vector field are well defined, independently of the choice of normal coordinates. Both r and ∂_r are smooth on $U \setminus \{p\}$, and r^2 is smooth on all of U.*

Riemann normal coordinate prop

Proof. Proposition 1.3.24 shows that any two normal coordinate charts on U are related by $\widetilde{x}^i = A^i_j x^j$ for some orthogonal matrix (A^i_j), and a straightforward computation shows that both r and ∂_r are invariant under such coordinate changes:

$$r(x) = |x| = |Ax|, \quad \frac{\partial}{\partial r} = \frac{\widetilde{x}^i}{r}\frac{\partial}{\partial \widetilde{x}^i} = \frac{A^i_j x^j}{r}\frac{\partial}{\partial \widetilde{x}^i} = \frac{x^j}{r}\frac{\partial \widetilde{x}^i}{\partial x^j}\frac{\partial}{\partial \widetilde{x}^i} = \frac{x^j}{r}\frac{\partial}{\partial x^j}.$$

The smoothness statements follow directly from the coordinate formulas. $\square$

The crux of the proof that geodesics are locally minimizing is the following deceptively simple geometric lemma.

Gauss lemma **Theorem 1.4.8 (The Gauss Lemma).** *Let (M, g) be a Riemannian manifold, let U be a geodesic ball centered at $p \in M$, and let ∂_r denote the radial vector field on $U \setminus \{p\}$. Then ∂_r is a unit vector field orthogonal to the geodesic spheres in $U \setminus \{p\}$.*

Proof. We will work entirely in normal coordinates (x^i) on U centered at p, using the properties of normal coordinates described in Proposition 1.3.24. Let $q \in U \setminus \{p\}$ be arbitrary, with coordinate representation $(q_1, \ldots, q_n)$. It follows that $\partial_r|_q$ has the coordinate representation

$$\frac{\partial}{\partial r}\bigg|_q = \frac{q^i}{r(q)} \frac{\partial}{\partial x^i}\bigg|_q.$$

Let $v = v^i \partial_i \in T_pM$ be the tangent vector at p with components $v^i = q^i / r(q)$. By Proposition 1.3.24(c), the radial geodesic with initial velocity v is given in these coordinates by

$$\gamma_v(t) = (tv^1, \ldots, tv^n).$$

It satisfies $\gamma(0) = p$, $\gamma(r(q)) = q$, and $\gamma_v'(r(q)) = v^i \partial_i|_q = \partial_r|_q$. Because g_p is equal to the Euclidean metric in these coordinates (Proposition 1.3.24(b)), we have

$$|\gamma_v'(0)|_g = |v|_g = \sqrt{(v^1)^2 + \cdots + (v^n)^2} = \frac{1}{r(q)} \sqrt{(q^1)^2 + \cdots + (q^n)^2} = 1.$$

so v is a unit vector, and thus γ_v is a unit-speed geodesic. It follows that $\partial_r = \gamma_v'(b)$ is also a unit vector.

To prove that ∂_r is orthogonal to the geodesic spheres, let q and v be as above, and let $\Sigma_{r(q)}$ be the geodesic sphere containing q. In these coordinates, $\Sigma_{r(q)} = \exp_p(\partial B_{r(q)}(0))$ is the set of points satisfying the equation $(x^1)^2 + \cdots + (x^n)^2 = r^2(q)$. Let $w \in T_qM$ be any vector tangent to $\Sigma_{r(q)}$ at q. We need to show that $\langle w, \partial_r|_q \rangle = 0$.

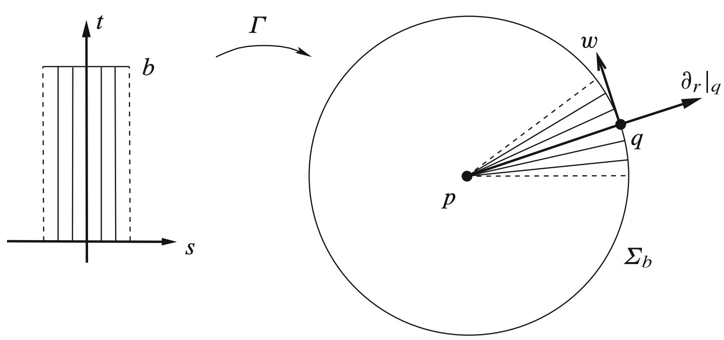

Figure 1.3: Proof of the Gauss lemma.

Choose a smooth curve $\sigma : (-\varepsilon, \varepsilon) \to \Sigma_{r(q)}$ satisfying $\sigma(0) = q$ and $\sigma'(0) = w$, and write its coordinate representation in (x^i)-coordinates as $\sigma(s) = (\sigma^1(s), \ldots, \sigma^n(s))$. The fact that $\sigma(s)$ lies in $\Sigma_{r(q)}$ for all s means that

$$(\sigma^1)^2 + \cdots + (\sigma^n)^2 = r^2(q). \tag{4.5}$$

Define a smooth map $\Gamma : (-\varepsilon, \varepsilon) \times [0, r(q)] \to M$ by

$$\Gamma(s,t) = \frac{t}{r(q)} \sigma(s).$$

For each $s \in (-\varepsilon, \varepsilon)$, Γ_s is a geodesic by Proposition 1.3.24(c). Its initial velocity is

$$\Gamma_s'(0) = \frac{\sigma(s)}{r(q)},$$

which is a unit vector by (4.5) and the fact that g_p is the Euclidean metric in coordinates; thus each Γ_s is a unit-speed geodesic.

Note that $\Gamma(0,t) = \gamma_v(t)$, so it now follows from the definitions that

$$\partial_s\Gamma(0,0) = \frac{d}{ds}\Big|_{s=0}\Gamma_s(0) = 0, \quad \partial_t\Gamma(0,0) = \frac{d}{dt}\Big|_{t=0}\gamma_v(t) = v;$$
$$\partial_s\Gamma(0,r(q)) = \frac{d}{ds}\Big|_{s=0}\sigma(s) = w, \quad \partial_t\Gamma(0,r(q)) = \frac{d}{dt}\Big|_{s=r(q)}\gamma_v(t) = \partial_r|_q.$$

Therefore $\langle \partial_s\Gamma, \partial_t\Gamma\rangle$ is zero when $(s,t) = (0,0)$ and equal to $\langle w, \partial_r|_q\rangle$ when $(s,t) = (0, r(q))$, so to prove the theorem it suffices to show that $\langle \partial_s\Gamma, \partial_t\Gamma\rangle$ is independent of t. We compute

$$\begin{aligned} \frac{d}{dt}\langle \partial_s\Gamma, \partial_t\Gamma\rangle &= \langle D_t\partial_s\Gamma, \partial_t\Gamma\rangle + \langle \partial_s\Gamma, D_t\partial_t\Gamma\rangle \\ &= \langle D_t\partial_s\Gamma, \partial_t\Gamma\rangle = \langle D_s\partial_t\Gamma, \partial_t\Gamma\rangle \\ &= \frac{1}{2}\frac{\partial}{\partial s}|\partial_t\Gamma|^2 = 0. \end{aligned} \tag{4.6}$$

where we use the symmetry lemma and the fact that each Γ_s is a unit-speed geodesic. This proves the theorem. □

Corollary 1.4.9. *Let U be a geodesic ball centered at $p \in M$, and let r and ∂_r be the radial distance and radial vector field. Then* $\operatorname{grad} r = \partial r$ *on* $U \setminus \{p\}$.

Proof. By the result of Exercise **??**, it suffices to show that ∂_r is orthogonal to the level sets of r and $\partial_r(r) \equiv |\partial_r|_g^2$. The first claim follows directly from the Gauss lemma, and the second from the fact that $\partial_r(r) = 1$ by direct computation in normal coordinates, which in turn is equal to $|\partial_r|_g^2$ by the Gauss lemma. □

Proposition 1.4.10. *Let (M,g) be a Riemannian manifold. Suppose $p \in M$ and q is contained in a geodesic ball around p. Then (up to reparametrization) the radial geodesic from p to q is the unique minimizing curve in M from p to q.*

Proof. Choose $\varepsilon > 0$ such that $\exp_p(B_\varepsilon(0))$ is a geodesic ball containing q. Let $\gamma : [0,c] \to M$ be the radial geodesic from p to q parametrized by arc length, and write $\gamma(t) = \exp_p(tv)$ for some unit vector $v \in T_pM$. Then $L_g(\gamma) = c$, since γ has unit speed.

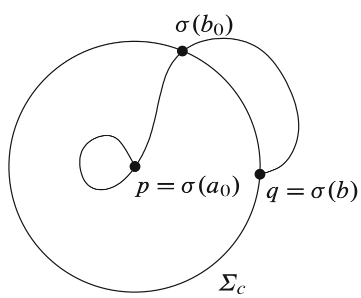

Figure 1.4: Radial geodesics are minimizing.

To show that γ is minimizing, we need to show that every other admissible curve from p to q has length at least c. Let $\sigma : [0,b] \to M$ be an arbitrary admissible curve from p to q, which we may assume to be parametrized by arc length as well. Let $a_0 \in [0,b]$ denote the last time that $\sigma(t) = p$, and $b_0 \in [0,b]$ the first time after a_0 that $\sigma(t)$ meets the geodesic sphere Σ_c of radius c around p. Then the composite function $r \circ \sigma$ is continuous on $[a_0, b_0]$

and piecewise smooth in (a_0, b_0), so we can apply the fundamental theorem of calculus to conclude that

$$\begin{aligned} r(\sigma(b_0)) - r(\sigma(a_0)) &= \int_{a_0}^{b_0} \frac{d}{dt} r(\sigma(t)) dt = \int_{a_0}^{b_0} \langle \operatorname{grad} r|_{\sigma(t)}, \sigma'(t) \rangle dt \\ &\leq \int_{a_0}^{b_0} |\operatorname{grad} r|_{\sigma(t)}| \cdot |\sigma'(t)| dt \qquad (4.7) \\ &= \int_{a_0}^{b_0} |\sigma'(t)| dt = L_g(\sigma|_{[a_0, b_0]}). \end{aligned}$$

Thus $L_g(\sigma) \geq r(\sigma(b_0)) - r(\sigma(a_0)) = c$, so γ is minimizing.

Now suppose $L_g(\sigma) = c$. Then $b = c$, and $a_0 = 0$ and $b_0 = b = c$, since otherwise the segments of σ before $t = a_0$ and after $t = b_0$ would contribute positive lengths. Moreover, we have that $\sigma'(t)$ is a positive multiple of $\operatorname{grad} r|_{\sigma(t)}$ for each t. Since we assume that σ has unit speed, we must have $\sigma'(t) = \operatorname{grad} r|_{\sigma(t)} = \partial_r|_{\sigma(t)}$. Thus σ and γ are both integral curves of ∂r passing through q at time $t = c$, so $\sigma = \gamma$. □

The next two corollaries show how radial distance functions, balls, and spheres in normal coordinates are related to their global metric counterparts.

Corollary 1.4.11. *Let (M, g) be a connected Riemannian manifold and $p \in M$. Within every open or closed geodesic ball around p, the radial distance function $r(x)$ is equal to the Riemannian distance from p to x in M.*

Proof. Since every closed geodesic ball is contained in an open geodesic ball of larger radius, we need only consider the open case. If x is in the open geodesic ball $\exp_p(B_c(0))$, the radial geodesic γ from p to x is minimizing by Proposition 1.4.10. Since its velocity is equal to ∂_r, which is a unit vector in both the g-norm and the Euclidean norm in normal coordinates, the g-length of γ is equal to its Euclidean length, which is $r(x)$. □

Corollary 1.4.12. *In a connected Riemannian manifold, every open or closed geodesic ball is also an open or closed metric ball of the same radius, and every geodesic sphere is a metric sphere of the same radius.*

Proof. Let (M, g) be a Riemannian manifold, and let $p \in M$ be arbitrary. First, let $V = \exp_p(\overline{B_c(0)}) \subseteq M$ be a closed geodesic ball of radius $c > 0$ around p. Suppose q is an arbitrary point of M. If $q \in V$, then Corollary 1.4.11 shows that q is also in the closed metric ball of radius c. Conversely, suppose $q \notin V$. Let S be the geodesic sphere $\exp_p(\partial B_c(0))$. The complement of S is the disjoint union of the open sets $\exp_p(B_c(0))$ and $M \setminus \exp_p(B_c(0))$ and hence disconnected. Thus if $\gamma : [a, b] \to M$ is any admissible curve from p to q, there must be a time $t_0 \in [a, b]$ when $\gamma(t_0) \in S$, and then Corollary 1.4.11 shows that the length of $\gamma|_{[a, a_0]}$ must be at least c. Since $\gamma|_{[a_0, b]}$ must have positive length, it follows that $d_g(p, q) > c$, so q is not in the closed metric ball of radius c around p.

Next, let $W = \exp_p(B_c(0))$ be an open geodesic ball of radius c. Since W is the union of all closed geodesic balls around p of radius less than c, and the open metric ball of radius c is similarly the union of all closed metric balls of smaller radii, the result of the preceding paragraph shows that W is equal to the open metric ball of radius c.

Finally, if $S = \exp_p(\partial B_c(0))$ is a geodesic sphere of radius c, the arguments above show that S is equal to the closed metric ball of radius c minus the open metric ball of radius c, which is exactly the metric sphere of radius c. □

The last corollary suggests a simplified notation for geodesic balls and spheres in M. From now on, we will use the notations

$$B_\varepsilon(p) = \exp_p(B_\varepsilon(0)), \quad \overline{B_\varepsilon(p)} = \exp_p(\overline{B_\varepsilon(0)}), \quad S_\varepsilon(p) = \exp_p(\partial B_\varepsilon(0))$$

for open and closed geodesic balls and geodesic spheres, which we now know are also open and closed metric balls and spheres.

We continue to let (M, g) be a Riemannian manifold. In order to prove that geodesics in M are locally minimizing, we need the following refinement of the concept of normal neighborhoods. A subset $W \subseteq M$ is called **uniformly normal** if there exists some $\delta > 0$ such that W is contained in a geodesic ball of radius δ around each of its points. If δ is any such constant, we will also say that W is **uniformly δ-normal**. Clearly every subset of a uniformly δ-normal set is itself uniformly δ-normal.

neighborhood **Lemma 1.4.13 (Uniformly Normal Neighborhood Lemma).** *Given $p \in M$ and any neighborhood U of p, there exists a uniformly normal neighborhood of p contained in U.*

Proof. Choose a normal coordinate chart $(U_0, (x^i))$ centered at p and contained in U, and let (x^i, v^i) be the corresponding natural coordinates for $\pi^{-1}(U_0) \subseteq TM$. Because this is a local question, we might as well identify U_0 with an open subset of $\mathbb{R}^n$, and identify TM with $U_0 \times \mathbb{R}^n$. The exponential map for the Riemannian manifold (U_0, g) is defined on an open subset $\mathcal{E} \subseteq U_0 \times \mathbb{R}^n$. Consider the map $E : \mathcal{E} \to U_0 \times U_0$ defined by

$$E(x, v) = (x, \exp_x v).$$

The differential of E at $(p, 0)$ is represented by the matrix

$$dE_{(p,0)} = \begin{pmatrix} \dfrac{\partial x^i}{\partial x^j} & \dfrac{\partial x^i}{\partial v^j} \\ \dfrac{\partial \exp^i}{\partial x^j} & \dfrac{\partial \exp^i}{\partial v^j} \end{pmatrix} = \begin{pmatrix} \mathrm{id} & 0 \\ * & \mathrm{id} \end{pmatrix}$$

which is invertible. By the inverse function theorem, therefore, there are neighborhoods $\mathcal{U} \subseteq U_0 \times \mathbb{R}^n$ of $(p, 0)$ and $\mathcal{V} \subseteq U_0 \times U_0$ of (p, p) such that E restricts to a diffeomorphism from $\mathcal{U}$ to $\mathcal{V}$. Shrinking both neighborhoods if necessary, we may assume that $\mathcal{U}$ is a product set of the form $W \times B_\varepsilon(0)$, where W is a neighborhood of p and $B_\varepsilon(0)$ is a Euclidean ball in v-coordinates. It follows that for each $x \in W$, $\exp_x$ maps $B_\varepsilon(0)$ smoothly onto the open set $\mathcal{V}_x = \{y : (x, y) \in \mathcal{V}\}$, and it is a diffeomorphism because its inverse is given explicitly by $\exp_x^{-1}(y) = \pi_2 \circ E^{-1}(x, y)$, where $\pi_2 : U_0 \times \mathbb{R}^n \to \mathbb{R}^n$ is the projection. Shrinking W still further if necessary, we may assume that the metric g satisfies an estimate of the form (**??**) Riemann metric-1 for all $x \in W$. This means that for each $x \in W$, the coordinate ball $B_\varepsilon(0) \subseteq T_xM$ contains a g_x-ball of radius at least ε/c. Put $\delta = \varepsilon/c$, we have shown that for each $x \in W$, there is a g-geodesic ball of radius δ in M centered at x.

Now, shrinking W once more, we may assume that its diameter (with respect to the metric d_g) is less than δ. It follows that for each $x \in W$, the entire set W is contained in the metric ball of radius δ around x, and Corollary 1.4.12 Riemann geodesic ball is metric ball shows that this metric ball is also a geodesic ball of radius δ. Thus W is the required uniformly normal neighborhood of p. □

ocal minimize **Theorem 1.4.14.** *Every Riemannian geodesic is locally minimizing.*

Proof. Let (M, g) be a Riemannian manifold. Suppose $\gamma : I \to M$ is a geodesic, which we may assume to be defined on an open interval, and let $t_0 \in I$. Let W be a uniformly normal neighborhood of $\gamma(t_0)$, and let $I_0 \subseteq I$ be the connected component of $\gamma^{-1}(W)$ containing t_0. If $a, b \in I_0$ with $a < b$, then the definition of uniformly normal neighborhood implies that $\gamma(b)$ is contained in a geodesic ball centered at $\gamma(a)$. Therefore, by Proposition 1.4.10, the radial geodesic segment from $\gamma(a)$ to $\gamma(b)$ is the unique minimizing curve segment between these points. However, the restriction of γ to $[a, b]$ is also a geodesic segment from $\gamma(a)$ to $\gamma(b)$ lying in the same geodesic ball, and thus $\gamma|_{[a,b]}$ must coincide with this minimizing geodesic. $\square$

It is interesting to note that the Gauss lemma and the uniformly normal neighborhood lemma also yield another proof that minimizing curves are geodesics, without using the first variation formula. On the principle that knowing more than one proof of an important fact always deepens our understanding of it, we present this proof for good measure.

Another proof of Theorem 1.4.4. Suppose $\gamma : [a, b] \to M$ is a minimizing admissible curve. Just as in the preceding proof, for every $t_0 \in [a, b]$ we can find a connected neighborhood I_0 of t_0 such that $\gamma(I_0)$ is contained in a uniformly normal neighborhood W. Then for every $a_0, b_0 \in I_0$, the same argument as above shows that the unique minimizing curve segment from $\gamma(a_0)$ to $\gamma(b_0)$ is a radial geodesic. Since the restriction of γ to $[a_0, b_0]$ is such a minimizing curve segment, it must coincide with this radial geodesic. Therefore γ solves the geodesic equation in a neighborhood of t_0. Since t_0 was arbitrary, γ is a geodesic. $\square$

Given a Riemannian manifold (M, g) (without boundary), for each point $p \in M$ we define the **injectivity radius** of M at p, denoted by $\mathrm{inj}(p)$, to be the supremum of all $\varepsilon > 0$ such that $\exp_p$ is a diffeomorphism from $B_\varepsilon(0) \subseteq T_pM$ onto its image. If there is no upper bound to the radii of such balls (as is the case, for example, on $\mathbb{R}^n$), then we set $\mathrm{inj}(p) = \infty$. Then we define the injectivity radius of M, denoted by $\mathrm{inj}(M)$, to be the infimum of $\mathrm{inj}(p)$ as p ranges over points of M. It can be zero, positive, or infinite.

compact inj>0 **Lemma 1.4.15.** *If (M, g) is a compact Riemannian manifold, then* $\mathrm{inj}(M)$ *is positive.*

Proof. For each $x \in M$, there is a positive number $\delta(x)$ such that x is contained in a uniformly $\delta(x)$-normal neighborhood W_x, and $\mathrm{inj}(x') \geq \mathrm{inj}(x)$ for each $x' \in W_x$. Since M is compact, it is covered by finitely many such neighborhoods $W_{x_1}, \ldots, W_{x_n}$. Therefore, $\mathrm{inj}(M)$ is at least equal to the minimum of $\delta(x_1), \ldots, \delta(x_n)$. It cannot be infinite, because a compact metric space is bounded, and a geodesic ball of radius c contains points whose distances from the center are arbitrarily close to c. $\square$

In addition to uniformly normal neighborhoods, there is another, more specialized, kind of normal neighborhood that is frequently useful. Let (M, g) be a Riemannian manifold. A subset $U \subseteq M$ is said to be **geodesically convex** if for each $p, q \in U$, there is a unique minimizing geodesic segment from p to q in M, and the image of this geodesic segment lies entirely in U. Now we show that every sufficiently small geodesic ball is geodesically convex.

ghborhood lem **Lemma 1.4.16.** *Let (M, g) be a Riemannian manifold. For each $p \in M$, there exists $c > 0$ such that any geodesic in M that is tangent at $q \in M$ to the geodesic sphere $S_r(p)$ of radius $r < c$ stays out of the geodesic ball $B_r(p)$ for some neighborhood of q.*

Proof. Let $p \in M$ be fixed, and let W be a uniformly normal neighborhood of p. For $\varepsilon > 0$ small enough that $B_\varepsilon(p) \subseteq W$, define a subset $W \subseteq TM \times \mathbb{R}$ by

$$W_\varepsilon = \{(q, v, t) \in TM \times \mathbb{R} : q \in B_\varepsilon(p), v \in T_q M, |v| = 1, |t| < 2\varepsilon\}.$$

Define $F : W_\varepsilon \to \mathbb{R}$ by

$$F(q, v, t) = d_g(p, \gamma_v(t))^2,$$

where $\gamma_v(t)$ is the geodesic starting at q with velocity v. Choose a normal coordinate (x^i) centered at p, then by Proposition 1.4.11 F is squre the radial distance from p to $\gamma_v(t)$, so it is smooth by Lemma 1.4.7. Moreover, in this coordinate, by Corollary 1.4.9 we have

$$\frac{\partial F}{\partial t} = 2\langle \gamma_v(t), \gamma_v'(t)\rangle, \quad \frac{\partial^2 F}{\partial t^2} = 2|\gamma_v'(t)|^2 + 2\langle \gamma_v(t), \gamma_v''(t)\rangle.$$

Note that by Proposition 1.3.24(d) we have

$$\frac{\partial^2 F}{\partial t^2}(p, v, 0) = 2|v|^2 = 2,$$

so there exists a number $c > 0$ such that $B_c(p) \subseteq W$ and $\partial^2 F/\partial^2 t > 0$ for all $q \in V$ and $v \in T_q M, |v| = 1$.

Now let $0 < r < c$. We may only consider geodesic parametrized by arc-length. If a geodesic γ is tangent at the point q to the geodesic sphere $S_r(p)$ with $\gamma(t_0) = q, \gamma'(t_0) = v$, then from the Guass lemma,

$$\langle \gamma(t_0), \gamma'(t_0)\rangle = 0.$$

that is, $\partial F/\partial t(q, v, 0) = 0$. Then since $\partial^2 F/\partial t^2 > 0$, we know that t_0 is a strict minimum point of $d_g(p, \gamma(t))$. Therefore γ stay out of the geodesic ball $B_r(p)$ for some neighborhood of q. □

Theorem 1.4.17. *Let (M, g) be a Riemannian manifold. For each $p \in M$, there exists $\varepsilon_0 > 0$ such that every geodesic ball centered at p of radius less than or equal to ε_0 is geodesically convex.*

Proof. Let c be given in Lemma 1.4.16. Choose $0 < \delta < c/2$ and let W be a uniformly δ-normal neighborhood of p. Let $q_1, q_2 \in B_\delta(p)$ and let γ be the unique geodesic of length $< 2\delta < c$ joining q_1 to q_2. It is clear that γ is contained in $B_c(p)$. Since $\operatorname{im}\gamma$ is compact, there is a point q in the interior of γ where $d_g(p, \gamma)$ attains its maximum, then γ is tangent at q to the geodesic sphere $S_r(p)$ where $r = d_g(p, q)$. Since $q \in B_c(p)$, this contradicts Lemma 1.4.16. □

1.4.2 Completeness

Suppose (M, g) is a connected Riemannian manifold. Now that we can view M as a metric space, it is time to address one of the most important questions one can ask about a metric space: Is it complete? In general, the answer is no: for example, if M is an open ball in $\mathbb{R}^n$ with its Euclidean metric, then every sequence in M that converges in $\mathbb{R}^n$ to a point in ∂M is Cauchy, but not convergent in M. We have introduced another notion of completeness for Riemannian and pseudo-Riemannian manifolds: recall that such a manifold is said to be **geodesically complete** if every maximal geodesic is defined for all $t \in \mathbb{R}$. For clarity, we will use the phrase **metrically complete** for a connected Riemannian manifold that is complete as a metric space with the Riemannian distance function, in the sense that every Cauchy sequence converges.

The HopfRinow theorem, which we will state and prove below, shows that these two notions of completeness are equivalent for connected Riemannian manifolds. Before we prove it, let us establish a preliminary result, which will have other important consequences besides the Hopf-Rinow theorem itself.

Lemma 1.4.18. *Suppose (M, g) is a connected Riemannian manifold, and there is a point $p \in M$ such that $\exp_p$ is defined on the whole tangent space T_pM. Then*

(a) Given any other $q \in M$, there is a minimizing geodesic segment from p to q.

(b) M is metrically complete.

Proof. Let $q \in M$ be arbitrary. If $\gamma : [a, b] \to M$ is a geodesic segment starting at p, let us say that γ aims at q if γ is minimizing and

$$d_g(p, q) = d_g(p, \gamma(b)) + d_g(\gamma(b), q). \tag{4.8}$$

(This would be the case, for example, if γ were an initial segment of a minimizing geodesic from p to q; but we are not assuming that.) To prove (a), it suffices to show that there is a geodesic segment $\gamma : [a, b] \to M$ that begins at p, aims at q, and has length equal to $d_g(p, q)$, for then the fact that γ is minimizing means that $d_g(p, \gamma(b)) = L_g(\gamma) = d_g(p, q)$, and (4.8) becomes

$$d_g(p, q) = d_g(p, q) + d_g(\gamma(b), q)$$

which implies $\gamma(b) = q$. Since γ is a segment from p to q of length $d_g(p, q)$, it is the desired minimizing geodesic segment.

Choose $\varepsilon > 0$ such that there is a closed geodesic ball $\overline{B_\varepsilon(p)}$ around p that does not contain q. Since the distance function on a metric space is continuous, there is a point x in the geodesic sphere $S_\varepsilon(p)$ where $d_g(x, q)$ attains its minimum on the compact set $S_\varepsilon(p)$. Let γ be the maximal unit-speed geodesic whose restriction to $[0, \varepsilon]$ is the radial geodesic segment from p to x; by assumption, γ is defined for all $t \in \mathbb{R}$.

We begin by showing that $\gamma|_{[0,\varepsilon]}$ aims at q. Since it is minimizing by Proposition 1.4.10 (noting that every closed geodesic ball is contained in a larger open one), we need only show that (4.8) holds with $b = \varepsilon$, or

$$d_g(p, q) = d_g(p, x) + d_g(x, q). \tag{4.9}$$

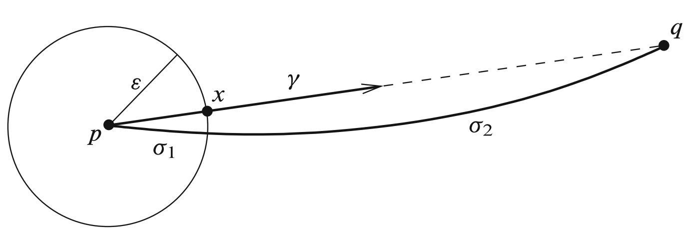

Figure 1.5: Proof that $\gamma|_{[0,\varepsilon]}$ aims at q.

To this end, let $\sigma : [a_0, b_0] \to M$ be any admissible curve from p to q. Let t_0 be the first time σ hits $S_\varepsilon(p)$, and let σ_1 and σ_2 denote the restrictions of σ to $[a_0, t_0]$ and $[t_0, b_0]$, respectively.

Since every point in $S_\varepsilon(p)$ is at a distance ε from p, we have

$$L_g(\sigma_1) \geq d_g(p, \sigma(t_0)) = d_g(p, x),$$

and by our choice of x we have

$$L_g(\sigma_2) \geq d_g(\sigma(t_0), q) \geq d_g(x, q).$$

Putting these two inequalities together yields

$$L_g(\sigma) = L_g(\sigma_1) + L_g(\sigma_2) \geq d_g(p, x) + d_g(x, q).$$

Taking the infimum over all such σ, we find that $d_g(p, q) \geq d_g(p, x) + d_g(x, q)$. The opposite inequality is just the triangle inequality, so (4.9) holds.

Now let $T = d_g(p, d)$ and

$$\mathcal{A} = \{b \in [0, T] : \gamma|_{[0,b]} \text{ aims at } q\}.$$

We have just shown that $\varepsilon \in \mathcal{A}$. Let $A = \sup \mathcal{A}$. By continuity of the distance function, it is easy to see that $\mathcal{A}$ is closed in $[0, T]$, and therefore $A \in \mathcal{A}$. If $A = T$, then $\gamma|_{[0,T]}$ is a geodesic of length $T = d_g(p, q)$ that aims at q, and by the remark above we are done. So we assume $A < T$ and derive a contradiction.

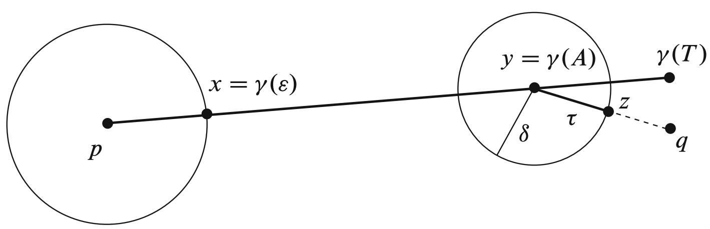

Figure 1.6: Proof that $A = T$.

Let $y = \gamma(A)$, and choose $\delta > 0$ such that there is a closed geodesic ball $\overline{B_\delta(y)}$ around y, small enough that it does not contain q. The fact that $A \in \mathcal{A}$ means that

$$d_g(p, q) = d_g(p, y) + d_g(y, q).$$

Let $z \in S_\delta(y)$ be a point where $d_g(z, q)$ attains its minimum, and let $\tau : [0, \delta] \to M$ be the unit-speed radial geodesic from y to z. By exactly the same argument as before, τ aims at q, so

$$d_g(y, q) = d_g(y, z) + d_g(z, q). \tag{4.10}$$

By the triangle inequality

$$\begin{aligned} d_g(p, z) &\geq d_g(p, q) - d_g(z, q) \\ &= d_g(p, y) + d_g(y, q) - d_g(z, q) \\ &= d_g(p, y) + d_g(y, z) \\ &\geq d_g(p, z). \end{aligned} \tag{4.11}$$

Therefore, the admissible curve consisting of $\gamma|_{[0,A]}$ (of length A) followed by τ (of length δ) is a minimizing curve from p to z. This means that it has no corners, so z must lie on γ, and in fact, $z = \gamma(A + \delta)$. But then (4.11) says that

$$d_g(p, q) = d_g(p, z) + d_g(q, z),$$

so $\gamma|_{[0,A+\delta]}$ aims at q and $A+\delta \in \mathcal{A}$, which is a contradiction. This completes the proof of (a).

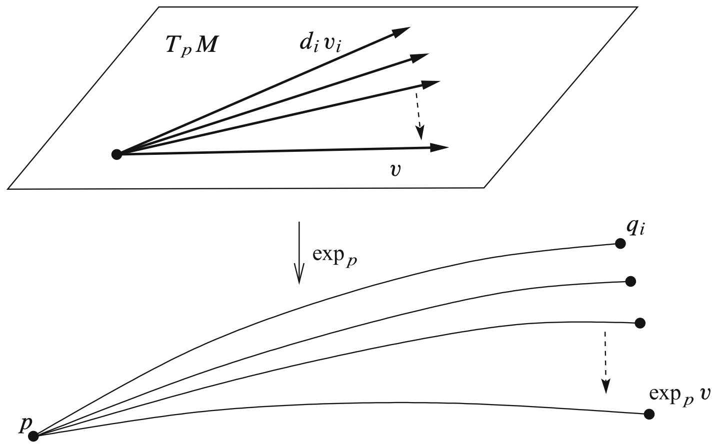

Figure 1.7: Cauchy sequences converge.

To prove (b), we need to show that every Cauchy sequence in M converges. Let (q_i) be a Cauchy sequence in M. For each i, let $\gamma_i(t) = \exp_p(tv_i)$ be a unit-speed minimizing geodesic from p to q_i, and let $d_i = d_g(p, q_i)$, so that $q_i = \exp_p(d_i v_i)$. The sequence (d_i) is bounded in $\mathbb{R}$ (because Cauchy sequences in a metric space are bounded), and the sequence (v_i) consists of unit vectors in T_pM, so the sequence of vectors $(d_i v_i)$ in T_pM is bounded. Therefore a subsequence $(d_{i_k} v_{i_k})$ converges to some $v \in T_pM$. By continuity of the exponential map, $q_{i_k} = \exp_p(d_{i_k} v_{i_k}) \to \exp_p(v)$ and since the original sequence (q_i) is Cauchy, it converges to the same limit. □

Theorem 1.4.19 (Hopf-Rinow). *Let (M, g) be a connected Riemannian manifold. Then the following statements are equivalent:*

(i) M is metrically complete.

(ii) M geodesically complete.

(iii) There exists $p \in M$ so that $\exp_p$ is defined on the whole tangent space T_pM.

(iv) Any bounded closed subset in M is compact.

Moreover, each of the previous statements implies that any two points in M can be joined by a minimizing geodesic segment

Remark 1.4.1. The condition that any tow points can be connected by a minimizing geodesic does not implies that M is complete. For example, this statement is true in the Euclidean ball $\mathbb{B}^n \subseteq \mathbb{R}^n$, but it is not complete.

Also, the equivalence of the completeness and the Heine-Borel property is not valid in general metric spaces. For example, one can consider a countable infinite set $\{x_n : n \in \mathbb{N}\}$ and define a metric on it via $d(x_i, x_j) = 1$ for all $i \neq j$ (discrete metric). In this space, the only Cauchy sequences are eventually-constant sequences which of course converge. However, the whole space is closed and bounded but not compact. So as metric spaces, Riemannian manifolds are special (and nice) metric spaces.

Riemann hopf-rinow le

Proof. Clearly we only need to prove the equivalence of (i)–(iv) in view of Lemma 1.4.18. Moreover, the implications (ii)⇒ (iii) ⇒(i) are immediate by the same lemma, and (iv)⇒(i) is a standard result for metric spaces.

Now we show (i)⇒(ii). Suppose M is metrically complete, and assume for the sake of contradiction that it is not geodesically complete. Then there is some unit-speed geodesic $\gamma : [0, b) \to M$ that has no extension to a geodesic on any interval $[0, b_0)$ for $b_0 > b$. Let (t_i) be any increasing sequence in $[0, b)$ that approaches b, and set $q_i = \gamma(t_i)$. Since γ is parametrized by arc length, the length of $\gamma|_{[t_i,t_j]}$ is exactly $t_{i+1} - t_i$, so $d_g(q_i, q_j) \le |t_i, t_j|$ and (q_i) is a Cauchy sequence inM. By completeness, (q_i) converges to some point $q \in M$.

Let W be a uniformly δ-normal neighborhood of q for some $\delta > 0$. Choose j large enough that $t_j > b - \delta$ and $q_j \in W$. The fact that $B_\delta(q_j)$ is a geodesic ball means that every unit-speed geodesic starting at q_j exists at least for $t \in [0, \delta)$. In particular, this is true of the geodesic σ with $\sigma(0) = q_j$ and $\sigma'(0) = \gamma'(t_j)$. Define $\widetilde{\gamma} : [0, t_j + \delta) \to M$ by

$$\widetilde{\gamma}(t) = \begin{cases} \gamma(t) & t \in [0, b); \\ \sigma(t - t_j) & t \in (t_j, t_j + \delta) \end{cases}$$

Note that both expressions on the right-hand side are geodesics, and they have the same position and velocity when $t = t_j$. Therefore, by uniqueness of geodesics, the two definitions agree where they overlap. Since $t_j + \delta > b$, $\widetilde{\gamma}$ is an extension of γ past b, which is a contradiction.

Finally, we show (iii)⇒(iv). Suppose that there exists $p \in M$ so that $\exp_p$ is defined on the whole tangent space T_pM. Let $K \subseteq M$ be a bounded closed set. Then there exists a constant $C > 0$ so that $d(p, k) < C$ for all $k \in K$. Then according to (iii), K is contained in the closed geodesic ball $\overline{B_C(p)}$, which is compact since $B_C(0)$ is compact in T_pM. Thus K, as a closed subset of a compact set, is compact. □

Corollary 1.4.20. *If M is a compact Riemannian manifold, then every maximal geodesic in M is defined for all time.*

Now we use the Hopf-Rinow theorem to prove the following important theorem about Riemannian covering maps.

lete is cover

Theorem 1.4.21. *Suppose $(\widetilde{M}, \widetilde{g})$ and (M, g) are connected Riemannian manifolds with $\widetilde{M}$ complete, and $\pi : \widetilde{M} \to M$ is a local isometry. Then M is complete and π is a Riemannian covering map.*

Proof. A fundamental property of covering maps is the path-lifting property: if π is a covering map, then every continuous path $\gamma : I \to M$ lifts to a path $\widetilde{\gamma}$ in $\widetilde{M}$ such that $pi \circ \widetilde{\gamma} = \gamma$. We begin by proving that π possesses the path-lifting property for geodesics: if $p \in M$ is a point in the image of π, $\gamma : I \to M$ is any geodesic starting at p, and $\widetilde{p}$ is any point in $\pi^{-1}(p)$, then γ has a unique lift starting at $\widetilde{p}$. The lifted curve is necessarily also a geodesic because γ is a local isometry.

To prove the path-lifting property for geodesics, suppose $p \in \pi(\widetilde{M})$ and $\widetilde{p} \in \pi^{-1}(M)$, and let $\gamma : I \to M$ be any geodesic with $p = \gamma(0)$. Let $v = \gamma'(0)$ and $\widetilde{v} = (d\pi_{\widetilde{p}})^{-1}(v) \in T_{\widetilde{p}}\widetilde{M}$ (which is well defined because $d\pi_{\widetilde{p}}$ is an isomorphism), and let $\widetilde{\gamma}$ be the geodesic in $\widetilde{M}$ with initial point $\widetilde{p}$ and initial velocity $\widetilde{v}$. Because $\widetilde{M}$ is complete, $\widetilde{\gamma}$ is defined on all of $\mathbb{R}$. Since π is a local isometry, it takes geodesics to geodesics; and since by construction $\pi(\widetilde{\gamma}(0)) = \gamma(0)$ and $d\pi_{\widetilde{p}}(\widetilde{\gamma}'(0)) = \gamma'(0)$, we must have $\pi \circ \widetilde{\gamma} = \gamma$ on I, so $\widetilde{\gamma}|_I$ is a lift of γ starting at $\widetilde{p}$.

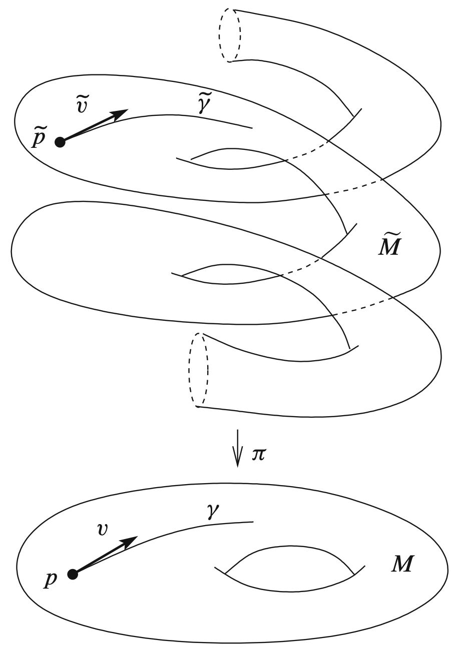

Figure 1.8: Lifting geodesics.

To show that M is complete, let p be any point in the image of π. If $\gamma : I \to M$ is any geodesic starting at p, then γ has a lift $\widetilde{\gamma} : I \to \widetilde{M}$. Because $\widetilde{M}$ is complete, $\pi \circ \widetilde{\gamma}$ is a geodesic defined for all t that coincides with γ on I, so γ extends to all of $\mathbb{R}$. Thus M is complete by Hopf-Rinow Theorem.

Next we show that γ is surjective. Choose some point $\widetilde{p} \in \widetilde{M}$, write $p = \pi(\widetilde{p})$, and let $q \in M$ be arbitrary. Because M is connected and complete, there is a minimizing unit-speed geodesic segment γ from p to q. Letting $\widetilde{\gamma}$ be the lift of γ starting at $\widetilde{p}$ and $r = d_g(p, q)$, we have $\pi(\widetilde{\gamma}(r)) = \gamma(r) = q$, so q is in the image of π.

To show that π is a smooth covering map, we need to show that every point of M has a neighborhood U that is evenly covered, which means that $\pi^{-1}(U)$ is a disjoint union of connected open sets $\widetilde{U}_\alpha$ such that $\pi|_{\widetilde{U}_\alpha}$ is a diffeomorphism. We will show, in fact, that every geodesic ball is evenly covered.

Let $p \in M$, and let $U = B_\varepsilon(p)$ be a geodesic ball centered at p. Write $\pi^{-1}(p) = \{\widetilde{p}_\alpha\}_{\alpha \in A}$, and for each α let $\widetilde{U}_\alpha$ denote the metric ball of radius ε around $\widetilde{p}_\alpha$ (we are not claiming that $\widetilde{U}_\alpha$ is a geodesic ball). The first step is to show that the various sets $\widetilde{U}_\alpha$ are disjoint. For every $\alpha \neq \beta$, there is a minimizing geodesic segment $\widetilde{\gamma}$ from $\widetilde{p}_\alpha$ to $\widetilde{p}_\beta$ because $\widetilde{M}$ is complete. The projected curve $\pi \circ \widetilde{\gamma}$ is a geodesic segment that starts and ends at p, whose length is the same as that of $\widetilde{\gamma}$. Such a geodesic must leave U and reenter it (since all geodesics passing through p and lying in U are radial line segments), and thus must have length at least 2ε. This means that $d_g(\widetilde{p}_\alpha, \widetilde{p}_\beta)$, and thus by the triangle inequality, $\widetilde{U}_\alpha \cap \widetilde{U}_\beta = \varnothing$.

The next step is to show that $\pi^{-1} = \bigcup_\alpha \widetilde{U}_\alpha$. If $\widetilde{q}$ is any point in $\widetilde{U}_\alpha$, then there is a geodesic $\widetilde{\gamma}$ of length less than ε from $\widetilde{p}_\alpha$ to $\widetilde{q}$, and then $\pi \circ \widetilde{\gamma}$ is a geodesic of the same length from p to $\pi(\widetilde{q})$, showing that $\pi(\widetilde{q}) \in U$. It follows that $\bigcup_\alpha \widetilde{U}_\alpha \subseteq \pi^{-1}(U)$.

Conversely, suppose $\widetilde{q} \in \pi^{-1}(U)$, and set $q = \pi(\widetilde{q})$. This means that $q \in U$, so there is a minimizing radial geodesic γ in U from q to p, and $r = d_g(q, p) < \varepsilon$. Let $\widetilde{\gamma}$ be the lift of

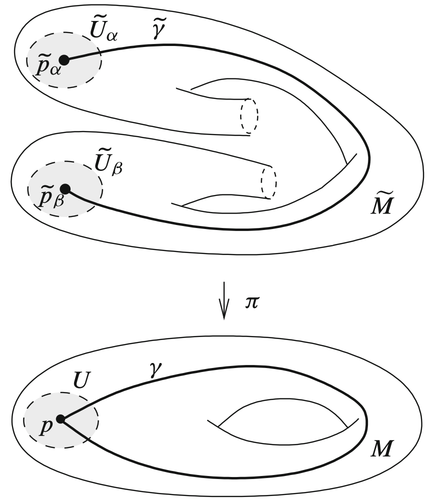

Figure 1.9: Proof that $\widetilde{U}_\alpha \cap \widetilde{U}_\beta = \varnothing$.

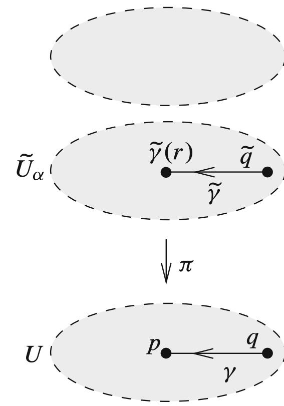

Figure 1.10: Proof that $\pi^{-1}(U) \subseteq \bigcup_\alpha \widetilde{U}_\alpha$.

γ starting at $\widetilde{q}$. It follows that $\pi(\widetilde{\gamma}(r)) = \gamma(r) = p$. Therefore $\widetilde{\gamma}(r) = \widetilde{p}_\alpha$ for some α, and $d_{\widetilde{g}}(\widetilde{q}, \widetilde{p}_\alpha) \le L_g(\widetilde{\gamma}) = r < \varepsilon$, so $\widetilde{q} \in \widetilde{U}_\alpha$.

It remains only to show that $\pi : \widetilde{U}_\alpha \to U$ is a diffeomorphism for each α. It is certainly a local diffeomorphism (because π is). It is bijective because its inverse can be constructed explicitly: it is the map sending each radial geodesic starting at p to its lift starting at $\widetilde{p}_\alpha$. This completes the proof. □

complete iff **Corollary 1.4.22.** *Suppose $\widetilde{M}$ and M are connected Riemannian manifolds, and $\pi : \widetilde{M} \to M$ is a Riemannian covering map. Then M is complete if and only if $\widetilde{M}$ is complete.*

Proof. A Riemannian covering map is, in particular, a local isometry. Thus if $\widetilde{M}$ is complete, π satisfies the hypotheses of Theorem Riemann local isometry complete is cover 1.4.21, which implies that M is also complete.

Conversely, suppose M is complete. Let $\widetilde{p} \in \widetilde{M}$ and $\widetilde{v} \in T_{\widetilde{p}}\widetilde{M}$ be arbitrary, and let $p = \pi(\widetilde{p})$ and $v = d\pi_{\widetilde{p}}(\widetilde{v})$. Completeness of M implies that the geodesic γ with $\gamma(0) = p$ and $\gamma'(0) = v$ is defined for all $t \in \mathbb{R}$, and then its lift $\widetilde{\gamma} : \mathbb{R} \to \widetilde{M}$ starting at $\widetilde{p}$ is a geodesic in $\widetilde{M}$ with initial velocity $\widetilde{v}$, also defined for all t. □

The Hopf-Rinow theorem shows that any two points in a complete, connected Riemannian manifold can be joined by a minimizing geodesic segment. The next proposition gives a refinement of that statement.

omotopy class **Proposition 1.4.23.** *Suppose (M, g) is a complete, connected Riemannian manifold, and $p, q \in M$. Every path-homotopy class of paths from p to q contains a geodesic segment γ that minimizes length among all admissible curves in the same path-homotopy class.*

Proof. Let $\pi : \widetilde{M} \to M$ be the universal covering manifold of M, endowed with the pullback metric $\widetilde{g} = \pi^* g$. Given $p, q \in M$ and a path $\sigma : [0,1] \to M$ from p to q, choose a point $\widetilde{p} \in \pi^{-1}(p)$ and let $\widetilde{\sigma} : [0,1] \to \widetilde{M}$ be the lift of σ starting at $\widetilde{p}$, and set $\widetilde{q} = \widetilde{\sigma}(1)$. By the Hopf-Rinow theorem, there is a minimizing $\widetilde{g}$-geodesic segment $\widetilde{\gamma}$ from $\widetilde{p}$ to $\widetilde{q}$, and because π is a local isometry, $\pi \circ \widetilde{\gamma}$ is a geodesic in M from p to q. If γ_1 is any other admissible curve from p

to q in the same path-homotopy class, then by the monodromy theorem, its lift $\widetilde{\gamma}_1$ starting at $\widetilde{p}$ also ends at $\widetilde{q}$. Thus $\widetilde{\gamma}_1$ is no longer than $\widetilde{\gamma}$, which implies γ_1 is no longer than γ. □

Suppose (M, g) is a connected Riemannian manifold. A **closed geodesic** in M is a non-constant geodesic segment $\gamma : [a, b] \to M$ such that $\gamma(a) = \gamma(b)$ and $\gamma'(a) = \gamma'(b)$.

Proposition 1.4.24. *A geodesic segment is closed if and only if it extends to a periodic geodesic defined on all of* $\mathbb{R}$.

Round spheres have the remarkable property that all of their geodesics are closed when restricted to appropriate intervals. Of course, this is not typically the case, even for compact Riemannian manifolds; but it is natural to wonder whether closed geodesics exist in more general manifolds. Much work has been done in Riemannian geometry to determine how many closed geodesics exist in various situations. Here we can only touch on the simplest case.

A continuous path $\sigma : [0, 1] \to M$ is called a **loop** if $\sigma(0) = \sigma(1)$. Two loops are said to be **freely homotopic** if they are homotopic through closed paths (but not necessarily preserving the base point). This is an equivalence relation on the set of all loops in M, and an equivalence class is called a **free homotopy class**. The **trivial free homotopy class** is the equivalence class of any constant path.

Lemma 1.4.25. *Given a connected manifold M and a point $x \in M$, then a loop based at x represents the trivial free homotopy class if and only if it represents the identity element of $\pi_1(M, x)$.*

Proof. Assume that a loop σ based at x is freely homotopic to the constant loop c_y for $y \in M$, that is, there exists a homotopy $H : [0, 1] \times [0, 1] \to M$ satisfying

$$H(s, 0) = \sigma(s), H(s, 1) = y \ \text{ for } \ s \in [0, 1], \quad H(0, t) = H(1, t) \ \text{ for } \ t \in [0, 1].$$

Let $\gamma(t) = H(0, t)$, then we have $[\overline{\gamma} * \sigma * \gamma] = [c_y]$. Then we get $[\sigma] = [\gamma] * [c_y] * [\overline{\gamma}] = [c_x]$, so σ represents the identity element of $\pi_1(M, x)$ □

The next proposition shows that closed geodesics are easy to find on compact Riemannian manifolds that are not simply connected.

Proposition 1.4.26. *Suppose (M, g) is a compact, connected Riemannian manifold. Every nontrivial free homotopy class in M is represented by a closed geodesic that has minimum length among all admissible loops in the given free homotopy class.*

1.4.3 Distance functions

Suppose (M, g) is a connected Riemannian manifold and $S \subseteq M$ is any subset. For each point $x \in M$, we define the **distance from x to S** to be

$$d_g(x, S) = \inf\{d_g(x, s) : s \in S\}.$$

function lem **Lemma 1.4.27.** *Suppose (M, g) is a connected Riemannian manifold and $S \subseteq M$ is any subset.*

(a) $d_g(x, S) \le d_g(x, y) + d_g(y, S)$ *for all* $x, y \in M$.

(b) $x \mapsto d_g(x, S)$ *is a continuous function on* M.

Proof. For all $x, y \in M$ and $s \in S$ we have

$$d_g(x, s) \le d_g(x, y) + d_g(y, s).$$

Therefore $d_g(x, S) \le d_g(x, y) + d_g(y, S)$. With this, we then get

$$|d_g(x, S) - d_g(y, S)| \le d_g(x, y),$$

so $d_g(\cdot, S)$ is continuous. □

The simplest example of a distance function occurs when the set S is just a singleton, $S = \{p\}$. Inside a geodesic ball around p, Corollary 1.4.11 shows that $d_g(x, S) = r(x)$, the radial distance function, and Corollary 1.4.9 shows that it has unit gradient where it it smooth (everywhere inside the geodesic ball except at p itself). The next theorem is a far-reaching generalization of that result.

Theorem 1.4.28. *Suppose (M, g) is a connected Riemannian manifold, $S \subseteq M$ is arbitrary, and $f : [0, \infty) \to M$ is the distance to S, that is, $f(x) = d_g(x, S)$ for all $x \in M$. If f is continuously differentiable on some open subset $U \subseteq M \setminus S$, then $|\operatorname{grad} f| \equiv 1$ on U.*

Proof. Suppose $U \subseteq M \setminus S$ is an open subset on which f is continuously differentiable, and $x \in U$. We will show first that $|\operatorname{grad} f|_x| \le 1$, and then that $|\operatorname{grad} f|_x| \ge 1$.

To prove the first inequality, we may assume $\operatorname{grad} f|_x \ne 0$, for otherwise the inequality is trivial. Let $v \in T_xM$ be any unit vector, and let γ be the unit-speed geodesic with $\gamma(0) = x$ and $\gamma'(0) = v$. Then for every positive t sufficiently small that $\gamma|_{[0,t]}$ is minimizing, Lemma 1.4.27 gives $d_g(x, S) \le d_g(x, \gamma(t)) + d_g(\gamma(t), S)$, or equivalently $f(x) \le t + f(\gamma(t))$. Therefore, since f is differentiable at x,

$$\frac{d}{dt}\Big|_{t=0} f(\gamma(t)) = \lim_{t \to 0^+} \frac{f(\gamma(t)) - f(\gamma(0))}{t} \le 1.$$

In particular, taking $v = (\operatorname{grad} f|_x)/|\operatorname{grad} f|_x|$ (the unit vector in the direction of $\operatorname{grad} f|_x$), we obtain

$$\frac{d}{dt}\Big|_{t=0} f(\gamma(t)) = df_x(\gamma'(0)) = \langle \operatorname{grad} f|_x, v\rangle = |\operatorname{grad} f|_x|.$$

This proves $|\operatorname{grad} f|_x| \le 1$.

To prove the reverse inequality, assume for the sake of contradiction that $|\operatorname{grad} f|_x| < 1$. Since we are assuming that $\operatorname{grad} f$ is continuous on U, there exist $\delta, \varepsilon > 0$ suchthat $\overline{B_\varepsilon(x)}$ is a closed geodesic ball contained in U and $|\operatorname{grad} f| \le 1 - \delta$ on it. Let c be a positive constant less than $\delta\varepsilon$. By definition of $d_g(x, S)$, there is an admissible curve $\sigma : [0, b] \to M$ (which we may assume to be parametrized by arc length) such that $\sigma(0) = x, \sigma(b) \in S$ and

$$b - c = L_g(\sigma) - c < d_g(x, S) < L_g(\sigma) = b.$$

Since we are assuming $\overline{B_\varepsilon(x)} \subseteq U \subseteq M \setminus S$, we have that $b > \varepsilon$, so $\sigma|_{[\varepsilon,b]}$ is an admissible curve from $\sigma(\varepsilon)$ to S. On the one hand,

$$f(\sigma(\varepsilon)) = d_g(\sigma(\varepsilon), S) \le L_g(\sigma|_{[\varepsilon,b]}) = b - \varepsilon < d_g(x, S) + c - \varepsilon = f(x) + c - \varepsilon. \tag{4.12}$$

On the other hand, for $t \in [0, \varepsilon]$, the fact that $\sigma(t) \in \overline{B_\varepsilon(x)}$ implies

$$\left|\frac{d}{dt} f(\sigma(t))\right| = |\langle \operatorname{grad} f|_{\sigma(t)}, \sigma'(t)\rangle| \le |\operatorname{grad} f|_{\sigma(t)}||\sigma'(t)| \le 1 - \delta.$$

Thus $f(\sigma(t)) \geq f(x) - (1-\delta)t$ for all such t. In particular, for $t = \varepsilon$, this means that

$$f(\sigma(\varepsilon)) \geq f(x) - (1-\delta)\varepsilon. \tag{4.13}$$

Combining (4.12) and (4.13) yields $c > \delta\varepsilon$, contradicting our choice of c. □

Motivated by the previous theorem, if (M, g) is a Riemannian manifold and $U \subseteq M$ is an open subset, we define a **local distance function** on U to be a continuously differentiable function $f : U \to \mathbb{R}$ such that $|\mathrm{grad}\, f|_g \equiv 1$ in U. First, we develop some important general properties of local distance functions.

Theorem 1.4.29. *Suppose (M, g) is a Riemannian manifold and f is a smooth local distance function on an open subset $U \subseteq M$. Then $\nabla_{\mathrm{grad}\, f}(\mathrm{grad}\, f) = 0$, and each integral curve of $\mathrm{grad}\, f$ is a unit-speed geodesic.*

Proof. Let $F \in \mathfrak{X}(U)$ denote the unit vector field $\mathrm{grad}\, f$. The definition of the gradient shows that for every vector field W, we have

$$Wf = df(W) = \langle F, W \rangle.$$

and therefore

$$Ff = \langle F, F \rangle = |\mathrm{grad}\, f|^2 \equiv 1.$$

For every smooth vector field W on U, we have

$$\begin{aligned}
\langle W, \nabla_F F \rangle &= F\langle W, F \rangle - \langle \nabla_F W, F \rangle \\
&= FWf - \langle \nabla_W F, F \rangle - \langle [F, W], F \rangle \\
&= FWf - [F, W]f - \frac{1}{2} W\langle F, F \rangle \\
&= WFf - \frac{1}{2} W|F|^2 \\
&= 0.
\end{aligned}$$

Since W is arbitrary, this proves that $\nabla_F F = 0$.

If $\gamma : I \to U$ is an integral curve of F, then the fact that γ' is extendible implies

$$D_t\gamma'(t) = \nabla_{\gamma'}\gamma'(t) = \nabla_F F|_{\gamma(t)} = 0,$$

so γ is a geodesic. □

Lemma 1.4.30. *Suppose (M, g) is a Riemannian manifold, $K \subseteq M$, and $f : K \to \mathbb{R}$ is a continuous function whose restriction to some open set $W \subseteq K$ is a smooth local distance function. For every admissible curve $\sigma : [a_0, b_0] \to K$ such that $\sigma(a_0, b_0) \subseteq W$, we have*

$$L_g(\sigma) \geq |f(\sigma(b_0)) - f(\sigma(a_0))|.$$

Proof. This is proved exactly as in (4.7), noting that the only properties of r we used in that computation were that it is continuous on the image of σ and continuously differentiable on $\sigma(a_0, b_0)$, and its gradient has unit length there. □

The next theorem and its corollary explain why the name "local distance function" is justified. Its proof is an adaptation of the proof of Proposition 1.4.10.

Theorem 1.4.31. *Suppose (M, g) is a Riemannian manifold, $U \subseteq M$ is an open subset, $S \subseteq U$, and $f : U \to [0, \infty)$ is a continuous function such that $f^{-1}(0) = S$ and f is a smooth local distance* *function on $U \setminus S$. Then there is a neighborhood $U_0 \subseteq U$ of S in which $f(x)$ is equal to the distance in M from x to S.*

Proof. For each $p \in S$, there are positive numbers ε_p, δ_p such that $B_{\varepsilon_p}(p)$ is a uniformly δ_p-normal geodesic ball and $B_{2\varepsilon_p}(p) \subseteq U$. This means that $B_{\varepsilon_p}(p)$ is contained in the open geodesic ball of radius δ_p around each of its points. In particular, $B_{\varepsilon_p}(p) \subseteq B_\delta(p)$, which means that $\varepsilon_p \le \delta_p$, and thus every geodesic starting at a point of $B_{\varepsilon_p}(p)$ is defined at least for $t \in (-\varepsilon_p, \varepsilon_p)$. Let U_0 be the union of all of the geodesic balls $B_{\varepsilon_p}(p)$ for $p \in S$, which is a neighborhood of S contained in U.

Let $x \in U_0$ be arbitrary, and let $c = f(x)$. We will show that $d_g(x, S) = f(x)$. If $x \in S$, then $d_g(x, S) = 0 = c$, so we may as well assume $x \notin S$.

There is some $p \in S$ such that $x \in B_{\varepsilon_p}(p)$, which means that $d_g(x, S) < \varepsilon_p$ and geodesics starting at x are defined at least on $(-\varepsilon_p, \varepsilon_p)$. Let $\sigma : [0, b] \to B_{\varepsilon_p}(p)$ be the radial geodesic segment from p to x, we define

$$\mathcal{A} = \{t \in [0, b] : \sigma(t) \in S\},$$

and set $a = \sup \mathcal{A}$. Since $x \notin S$ and S is closed, we have $a < b$, and $\sigma(a) \in S$. It follows from Lemma 1.4.30 that

$$L_g(\sigma) \ge L_g(\sigma|_{[a,b]}) \ge |f(x) - f(\sigma(a))| = c,$$

and we conclude that $c \le L_g(\sigma) < \varepsilon_p$ as well.

Let $\gamma : (-\varepsilon_p, \varepsilon_p) \to U$ be the unit-speed geodesic starting at x with initial velocity equal to $-\operatorname{grad} f|_x$. By Theorem 1.4.29 and uniqueness of geodesics, γ coincides with an integral curve of $-\operatorname{grad} f$ as long as $\gamma(t) \in U \setminus S$, which is to say as long as $f(\gamma(t)) \ne 0$. For all such t we have

$$\frac{d}{dt} f(\gamma(t)) = \langle \operatorname{grad} f|_{\gamma(t)}, \gamma'(t)\rangle = -|\operatorname{grad} f|_{\gamma(t)}|^2 = -1.$$

so $f(\gamma(t)) = c - t$ as long as $t < c$, and by continuity, $f(\gamma(c)) = 0$. This means that $f(\gamma(c)) \in S$, and $\gamma|_{[0,c]}$ is a curve segment of length c connecting x with S, so $d_g(x, S) \le c$.

To prove the reverse inequality, suppose $\sigma : [a, b] \to M$ is any admissible curve starting at x and ending at a point of S. Assume first that $\sigma(t) \in U$ for all $t \in [a, b]$, and let $b_0 \in [a, b]$ be the first time that $\sigma(b_0) \in S$. Then Lemma 1.4.30 shows that

$$L_g(\sigma) \ge L_g(\sigma|_{[a,b_0]}) = |f(\sigma(b_0)) - f(\sigma(a))| = c.$$

On the other hand, suppose $\sigma(t) \in M \setminus U$ for some t. The fact $B_{\varepsilon_p}(p) \subseteq B_{2\varepsilon_p}(p) \subseteq U$ and the triangle inequality imply that there is a first time $b_0 \in [a, b]$ such that $d_g(x, \sigma(b_0)) \ge \varepsilon_p$. Then

$$L_g(\sigma) \ge L_g(\sigma|_{[a,b_0]}) \ge d_g(x, \sigma(b_0)) \ge \varepsilon_p > c.$$

Taken together, these two inequalities show that $L_g(\sigma) \ge c$ for every such σ, which implies $d_g(x, S) \ge c$. □

Corollary 1.4.32. *Let (M, g) be a Riemannian manifold, and let f be a smooth local distance function on an open subset $U \subseteq M$. If c is a real number such that $S = f^{-1}(c)$ is nonempty, then there is a neighborhood U_0 of S in U on which $|f(x) - c|$ is equal to the distance in M from x to S.*

Proof. Observe that

$$\operatorname{grad} |f(x) - c| = \frac{f(x) - c}{|f(x) - c|} \cdot \operatorname{grad} f.$$

Therefore $|f(x) - c|$ is a local distance function on $U \setminus S$. □

The following result is a global form of Theorem 1.4.31.

Theorem 1.4.33. *Suppose (M, g) is a complete, connected Riemannian manifold, $S \subseteq M$ is a closed subset and $f : M \to \mathbb{R}$ is a continuous function such that $f^{-1}(0) = S$ and f is smooth with unit gradient on $M \setminus S$. Then $f(x) = d_g(x, S)$ for all $x \in M$.*

Proof. Let $x \in M \setminus S$ be arbitrary, and le $c = f(x)$. Let $\gamma : \mathbb{R} \to M$ be the unit-speed geodesic starting at x with initial velocity equal to $-\operatorname{grad} f|_x$. By Theorem 1.4.29 and uniqueness of geodesics, γ coincides with an integral curve of $-\operatorname{grad} f$ as long as $\gamma(t) \in M \setminus S$, which is to say as long as $f(\gamma(t)) \neq 0$. For all such t we have

$$\frac{d}{dt} f(\gamma(t)) = \langle \operatorname{grad} f|_{\gamma(t)}, \gamma'(t)\rangle = -|\operatorname{grad} f|_{\gamma(t)}|^2 = -1.$$

so $f(\gamma(t)) = c - t$ as long as $t < c$, and by continuity, $f(\gamma(c)) = 0$. This means that $f(\gamma(c)) \in S$, and $\gamma|_{[0,c]}$ is a curve segment of length c connecting x with S, so $d_g(x, S) \leq c$.

To prove the reverse inequality, suppose $\sigma : [a, b] \to M$ is any admissible curve starting at x and ending at a point of S. Let $b_0 \in [a, b]$ be the first time that $\sigma(b_0) \in S$. Then Lemma 1.4.30 shows that

$$L_g(\sigma) \geq L_g(\sigma|_{[a,b_0]}) = |f(\sigma(b_0)) - f(\sigma(a))| = c.$$

Therefore $L_g(\sigma) \geq c$ for every such σ, which implies $d_g(x, S) \geq c$. □

Distance functions for embedded submanifolds

The most important local distance functions are those associated with embedded submanifolds. As we will see, such distance functions are always smooth near the manifold.

Suppose (M, g) is a Riemannian n-manifold (without boundary) and $P \subseteq M$ is an embedded k-dimensional submanifold. Let NP denote the normal bundle of P in M, and let $U \subseteq M$ be a normal neighborhood of P in M, which is the diffeomorphic image of a certain open subset $V \subseteq NP$ under the normal exponential map. (Such a neighborhood always exists by Theorem 1.3.25.) We begin by constructing generalizations of the radial distance function and radial vector field.

Proposition 1.4.34. *Let P be an embedded submanifold of a Riemannian manifold (M, g) and let U be any normal neighborhood of P in M. There exist a unique continuous function $r : U \to [0, \infty)$ and smooth vector field ∂_r on $U \setminus P$ that have the following coordinate representations in terms of any Fermi coordinates (x, v) for P on a subset $U_0 \subseteq U$:*

$$r(x, v) = \sqrt{(v^1)^2 + \cdots + (v^{n-k})^2}, \tag{4.14}$$

$$\partial_r = \frac{v^1}{r(x,v)} \frac{\partial}{\partial v^1} + \cdots + \frac{v^{n-k}}{r(x,v)} \frac{\partial}{\partial v^{n-k}}. \tag{4.15}$$

The function r is smooth on $U \setminus P$, and r^2 is smooth on all of U.

Proof. The uniqueness, continuity, and smoothness claims follow immediately from the coordinate expressions (4.14) and (4.15), so we need only prove that r and ∂_r can be globally defined so as to have the indicated coordinate expressions in any Fermi coordinates.

Let $V \subseteq NP$ be the subset that is mapped diffeomorphically onto U by the normal exponential map E. Define a function $\rho : V \to [0, \infty)$ by $\rho(p, v) = |v|_g$, and define $r : U \to [0, \infty)$

by $r = \rho \circ E^{-1}$. Any Fermi coordinates for P are defined by choosing local coordinates $(x^1, \dots, x^k)$ for P and a local orthonormal frame (E_j) for NP, and assigning the coordinates $(x^1, \dots, x^k, v^1, \dots, v^{n-k})$ to the point $E(p, v^\alpha E_\alpha|_p)$. (Here we are using the summation convention with Greek indices running from 1 to $n-k$.) Because the frame is orthonormal, for each $(p, v) = (p, v^\alpha E_\alpha|_p) \in V$ we have

$$r(E(p,v))^2 = \rho(p,v)^2 = (v^1)^2 + \cdots + (v^{n-k})^2,$$

which shows that r has the coordinate representation (4.14).

To define ∂_r, let q be an arbitrary point in $U \setminus P$. Then $q = \exp_p(tv)$ for a unique $(p,v) \in V$, and the curve $\gamma : [0,1] \to U$ given by $\gamma(t) = \exp_p(tv)$ is a geodesic from p to q. Define

$$\partial_r|_q = \frac{1}{r(q)} \gamma'(1), \tag{4.16}$$

which is independent of the choice of coordinates. Proposition 1.3.26 shows that in any Fermi coordinates, if we write $v = v^\alpha E_\alpha|_p$, then γ has the coordinate formula

$$\gamma(t) = (x^1(p), \dots, x^k(p), tv^1, \dots, tv^{n-k}),$$

and therefore $\gamma'(t) = v^\alpha \partial/\partial v^\alpha|_{\gamma(t)}$. It follows that ∂_r has the coordinate formula (4.15). □

By analogy with the special case in which P is a point, we call r the **radial distance function for P** and ∂_r the **radial vector field for P**.

Theorem 1.4.35 (Gauss Lemma for Submanifolds). *Let P be an embedded submanifold of a Riemannian manifold (M,g), let U be a normal neighborhood of P in M, and let r and ∂_r be defined as in Proposition 1.4.34. On $U \setminus P$, ∂_r is a unit vector field orthogonal to the level sets of r.*

Proof. The proof is a dressed-up version of the proof of the ordinary Gauss lemma. Let $q \in U \setminus P$ be arbitrary, and let $(x^1, \dots, x^k, v^1, \dots, v^{n-k})$ be the coordinate representation of q in some choice of Fermi coordinates associated with a local orthonormal frame (E_j) for NP. As in the proof of Proposition 1.4.34, $q = \gamma(1)$ where γ is the geodesic $\exp_p(tv)$ for some $p \in P$ and $v = v^\alpha E_\alpha|_p \in N_pM$. Since the frame (E_j) is orthonormal, we have

$$|\gamma'(0)| = |v|_g = \sqrt{(v^1)^2 + \cdots + (v^{n-k})^2} = r(q).$$

Since geodesics have constant speed, it follows that $|\gamma'(1)|_g = r(q)$ as well, and then (4.16) shows that $\partial_r|_q$ is a unit vector.

Next we show that ∂_r is orthogonal to the level sets of r. Suppose $q \in U \setminus P$, and write $q = \exp_{p_0}(v_0)$ for some $p_0 \in P$ and $v_0 \in N_{p_0}P$ with $v_0 \neq 0$. The coordinate representation (4.14) shows that $r^{-1}(r(q))$ is a regular level set, and hence an embedded submanifold of U.

Let $w \in T_qM$ be an arbitrary vector tangent to this level set, and let $\sigma : (-\varepsilon, \varepsilon) \to U$ be a smooth curve lying in the same level set, with $\sigma(0) = q$ and $\sigma'(0) = w$. We can write $\sigma(s) = \exp_{x(s)}(v(s))$, where $x(s) \in P$ and $v(s) \in N_{x(s)}P$ with $|v(s)|_g = r(q)$. The initial condition $\sigma(0) = q$ translates to $x(0) = p_0$ and $v(0) = v_0$. Define a smooth one-parameter family of curves $\Gamma : (-\varepsilon, \varepsilon) \times [0, r(q)] \to M$ by

$$\Gamma(s,t) = \exp_{x(s)}\left(\frac{t}{r(q)} v(s)\right).$$

Since $|v(s)|_g = r(q)$, each Γ_s is a unit-speed geodesic.

Note that $\Gamma(0,t) = \exp_{p_0}(tv_0/r(q))$, we have the following endpoint conditions:

$$\begin{aligned}
\partial_s\Gamma(0,0) &= \frac{d}{ds}\Big|_{s=0} \exp_{x(s)}(0) = \frac{d}{ds}\Big|_{s=0} x(s) = x'(0);\\
\partial_t\Gamma(0,0) &= \frac{d}{dt}\Big|_{t=0} \exp_{p_0}\left(\frac{v_0}{r(q)}t\right) = \frac{v_0}{r(q)};\\
\partial_s\Gamma(0,r(q)) &= \frac{d}{ds}\Big|_{s=0} \exp_{x(s)}(v(s)) = \frac{d}{ds}\Big|_{s=0} \sigma(s) = w;\\
\partial_t\Gamma(0,r(q)) &= \frac{d}{dt}\Big|_{t=r(q)} \exp_{p_0}\left(\frac{v_0}{r(q)}t\right) = \partial_r|_q.
\end{aligned}$$

Then the same computation as in (4.6) shows that $(d/dt)\langle\partial_s\Gamma, \partial_t\Gamma\rangle = 0$, and therefore

$$\langle w, \partial_r|_q\rangle = \langle\partial_s\Gamma(0,r(q)), \partial_t\Gamma(0,r(q))\rangle = \langle\partial_s\Gamma(0,0), \partial_t\Gamma(0,0)\rangle = \frac{1}{r(q)}\langle x'(0), v_0\rangle,$$

which is zero because $x'(0)$ is tangent to P and v_0 is normal to it. This proves that ∂_r is orthogonal to the level sets of r. □

Corollary 1.4.36. *Assume the hypotheses of Theorem 1.4.35.*

(a) *∂_r is equal to the gradient of r on $U \setminus P$.*

(b) *r is a local distance function.*

(c) *Each unit-speed geodesic $\gamma : [a,b) \to U$ with $\gamma'(a)$ normal to P coincides with an integral curve of ∂_r on (a,b).*

(d) *P has a tubular neighborhood in which the distance in M to P is equal to r.*

Proof. By direct computation in Fermi coordinates using formulas (4.14) and (4.15), $\partial_r(r) = 1$, which is equal to $|\partial_r|^2$ by the previous theorem. Thus $\partial_r = \operatorname{grad} r$ on $U \setminus P$ by Exercise **??**. Because $\operatorname{grad} r$ is a unit vector field, r is a local distance function. By Proposition 1.4.29, the geodesics in U that start normal to P are represented in any Fermi coordinates by $(x^1(p), \ldots, x^k(p), tv^1, \ldots, tv^{n-k})$, and such a geodesic has unit speed if and only if $(v^1)^2 + \cdots + (v^{n-k})^2 = 1$. Another direct computation shows that each such curve is an integral curve of ∂_r wherever $r \neq 0$.

Finally, to prove (d), note that Theorem 1.4.31 shows that there is some neighborhood $\widetilde{U}_0$ of P in M on which $r(x) = d_g(x,P)$; if we take U_0 to be a tubular neighborhood of P in $\widetilde{U}_0$, then U_0 satisfies the conclusion. □

When P is compact, we can say more.

Theorem 1.4.37. *Suppose (M,g) is a connected Riemannian manifold, $P \subseteq M$ is a compact submanifold, and U_ε is an ε-tubular neighborhood of P. Then U_ε is also an ε-neighborhood in the metric space sense, and inside U_ε, the distance in M to P is equal to the function r defined in Proposition 1.4.34.*

Proof. First we show that r can be extended continuously to $barU_\varepsilon$ by setting $r(q) = \varepsilon$ for $q \in \partial U_\varepsilon$. Indeed, suppose $q \in \partial U_\varepsilon$ and q_i is any sequence of points in U_ε converging to q. Then $\overline{\lim}_i\, r(q_i) \leq \varepsilon$ because $r(q_i) < \varepsilon$ for each i. Let $c = \underline{\lim}_i\, r(q_i)$, we will prove the result by showing that $c = \varepsilon$. Suppose for the sake of contradiction that $c < \varepsilon$. By passing to a subsequence, we may assume that $r(q_i) \to c$. We can write $q_i = \exp_{p_i}(v_i)$ for $p_i \in P$ and $v_i \in N_{p_i}P$, and because P is compact and $\lim_i |v|_i = \lim_i r(q_i) \to c$, we can pass to a further subsequence and assume that $(p_i, v_i) \to (p, v) \in NP$ with $|v|_g = c < \varepsilon$. Then we have $q = \lim_i q_i = \lim_i \exp_{p_i}(v_i) = \exp_p v$, which lies in the open set U_ε, contradicting our assumption that $q = \partial U_\varepsilon$. Henceforth, we regard r as a continuous function on $\overline{U}_\varepsilon$.

Now to prove the theorem, let W_ε denote the ε-neighborhood of P in the metric space sense. Suppose first that $q \in M \setminus U_\varepsilon$, and suppose $\sigma : [a, b] \to M$ is any admissible curve from a point of P to q. There is a first time $b_0 \in [a, b]$ that $\sigma(b_0) \in \partial U_\varepsilon$, and then Lemma 1.4.30 shows that

$$L_g(\sigma) \geq L_g(\sigma|_{[a,b_0]}) \geq |r(\sigma(b_0)) - r(\sigma(a))| = \varepsilon.$$

Thus $q \notin U\varepsilon \Rightarrow q \notin W_\varepsilon$, or equivalently $W\varepsilon \subseteq U_\varepsilon$.

Conversely, suppose $q \in U_\varepsilon$. Then q is connected to P by a geodesic segment of length $r(q)$, so $d_g(q, P) \leq r(q)$. To prove the reverse inequality, suppose $\sigma : [a, b] \to M$ is any admissible curve starting at a point of P and ending at q. If $\sigma(t)$ remains in U_ε for all $t \in [a, b]$, then Lemma 1.4.30 shows that

$$L_g(\sigma) \geq L_g(\sigma|_{[a_0,b]}) \geq |r(\sigma(b)) - r(\sigma(a_0))| = r(q),$$

where a_0 is the last time that $\sigma(a_0) \in P$. On the other hand, if $\sigma(t)$ does not remain in U_ε, then there is a first time b_0 such that $\sigma(b_0) \in \partial U_\varepsilon$, and the argument in the preceding paragraph shows that $L_g(\sigma) \geq \varepsilon > r(q)$. Thus $d_g(q, P) = r(q)$ for all $q \in U_\varepsilon$. Since $r(q) < \varepsilon$ for all such q, it follows also that $U_\varepsilon \subseteq W_\varepsilon$. $\square$

1.4.4 Semigeodesic coordinates

Local distance functions can be used to build coordinate charts near submanifolds in which the metric has a particularly simple form. We begin by describing the kind of coordinates we are looking for.

Let (M, g) be an n-dimensional Riemannian manifold. Smooth local coordinates $(x^1, \dots, x^n)$ on an open subset $U \subseteq M$ are called **semigeodesic coordinates** if each x^n-coordinate curve $t \mapsto (X^1, \dots, x^{n-1}, t)$ is a unit-speed geodesic that meets each level set of x^n orthogonally.

Because of the distinguished role played by the last coordinate function, we will use the summation convention with Latin indices running from 1 to n and Greek indices running from 1 to $n-1$.

We will see below that semigeodesic coordinates are easy to construct. But first, let us develop some alternative characterizations of them.

Proposition 1.4.38 (Characterizations of Semigeodesic Coordinates). *Let (M, g) be a Riemannian n-manifold, and let $(x^1, \dots, x^n)$ be smooth coordinates on an open subset of M. The following are equivalent:*

(a) (x^i) *are semigeodesic coordinates.*

(b) $|\partial_n|_g \equiv 1$ *and* $\langle \partial_\alpha, \partial_n \rangle \equiv 0$ *for* $\alpha = 1, \dots, n-1$.

(c) $|dx^n|_g \equiv 1$ *and* $\langle dx^\alpha, dx^n \rangle \equiv 0$ *for* $\alpha = 1, \dots, n-1$.

(d) $|\operatorname{grad} x^n|_g \equiv 1$ *and* $\langle \operatorname{grad} x^\alpha, \operatorname{grad} x^n \rangle \equiv 0$ *for* $\alpha = 1, \dots, n-1$.

(e) x^n *is a local distance function and* $(x^1, \dots, x^{n-1})$ *are constant along the integral curves of* $\operatorname{grad} x^n$.

(f) $\operatorname{grad} x^n \equiv \partial_n$.

Proof. We begin by showing that $(b) \Leftrightarrow (c) \Leftrightarrow (d) \Leftrightarrow (e)$ and $(c) \Leftrightarrow (f)$. Note that (b) is equivalent to the coordinate matrix of g having the block form $\left(\begin{smallmatrix} * & 0 \\ 0 & 1 \end{smallmatrix}\right)$, where the asterisk represents an arbitrary $(n-1) \times (n-1)$ positive definite symmetric matrix, while (c) is equivalent to the inverse matrix having the same form. It follows from Cramer's rule that the matrix of g has this form if and only if its inverse does, and thus (b) is equivalent to (c).

The equivalence of (c) and (d) follows from the definitions of the gradient and of the inner product on 1-forms: for all $1 \le i, j \le n$,

$$\langle dx^i, dx^j \rangle = \langle (dx^i)^\sharp, (dx^j)^\sharp \rangle = \langle \operatorname{grad} x^i, \operatorname{grad} x^j \rangle.$$

The equivalence of (d) and (e) also follows from the definition of the gradient:

$$\langle \operatorname{grad} x^\alpha, \operatorname{grad} x^n \rangle = dx^\alpha(\operatorname{grad} x^n) = (\operatorname{grad} x^n)(x^\alpha)$$

for each α, which means that x^α is constant along the $\operatorname{grad} x^n$ integral curves if and only if $\langle \operatorname{grad} x^\alpha, \operatorname{grad} x^n \rangle = 0$. Finally, by examining the individual components of the coordinate formula $\operatorname{grad} x^n = g^{nj}\partial_j$, we see that (c) is also equivalent to (f).

To complete the proof, we show that $(a) \Leftrightarrow (b)$. Assume first that (a) holds. Because the x^n-coordinate curves have unit speed, it follows that $|\partial_n|_g \equiv 1$. The tangent space to any x^n-level set is spanned at each point by $\partial_1, \dots, \partial_{n-1}$, and (a) guarantees that ∂_n is orthogonal to each of these, showing that (b) holds. Conversely, if we assume (b), the first part of the proof shows that (f) holds as well, so $|\operatorname{grad} x^n|_g = |\partial_n|_g \equiv 1$, showing that x^n is a local distance function. Thus the x^n-coordinate curves are also integral curves of $\operatorname{grad} x^n$ and hence are unit-speed geodesics by Theorem 1.4.29. The fact that $\langle \partial_\alpha, \partial_n \rangle = 0$ for $\alpha = 1, \dots, n-1$ implies that these geodesics are orthogonal to the level sets of x^n, thus proving (a). □

Part (b) of this proposition leads to the following simplified coordinate representations for the metric and Christoffel symbols in semigeodesic coordinates. Recall that implied summations with Greek indices run from 1 to $n-1$.

Corollary 1.4.39. *Let (x^i) be semigeodesic coordinates on an open subset of a Riemannian n-manifold (M, g).*

(a) The metric has the following coordinate expression:

$$g = (dx^n)^2 + g_{\alpha\beta}dx^\alpha dx^\beta$$

(b) The Christoffel symbols of g have the following coordinate expressions:

$$\begin{aligned}
\Gamma^n_{nn} &= \Gamma^\alpha_{nn} = \Gamma^n_{n\alpha} = \Gamma^n_{\alpha n} = 0, \\
\Gamma^n_{\alpha\beta} &= \Gamma^n_{\beta\alpha} = -\frac{1}{2}\partial_n g_{\alpha\beta}, \\
\Gamma^\beta_{\alpha n} &= \Gamma^\beta_{n\alpha} = \frac{1}{2} g^{\beta\gamma}\partial_n g_{\gamma\alpha}, \\
\Gamma^\gamma_{\alpha\beta} &= \widehat{\Gamma}^\gamma_{\alpha\beta}.
\end{aligned} \tag{4.17}$$

where for each fixed value of x^n, the quantities $\widehat{\Gamma}^\gamma_{\alpha\beta}$ are the Christoffel symbols in (x^α) coordinates for the induced metric $\widehat{g}$ on the level set $x^n = constant$.

Proof. Part (a) follows immediately from part (b) of Proposition 1.4.38, and (b) is proved by inserting $g_{nn} = 1$ and $g_{\alpha n} = g_{n\alpha} = 0$ into formula (3.9) for the Christoffel symbols. □

Proposition 1.4.38(e) gives us an effective way to construct semigeodesic coordinates: if r is any smooth local distance function (for example, the distance from a point or a smooth submanifold), just set $x^n = r$, choose any local coordinates $(x^1, \dots, x^n)$ for a level set of r, and then extend them to be constant along the integral curves of $\operatorname{grad} r$. Here are some explicit examples.

Example 1.4.40 (Examples of Semigeodesic Coordinates).

(a) **Fermi Coordinates for a Hypersurface**. Suppose P is an embedded hypersurface in a Riemannian manifold (M, g), and let $(x^1, \dots, x^{n-1}, v)$ be any Fermi coordinates for P on an open subset $U \subseteq M$. In this case, the function r defined by (4.14) is just $r(x, v) = \sqrt{v^2} = |v|$, so v is a local distance function on $U \setminus P$. It follows from Corollary 1.4.36 that there is a neighborhood U_0 of P on which $|v|$ is equal to the distance from P. Moreover, (4.15) reduces to $\partial_r = \pm\partial/\partial_v$, which is equal to $\operatorname{grad}|v|$ by Corollary 1.4.36, so it follows from Proposition 1.4.38(f) that Fermi coordinates for a hypersurface are automatically semigeodesic coordinates.

(b) **Boundary Normal Coordinates**. Suppose (M, g) is a smooth Riemannian manifold with nonempty boundary. The results in this section do not apply directly to manifolds with boundary, but we can embed M in its double $D(M)$ (Example **??**), extend the metric smoothly to $D(M)$, and construct Fermi coordinates $(x^1, \dots, x^{n-1}, v)$ for ∂M in $D(M)$. By replacing v with $-v$ if necessary, we can arrange that $v > 0$ in $\operatorname{Int} M$, and then these Fermi coordinates restrict to smooth boundary coordinates for M that are also semigeodesic coordinates. Such coordinates are called **boundary normal coordinates** for M.

(c) **Polar coordinates**. Polar coordinates for $\mathbb{R}^n$ are constructed by choosing a smooth local parametrization $\widehat{\psi} : \widehat{U} \to U \subseteq S^{n-1}$ for an open subset U of S^{n-1}, and defining $\widehat{\Psi} : \widehat{U} \times \mathbb{R}^+ \to \mathbb{R}^n$ by

$$\widehat{\Psi}(\theta^1, \dots, \theta^{n-1}, r) = r\widehat{\psi}(\theta^1, \dots, \theta^{n-1}).$$

It is straightforward to show that the differential of $\widehat{\Psi}$ vanishes nowhere, so $\widehat{\Theta} = \widehat{\Psi}^{-1}$ is a smooth coordinate map on the open subset $\mathcal{U} = \widehat{\Psi}(\widehat{U} \times \mathbb{R}^+) \subseteq \mathbb{R}^n \setminus \{0\}$. Familiar examples are ordinary polar coordinates in the plane and spherical coordinates in $\mathbb{R}^3$. Such coordinates have the property that the last coordinate function is $r(x) = |x|$.

Now let (M, g) be a Riemannian n-manifold, p a point in M, and φ any normal coordinate chart defined on a normal neighborhood V of p. For every choice of polar coordinates $(\mathcal{U}, \widehat{\Theta})$ for $\mathbb{R}^n \setminus \{0\}$, we obtain a smooth coordinate map $\Theta = \widehat{\Theta} \circ \varphi$ on an open subset of $V \setminus \{p\}$. Such coordinates are called **polar normal coordinates**. They have the property that the last coordinate function r is the radial distance function on V, and the other coordinates are constant along the integral curves of $\operatorname{grad} r$, so they are semigeodesic coordinates.

(d) **Polar Fermi Coordinates**. Now let P be an embedded submanifold of (M, g), and let $\varphi = (x^1, \dots, x^k, v^1, \dots, v^{n-k})$ be Fermi coordinates on a neighborhood U_0 of a point $p \in P$. Then any polar coordinate map $\widehat{\Theta}$ for $\mathbb{R}^{n-k}$ can be applied to the variables $(v^1, \dots, v^{n-k})$ to yield a coordinate chart $\Theta = (\mathrm{id}_{\mathbb{R}^k} \times \widehat{\Theta}) \circ \varphi$ on an open subset of $U_0 \setminus P$, taking values in $\mathbb{R}^k \times \mathbb{R}^{n-k-1} \times \mathbb{R}^+$. If we write the coordinate functions as

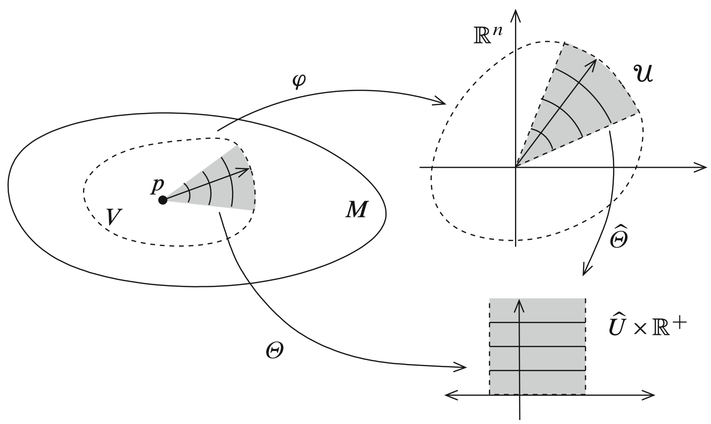

Figure 1.11: Polar normal coordinates.

Riemann Fermi coordinate prop

$(x^1, \ldots, x^k, \theta^1, \ldots, \theta^{n-k-1}, r)$, it follows from Proposition 1.3.26 that each coordinate curve $t \mapsto (x^1, \ldots, x^k, \theta^1, \ldots, \theta^{n-k-1}, t)$ is a unit-speed geodesic. Thus these are semi-geodesic coordinates, called polar Fermi coordinates. The polar normal coordinates described above are just the special case $P = \{p\}$.

Chapter 2

Curvature

2.1 The definition of curvature and basic properties

2.1.1 The curvature tensor

Let (M, g) be a Riemannian manifold and ∇ the Riemannian connection. The curvature tensor is the $(1,3)$-tensor defined by (Proposition connection second derivative 1.2.17)

$$\begin{aligned} R(X,Y)Z &= \nabla^2_{X,Y}Z - \nabla^2_{Y,X}Z \\ &= \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z \\ &= [\nabla_X, \nabla_Y]Z - \nabla_{[X,Y]}Z \end{aligned}$$

on vector fields X, Y, Z. First let's verify that this indeed defines a tensor on M.

ure is tensor **Proposition 2.1.1.** *The map R defined above is multilinear over $C^\infty(M)$, and thus defines a $(1,3)$-tensor field on M.*

Proof. The map R is obviously multilinear over $\mathbb{R}$. For $f \in C^\infty(M)$,

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

The same proof shows that R is linear over $C^\infty(M)$ in X, because $R(X,Y) = -R(Y,X)$ from the definition. For the variable Z, we first note that

$$\begin{aligned} \nabla^2_{X,Y}(fZ) &= \nabla_X\nabla_Y(fZ) - \nabla_{\nabla_X Y}(fZ) \\ &= \nabla_X((Yf)Z + f\nabla_Y Z) - \big((\nabla_X Y)f\big)Z - f\nabla_{\nabla_X Y}Z \\ &= (XYf)Z + (Yf)\nabla_X Z + (Xf)(\nabla_Y Z) + f\nabla_X\nabla_Y Z - \big((\nabla_X Y)f\big)Z - f\nabla_{\nabla_X Y}Z \\ &= \big(XYf - (\nabla_X Y)f\big)Z + (Yf)\nabla_X Z + (Xf)(\nabla_Y Z) + f\nabla^2_{X,Y}Z \\ &= (\nabla^2_{X,Y}f)Z + (Yf)\nabla_X Z + (Xf)(\nabla_Y Z) + f\nabla^2_{X,Y}Z. \end{aligned}$$

Repalce the place of X, Y and substrating, we get

$$R(X,Y)(fZ) = \nabla_{X,Y}(fZ) - \nabla_{Y,X}(fZ) = (R_{X,Y}f)Z + fR(X,Y)Z = fR(X,Y)Z.$$

where we use the fact that $R_{X,Y}f = 0$ for $f \in C^\infty(M)$. This completes the proof. $\square$

Thanks to this proposition, for each pair of vector fields $X, Y \in \mathfrak{X}(M)$, the map $R_{X,Y} : \mathfrak{X}(M) \to \mathfrak{X}(M)$ given by $Z \mapsto R(X,Y)Z$ is a smooth bundle endomorphism of TM, called the **curvature endomorphism determined by X and Y**. The tensor field R itself is called the **(Riemann) curvature endomorphism** or the **$(1,3)$-curvature tensor**.

The next proposition gives one reason for our interest in the curvature tensor.

Proposition 2.1.2. *The curvature tensor is a local isometry invariant: if (M,g) and $(\widetilde{M},\widetilde{g})$ are Riemannian or pseudo-Riemannian manifolds and $\varphi : M \to \widetilde{M}$ is a local isometry, then $\varphi^*\widetilde{R} = R$. More precisely,*

$$\widetilde{R}_{\varphi(p)}(d\varphi_p(X_p), d\varphi_p(Y_p))d\varphi_p(Z_p) = d\varphi_p(R_p(X_p,Y_p)Z_p)$$

for every $p \in M$ and $X, Y, Z \in X(M)$.

Proof. This question is local, so we can assume that φ is a isometry. In this case, the pull back connection $\varphi^*\widetilde{\nabla}$ is the Levi-civita connection on M (Proposition connection pull back 1.2.30), therefore the claim follows. □

As a $(1,3)$-tensor field, the curvature endomorphism can be written in terms of any local frame with one upper and three lower indices. We adopt the convention that the last index is the contravariant (upper) one. (This is contrary to our default assumption that covector arguments come first.) Thus, for example, the curvature endomorphism can be written in terms of local coordinates (x^i) as

$$R = R_{ijk}^l dx^i \otimes dx^j \otimes dx^k \otimes \partial_l.$$

where the coefficients R_{ijk}^l are defined by

$$R(\partial_i, \partial_j)\partial_k = R_{ijk}^l \partial_l.$$

The next proposition shows how to compute the components of R in coordinates.

Proposition 2.1.3. *Let (M,g) be a Riemannian or pseudo-Riemannian manifold. In terms of any smooth local coordinates, the components of the $(1,3)$-curvature tensor are given by*

$$R_{ijk}^l = \partial_i \Gamma_{jk}^l - \partial_j \Gamma_{ik}^l + \Gamma_{is}^l \Gamma_{jk}^s - \Gamma_{js}^l \Gamma_{ik}^s.$$

Proof. Note that

$$\begin{aligned}
R(\partial_i, \partial_j)\partial_k &= \nabla_{\partial_i}\nabla_{\partial_j}\partial_k - \nabla_{\partial_j}\nabla_{\partial_i}\partial_k - \nabla_{[\partial_i,\partial_j]}\partial_k \\
&= \nabla_{\partial_i}(\Gamma_{jk}^s \partial_s) - \nabla_{\partial_j}(\Gamma_{ik}^s \partial_s) \\
&= \partial_i \Gamma_{jk}^s \partial_s + \Gamma_{jk}^s \nabla_{\partial_i}\partial_s - \partial_j \Gamma_{ik}^s \partial_s - \Gamma_{ik}^s \nabla_{\partial_j}\partial_s \\
&= \partial_i \Gamma_{jk}^s \partial_s + \Gamma_{jk}^s \Gamma_{is}^l \partial_l - \partial_j \Gamma_{ik}^s \partial_s - \Gamma_{ik}^s \Gamma_{js}^l \partial_l \\
&= (\partial_i \Gamma_{jk}^l - \partial_j \Gamma_{ik}^l + \Gamma_{is}^l \Gamma_{jk}^s - \Gamma_{js}^l \Gamma_{ik}^s)\partial_l
\end{aligned}$$

therefore the claim follows. □

From the proof of Proposition Riemann curvature is tensor 2.1.1, we see that the curvature tensor acts trivially on functions. Now we want to extend this result. Recall that given vector fields X and Y, we have the dual map $R_{X,Y}^*$ defined by

$$(R_{X,Y}^*\omega)(Z) = \omega(R_{X,Y}Z).$$

Theorem 2.1.4 (Ricci Identities). *On a Riemannian or pseudo-Riemannian manifold M, the second total covariant derivatives of vector and tensor fields satisfy the following identities.*

(a) If Z is a smooth vector field, then

$$(\nabla^2_{X,Y} - \nabla^2_{Y,X})Z = R_{X,Y}Z. \tag{1.1}$$

(b) If ω is a smooth 1-form, then

$$R_{X,Y}\omega = \nabla^2_{X,Y}\omega - \nabla^2_{Y,X}\omega = -R^*_{X,Y}\omega \tag{1.2}$$

(c) If F is a smooth (r,s)-tensor field, then

$$\begin{aligned} &R_{X,Y}F(\omega^1,\dots,\omega^r,V_1,\dots,V_s) = (\nabla^2_{X,Y}F - \nabla^2_{Y,X}F)(\omega^1,\dots,\omega^r,V_1,\dots,V_s) \\ &= \sum_{i=1}^r F(\omega^1,\dots,R^*_{X,Y}\omega^i,\dots,\omega^r,Y_1,\dots,Y_s) - \sum_{j=1}^s F(\omega^1,\dots,\omega^r,Y_1,\dots,R_{X,Y}Y_j,\dots,Y_s). \end{aligned} \tag{1.3}$$

Proof. We first prove (1.2). Using (1.2.17) repeatedly, we compute

$$\begin{aligned} (\nabla_X\nabla_Y\omega)(Z) &= X((\nabla_Y\omega)(Z)) - (\nabla_Y\omega)(\nabla_X Z) \\ &= X\big(Y(\omega(Z)) - \omega(\nabla_Y Z)\big) - (\nabla_Y\omega)(\nabla_X Z) \\ &= XY(\omega(Z)) - \omega(\nabla_X\nabla_Y Z) - (\nabla_X\omega)(\nabla_Y Z) - (\nabla_Y\omega)(\nabla_X Z). \end{aligned}$$

Reversing the roles of X and Y, we get

$$(\nabla_X\nabla_Y\omega)(Z) = YX(\omega(Z)) - \omega(\nabla_Y\nabla_X Z) - (\nabla_Y\omega)(\nabla_X Z) - (\nabla_X\omega)(\nabla_Y Z).$$

and applying (1.2.17) one more time yields

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

Therefore

$$\begin{aligned} (\nabla_X\nabla_Y\omega - \nabla_Y\nabla_X\omega - \nabla_{[X,Y]}\omega)(Z) &= -\omega(\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z) + \omega(\nabla_{[X,Y]}Z) \\ &= -\omega(R_{X,Y}Z) = -(R^*_{X,Y}\omega)(Z). \end{aligned}$$

Next consider the action of $R_{X,Y}$ on an arbitrary tensor product: (see the calculation in Proposition 2.1.1)

$$\begin{aligned} \nabla^2_{X,Y}(F\otimes G) &= (\nabla_X\nabla_Y - \nabla_{\nabla_X Y})(F\otimes G) \\ &= (\nabla^2_{X,Y}F)\otimes G + (\nabla_Y F)\otimes(\nabla_X G) + (\nabla_X F)\otimes(\nabla_Y G) + F\otimes\nabla^2_{X,Y}G. \end{aligned}$$

Therefore

$$R_{X,Y}(F\otimes G) = (\nabla^2_{X,Y} - \nabla^2_{Y,X})(F\otimes G) = \big((\nabla^2_{X,Y} - \nabla_{Y,X})F\big)\otimes G + F\otimes\big((\nabla^2_{X,Y} - \nabla^2_{Y,X})G\big).$$

A simple induction using this relation together with (1.1) and (1.2) gives (1.3). □

For many purposes, the information contained in the curvature endomorphism is much more conveniently encoded in the form of a covariant 4-tensor. We define the (Riemann) curvature tensor to be the $(0,4)$-tensor field

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

In terms of any smooth local coordinates it is written

$$R = R_{ijkl}\,dx^i\otimes dx^j\otimes dx^k\otimes dx^l.$$

where $R_{ijkl} = R^s_{ijk}g_{sl}$. We justify next why the variables are treated on a more equal footing in this formula by showing several important symmetry properties.

ture symmetry **Proposition 2.1.5 (Symmetries of the Curvature Tensor).** *The curvature tensor $R(X, Y, Z, W)$ satisfies the following properties:*

(a) R is skew-symmetric in the first two and last two entries:

$$R(X, Y, Z, W) = -R(Y, X, Z, W), \quad R(X, Y, Z, W) = -R(X, Y, W, Z).$$

(b) R is symmetric between the first two and last two entries:

$$R(X, Y, Z, W) = R(Z, W, X, Y).$$

(c) R satisfies a cyclic permutation property called ***algebraic Bianchi identity****:*

$$R(X, Y)Z + R(Y, Z)X + R(Z, X)Y = 0.$$

Proof. The first part of (a) has already been established. For part two of (a) we compute

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

Now (a) follows by polarizing the identity $R(X, Y, Z, Z)$ in Z:

$$0 = R(X, Y, Z + W, Z + W) = R(X, Y, Z, W) + R(X, Y, W, Z).$$

Part (c) relies on the torsion free property and the definition of R and the symmetry of the connection:

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

Now we show that identity (b) follows from (a) and (c). Writing the algebraic Bianchi identity four times with indices cyclically permuted gives

$$\begin{aligned}
R(X, Y, Z, W) + R(Y, Z, X, W) + R(Z, X, Y, W) &= 0, \\
R(Y, Z, W, X) + R(Z, W, Y, X) + R(W, Y, Z, X) &= 0, \\
R(Z, W, X, Y) + R(W, X, Z, Y) + R(X, Z, W, Y) &= 0, \\
R(W, X, Y, Z) + R(X, Y, W, Z) + R(Y, W, X, Z) &= 0.
\end{aligned}$$

Now add up all four equations. Applying (a) four times makes all the terms in the first two columns cancel. Then applying (a) in the last column yields

$$2R(Z, X, Y, W) + 2R(W, Y, Z, X) = 0,$$

which is equivalent to (b). $\square$

Proposition 2.1.6 (Differential Bianchi Identity). *The total covariant derivative of the curvature tensor satisfies the following identity:*

$$(\nabla_Z R)(X,Y)W + (\nabla_X R)(Y,Z)W + (\nabla_Y R)(Z,X)W = 0.$$

Proof. We first give an equation about $(\nabla_Z)R(X,Y)W$. Using Proposition 1.2.12, we find that

$$\begin{aligned}(\nabla_Z R^\flat)(X,Y,W,V) &= X\langle R(X,Y)W,V\rangle \\ &\quad - R(\nabla_Z X,Y,W,V) - R(X,\nabla_Z Y,W,V) - R(X,Y,\nabla_Z W,V) - R(X,Y,W,\nabla_Z V).\end{aligned}$$

Therefore, by the metric compactibility,

$$\begin{aligned}\langle\nabla_Z(R(X,Y)W),V\rangle &= X\langle R(X,Y)W,V\rangle - \langle R(X,Y)W,\nabla_Z V\rangle \\ &= (\nabla_Z R^\flat)(X,Y,W,V) + R(\nabla_Z X,Y,W,V) + R(X,\nabla_Z Y,W,V) + R(X,Y,\nabla_Z W,V) \\ &= (\nabla_Z R)^\flat(X,Y,W,V) + R(\nabla_Z X,Y,W,V) + R(X,\nabla_Z Y,W,V) + R(X,Y,\nabla_Z W,V) \\ &= \langle(\nabla_Z R)(X,Y)W,V\rangle + R(\nabla_Z X,Y,W,V) + R(X,\nabla_Z Y,W,V) + R(X,Y,\nabla_Z W,V).\end{aligned}$$

where we use the fact $(\nabla R)^\flat = \nabla(R^\flat)$ (Proposition 1.3.18). Then since V is arbitrary, we conclude that

$$\begin{aligned}\nabla_Z(R(X,Y)W) &= (\nabla_Z R)(X,Y)W + R(\nabla_Z X,Y)W + R(X,\nabla_Z Y)W + R(X,Y)\nabla_Z W, \\ \nabla_X(R(Y,Z)W) &= (\nabla_X R)(Y,Z)W + R(\nabla_X Y,Z)W + R(Y,\nabla_X Z)W + R(Y,Z)\nabla_X W, \\ \nabla_Y(R(Z,X)W) &= (\nabla_Y R)(Z,X)W + R(\nabla_Y Z,X)W + R(Z,\nabla_Y X)W + R(Z,X)\nabla_Y W.\end{aligned} \tag{1.4}$$

Recall that $R(X,Y) = -R(Y,X)$, so adding the third and fourth columns of (1.4) yeilds

$$\begin{aligned}&R(\nabla_Z X,Y)W + R(X,\nabla_Z Y)W + R(\nabla_X Y,Z)W + R(Y,\nabla_X Z)W + R(\nabla_Y Z,X)W + R(Z,\nabla_Y X)W \\ &= R(\nabla_Y Z - \nabla_Z Y,X)W + R(\nabla_X Y - \nabla_Y X,Z)W + R(\nabla_Y Z - \nabla_Z Y,X) \\ &= R([Y,Z],X)W + R([X,Y],Z)W + R([Y,Z],X)W.\end{aligned} \tag{1.5}$$

Also, adding the left side of (1.4) gives

$$\begin{aligned}&\nabla_Z(R(X,Y)W) + \nabla_X(R(Y,Z)W) + \nabla_Y(R(Z,X)W) \\ &= (\nabla_Z\nabla_X\nabla_Y W - \nabla_Z\nabla_Y\nabla_X W - \nabla_Z\nabla_{[X,Y]}W) + (\nabla_X\nabla_Y\nabla_Z W - \nabla_X\nabla_Z\nabla_Y W - \nabla_X\nabla_{[Y,Z]}W) \\ &\quad + (\nabla_Y\nabla_Z\nabla_X W - \nabla_Y\nabla_X\nabla_Z W - \nabla_Y\nabla_{[Z,X]}W) \\ &= (\nabla_X\nabla_Y - \nabla_Y\nabla_X)\nabla_Z W - \nabla_Z\nabla_{[X,Y]}W + (\nabla_Y\nabla_Z - \nabla_Z\nabla_Y)\nabla_X W - \nabla_X\nabla_{[Y,Z]}W \\ &\quad + (\nabla_Z\nabla_X - \nabla_X\nabla_Z)\nabla_Y W - \nabla_Y\nabla_{[Z,X]}W \\ &= \nabla_{[X,Y]}\nabla_Z W - \nabla_Z\nabla_{[X,Y]}W + \nabla_{[Y,Z]}\nabla_X W - \nabla_X\nabla_{[Y,Z]}W + \nabla_{[Z,X]}\nabla_Y W - \nabla_Y\nabla_{[Z,X]}W \\ &\quad + R(X,Y)\nabla_Z W + R(Y,Z)\nabla_X W + R(Z,X)\nabla_Y W \\ &= R([X,Y],Z)W + R([Y,Z],X)W + R([Z,X],Y)W + \nabla_{[[X,Y],Z]}W + \nabla_{[[Y,Z],X]}W + \nabla_{[[Z,X],Y]}W \\ &\quad + R(X,Y)\nabla_Z W + R(Y,Z)\nabla_X W + R(Z,X)\nabla_Y W \\ &= R([X,Y],Z)W + R([Y,Z],X)W + R([Z,X],Y)W + R(X,Y)\nabla_Z W + R(Y,Z)\nabla_X W + R(Z,X)\nabla_Y W.\end{aligned} \tag{1.6}$$

Then adding (1.4) together and apply (1.5) and (1.6), we finally get

$$(\nabla_Z R)(X,Y)W + (\nabla_X R)(Y,Z)W + (\nabla_Y R)(Z,X)W = 0.$$

This completes the proof. □

Corollary 2.1.7. *The covariant derivative of the curvature tensor $R(X,Y,V,W)$ satisfies the following differential Bianchi identity:*

$$(\nabla R)(X,Y,V,W,Z) + (\nabla R)(X,Y,W,Z,V) + (\nabla R)(X,Y,Z,V,W) = 0.$$

In components, this is

$$R_{ijkl;s} + R_{ijls;k} + R_{ijsk;l} = 0. \tag{1.7}$$

Riemann diffe

Proof. Since $(\nabla R)^\flat = \nabla(R^\flat)$, by the differential Bianchi identity of $R(X,Y)Z$ we have

$$(\nabla_Z R)(V,W,X,Y) + (\nabla_V R)(W,Z,X,Y) + (\nabla_W R)(Z,V,X,Y) = 0.$$

Using the symmetric property of $R(X,Y,V,W)$, we get the claim. □

Remark 2.1.1. Since the $(0,4)$-tensor $R(X,Y,V,W)$ is used more frequently, without further remark, we will use R to denote this tensor. Then $(1,3)$ curvature tensor will be denote by $R(X,Y)Z$, typically.

2.1.2 Curvature and differential forms

Since the curvature operator involves the second derivative and the christoffel symbol, it is often hard to compute it. Now we use differentiable forms to give an computational accessible expression of the curvature of a Riemannian manifold.

Let M be a smooth n-manifold and ∇ a connection in TM, let (E_i) be a local frame on some open subset $U \subseteq M$, and let (ε^i) be the dual coframe. Let $X \in \mathfrak{X}(M)$, then the connection $\nabla_X E_i$ can written into a linear combination of E_j:

$$\nabla_X E_i = \omega_i^j(X) E_j.$$

The $n \times n$ matrix of smooth 1-forms (ω_i^j) on U is then uniquely determined by the connection ∇ and (E_i), and is called the **connection 1-forms** for this frame.

Similarly, for $X, Y \in \mathfrak{X}(M)$, the vector field $R(X,Y)E_i$ is a linear combination of (E_j):

$$R(X,Y)E_i = \Omega_i^j(X,Y)E_j.$$

The $n \times n$ matrix of smooth 2-forms (Ω_i^j) is called the **curvature** **2-forms** for this frame.

Proposition 2.1.8. *The connection form and the curvature form are given by*

$$\omega_i^j(X) = \Gamma_{ki}^j \varepsilon^k, \quad \Omega_i^j = \frac{1}{2} R_{kli}^j \, \varepsilon^k \wedge \varepsilon^l,$$

where (ε^j) is the coframe dual to (E_i).

Proof. Write $X = X^k E_k$, then

$$\nabla_X E_i = X_k \nabla_{E_k} E_i = X^k \Gamma_{ki}^j E_j.$$

Therefore $\omega_i^j(X) = \Gamma_{ki}^j X^k$, that is, $\omega_i^j = \Gamma_{ki}^j \varepsilon^k$. Similarly, we have

$$R(X,Y)E_i = X^k Y^l R(E_k, E_l)E_i = X^k Y^l R_{kli}^j E_j.$$

Since $R(X,Y)$ is antisymmetric, we find

$$(R_{kli}^j \varepsilon^k \wedge \varepsilon^l)(X,Y) = R_{kli}^j (X^k Y^l - Y^k X^l) = R_{kli}^j X^k Y^l - R_{kli}^j Y^k X^l = 2R_{kli}^j X^k Y^l.$$

Therefore we get the second equation. □

For $X, Y \in \mathfrak{X}(M)$, the torsion $\tau(X,Y)$ is a smooth 2-tensor, so we can write $\tau(X,Y) = \tau^i(X,Y)E_i$ for some 2-forms on M, called the **torsion forms** of the connection ∇.

Proposition 2.1.9. *Let ω and Ω be the connection and curvature forms, then we have the following structural equations:*

*(i) (**Cartan's first sturcture equation**)*

$$d\varepsilon^j = \varepsilon^i \wedge \omega_i^j + \tau^j. \tag{1.8}$$

Riemann sturc

*(ii) (**Cartan's second sturcture equation**)*

$$d\omega_i^j = \omega_i^k \wedge \omega_k^j + \Omega_i^j. \tag{1.9}$$

Riemann sturc

Proof. For (i), by the definition of ω_i^j, we have

$$\begin{aligned}\nabla_X Y &= \nabla_X(Y^i E_i) = \nabla_X(\varepsilon^i(Y)E_i)\\ &= X(\varepsilon^i(Y))E_i + \varepsilon^i(Y)\nabla_X E_i\\ &= X(\varepsilon^j(Y))E_j + \varepsilon^i(Y)\omega_i^j(X)E_j.\end{aligned}$$

Interchanging X and Y, we get $\nabla_Y X = Y(\varepsilon^j(Y))E_j + \varepsilon^i(X)\omega_i^j(Y)E_j$, and therefore

$$\begin{aligned}\tau(X,Y) &= \big(X(\varepsilon^j(Y)) - Y(\varepsilon^j(X)) - \varepsilon^j([X,Y])\big)E_i + \big(\varepsilon^i(Y)\omega_i^j(X) - \varepsilon^i(X)\omega_i^j(Y)\big)E_j\\ &= d\varepsilon^j(X,Y)E_j - (\varepsilon^i \wedge \omega_i^j)(X,Y)E_j.\end{aligned}$$

This gives $\tau^j = d\varepsilon^j - \varepsilon^i \wedge \omega_i^j$, which is (i).

For (ii), note that

$$\begin{aligned}\nabla_X \nabla_Y E_i &= \nabla_X(\omega_i^k(Y)E_k)\\ &= X(\omega_i^k(Y))E_k + \omega_i^k(Y)\nabla_X E_k\\ &= X(\omega_i^j(Y))E_j + \omega_i^k(Y)\omega_k^j(X)E_j\end{aligned}$$

Interchanging X and Y, we get $\nabla_Y \nabla_X E_i = Y(\omega_i^j(X))E_j + \omega_i^k(Y)\omega_k^j(Y)E_j$. Furthermore,

$$\nabla_{[X,Y]}E_i = \omega_i^j([X,Y])E_j.$$

Therefore,

$$\begin{aligned}R(X,Y)E_i &= \big(X(\omega_i^j(Y)) - Y(\omega_i^j(X)) - \omega_i^j([X,Y])\big)E_j + \big(\omega_i^k(Y)\omega_k^j(X) - \omega_i^k(X)\omega_k^j(Y)\big)E_j\\ &= d\omega_i^j(X,Y)E_j - (\omega_i^k \wedge \omega_k^j)(X,Y)E_j.\end{aligned}$$

Comparing this with the definition of the curvature form Ω_i^j gives (ii). □

Corollary 2.1.10. *Let (M,g) be a Riemannian manifold and ∇ be its Levi-Civita connection, then the structural equations of ∇ are*

$$d\varepsilon^j = \varepsilon^i \wedge \omega_i^j, \quad d\omega_i^j = \Omega_i^j + \omega_i^k \wedge \omega_k^j. \tag{1.10}$$

Riemann Levi-

Proof. Recall that the Levi-Civita connection is torsion-free. □

If $A = [\alpha_i^j]$ and $B = [\beta_i^j]$ are matrices of differential forms on M with the number of columns of A equal to the number of rows of B, then their wedge product $A \wedge B$ is defined to be the matrix of differential forms whose (i,j)-entry is

$$(A \wedge B)_i^j = \alpha_i^k \wedge \beta_k^j$$

and dA is defined to be $[d\alpha_i^j]$. In matrix notation, we write

$$\tau = \begin{bmatrix}\tau^1\\ \vdots\\ \tau^n\end{bmatrix} \quad \varepsilon = \begin{bmatrix}\varepsilon^1\\ \vdots\\ \varepsilon^n\end{bmatrix} \quad \omega = [\omega_i^j] \quad \Omega = [\Omega_i^j].$$

Then the structural equations can be written as

$$d\varepsilon = \varepsilon \wedge \omega + \tau, \quad d\omega = \omega \wedge \omega + \Omega.$$

Apart from the structural equations, we now explore more properties of the connection and curvature forms.

Proposition 2.1.11. *Let (M, g) be a Riemannian manifold and ∇ a connection on TM.*

(a) If the connection ∇ is compatible with the metric, then its connection matrix $[\omega_i^j]$ relative to any local orthonormal frame (E_i) over an open set $U \subseteq M$ is skew-symmetric.

(b) If every point $p \in M$ has a local orthonormal frame on a neighborhood $U \subseteq M$ such that the connection matrix $[\omega_i^j]$ relative to (E_i) is skew-symmetric, then the connection ∇ is compatible with the metric.

Proof. Part (a) is straightforward, just note that

$$0 = X\langle E_i, E_j\rangle = \langle \nabla_X E_i, E_j\rangle + \langle E_i, \nabla_X E_j\rangle = \omega_i^j + \omega_j^i.$$

For part (b), we note first that compatibility with the metric is a local condition, so ∇ is compatible with the metric if and only if its restriction ∇^U to any open set U is compatible with the metric. Suppose $\omega_i^j = -\omega_j^i$. Let $Y = X^i E_i$ and $Z = Z^j E_j$, with $Y^i, Z^j \in C^\infty(U)$. Then

$$\begin{aligned}
\nabla_X Y &= \nabla_X(Y^i E_i) = X(Y^i)E_i + Y^i \omega_i^j(X) E_j, \\
\nabla_X Z &= \nabla_X(Z^j E_j) = X(Z^j)E_j + Z^j \omega_j^i(X) E_i.
\end{aligned}$$

Therefore

$$\begin{aligned}
\langle \nabla_X Y, Z\rangle &= \sum_i X(Y^i)Z^i + Y^i Z^j \omega_i^j(X), \\
\langle Y, \nabla_X Z\rangle &= \sum_j X(Z^j)Y^j + Y^i Z^j \omega_j^i(X).
\end{aligned}$$

Since $\omega_i^j + \omega_j^i = 0$, we then find

$$\langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle = \sum_i X(Y^i)Z^i + \sum_j X(Z^j)Y^j = X\langle Y, Z\rangle.$$

Therefore ∇ is a metric connection. $\square$

Proposition 2.1.12. *If the connection matrix $[\omega_i^j]$ relative to a local frame (E_i) of an connection is skew-symmetric, then so is the curvature matrix $[\Omega_i^j]$.*

Proof. By the structural equation,

$$\Omega_i^j = d\omega_i^j - \omega_i^k \wedge \omega_k^j = -d\omega_j^i - \omega_k^i \wedge \omega_j^k = -(d\omega_j^i - \omega_j^k \wedge \omega_k^i) = -\Omega_j^i.$$

Therefore $[\Omega_i^j]$ is also skew-symmetric. $\square$

Example 2.1.13. As a special case, if $n = 2$ and (E_1, E_2) is a local orthonormal frame for M with dual frame $(\varepsilon^1, \varepsilon^2)$, then $[\omega_i^j]$ and $[\Omega_i^j]$ are both skew-symmetric, and in particular the curvature forms of M reduce to the 2-form Ω_1^2. Moreover, by the structural equation (1.9), we have

$$\Omega_1^2 = d\omega_1^2 - \omega_1^1 \wedge \omega_1^2 - \omega_1^2 \wedge \omega_2^2 = d\omega_1^2.$$

Also, by the structural equation (1.9), ω_1^2 can be computed by

$$d\varepsilon^1 = \varepsilon^2 \wedge \omega_2^1 = -\varepsilon^2 \wedge \omega_1^2, \quad d\varepsilon^2 = \varepsilon^1 \wedge \omega_1^2.$$

Since $\varepsilon^1 \wedge \varepsilon^2$ is a basis for $\Omega^2(M)$, from these two equalities we can solve ω_1^2, and then get Ω_1^2.

Example 2.1.14. The Poincaré disk is the open unit disk $\mathbb{D}$ in the complex plane with Riemannian metric

$$g = 4\frac{dx^2 + dy^2}{(1-|z|^2)^2} = 4\frac{dx^2 + dy^2}{(1-x^2-y^2)^2}.$$

An orthonormal frame for $\mathbb{D}$ is

$$E_1 = \frac{1}{2}(1-|z|^2)\frac{\partial}{\partial x}, \quad E_2 = \frac{1}{2}(1-|z|^2)\frac{\partial}{\partial y},$$

with dual frames

$$\varepsilon^1 = \frac{2}{1-|z|^2}dx, \quad \varepsilon^2 = \frac{2}{1-|z|^2}dy.$$

Now we find the connection form ω_i^j for this frame. First we have

$$d\varepsilon^1 = -\frac{4y}{(1-|z|^2)^2}dx \wedge dy, \quad d\varepsilon^2 = \frac{4x}{(1-|z|^2)^2}dx \wedge dy.$$

Therefore by (1.8) and the relation $\omega_1^2 = -\omega_2^1$ we then get

$$\omega_1^2 = \frac{2x}{1-|z|^2}dy - \frac{2y}{1-|z|^2}dx.$$

With this, we then use (1.9) to compute Ω_1^2:

$$\Omega_1^2 = d\omega_1^2 = \frac{4}{(1-|z|^2)^2}dx \wedge dy.$$

Example 2.1.15. Let $\mathbb{H}$ be the Poincaré half-plane wohse metric is given by

$$g = \frac{dx^2 + dy^2}{y^2}.$$

With this metric, an orthonormal frame is

$$E_1 = y\frac{\partial}{\partial y}, \quad E_2 = y\frac{\partial}{\partial y}$$

with dual frames

$$\varepsilon^1 = \frac{1}{y}dx, \quad \varepsilon^2 = \frac{1}{y}dy.$$

Then we have

$$d\varepsilon^1 = \frac{1}{y^2}dx \wedge dy, \quad d\varepsilon^2 = 0,$$

and therefore $\omega_1^2 = \frac{1}{y}dx$. With this, the curvature form is given by

$$\Omega_1^2 = d\omega_1^2 = \frac{1}{y^2}dx \wedge dy.$$

Example 2.1.16. Even more general, let M be a two-dimensional Riemannian manifold that is locally conformally flat. Then there are local charts (x, y) at each point of M such that

$$g = f(x)(dx^2 + dy^2)$$

under the local frames $(\partial/\partial x, \partial/\partial y)$, where f is a smooth function. Since f is then positive, the vector frames

$$E_1 = \frac{1}{\sqrt{f}}\frac{\partial}{\partial x}, \quad E_2 = \frac{1}{\sqrt{f}}\frac{\partial}{\partial y}$$

are then orthonormal. The corresponding dual frames are $\varepsilon^1 = \sqrt{f}dx, \varepsilon^2 = \sqrt{f}dy$, so we have

$$d\varepsilon^1 = -\frac{f_y}{2\sqrt{f}}dx \wedge dy, \quad d\varepsilon^2 = \frac{f_x}{2\sqrt{f}}dx \wedge dy.$$

This then implies

$$\omega_1^2 = -\frac{f_y}{2f}dx + \frac{f_x}{2f}dy, \quad \Omega_1^2 = \frac{f(f_{xx}+f_{yy}) - f_x^2 - f_y^2}{2f^2}dx \wedge dy.$$

Example 2.1.17. If we parametrize S^2 by

$$(x, y, z) = (\cos\theta \sin\phi, \sin\theta\sin\phi, \cos\phi)$$

Then the metric is given by

$$g = \sin^2\phi\, d\theta^2 + d\phi^2.$$

Therefore an orthonormal frame is given by $(1/\sin\phi)\partial_\theta, \partial_\phi$ with dual frames $(\sin\phi\, d\theta, d\phi)$. We compute that

$$d(\sin\phi\, d\theta) = -\cos\phi\, d\theta \wedge d\phi, \quad d(d\phi) = 0.$$

Therefore the connection form and the curvature form are given by

$$\omega_1^2 = -\cos\phi d\theta, \quad \Omega_1^2 = -\sin\phi\, d\theta \wedge d\phi.$$

2.1.3 Flat manifolds

Recall that the curvature operator endodes the failure of the exchange of covariant derivatives. Importantly for our purposes, we show that it also measures the failure of second covariant derivatives along families of curves to commute. Given a smooth one-parameter family of curves $\Gamma : J \times I \to M$, recall that the velocity fields $\partial_t\Gamma(s,t) = (\Gamma_s)'(t)$ and $\partial_s\Gamma(s,t) = (\Gamma^t)'(s)$ are smooth vector fields along Γ.

ivatire curve **Proposition 2.1.18.** *Suppose (M, g) is a smooth Riemannian or pseudo-Riemannian manifold and $\Gamma : J \times I \to M$ is a smooth one-parameter family of curves in M. Then for every smooth vector field V along γ,*

$$D_sD_tV - D_tD_sV = R(\partial_s\Gamma, \partial_t\Gamma)V. \tag{1.11}$$

Riemann curva

Proof. This is a local question, so for each $(s,t) \in J \times I$, we can choose smooth coordinates (x^i) defined on a neighborhood of $\Gamma(s,t)$ and write

$$\Gamma(s,t) = (\gamma^1(s,t), \ldots, \gamma^n(s,t)) \quad V(s,t) = V^i(s,t)\partial_i|_{\Gamma(s,t)}.$$

connection derivative curve-1

Formula (2.12) yields

$$D_tV = \frac{\partial V^i}{\partial t}\partial_i + V^iD_t\partial_i.$$

connection derivative curve-1

Therefore, applying (2.12) again, we get

$$D_sD_tV = \frac{\partial^2 V^i}{\partial s\partial t}\partial_i + \frac{\partial V^i}{\partial t}D_s\partial_i + \frac{\partial V^i}{\partial s}D_t\partial_i + V^iD_sD_t\partial_i.$$

Interchanging s and t and subtracting, we see that all the terms except the last cancel:

$$D_sD_tV - D_tD_sV = V^i(D_sD_t\partial_i - D_tD_s\partial_i) \tag{1.12}$$

Riemann curva

Now we need to compute the commutator in parentheses. Because ∂_i is extendible,

$$D_t\partial_i = \nabla_{\partial_i\Gamma}\partial_i = \frac{\partial\gamma^j}{\partial t}\nabla_{\partial_j}\partial_i.$$

and therefore, because $\nabla_{\partial_j}\partial_i$ is also extendible,

$$\begin{aligned} D_sD_t\partial_i = D_s\Big(\frac{\partial\gamma^j}{\partial t}\nabla_{\partial_j}\partial_i\Big) &= \frac{\partial^2\gamma^j}{\partial s\partial t}\nabla_{\partial_j}\partial_i + \frac{\partial\gamma^j}{\partial t}\nabla_{\partial_s\Gamma}(\nabla_{\partial_j}\partial_i) \\ &= \frac{\partial^2\gamma^j}{\partial s\partial t}\nabla_{\partial_j}\partial_i + \frac{\partial\gamma^j}{\partial t}\frac{\partial\gamma^k}{\partial s}\nabla_{\partial_k}\nabla_{\partial_j}\partial_i. \end{aligned}$$

Interchanging $s \leftrightarrow t$ and $j \leftrightarrow k$ and subtracting, we find that the first terms cancel, and we get

$$\begin{aligned} D_sD_t\partial_i - D_tD_s\partial_i &= \frac{\partial\gamma^j}{\partial t}\frac{\partial\gamma^k}{\partial s}(\nabla_{\partial_k}\nabla_{\partial_j}\partial_i - \nabla_{\partial_j}\nabla_{\partial_k}\partial_i) \\ &= \frac{\partial\gamma^j}{\partial t}\frac{\partial\gamma^k}{\partial s}R(\partial_k, \partial_j)\partial_i \\ &= R(\partial_s\Gamma, \partial_t\Gamma)\partial_i. \end{aligned}$$

Riemann curvature second derivatire curve-2

Finally, inserting this into (1.12) yields the result. □

Recall that a Riemannian manifold is said to be **flat** if it is locally isometric to a Euclidean space, that is, if every point has a neighborhood that is isometric to an open set in Rn with its Euclidean metric. Similarly, a pseudo-Riemannian manifold is flat if it is locally isometric to a pseudo-Euclidean space. We say that a connection ∇ on a smooth manifold M satisfies the flatness criterion if whenever X, Y, Z are smooth vector fields defined on an open subset of M, the following identity holds:

$$\nabla_X\nabla_Y - \nabla_Y\nabla_X = \nabla_{[X,Y]}. \tag{1.13}$$

Riemann flat

Proposition 2.1.19. *If (M, g) is a flat Riemannian or pseudo-Riemannian manifold, then its Levi-Civita connection satisfies the flatness criterion.*

Riemann flat crition

Proof. It is clear that the Euclidean connection on $\mathbb{R}^n$ satisfies (1.13). By naturality, the Levi-Civita connection on every manifold that is locally isometric to a Euclidean or pseudo-Euclidean space must also satisfy the same identity. □

To give a qualitative geometric interpretation to the curvature tensor, we will show that it is precisely the obstruction to being locally isometric to Euclidean (or pseudo-Euclidean) space. The crux of the proof is the following lemma.

extension lem

Lemma 2.1.20. *Suppose M is a smooth manifold, and ∇ is any connection on M satisfying the flatness criterion. Given $p \in M$ and any vector $v \in T_pM$, there exists a parallel vector field V on a neighborhood of p such that $V_p = v$.*

Proof. Let $p \in M$ and $v \in T_pM$ be arbitrary, and let (x^i) be any smooth coordinates for M centered at p. By shrinking the coordinate neighborhood if necessary, we may assume that the image of the coordinate map is an open cube $C_\varepsilon = \{x : |x^i| \leq \varepsilon, 1 \leq i \leq n\}$. We use the coordinate map to identify the coordinate domain with C_ε.

Begin by parallel transporting v along the x^1-axis; then from each point on the x^1-axis, parallel transport along the coordinate line parallel to the x^2-axis; then successively parallel transport along coordinate lines parallel to the x^3 through x^n-axes. The result is a vector field V defined in C_ε. The fact that V is smooth follows from an inductive application of Theorem **??** to vector fields of the form $W_k|_{(x,v)} = \partial/\partial x^k - v^i \Gamma^j_{ki}(x)\partial/\partial v^j$ on $C_\varepsilon \times \mathbb{R}^n$.

Since $\nabla_X V$ is linear over $C^\infty(M)$ in X, to show that V is parallel, it suffices to show that $\nabla_{\partial_i} V$ for each $1 \le i \le n$. By construction, $\nabla_{\partial_1} V = 0$ on the x^1-axis, $\nabla_{\partial_2} V = 0$ on the (x^1, x^2)-plane, and in general $\nabla_{\partial_k} V = 0$ on the slice $M_k \subseteq C_\varepsilon$ defined by $x^{k+1} = \cdots = x^n = 0$. We will prove the following fact by induction on k:

$$\nabla_{\partial_1} V = \cdots = \nabla_{\partial_k} V = 0 \quad \text{on} \quad M_k. \tag{1.14}$$

For $k = 1$, this is true by construction, and for $k = n$, it means that V is parallel on the whole cube C_ε. So assume that (1.14) holds for some k. By construction, $\nabla_{k+1} V = 0$ on all of M_{k+1}, and for $i \le k$, the inductive hypothesis shows that $\nabla_{\partial_i} V = 0$ on the hyperplane $M_k \subseteq M_{k+1}$. Since $[\partial_{k+1}, \partial_i] = 0$, the flatness criterion gives

$$\nabla_{\partial_{k+1}}(\partial_{\partial_i} V) = \partial_{\partial_i}(\partial_{k+1} V) = 0 \quad \text{on} \quad M_k.$$

This shows that $\nabla_{\partial_i} V$ is parallel along the x^{k+1}-curves starting on M_k. Since $\nabla_{\partial_i} V$ vanishes on M_k and the zero vector field is the unique parallel transport of zero, we conclude that $\nabla_{\partial_i} V$ is zero on each x^{k+1}-curve. Since every point of M_{k+1} is on one of these curves, it follows that $\nabla_{\partial_i} V = 0$ on all of M_{k+1}. This completes the inductive step to show that V is parallel. □

Theorem 2.1.21. *A Riemannian or pseudo-Riemannian manifold is flat if and only if its curvature tensor vanishes identically.*

Proof. One direction is immediate: Proposition 1.3.12 showed that the Levi-Civita connection of a flat metric satisfies the flatness criterion, so its curvature endomorphism is identically zero, which implies that the curvature tensor is also zero.

Now suppose (M, g) has vanishing curvature tensor. This means that the curvature endomorphism vanishes as well, so the Levi-Civita connection satisfies the flatness criterion. We begin by showing that g shares one important property with Euclidean and pseudo-Euclidean metrics: it admits a parallel orthonormal frame in a neighborhood of each point.

Let $p \in M$, and choose an orthonormal basis $(e_1, \ldots, e_n)$ for T_pM. In the pseudo case, we may assume that the basis is in standard order (with positive entries before negative ones in the matrix $g_{ij} = g_p(e_i, e_j)$). Lemma 2.1.20 shows that there exist parallel vector fields $(E_1, \ldots, E_n)$ on a neighborhood U of p such that $E_i|_p = e_i$ for each $i = 1, \ldots, n$. Because parallel transport preserves inner products, the vector fields (E_i) are orthonormal (and hence linearly independent) in all of U. Because the Levi-Civita connection is symmetric, we have

$$[E_i, E_j] = \nabla_{E_i} E_j - \nabla_{E_j} E_i = 0.$$

Thus the vector fields $(E_1, \ldots, E_n)$ form a commuting orthonormal frame on U. The canonical form theorem for commuting vector fields shows that there are coordinates $(y^1, \ldots, y^n)$ on a (possibly smaller) neighborhood of p such that $E_i = \partial/\partial y^i$ for $i = 1, \ldots, n$. In any such coordinates, $g_{ij} = g(\partial_i, \partial_j) = \pm\delta_{ij}$, so the map $y = (y^1, \ldots, y^n)$ is an isometry from a neighborhood of p to an open subset of the appropriate Euclidean or pseudo-Euclidean space. □

Using similar ideas, we can give a more precise interpretation of the meaning of the curvature tensor: it is a measure of the extent to which parallel transport around a small rectangle fails to be the identity map.

Theorem 2.1.22. *Let (M, g) be a Riemannian or pseudo-Riemannian manifold; let I be an open interval containing 0; let $\Gamma : I \times I \to M$ be a smooth one-parameter family of curves; and let $p = \Gamma(0,0)$, $x = \partial_s\Gamma(0,0)$, and $y = \partial_t\Gamma(0,0)$. For any $s_1, s_2, t_1, t_2 \in I$, let*

$$P^{s_1,t_2}_{s_1,t_1} : T_{\Gamma(s_1,t_1)}M \to T_{\Gamma(s_2,t_2)}M$$

denote parallel transport along the curve $t \mapsto \Gamma(s_1, t)$ from time t_1 to time t_2, and let

$$P^{s_2,t_1}_{s_1,t_1} : T_{\Gamma(s_1,t_1)}M \to T_{\Gamma(s_2,t_2)}M$$

denote parallel transport along the curve $s \mapsto \Gamma(s, t_1)$ from time s_1 to time s_2. Then for every $z \in T_pM$,

$$R(x,y)z = \lim_{\delta,\varepsilon\to 0} \frac{P^{0,0}_{\delta,0} \circ P^{\delta,0}_{\delta,\varepsilon} \circ P^{\delta,\varepsilon}_{0,\varepsilon} \circ P^{0,\varepsilon}_{0,0}(z) - z}{\delta\varepsilon}. \tag{1.15}$$

Proof. Define a vector field Z along Γ by first parallel transporting z along the curve $t \mapsto \Gamma(0,t)$ and then for each t, parallel transporting $Z(0,t)$ along the curve $s \mapsto \Gamma(s,t)$. The resulting vector field along Γ is smooth by application of Theorem **??** as in the proof of Lemma 2.1.20; and by construction, it satisfies $D_tZ(0,t) = 0$ for all $t \in I$, and $D_sZ(s,t) = 0$ for all $(s,t) \in I \times I$. Proposition 2.1.18 shows that

$$R(x,y)z = D_sD_tZ(0,0) - D_tD_sZ(0,0) = D_sD_t\Gamma(0,0).$$

Thus we need only show that $D_sD_tZ(0,0)$ is equal to the limit on the right-hand side of (1.15).

From Theorem 1.2.27, we have

$$D_tZ(s,0) = \lim_{\varepsilon\to 0} \frac{P^{s,0}_{s,\varepsilon}(Z(s,\varepsilon)) - Z(s,0)}{\varepsilon}, \tag{1.16}$$

$$D_sD_tZ(0,0) = \lim_{\delta\to 0} \frac{P^{0,0}_{\delta,0}(D_tZ(\delta,0)) - D_tZ(0,0)}{\delta}. \tag{1.17}$$

Evaluating (1.16) first at $s = \delta$ and then at $s = 0$, and inserting the resulting expressions into (1.17), we obtain

$$D_sD_tZ(0,0) = \lim_{\varepsilon,\delta\to 0} \frac{P^{0,0}_{\delta,0} \circ P^{\delta,0}_{\delta,\varepsilon}(Z(\delta,\varepsilon)) - P^{0,0}_{\delta,0}(Z(\delta,0)) - P^{0,0}_{0,\varepsilon}(Z(0,\varepsilon)) + Z(0,0)}{\delta\varepsilon}. \tag{1.18}$$

Here we have used the fact that parallel transport is linear, so the ε-limit can be pulled past $P^{0,0}_{\delta,0}$.

Now, the fact that Z is parallel along $t \mapsto \Gamma(0,t)$ and along all of the curves $s \mapsto \Gamma(s,t)$ implies

$$P^{0,0}_{\delta,0}(Z(\delta,0)) = P^{0,0}_{0,\varepsilon}(Z(0,\varepsilon)) = Z(0,0) = z,$$

$$Z(\delta,\varepsilon) = P^{\delta,\varepsilon}_{0,\varepsilon}(Z(0,\varepsilon)) = P^{0,\varepsilon}_{0,0} \circ P^{0,\varepsilon}_{0,0}(z).$$

Inserting these relations into (1.18) yields (1.15). □

2.1.4 Ricci and scalar curvatures

Suppose (M, g) is an n-dimensional Riemannian or pseudo-Riemannian manifold. Because 4-tensors are so complicated, it is often useful to construct simpler tensors that summarize some of the information contained in the curvature tensor. The most important such tensor is the **Ricci curvature** or **Ricci tensor**, denoted by Ric, which is the covariant 2-tensor field defined as the trace of the curvature endomorphism on its first and last indices. Thus for vector fields X, Y

$$\mathrm{Ric}(X, Y) = \mathrm{tr}(Z \mapsto R(Z, X)Y)$$

The components of Ric are usually denoted by R_{ij}, so that

$$R_{ij} = R^k_{kij} = g^{kl} R_{kijl}.$$

The **scalar curvature** is the function S defined as the trace of the Ricci tensor:

$$S = \mathrm{tr}_g \mathrm{Ric} = R^i_i = g^{ij} R_{ij}.$$

First we establish some properties of the Ricci tensor.

Lemma 2.1.23. *The Ricci curvature is a symmetric 2-tensor field. It can be expressed in any of the following ways:*

$$R_{ij} = R_{kij}{}^k = R_{ik}{}^k{}_j = -R_{ki}{}^k{}_j = -R_{ikj}{}^k.$$

Proof. Let $E_1, \ldots, E_n$ be a local orthonormal frame of M, then by the symmetry properties of R, we have

$$\begin{aligned} R(X, Y) &= \mathrm{tr}(R(-, X), Y) = \langle R(E_i, X)Y, E_i \rangle \\ &= R(E_i, X, Y, E_i) = R(Y, E_i, E_i, X) \\ &= R(E_i, Y, X, E_i) = R(Y, X). \end{aligned}$$

The other equations follows from the symmetric properties of R_{ijkl}. □

It is sometimes useful to decompose the Ricci tensor into a multiple of the metric and a complementary piece with zero trace. Define the traceless Ricci tensor of g as the following symmetric 2-tensor:

$$\mathring{\mathrm{Ric}} = \mathrm{Ric} - \frac{1}{n} S g.$$

Proposition 2.1.24. *Let (M, g) be a Riemannian or pseudo-Riemannian n-manifold. Then* $\mathrm{tr}\mathring{\mathrm{Ric}} = 0$*, and the Ricci tensor decomposes orthogonally as*

$$\mathrm{Ric} = \mathring{\mathrm{Ric}} + \frac{1}{n} S g. \tag{1.19}$$

Therefore, in the Riemannian case,

$$|\mathrm{Ric}|_g^2 = |\mathring{\mathrm{Ric}}|_g^2 + \frac{1}{n} S^2. \tag{1.20}$$

Proof. Note that in every local frame, we have

$$\mathrm{tr}_g g = g_{ij} g^{ji} = \delta^i_i = n.$$

It then follows directly from the definition of $\mathring{\mathrm{Ric}}$ that $\mathrm{tr}_g \mathring{\mathrm{Ric}} = 0$ and (1.19) holds. The fact that the decomposition is orthogonal follows easily from the fact that for every symmetric 2-tensor h, we have

$$\langle h, g \rangle = g^{ik} g^{jl} h_{ij} g_{kl} = g^{ij} h_{ij} = \mathrm{tr}_g h.$$

and therefore $\langle \mathring{\mathrm{Ric}}, g \rangle = 0$. Finally, (1.20) follows from (1.19) and the fact that $\langle g, g \rangle = \mathrm{tr}_g g = n$. □

The next proposition, which follows directly from the differential Bianchi identity, expresses some important relationships among the covariant derivatives of the various curvature tensors. To express it concisely, it is useful to introduce another operator on tensor fields.

Recall that for a smooth 1-form ω, the exterior derivative $d\omega$ can be expressed using ∇ by

$$(d\omega)(X, Y) = -(\nabla\omega)(X, Y) + (\nabla\omega)(Y, X).$$

Mimiking this equation, we define the **exterior covariant derivative** of a 2-tensor T to be the 3-tensor field DT defined by

$$(DT)(X, Y, Z) = -(\nabla T)(X, Y, Z) + (\nabla T)(X, Z, Y). \tag{1.21}$$

In terms of components, this is

$$(DT)_{ijk} = -T_{ij;k} + T_{ik;j}.$$

Proposition 2.1.25 (Contracted Bianchi Identities). *Let (M, g) be a Riemannian or pseudo-Riemannian manifold. The covariant derivatives of the Riemann, Ricci, and scalar curvatures of g satisfy the following identities:*

$$\mathrm{div} R = \mathrm{tr}_g(\nabla R) = -D(\mathrm{Ric}), \tag{1.22}$$

$$\mathrm{div}\mathrm{Ric} = \mathrm{tr}_g(\nabla \mathrm{Ric}) = \frac{1}{2} dS. \tag{1.23}$$

where the trace in each case is on the first and last indices. In components, this is

$$R_{ijkl;}{}^{i} = R_{jk;l} - R_{jl;k}, \tag{1.24}$$

$$R_{il;}{}^{i} = \frac{1}{2} S_{;l}. \tag{1.25}$$

Proof. Recall the component form (1.7) of the differential Bianchi identity:

$$R_{ijkl;s} + R_{ijls;k} + R_{ijsk;l} = 0.$$

Raise the index s and contract on the indices i, s, we then obtain

$$R_{ijkl;}{}^{i} + R_{ijl}{}^{i}{}_{;k} + R_{ij}{}^{i}{}_{k;l} = R_{ijkl;}{}^{i} + R_{jl;k} - R_{jk;l} = 0. \tag{1.26}$$

which is (1.24). (Note that covariant differentiation commutes with contraction and with the musical isomorphisms, so it does not matter whether the indices that are raised and contracted come before or after the semicolon.)

Raise the index k and contract on the indices j, k in (1.26), we get

$$R_{ik}{}^{k}{}_{l;}{}^{i} + R_{kl;}{}^{k} - R_{k}{}^{k}{}_{;l} = 2R_{il;}{}^{i} - S_{;l} = 0.$$

This gives (1.25). The coordinate-free formulas (1.22) and (1.23) follow by expanding everything out in components. □

A Riemannian or pseudo-Riemannian metric is said to be an **Einstein metric** if its Ricci tensor is a constant multiple of the metric—that is,

$$\mathrm{Ric} = \lambda g \tag{1.27}$$

for some constant $\lambda \in \mathbb{R}$. This equation is known as the Einstein equation. As the next proposition shows, for connected manifolds of dimension greater than 2, it is not necessary to assume that λ is constant; just assuming that the Ricci tensor is a function times the metric is sufficient.

Proposition 2.1.26 (Schur's Lemma). *Suppose (M, g) is a connected (pseudo) Riemannian manifold of dimension $n \geq 3$ whose Ricci tensor satisfies $\mathrm{Ric} = fg$ for some smooth real-valued function f. Then f is constant and g is an Einstein metric.*

Proof. Taking traces of both sides of $\mathrm{Ric} = fg$ shows that $f = S/n$, so the traceless Ricci tensor is identically zero. It follows that $\nabla \mathring{\mathrm{Ric}} \equiv 0$. Because the covariant derivative of the metric is zero (Proposition 1.3.4), this implies the following equation in any coordinate chart:

$$0 = R_{ij;k} - \frac{1}{n} S_{;k} g_{ij}.$$

Tracing this equation on i and k, and comparing with the contracted Bianchi identity (1.25), we conclude that

$$0 = \frac{1}{2} S_{;l} - \frac{1}{n} S_{;j}.$$

Because $n \geq 3$, this implies $S_{;j} = 0$. But $S_{;j}$ is the component of $\nabla S = dS$, so connectedness of M implies that S is constant and thus so is f. $\square$

2.1.5 The Weyl tensor

As noted above, the Ricci and scalar curvatures contain only part of the information encoded into the curvature tensor. Now we introduce a tensor field called the **Weyl tensor**, which encodes all the rest.

We begin by considering some linear-algebraic aspects of tensors that have the symmetries of the curvature tensor. Suppose V is an n-dimensional real vector space. Let $\mathcal{R}(V^*) \subseteq T^4(V^*)$ denote the vector space of all covariant 4-tensors T on V that have the symmetries of the $(0, 4)$-Riemann curvature tensor:

(a) $T(x, y, v, w) = -T(y, x, v, w)$.

(b) $T(x, y, v, w) = -T(x, y, w, v)$.

(c) $T(x, y, v, w) = T(v, w, x, y)$.

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

An element of $\mathcal{R}(V^*)$ is called an **algebraic curvature tensor** on V.

Proposition 2.1.27. *If the vector space V has dimension n, then*

$$\dim \mathcal{R}(V^*) = \frac{n^2(n^2-1)}{12}. \tag{1.28}$$

Proof. Let $\mathcal{B}(V^*)$ denote the linear subspace of $T^4(V^*)$ consisting of tensors satisfying properties (a)–(c), and let $\Sigma^2(\bigwedge^2(V)^*)$ denote the space of symmetric bilinear forms on the vector space $\bigwedge^2(V)$ of alternating contravariant 2-tensors on V. Define a map $\Phi : \Sigma^2(\bigwedge^2(V)^*) \to \mathcal{B}(V^*)$ as follows:

$$\Phi(B)(x, y, v, w) = B(x \wedge y, v \wedge w).$$

It is easy to check that $\Phi(B)$ satisfies (a)–(c), so $\Phi(B) \in \mathcal{B}(V^*)$, and that Φ is a linear map. In fact, it is an isomorphism, which we prove by constructing an inverse for it. Choose a basis $(b_1, \ldots, b_n)$ for V, so the collection $\{b_i \wedge b_j : i < j\}$ is a basis for $\bigwedge^2(V)$. Define a map $\Psi : \Sigma^2(\bigwedge^2(V)^*) \to \mathcal{B}(V^*)$ by setting

$$\Psi(T)(b_i \wedge b_j, b_k \wedge b_l) = T(b_i, b_j, b_k, b_l).$$

when $i < j$ and $k < l$, and extending by bilinearity. A straightforward computation shows that Ψ is an inverse for Φ.

The upshot of the preceding construction is that

$$\dim \mathcal{B}(V^*) = \dim \Sigma^2(\bigwedge^2(V)^*) = \frac{\binom{n}{2}(\binom{n}{2}+1)}{2} = \frac{n(n-1)(n^2-n+2)}{8}.$$

where we have used the facts that $\dim \bigwedge^2(V) = \binom{n}{2}$ and the dimension of the space of symmetric bilinear forms on a vector space of dimension k is $k(k+1)/2$.

Now consider the linear map $\pi : \mathcal{B}(V^*) \to T^4(V^*)$ defined by

$$\pi(T)(x, y, v, w) = \frac{1}{3}(T(x, y, v, w) + T(y, v, x, w) + T(v, x, y, w)).$$

By definition, $\mathcal{R}(V^*)$ is equal to the kernel of π. In fact, π is equal to the restriction to $\mathcal{B}(V^*)$ of the operator $\mathrm{Alt} : T^4(V^*) \to \bigwedge^4(V^*)$: thanks to the symmetries (a)-(c), the 24 terms in the definition of $\mathrm{Alt}(T)$ can be arranged in three groups of eight in such a way that all the terms in each group reduce to one of the terms in the definition of π. Thus the image of π is contained in $\bigwedge^4(V^*)$. In fact, the image is all of $\bigwedge^4(V^*)$: every $T \in \bigwedge^4(V^*)$ satisfies (a)–(c) and thus lies in $\mathcal{B}(V^*)$, and $\pi(T) = \mathrm{Alt}T = T$ for each such tensor.

Therefore, using the rank-nullity theorem of linear algebra, we conclude that

$$\dim \mathcal{R}(V^*) = \dim \mathcal{B}(V^*) - \dim \bigwedge^4(V^*) = \frac{n(n-1)(n^2-n+2)}{8} - \binom{n}{4} = \frac{n^2(n^2-1)}{12}.$$

This finishes the proof. □

Let us now assume that our vector space V is endowed with a (not necessarily positive definite) scalar product $g \in \Sigma^2(V^*)$. Let $\mathrm{tr}_g : \mathcal{R}(V^*) \to \Sigma^2(V^*)$ denote the trace operation (with respect to g) on the first and last indices (so that, for example, $\mathrm{Ric} = \mathrm{tr}_g R$). It is natural to wonder whether this operator is surjective and what its kernel is, as a way of asking how much of the information contained in the Riemann curvature tensor is captured by the Ricci tensor. One way to try to answer the question is to attempt to construct a right inverse for the trace operator—a linear map $G : \Sigma^2(V^*) \to \mathcal{R}(V^*)$ such that $\mathrm{tr}_g(G(S)) = S$ for all $S \in \Sigma^2(V^*)$.

Such an operator must start with a symmetric 2-tensor and construct a 4-tensor, using only the given 2-tensor and the metric. It turns out that there is a natural way to construct an algebraic curvature tensor out of two symmetric 2-tensors, which we now describe. Given $h, k \in \Sigma^2(V^*)$, we define a covariant 4-tensor $h \owedge k$, called the **Kulkarni-Nomizu product** of h and k, by the following formula:

$$h \owedge k(x, y, v, w) = \begin{vmatrix} h(x,w) & h(y,w) \\ k(x,v) & k(y,v) \end{vmatrix} + \begin{vmatrix} k(x,w) & k(y,w) \\ h(x,v) & h(y,v) \end{vmatrix}.$$

In terms of any basis, the components of are

$$(h \owedge k)_{ijkl} = h_{il}k_{jk} - h_{jl}k_{ik} + k_{il}h_{jk} - k_{jl}h_{ik}.$$

i-Nomizu prop

Lemma 2.1.28 (Properties of the Kulkarni-Nomizu Product). *Let V be an n-dimensional vector space endowed with a scalar product g, let h and k be symmetric 2-tensors on V, let T be an algebraic curvature tensor on V, and let tr_g denote the trace on the first and last indices.*

(a) $h \owedge k$ *is an algebraic curvature tensor.*

(b) $h \owedge k = k \owedge h$.

(c) $\operatorname{tr}_g(h \owedge g) = (n-2)h + (\operatorname{tr}_g h)g$.

(d) $\operatorname{tr}_g(g \owedge g) = 2(n-1)g$.

(e) $\langle T, h \owedge g\rangle_g = 4\langle \operatorname{tr}_g T, h\rangle$.

(f) In case g is positive definite, $|h \owedge g|_g^2 = 4(n-2)|h|_g^2 + 4(\operatorname{tr}_g h)^2$.

Proof. It is evident from the definition that $h \owedge k$ has three of the four symmetries of an algebraic curvature tensor: it is antisymmetric in its first two arguments and also in its last two, and its value is unchanged when the first two and last two arguments are interchanged. Thus to prove (a), only the algebraic Bianchi identity needs to be checked. This is a straightforward computation:

$$\begin{aligned}
&(h \owedge k)_{ijkl} + (h \owedge k)_{jkil} + (h \owedge k)_{kijl}\\
&= (h_{il}k_{jk} - h_{jl}k_{ik} + k_{il}h_{jk} - k_{jl}h_{ik}) + (h_{jl}k_{ki} - h_{kl}k_{ji} + k_{jl}h_{ki} - k_{kl}h_{ji})\\
&+ (h_{kl}k_{ij} - h_{il}k_{kj} + k_{kl}h_{ij} - k_{il}h_{kj})\\
&= h_{il}(k_{jk} - k_{kj}) - h_{jl}(k_{ik} - k_{ki}) + k_{il}(h_{jk} - h_{kj}) - k_{jl}(h_{ik} - h_{ki})\\
&- h_{kl}(k_{ji} - k_{ij}) - k_{kl}(h_{ji} - h_{ij})\\
&= 0.
\end{aligned}$$

Part (b) is immediate from the definition. To prove (c), choose a basis and use the definition to compute

$$\begin{aligned}
(\operatorname{tr}_g(h \owedge g))_{jk} &= g^{il}(h \owedge g)_{ijkl} = g^{il}(h_{il}g_{jk} - h_{jl}g_{ik} + g_{il}h_{jk} - g_{jl}h_{ik})\\
&= g^{il}h_{il}g_{jk} - h_{jk} + nh_{jk} - h_{jk}.
\end{aligned}$$

which is equivalent to (c). Then (d) follows from (c) and the fact that $\operatorname{tr}_g g = n$.

Part (e) is also a direct computation:

$$\begin{aligned}
&\langle T, h \owedge g\rangle_g = g^{pi}g^{qj}g^{rk}g^{sl}T_{pqrs}(h \owedge g)_{ijkl}\\
&= g^{pi}g^{qj}g^{rk}g^{sl}T_{pqrs}(h_{il}g_{jk} - h_{jl}g_{ik} + g_{il}h_{jk} - g_{jl}h_{ik})\\
&= g^{pi}g^{sl}T_{pqrs}h_{il}(g^{qj}g^{rk}g_{jk}) - g^{qj}g^{sl}T_{pqrs}h_{jl}(g^{pi}g^{rk}g_{ik}) + g^{qj}g^{rk}T_{pqrs}h_{jk}g^{pi}g^{sl}g_{il}\\
&- g^{pi}g^{rk}T_{pqrs}h_{ik}g^{qj}g^{sl}g_{jl}\\
&= g^{pi}g^{sl}g^{rk}T_{pkrs}h_{il} - g^{qj}g^{sl}g^{pi}T_{pqis}h_{jl} + g^{qj}g^{rk}g^{pi}T_{pqri}h_{jk} - g^{pi}g^{rk}g^{qj}T_{pqrj}h_{ik}\\
&= g^{pi}g^{sl}(g^{rk}T_{rspk})h_{il} + g^{qj}g^{sl}(g^{pi}T_{pqsi})h_{jl} + g^{qj}g^{rk}(g^{pi}T_{pqri})h_{jk} + g^{pi}g^{rk}(g^{qj}T_{qprj})h_{ik}\\
&= 4g^{pi}g^{qj}(g^{rk}T_{rpqs})h_{ij} = 4\langle \operatorname{tr}_g T, h\rangle.
\end{aligned}$$

Finally, (f) follows from (e) by

$$\begin{aligned} |h \owedge g|_g^2 &= \langle h \owedge g, h \owedge g\rangle = 4\langle \operatorname{tr}_g(h \owedge g), h\rangle = 4\langle (n-2)h + (\operatorname{tr}_g h)g, h\rangle \\ &= 4(n-2)|h|_g^2 + 4\operatorname{tr}_g h\langle g, h\rangle \\ &= 4(n-2)|h|_g^2 + 4(\operatorname{tr}_g h)^2. \end{aligned}$$

This finishes the proof. □

Here is the primary application of the Kulkarni-Nomizu product.

ght inv trace **Proposition 2.1.29.** *Let (V, g) be an n-dimensional scalar product space with $n \geq 3$, and define a linear map $G : \Sigma^2(V^*) \to \mathcal{R}(V^*)$ by*

$$G(h) = \frac{1}{n-2}\Big(h - \frac{\operatorname{tr}_g h}{2(n-1)} g\Big) \owedge g.$$

Then G is a right inverse for tr_g, and its image is the orthogonal complement of the kernel of tr_g in $\mathcal{R}(V^*)$.

Proof. The fact that G is a right inverse is a straightforward computation based on the definition and Lemma 2.1.28(c, d). This implies that G is injective and tr_g is surjective, so the dimension of $\operatorname{im} G$ is equal to the codimension of $\ker \operatorname{tr}_g$, which in turn is equal to the dimension of $\ker G^\perp$. If $T \in \mathcal{R}(V^*)$ is an algebraic curvature tensor such that $\operatorname{tr}_g T = 0$, then Lemma 2.1.28(e) shows that $\langle T, G(h)\rangle = 0$, so it follows by dimensionality that $\operatorname{im} G = (\ker \operatorname{tr}_g)^\perp$. □

Now suppose g is a Riemannian or pseudo-Riemannian metric. Define the Schouten tensor of g, denoted by P, to be the following symmetric 2-tensor field:

$$P = \frac{1}{n-2}\Big(\operatorname{Ric} - \frac{S}{2(n-1)} g\Big),$$

and define the **Weyl tensor** of g to be the following algebraic curvature tensor field:

$$W = R - P \owedge g = R - \frac{1}{n-2}\Big(\operatorname{Ric} - \frac{S}{2(n-1)} g\Big) \owedge g.$$

sor traceless **Proposition 2.1.30.** *For every Riemannian or pseudo-Riemannian manifold (M, g) of dimension $n \geq 3$, the trace of the Weyl tensor is zero, and $R = W + P \owedge g$ is the orthogonal decomposition of R corresponding to $\mathcal{R}(V^*) = \ker \operatorname{tr}_g \oplus (\ker \operatorname{tr}_g)^\perp$.*

Proof. This follows immediately from Proposition 2.1.29 and the fact that $P \owedge g = G(\operatorname{Ric}) = G(\operatorname{tr}_g R)$. □

These results lead to some important simplifications in low dimensions.

ure dim=1,2,3 **Corollary 2.1.31.** *Let V be an n-dimensional real vector space.*

(a) If $n = 0$ or 1, then $\mathcal{R}(V^*) = \{0\}$.

(b) If $n = 2$, then $\mathcal{R}(V^*)$ is 1-dimensional, spanned $g \owedge g$.

(c) If $n = 3$, then $\mathcal{R}(V^*)$ is 6-dimensional, and $G : \Sigma^2(V^*) \to \mathcal{R}(V^*)$ is an isomorphism.

Proof. The dimensional results follow immediately from Proposition 1.28. In the case $n = 2$, Lemma 2.1.28(d) shows that $\mathrm{tr}_g(g \owedge g) = 2g \neq 0$, which implies that $g \owedge g$ is nonzero and therefore spans the 1-dimensional space $\mathcal{R}(V^*)$.

Now consider $n = 3$. Proposition 2.1.29 shows that $\mathrm{tr}_g \circ G$ is the identity, which means that G is injective. On the other hand, Proposition 1.28 shows that $\dim \mathcal{R}(V^*) = 6$, so G is also surjective. □

The next corollary shows that the entire curvature tensor is determined by the Ricci tensor in dimension 3.

Corollary 2.1.32. *On every Riemannian or pseudo-Riemannian manifold (M, g) of dimension 3, the Weyl tensor is zero, and the Riemann curvature tensor is determined by the Ricci tensor via the formula*

$$R = P \owedge g = \mathrm{Ric} \owedge g - \frac{1}{4} S g \owedge g. \tag{1.29}$$

Proof. Corollary 2.1.31 shows that $G : \Sigma^2(V^*) \to \mathcal{R}(V^*)$ is an isomorphism in dimension 3. Since $\mathrm{tr}_g \circ G$ is the identity, it follows that tr_g is also an isomorphism. Because $\mathrm{tr}_g W$ is always zero by Proposition 2.1.30, it follows that W is always zero. Formula (1.29) then follows from the definition of the Weyl tensor. □

In dimension 2, the definitions of the Weyl and Schouten tensors do not make sense; but we have the following analogous result instead.

Corollary 2.1.33. *On every Riemannian or pseudo-Riemannian manifold (M, g) of dimension 2, the Riemann and Ricci tensors are determined by the scalar curvature as follows:*

$$R = \frac{1}{4} S g \owedge g, \quad \mathrm{Ric} = \frac{1}{2} S g. \tag{1.30}$$

Proof. In dimension 2, it follows from Corollary 2.1.31(b) that there is some scalar function f such that $R = fg \owedge g$. Taking traces, we find from Lemma 2.1.28(d) that $\mathrm{Ric} = \mathrm{tr}_g R = 2fg$, and then $S = \mathrm{tr}_g \mathrm{Ric} = 2f \mathrm{tr}_g g = 4f$. The results follow by substituting $f = S/4$ back into these equations. □

Although the traceless Ricci tensor is always zero on a 2-manifold, this does not imply that S is constant, since the proof of Schur's lemma fails in dimension 2. Einstein metrics in dimension 2 are simply those with constant scalar curvature.

Returning now to dimensions greater than 2, we can use (7.27) to further decompose the Schouten tensor into a part determined by the traceless Ricci tensor and a purely scalar part. The next proposition is the analogue of Proposition 2.1.24 for the full curvature tensor.

Proposition 2.1.34 (The Ricci Decomposition of the Curvature Tensor). *Let (M, g) be a Riemannian or pseudo-Riemannian manifold of dimension $n \geq 3$. Then the $(0, 4)$-curvature tensor of g has the following orthogonal decomposition:*

$$R = W + \frac{1}{n-2} \mathring{\mathrm{Ric}} \owedge g + \frac{1}{2n(n-1)} S g \owedge g. \tag{1.31}$$

Therefore, in the Riemannian case,

$$\begin{aligned} |R|_g^2 &= |W|_g^2 + \frac{1}{(n-2)^2}|\mathring{\mathrm{Ric}} \owedge g|_g^2 + \frac{1}{4n^2(n-1)^2}|Sg \owedge g|_g^2 \\ &= |W|_g^2 + \frac{4}{(n-2)}|\mathring{\mathrm{Ric}}|_g^2 + \frac{2}{n(n-1)}S^2 \end{aligned} \tag{1.32}$$

Proof. The decomposition (1.31) follows immediately by substituting (1.19) into the definition of the Weyl tensor and simplifying. The decomposition is orthogonal thanks to Lemma 2.1.28(e), and (1.32) follows from Lemma 2.1.28(f). ☐

2.1.6 Curvature of a surface

In this part we will introduce the Gaussian curvature for a surface in $\mathbb{R}^3$. This will serve as our motivation for the definition of curvature.

Gaussian curvature for surfaces in $\mathbb{R}^3$

By a surface in $\mathbb{R}^3$, we mean an embedded 2-dimensional smooth submanifold of $\mathbb{R}^3$. Let p be a point on a surface M in $\mathbb{R}^3$. A normal vector to M at p is a vector $N_p \in T_p\mathbb{R}^3$ that is orthogonal to the tangent plane T_pM. A normal vector field on M is a function N that assigns to each $p \in M$ a normal vector N_p at p. If N is a normal vector field on M, then at each point $p \in M$, we can write $N = N^i\partial_i$. The normal vector field N on M is said to be smooth if the coefficient functions are smooth functions on M.

Let N be a smooth unit normal vector field on a neighborhood of p in M. Denote by N_p the unit normal vector at p. Under the canonical identification of $T_p\mathbb{R}^3$ with $\mathbb{R}^3$, every plane P through N_p slices the surface M along a plane curve $P \cap M$ through p. We call such a plane curve a **normal section** of the surface through p.

Proposition 2.1.35. *Let M be a surface in $\mathbb{R}^3$ and $p \in M$. A normal section of M at p is a smooth submanifold of dimension one in a neighborhood of p.*

Proof. The plane P is the zero set of the linear function $f(x) = \langle x, Y_p\rangle + d$, where Y_p is the normal vector of P and $d \in \mathbb{R}$ is such that $f(p) = 0$. Let $\widetilde{f} : M \to \mathbb{R}$ be the restriction of f to M. Then the normal section C is precisely the level set $\widetilde{f}^{-1}(0)$. Now note that $\operatorname{grad} f = Y_p$ is non-zero, and since P is spanned by N_p and X_p, the vector Y_p is contained in the tangent space T_pM. This then means $\operatorname{grad}\widetilde{f} =$ tangent part of Y_p is nonzero, so p is a regular point of $\widetilde{f}$. This then implies that $P \cap M$ is a smooth submanifold of dimension one in a neighborhood of p. ☐

Now, since the normal section is a smooth curve around p, we can compute the curvature of a normal section with respect to N_p. The collection of the curvatures at p of all the normal sections gives a fairly good picture of how the surface curves at p.

More precisely, each unit tangent vector X_p to the surface M at p determines together with N_p a plane that slices M along a normal section. Moreover, the unit tangent vector X_p determines an orientation of the normal section. Let $\gamma(s)$ be the arc length parametrization of this normal section with initial point $\gamma(0) = p$ and initial vector $\gamma'(0) = X_p$. Note that $\gamma(s)$

is completely determined by the unit tangent vector X_p. Define the normal curvature of the normal section $\gamma(s)$ at p with respect to N_p by

$$\kappa(X_p) = \langle \gamma''(0), N_p \rangle. \tag{1.33}$$

Since the set of all unit vectors in T_pM is a circle, we have a function $\kappa : S^1 \to \mathbb{R}$. Clearly, $\kappa(-X_p) = \kappa(X_p)$ for $X_p \in S^1$, because replacing a unit tangent vector by its negative simply reverses the orientation of the normal section, which reverses the sign of the first derivative $\gamma'(s)$ but does not change the sign of the second derivative $\gamma''(s)$.

The maximum and minimum values κ_1, κ_2 of the function κ are the **principal curvatures** of the surface at p. Their average $(\kappa_1+\kappa_2)/2$ is the **mean curvature** H, and their product $\kappa_1\kappa_2$ the **Gaussian curvature** K. A unit direction $X_p \in T_pM$ along which a principal curvature occurs is called a principal direction. Note that if X_p is a principal direction, then so is $-X_p$, since $\kappa(-X_p) = \kappa(X_p)$. If κ_1 and κ_2 are equal, then every unit vector in T_pM is a principal direction.

Remark 2.1.2. Using $-N_p$ instead of N_p reverses the signs of all the normal curvatures at p, as one sees from (1.33). This will change the sign of the mean curvature, but it leaves invariant the Gaussian curvature. Thus, the Gaussian curvature K is independent of the choice of the unit normal vector field N.

Example 2.1.36. Every normal section of a sphere of radius R is a circle of radius R. With respect to the unit inward-pointing unit normal vector field, the principal curvatures are both $1/R$, the mean curvature is $H = 1/R$ and the Gaussian curvature is $K = 1/R^2$.

Example 2.1.37. For a plane M in $\mathbb{R}^3$ the principal curvatures, mean curvature, and Gaussian curvature are all zero.

Example 2.1.38. For a cylinder of radius a with a unit inward normal, it appears that the principal curvatures are 0 and $1/R$, corresponding to normal sections that are a line and a circle, respectively. We will establish this rigorously later. Hence, the mean curvature is $1/2R$ and the Gaussian curvature is 0. If we use the unit outward normal on the cylinder, then the principal curvatures are 0 and $-1/R$, and the mean curvature is $-1/2R$, but the Gaussian curvature is still 0.

Since the plane is locally isometric to a cylinder, the above tow examples show that the principal curvatures κ_1, κ_2 and the mean curvature H are not isometric invariants. It is an astonishing fact that although neither κ_1 nor κ_2 is invariant under isometries, their product, the Gaussian curvature K, is. This is the content of the Theorema Egregium of Gauss. We will prove this later.

The shape operator

The definition above provides an geometric intuition for Gaussian curvature, but is rather hard to compute. Now we use the Levi-Civita connection of M to give an alternative description for the Gaussian curvature.

Let p be a point on a surface M in $\mathbb{R}^3$ and let N be a smooth unit normal vector field on M. For any tangent vector$X_p \in T_oM$, define

$$L_p(X_p) = -\overline{\nabla}_{X_p} N,$$

where $\overline{\nabla}$ is the Euclidean connection of $\mathbb{R}^3$. Applying the vector X_p to $\langle N, N\rangle \equiv 1$ gives $\langle \overline{\nabla}_{X_p} N, N_p\rangle = 0$, thus $\overline{\nabla}_{X_p} N$ is perpendicular to N_p at p and is therefore in the tangent plane T_pM. So L_p is a linear map $T_pM \to T_pM$. It is called the **shape operator** or the **Weingarten map** of the surface M at p. Note that the shape operator depends on the unit normal vector field N and the point p. With the unit normal vector field N on M fixed, as the point p varies in M, there is a different shape operator L_p at each p, therefore we often omit p. The shape operator, being the directional derivative of a unit normal vector field on a surface, should encode in it information about how the surface bends at p.

Lemma 2.1.39. *Let M be a surface in $\mathbb{R}^3$ having a smooth unit normal vector field N. For $X, Y \in \mathfrak{X}(M)$,*

$$\langle L(X), Y\rangle = \langle \overline{\nabla}_X Y, N\rangle.$$

Proof. Since Y is tangent to M, the inner product $\langle Y, N\rangle$ is identically zero on U. Differentiating the equation $\langle Y, N\rangle \equiv 0$ with respect to X yields $\langle \overline{\nabla}_X Y, N\rangle + \langle Y, \overline{\nabla}_X N\rangle = 0$, so the claim follows. □

With this lemma, we can prove the following important result.

Proposition 2.1.40. *The shape operator is self-adjoint: for any $X_p, Y_p \in T_pM$,*

$$\langle L_p(X_p), Y_p\rangle = \langle X_p, L_p(Y_p)\rangle.$$

Therefore it has real eigenvalues.

Proof. Suppose the smooth unit normal vector field N is defined on a neighborhood U of p. Let X, Y be vector fields on U that extend the vectors X_p, Y_p at p. Then by Lemma 2.1.39 we have

$$\langle L(X), Y\rangle = \langle N, \overline{\nabla}_X Y\rangle,$$
$$\langle X, L(Y)\rangle = \langle N, \overline{\nabla}_Y X\rangle.$$

Substrating these two equations, we get

$$\langle L(X), Y\rangle - \langle X, L(Y)\rangle = \langle N, \overline{\nabla}_X Y - \overline{\nabla}_Y X\rangle = \langle N, [X, Y]\rangle,$$

which is zero since $[X, Y]$ is tangent to M and N is normal to M. □

Now we will see the connection between the Gaussian curvature and the shape operator. Consider as before a surface M in $\mathbb{R}^3$ having a smooth unit normal vector field N. Lemma 2.1.39 on the shape operator has a counterpart for vector fields along a curve.

Proposition 2.1.41. *Let $\gamma : [a, b] \to M$ be a curve in M and let V be a vector field in M along γ. Then*

$$\langle L(\gamma'(t), V)\rangle = \langle \overline{D}_t V, N_{\gamma(t)}\rangle.$$

Proof. This is proved in the same way as Lemma 2.1.39. □

Proposition 2.1.42. *Suppose $\gamma(s)$ is a normal section, parametrized by arc length, determined by a unit tangent vector $X_p \in T_pM$ and the unit normal vector N_p. Then the normal curvature of $\gamma(s)$ with respect to N_p at p is given by the* ***second fundamental form****:*

$$\kappa(X_p) = \langle L(X_p), X_p\rangle =: \mathrm{II}(X_p, X_p).$$

Proof. By definition, $\gamma(0) = p$ and $\gamma'(0) = X_p$. Let $T(s) := \gamma'(s)$ be the unit tangent vector field along $\gamma(s)$. Then the curvature of the normal section $\gamma(s)$ is

$$\begin{aligned}\kappa(\gamma'(s)) &= \langle \gamma''(s), N_{\gamma(s)} \rangle \\ &= \langle \frac{dT}{ds}, N_{\gamma(s)} \rangle \\ &= \langle L(\gamma'(s)), T \rangle \\ &= \langle L(T), T \rangle.\end{aligned}$$

Evaluating at $s = 0$ gives the claim. □

Corollary 2.1.43. *The principal directions of the surface M in $\mathbb{R}^3$ at p are the unit eigenvectors of the shape operator L; the principal curvatures are the eigenvalues of L, and the Gaussian curvature of a surface M in $\mathbb{R}^3$ is the determinant of the shape operator.*

Proof. The principal curvatures at p are the maximum and minimum values of the function

$$\kappa(X_p) = \langle L(X_p), X_p \rangle = \mathrm{II}(X_p, X_p).$$

for $X_p \in T_pM$ satisfying $\langle X_p, X_p \rangle = 1$, which turns out to be the eigenvalues of L_p. □

In this situation let's introduce the first and second fundamental forms of a surface. Let M be a surface of $\mathbb{R}^3$ and $p \in M$, the Euclidean Riemannian metric for M at p is called the **first fundamental form** of M at p. If $L : T_pM \to T_pM$ is the shape operator, the symmetric bilinear form

$$\mathrm{II}(X_p, Y_p) = \langle L_p(X_p), Y_p \rangle$$

is called the second fundamental form of the surface M at p. The first fundamental form is the metric and the second fundamental form is essentially the shape operator.

Let e_1, e_2 be a basis for the tangent space T_pM. We set

$$E = \langle e_1, e_1 \rangle, \quad F = \langle e_1, e_2 \rangle, \quad G = \langle e_2, e_2 \rangle.$$

The three numbers E, F, G determine completely the first fundamental form of M at p. They are called the **coefficients of the first fundamental form** relative to e_1, e_2. Similarly, the three numbers

$$e = \mathrm{II}(e_1, e_1), \quad f = \mathrm{II}(e_1, e_2), \quad f = \mathrm{II}(e_2, e_2)$$

determine completely the second fundamental form of M at p, and are called the **coefficients of the second fundamental form** relative to e_1, e_2.

amental forms **Proposition 2.1.44.** *Suppose M is an oriented surface in $\mathbb{R}^3$, $p \in M$, and e_1, e_2 is a basis for T_pM. Let E, F, G, e, f, g be the coefficients of the first and second fundamental forms of M at p relative to e_1, e_2. Then the shape operator is givne by the matrix*

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} e & f \\ f & g \end{pmatrix}$$

under e_1, e_2.

Proof. If the shape operator has matrix L, then we have

$$\begin{pmatrix} \langle L(e_1), e_1 \rangle & \langle L(e_1), e_2 \rangle \\ \langle L(e_2), e_1 \rangle & \langle L(e_2), e_2 \rangle \end{pmatrix} = \begin{pmatrix} \langle e_1, e_1 \rangle & \langle e_1, e_2 \rangle \\ \langle e_2, e_1 \rangle & \langle e_2, e_2 \rangle \end{pmatrix} \begin{pmatrix} L(e_1) & L(e_2) \end{pmatrix}$$

This imples the claim. □

Theorem 2.1.45. *Suppose M and $\widetilde{M}$ are two surfaces of $\mathbb{R}^3$, and $\varphi : M \to \widetilde{M}$ is a diffeomorphism. Let E, F, G be the coefficients of the first fundamental form relative to a frame X_1, X_2 on M, and $\widetilde{E}, \widetilde{G}, \widetilde{F}$ the corresponding coefficients relative to $\widetilde{X}_1 := \varphi_* X_1$, $\widetilde{X}_2 := \varphi_* X_2$ on $\widetilde{M}$. Then φ is an isometry if and only if E, F, G at p are equal to $\widetilde{E}, \widetilde{G}, \widetilde{F}$ at $\varphi(p)$, respectively, for all $p \in M$.*

Proof. The diffeomorphism φ is an isometry if and only if

$$\langle \varphi_* X_p, \varphi_* Y_p \rangle_{\varphi(p)} = \langle X_p, Y_p \rangle_p$$

for all $p \in M$. This condition holds if and only if E, F, G at p are equal to $\widetilde{E}, \widetilde{G}, \widetilde{F}$ at $\varphi(p)$. □

Now we consider the computation of Gaussian curvature. We have the following result.

ian curvature

Proposition 2.1.46. *Let $M \subseteq \mathbb{R}^3$ be a surface parametrized by*

$$X : U \to \mathbb{R}^3, \quad (u, v) \mapsto (x, y, z).$$

Then the first and second fundamental forms of M are given by

$$\mathrm{I} = \begin{pmatrix} \langle X_u, X_u \rangle & \langle X_u, X_v \rangle \\ \langle X_v, X_u \rangle & \langle X_v, X_v \rangle \end{pmatrix} \quad \mathrm{II} = \begin{pmatrix} \langle X_{uu}, N \rangle & \langle X_{uv}, N \rangle \\ \langle X_{uv}, N \rangle & \langle X_{vv}, N \rangle \end{pmatrix}$$

under the frame (∂_u, ∂_v) and the unit normal vector field $N = X_u \times X_v / |X_u \times X_v|$.

Proof. The frame (∂_u, ∂_v) is linear independent every where in U since X is a smooth parametrization, therefore we can compute L under it. We choose the normal unit vector field

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

Riemann shape operator lem

Recall that by Lemma 2.1.39 we have

$$\mathrm{II}(X, Y) = \langle L(X), Y \rangle = \langle \overline{\nabla}_X Y, N \rangle.$$

Apply this on ∂_u we get

$$\mathrm{II}(\partial_u, \partial_u) = \langle \overline{\nabla}_{\partial_u} \partial_u, N \rangle.$$

Note that

$$\overline{\nabla}_{\partial_u} \partial_u = \frac{\partial}{\partial u}\Big(X_u^1 \frac{\partial}{\partial x} + X_u^2 \frac{\partial}{\partial y} + X_u^3 \frac{\partial}{\partial z} \Big) = X_{uu}^1 \frac{\partial}{\partial x} + X_{uu}^2 \frac{\partial}{\partial y} + X_{uu}^3 \frac{\partial}{\partial z} = X_{uu},$$

therefore the claim follows. □

Example 2.1.47 (Shape operator of a sphere). Let M be the sphere of radius R in $\mathbb{R}^3$. Parametrize the sphere using spherical coordinates:

$$(x, y, z) = (R \cos\theta \sin\phi, R \sin\theta \sin\phi, R \cos\phi).$$

Then the unit normal vector is $N = (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi)$. This enables us to compute

Riemann parametrized surface Guassian curvature

the Guassian curvature by Proposition 2.1.46:

$$X_\theta = (-R \sin\theta \sin\phi, R \cos\theta \sin\phi, 0),$$
$$X_\phi = (R \cos\theta \cos\phi, R \sin\theta \cos\phi, -R \sin\phi),$$

and

$$\begin{aligned}
X_{\theta\theta} &= (-R\cos\theta\sin\phi, -R\sin\theta\sin\phi, 0),\\
X_{\theta\phi} &= (-R\sin\theta\cos\phi, R\cos\theta\cos\phi, 0),\\
X_{\phi\phi} &= (-R\cos\theta\sin\phi, -R\sin\theta\sin\phi, -R\cos\phi).
\end{aligned}$$

Therefore

$$E = R^2\sin^2\phi, \quad F = 0, \quad G = R^2.$$

$$e = -R\sin^2\phi, \quad f = 0, \quad g = -R.$$

Therefore the shape operator is given by

$$\begin{pmatrix} \frac{1}{R^2\sin^2\phi} & \\ & \frac{1}{R^2} \end{pmatrix}\begin{pmatrix} -R\sin^2\phi & \\ & -R \end{pmatrix} = \begin{pmatrix} -\frac{1}{R} & \\ & -\frac{1}{R} \end{pmatrix}$$

and the Gaussian curvature is

$$K = \frac{R^2\sin^2\phi}{R^4\sin^2\phi} = \frac{1}{R^2}.$$

Example 2.1.48 (Shape operator of a cylinder). Let M be the cylinder of radius R in $\mathbb{R}^3$ defined by $x^2+y^2=R^2$. At $p=(x,y,z)\in M$, the vectors

$$X = \frac{1}{R}\Big(-y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}\Big), \quad Y = \frac{\partial}{\partial z}$$

form an orthonormal frame of M. Let N be the unit normal vector field

$$N = \frac{1}{R}\Big(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}\Big).$$

Then we can compute that $\overline{\nabla}_X N = (1/R)X$ and $\overline{\nabla}_Y N = 0$, therefore the shape operator has a matrix expression:

$$\begin{pmatrix} -\frac{1}{R} & 0 \\ 0 & 0 \end{pmatrix}$$

Also, this gives $K=0$.

Example 2.1.49 (The first and second fundamental forms of a graph). Let M be the graph of a smooth function $z=f(x,y)$ in $\mathbb{R}^3$. Then M can be parametrized by $F(x,y)=(x,y,f(x,y))$. We compute that

$$\begin{aligned}
&F_x = (1,0,f_x), \quad F_y = (0,1,F_y), \quad N = \frac{F_x\times F_y}{|F_x\times F_y|} = \frac{(-f_x,-f_y,1)}{\sqrt{1+f_x^2+f_y^2}},\\
&F_{xx} = (0,0,f_{xx}), \quad F_{xy} = (0,0,f_{xy}), \quad F_{yy} = (0,0,f_{yy}).
\end{aligned}$$

Therefore the first and second fundamental form are given by

$$E = 1+f_x^2, \quad F = f_xf_y, \quad G = 1+f_y^2,$$

and

$$e = \frac{f_{xx}}{\sqrt{1+f_x^2+f_y^2}}, \quad f = \frac{f_{xy}}{\sqrt{1+f_x^2+f_y^2}}, \quad g = \frac{f_{yy}}{\sqrt{1+f_x^2+f_y^2}}.$$

This then implies

$$K = \frac{f_{xx}f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2}$$

Example 2.1.50 (The first and second fundamental forms of a surface of revolution). Let C be an embedded one dimensional submanifold of the half-plane $\{(r,z) : r > 0\}$ and let S_C be the surface of revolution generated by C as described in Example surfaces of revolution **??**. To compute the induced metric on S_C, choose any smooth local parametrization $\gamma(t) = (a(t), b(t))$ for C, and note that the map

$$X(t,\theta) = (a(t)\cos\theta, a(t)\sin\theta, b(t))$$

yields a smooth local parametrization of S_C, provided that (t,θ) is restricted to a sufficiently small open subset of the plane. Thus we can compute

$$X_t = (a'(t)\cos\theta, a'(t)\sin\theta, b'(t)), \quad X_\theta = (-a(t)\sin\theta, a(t)\cos\theta, 0).$$

Therefore the normal unit vector field is given by

$$N = \frac{X_t \times X_\theta}{|X_t \times X_\theta|} = \frac{(-b'(t)\cos\theta, -b'(t)\sin\theta, a'(t))}{\sqrt{a'(t)^2 + b'(t)^2}}.$$

Also, the second derivatives are

$$\begin{aligned} X_{tt} &= (a''(t)\cos\theta, a''(t)\sin\theta, b''(t)), \\ X_{t\theta} &= (-a'(t)\sin\theta, a'(t)\cos\theta, 0), \\ X_{\theta\theta} &= (-a(t)\cos\theta, -a(t)\sin\theta, 0). \end{aligned}$$

With these preparations, we can now derive that

$$E = a'(t)^2 + b'(t)^2, \quad F = 0, \quad G = a(t)^2.$$

and

$$e = \frac{a'(t)b''(t) - a''(t)b'(t)}{\sqrt{a'(t)^2 + b'(t)^2}}, \quad f = 0, \quad g = \frac{a(t)b'(t)}{\sqrt{a'(t)^2 + b'(t)^2}}.$$

Therefore the Gaussian curvature is given by

$$K = \frac{b'(t)(a'(t)b''(t) - b'(t)a''(t))}{a(t)(a'(t)^2 + b'(t)^2)^3}.$$

Gauss's theorema egregium

For a surface in $\mathbb{R}^3$ we defined its Gaussian curvature K at a point p by taking normal sections of the surface, finding the maximum κ_1 and the minimum κ_2 of the curvature of the normal sections, and setting K to be the product of κ_1 and κ_2. So defined, the Gaussian curvature evidently depends on how the surface is isometrically embedded in $\mathbb{R}^3$.

On the other hand, an abstract Riemannian manifold has a unique Riemannian connection. The curvature tensor $R(X,Y)$ of the Riemannian connection is then completely determined by the Riemannian metric and so is an intrinsic invariant of the Riemannian manifold, independent of any embedding. We think of a surface in $\mathbb{R}^3$ as a particular isometric embedding of an abstract Riemannian manifold of dimension 2. For example, both a plane and a cylinder are locally isometric embeddings of the same abstract surface, as one sees by simply bending a piece of paper. We will show that the Gaussian curvature of a surface in $\mathbb{R}^3$ is expressible in terms of the curvature tensor $R(X,Y)$ and the metric. Hence, it too depends only on the metric, not on the particular embedding into $\mathbb{R}^3$.

Suppose M is a smooth submanifold of $\mathbb{R}^n$ and ∇ is its Levi-Civita connection. Let $X, Y \in \mathfrak{X}(M)$, then ∇ is defined by

$$\nabla_X Y = \pi^\top(\overline{\nabla}_X Y)$$

where $\pi^\top : T\mathbb{R}^n|_M \to TM$ is the tangent map. Let N be a unit normal vector field on M, then the connection $\nabla_X Y$ can be expressed as

$$\nabla_X Y = \overline{\nabla}_X Y - \langle \overline{\nabla}_X Y, N\rangle N.$$

Recall the definition of the shape operator and Lemma 2.1.39 (Riemann shape operator lem), we then find

$$\overline{\nabla}_X Y = \nabla_X Y + \langle \overline{\nabla}_X Y, N\rangle N = \nabla_X Y + \langle L(X), Y\rangle N.$$

This leads to the following result.

Theorem 2.1.51. *Let M be an oriented surface in $\mathbb{R}^3$, ∇ its Levi-civita connection, R the curvature operator of ∇, and L the shape operator on M. For $X, Y, Z \in \mathfrak{X}(M)$,*

*(i) (**Gauss curvature equation**)*

$$R(X,Y)Z = \langle L(Y), Z\rangle L(X) - \langle L(X), Y\rangle L(Y) = \det\begin{pmatrix} \langle L(X), \cdot\,\rangle & \langle L(X), Z\rangle \\ \langle L(Y), \cdot\,\rangle & \langle L(Y), Z\rangle \end{pmatrix}$$

*(ii) (**Codazzi-Mainardi equation**)*

$$\nabla_X L(Y) - \nabla_Y L(X) - L([X,Y]) = 0.$$

Proof. Decomposing $\overline{\nabla}_X Y$ into its tangential and normal components, one has

$$\overline{\nabla}_X Y = \nabla_X Y + \langle L(X), Y\rangle N. \tag{1.34}$$ Riemann surfa

Hence,

$$\begin{aligned}
\overline{\nabla}_X \overline{\nabla}_Y Z &= \overline{\nabla}_X(\nabla_Y Z + \langle L(Y), Z\rangle N) \\
&= \overline{\nabla}_X \nabla_Y Z + \overline{\nabla}_X(\langle L(Y), Z\rangle N) \\
&= \nabla_X \nabla_Y Z + \langle L(X), \nabla_Y Z\rangle N + \overline{\nabla}_X \langle L(Y), Z\rangle N + \langle L(Y), Z\rangle \overline{\nabla}_X N \\
&= \nabla_X \nabla_Y Z - \langle L(Y), Z\rangle L(X) + \langle L(X), \nabla_Y Z\rangle N + \overline{\nabla}_X \langle L(Y), Z\rangle N.
\end{aligned}$$

Interchanging X and Y gives

$$\overline{\nabla}_Y \overline{\nabla}_X Z = \nabla_Y \nabla_X Z - \langle L(X), Z\rangle L(Y) + \langle L(Y), \nabla_X Z\rangle N + \overline{\nabla}_Y \langle L(X), Z\rangle N.$$

Also,

$$\overline{\nabla}_{[X,Y]} Z = \nabla_{[X,Y]} Z + \langle L([X,Y]), Z\rangle N.$$

Therefore one gets

$$\begin{aligned} 0 = \overline{R}(X,Y)Z = R(X,Y)Z - \langle L(Y), Z\rangle L(X) + \langle L(X), Z\rangle L(Y) \\ + \text{normal part}. \end{aligned} \tag{1.35}$$ Riemann surfa

Therefore,

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

which is (i).

Note that the normal part of (1.35) is

$$\begin{aligned}
&\big(\langle L(X), \nabla_Y Z\rangle - \langle L(Y), \nabla_X Z\rangle + \overline{\nabla}_X\langle L(Y), Z\rangle - \overline{\nabla}_Y\langle L(X), Z\rangle - \langle L([X,Y]), Z\rangle\big)N\\
&\qquad = \big(\langle L(X), \nabla_Y Z - \overline{\nabla}_X Z\rangle - \langle L(Y), \nabla_X Z - \overline{\nabla}_X Z\rangle\\
&\qquad\quad + \langle \overline{\nabla}_X L(Y), Z\rangle - \langle \overline{\nabla}_Y L(X), Z\rangle - \langle L([X,Y]), Z\rangle\big)N\\
&\qquad = \big(\langle \overline{\nabla}_X L(Y), Z\rangle - \langle \overline{\nabla}_Y L(X), Z\rangle - \langle L([X,Y]), Z\rangle\big)N\\
&\qquad = \big(\langle \nabla_X L(Y), Z\rangle - \langle \nabla_Y L(X), Z\rangle - \langle L([X,Y]), Z\rangle\big)N
\end{aligned}$$

where we use (1.34). This gives (ii). □

Theorem 2.1.52 (Theorema Egregium). *Let M be a surface in $\mathbb{R}^3$ and p a point in M.*

(i) If e_1, e_2 is an orthonormal basis for the tangent plane T_pM, then the Gaussian curvature K_p of M at p is

$$K_p = \langle R(e_1, e_2)e_2, e_1\rangle \tag{1.36}$$

(ii) The Gaussian curvature is invariant under local isometry of smooth surfaces in $\mathbb{R}^3$.

Proof. In Proposition 2.1.44 we found a formula for the Gaussian curvature K_p in terms of the shape operator L and an orthonormal basis e_1, e_2 for T_pM:

$$K_p = \det\begin{pmatrix} \langle L(e_1), e_1\rangle & \langle L(e_1), e_2\rangle \\ \langle L(e_2), e_1\rangle & \langle L(e_2), e_2\rangle \end{pmatrix} = \langle L(e_1), e_1\rangle\langle L(e_2), e_2\rangle - \langle L(e_1), e_2\rangle\langle L(e_2), e_1\rangle.$$

By the Gauss curvature equation,

$$R(e_1, e_2)e_1 = \langle L(e_1), e_1\rangle L(e_2) - \langle L(e_1), e_2\rangle L(e_2).$$

Taking the inner product with e_1 gives (i).

Since $R(e_1, e_2)$ is determined completely by the metric, by (1.36) the same can be said of K_p. □

The Theorema Egregium gives a formula for the Gaussian curvature of a surface in terms of an orthonormal basis for the tangent plane at a point. From it, one can derive a formula for the Gaussian curvature in terms of an arbitrary basis.

Proposition 2.1.53. *Let M be a smooth surface in $\mathbb{R}^3$ and $p \in M$. If u, v is any basis for the tangent plane T_pM, then the Gaussian curvature at p is*

$$K_p = \frac{\langle R(u,v)v, u\rangle}{\langle u,u\rangle\langle v,v\rangle - \langle u,v\rangle^2}.$$

Proof. This follows from Proposition 2.1.44. □

Recall that if (E_1, E_2) is a local orthonormal frame for a surface M in $\mathbb{R}^3$, then the curvature tensor of M is determined by the 2-forms Ω_i^j, where

$$R(X,Y)E_i = \Omega_i^j(X,Y)E_j.$$

Since the matrix $[\Omega_i^j]$ is skew-symmetric, by Proposition 2.1.53, we see

$$K_p = \langle R(E_1, E_2)E_2, E_1\rangle = \langle \Omega_2^1(E_1, E_2)E_1, E_1\rangle = -\Omega_1^2(E_1, E_2).$$

If $(\varepsilon^1, \varepsilon^2)$ is the dual frame, then by (1.8) this can be written as

$$d\omega_1^2 = \Omega_1^2 = -K_p\, \varepsilon^1 \wedge \varepsilon^2.$$

This gives an effective way to compute the Gauss curvature for surfaces in $\mathbb{R}^3$.

2.2 Riemannian submanifolds

2.2.1 The second fundamental form

Suppose (M, g) is a Riemannian submanifold of a Riemannian manifold $(\widetilde{M}, \widetilde{g})$. Recall that this means that M is a submanifold of $\widetilde{M}$ endowed with the induced metric $g = \iota_M^* \widetilde{g}$ (where $\iota_M : M \hookrightarrow \widetilde{M}$ is the inclusion map). Our goal is to study the relationship between the geometry of M and that of $\widetilde{M}$.

Throughout this section, we assume that $(\widetilde{M}, \widetilde{g})$ is a Riemannian or pseudo-Riemannian manifold of dimension m, and (M, g) is an embedded n-dimensional Riemannian submanifold of $\widetilde{M}$. We call $\widetilde{M}$ the ambient manifold. We will denote covariant derivatives and curvatures associated with (M, g) in the usual way, and write those associated with $\widetilde{M}$ with tildes. We can unambiguously use the inner product notation $\langle v, w\rangle$ to refer either to g or to $\widetilde{g}$, since g is just the restriction of $\widetilde{g}$ to pairs of vectors in TM.

Our first main task is to compare the Levi-Civita connection of M with that of $\widetilde{M}$. The starting point for doing so is the orthogonal decomposition of sections of the ambient tangent bundle $T\widetilde{M}|_M$ into tangential and orthogonal components. Just as we did for submanifolds of $\mathbb{R}^n$, we define orthogonal projection maps called **tangential** and **normal projections**:

$$\pi^\top : T\widetilde{M}|_M \to TM, \quad \pi^\perp : T\widetilde{M}|_M \to NM.$$

If X is a section of $T\widetilde{M}|_M$, we often use the shorthand notations $X^\top = \pi^\top X$ and $X\perp = \pi^\perp X$ for its tangential and normal projections.

If X, Y are vector fields in $\mathfrak{X}(M)$, we can extend them to vector fields on an open subset of $\widetilde{M}$ (still denoted by X and Y), apply the ambient covariant derivative operator $\widetilde{\nabla}$, and then decompose at points of M to get

$$\widetilde{\nabla}_X Y = (\widetilde{\nabla}_X Y)^\top + (\widetilde{\nabla}_X Y)^\perp. \tag{2.1}$$ Riemann conne

We wish to interpret the two terms on the right-hand side of this decomposition.

Let us focus first on the normal component. We define the **second fundamental form** of M to be the map $\mathrm{II} : \mathfrak{X}(M) \to \mathfrak{X}(M) \to \Gamma(NM)$ given by

$$\mathrm{II}(X, Y) = (\widetilde{\nabla}_X Y)^\perp.$$

where X and Y are extended arbitrarily to an open subset of $\widetilde{M}$. Since $\pi^\perp$ maps smooth sections to smooth sections, $\mathrm{II}(X, Y)$ is a smooth section of NM.

ond form prop **Proposition 2.2.1 (Properties of the Second Fundamental Form).** *Suppose (M, g) is an embedded Riemannian submanifold of a Riemannian or pseudo-Riemannian manifold $(\widetilde{M}, \widetilde{g})$, and let $X, Y \in \mathfrak{X}(M)$.*

(a) $\mathrm{II}(X, Y)$ *is independent of the extensions of X and Y to an open subset of $\widetilde{M}$.*

(b) $\mathrm{II}(X, Y)$ *is bilinear over $C^\infty(M)$ in X and Y.*

(c) $\mathrm{II}(X, Y)$ *is symmetric in X and Y.*

(d) The value of $\mathrm{II}(X, Y)$ *at a point $p \in M$ depends only on X_p and Y_p.*

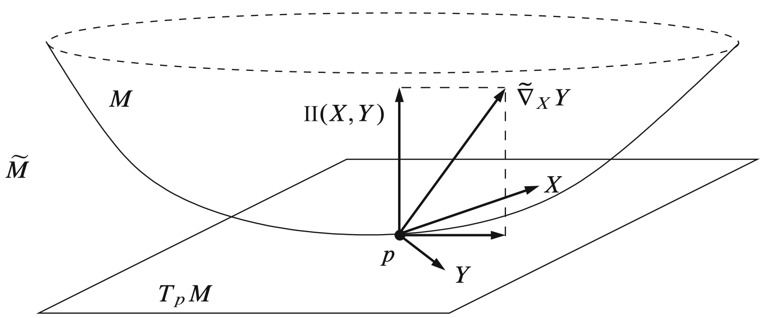

Figure 2.1: The second fundamental form.

Proof. Choose particular extensions of X and Y to a neighborhood of M in $\widetilde{M}$, and for simplicity denote the extended vector fields also by X and Y. We begin by proving that $\mathrm{II}(X, Y)$ is symmetric in X and Y when defined in terms of these extensions. The symmetry of the connection $\widetilde{\nabla}$ implies

$$\mathrm{II}(X, Y) - \mathrm{II}(Y, X) = (\widetilde{\nabla}_X Y)^\perp - (\widetilde{\nabla}_Y X)^\perp = [X, Y]^\perp.$$

Since X and Y are tangent to M at all points of M, so is their Lie bracket (Corollary Lie bracket tangent **??**). Therefore $[X, Y]^\perp = 0$, so II is symmetric.

Because $\widetilde{\nabla}_X Y$ depends only on X_p, it follows that the value of $\mathrm{II}(X, Y)$ at p depends only on X_p, and in particular is independent of the extension chosen for X. Because $\widetilde{\nabla}_X Y$ is linear over $C^\infty(\widetilde{M})$ in X, and every $f \in C^\infty(M)$ can be extended to a smooth function on a neighborhood of M in $\widetilde{M}$, it follows that $\mathrm{II}(X, Y)$ is linear over $C^\infty(M)$ in X. By symmetry, the same claims hold for Y. □

As a consequence of the preceding proposition, for every $p \in M$ and all vectors $v, w \in T_pM$, it makes sense to interpret $\mathrm{II}(v, w)$ as the value of $\mathrm{II}(V, W)$ at p, where V and W are any vector fields on M such that $V_p = v$ and $W_p = w$, and we will do so from now on without further comment.

We have not yet identified the tangential term in the decomposition of $\widetilde{\nabla}_X Y$. Proposition Levi-Civita Euclidean 1.3.12(b) showed that in the special case of a submanifold of a Euclidean or pseudo-Euclidean space, it is none other than the covariant derivative with respect to the Levi-Civita connection of the induced metric on M. The following theorem shows that the same is true in the general case. Therefore, we can interpret the second fundamental form as a measure of the difference between the intrinsic Levi-Civita connection on M and the ambient Levi-Civita connection on $\widetilde{M}$.

ivita submani **Theorem 2.2.2 (Guass formula).** *Suppose (M, g) is an embedded Riemannian submanifold of a Riemannian or pseudo-Riemannian manifold $(\widetilde{M}, \widetilde{g})$. If $X, Y \in \mathfrak{X}(M)$ are extended arbitrarily to smooth vector fields on a neighborhood of M in $\widetilde{M}$, the following formula holds along M:*

$$\widetilde{\nabla}_X Y = \nabla_X Y + \mathrm{II}(X, Y).$$

Proof. Because of the decomposition (Riemann connection submani decomp 2.1) and the definition of the second fundamental form, it suffices to show that $(\widetilde{\nabla}_X Y)^\top = \nabla_X Y$ at all points of M.

Define a map $\nabla^\top : \mathfrak{X}(M) \to \mathfrak{X}(M) \to \mathfrak{X}(M)$ by

$$\nabla^\top_X Y = (\widetilde{\nabla}_X Y)^\top.$$

where X and Y are extended arbitrarily to an open subset of $\widetilde{M}$. We examined a special case of this construction, in which $\widetilde{g}$ is a Euclidean or pseudo-Euclidean metric, in Example 1.2.7. It follows exactly as in that example that $\nabla^\top$ is a connection on M, and exactly as in the proofs of Propositions 1.3.6 and 1.3.7 that it is symmetric and compatible with g. The uniqueness of the Riemannian connection on M then shows that $\nabla^\top = \nabla$. □

The Gauss formula can also be used to compare intrinsic and extrinsic covariant derivatives along curves. If $\gamma : I \to M$ is a smooth curve and X is a vector field along γ that is everywhere tangent to M, then we can regard X as either a vector field along γ in $\widetilde{M}$ or a vector field along γ in M. We let $\widetilde{D}_t X$ and $D_t X$ denote its covariant derivatives along γ as a curve in $\widetilde{M}$ and as a curve in M, respectively.

The next corollary shows how the two covariant derivatives are related.

Corollary 2.2.3 (The Gauss Formula Along a Curve). *Suppose (M, g) is an embedded Riemannian submanifold of a Riemannian or pseudo-Riemannian manifold $(\widetilde{M}, \widetilde{g})$, and $\gamma : I \to M$ is a smooth curve. If X is a smooth vector field along γ that is everywhere tangent to M, then*

$$\widetilde{D}_t X = D_t X + \mathrm{II}(\gamma', X). \tag{2.2}$$

Proof. For each $t_0 \in I$, we can find an adapted orthonormal frame $(E_1, \ldots, E_m)$ in a neighborhood of $\gamma(t_0)$. (Recall that our default assumption is that $\dim M = m$ and $\dim M = n$.) In terms of this frame, X can be written $X(t) = \sum_{i=1}^n X^i E_i|_{\gamma(t)}$. Applying the product rule and the Gauss formula, and using the fact that each vector field E_i is extendible, we get

$$\begin{aligned}
\widetilde{D}_t X &= \sum_{i=1}^n (\dot{X}^i E_i + X^i \widetilde{\nabla}_{\gamma'} E_i) \\
&= \sum_{i=1}^n (\dot{X}^i E_i + X^i \nabla_{\gamma'} E_i + X^i \mathrm{II}(\gamma', E_i)) \\
&= D_t X + \mathrm{II}(\gamma', X).
\end{aligned}$$

This gives the claim. □

Although the second fundamental form is defined in terms of covariant derivatives of vector fields tangent to M, it can also be used to evaluate extrinsic covariant derivatives of normal vector fields, as the following proposition shows. To express it concisely, we introduce one more notation. For each normal vector field $N \in NM$, we obtain a scalar-valued symmetric bilinear form $\mathrm{II}_N : \mathfrak{X}(M) \times \mathfrak{X}(M) \to C^\infty(M)$ by

$$\mathrm{II}_N(X, Y) = \langle \mathrm{II}(X, Y), N \rangle.$$

Let $W_N : \mathfrak{X}(M) \to \mathfrak{X}(M)$ denote the self-adjoint linear map associated with this bilinear form, characterized by

$$\langle W_N(X), Y \rangle = \mathrm{II}_N(X, Y) = \langle \mathrm{II}(X, Y), N \rangle. \tag{2.3}$$

The map W_N is called the **Weingarten map in the direction of N**. Because the second fundamental form is bilinear over $C^\infty(M)$, it follows that W_N is linear over $C^\infty(M)$ and thus defines a smooth bundle homomorphism from TM to itself.

Proposition 2.2.4 (The Weingarten Equation). *Suppose (M, g) is an embedded Riemannian submanifold of a Riemannian or pseudo-Riemannian manifold $(\widetilde{M}, \widetilde{g})$. For every $X \in \mathfrak{X}(M)$ and*

$N \in NM$, the following equation holds:

$$W_N(X) = -(\widetilde{\nabla}_X N)^\top. \tag{2.4}$$

when N is extended arbitrarily to an open subset of $\widetilde{M}$.

Proof. Note that at points of M, the covariant derivative $\widetilde{\nabla}_X N$ is independent of the choice of extensions of X and N by Proposition 1.2.21. Let $Y \in \mathfrak{X}(M)$ be arbitrary, extended to a vector field on an open subset of $\widetilde{M}$. Since $\langle Y, N\rangle$ vanishes identically along M and X is tangent to M, the following holds at points of M:

$$\begin{aligned}
0 &= X\langle N, Y\rangle \\
&= \langle \widetilde{\nabla}_X N, Y\rangle + \langle N, \widetilde{\nabla}_X Y\rangle \\
&= \langle \widetilde{\nabla}_X N, Y\rangle + \langle N, \nabla_X Y + \mathrm{II}(X,Y)\rangle \\
&= \langle \widetilde{\nabla}_X N, Y\rangle + \langle N, \mathrm{II}_N(X,Y)\rangle \\
&= \langle \widetilde{\nabla}_X N, Y\rangle + \langle W_N(X), Y\rangle.
\end{aligned}$$

Since Y was an arbitrary vector field tangent to M, this implies

$$0 = (\widetilde{\nabla}_X N + W_N(X))^\top = (\widetilde{\nabla}_X N)^\top + W_N(X).$$

which is equivalent to (2.4). □

In addition to describing the difference between the intrinsic and extrinsic connections, the second fundamental form plays an even more important role in describing the difference between the curvature tensors of $\widetilde{M}$ and M. The explicit formula, also due to Gauss, is given in the following theorem.

Theorem 2.2.5 (The Gauss Equation). *Suppose (M, g) is an embedded Riemannian submanifold of a Riemannian or pseudo-Riemannian manifold $(\widetilde{M}, \widetilde{g})$. For all $X, Y, V, W \in \mathfrak{X}(M)$, the following equation holds:*

$$R(X,Y,V,W) = \langle \mathrm{II}(X,W), \mathrm{II}(Y,V)\rangle - \langle \mathrm{II}(X,V), \mathrm{II}(Y,W)\rangle + \widetilde{R}(X,Y,V,W). \tag{2.5}$$

Proof. Let X, Y, V, W be extended arbitrarily to an open subset of $\widetilde{M}$. At points of M, using the definition of the curvature and the Gauss formula, we get

$$\begin{aligned}
\langle \widetilde{\nabla}_X \widetilde{\nabla}_Y V, W\rangle &= \langle \widetilde{\nabla}_X(\nabla_Y V + \mathrm{II}(Y,V)), W\rangle \\
&= \langle \widetilde{\nabla}_X \nabla_Y V, W\rangle + \langle \widetilde{\nabla}_X \mathrm{II}(Y,V), W\rangle \\
&= \langle \nabla_X \nabla_Y V, W\rangle + \langle \mathrm{II}(X, \nabla_Y V), W\rangle + \widetilde{\nabla}_X\langle \mathrm{II}(Y,V), W\rangle - \langle \mathrm{II}(Y,V), \widetilde{\nabla}_X W\rangle \\
&= \langle \nabla_X \nabla_Y V, W\rangle - \langle \mathrm{II}(Y,V), \widetilde{\nabla}_X W\rangle \\
&= \langle \nabla_X \nabla_Y V, W\rangle - \langle \mathrm{II}(Y,V), \nabla_X W\rangle - \langle \mathrm{II}(Y,V), \mathrm{II}(X,W)\rangle \\
&= \langle \nabla_X \nabla_Y V, W\rangle - \langle \mathrm{II}(Y,V), \mathrm{II}(X,W)\rangle,
\end{aligned}$$

and

$$\langle \widetilde{\nabla}_{[X,Y]} V, W\rangle = \langle \nabla_{[X,Y]} V, W\rangle + \langle \mathrm{II}([X,Y], V), W\rangle = \langle \nabla_{[X,Y]} V, W\rangle.$$

These together give

$$\begin{aligned}\widetilde{R}(X,Y,V,W) &= \langle \nabla_X\nabla_Y V - \nabla_Y\nabla_X V - \nabla_{[X,Y]}V, W\rangle - \langle \mathrm{II}(Y,V), \mathrm{II}(X,W)\rangle \\ &\quad + \langle \mathrm{II}(X,V), \mathrm{II}(Y,W)\rangle \\ &= R(X,Y,V,W) - \langle \mathrm{II}(Y,V), \mathrm{II}(X,W)\rangle + \langle \mathrm{II}(X,V), \mathrm{II}(Y,W)\rangle\end{aligned}$$

Riemann submani Guass equation

which is equivalent to (2.5). □

There is one other fundamental submanifold equation, which relates the normal part of the ambient curvature endomorphism to derivatives of the second fundamental form. We will not have need for it, but we include it here for completeness. To state it, we need to introduce a connection on the normal bundle of a Riemannian submanifold.

If (M,g) is a Riemannian submanifold of a Riemannian or pseudo-Riemannian manifold $(\widetilde{M},\widetilde{g})$, the **normal connection** $\nabla^\perp : \mathfrak{X}(M)\times\Gamma(NM)\to\Gamma(NM)$ is defined by

$$\nabla^\perp_X N = (\widetilde{\nabla}_X N)^\perp.$$

where N is extended to a smooth vector field on a neighborhood of M in $\widetilde{M}$.

Proposition 2.2.6. *If (M,g) is an embedded Riemannian submanifold of a Riemannian or pseudo-Riemannian manifold $(\widetilde{M},\widetilde{g})$, then $\nabla^\perp$ is a well-defined connection in NM, which is compatible with $\widetilde{g}$ in the sense that for any two sections N_1, N_2 of NM and every $X\in\mathfrak{X}(M)$, we have*

$$X\langle N_1,N_2\rangle = \langle \nabla^\perp_X N_1, N_2\rangle + \langle N_1, \nabla^\perp_X N_2\rangle.$$

Proof. The map $\pi^\perp : T\widetilde{M}|_M \to NM$ is a smooth bundle homomorphism, so it is clear that $\nabla^\perp$ satisfies the conditions of connections since ∇ does. The metric compactibility also follows from this. □

We need the normal connection primarily to make sense of tangential covariant derivatives of the second fundamental form. To do so, we make the following definitions. Let $E\to M$ denote the bundle whose fiber at each point $p\in M$ is the set of bilinear maps $T_pM\times T_pM\to N_pM$. It is easy to check that E is a smooth vector bundle over M, and that smooth sections of E correspond to smooth maps $\mathfrak{X}(M)\times\mathfrak{X}(M)\to\Gamma(NM)$ that are bilinear over $C^\infty(M)$, such as the second fundamental form. Define a connection ∇^E in E as follows: if B is any smooth section of E, let $\nabla^E_X B$ be the smooth section of E defined by

$$(\nabla^E_X B)(Y,Z) = \nabla^\perp_X(B(Y,Z)) - B(\nabla_X Y, Z) - B(Y,\nabla_X Z).$$

Now we are ready to state the last of the fundamental equations for submanifolds.

Theorem 2.2.7 (The Codazzi Equation). *Suppose (M,g) is an embedded Riemannian submanifold of a Riemannian or pseudo-Riemannian manifold $(\widetilde{M},\widetilde{g})$. For all $X,Y,Z\in\mathfrak{X}(M)$ the following equation holds:*

$$(\widetilde{R}(X,Y)Z)^\perp = (\nabla^E_X\mathrm{II})(Y,Z) - (\nabla^E_Y\mathrm{II})(X,Z). \tag{2.6}$$

Riemann Coda

Riemann Codazzi equation-1

Proof. It suffices to show that both sides of (2.6) give the same result when we take their inner products with an arbitrary smooth normal vector field N along M:

$$\langle \widetilde{R}(X,Y)Z, N\rangle = \langle (\nabla^E_X)(Y,Z), N\rangle - \langle (\nabla^E_Y)(X,Z), N\rangle \tag{2.7}$$

Riemann Coda

We begin as in the proof of the Gauss equation: after extending the vector fields to a neighborhood of M and applying the Gauss formula, we obtain

$$\widetilde{R}(X,Y,Z,N) = \langle \widetilde{\nabla}_X(\nabla_Y Z + \mathrm{II}(Y,Z)) - \widetilde{\nabla}_Y(\nabla_X Z + \mathrm{II}(X,Z)) - \widetilde{\nabla}_{[X,Y]}Z, N\rangle$$

Now when we expand the covariant derivatives, we need only pay attention to the normal components. This yields

$$\begin{aligned}\widetilde{R}(X,Y,Z,N) &= \langle \mathrm{II}(X,\nabla_Y Z) + (\nabla^E_X \mathrm{II})(Y,Z) + \mathrm{II}(\nabla_X Y, Z) + \mathrm{II}(Y,\nabla_X Z), N\rangle \\ &\quad - \langle \mathrm{II}(Y,\nabla_X Z) + (\nabla^E_Y \mathrm{II})(X,Z) + \mathrm{II}(\nabla_Y X, Z) + \mathrm{II}(X,\nabla_Y Z), N\rangle \\ &\quad - \langle \mathrm{II}([X,Y],Z), N\rangle \\ &= \langle (\nabla^E_X)(Y,Z), N\rangle - \langle (\nabla^E_Y)(X,Z), N\rangle + \langle \nabla_X Y - \nabla_Y X - [X,Y], N\rangle \\ &= \langle (\nabla^E_X)(Y,Z), N\rangle - \langle (\nabla^E_Y)(X,Z), N\rangle.\end{aligned}$$

This finishes the proof. □

Curvature of a curve

By studying the curvatures of curves, we can give a more geometric interpretation of the second fundamental form. Suppose (M,g) is a Riemannian or pseudo-Riemannian manifold, and $\gamma : I \to M$ is a smooth unit-speed curve in M. We define the **geodesic curvature** of γ as the length of the acceleration vector field, which is the function $\kappa : I \to \mathbb{R}$ given by

$$\kappa(t) = |D_t\gamma'(t)|.$$

If γ is an arbitrary regular curve in a Riemannian manifold (not necessarily of unit speed), we first find a unit-speed reparametrization $\widetilde{\gamma} = \gamma \circ \varphi$, and then define the curvature of γ at t to be the curvature of $\widetilde{\gamma}$ at $\varphi^{-1}(t)$. In a pseudo-Riemannian manifold, the same approach works, but we have to restrict the definition to curves γ such that $|\gamma'(t)|$ is everywhere nonzero.

From the definition, it follows that a smooth unit-speed curve has vanishing geodesic curvature if and only if it is a geodesic, so the geodesic curvature of a curve can be regarded as a quantitative measure of how far it deviates from being a geodesic. If $M = \mathbb{R}^n$ with the Euclidean metric, the geodesic curvature agrees with the notion of curvature introduced in advanced calculus courses.

Now suppose $(\widetilde{M},\widetilde{g})$ is a Riemannian or pseudo-Riemannian manifold and (M,g) is a Riemannian submanifold. Every regular curve $\gamma : I \to M$ has two distinct geodesic curvatures: its **intrinsic curvature** κ as a curve in M, and its **extrinsic curvature** $\widetilde{\kappa}$ as a curve in $\widetilde{M}$. The second fundamental form can be used to compute the relationship between the two.

Proposition 2.2.8. *Suppose (M,g) is an embedded Riemannian submanifold of a Riemannian or pseudo-Riemannian manifold $(\widetilde{M},\widetilde{g})$, $p \in M$, and $v \in T_pM$.*

(a) $\mathrm{II}(v,v)$ *is the $\widetilde{g}$-acceleration at p of the g-geodesic γ_v.*

(b) *If v is a unit vector, then $|\mathrm{II}(v,v)|$ is the $\widetilde{g}$-curvature of γ_v at p.*

Proof. Suppose $\gamma : (-\varepsilon,\varepsilon) \to M$ is any regular curve with $\gamma(0) = p$ and $\gamma'(0) = v$. Applying the Gauss formula (Corollary 2.2) to the vector field γ' along γ, we obtain

$$\widetilde{D}_t\gamma' = D_t\gamma + \mathrm{II}(\gamma',\gamma').$$

If γ is a g-geodesic in M, this formula simplifies to

$$\widetilde{D}_t \gamma' = \mathrm{II}(\gamma', \gamma').$$

Both conclusions of the proposition follow from this. $\square$

Note that the second fundamental form is completely determined by its values of the form $\mathrm{II}(v, v)$ for unit vectors v, by the following lemma.

ic form lemma **Lemma 2.2.9.** *Suppose V is an inner product space, W is a vector space, and $B, B' : V \times V \to W$ are symmetric and bilinear. If $B(v, v) = B'(v, v)$ for every unit vector $v \in V$, then $B = B'$.*

Proof. Every vector $v \in V$ can be written $v = |v|\hat{v}$ for some unit vector $\hat{v}$, so the bilinearity of B and B' implies $B(v, v) = B'(v, v)$ for every v, not just unit vectors. The result then follows from the following polarization identity for inner products:

$$B(v, w) = \frac{1}{4}\big(B(v + w, v + w) - B(v - w, v - w)\big).$$

$\square$

Because the intrinsic and extrinsic accelerations of a curve are usually different, it is generally not the case that a $\widetilde{g}$-geodesic that starts tangent to M stays in M; just think of a sphere in Euclidean space, for example. A Riemannian submanifold (M, g) of $(\widetilde{M}, \widetilde{g})$ is said to be **totally geodesic** if every $\widetilde{g}$-geodesic that is tangent to M at some time t_0 stays in M for all t in some interval $(t_0 - \varepsilon, t_0 + \varepsilon)$.

Proposition 2.2.10. *Suppose (M, g) is an embedded Riemannian submanifold of a Riemannian or pseudo-Riemannian manifold $(\widetilde{M}, \widetilde{g})$, The following are equivalent:*

(a) M is totally geodesic in $\widetilde{M}$.

(b) Every g-geodesic in M is also a $\widetilde{g}$-geodesic in $\widetilde{M}$.

(c) The second fundamental form of M vanishes identically.

(d) If $\gamma : I \to M$ is a curve in M, $t_0 \in I$ and $v \in T_{\gamma(t_0)}M$. Then the parallel transport of v along γ is the same for M and for $\widetilde{M}$.

Proof. We will prove $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a)$. First assume that M is totally geodesic. Let $\gamma : I \to M$ be a g-geodesic. For each $t_0 \in I$, let $\widetilde{\gamma} : \widetilde{I} \to \widetilde{M}$ be the $\widetilde{g}$-geodesic with $\widetilde{\gamma}(t_0) = \gamma(t_0)$ and $\widetilde{\gamma}'(t_0) = \gamma'(t_0)$. The hypothesis implies that there is some open interval I_0 containing t_0 Riemann connection submani curve-1 such that $\widetilde{\gamma}(t) \in M$ for $t \in I_0$. On I_0, the Gauss formula (2.2) for $\widetilde{\gamma}'$ reads

$$0 = \widetilde{D}_t \widetilde{\gamma}' = D_t \widetilde{\gamma}' + \mathrm{II}(\widetilde{\gamma}', \widetilde{\gamma}').$$

Because the first term on the right is tangent to M and the second is normal, the two terms must vanish individually. In particular, $D_t \widetilde{\gamma}' \equiv 0$ on I_0, which means that $\widetilde{\gamma}$ is also a g-geodesic there. By uniqueness of geodesics, therefore, $\gamma = \widetilde{\gamma}$ on I_0, so it follows in turn that γ is a $\widetilde{g}$-geodesic there. Since the same is true in a neighborhood of every $t_0 \in I$, it follows that γ is a $\widetilde{g}$-geodesic on its whole domain.

Next assume that every g-geodesic is a $\widetilde{g}$-geodesic. Let $p \in M$ and $v \in T_pM$ be arbitrary, and let $\gamma : I \to M$ be the g-geodesic with $\gamma(0) = p$ and $\gamma'(0) = v$. The hypothesis implies that γ is also a $\widetilde{g}$-geodesic. Thus $\widetilde{D}_t\gamma' = D_t\gamma' = 0$, so the Gauss formula yields $\mathrm{II}(\gamma',\gamma') = 0$ along γ. In particular, $\mathrm{II}(v,v) = 0$. By Lemma 2.2.9, this implies that II is identically zero.

Now assume that $\mathrm{II} = 0$, and let $\widetilde{\gamma} : \widetilde{I} \to \widetilde{M}$ be a $\widetilde{g}$-geodesic such that $\widetilde{\gamma}(t_0) \in M$ and $\widetilde{\gamma}'(t_0) \in TM$ for some $t_0 \in I$. Let $\gamma : I \to M$ be the g-geodesic with the same initial conditions: $\gamma(t_0) = \widetilde{\gamma}(t_0)$ and $\gamma'(t_0) = \widetilde{\gamma}'(t_0)$. The Gauss formula together with the hypothesis $\mathrm{II} = 0$ implies that $\widetilde{D}_t\gamma' = D_t\gamma' = 0$, so γ is also a $\widetilde{g}$-geodesic. By uniqueness of geodesics, therefore, $\widetilde{\gamma} = \gamma$ on the intersection of their domains, which implies that $\widetilde{\gamma}$ lies in M for t in some open interval around t_0.

Finally, recall that the parallel transport is defined by the unique vector field V along γ such that $D_tV \equiv 0$ and $V(t_0) = v$. In view of the Guass formula (2.2), it is clear that $(c) \Rightarrow (d)$ and $(d) \Rightarrow (b)$. Therefore we are done. $\square$

2.2.2 Hypersurfaces

Now we specialize the preceding considerations to the case in which M is a hypersurface (i.e., a submanifold of codimension 1) in $\widetilde{M}$. Throughout this part, our default assumption is that (M, g) is an embedded n-dimensional Riemannian submanifold of an $n+1$-dimensional Riemannian manifold $(\widetilde{M}, \widetilde{g})$. In this situation, at each point of M there are exactly two unit normal vectors. In terms of any local adapted orthonormal frame $(E_1, \ldots, E_{n+1})$, the two choices are E_{n+1}. In a small enough neighborhood of each point of M, therefore, we can always choose a smooth unit normal vector field along M.

If both M and $\widetilde{M}$ are orientable, we can use an orientation to pick out a global smooth unit normal vector field along all of M. In general, though, this might or might not be possible. Since all of our computations in this chapter are local, we will always assume that we are working in a small enough neighborhood that a smooth unit normal field exists. We will address as we go along the question of how various quantities depend on the choice of normal vector field.

The scalar second fundamental form and the shape operator

Having chosen a distinguished smooth unit normal vector field N on the hypersurface $M \subseteq \widetilde{M}$, we can replace the vector-valued second fundamental form II by a simpler scalar-valued form. The **scalar second fundamental form** of M is the symmetric covariant 2-tensor field $h \in \Gamma(\Sigma^2T^*M)$ defined by $h = \mathrm{II}_N$; in other words,

$$h(X,Y) = \langle \mathrm{II}(X,Y), N\rangle.$$

Using the Gauss formula $\widetilde{\nabla}_XY = \nabla_XY + \mathrm{II}(X,Y)$ and noting that ∇_XY is orthogonal to N, we can rewrite the definition as

$$h(X,Y) = \langle \widetilde{\nabla}_XY, N\rangle.$$

Also, since N is a unit vector spanning NM at each point, the definition of h is equivalent to

$$\mathrm{II}(X,Y) = h(X,Y)N. \tag{2.8}$$

Note that replacing N by $-N$ multiplies h by -1, so the sign of h depends on which unit normal is chosen; but h is otherwise independent of the choices.

The choice of unit normal field also determines a Weingarten map $W_N : \mathfrak{X}(M) \to \mathfrak{X}(M)$ by (2.3); in the case of a hypersurface, we use the notation $s = W_N$ and call it the **shape**

operator of M. Alternatively, we can think of s as the $(1,1)$-tensor field on M obtained from h by raising an index. It is characterized by

$$\langle s(X), Y\rangle = h(X,Y) \text{ for } X, Y \in \mathfrak{X}(M).$$

Because h is symmetric, s is a self-adjoint endomorphism of TM, that is,

$$\langle s(X), Y\rangle = \langle X, s(Y)\rangle \text{ for } X, Y \in \mathfrak{X}(M).$$

As with h, the sign of s depends on the choice of N.

Theorem 2.2.11 (Fundamental Equations for a Hypersurface). *Suppose (M,g) is a Riemannian hypersurface in a Riemannian manifold $(\widetilde{M},\widetilde{g})$, and N is a smooth unit normal vector field along M.*

(a) The Guass formula for a hypersurface: If $X, Y \in \mathfrak{X}(M)$ are extended to an open subset of $\widetilde{M}$, then

$$\widetilde{\nabla}_X Y = \nabla_X Y + h(X,Y)N.$$

(b) The Guass formula for a curve in a hypersurface: If $\gamma : I \to M$ is a smooth curve and $X : I \to TM$ is a smooth vector field along γ, then

$$\widetilde{D}_t X = D_t X + h(\gamma', X)N.$$

(c) The Weingarten equation for a hypersurface: For every $X \in \mathfrak{X}(M)$,

$$sX = -\widetilde{\nabla}_X N. \tag{2.9}$$

Riemann Weiga

(d) The Guass equation for a hypersurface: For all $X, Y, V, W \in \mathfrak{X}(M)$,

$$R(X,Y,V,W) = \widetilde{R}(X,Y,V,W) + \frac{1}{2}(h \owedge h)(X,Y,V,W). \tag{2.10}$$

Riemann Guass

(e) The Codazzi equation for a hypersurface: For all $X, Y, V, W \in \mathfrak{X}(M)$,

$$\widetilde{R}(X,Y,Z,N) = (Dh)(Z,X,Y). \tag{2.11}$$

Riemann Codaz

Riemann shapre operator hypersurface

Proof. Parts (a), (b), and (d) follow immediately from substituting (2.8) into the general versions of the Gauss formula and Gauss equation. To prove (c), note first that the general version of theWeingarten equation can be written $(\widetilde{\nabla}_X N)^\top = -sX$. Since $\langle \widetilde{\nabla}_X N, N\rangle = X(|N|^2)/2 = 0$, it follows that $\nabla_X N$ is tangent to M, so (c) follows.

To prove the hypersurface Codazzi equation, note that the fact that N is a unit vector field implies

$$0 = X|N|^2 = 2\langle \nabla_X^\perp N, N\rangle.$$

Since N spans the normal bundle, this implies that N is parallel with respect to the normal connection. Moreover

$$\begin{aligned}
(\nabla_X^E \mathrm{II})(Y,Z) &= \nabla_X^\perp Y(\mathrm{II}(Y,Z)) - \mathrm{II}(\nabla_X Y, Z) - \mathrm{II}(Y, \nabla_X Z) \\
&= \nabla_X^\perp Y(h(Y,Z)N) - \mathrm{II}(\nabla_X Y, Z) - \mathrm{II}(Y, \nabla_X Z) \\
&= \big(X(h(X,Y)) - h(\nabla_X Y, Z) - h(Y, \nabla_X Z)\big)N \\
&= (\nabla_X h)(Y,Z)N.
\end{aligned}$$

Riemann Codazzi equation-1

Inserting this into the general form (2.6) of the Codazzi equation and using the fact that ∇h

is symmetric in its first two indices yields

$$\begin{aligned}\widetilde{R}(X,Y,Z,N) &= (\nabla_X h)(Y,Z) - (\nabla_Y h)(X,Z)\\ &= (\nabla h)(Y,Z,X) - (\nabla h)(X,Z,Y)\\ &= -(\nabla h)(Z,X,Y) + (\nabla h)(Z,Y,X).\end{aligned}$$

which is equivalent to (2.11). □

Principal curvatures

At every point $p \in M$, we have seen that the shape operator s is a self-adjoint linear endomorphism of the tangent space T_pM. To analyze such an operator, we recall some linear-algebraic facts about self-adjoint endomorphisms.

Lemma 2.2.12. *Suppose V is a finite-dimensional inner product space and $s : V \to V$ is a self-adjoint linear endomorphism. Let S denote the set of unit vectors in V. There is a vector $v_0 \in S$ where the function $v \mapsto \langle sv, v\rangle$ achieves its maximum among elements of S, and every such vector is an eigenvector of s with eigenvalue $\lambda_0 = \langle sv_0, v_0\rangle$.*

Proposition 2.2.13. *Suppose V is a finite-dimensional inner product space and $s : V \to V$ is a self-adjoint linear endomorphism. Then V has an orthonormal basis of s-eigenvectors, and all of the eigenvalues are real.*

Applying this proposition to the shape operator $s : T_pM \to T_pM$, we see that s has real eigenvalues $\kappa_1, \dots, \kappa_n$, and there is an orthonormal basis $(e_1, \dots, e_n)$ for T_pM consisting of s-eigenvectors, with $se_i = \lambda_i e_i$ for each i (no summation). In this basis, both h and s are represented by diagonal matrices, and h has the expression

$$h(v,w) = \kappa_1 v^1 w^1 + \cdots + \kappa_n v^n w^n.$$

The eigenvalues of s at a point $p \in M$ are called the principal curvatures of M at p, and the corresponding eigenspaces are called the **principal directions**. The principal curvatures all change sign if we reverse the normal vector, but the principal directions and principal curvatures are otherwise independent of the choice of coordinates or bases.

There are two combinations of the principal curvatures that play particularly important roles for hypersurfaces. The **Gaussian curvature** is defined as $K = \det(s)$, and the **mean curvature** as $H = \mathrm{tr}(s)/n = \mathrm{tr}_g(h)/n$. Since the determinant and trace of a linear endomorphism are basis-independent, these are well defined once a unit normal is chosen. In terms of the principal curvatures, they are

$$K = \kappa_1 \cdots \kappa_n, \quad H = \frac{1}{n}(\kappa_1 + \cdots + \kappa_n).$$

as can be seen by expressing s in terms of an orthonormal basis of eigenvectors. If N is replaced by $-N$, then H changes sign, while K is multiplied by $(-1)^n$.

Computations in semigeodesic coordinates

Semigeodesic coordinates (Proposition 1.4.38) provide an extremely convenient tool for computing the invariants of hypersurfaces.

Let $(\widetilde{M}, \widetilde{g})$ be an $(n+1)$-dimensional Riemannian manifold, and let $(x^1, \dots, x^n, v)$ be semigeodesic coordinates on an open subset $U \subseteq \widetilde{M}$. (For example, they might be Fermi coordinates for the hypersurface $M = v^{-1}(0)$.) For each real number a such that $v^{-1}(a) \neq \varnothing$, the level set $M_a = v^{-1}(a)$ is a properly embedded hypersurface in U. Let g_a denote the induced metric on M_a. Corollary 1.4.39 shows that $\widetilde{g}$ is given by

$$\widetilde{g} = dv^2 + g_{\alpha\beta} dx^\alpha dx^\beta. \tag{2.12}$$

The restrictions of $(x^1, \dots, x^n)$ give smooth coordinates for each hypersurface M_a, and in those coordinates the induced metric g_a is given by $g_{\alpha\beta} dx^\alpha dx^\beta$. (Here we use the summation convention with Greek indices running from 1 to n.) The vector field $\partial_v = \partial_{n+1}$ restricts to a unit normal vector field along each hypersurface M_a.

As the next proposition shows, semigeodesic coordinates give us a simple formula for the second fundamental forms of all of the submanifolds M_a at once.

Proposition 2.2.14. *With notation as above, the components in $(x^1, \dots, x^n)$-coordinates of the scalar second fundamental form, the shape operator, and the mean curvature of (M_a, g_a) (denoted by h_a, s_a and H_a, respectively) with respect to the normal $N = \partial_v$ are given by*

$$(h_a)_{\alpha\beta} = -\frac{1}{2} \partial_v g_{\alpha\beta}|_{v=a}, \quad (s_a)^\alpha_\beta = -\frac{1}{2} g^{\alpha\gamma} \partial_v g_{\gamma\beta}|_{v=a}, \quad H = -\frac{1}{2n} g^{\alpha\beta} \partial_v g_{\alpha\beta}|_{v=a}.$$

Proof. The normal component of $\widetilde{\nabla}_{\partial_\alpha} \partial_\beta$ is $\widetilde{\Gamma}^{n+1}_{\alpha\beta} \partial|_v$, which Corollary 1.4.39 shows is equal to $-1/2 \partial_v g_{\alpha\beta} \partial_r$ (noting that the roles of g and $\widehat{g}$ in that corollary are being played here by $\widetilde{g}$ and g_a, respectively). Equation (4.17) evaluated at points of M_a gives

$$(h_a)_{\alpha\beta} = \langle \widetilde{\nabla}_{\partial_\alpha} \partial_\beta, N \rangle = \langle -\frac{1}{2} \partial_v g_{\alpha\beta} \partial_v, \partial_v \rangle = -\frac{1}{2} \partial_v g_{\alpha\beta}.$$

The formulas for s_a and H_a follow by using $(g^{\alpha\gamma})$ (the inverse matrix of $(g_{\alpha\gamma})$) to raise an index and then taking the trace. □

Minimal hypersurfaces

A natural question that has received a great deal of attention over the past century is this: Given a simple closed curve C in $\mathbb{R}^3$, is there an embedded or immersed surface M with $\partial M = C$ that has least area among all surfaces with the same boundary? If so, what is it? Such surfaces are models of the soap films that are produced when a closed loop of wire is dipped in soapy water.

More generally, we can consider the analogous question for hypersurfaces in Riemannian manifolds. Suppose M is a compact codimension-1 submanifold with nonempty boundary in an $n+1$-dimensional Riemannian manifold $(\widetilde{M}, \widetilde{g})$. By analogy with the case of surfaces in $\mathbb{R}^3$, it is traditional to use the term area to refer to the n-dimensional volume of M with its induced Riemannian metric, and to say that M is **area-minimizing** if it has the smallest area among all compact embedded hypersurfaces in $\widetilde{M}$ with the same boundary. One key observation is the following theorem, which is an analogue for hypersurfaces of Theorem 1.4.4 about length-minimizing curves.

Theorem 2.2.15. *Let M be a compact codimension-1 submanifold with nonempty boundary in an $(n+1)$-dimensional Riemannian manifold $(\widetilde{M}, \widetilde{g})$. If M is area minimizing, then its mean curvature is identically zero.*

Proof. Let g denote the induced metric on M. The fact that M minimizes area among hypersurfaces with the same boundary means, in particular, that it minimizes area among small perturbations of M in a neighborhood of a single point. We will exploit this idea to prove that M must have zero mean curvature everywhere.

Let $p \in \operatorname{Int} M$ be arbitrary, let $(x^1, \dots, x^n, v)$ be Fermi coordinates for M on an open set $\widetilde{U} \subseteq \widetilde{M}$ containing p, and let $U = \widetilde{U} \cap M$. By taking U sufficiently small, we can arrange that U is a regular coordinate ball in M and $\widetilde{U} \cap \partial M = \varnothing$. We use $(x^1, \dots, x^n)$ as coordinates on M, and observe the summation convention with Greek indices running from 1 to n.

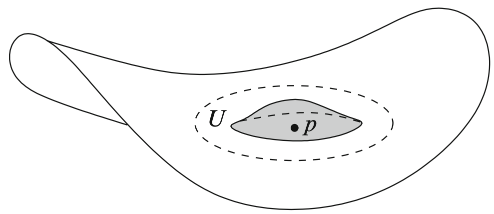

Figure 2.2: The hypersurface M_t.

Let φ be an arbitrary smooth real-valued function on M with compact support in U. For sufficiently small t, define a set $M_t \subseteq \widetilde{M}$ by

$$M_t = (M \setminus U) \cup \{z \in \widetilde{U} : v(z) = t\varphi(x^1(z), \dots, x^n(z))\}.$$

Then M_t is an embedded smooth hypersurface in $\widetilde{M}$, which agrees with M outside of U and which coincides with the graph of $v = t\varphi$ in $\widetilde{U}$. Let $f_t : U \to \widetilde{U}$ be the graph parametrization of $M_t \cap \widetilde{U}$, given in Fermi coordinates by

$$f_t(x^1, \dots, x^n) = (x^1, \dots, x^n, t\varphi(x)). \tag{2.13}$$

Riemann minim

Using this map, for each t we can define a diffeomorphism $F_t : M \to M_t$ by

$$F_t(z) = \begin{cases} z, & z \in M \setminus \operatorname{supp}(\varphi), \\ f_t(z), & z \in U. \end{cases}$$

For each t, let $\widehat{g}_t = \iota_{M_t}^* \widetilde{g}$ denote the induced Riemannian metric on M_t, and let $g_t = F_t^* \widehat{g}_t = F^* \widetilde{g}$ denote the pulled-back metric on M. When $t = 0$, we have $M_0 = M$, and both g_0 and $\widetilde{g}_0$ are equal to the induced metric g on M. Since $\widetilde{g}$ is given by (2.12) in Fermi coordinates, a simple computation shows that in U, $g_t = F_t^* \widetilde{g}$ has the coordinate expression $g_t = (g_t)_{\alpha\beta} dx^\alpha dx^\beta$, where

Riemann hypersurface semigeodesic metric

$$(g_t)_{\alpha\beta} = t^2 \frac{\partial \varphi}{\partial x^\alpha}(x) \frac{\partial \varphi}{\partial x^\beta}(x) + g_{\alpha\beta}(x, t\varphi(x)). \tag{2.14}$$

Riemann minim

while on $M \setminus U$, g_t is equal to g and thus is independent of t.

Since each M_t is a smooth hypersurface with the same boundary as M, our hypothesis guarantees that $\operatorname{Area}(M_t, \widehat{g}_t)$ achieves a minimum at $t = 0$. Because F_t is an isometry from (M, g_t) to $(M_t, \widehat{g}_t)$, we can express this area as follows:

$$\operatorname{Area}(M_t, \widehat{g}_t) = \operatorname{Area}(M, g_t) = \operatorname{Area}(M \setminus U, g) + \operatorname{Area}(U, g_t).$$

The first term on the right is independent of t, and we can compute the second term explicitly in coordinates $(x^1, \dots, x^n)$ on U:

$$\operatorname{Area}(U, g_t) = \int_U \sqrt{\det g_t}\, dx^1 \cdots dx^n.$$

Riemann minimal surface-2

where $\det g_t$ denotes the determinant of the matrix g_t defined by (2.14). The integrand above is a smooth function of t and $(x^1, \dots, x^n)$, so the area is a smooth function of t. We have

$$\begin{aligned} \frac{d}{dt}\bigg|_{t=0} \operatorname{Area}(M_t, \widehat{g}_t) &= \int_U \frac{\partial}{\partial t}\bigg|_{t=0} \sqrt{\det g_t}\, dx^1 \cdots dx^n \\ &= \int_U \frac{1}{2\sqrt{\det g}} \frac{\partial}{\partial t}\bigg|_{t=0} (\det g_t) dx^1 \cdots dx^n. \end{aligned} \tag{2.15}$$

(The differentiation under the integral sign is justified because the integrand is smooth and has compact support in U.) To compute the derivative of the determinant, note that the expansion by minors along, say, row α shows that the partial derivative of $\det$ with respect to the matrix entry in position $\alpha\beta$ is equal to the cofactor $A_{\alpha\beta}$, and thus by the chain rule,

$$\frac{\partial}{\partial t}(\det g_t) = A_{\alpha\beta} \frac{\partial}{\partial t}(g_t)_{\alpha\beta}. \tag{2.16}$$

On the other hand, Cramer's rule shows that the (α, β) component of the inverse matrix is given by $g^{\alpha\beta} = (\det g)^{-1} A_{\beta\alpha}$. Thus from (2.16) and (2.14) we obtain

$$\frac{\partial}{\partial t}\bigg|_{t=0} (\det g_t) = (\det g) g^{\beta\alpha} \frac{\partial}{\partial t}\bigg|_{t=0} (g_t)_{\alpha\beta} = (\det g) g^{\alpha\beta} \frac{\partial g_{\alpha\beta}}{\partial v}\bigg|_{v=0} \varphi.$$

Inserting this into (2.15) and using the result of Proposition 2.2.14, we conclude that

$$\begin{aligned} \frac{d}{dt}\bigg|_{t=0} \operatorname{Area}(M_t, \widehat{g}_t) &= \frac{1}{2} \int_U (g^{\alpha\beta} \partial_v g_{\alpha\beta}|_{v=0}) \varphi \sqrt{\det g}\, dx^1 \cdots dx^n \\ &= -n \int_U H \varphi \, dV_g. \end{aligned} \tag{2.17}$$

where H is the mean curvature of (M, g). Since $\operatorname{Area}(M_t, \widehat{g}_t)$ attains a minimum at $t = 0$, we conclude that $\int_U H\varphi \, dV_g = 0$ for every such φ.

Now suppose for the sake of contradiction that $H(p) \neq 0$. If $H(p) > 0$, we can let φ be a smooth nonnegative bump function that is positive at p and supported in a small neighborhood of p on which $H > 0$. The argument above shows that $\int_U H\varphi \, dV_g = 0$, which is impossible because the integrand is nonnegative on U and positive on an open set. A similar argument rules out $H(p) < 0$. Since p was an arbitrary point in $\operatorname{Int} M$, we conclude that $H \equiv 0$ on $\operatorname{Int} M$, and then by continuity on all of M. $\square$

Because of the result of Theorem 2.2.15, a hypersurface (immersed or embedded, with or without boundary) that has mean curvature identically equal to zero is called a **minimal hypersurface** (or a **minimal surface** when it has dimension 2). It is an unfortunate historical accident that the term minimal hypersurface is defined in this way, because in fact, a minimal hypersurface is just a critical point for the area, not necessarily area-minimizing. It can be shown that as in the case of geodesics, a small enough piece of every minimal hypersurface is area-minimizing.

As a complement to the above theorem about hypersurfaces that minimize area with fixed boundary, we have the following result about hypersurfaces that minimize area while enclosing a fixed volume.

Theorem 2.2.16. *Suppose $(\widetilde{M}, \widetilde{g})$ is a Riemannian $(n+1)$-manifold, $D \subseteq \widetilde{M}$ is a compact regular domain, and $M = \partial D$. If M has the smallest surface area among boundaries of compact regular domains with the same volume as D, then M has constant mean curvature (computed with respect to the outward unit normal).*

Proof. Let g denote the induced metric on M. Assume for the sake of contradiction that the mean curvature H of M is not constant, and let $p, q \in M$ be points such that $H(p) < H(q)$.

Since M is compact, it has an ε-tubular neighborhood for some $\varepsilon > 0$ by Theorem 1.3.25. As in the previous proof, let $(x^1, \dots, x^n, v)$ be Fermi coordinates for M on an open set $\widetilde{U} \subseteq \widetilde{M}$ containing p, and let $U = \widetilde{U} \cap M$. We may assume that U is a regular coordinate ball in M and the image of the chart is a set of the form $\widehat{U} \times (-\varepsilon, \varepsilon)$ for some open subset $\widehat{U} \subseteq \mathbb{R}^n$. Similarly, let $(y^1, \dots, y^n, w)$ be Fermi coordinates for M on an open set $\widetilde{W} \subseteq \widetilde{M}$ containing q and satisfying the analogous conditions, and let $W = \widetilde{W} \cap M$. By replacing v with its negative if necessary, we can arrange that $D \cap \widetilde{U}$ is the set where $v \le 0$, and similarly for w. Also, by shrinking both domains, we can assume that the mean curvature of M satisfies $H \le H_1$ on U and $H \ge H_2$ on W , where H_1, H_2 are constants such that $H(p) < H_1 < H_2 < H(q)$.

Let φ and ψ be positive smooth real-valued functions on M, with φ compactly supported in U and compactly supported in W, and satisfying $\int_U \varphi \, dV_g = \int_W \psi \, dV_g = 1$. For sufficiently small $s, t \in \mathbb{R}$, define a subset $D_{s,t} \subseteq \widetilde{M}$ as follows:

$$\begin{aligned} D_{s,t} =& \{z \in \widetilde{U} : v(z) \le s\varphi(x^1(z), \dots, x^n(z))\} \\ & \cup \{z \in \widetilde{W} : w(z) \le t\psi(y^1(z), \dots, y^n(z))\} \cup (D \setminus (\widetilde{U} \cup \widetilde{W})). \end{aligned}$$

and let $M_{s,t} = \partial D_{s,t}$, so $D_{0,0} = D$ and $M_{0,0} = M$. For sufficiently small s and t, the set $D_{s,t}$ is a regular domain and $M_{s,t}$ is a compact smooth hypersurface, and $\operatorname{Vol}(D_{s,t})$ and $\operatorname{Area}(M_{S,t})$ are both smooth functions of (s,t). For convenience, write $V(s,t) = \operatorname{Vol}(D_{s,t})$ and $A(s,t) = \operatorname{Area}(M_{s,t})$. The same argument that led to (2.17) shows that

$$\frac{\partial A}{\partial s}(0,0) = -n \int_U H\varphi \, dV_g, \quad \frac{\partial A}{\partial t}(0,0) = -n \int_W H\psi \, dV_g.$$

To compute the partial derivatives of the volume, we just note that if we hold $t = 0$ fixed and let s vary, the only change in volume occurs in the part of $D_{s,t}$ contained in $\widetilde{U}$, so the fundamental theorem of calculus gives

$$\begin{aligned} \frac{\partial V}{\partial s}(0,0) &= \frac{d}{ds}\Big|_{s=0} \operatorname{Vol}(D_{s,0} \cap \widetilde{U}) \\ &= \frac{d}{ds}\Big|_{s=0} \int_U \Big(\int_{-\varepsilon}^{s\varphi(x)} \sqrt{\det \widetilde{g}(x,v)} \, dv \Big) dx^1 \cdots dx^n \\ &= \int_U \frac{d}{ds}\Big|_{s=0} \Big(\int_{-\varepsilon}^{s\varphi(x)} \sqrt{\det \widetilde{g}(x,v)} \, dv \Big) dx^1 \cdots dx^n \\ &= \int_U \varphi(x) \sqrt{\det \widetilde{g}(x,0)} \, dx^1 \cdots dx^n = \int_U \varphi \, dV_g = 1. \end{aligned}$$

where the differentiation under the integral sign in the third line is justified just like (2.17), and in the last line we used the fact that $g_{\alpha\beta}(x) = \widetilde{g}_{\alpha\beta}(x,0)$ in these coordinates. Similarly, $\partial V / \partial t(0,0) = 1$.

Because $V(0,0) = \operatorname{Vol}(D)$ and $\partial V/\partial t(0,0) \ne 0$, the implicit function theorem guarantees that there is a smooth function $\lambda : (-\delta, \delta) \to \mathbb{R}$ for some $\delta > 0$ such that $V(s, \lambda(s)) \equiv \operatorname{Vol}(D)$. The chain rule then implies

$$0 = \frac{d}{ds}\Big|_{s=0} V(s, \lambda(s)) = \frac{\partial V}{\partial s}(0,0) + \lambda'(0)\frac{\partial V}{\partial t}(0,0) = 1 + \lambda'(0).$$

Thus $\lambda'(0) = -1$.

Our hypothesis that M minimizes area implies that

$$0 = \frac{d}{ds}A(s,\lambda(s)) = \frac{\partial A}{\partial s}(0,0) + \lambda'(0)\frac{\partial A}{\partial t}(0,0) = -n\int_U H\varphi\, dV_g + n\int_W H\psi\, dV_g.$$

and thus $\int_U H\varphi\, dV_g = \int_W H\psi\, dV_g$. But our choice of U and V together with the fact that $\int_U \varphi dV_g = \int_W \psi dV_g = 1$ guarantees that

$$\int_U H\varphi\, dV_g \le H_1 < H_2 \le \int_W H\psi\, dV_g,$$

which is a contradiction. □

2.2.3 Hypersurfaces in Euclidean space

Now we specialize even further, to hypersurfaces in Euclidean space. In this part, we assume that $M \subseteq \mathbb{R}^{n+1}$ is an embedded n-dimensional submanifold with the induced Riemannian metric. The Euclidean metric will be denoted as usual by $\bar{g}$, and covariant derivatives and curvatures associated with $\bar{g}$ will be indicated by a bar. The induced metric on M will be denoted by g.

In this setting, because $\overline{R} = 0$, the Gauss and Codazzi equations take even simpler forms:

$$R = \frac{1}{2} h \owedge h, \quad Dh = 0. \tag{2.18}$$ Riemann Eucli

or in terms of a local frame for M,

$$R_{ijkl} = h_{il}h_{jk} - h_{jl}h_{ik}, \quad h_{ij;k} - h_{ik;j} = 0. \tag{2.19}$$ Riemann Eucli

In the setting of a hypersurface $M \subseteq \mathbb{R}^{n+1}$, we can give some very concrete geometric interpretations of the quantities we have defined so far. We begin with curves. For every unit vector $v \in T_pM$, let $\gamma = \gamma_v : I \to M$ be the g-geodesic in M with initial velocity v. Then the Gauss formula shows that the ordinary Euclidean acceleration of γ at 0 is $\gamma''(0) = \overline{D}_t\gamma' = h(v,v)N_p$. Thus $h(v,v)$ is the Euclidean curvature of γ at 0, and $h(v,v) = \langle \gamma''(0), N_p\rangle > 0$ if and only if $\gamma''(0)$ points in the same direction as N_p. In other words, $h(v,v)$ is positive if γ is curving in the direction of N_p, and negative if it is curving away from N_p.

The next proposition shows that this Euclidean curvature can be interpreted in terms of the radius of the "best circular approximation".

Proposition 2.2.17. *Suppose $\gamma : I \to \mathbb{R}^m$ is a unit-speed curve, $t_0 \in I$, and $\kappa(t_0) \neq 0$.*

(a) *There is a unique unit-speed parametrized circle $c : \mathbb{R} \to \mathbb{R}^m$, called the **osculating circle at $\gamma(t_0)$**, with the property that c and γ have the same position, velocity, and acceleration at $t = t_0$.*

(b) *The Euclidean curvature of γ at t_0 is $\kappa(t_0) = 1/R$, where R is the radius of the osculating circle.*

Proof. An easy geometric argument shows that every circle in $\mathbb{R}^m$ with center q and radius R has a unit-speed parametrization of the form

$$c(t) = q + R\cos\left(\frac{t-t_0}{R}\right)v + R\sin\left(\frac{t-t_0}{R}\right)w,$$

where (v, w) is a pair of orthonormal vectors in $\mathbb{R}^m$. By direct computation, such a parametrization satisfies

$$c(t_0) = q + Rv, \quad c'(t_0) = w, \quad c''(t_0) = -\frac{1}{R}v.$$

Thus if we put

$$R = \frac{1}{|\gamma''(t_0)|} = \frac{1}{\kappa(t_0)}, \quad v = -R\gamma''(t_0), \quad q = \gamma(t_0) - Rv.$$

we obtain a circle satisfying the required conditions, and its radius is equal to $1/\kappa(t_0)$ by construction. Uniqueness is clear. $\square$

Computations in Euclidean space

When we wish to compute the invariants of a Euclidean hypersurface $M \subseteq \mathbb{R}^{n+1}$, it is usually unnecessary to go to all the trouble of computing Christoffel symbols. Instead, it is usually more effective to use either a defining function or a parametrization to compute the scalar second fundamental form, and then use (2.18) to compute the curvature. Here we describe several contexts in which this computation is not too hard.

Riemann Euclidean Guass Codazzi-1

Usually the computations are simplest if the hypersurface is presented in terms of a local parametrization. Suppose $M \subseteq \mathbb{R}^{n+1}$ is a smooth embedded hypersurface, and let $X : U \to \mathbb{R}^{n+1}$ be a smooth local parametrization of M. The coordinates $(u^1, \ldots, u^n)$ on $U \subseteq \mathbb{R}^n$ thus give local coordinates for M. The coordinate vector fields $\partial_i = \partial/\partial u^i$ push forward to vector fields $dX(\partial_i)$ on M, which we can view as sections of the restricted tangent bundle $T\mathbb{R}^{n+1}|_M$, or equivalently as $\mathbb{R}^{n+1}$-valued functions. If we think of $X(u) = (X^1(u), \ldots, X^{n+1}(u))$ as a vector-valued function of u, these vectors can be written as

$$dX_u(\partial_i) = \partial_i X(u) = (\partial_i X^1(u), \ldots, \partial_i X^{n+1}(u)).$$

For simplicity, write $X_i = \partial_i X$.

Once these vector fields are computed, a unit normal field can be computed as follows: Choose any coordinate vector field $\partial/\partial x^{j_0}$ that is not contained in $\operatorname{span}(X_1, \ldots, X_n)$ (there will always be one, at least in a neighborhood of each point). Then apply the GramSchmidt algorithm to the local frame $(X^1, \ldots, X^n, \partial/\partial x^{j_0})$ along M to obtain an adapted orthonormal frame $(E_1, \ldots, E_n, E_{n+1})$. The two choices of unit normal are $N = \pm E_{n+1}$.

The next proposition gives a formula for the second fundamental form that is often easy to use for computation.

face two form

Proposition 2.2.18. *Suppose $M \subseteq \mathbb{R}^{n+1}$ is an embedded hypersurface, $X : U \to M$ is a smooth local parametrization of M, $(X_1, \ldots, X_n)$ is the local frame for TM determined by X, and N is a unit normal field on M. Then the scalar second fundamental form is given by*

$$h(X_i, X_j) = \left\langle \frac{\partial^2 X}{\partial u^i \partial u^j}, N \right\rangle \tag{2.20}$$

Riemann Eucli

Proof. Let $u_0 = (u_0^1, \ldots, u_0^n)$ be an arbitrary point of U and let $p = X(u_0) \in M$. For each $i \in \{1, \ldots, n\}$, the curve $\gamma(t) = X(u_0^1, \ldots, u_0^i + t, \cdots, u_0^n)$ is a smooth curve in M whose initial velocity is X_i. Regarding the normal field N as a smooth map from M to $\mathbb{R}^{n+1}$, we have

$$\frac{\partial}{\partial u^i} N(X(u_0)) = (N \circ \gamma)'(0) = \overline{\nabla}_{X_i} N(X(u_0)).$$

Because $X_j = \partial X/\partial u^j$ is tangent to M and N is normal, the following expression is zero for all $u \in U$:

$$\left\langle \frac{\partial X}{\partial u^j}(u), N(X(u)) \right\rangle.$$

Differentiating with respect to u^i and using the product rule for ordinary inner products in $\mathbb{R}^{n+1}$ yields

$$0 = \left\langle \frac{\partial^2 X}{\partial u^i \partial u^j}(u), N(X(u)) \right\rangle + \left\langle \frac{\partial X}{\partial u^j}(u), \overline{\nabla}_{X_i(u)} N(X(u)) \right\rangle.$$

By the Weingarten equation (2.9), the last term on the right becomes

$$\langle X_j(u), -s(X_i(u)) \rangle = -h(X_j(u), h_i(u)).$$

Inserting this above yields (2.20). □

Here is an application of this formula: it shows how the principal curvatures give a concise description of the local shape of an embedded hypersurface by approximating the surface with the graph of a quadratic function.

Proposition 2.2.19. *Suppose $M \subseteq \mathbb{R}^{n+1}$ is a Riemannian hypersurface. Let $p \in M$, and let $\kappa_1, \ldots, \kappa_n$ denote the principal curvatures of M at p with respect to some choice of unit normal. Then there is an isometry $\varphi : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ that takes p to the origin and takes a neighborhood of p in M to a graph of the form $x^{n+1} = f(x^1, \ldots, x^n)$, where*

$$f(x) = \frac{1}{2}(\kappa_1 (x^1)^2 + \cdots + \kappa_n (x^n)^2) + o(|x|^3). \tag{2.21}$$

Proof. Replacing M by its image under a translation and a rotation (which are Euclidean isometries), we may assume that p is the origin and T_pM is equal to the span of $(\partial_1, \ldots, \partial_n)$. Then after reflecting in the $(x^1, \ldots, x^n)$-hyperplane if necessary, we may assume that the chosen unit normal is $(0, \ldots, 0, 1)$. By an orthogonal transformation in the first n variables, we can also arrange that the scalar second fundamental form at 0 is diagonal with respect to the basis $(\partial_1, \ldots, \partial_n)$, with diagonal entries $(\kappa_1, \ldots, \kappa_n)$.

It follows from the implicit function theorem that there is some neighborhood U of 0 such that $M \cap U$ is the graph of a smooth function of the form $x^{n+1} = f(x^1, \ldots, x^n)$ with $f(0) = 0$. A smooth local parametrization of M is then given by $X(u) = (u^1, \ldots, u^n, f(u))$, and the fact that T_0M is spanned by $(\partial/\partial x^1, \ldots, \partial/\partial x^n)$ guarantees that $\partial_1 f(0) = \cdots = \partial_n f(0) = 0$. Because $X_i = \partial/\partial x^i$ at 0, Proposition 2.2.18 then yields

$$h\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \frac{\partial^2 f}{\partial x^i \partial x^j}.$$

It follows from our normalization that the matrix of second derivatives of f at 0 is diagonal, and its diagonal entries are the principal curvatures of M at that point. Then (2.21) follows from Taylor's theorem. □

Here is another approach. When it is practical to write down a smooth vector field $N = N_i \partial_i$ on an open subset of $\mathbb{R}^{n+1}$ that restricts to a unit normal vector field along M, then the shape operator can be computed straightforwardly using the Weingarten equation and observing that the Euclidean covariant derivatives of N are just ordinary directional derivatives in Euclidean space. Thus for every vector $X = X^j \partial_j$ tangent to M, we have

$$sX = -\overline{\nabla}_X N = -\sum_{i,j=1}^{n+1} X^j (\partial_j N^i) \partial_i. \tag{2.22}$$

Riemann Eucli

One common way to produce such a smooth vector field is to work with a local defining function for M: Recall that this is a smooth real-valued function defined on some open subset $U \subseteq \mathbb{R}^{n+1}$ such that $U \cap M$ is a regular level set of F. The definition ensures that $\operatorname{grad} F$ (the gradient of F with respect to $\overline{g}$) is nonzero on some neighborhood of $M \cap U$, so a convenient choice for a unit normal vector field along M is

$$N = \frac{\operatorname{grad} F}{|\operatorname{grad} F|}.$$

Here is an application.

erator sphere

Example 2.2.20 (Shape Operators of Spheres). The function $F : \mathbb{R}^{n+1} \to \mathbb{R}$ defined by $F(x) = |x|^2$ is a smooth defining function for each sphere $S^n(R)$. The gradient of this function is $\operatorname{grad} F = 2x^i \partial_i$, which has length $2R$ along $S^n(R)$. The smooth vector field

$$N = \frac{1}{R} \sum_{i=1}^{n+1} x^i \partial_i$$

thus restricts to a unit normal along $S^n(R)$. (It is the outward pointing normal.) The shape operator is now easy to compute:

$$sX = -\frac{1}{R} \sum_{i,j=1}^{n+1} X^j (\partial_j x^i) \partial_i = -\frac{1}{R} X.$$

Therefore $s = -1/R\mathrm{id}$. The principal curvatures, therefore, are all equal to $-1/R$, and it follows that the mean curvature is $H = -1/R$ and the Gaussian curvature is $(-1/R)^n$.

For surfaces in $\mathbb{R}^3$, either of the above methods can be used. When a parametrization X is given, the normal vector field is particularly easy to compute: because X_1 and X_2 span the tangent space to M at each point, their cross product is a nonzero normal vector, so one choice of unit normal is

$$N = \frac{X_1 \times X_2}{|X_1 \times X_2|}.$$

The Gaussian curvature of a surface is intrinsic

Because the Gaussian and mean curvatures are defined in terms of a particular embedding of M into $\mathbb{R}^{n+1}$, there is little reason to suspect that they have much to do with the intrinsic Riemannian geometry of (M, g). The amazing discovery made by Gauss was that the Gaussian curvature of a surface in $\mathbb{R}^3$ is actually an intrinsic invariant of the Riemannian manifold (M, g).

Theorem 2.2.21 (Gauss's Theorema Egregium). *Suppose (M, g) is an embedded 2-dimensional Riemannian submanifold of $\mathbb{R}^3$. For every $p \in M$, the Gaussian curvature of M at p is equal to one-half the scalar curvature of g at p, and thus the Gaussian curvature is a local isometry invariant of (M, g).*

Proof. Let $p \in M$ be arbitrary, and choose an orthonormal basis (e_1, e_2) for T_pM. In this basis g is represented by the identity matrix, and the shape operator has the same matrix as the scalar second fundamental form. Thus $K_p = \det s = \det(h_{ij})$, and the Gauss equation (2.10) reads

$$R_p(e_1, e_2, e_2, e_1) = h_{11}h_{22} - h_{12}h_{21} = \det(h_{ij}) = K_p.$$

On the other hand, Corollary 2.1.33 shows that $R = 1/4 Sg \owedge g$, and thus by the definition of the Kulkarni-Nomizu product we have

$$R_p(e_1, e_2, e_2, e_1) = \frac{1}{4}S(p)(2g_{11}g_{22} - 2g_{12}g_{21}) = \frac{1}{2}S(p).$$

This gives the claim. □

Motivated by the Theorema Egregium, for an abstract Riemannian 2-manifold (M, g), not necessarily embedded in $\mathbb{R}^3$, we define the **Gaussian curvature** to be $K = S/2$, where S is the scalar curvature. If M is a Riemannian submanifold of $\mathbb{R}^3$, then the Theorema Egregium shows that this new definition agrees with the original definition of K as the determinant of the shape operator. The following result is a restatement of Corollary 2.1.33 using this new definition.

Proposition 2.2.22. *If (M, g) is a Riemannian 2-manifold, the following relationships hold:*

$$R = \frac{1}{2}Kg \owedge g, \quad \mathrm{Ric} = Kg, \quad S = 2K.$$

2.2.4 Sectional curvatures

Now, finally, we can give a quantitative geometric interpretation to the curvature tensor in dimensions higher than 2. Suppose M is a Riemannian n-manifold (with $n \geq 2$), p is a point of M, and $V \subseteq T_pM$ is a star-shaped neighborhood of zero on which $\exp_p$ is a diffeomorphism onto an open set $U \subseteq M$. Let Π be any 2-dimensional linear subspace of T_pM. Since $\Pi \cap V$ is an embedded 2-dimensional submanifold of V, it follows that $S_\Pi = \exp_p(V \cap \Pi)$ is an embedded 2-dimensional submanifold of $U \subseteq M$ containing p, called the **plane section determined by Π**. Note that S_Π is just the set swept out by geodesics whose initial velocities lie in Π, and T_pS_Π is exactly Π.

We define the **sectional curvature of Π**, denoted by $\sec(\Pi)$, to be the intrinsic Gaussian curvature at p of the surface S_Π with the metric induced from the embedding S_Π. If (v, w) is any basis for Π, we also use the notation $\sec(v, w)$ for $\sec(\Pi)$.

The next theorem shows how to compute the sectional curvatures in terms of the curvature of (M, g). To make the formula more concise, we introduce the following notation. Given vectors v, w in an inner product space V, we set

$$|v \wedge w|^2 = \langle v, v\rangle\langle w, w\rangle - \langle v, w\rangle^2.$$

This is the squre of the volume of the parallelogram spanned by v and w, so $|v \wedge w| > 0$ if v and w are linearly dependent, and $|v \wedge w| = 1$ when v and w are orthonormal.

Proposition 2.2.23 (Formula for the Sectional Curvature). *Let (M, g) be a Riemannian manifold and $p \in M$. If v, w are linearly independent vectors in T_pM, then the sectional curvature of the* *plane spanned by v and w is given by*

$$\sec(v,w) = \frac{R(v,w,w,v)}{|v \wedge w|^2} \tag{2.23}$$

Riemann secti

Proof. Let $\Pi \subseteq T_pM$ be the subspace spanned by (v,w). For this proof, we denote the induced metric on S_Π by $\widehat{g}$, and its associated curvature tensor by $\widehat{R}$. By definition, $\sec(v,w)$ is equal to $\widehat{K}(p)$, the Gaussian curvature of $\widehat{g}$ at p.

We show first that the second fundamental form of S_Π in M vanishes at p. To see why, let $z \in \Pi$ be arbitrary, and let $\gamma = \gamma_z$ be the g-geodesic with initial velocity z, whose image lies in S_Π for t sufficiently near 0. By the Gauss formula for vector fields along curves,

$$0 = D_t\gamma' = \widehat{D}_t\gamma' + \mathrm{II}(\gamma', \gamma').$$

Since the two terms on the right-hand side are orthogonal, each must vanish identically. Evaluating at $t = 0$ gives $\mathrm{II}(z,z) = 0$. Since z was an arbitrary element of $\Pi = T_p(S_\Pi)$ and II is symmetric, polarization shows that $\mathrm{II} = 0$ at p. (We cannot in general expect II to vanish at other points of S_Π—it is only at p that all g-geodesics starting tangent to S_Π remain in S_Π.) The Gauss equation then tells us that the curvature tensors of M and S_Π are related at p by

$$R_p(x,y,v,w) = \widehat{R}_p(x,y,v,w).$$

whenever $x,y,v,w \in \Pi$.

Now choose an orthonormal basis (e_1,e_2) for Π. Based on the observations above, we see that the sectional curvature of Π is

$$\begin{aligned}
\widehat{K}(p) = \frac{1}{2}\widehat{S}(p) &= \frac{1}{2}\sum_{i,j=1}^{2} \widehat{R}_p(e_i,e_j,e_j,e_i) \\
&= \frac{1}{2}\widehat{R}_p(e_1,e_2,e_2,e_1) + \frac{1}{2}\widehat{R}_p(e_2,e_1,e_1,e_2) \\
&= \widehat{R}_p(e_1,e_2,e_2,e_1) = R_p(e_1,e_2,e_2,e_1).
\end{aligned}$$

To see how to compute this in terms of an arbitrary basis, let (v,w) be any basis for Π. Then we can write

$$v = ae_1 + be_2, \quad w = ce_1 + de_2,$$

and so

$$\begin{aligned}
\frac{R_p(v,w,w,v)}{|v|^2|w|^2 - \langle v,w\rangle^2} &= \frac{R_p(ae_1+be_2, ce_1+de_2, ce_1+de_2, ae_1+be_2)}{|ae_1+be_2|^2|ce_1+de_2|^2 - \langle ae_1+be_2, ce_1+de_2\rangle^2} \\
&= \frac{(a^2d^2+b^2c^2-2abcd)R_p(e_1,e_2,e_2,e_1)}{(a^2d^2+b^2c^2-2abcd)} = R_p(e_1,e_2,e_2,e_1).
\end{aligned}$$

This finishes the proof. □

Riemann sectional curvature formula-1

Proposition 2.23 shows that one important piece of quantitative information provided by the curvature tensor is that it encodes the sectional curvatures of all plane sections. It turns out, in fact, that this is all of the information contained in the curvature tensor: as the following proposition shows, the sectional curvatures completely determine the curvature tensor.

ure determine

Proposition 2.2.24. *Suppose R_1 and R_2 are algebraic curvature tensors on a finite-dimensional in-*

ner product space V. If for every pair of linearly independent vectors $v, w \in V$,

$$\frac{R_1(v,w,w,v)}{|v \wedge w|^2} = \frac{R_2(v,w,w,v)}{|v \wedge w|^2},$$

then $R_1 = R_2$.

Proof. Let R_1 and R_2 be tensors satisfying the hypotheses, and set $D = R_1 - R_2$. Then D is an algebraic curvature tensor, and $D(v,w,w,v) = 0$ for all $v, w \in V$. (This is true by hypothesis when v and w are linearly independent, and it is true by the symmetries of D when they are not.) We need to show that $D = 0$.

For all vectors v, w, x, the symmetries of D give

$$\begin{aligned}
0 &= D(v+w,x,x,v+w) = D(v,x,x,v) + D(v,x,x,w) + D(w,x,x,v) + D(w,x,x,w) \\
&= D(v,x,x,w) + D(w,x,x,v) \\
&= 2D(v,x,x,w).
\end{aligned}$$

From this it follows that

$$\begin{aligned}
0 &= D(v,x+y,x+y,w) = D(v,x,x,w) + D(v,x,y,w) + D(v,y,x,w) + D(v,y,y,w) \\
&= D(v,x,y,w) + D(v,y,x,w).
\end{aligned}$$

Therefore D is antisymmetric in every adjacent pair of arguments. Now the algebraic Bianchi identity yields

$$\begin{aligned}
0 &= D(x,y,v,w) + D(y,v,x,w) + D(v,x,y,w) \\
&= D(x,y,v,w) + D(x,y,v,w) + D(x,y,v,w) \\
&= 3D(x,y,v,w).
\end{aligned}$$

Therefore $D = 0$ and we are done. □

We can also give a geometric interpretation of the Ricci and scalar curvatures on a Riemannian manifold. Since the Ricci tensor is symmetric and bilinear, Lemma 2.2.9 shows that it is completely determined by its values of the form $\mathrm{Ric}(v,v)$ for unit vectors v. Bilinear symmetric form lemma

nterpretation **Proposition 2.2.25 (Geometric Interpretation of Ricci and Scalar Curvatures).** *Let (M,g) be a Riemannian n-manifold and $p \in M$.*

(a) For every unit vector $v \in T_pM$, $\mathrm{Ric}_p(v,v,)$ is the sum of the sectional curvatures of the 2-planes spanned by $(v,e_2),\dots,(v,e_n)$, where $(e_1,\dots,e_n)$ is any orthonormal basis for T_pM with $e_1 = v$.

(b) The scalar curvature at p is the sum of all sectional curvatures of the 2-planes spanned by ordered pairs of distinct basis vectors in any orthonormal basis.

Proof. Given any unit vector $v \in T_pM$, let $(e_1,\dots,e_n)$ be as in the hypothesis. Then $\mathrm{Ric}(v,v)$ is given by

$$\mathrm{Ric}_p(v,v) = R_{11}(p) = R_{i11}{}^{i}(p) = \sum_{i=1}^{n} R_p(e_i,v,v,e_i) = \sum_{i=2}^{n} \sec(v,e_i).$$

For the scalar curvature, we let $\{e_1,\dots,e_n\}$ be any orthonormal basis for T_pM, and compute

$$S(p) = R_j{}^{j}(p) = \sum_{j=1}^{n} \mathrm{Ric}_p(e_j,e_j) = \sum_{i,j=1}^{n} R_p(e_i,e_j,e_j,e_i) = \sum_{j \neq k} \sec(e_i,e_j).$$

From these equations, the claims of the proposition are clear. □

One consequence of this proposition is that if (M, g) is a Riemannian manifold in which all sectional curvatures are positive, then the Ricci and scalar curvatures are both positive as well. The analogous statement holds if "positive" is replaced by "negative", "nonpositive", or "nonnegative".

Sectional curvatures of the model spaces

We can now compute the sectional curvatures of our three families of framehomogeneous model spaces. A Riemannian metric or Riemannian manifold is said to have **constant sectional curvature** if the sectional curvatures are the same for all planes at all points.

Lemma 2.2.26. *If a Riemannian manifold (M, g) is frame-homogeneous, then it has constant sectional curvature.*

Proof. Frame homogeneity implies, in particular, that given two 2-planes at the same or different points, there is an isometry taking one to the other. The result follows from the isometry invariance of the curvature tensor. □

Thus to compute the sectional curvature of one of our model spaces, it suffices to compute the sectional curvature for one plane at one point in each space.

Theorem 2.2.27 (Sectional Curvatures of the Model Spaces). *The following Riemannian manifolds have the indicated constant sectional curvatures:*

(a) $(\mathbb{R}^n, \bar{g})$ *has constant sectional curvature* 0.

(b) $(S^n(R), \mathring{g})$ *has constant sectional curvature* $1/R^2$.

(c) $(\mathbb{H}^n(R), \breve{g})$ *has constant sectional curvature* $-1/R^2$.

Proof. First we consider the simplest case: Euclidean space. Since the curvature tensor of $\mathbb{R}^n$ is identically zero, clearly all sectional curvatures are zero. This is also easy to see geometrically, since each plane section is actually a plane, which has zero Gaussian curvature.

Next consider the sphere $S^n(R)$. We need only compute the sectional curvature of the plane Π spanned by (∂_1, ∂_2) at the point $(0, \dots, 0, 1)$. The geodesics with initial velocities in Π are great circles in the (x^1, x^2, x^{n+1}) subspace. Therefore S_Π is isometric to the round 2-sphere of radius R embedded in $\mathbb{R}^3$. As Example 2.2.20 showed, $S^2(R)$ has Gaussian curvature $1/R^2$. Therefore $S^n(R)$ has constant sectional curvature equal to $1/R^2$. The proof for hyperbolic spaces is similar. □

Since the sectional curvatures determine the curvature tensor, one would expect to have an explicit formula for the full curvature tensor when the sectional curvature is constant. Such a formula is provided in the following proposition.

Proposition 2.2.28. *A Riemannian metric g has constant sectional curvature λ if and only if its curvature tensor satisfies*

$$R = \frac{\lambda}{2} g \owedge g.$$

In this case, the Ricci tensor and scalar curvature of g are given by the formulas

$$\mathrm{Ric} = (n-1)\lambda g, \quad S = n(n-1)\lambda.$$

and the Riemann curvature endomorphism is

$$R(x,y)v = \lambda(\langle y,v\rangle x - \langle x,v\rangle y).$$

In terms of any basis,

$$R_{ijkl} = \lambda(g_{il}g_{jk} - g_{jl}g_{ik}), \quad R_{ij} = (n-1)\lambda g_{ij}.$$

Proof. We prove that the algebriac tensor $T := (\lambda/2)g \owedge g$ satisfies that condition of Proposition 2.2.24, then the rest is clear. In fact, for any orthonormal vectors $v, w \in T_pM$, we have

$$\frac{\lambda}{2}(g \owedge g)(v,w,w,v) = \lambda g(v,v)g(w,w) = \lambda = \sec(v,w).$$

Therefore we are done. □

2.3 The Gauss-Bonnet theorem

2.3.1 Some plane geometry

Throughout this part, $\gamma : [a,b] \to \mathbb{R}^2$ is an admissible curve in the plane. We say that γ is a **simple closed curve** if $\gamma(a) = \gamma(b)$ but γ is injective on $[a,b)$. We donot assume that γ has unit speed; instead, we define the unit **tangent vector field of γ** as the vector field T along each smooth segment of γ given by

$$T = \frac{\gamma'(t)}{|\gamma'(t)|}.$$

Since each tangent space to $\mathbb{R}^2$ is naturally identified with $\mathbb{R}^2$ itself, we can think of T as a map into $\mathbb{R}^2$, and since T has unit length, it takes its values in S^1.

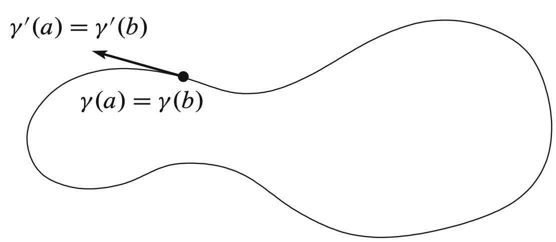

Figure 2.1: A closed curve with $\gamma'(a) = \gamma'(b)$.

If γ is smooth (or at least continuously differentiable), we define a **tangent angle function for γ** to be a continuous function $\theta : [a,b] \to \mathbb{R}$ such that $T(t) = (\cos\theta(t), \sin\theta(t))$ for all $t \in [a,b]$. It follows from the theory of covering spaces that such a function exists: the map $p : \mathbb{R} \to S^1$ given by $p(t) = (\cos t, \sin t)$ is a smooth covering map, and the path-lifting property of covering maps ensures that there is a continuous function $\theta : [a,b] \to \mathbb{R}$ that satisfies $p(\theta(t)) = T(t)$. By the unique lifting property, a lift is uniquely determined once its value at any single point is determined, and thus any two lifts differ by a constant integral multiple of 2π.

If γ is a continuously differentiable simple closed curve such that $\gamma'(a) = \gamma'(b)$, then $(\cos\theta(a), \sin\theta(a)) = (\cos\theta(b), \sin\theta(b))$, so $\theta(b) - \theta(a)$ must be an integral multiple of 2π. For such a curve, we define the **rotation index** of γ to be the following integer:

$$\rho(\gamma) = \frac{1}{2\pi}(\theta(b) - \theta(a)).$$

where θ is any tangent angle function for γ. For any other choice of tangent angle function, $\gamma(a)$ and $\gamma(b)$ would change by addition of the same constant, so the rotation index is independent of the choice of θ. We would also like to extend the definition of the rotation index to certain piecewise regular closed curves. For this purpose, we have to take into account the "jumps" in the tangent angle function at corners. To do so, suppose $\gamma : [a, b] \to \mathbb{R}^2$ is an admissible simple closed curve, and let $(a_0, \ldots, a_k)$ be an admissible partition of $[a, b]$. The points $\gamma(a_i)$ are called the **vertices** of γ, and the curve segments $\gamma|_{[a_{i-1}, a_i]}$ are called its **edges** or **sides**.

At each vertex $\gamma(a_i)$, recall that γ has left-hand and right-hand velocity vectors denoted by $\gamma'(a_i^-)$ and $\gamma'(a_i^+)$, respectively; let $T(a_i^-)$ and $T(a_i^+)$ denote the corresponding unit vectors. We classify each vertex into one of three categories:

- If $T(a_i^-) \neq \pm T(a_i^+)$, then $\gamma(a_i)$ is an **ordinary vertex**.
- If $T(a_i^- = T(a_i^+)$, then $\gamma(a_i)$ is an **flat vertex**.
- If $T(a_i^-) = -T(a_i^+)$, then $\gamma(a_i)$ is a **cusp vertex**.

At each ordinary vertex, define the **exterior angle at $\boldsymbol{\gamma(a_i)}$** to be the oriented measure ε_i of the angle from $T(a_i^-)$ to $T(a_i^+)$, chosen to be in the interval $(-\pi, \pi)$, with a positive sign if $(T(a_i^-), T(a_i^+))$ is an oriented basis for $\mathbb{R}^2$, and a negative sign otherwise. At a flat vertex, the exterior angle is defined to be zero. At a cusp vertex, there is no simple way to choose unambiguously between π and $-\pi$, so we leave the exterior angle undefined. The vertex $\gamma(a) = \gamma(b)$ is handled in the same way, with $T(b)$ and $T(a)$ playing the roles of $T(a^-)$ and $T(a^+)$, respectively. If $\gamma(a_i)$ is an ordinary or a flat vertex, the **interior angle at $\boldsymbol{\gamma(a_i)}$** is defined to be $\theta_i = \pi - \varepsilon_i$; our conventions ensure that $0 < \theta_i < 2\pi$.

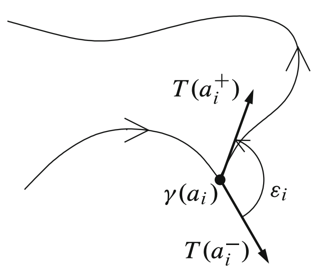

Figure 2.2: An exterior angle.

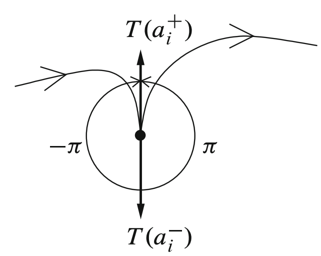

Figure 2.3: A cusp vertex.

The curves we wish to consider are of the following type: a **curved polygon** in the plane is an admissible simple closed curve without cusp vertices, whose image is the boundary of a precompact open set $\Omega \subseteq \mathbb{R}^2$. The set Ω is called the interior of γ (not to be confused with the topological interior of its image as a subset of $\mathbb{R}^2$, which is the empty set).

Suppose $\gamma : [a, b] \to \mathbb{R}^2$ is a curved polygon. If γ is parametrized so that at smooth points, γ' is positively oriented with respect to the induced orientation on $\partial\Omega$ in the sense of Stokes's theorem, we say that γ is **positively oriented**. Intuitively, this means that γ is parametrized in the counterclockwise direction, or that Ω is always to the left of γ.

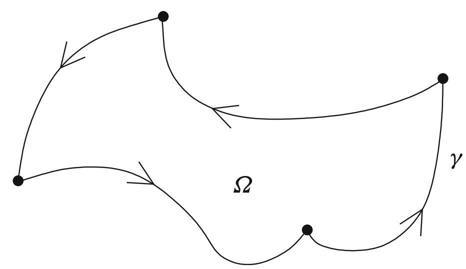

Figure 2.4: A positively oriented curved polygon.

We define a **tangent angle function for a curved polygon $\boldsymbol{\gamma}$** to be a piecewise continuous function $\theta : [a, b] \to \mathbb{R}$ that satisfies $T(t) = (\cos\theta(t), \sin\theta(t))$ at each point t where γ is smooth, that is continuous from the right at each a_i with

$$\theta(a_i) = \lim_{t \to a_i^-} \theta(t) + \varepsilon_i, \tag{3.1}$$

and that satisfies

$$\theta(b) = \lim_{t \to b} \theta(t) + \varepsilon_k, \tag{3.2}$$

where ε_k is the exterior angle at $\gamma(b)$. Such a function always exists: start by defining $\theta(t)$ for $t \in [a, a_1)$ to be any lift of T on that interval; then on $[a_1, a_2)$ define $\theta(t)$ to be the unique lift that satisfies (3.1), and continue by induction, ending with $\gamma(b)$ defined by (3.2).

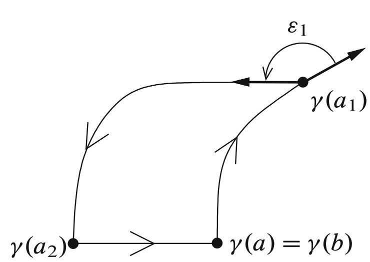

Figure 2.5: Tangent angle at a vertex.

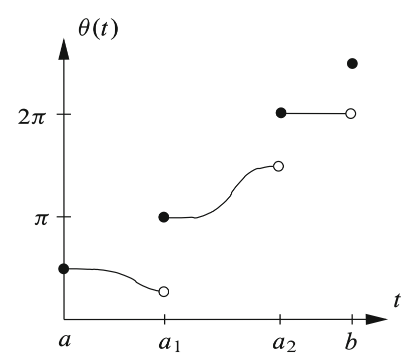

Figure 2.6: Tangent angle function.

Once again, the difference between any two such functions is a constant integral multiple of 2π. We define the rotation index of γ to be

$$\rho(\gamma) = \frac{1}{2\pi}(\theta(b) - \theta(a)),$$

just as in the smooth case. As before, $\rho(\gamma)$ is an integer, because the definition ensures that $(\cos\theta(b), \sin\theta(b)) = (\cos\theta(a), \sin\theta(a))$.

Theorem 2.3.1 (Rotation Index Theorem). *The rotation index of a positively oriented curved polygon in the plane is $+1$.*

Proof. Let $\gamma : [a, b] \to \mathbb{R}^2$ be such a curved polygon. Assume first that all the vertices of γ are flat. This means, in particular, that γ' is continuous and $\gamma'(a) = \gamma'(b)$. Since $\gamma(a) = \gamma(b)$, we can extend γ to a continuous map from $\mathbb{R}$ to $\mathbb{R}^2$ by requiring it to be periodic of period $b - a$, and our hypothesis $\gamma'(a) = \gamma'(b)$ guarantees that the extended map still has continuous first derivatives. Define $T(t) = \gamma'(t)/|\gamma'(t)|$ as before.

Let $\theta : \mathbb{R} \to S^1$ be any lift of $T : \mathbb{R} \to S^1$. Then $\theta|_{[a,b]}$ is a tangent angle function for γ, and thus $\theta(b) = \theta(a) + 2\pi\rho(\gamma)$. If we set $\widetilde{\theta}(t) = \theta(t + b - a) - 2\pi\rho(\gamma)$, then

$$(\cos\widehat{\theta}(t), \sin\widehat{\theta}(t)) = (\cos\theta(t + b - a), \sin\theta(t + b - a)) = T(t + b - a) = T.$$

so $\widetilde{\theta}$ is also a lift of T. Because $\widetilde{\theta}(a) = \theta(a)$, it follows that $\widetilde{\theta} \equiv \theta$, or in other words the following equation holds for all $t \in \mathbb{R}$:

$$\theta(t + b - a) = \theta(t) + 2\pi\rho(\gamma). \tag{3.3}$$

If a_1 is an arbitrary point in $[a, b]$ and $b_1 = a_1 + b - a$, then $\gamma|_{[a_1,b_1]}$ is also a positively oriented curved polygon with only flat vertices, and $\theta|_{[a_1,b_1]}$ is a tangent angle function for it. Note that (3.3) implies

$$\theta(b_1) - \theta(a_1) = \theta(a_1 + b - a) - \theta(a_1) = \theta(a_1) + 2\pi\rho(\gamma) - \theta(a_1) = 2\pi\rho(\gamma).$$

so $\gamma|_{[a_1,b_1]}$ has the same rotation index as $\gamma|_{[a,b]}$. Thus we obtain the same result by restricting γ to any closed interval of length $b - a$.

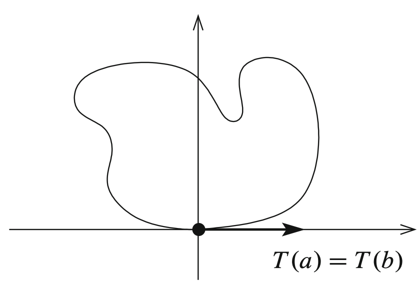

Figure 2.7: Changing the parameter interval and translating $\gamma(a)$ to the origin.

Using this freedom, we can assume that the parameter interval $[a, b]$ has been chosen so that the y-coordinate of γ achieves its minimum at $t = a$. Moreover, by a translation in the xy-plane (which does not change γ' or γ), we may as well assume that $\gamma(a)$ is the origin. With these adjustments, the image of γ remains in the closed upper half-plane, and $T(a) = T(b) = (1, 0)$. By adding a constant integral multiple of 2π to θ if necessary, we can also assume that $\theta(a) = 0$.

Next, we define a continuous **secant angle function**, denoted by $\varphi(t_1, t_2)$, representing the angle between the positive x-direction and the vector from $\gamma(t_1)$ to $\gamma(t_2)$. To be precise, let $\Delta \subseteq \mathbb{R}^2$ be the triangular region $\Delta = \{(t_1, t_2) : a \le t_1 \le t_2 \le b\}$, and define a map $V : \Delta \to S^1$ by

$$V(t_1, t_2) = \begin{cases} \dfrac{\gamma(t_2) - \gamma(t_1)}{|\gamma(t_2) - \gamma(t_1)|}, & t_1 < t_2 \text{ and } (t_1, t_2) \neq (a, b); \\ T(t_1), & t_1 = t_2; \\ -T(b), & (t_1, t_2) = (a, b). \end{cases}$$

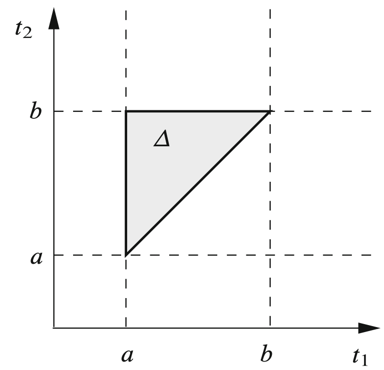

Figure 2.8: The domain of φ.

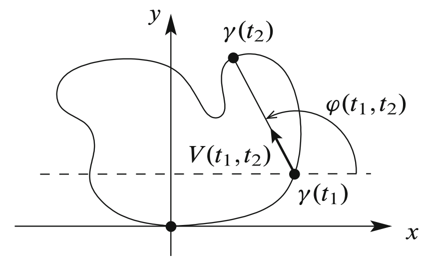

Figure 2.9: The secant angle function.

The function V is continuous at points where $t_1 < t_2$ and $(t_1, t_2) \neq (a, b)$, because γ is continuous and injective there. To see that it is continuous elsewhere, note that for $t_1 < t_2$, the fundamental theorem of calculus gives

$$\gamma(t_2) - \gamma(t_1) = \int_0^1 \frac{d}{ds}\gamma(t_1 + s(t_2 - t_1))ds = \int_0^1 \gamma'(t_1 + s(t_2 - t_1))(t_2 - t_1)ds.$$

and thus

$$\Big|\frac{\gamma(t_2) - \gamma(t_1)}{t_2 - t_1} - \gamma'(t)\Big| \leq \int_0^1 |\gamma'(t_1 + s(t_2 - t_1)) - \gamma'(t)|ds.$$

Because γ' is uniformly continuous on the compact set $[a, b]$, this last expression can be made as small as desired by taking (t_1, t_2) close to (t, t). It follows that

$$\lim_{\substack{(t_1,t_2)\to(t,t)\\ t_1<t_2}} \frac{\gamma(t_2) - \gamma(t_1)}{t_2 - t_1} = \gamma'(t),$$

and therefore

$$\lim_{\substack{(t_1,t_2)\to(t,t)\\ t_1<t_2}} V(t_1, t_2) = \lim_{\substack{(t_1,t_2)\to(t,t)\\ t_1<t_2}} \frac{\gamma(t_2) - \gamma(t_1)}{|\gamma(t_2) - \gamma(t_1)|} = \frac{\gamma'(t)}{\gamma'(t)|} = T(t) = V(t, t).$$

Similarly, because T is continuous,

$$\lim_{\substack{(t_1,t_2)\to(t,t)\\ t_1=t_2}} V(t_1, t_2) = \lim_{t_1\to t} T(t_1) = T(t) = V(t, t).$$

It follows that V is continuous at (t, t).

To prove that V is continuous at (a, b), recall that we have extended γ to be periodic of period $b - a$. The argument above gives

$$\begin{aligned} \lim_{\substack{(t_1,t_2)\to(a,b)\\ t_1<t_2}} V(t_1, t_2) &= \lim_{\substack{(t_1,t_2)\to(a,b)\\ t_1<t_2}} \frac{\gamma(t_2) - \gamma(t_1 + b - a)}{|\gamma(t_2) - \gamma(t_1 + b - a)|} \\ &= \lim_{\substack{(s_1,s_2)\to(b,b)\\ s_1>2}} \frac{\gamma(s_2) - \gamma(s_1)}{|\gamma(s_2) - \gamma(s_1)|} = -T(b) = V(a, b). \end{aligned}$$

Thus V is continuous.

Since Δ is simply connected, the map $V : \Delta \to S^1$ has a continuous lift $\varphi : \Delta \to \mathbb{R}$, which is unique if we require $\varphi(a, a) = 0$. This is our secant angle function.

We can express the rotation index in terms of the secant angle function as follows:

$$\rho(\gamma) = \frac{1}{2\pi}(\theta(b) - \theta(a)) = \frac{1}{2\pi}(\varphi(b,b) - \varphi(a,a)) = \frac{1}{2\pi}\varphi(b,b).$$

Observe that along the side of Δ where $t_1 = a$ and $t_2 \in [a,b]$, the vector $V(a,t_2)$ has its tail at the origin and its head in the upper half-plane. Since we stipulate that $\varphi(a,a) = 0$, we must have $\varphi(a,t_2) \in [0,\pi]$ on this segment. By continuity, therefore, $\varphi(a,b) = \pi$ (since $\varphi(a,b)$ represents the tangent angle of $-T(b) = (-1,0)$). Similarly, on the side where $t_2 = b$, the vector $V(t_1,b)$ has its head at the origin and its tail in the upper half-plane, so $\varphi(t_1,b) \in [\pi, 2\pi]$ Therefore, since $\varphi(b,b)$ represents the tangent angle of $T(b) = (1,0)$, we must have $\varphi(b,b) = 2\pi$ and therefore $\rho(\gamma) = 1$. This completes the proof for the case in which γ' is continuous.

Now suppose γ has one or more ordinary vertices. It suffices to show there is a curve with a continuous velocity vector field that has the same rotation index as γ. We will construct such a curve by "rounding the corners" of γ. It will simplify the proof somewhat if we choose the parameter interval $[a,b]$ so that $\gamma(a) = \gamma(b)$ is not one of the ordinary vertices.

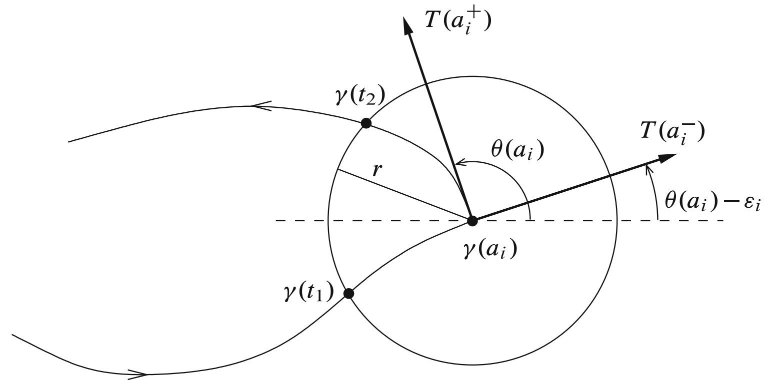

Figure 2.10: Isolating the change in the tangent angle at a vertex.

Let $\gamma(a_i)$ be any ordinary vertex, let ε_i be its exterior angle, and let α be a positive number less than $(\pi - |\varepsilon_i|)/2$. Recall that γ is continuous from the right at a_i and $\lim_{t\to a_i^-}\theta(t) = \theta(a_i) - \varepsilon_i$. Therefore, we can choose δ small enough that $|\theta(t) - \theta(a_i) - \varepsilon_i| < \alpha$ when $t \in (a_i - \delta, a_i)$, and $|\theta(t) - \theta(a_i)| < \alpha$ when $t \in (a_i, a_i + \delta)$.

The image under $\gamma$of $[a,b] \setminus (a_i - \delta, a_i + \delta)$ is a compact set that does not contain $\gamma(a_i)$, so we can choose r small enough that γ does not enter $\overline{B_r(\gamma(a_i)}$ except when $t \in (a_i - \delta, a_i + \delta)$. Let $t_1 \in (a_i - \delta, a_i)$ denote a time when γ enters $\overline{B_r(\gamma(a_i))}$, and $t_2 \in (a_i, a_i + \delta)$ a time when it leaves. By our choice of δ, the total change in $\theta(t)$ is not more than α when $t \in [t_1, a_i)$, and again not more than α when $t \in (a_i, t_2]$. Therefore, the total change $\Delta\theta$ in $\gamma(t)$ during the time interval $[t_1, t_2]$ is between $\varepsilon - 2\alpha$ and $\varepsilon + 2\alpha$ which implies $-\pi < \Delta\theta < \pi$.

Now we simply replace $\gamma|_{[t_1,t_2]}$ with a smooth curve segment σ that has the same velocity as γ at $\gamma(t_1)$ and $\gamma(t_2)$, and whose tangent angle increases or decreases monotonically from $\gamma(t_1)$ to $\gamma(t_2)$; an arc of a hyperbola will do. Since the change in tangent angle of σ is between $-\pi$ and π and represents the angle between $T(t_1)$ and $T(t_2)$, it must be exactly $\Delta\theta$. Repeating this process for each vertex, we obtain a new curved polygon with a continuous velocity vector field whose rotation index is the same as that of γ, thus proving the theorem. $\square$

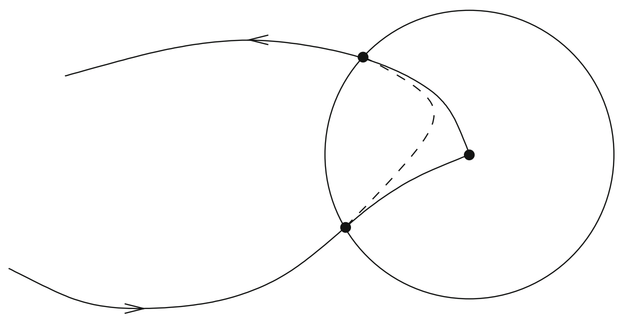

Figure 2.11: Rounding a corner.

2.3.2 The Gauss-Bonnet formula

We now direct our attention to the case of an oriented Riemannian 2-manifold (M, g). In this setting, an admissible simple closed curve $\gamma : [a, b] \to M$ is called a **curved polygon in M** if the image of γ is the boundary of a precompact open set $\Omega \subseteq M$, and there is an oriented smooth coordinate disk containing $\overline{\Omega}$ under whose image γ is a curved polygon in the plane. As in the planar case, we call Ω the **interior of γ**. A curved polygon whose edges are all geodesic segments is called a **geodesic polygon**.

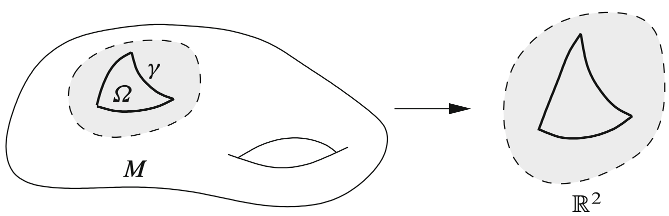

Figure 2.12: A curved polygon on a surface.

For a curved polygon γ in M, our previous definitions go through almost unchanged. We say that γ is **positively oriented** if it is parametrized in the direction of its Stokes orientation as the boundary of Ω. On each smooth segment of γ, we define the **unit tangent vector field** $T(t) = \gamma'(t)/|\gamma'(t)|$. If $\gamma(a_i)$ is an ordinary or flat vertex, we define the **exterior angle of γ** at $\gamma(a_i)$ as the oriented measure ε_i of the angle from $T(a_i^-)$ to $T(a_i^+)$ with respect to the g-inner product and the given orientation of M; explicitly, this is

$$\varepsilon_i = \frac{dV_g(T(a_i^-), T(a_i^+))}{|dV_g(T(a_i^-), T(a_i^+))|} \arccos\langle T(a_i^-), T(a_i^+)\rangle. \tag{3.4}$$

Riemann exter

The corresponding **interior angle of γ at $\gamma(a_i)$** is $\theta_i = \pi - \varepsilon_i$. Exterior and interior angles at $\gamma(a) = \gamma(b)$ are defined similarly.

We need a version of the rotation index theorem for curved polygons in M. Suppose $\gamma : [a, b] \to M$ is a curved polygon and Ω is its interior, and let (U, φ) be an oriented smooth chart such that U contains $\overline{\Omega}$. Using the coordinate map φ to transfer γ, Ω, and g to the plane, we may as well assume that g is a metric on some open subset $\widehat{U} \subseteq \mathbb{R}^2$, and γ is a curved polygon in $\widehat{U}$. Let (E_1, E_2) be the oriented orthonormal frame for g obtained by applying the Gram-Schmidt algorithm to (∂_x, ∂_y), so that E_1 is a positive scalar multiple of ∂_x everywhere in $\widehat{U}$.

We define a **tangent angle function for $\boldsymbol{\gamma}$** to be a piecewise continuous function $\theta : [a,b] \to \mathbb{R}$ that satisfies

$$T(t) = \cos\theta(t) E_1|_{\gamma(t)} + \sin\theta(t) E_2|_{\gamma(t)}.$$

at each t where γ' is continuous, and that is continuous from the right and satisfies (3.2) and (3.2) at vertices. The existence of such a function follows as in the planar case, using the fact that

$$T(t) = u_1 E_1|_{\gamma(t)} + u_2 E_2|_{\gamma(t)}$$

for a pair of piecewise continuous functions $u_1, u_2 : [a,b] \to \mathbb{R}$ that can be regarded as the coordinate functions of a map $(u_1, u_2) : [a,b] \to S^1$ because T has unit length.

The **rotation index of $\boldsymbol{\gamma}$** is $(\theta(b) - \theta(a))/2\pi$. Because of the role played by the specific frame (E_1, E_2) in the definition, it is not obvious that the rotation index has any coordinate-independent meaning; however, the following easy consequence of the rotation index theorem shows that it does not depend on the choice of coordinates.

Lemma 2.3.2. *If M is an oriented Riemannian 2-manifold, the rotation index of every positively oriented curved polygon in M is $+1$.*

Proof. If we use the given oriented coordinate chart to regard γ as a curved polygon in the plane, we can compute its tangent angle function either with respect to g or with respect to the Euclidean metric $\bar{g}$. In either case, $\rho(\gamma)$ is an integer because $\theta(a)$ and $\theta(b)$ both represent the angle between the same two vectors, calculated with respect to some inner product. Now for $0 \le s \le 1$, let $g_s = sg + (1-s)\bar{g}$. By the same reasoning, the rotation index $\rho_{g_s}(\gamma)$ with respect to g_s is also an integer for each s, so the function $f(s) = \rho_{g_s}(\gamma)$ is integer-valued.

In fact, the function f is continuous in s, as can be deduced easily from the following observations: (1) Our preferred g_s-orthonormal frame $(E_1^{(s)}, E_2^{(s)})$ depends continuously on s, as can be seen from the formulas of the GramSchmidt algorithm. (2) On every interval $[a_{i-1}, a_i]$ where γ is smooth, the functions u_1 and u_2 satisfying the g_s-analogue of () can be expressed as $u_j(t,s) = \langle T_s(t), E_j^{(s)|_{\gamma(t)}} \rangle_{g_s}$, where $T_s(t) = \gamma'(t)/|\gamma'(t)|_{g_s}$. Thus u_1 and u_2 depend continuously on $(t,s) \in [a_{i-1}, a_i] \times [0,1]$, so the function $(u_1, u_2) : [a_{i-1}, a_i] \times [0,1] \to S^1$ has a continuous lift $\theta : [a_{i-1}, a_i] \times [0,1] \to \mathbb{R}$, uniquely determined by its value at one point. (3) At each vertex, it follows from formula (3.4) that the exterior angle depends continuously on g_s.

Because f is continuous and integer-valued, it follows that $\rho(\gamma) = f(1) = f(0) = \rho_{\bar{g}}(\gamma) = 1$, which was to be proved. □

From this point onward, we assume for convenience that our curved polygon γ is given a unit-speed parametrization, so the unit tangent vector field $T(t)$ is equal to $\gamma'(t)$. There is a unique unit normal vector field N along the smooth portions of γ such that $(\gamma'(t), N(t))$ is an oriented orthonormal basis for $T_{\gamma(t)}M$ for each t. If γ is positively oriented as the boundary of Ω, this is equivalent to N being the inward-pointing normal to ∂N. We define the signed curvature of γ at smooth points of γ by

$$\kappa_N = \langle D_t\gamma'(t), N(t) \rangle_g.$$

By differentiating $|\gamma'(t)|_g \equiv 1$, we see that $D_t\gamma'(t)$ is orthogonal to $\gamma'(t)$, and therefore we can write $D_t\gamma'(t) = \kappa_N N(t)$, and the (unsigned) geodesic curvature of γ is $\kappa(t) = |\kappa_N(t)|$. The

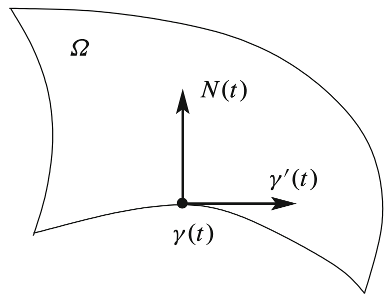

Figure 2.13: $N(t)$ is the inward-pointing normal.

sign of $N(t)$ is positive if γ is curving toward Ω, and negative if it is curving away.

Theorem 2.3.3 (The Gauss-Bonnet Formula). *Let (M, g) be an oriented Riemannian 2-manifold. Suppose γ is a positively oriented curved polygon in M, and Ω is its interior. Then*

$$\int_\Omega K dA + \int_\gamma \kappa_N ds + \sum_{i=1}^k \varepsilon_i = 2\pi. \tag{3.5}$$

Guass Bonnet

where K is the Gaussian curvature of g, dA is its Riemannian volume form, $\varepsilon_1, \ldots, \varepsilon_k$ are the exterior angles of γ.

Proof. Let $(a_0, \ldots, a_k)$ be an admissible partition of $[a, b]$, and let (x, y) be oriented smooth coordinates on an open set U containing $\overline{\Omega}$. Let $\theta : [a, b] \to \mathbb{R}$ be a tangent angle function for γ. Using the rotation index theorem and the fundamental theorem of calculus, we can write

$$2\pi = \theta(b) - \theta(a) = \sum_{i=1}^k \varepsilon_i + \sum_{i=1}^k \int_{a_{i-1}}^{a_i} \theta'(t) dt. \tag{3.6}$$

Guass Bonnet

Guass Bonnet formula-1

To prove (3.5), we need to derive a relationship among θ', κ_N and K. Let (E_1, E_2) be the oriented g-orthonormal frame obtained by applying the Gram-Schmidt algorithm to (∂_x, ∂_y) as before. Then by definition of θ and N, the following formulas hold at smooth points of γ:

$$\gamma'(t) = \cos\theta(t) E_1|_{\gamma(t)} + \sin\theta(t) E_2|_{\gamma(t)}, \quad N(t) = -\sin\theta(t) E_1|_{\gamma(t)} + \cos\theta(t) E_2|_{\gamma(t)}.$$

Differentiating γ' (and omitting the t dependence from the notation for simplicity), we get

$$\begin{aligned} D_t\gamma' &= -(\sin\theta)\theta' E_1 + (\cos\theta)\nabla_{\gamma'} E_1 + (\cos\theta)\theta' E_2 + (\sin\theta)\nabla_{\gamma'} E_2 \\ &= \theta' N + (\cos\theta)\nabla_{\gamma'} E_1 + (\sin\theta)\nabla_{\gamma'} E_2. \end{aligned}$$

Next we analyze the covariant derivatives of E_1 and E_2. Let $\omega = \omega_2^1$ be the connection one form, that is,

$$\nabla_v E_2 = \omega(v) E_1, \quad \nabla_v E_1 = -\omega(v) E_2.$$

Using this, we can compute

$$\begin{aligned} \kappa_N &= \langle D_t\gamma', N\rangle \\ &= \langle \theta' N, N\rangle + \cos\theta \langle \nabla_{\gamma'} E_1, N\rangle + \sin\theta \langle \nabla_{\gamma'} E_2, N\rangle \\ &= \theta' - \cos\theta \langle \omega(\gamma') E_2, N\rangle + \sin\theta \langle \omega(\gamma') E_1, N\rangle \\ &= \theta' - \cos^2\theta\, \omega(\gamma') - \sin^2\theta\, \omega(\gamma') \\ &= \theta' - \omega(\gamma'). \end{aligned}$$

Therefore, (3.6) becomes

$$\begin{aligned}2\pi &= \sum_{i=1}^{k}\varepsilon_i + \sum_{i=1}^{k}\int_{a_{i-1}}^{a_i}\kappa_N(t)dt + \sum_{i=1}^{k}\int_{a_{i-1}}^{a_i}\omega(\gamma'(t))dt \\ &= \sum_{i=1}^{k}\varepsilon_i + \int_\gamma \kappa_N ds + \int_\gamma \omega.\end{aligned}$$

The theorem will therefore be proved if we can show that

$$\int_\gamma \omega = \int_\Omega K dA. \tag{3.7}$$

Because Ω is a smooth manifold with corners, we can apply Stokes's theorem and conclude that the left-hand side of (3.7) is equal to $\int_\Omega d\omega$. The last step of the proof is to show that $d\omega = K dA$. This follows from the general formula relating the curvature tensor and the connection 1-forms given in Proposition 1.10:

$$\Omega_2^1 = d\omega_2^1 = d\omega.$$

and therefore

$$K dA(E_1,E_2) = K = \langle R(E_1,E_2)E_2, E_1\rangle = \langle \Omega_2^1(E_1,E_2)E_1, E_1\rangle = \Omega_2^1(E_1,E_2) = d\omega(E_1,E_2).$$

This completes the proof. □

The following local-to-global theorems of plane geometry follow immediately from the Gauss-Bonnet formula.

Corollary 2.3.4 (Angle-Sum Theorem). *Let (M,g) be an oriented Riemannian 2-manifold. Suppose γ is a positively oriented curved triangle in M, and Ω is its interior. Then*

$$\sum_{i=1}^{3}\theta_i = \pi + \int_\Omega K dA + \int_\gamma \kappa_N ds. \tag{3.8}$$

where K is the Gaussian curvature of g and $\theta_1, \theta_2, \theta_3$ are the interior angles of γ.

Corollary 2.3.5 (Total Curvature Theorem). *If $\gamma : [a,b] \to \mathbb{R}^2$ is a smooth, unitspeed, simple closed curve such that $\gamma'(a) = \gamma'(b)$, and N is the inward-pointing normal, then*

$$\int_a^b \kappa_N ds = 2\pi.$$

2.3.3 The Gauss-Bonnet theorem

Let M be a smooth, compact 2-manifold. A **curved triangle in M** is a curved polygon with exactly three edges and three vertices. A **smooth triangulation of M** is a finite collection of curved triangles with disjoint interiors such that the union of the triangles with their interiors is M, and the intersection of any pair of triangles (if not empty) is either a single vertex of each or a single edge of each. Every smooth, compact surface possesses a smooth triangulation. In fact, it was proved by Tibor Radó in 1925 that every compact topological 2-manifold possesses a triangulation (without the assumption of smoothness of the edges, of course), in which every edge belongs to exactly two triangles. There is a proof for the smooth case that is not terribly hard, based on choosing geodesic triangles contained in convex geodesic balls.

If M is a triangulated 2-manifold, the Euler characteristic of M (with respect to the given triangulation) is defined to be

$$\chi(M) = V - E + F,$$

where V is the number of vertices in the triangulation, E is the number of edges, and F is the number of faces (the interiors of the triangles). It is an important result of algebraic topology that the Euler characteristic is in fact a topological invariant, and is independent of the choice of triangulation.

Theorem 2.3.6 (**The Gauss-Bonnet Theorem**). *If (M, g) is a smoothly triangulated compact Riemannian 2-manifold, then*

$$\int_M K dA = 2\pi\chi(M).$$

Proof. We may as well assume that M is connected, because if not we can prove the theorem for each connected component and add up the results.

First consider the case in which M is orientable. In this case, we can choose an orientation for M, and then $\int_M K dA$ gives the same result whether we interpret dA as the Riemannian density or as the Riemannian volume form, so we will use the latter interpretation for the proof. Let $\{\Omega_i : i = 1, \dots, F\}$ denote the faces of the triangulation, and for each i, let $\{\gamma_{ij} : j = 1, 2, 3\}$ be the edges of Ω_i and $\{\theta_{ij} : j = 1, 2, 3\}$ its interior angles. Since each exterior angle is π minus the corresponding interior angle, applying the Gauss-Bonnet formula to each triangle and summing over i gives

$$\sum_{i=1}^{F}\sum_{j=1}^{3}\theta_{ij} = \pi F + \sum_{i=1}^{F}\int_{\Omega_i} K dA + \sum_{i=1}^{F}\sum_{j=1}^{3}\int_{\gamma_{ij}}\kappa_N ds. \tag{3.9}$$

Note that each edge integral appears exactly twice in the above sum, with opposite orientations, so the integrals of κ_N all cancel out. Thus (3.9) becomes

$$\sum_{i=1}^{F}\sum_{j=1}^{3}\theta_{ij} = \pi F + \int_M K dA. \tag{3.10}$$

Note also that each interior angle θ_{ij} appears exactly once. At each vertex, the angles that touch that vertex must have interior measures that add up to 2π; thus the angle sum can be rearranged to give exactly $2\pi V$. Equation (3.10) thus can be written

$$2\pi V = \pi F + \int_M K dA. \tag{3.11}$$

Finally, since each edge appears in exactly two triangles, and each triangle has exactly three edges, the total number of edges counted with multiplicity is $2E = 3F$, where we count each edge once for each triangle in which it appears. This means that $F = 2E - 2F$, so (3.11) finally becomes

$$\int_M K dA = 2\pi V - \pi F = 2\pi V - 2\pi E + 2\pi F = 2\pi\chi(M).$$

Now suppose M is nonorientable. Then there is an orientable connected smooth manifold $\widehat{M}$ that admits a 2-sheeted smooth covering $\widehat{\pi} : \widehat{M} \to M$, and then $\widehat{M}$ is compact since M is. If we endow $\widehat{M}$ with the metric $\widehat{g} = \widehat{\pi}^* g$, then the Riemannian density of $\widehat{g}$ is given by

$\widehat{dA} = \widehat{\pi}^* dA$, and its Gaussian curvature is $\widehat{K} = \widehat{\pi}^* K$, so $\widehat{\pi}^*(K dA) = \widehat{K}\widehat{dA}$ and we get

$$\int_{\widehat{M}} \widehat{K}\widehat{dA} = 2\int_M K dA.$$

To compare the Euler characteristics, we will show that the given triangulation of M "lifts" to a smooth triangulation of $\widehat{M}$. To see this, let γ be any curved triangle in M and let Ω be its interior. By definition, this means that there exists a smooth chart (U, φ) whose domain contains $\overline{\Omega}$ and whose image is a disk $D \subseteq \mathbb{R}^2$, and such that $\varphi(\overline{\Omega}) = \overline{\Omega}_0$, where Ω_0 is the interior of a curved triangle γ_0 in $\mathbb{R}^2$. Then φ^{-1} is an embedding of D into M, which restricts to a diffeomorphism $F : \overline{\Omega}_0 \to \overline{\Omega}$. Because D is simply connected, it follows that φ^{-1} (and therefore also F) has a lift to $\widehat{M}$, which is smooth because $\widehat{\pi}$ is a local diffeomorphism; and because the covering is two-sheeted, there are exactly two such lifts F_1, F_2. Each lift is injective because $\overline{\pi} \circ F_i = F$, which is injective, and their images are disjoint because if two lifts agree at a point, they must be identical. From this it is straightforward to verify that the lifted curved triangles form a triangulation of $\widehat{M}$ with twice as many vertices, edges, and faces as that of M, and thus $\chi(\widehat{M}) = 2\chi(M)$. Substituting these relations into the Gauss-Bonnet theorem for $\widehat{M}$ and dividing through by 2, we obtain the relation for M. □

The significance of this theorem cannot be overstated. Together with the classification theorem for compact surfaces, it gives us very detailed information about the possible Gaussian curvatures for metrics on compact surfaces. The classification theorem says that every compact, connected, orientable 2-manifold M is homeomorphic to a sphere or a connected sum of n tori, and every nonorientable one is homeomorphic to a connected sum of n copies of the real projective plane $\mathbb{RP}^2$; the number n is called the genus of M. (The sphere is said to have genus zero.) By constructing simple triangulations, one can show that the Euler characteristic of an orientable surface of genus n is $2 - 2n$, and that of a nonorientable one is $2 - n$. The following corollary follows immediately from the Gauss-Bonnet theorem.

to curvature **Corollary 2.3.7.** *Let (M, g) be a compact Riemannian 2-manifold and let K be its Gaussian curvature.*

(a) If M is homeomorphic to the sphere or the projective plane, then $K > 0$ somewhere.

(b) If M is homeomorphic to the torus or the Klein bottle, then either $K \equiv 0$ or K takes on both positive and negative values.

(c) If M is any other compact surface, then $K < 0$ somewhere.

Proof. Note that $\chi(S^2) = 2$, $\chi(\mathbb{RP}^2) = 1$, $\chi(T^2) = 0$, and other surfaces possess negative Euler characteristics. □

Riemann surface topology to curvature

In Corollary 2.3.7 we assumed we knew the topology of M and drew conclusions about the possible curvatures it could support. In the following corollary we reverse our point of view, and use assumptions about the curvature to draw conclusions about the topology of the manifold.

e to topology **Corollary 2.3.8.** *Let (M, g) be a compact Riemannian 2-manifold and K its Gaussian curvature.*

(a) If $K > 0$ everywhere on M, then the universal covering manifold of M is homeomorphic to S^2, and $\pi_1(M)$ is either trivial or isomorphic to the two element group $\mathbb{Z}/2\mathbb{Z}$.

(b) If $K \leq 0$ everywhere on M, then the universal covering manifold of M is homeomorphic to $\mathbb{R}^2$, and $\pi_1(M)$ is infinite.

Proof. Suppose first that M has positive Gaussian curvature. From the Gauss-Bonnet theorem, M has positive Euler characteristic. The classification theorem for compact surfaces shows that the only such surfaces are the sphere (with trivial fundamental group) and the projective plane (with fundamental group isomorphic to $\mathbb{Z}/2\mathbb{Z}$), both of which are covered by the sphere.

On the other hand, suppose M has nonpositive Gaussian curvature. Then its Euler characteristic is nonpositive, so it is either an orientable surface of genus $n \geq 1$ or a nonorientable one of genus $n \geq 2$. Thus the universal covering space of M is $\mathbb{R}^2$ if M is the torus or the Klein bottle, and $\mathbb{B}^2$ in all other cases, both of which are homeomorphic to $\mathbb{R}^2$. The fact that the universal covering space is noncompact implies that the universal covering map has infinitely many sheets, and therefore $\pi_1(M)$ is infinite. $\square$

2.4 Jacobi fields

Our goal for the remainder of this chapter is to generalize to higher dimensions some of the geometric and topological consequences of the Gauss-Bonnet theorem. We need to develop a new approach: instead of using Stokes's theorem and differential forms to relate the curvature to global topology as in the proof of the Gauss-Bonnet theorem, we study the way that curvature affects the behavior of nearby geodesics. Roughly speaking, positive curvature causes nearby geodesics to converge, while negative curvature causes them to spread out. In order to draw topological consequences from this fact, we need a quantitative way to measure the effect of curvature on a one-parameter family of geodesics.

We begin by deriving the Jacobi equation, which is an ordinary differential equation satisfied by the variation field of any one-parameter family of geodesics. A vector field satisfying this equation along a geodesic is called a **Jacobi field**. We then introduce conjugate points, which are pairs of points along a geodesic where some nontrivial Jacobi field vanishes. Intuitively, if p and q are conjugate along a geodesic, one expects to find a one-parameter family of geodesic segments that start at p and end (almost) at q.

After defining conjugate points, we prove a simple but essential fact: the points conjugate to p are exactly the points where $\exp_p$ fails to be a local diffeomorphism. We then derive an expression for the second derivative of the length functional with respect to proper variations of a geodesic, called the second variation formula. Using this formula, we prove another essential fact about conjugate points: once a geodesic passes its first conjugate point, it is no longer minimizing. The converse of this statement, however, is untrue: a geodesic can cease to be minimizing before it reaches its first conjugate point. Finally, we study the set of points where geodesics starting at a given point p cease to minimize, called the cut locus of p.

2.4.1 The Jacobi equation

Let (M, g) be an n-dimensional Riemannian or pseudo-Riemannian manifold. In order to study the effect of curvature on nearby geodesics, we focus on variations through geodesics. Suppose, therefore, that $I, K \subseteq \mathbb{R}$ are intervals, $\gamma : I \to M$ is a geodesic, and $\Gamma : K \times I \to M$ is a variation of γ. We say that Γ is a variation through geodesics if each of the main curves

$\Gamma_s(t) = \Gamma(s,t)$ is also a geodesic. (In particular, this requires that Γ be smooth.) Our first goal is to derive an equation that must be satisfied by the variation field of a variation through geodesics.

Theorem 2.4.1 (The Jacobi Equation). *Let (M, g) be a Riemannian or pseudo-Riemannian manifold, let γ be a geodesic in M, and let J be a vector field along γ. If J is the variation field of a variation through geodesics, then J satisfies the following equation, called the **Jacobi equation**:*

$$D_t^2 J + R(J, \gamma')\gamma' = 0. \tag{4.1}$$

Proof. The geodesic equation tells us that $D_t \partial_t \Gamma \equiv 0$ for all (s,t). We can take the covariant derivative of this equation with respect to s, yielding $D_s D_t \partial_t \Gamma \equiv 0$. Using Proposition 2.1.18 to commute the covariant derivatives along γ, we compute

$$0 = D_s D_t \partial_t \Gamma = D_t D_s \partial_t \Gamma + R(\partial_s \Gamma, \partial_t \Gamma)\partial_t \Gamma = D_t D_t \partial_s \Gamma + R(\partial_s \Gamma, \partial_t \Gamma)\partial_t \Gamma$$

where the last step follows from the symmetry lemma. Evaluating at $s = 0$, where $\partial_s \Gamma(0,t) = J(t)$ and $\partial_t \Gamma(0,t) = \gamma'(t)$, we get (2.4.1). $\square$

A smooth vector field along a geodesic that satisfies the Jacobi equation is called a **Jacobi field**. As the following proposition shows, the Jacobi equation can be written as a system of second-order linear ordinary differential equations, so it has a unique solution given initial values for J and $D_t J$ at one point.

Proposition 2.4.2. *Let (M, g) be a Riemannian or pseudo-Riemannian manifold. Suppose $I \subseteq \mathbb{R}$ is an interval, $\gamma : I \to \mathbb{R}$ is a geodesic, $a \in I$, and $p = \gamma(a)$. For every pair of vectors $v, w \in T_pM$, there is a unique Jacobi field J along γ satisfying the initial conditions*

$$J(a) = v, \quad D_t J(a) = w.$$

Proof. Choose a parallel orthonormal frame (E_i) along γ, and write $v = v^i E_i(a)$, $w = w^i E_i(a)$, and $\gamma'(t) = y^i(t) E_i(t)$ in terms of this frame. Writing an unknown vector field $J \in \mathfrak{X}(\gamma)$ as $J(t) = J^i(t) E_i(t)$, we can express the Jacobi equation as

$$\ddot{J}^l(t) + R_{ijk}^{l}(\gamma(t)) J^i(t) y^j(t) y^k(t) = 0.$$

This is a system of n linear second-order ODEs for the n functions $J^i : I \to \mathbb{R}$. Making the substitution $W^i = \dot{J}^i$ converts it to the following equivalent first-order linear system for the $2n$ unknown functions (J^i, W^i):

$$\begin{aligned} \dot{J}^i(t) &= W^i(t), \\ \dot{W}^i(t) &= -R_{ijk}^{l}(\gamma(t)) J^i(t) y^j(t) y^k(t). \end{aligned}$$

Then Theorem 1.2.24 guarantees the existence and uniqueness of a smooth solution on the whole interval I with arbitrary initial conditions $J^i(a) = v^i$, $W^i(a) = w^i$. Since $D_t J(a) = \dot{J}^i(a) E_i(a) = W^i(a) E_i(a) = w$, it follows that $J(t) = J^i(t) E_i(t)$ is the desired Jacobi field. $\square$

Given a geodesic γ, let $\mathfrak{J}(\gamma) \subseteq \mathfrak{X}(\gamma)$ denote the set of Jacobi fields along γ.

Corollary 2.4.3. *Suppose (M, g) is a Riemannian or pseudo-Riemannian manifold of dimension n, and γ is any geodesic in M. Then $\mathfrak{J}(\gamma)$ is a $2n$-dimensional linear subspace of $\mathfrak{X}(\gamma)$.*

Proof. Because the Jacobi equation is linear, $\mathfrak{J}(\gamma)$ is a linear subspace of $\mathfrak{X}(\gamma)$. Let $p = \gamma(a)$ be any point on γ, and consider the linear map from $J(\gamma)$ to $T_pM \times T_pM$ by sending J to $(J(a), D_t J(a))$. The preceding proposition says precisely that this map is bijective. $\square$

The following proposition is a converse to Theorem 2.4.1; it shows that each Jacobi field along a geodesic segment tells us how some family of geodesics behaves, at least to first order along γ.

Proposition 2.4.4. *Let (M, g) be a Riemannian or pseudo-Riemannian manifold, and let $\gamma : I \to M$ be a geodesic. If M is complete or I is a compact interval, then every Jacobi field along γ is the variation field of a variation of γ through geodesics.*

Proof. Let J be a Jacobi field along γ. After applying a translation in t (which does not affect either the fact that γ is a geodesic or the fact that J is a Jacobi field), we can assume that the interval I contains 0, and write $p = \gamma(0)$ and $v = \gamma'(0)$. Note that this implies $\gamma(t) = \exp_p(tv)$ for all $t \in I$.

Choose a smooth curve $\sigma : (-\varepsilon, \varepsilon) \to M$ and a smooth vector field V along σ satisfying

$$\begin{aligned} \sigma(0) &= p & V(0) &= v \\ \sigma'(0) &= J(0) & D_s V(0) &= D_t J(0) \end{aligned}$$

where D_s and D_t denote covariant differentiation along σ and γ, respectively. (They are easily constructed in local coordinates around p.) We wish to define a variation of γ by setting

$$\Gamma(s, t) = \exp_{\sigma(s)}(tV(s)). \tag{4.2}$$

If M is geodesically complete, this is defined for all $(s, t) \in (-\varepsilon, \varepsilon) \times I$. On the other hand, if I is compact, the fact that the domain of the exponential map is an open subset of TM that contains the compact set $\{(p, tv) : t \in I\}$ guarantees that there is some $\delta > 0$ such that $\Gamma(s, t)$ is defined for all $(s, t) \in (-\delta, \delta) \times I$. (This is the tube lemma.)

Note that

$$\Gamma(0, t) = \exp_{\sigma(0)}(tV(0)) = \exp_p(tv) = \gamma(t), \tag{4.3}$$

$$\Gamma(s, 0) = \exp_{\sigma(s)}(0) = \sigma(s). \tag{4.4}$$

In particular, (4.3) shows that Γ is a variation of γ. The properties of the exponential map guarantee that Γ is a variation through geodesics, and therefore its variation field $W(t) = \partial_s \Gamma(0, t)$ is a Jacobi field along γ.

Now, (4.4) implies

$$W(0) = \frac{\partial}{\partial s}\Big|_{s=0} \Gamma(s, 0) = \sigma'(0) = J(0).$$

If we can show that $D_t W(0) = D_t J(0)$ as well, it then follows from the uniqueness of Jacobi fields that $W \equiv J$, and the proposition is proved.

Formula (4.2) shows that each main curve $\Gamma_s(t)$ is a geodesic whose initial velocity is $V(s)$, so

$$\partial_t \Gamma(s, 0) = \frac{\partial}{\partial t}\Big|_{t=0} \Gamma_s(t) = V(s).$$

It follows from the symmetry lemma that $D_t \partial_s \Gamma = D_s \partial_t \Gamma$, and our choice of V gives

$$D_t W(0) = D_t \partial_s \Gamma(0, 0) = D_s \partial_t \Gamma(0, 0) = D_s V(0) = D_t J(0).$$

This finishes the proof. $\square$

Proposition 2.4.5 (Local Isometry Invariance of Jacobi Fields). *Suppose (M, g) and $(\widetilde{M}, \widetilde{g})$ are Riemannian or pseudo-Riemannian manifolds and $\varphi : M \to \widetilde{M}$ is a local isometry. Let $\gamma : I \to M$*

and $\widetilde{\gamma} : I \to \widetilde{M}$ be geodesics related by $\widetilde{\gamma} = \varphi \circ \gamma$, and let $J \in \mathfrak{X}(\gamma)$, $\widetilde{J} \in \mathfrak{X}(\widetilde{\gamma})$ be related by $d\varphi_{\gamma(t)}(J(t)) = \widetilde{J}(t)$ for all $t \in I$. Then J is a Jacobi field if and only if $\widetilde{J}$ is.

Proof. The curvature tensor and covariant derivatire are preserved by a local isometry, therefore the Jacobi equation is also preserved. □

2.4.2 Basic computations with Jacobi fields

There are various situations in which Jacobi fields can be computed explicitly. We begin by describing the most important of these.

Tangential and normal Jacobi fields

Along every geodesic $\gamma : I \to M$, there are always two Jacobi fields that we can write down immediately. Because $D_t\gamma' \equiv 0$ and $R(\gamma', \gamma')\gamma' \equiv 0$ by antisymmetry of R, the vector fields $J_0(t) = \gamma'(t)$ and $J_1(t) = t\gamma'(t)$ both satisfy the Jacobi equation by direct computation. If I is compact or M is complete, the vector field J_0 is the variation field of the variation $\Gamma(s,t) = \Gamma(s+t)$, while J_1 is the variation field of $\Gamma(s,t) = \gamma((1+s)t)$. Therefore, these two Jacobi fields just reflect the possible reparametrizations of γ, and do not tell us anything about the behavior of geodesics other than γ itself.

To distinguish these trivial cases from more informative ones, we make the following definitions. Given a regular curve $\gamma : I \to M$, for each $t \in I$ we let $T^{\top}_{\gamma(t)}M \subseteq T_{\gamma(t)}M$ denote the one-dimensional subspace spanned by $\gamma'(t)$, and $T^{\perp}_{\gamma(t)}M$ its orthogonal complement. A **tangential vector field along $\boldsymbol{\gamma}$** is a vector field $V \in \mathfrak{X}(\gamma)$ such that $V(t) \in T^{\top}_{\gamma(t)}M$ for all t, and a **normal vector field along $\boldsymbol{\gamma}$** is one such that $V(t) \in T^{\perp}_{\gamma(t)}M$ for all t. Thus a **normal Jacobi field along $\boldsymbol{\gamma}$** is a Jacobi field J satisfying $J(t) \perp \gamma'(t)$ for all t. Let $\mathfrak{X}^{\top}(\gamma)$ and $X^{\perp}(\gamma)$ denote the spaces of smooth normal and tangential vector fields along γ, respectively. When γ is a geodesic, $\mathfrak{J}^{\top}(\gamma)$ and $\mathfrak{J}^{\perp}(\gamma)$ denote the spaces of normal and tangential Jacobi fields along γ, respectively.

bi field iff **Proposition 2.4.6.** *Let (M, g) be a Riemannian or pseudo-Riemannian manifold. Suppose $\gamma : I \to M$ is a geodesic and J is a Jacobi field along γ. Then the following are equivalent:*

(a) J is a normal Jacobi field.

(b) J is orthogonal to γ' at two distinct points.

(c) Both J and D_tJ are orthogonal to γ' at one point.

(d) Both J and D_tJ are orthogonal to γ' everywhere along γ.

Proof. Define a function $f : I \to \mathbb{R}$ by $f(t) = \langle J(t), \gamma'(t)\rangle$, so that $f(t) \equiv 0$ if and only if $J(t) \perp \gamma'(t)$. Using compatibility with the metric and the fact that $D_t\gamma'(t) \equiv 0$, we compute

$$f''(t) = \langle D_t^2 J, \gamma'\rangle = -\langle R(J, \gamma')\gamma', \gamma'\rangle = -R(J, \gamma', \gamma', \gamma') = 0$$

by the symmetries of the curvature tensor. Thus, by elementary calculus, f is an affine function of t.

Note that $f'(t) = \langle D_tJ(t), \gamma'(t)\rangle$, which vanishes at t if and only if $D_tJ(t) \perp \gamma'(t)$. It follows that $J(a)$ and $D_tJ(a)$ are orthogonal to $\gamma'(a)$ for some $a \in I$ if and only if f and its first derivative vanish at a, which happens if and only if $f \equiv 0$. Similarly, J is orthogonal to γ' at two points if and only if f vanishes at two points, which happens if and only if f is identically zero. If this is the case, then $f' \equiv 0$ as well, which implies that both J and $D_t J$ are orthogonal to γ' for all t. □

Corollary 2.4.7. *Suppose (M, g) is a Riemannian or pseudo-Riemannian n-manifold and $\gamma : I \to M$ is any nonconstant geodesic. Then $\mathfrak{J}^{\perp}(\gamma)$ is a $(2n-2)$-dimensional subspace of $\mathfrak{J}(\gamma)$, and $J^{\top}(\gamma)$ is a 2-dimensional subspace. Every Jacobi field can be uniquely decomposed as a sum of a tangential Jacobi field plus a normal Jacobi field.*

Proof. As we noted in the proof of Corollary 2.4.3, for every $a \in I$, the map from $\mathfrak{J}((\gamma)$ to $T_{\gamma(a)}M \oplus T_{\gamma(a)}M$ given by $J \mapsto (J(a), D_t J(a))$ is an isomorphism, and Proposition 2.4.6 shows that $\mathfrak{J}^{\perp}(\gamma)$ is exactly the preimage of the subspace consisting of all pairs $(v, w) \in T_{\gamma(a)}M \oplus T_{\gamma(a)}M$ such that $\langle v, \gamma'(a)\rangle = \langle w, \gamma'(a)\rangle = 0$. Because this subspace has dimension $2n-2$, it follows that $\mathfrak{J}^{\perp}(\gamma)$ has the same dimension.

On the other hand, $\mathfrak{J}^{\top}(\gamma)$ contains $J_0(t) = \gamma'(t)$ and $J_1(t) = t\gamma'(t)$, which are linearly independent over $\mathbb{R}$ because $\gamma'(t)$ never vanishes, so it is a subspace of dimension at least 2. Because $\mathfrak{J}^{\top}(\gamma) \cap \mathfrak{J}^{\perp}(\gamma) = \{0\}$, the dimension of $\mathfrak{J}^{\top}(\gamma)$ must be exactly 2, and we have a direct sum decomposition $\mathfrak{J}(\gamma) = \mathfrak{J}^{\top}(\gamma) \oplus \mathfrak{J}^{\perp}(\gamma)$. This implies that every $J \in \mathfrak{J}(\gamma)$ has a unique decomposition $J = J^{\top} + J^{\perp}$, with $J^{\top} \in \mathfrak{J}^{\top}(\gamma)$ and $J^{\perp} \in \mathfrak{J}^{\perp}(\gamma)$. □

Jacobi fields vanishing at a point

For many purposes, we will be primarily interested in Jacobi fields that vanish at a particular point. For these, there are some useful explicit formulas.

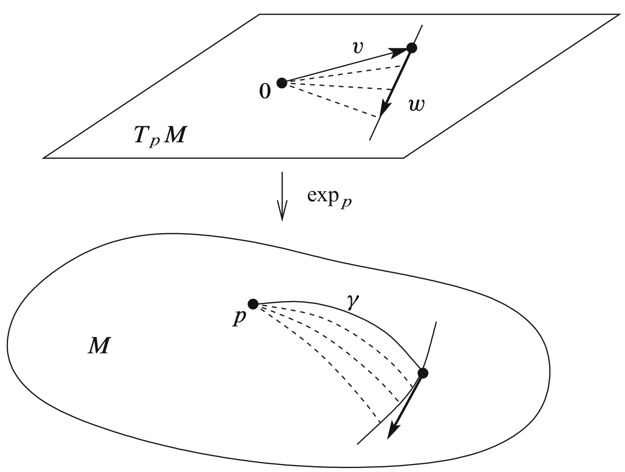

Figure 2.1: The variation of Lemma 2.4.8.

Lemma 2.4.8. *Let (M, g) be a Riemannian or pseudo-Riemannian manifold, $I \subseteq \mathbb{R}$ an interval containing 0, and $\gamma : I \to M$ a geodesic. Suppose $J : I \to M$ is a Jacobi field such that $J(0) = 0$. If M is geodesically complete or I is compact, then J is the variation field of the following variation of γ through geodesics:*

$$\Gamma(s, t) = \exp_p(t(v + sw)). \tag{4.5}$$

where $p = \gamma(0)$, $v = \gamma'(0)$, *and* $w = D_t J(0)$.

Proof. The proof of Proposition 2.4.4 showed that J is the variation field of a variation Γ of the form (4.2), with σ any smooth curve satisfying $\sigma(0) = p$ and $\sigma'(0) = 0$, and V a smooth vector field along σ with $V(0) = v$ and $D_s V(0) = w$. In this case, we can take $\sigma(s) \equiv p$ and $V(s) = v + sw \in T_pM$, yielding (4.5). □

This result leads to some explicit formulas for all of the Jacobi fields vanishing at a point.

Proposition 2.4.9 (Jacobi Fields Vanishing at a Point). *Let* (M, g) *be a Riemannian or pseudo-Riemannian n-manifold and* $p \in M$. *Suppose* $\gamma : I \to M$ *is a geodesic such that* $0 \in I$ *and* $\gamma'(0) = p$. *For every* $w \in T_pM$, *the Jacobi field* J *along* γ *such that* $J(0) = 0$ *and* $D_t J(0) = w$ *is given by*

$$J(t) = d(\exp_p)_{tv}(tw), \tag{4.6}$$

where $v = \gamma'(0)$, *and we regard* tw *as an element of* $T_{tv}(T_pM)$ *by means of the canonical identification* $T_{tv}(T_pM) \cong T_pM$. *If* (x^i) *are normal coordinates on a normal neighborhood of* p *containing the image of* γ, *then* J *is given by the formula*

$$J(t) = tw^i \partial_i|_{\gamma(t)}, \tag{4.7}$$

where $w^i \partial_i|_0$ *is the coordinate representation of* w.

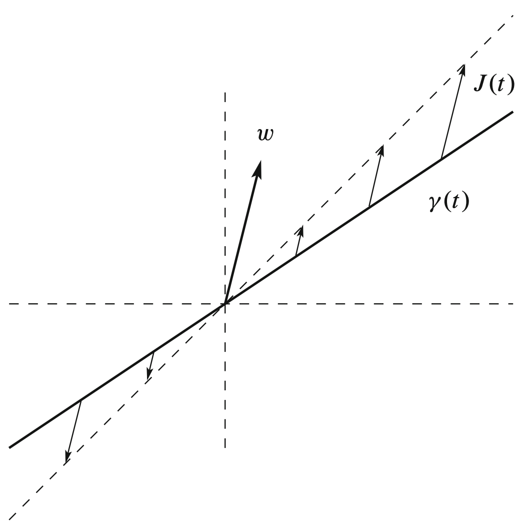

Figure 2.2: A Jacobi field in normal coordinates.

Proof. Under the given hypotheses, Lemma 2.4.8 showed that the restriction of J to any compact interval containing 0 is the variation field of a variation Γ through geodesics of the form (4.5). Using the chain rule to compute $J(t) = \partial_s \Gamma(0, t)$, we arrive at (4.6). Because every t in the domain of Γ is contained in some such compact interval, the formula holds for all such t.

In normal coordinates, the coordinate representation of the exponential map is the identity, so Γ can be written explicitly in coordinates as

$$\Gamma(s, t) = (t(v^1 + sw^1), \ldots, t(v^n + sw^n)).$$

Differentiating $\Gamma(s,t)$ with respect to s and setting $s=0$ shows that its variation field J is given by (4.7). $\square$

Corollary 2.4.10. *Suppose (M,g) is a Riemannian or pseudo-Riemannian manifold and U is a normal neighborhood of $p \in M$. For each $q \in U \setminus \{p\}$, every vector in T_qM is the value of a Jacobi field J along a radial geodesic such that J vanishes at p.*

Proof. Let (x^i) be normal coordinates on U. Given $q = (q^1, \dots, q^n) \in U \setminus \{p\}$ and $w = w^i w^i \partial_i|_q \in T_qM$, the curve $\gamma(t) = (tq^1, \dots, tq^n)$ is a radial geodesic satisfying $\gamma(0) = p$ and $\gamma(1) = q$. The previous proposition showed that $J(t) = tw^i \partial_i|_{\gamma(t)}$ is a Jacobi field along γ. Because $J(0) = 0$ and $J(1) = w$, the result follows. $\square$

Jacobi fields in constant-curvature spaces

For metrics with constant sectional curvature, we have a different kind of explicit formula for Jacobi fields—this one expresses a Jacobi field as a scalar multiple of a parallel vector field. To handle the various cases concisely, for each $c \in \mathbb{R}$, let us define a function $s_c : \mathbb{R} \to \mathbb{R}$ by

$$s_c(t) = \begin{cases} t, & \text{if } c = 0; \\ R \sin \dfrac{t}{R}, & \text{if } c = \dfrac{1}{R^2} > 0; \\ R \sinh \dfrac{t}{R}, & \text{if } c = -\dfrac{1}{R^2} < 0. \end{cases} \tag{4.8}$$

Proposition 2.4.11 (Jacobi Fields in Constant Curvature). *Suppose (M,g) is a Riemannian manifold with constant sectional curvature c, and γ is a unit-speed geodesic in M. The normal Jacobi fields along γ vanishing at $t=0$ are the vector fields of the form*

$$J(t) = ks_c(t)E(t), \tag{4.9}$$

wher E is any parallel unit normal vector field along γ, k is an arbitrary constant, and s_c is defined by (4.8). The initial derivative of such a Jacobi field is

$$D_tJ(0) = kE(0). \tag{4.10}$$

and its norm is

$$|J(t)| = |s_c(t)||D_tJ(0)|. \tag{4.11}$$

Proof. Since g has constant curvature, its curvature endomorphism is given by the formula of Proposition 2.2.28:

$$R(x,y)v = c(\langle y, v\rangle x - \langle x, v\rangle y).$$

Substituting this into the Jacobi equation, we find that a normal Jacobi field J satisfies

$$0 = D_t^2 J + c(\langle \gamma', \gamma'\rangle J - \langle J, \gamma'\rangle \gamma') = D_t^2 J + cJ, \tag{4.12}$$

where we have used the facts that $|\gamma'| = 1$ and $\langle J, \gamma'\rangle = 0$.

Since (4.12) says that the second covariant derivative of J is a multiple of J itself, it is reasonable to try to construct a solution by choosing an arbitrary parallel unit normal vector field E along γ and setting $J(t) = u(t)E(t)$ for some function u to be determined. Plugging this into (4.12), we find that J is a Jacobi field if and only if u is a solution to the differential equation

$$u''(t) + cu(t) = 0.$$

It is an easy matter to solve this ODE explicitly. In particular, the solutions satisfying $u(0) = 0$ are constant multiples of s_c. This construction yields all the normal Jacobi fields vanishing at 0, since there is an $(n-1)$-dimensional space of them, and the space of parallel normal vector fields has the same dimension.

To prove the last two statements, suppose J is given by (4.9), and compute

$$D_t J(0) = k s_c'(0) E(0) = k E(0).$$

since $s_c'(0) = 1$ in every case. Because E is a unit vector field, $|D_t J(0)| = |k|$ and (4.10) follows. $\square$

Here is our first significant application of Jacobi fields. Because every tangent vector in a normal neighborhood is the value of a Jacobi field vanishing at the origin by Corollary 2.4.10, Proposition 2.4.11 yields explicit formulas for constant curvature metrics in normal coordinates. To set the stage, we will rewrite the Euclidean metric on $\mathbb{R}^n$ in a form that is somewhat more convenient for these computations.

Let $\pi : \mathbb{R}^n \setminus \{0\} \to S^{n-1}$ be the radial projection

$$\pi(x) = \frac{x}{|x|}, \tag{4.13}$$

and define a symmetric 2-tensor field on $\mathbb{R}^n \setminus \{0\}$ by

$$\widehat{g} = \pi^* \mathring{g}, \tag{4.14}$$

where $\mathring{g}$ is the round metric on S^{n-1}.

Lemma 2.4.12. *On $\mathbb{R}^n \setminus \{0\}$, the metric $\widehat{g} = \pi^*\mathring{g}$ and the Euclidean metric $\bar{g}$ are related by*

$$\bar{g} = dr^2 + r^2 \widehat{g}. \tag{4.15}$$

Proof. Example **??** observed that the map $\varPhi : \mathbb{R}^+ \times S^{n-1} \to \mathbb{R}^n \setminus \{0\}$ given by

$$\varPhi(\rho, \omega) = \rho\omega. \tag{4.16}$$

is an isometry when $\mathbb{R}^+ \times S^{n-1}$ has the warped product metric $d\rho^2 \oplus \rho^2 \mathring{g}$ and $\mathbb{R}^n \setminus \{0\}$ has the Euclidean metric. Because $\varPhi^{-1}(x) = (r(x), \pi(x))$, this means that $\bar{g} = (\varPhi^{-1})^*(d\rho^2 \oplus \rho^2 \mathring{g}) = dr^2 + r^2 \widehat{g}$. $\square$

Theorem 2.4.13 (Constant-Curvature Metrics in Normal Coordinates). *Suppose (M, g) is a Riemannian manifold with constant sectional curvature c. Given $p \in M$, let be normal coordinates on a normal neighborhood U of p; let r be the radial distance function on U defined by (4.3); and let $\widehat{g}$ be the symmetric 2-tensor defined in x-coordinates by (4.14). On $U \setminus \{p\}$, the metric g can be written*

$$g = dr^2 + s_c(r)^2 \widehat{g}, \tag{4.17}$$

where s_c is defined by (4.8).

Proof. Let $\bar{g}$ denote the Euclidean metric in x-coordinates, and let g_c denote the metric defined by the formula on the right-hand side of (4.17). By the properties of normal coordinates, at points of $U \setminus \{p\}$, all three metrics g, $\bar{g}$, and g_c make the radial vector field ∂_r a unit vector orthogonal to the level sets of r. Thus we need only show that $g(w_1, w_2) = g_c(w_1, w_2)$ when w_1, w_2 are tangent to a level set of r, and by polarization it suffices to show that $g(w, w) = g_c(w, w)$ for every such vector w. Note that if w is tangent to a level set $r = b$, then formulas (4.17) and (4.15) imply

$$g_c(w,w) = s_c(b)\widehat{g}(w,w) = \frac{s_c(b)^2}{b^2}\overline{g}(w,w).$$

Let $q \in U \setminus \{p\}$ and $w \in T_qM$, and assume that w is tangent to the r-level set containing q. Let $b = d_g(p,q)$, and let $\gamma : [0,b] \to U$ be the unit-speed radial geodesic from p to q, so the coordinate representation of γ is

$$\gamma(t) = \Big(\frac{t}{b}q^1, \dots, \frac{t}{q}q^n\Big).$$

where $(q^1, \dots, q^n)$ is the coordinate representation of q. Let $J \in \mathfrak{X}(\gamma)$ be the vector field along γ given by

$$J(t) = \frac{t}{b}w^i\partial_i|_{\gamma(t)}, \tag{4.18}$$

where $w^i\partial_i|_q$ is the coordinate representation for w. By Proposition 2.4.9, J is a Jacobi field satisfying $D_tJ(0) = (1/b)w^i\partial_i|_p$, and it follows from the definition that $J(b) = w$. Because J is orthogonal to γ' at p and q, it is normal by Proposition 2.4.6. Thus by Proposition 2.4.11,

$$|w|_g^2 = |J(b)|_g^2 = s_c(b)^2|D_tJ(0)|_g^2 = s_c(b)^2\frac{1}{b^2}|w^i\partial_i|_p|_g^2 = s_c(b)^2\frac{1}{b^2}|w|_{\overline{g}}^2 = |w|_{g_c}^2.$$

where we use the fact that $g = \overline{g}$ at the point p. □

Corollary 2.4.14 (Local Uniqueness of Constant-Curvature Metrics). *Let (M,g) and $(\widetilde{M},\widetilde{g})$ be Riemannian manifolds of the same dimension with constant sectional curvature c. For all points $p \in M$, $\widetilde{p} \in \widetilde{M}$ there exist neighborhoods U of p and $\widetilde{U}$ of $\widetilde{p}$ and an isometry $\varphi : U \to \widetilde{U}$.*

Proof. Choose $p \in M$ and $\widetilde{p} \in \widetilde{M}$, and let U and $\widetilde{U}$ be geodesic balls of small radius ε around p and $\widetilde{p}$, respectively. Riemannian normal coordinates give maps $\psi : U \to B_\varepsilon \subset \mathbb{R}^n$ and $\widetilde{\psi} : \widetilde{U} \to \mathbb{R}^n$, under which both metrics are given by formula (4.17) on the complement of the origin. At the origin, $g_{ij} = \widetilde{g}_{ij} = \delta_{ij}$. Therefore $\widetilde{\psi}^{-1} \circ \psi$ is the required local isometry. □

Corollary 2.4.15 (Constant-Curvature Metrics as Warped Products). *Suppose (M,g) is a Riemannian manifold with constant sectional curvature c, and U is a geodesic ball of radius b centered at $p \in M$. Then $U \setminus \{p\}$ is isometric to a warped product of the form $(0,b) \times_{s_c} S^{n-1}$, where $(0,b) \subseteq \mathbb{R}$ has the Euclidean metric, and S^{n-1} is the unit sphere with the round metric $\mathring{g}$.*

Proof. By virtue of Theorem 2.4.13, we may consider g to be a metric on the ball of radius b in $\mathbb{R}^n$ given by formula (4.17). Let $\Phi : (0,b) \times S^{n-1} \to U \setminus \{p\}$ and $\pi : \mathbb{R}^n \setminus \{0\} \to S^{n-1}$ be the maps defined by (4.16) and (4.13). Because $\pi \circ \Phi$ restricts to the identity on $\{\rho\} \times S^{n-1}$ for each fixed ρ, it follows that $\Phi^*\widehat{g} = \Phi^*\pi^*\mathring{g} = \mathring{g}$, and thus

$$\Phi^*\rho = d\rho^2 \oplus s_c(\rho)^2\mathring{g}$$

as desired. □

Corollary 2.4.16 (Polar Decomposition of Integrals). *Suppose (M,g) is a Riemannian manifold with constant sectional curvature c, and U is an open or closed geodesic ball of radius b around a point $p \in M$. If $f : U \to \mathbb{R}$ is any bounded integrable function, then the integral of f over U can be expressed as*

$$\int_U f\,dV_g = \int_{S^{n-1}}\int_0^b f\circ\Phi(\rho,\omega)s_c(\rho)^{n-1}d\rho dV_{\mathring{g}},$$

where dV_g is the Riemannian density of g, and $\Phi : (0, b) \times S^{n-1} \to U \setminus \{p\}$ is defined in normal coordinates by (4.16).

Proof. Because every geodesic ball is orientable, we might as well choose an orientation on U and interpret dV_g as a differential form. Since the boundary of a geodesic ball has measure zero, it does not matter whether U is open or closed. Similarly, integrating over $U \setminus \{p\}$ instead of U does not change the value of the integral. The claim therefore follows from Corollary 2.4.15 together with the fact that the volume form of the warped product metric $d\rho^2 \oplus s_c(\rho)^2 \mathring{g}$ can be written $s_c(\rho)^{n-1} d\rho \wedge dV_{\mathring{g}}$. □

2.4.3 Conjugate points

Our next application of Jacobi fields is to study the question of when the exponential map is a local diffeomorphism.

Suppose (M, g) is a Riemannian or pseudo-Riemannian manifold and $p \in M$. The restricted exponential map $\exp_p$ is defined on an open subset $\mathcal{E}_p \subseteq T_pM$, and because it is a smooth map between manifolds of the same dimension, the inverse function theorem guarantees that it is a local diffeomorphism in a neighborhood of each of its regular points (points $v \in T_pM$ where $d(\exp_p)_v$ is surjective and thus invertible). To see where this fails, we need to identify the critical points of $\exp_p$ (the points where its differential is singular). Proposition 1.3.20(d) guarantees that 0 is a regular point, but it may well happen that it has critical points elsewhere in $\mathcal{E}_p$.

An enlightening example is provided by the sphere $S^n(R)$. All geodesics starting at a given point p meet at the antipodal point, which is at a distance of πR along each geodesic. The exponential map is a diffeomorphism on the ball $B_{\pi R}(0) \subseteq T_pS^n$, but every point on the boundary of that ball is a critical point. Moreover, Proposition 2.4.13 shows that each Jacobi field on $S^n(R)$ vanishing at p has its first zero precisely at distance πR.

On the other hand, formula (4.6) shows that if U is a normal neighborhood of p (the image of a star-shaped open set on which $\exp_p$ is a diffeomorphism), then no Jacobi field that vanishes at p can vanish at any other point in U. We might thus be led to expect a relationship between zeros of Jacobi fields and critical points of the exponential map.

Let (M, g) be a Riemannian or pseudo-Riemannian manifold, $\gamma : I \to M$ a geodesic, and $p = \gamma(a)$, $q = \gamma(b)$ for some $a, b \in I$. We say that p and q are conjugate along γ if there is a Jacobi field along γ vanishing at $t = a$ and $t = b$ but not identically zero. The **order** (or **multiplicity**) of conjugacy is the dimension of the space of Jacobi fields vanishing at a and b. From the existence and uniqueness theorem for Jacobi fields, there is an n-dimensional space of Jacobi fields that vanish at a; since tangential Jacobi fields vanish at most at one point (dimensional consideration), the order of conjugacy of two points along γ can be at most $n - 1$. This bound is sharp: Proposition 2.4.11 shows that if γ is a geodesic joining antipodal points p and q on $S^n(R)$, then there is a Jacobi field vanishing at p and q for each parallel normal vector field along γ; thus in that case p and q are conjugate to order exactly $n - 1$.

The most important fact about conjugate points is that they are the images of critical points of the exponential map, as the following proposition shows.

Proposition 2.4.17. *Suppose (M, g) is a Riemannian or pseudo-Riemannian manifold, $p \in M$, and $v \in \mathcal{E}_p \subseteq T_pM$. Let $\gamma = \gamma_v : I \to M$ be the geodesic segment $\gamma(t) = \exp_p(tv)$, and let $q = \gamma(1) =$*

$\exp_p(v)$. *Then v is a critical point of $\exp_p$ if and only if q is conjugate to p along γ.*

Proof. Suppose first that v is a critical point of $\exp_p$. Then there is a nonzero vector $w \in T_v(T_pM)$ such that $d(\exp_p)_v(w) = 0$. Because T_pM is a vector space, we can identify $T_v(T_pM)$ with T_pM as usual and regard w as a vector in T_pM. Let Γ be the variation of γ defined by (4.5), and let $J(t) = \partial_s\Gamma(0,t)$ be its variation field. We can compute $J(1)$ as follows:

$$J(1) = \partial_s\Gamma(0,1) = \frac{\partial}{\partial s}\Big|_{s=0}\Gamma(s,1) = \frac{\partial}{\partial s}\Big|_{s=0}\exp_p(v+sw) = d(\exp_p)_v(w) = 0.$$

Thus J is a nontrivial Jacobi field vanishing at $t=0$ and $t=1$, so q is conjugate to p along γ.

Conversely, if q is conjugate to p along γ, then there is some nontrivial Jacobi field J along γ such that $J(0) = 0$ and $J(1) = 0$. Lemma 2.4.8 shows that J is the variation field of a variation of γ of the form (4.5) with $w = D_tJ(0) \in T_pM$, and the computation in the preceding paragraph shows that $d(\exp_p)_v(w) = J(1) = 0$. Thus v is a critical point for $\exp_p$. □

As Proposition 2.4.2 shows, the "natural" way to specify a unique Jacobi field is by giving its initial value and initial derivative. However, in Corollary 2.4.10 and Proposition 2.4.17, we had to construct Jacobi fields along a geodesic satisfying $J(0) = 0$ and $J(1) = w$ for some specific vector w. More generally, one can pose the two-point boundary problem for Jacobi fields: given $v \in T_{\gamma(a)}M$ and $w \in T_{\gamma(b)}M$, find a Jacobi field J along γ such that $J(a) = v$ and $J(b) = w$. Another interesting property of conjugate points is that they are the obstructions to solving the two-point boundary problem, as the next proposition shows.

Proposition 2.4.18 (The Two-Point Boundary Problem for Jacobi Fields). *Suppose (M,g) is a Riemannian or pseudo-Riemannian manifold, and $\gamma : [a,b] \to M$ is a geodesic segment. The two-point boundary problem for Jacobi fields along γ is uniquely solvable for every pair of vectors $v \in T_{\gamma(a)}M$ and $w \in T_{\gamma(b)}M$ if and only if $\gamma(a)$ and $\gamma(b)$ are not conjugate along γ.*

Proof. It is clear that if $\gamma(a)$ and $\gamma(b)$ are conjugate then the solution is not unique. Now we assume the converse and prove the existence and uniqueness. □

2.4.4 The second variation formula

Our next task is to study the question of which geodesic segments are minimizing. In the remainder of the section, because of the complications involved in studying lengths on pseudo-Riemannian manifolds, we restrict our attention to the Riemannian case.

In our proof that every minimizing curve is a geodesic, we imitated the first derivative test of elementary calculus: if a geodesic γ is minimizing, then the first derivative of the length functional must vanish for every proper variation of γ. Now we imitate the second-derivative test: if γ is minimizing, the second derivative must be nonnegative. First, we must compute this second derivative. In keeping with classical terminology, we call it the **second variation** of the length functional.

Theorem 2.4.19 (Second Variation Formula). *Suppose (M,g) is a Riemannian manifold. Let $\gamma : [a,b] \to M$ be a unit-speed geodesic segment, $\Gamma : J \times [a,b] \to M$ a proper variation of γ, and V its variation field. The second variation of $L_g(\Gamma_s)$ is given by the following formula:*

$$\frac{d^2}{ds^2}\Big|_{s=0} L_g(\Gamma_s) = \int_a^b \Big(|D_tV^\perp|^2 - R(V^\perp,\gamma',\gamma',V^\perp)\Big)dt, \tag{4.19}$$

where $V^\perp$ is the normal component of V.

Proof. As usual, let $(a_0, \dots, a_k)$ be an admissible partition for γ. We begin, as we did when computing the first variation formula, by restricting to a rectangle $J \times [a_{i-1}, a_i]$ where γ is smooth. From Theorem 1.4.3 we have, for every s,

$$\frac{d}{ds} L_g(\Gamma_s|_{a_{i-1},a_i}) = \int_{a_{i-1}}^{a_i} \frac{\langle D_t \partial_s \Gamma, \partial_t \Gamma \rangle}{|\partial_t \Gamma|} dt.$$

Differentiating again with respect to s, and using the symmetry lemma and Proposition 2.1.18, we obtain

$$\begin{aligned}
&\frac{d^2}{ds^2} L_g(\Gamma_s|_{[a_{i-1},a_i]}) \\
&= \int_{a_{i-1}}^{a_i} \Big(\frac{\langle D_s D_t \partial_s \Gamma, \partial_t \Gamma \rangle}{|\partial_t \Gamma|} + \frac{\langle D_t \partial_s \Gamma, D_s \partial_t \Gamma \rangle}{|\partial_t \Gamma|} - \frac{\langle D_t \partial_s \Gamma, \partial_t \Gamma \rangle \cdot \langle D_s \partial_t \Gamma, \partial_t \Gamma \rangle}{|\partial_t \Gamma|^3} \Big) \\
&= \int_{a_{i-1}}^{a_i} \Big(\frac{\langle D_t D_s \partial_s \Gamma + R(\partial_s \Gamma, \partial_t \Gamma) \partial_s \Gamma, \partial_t \Gamma \rangle}{|\partial_t \Gamma|} + \frac{\langle D_t \partial_s \Gamma, D_t \partial_s \Gamma \rangle}{|\partial_t \Gamma|} - \frac{\langle D_t \partial_s \Gamma, \partial_t \Gamma \rangle^2}{|\partial_t \Gamma|^3} \Big).
\end{aligned}$$

Now restrict to $s = 0$, where $|T| = 1$:

$$\begin{aligned}
&\frac{d^2}{ds^2}\Big|_{s=0} L_g(\Gamma_s|_{[a_{i-1},a_i]}) \\
&= \int_{a_{i-1}}^{a_i} \big(\langle D_t D_s \partial_s \Gamma, \partial_t \Gamma \rangle - R(\partial_s \Gamma, \partial_t \Gamma, \partial_t \Gamma, \partial_s \Gamma) + |D_t \partial_s \Gamma|^2 - \langle D_t \partial_s \Gamma, \partial_t \Gamma \rangle^2) \big) dt \Big|_{s=0}.
\end{aligned} \tag{4.20}$$

Because $D_t \partial_t \Gamma = D_t \gamma' = 0$ when $s = 0$, the first term in (10.20) can be integrated as follows:

$$\int_{a_{i-1}}^{a_i} \langle D_t D_s \partial_s \Gamma, \partial_t \Gamma \rangle dt = \int_{a_{i-1}}^{a_i} \frac{d}{dt} \langle D_s \partial_s \Gamma, \partial_t \Gamma \rangle dt = \langle D_s \partial_s \Gamma, \partial_t \Gamma \rangle \Big|_{t=a_{i-1}}^{t=a_i}. \tag{4.21}$$

Notice that $\partial_s \Gamma(s,t) = 0$ for all s at the endpoints $t = a_0 = a$ and $t = a_k = b$ because Γ is a proper variation, so $D_s \Gamma = 0$ there. Moreover, along the boundaries $\{t = a_i\}$ of the smooth regions, $D_s \partial_s \Gamma = D_s(\partial_s \Gamma)$ depends only on the values of Γ when $t = a_i$, and it is smooth up to the line $\{t = a_i\}$ from both sides; therefore $D_s \partial_s \Gamma$ is continuous for all (s,t). Thus when we insert (4.21) into (4.20) and sum over i, the boundary contributions from the first term all cancel, and we get

$$\begin{aligned}
\frac{d^2}{ds^2}\Big|_{s=0} L_g(\Gamma_s) &= \int_a^b \big(|D_t \partial_s \Gamma|^2 - \langle D_t \partial_s \Gamma, \partial_t \Gamma \rangle^2 - R(\partial_s \Gamma, \partial_t \Gamma, \partial_t \Gamma, \partial_s \Gamma) \big) dt \Big|_{s=0}. \\
&= \int_{a_{i-1}}^{a_i} \big(|D_t V|^2 - \langle D_t V, \gamma' \rangle^2 - R(V, \gamma', \gamma', V) \big) dt
\end{aligned} \tag{4.22}$$

Every vector field V along γ can be written uniquely as $V = V^\top + V^\perp$, where $V^\top$ is tangential and $V^\perp$ is normal. Explicitly,

$$V^\top = \langle V, \gamma' \rangle \gamma', \quad V^\perp = V - V^\top.$$

Because $D_t \gamma' = 0$, it follows that

$$D_t(V^\top) = \langle D_t V, \gamma' \rangle \gamma' = (D_t V)^\top, \quad D_t(V^\perp) = D_t V - D_t(V^\top) = (D_t V)^\perp.$$

Therefore,

$$|D_t V|^2 = |(D_t V)^\top|^2 + |(D_t V)^\perp|^2 = \langle D_t V, \gamma' \rangle^2 + |D_t V^\perp|^2.$$

Also, the fact that $R(\gamma', \gamma', \cdot, \cdot) = R(\cdot, \cdot, \gamma', \gamma') = 0$ implies

$$R(V, \gamma', \gamma', V) = R(V^\perp, \gamma', \gamma', V^\perp).$$

Substituting these relations into (4.22) gives (4.19). $\square$

It should come as no surprise that the second variation depends only on the normal component of V, because the tangential component of V contributes only to a reparametrization of γ, and length is independent of parametrization. For this reason, we will generally restrict our attention to variations of the following type: if γ is an admissible curve, a variation of γ is called a normal variation if its variation field is a normal vector field along γ. Given a geodesic segment $\gamma : [a, b] \to M$, we define a symmetric bilinear form I, called the **index form of γ**, on the space of normal vector fields along γ by

$$I(V, W) = \int_a^b \big(\langle D_t V, D_t W\rangle - R(V, \gamma', \gamma', W)\big)\, dt. \tag{4.23}$$

You should think of $I(V, W)$ as a sort of Hessian or second derivative of the length functional. Because every proper normal vector field along γ is the variation field of some proper normal variation, the preceding theorem can be rephrased in terms of the index form in the following way.

Corollary 2.4.20. *Suppose (M, g) is a Riemannian manifold. Let $\gamma : [a, b] \to M$ be a unit-speed geodesic, γ a proper normal variation of γ, and V its variation field. The second variation of $L_g(\Gamma_s)$ is $I(V, W)$. If γ is minimizing, then $I(V, V) \geq 0$ for every proper normal vector field along γ.*

The next proposition gives another expression for I, which makes the role of the Jacobi equation more evident.

Proposition 2.4.21. *Let (M, g) be a Riemannian manifold and let $\gamma : [a, b] \to M$ be a geodesic segment. For every pair of piecewise smooth normal vector fields V, W along γ,*

$$I(V, W) = -\int_a^b \langle D_t^2 V + R(V, \gamma')\gamma', W\rangle dt + \langle D_t V, W\rangle\Big|_{t=a}^{t=b} - \sum_{i=1}^{k-1} \langle \Delta_i D_t V, W(a_i)\rangle \tag{4.24}$$

where $(a_0, \ldots, a_k)$ is an admissible partition for V and W, and $\Delta_i D_t V$ is the jump in $D_t V$ at $t = a_i$.

Proof. On every subinterval $[a_{i-1}, a_i]$ where V and W are smooth,

$$\frac{d}{dt}\langle D_t, W\rangle = \langle D_t^2 V, W\rangle + \langle D_t V, D_t W\rangle.$$

Thus, by the fundamental theorem of calculus,

$$\int_{a_{i-1}}^{a_i} \langle D_t V, D_t W\rangle dt = -\int_{a_{i-1}}^{a_i} \langle D_t^2 V, W\rangle dt + \langle D_t V, W\rangle\Big|_{a_{i-1}}^{a_i}.$$

Summing over i, and noting that W is continuous at $t = a_i$ for $i = 1, \ldots, k-1$, we get (4.24). $\square$

Corollary 2.4.22. *If γ is a geodesic segment and V is a proper normal piecewise smooth vector field along γ, then $I(V, W) = 0$ for every proper normal piecewise smooth vector field W along γ if and only if V is a Jacobi field.*

Proof. The proof is similar to Proposition 1.4.4. $\square$

Geodesics do not minimize past conjugate points

We can use the second variation formula to prove another extremely important fact about conjugate points: no geodesic is minimizing past its first conjugate point. The geometric intuition is as follows. Suppose $\gamma : [a, c] \to M$ is a minimizing geodesic segment, and $\gamma(b)$ is conjugate to $\gamma(a)$ along γ for some $a < b < c$. If J is a Jacobi field along γ that vanishes at $t = a$ and $t = b$, then there is a variation of γ through geodesics, all of which start at $\gamma(a)$. Since $J(b) = 0$, we can expect them to end "almost" at $\gamma(b)$. If they really did all end at $\gamma(b)$, we could construct a broken geodesic by following some Γ_s from $\gamma(a)$ to $\gamma(b)$ and then following γ from $\gamma(b)$ to $\gamma(c)$, which would have the same length and thus would also be a minimizing curve. But this is impossible: as the proof of Theorem Riemann minimizing is geodesic 1.4.4 shows, a broken geodesic can always be shortened by rounding the corner.

The problem with this heuristic argument is that there is no guarantee that we can construct a variation through geodesics that actually end at $\gamma(b)$. The proof of the following theorem is based on an "infinitesimal" version of rounding the corner to obtain a shorter curve.

Given a geodesic segment $\gamma : [a, c] \to M$, we say that **γ has a conjugate point** if there is some $b \in (a, c]$ such that $\gamma(b)$ is conjugate to $\gamma(a)$ along γ, and γ has an **interior conjugate point** if there is such a $b \in (a, c)$.

no minimizing **Theorem 2.4.23.** *Let (M, g) be a Riemannian manifold and $p, q \in M$. If γ is a unitspeed geodesic segment from p to q that has an interior conjugate point, then there exists a proper normal vector field X along γ such that $I(X, X) < 0$. Therefore, γ is not minimizing.*

Proof. Suppose $\gamma : [a, c] \to M$ is a unit-speed geodesic segment, and $\gamma(b)$ is conjugate to $\gamma(a)$ along γ for some $a < b < c$. This means that there is a nontrivial normal Jacobi field J along γ that vanishes at $t = a$ and $t = b$. Define a vector field V along all of γ by

$$V(t) = \begin{cases} J(t), & t \in [a, b]; \\ 0, & t \in [b, c]. \end{cases}$$

This is a proper, normal, piecewise smooth vector field along γ.

Let W be a smooth proper normal vector field along γ such that $W(b)$ is equal to the jump $\Delta D_t V$ at $t = b$. Such a vector field is easily constructed with the help of an orthonormal frame along γ and a bump function. Note that $\Delta D_t V = 0 - D_t J(b)$ is not zero, because otherwise J would be a Jacobi field satisfying $J(b) = D_t J(b) = 0$, and thus would be identically zero.

For small positive ε, let $X_\varepsilon = V + \varepsilon W$. Then

$$I(X_\varepsilon, X_\varepsilon) = I(V + \varepsilon W, V + \varepsilon W) = I(V, V) + 2\varepsilon I(V, W) + \varepsilon^2 I(W, W).$$

Since V satisfies the Jacobi equation on each subinterval $[a, b]$ and $[b, c]$, and $V(b) = 0$, (Riemann index form 4.24) gives

$$I(V, V) = -\langle \Delta D_t V, V(b) \rangle = 0.$$

Similarly,

$$I(V, W) = -\langle \Delta D_t V, W(b) \rangle = -|W(b)|^2.$$

Thus

$$I(X_\varepsilon, X_\varepsilon) = -2\varepsilon |W(b)|^2 + \varepsilon^2 I(W, W).$$

If we choose ε small enough, this is strictly negative. □

There is a partial converse to the preceding theorem, which says that a geodesic without conjugate points has the shortest length among all nearby curves in any proper variation. Before we prove it, we need the following technical lemma.

conjugate lem **Lemma 2.4.24.** *Let $\gamma : [a,b] \to M$ be a geodesic segment, and suppose J_1 and J_2 are Jacobi fields along γ. Then $\langle D_t J_1(t), J_2(t)\rangle - \langle J_1(t), D_t J_2(t)\rangle$ is constant along γ.*

Proof. Let $f(t) = \langle D_t J_1(t), J_2(t)\rangle - \langle J_1(t), D_t J_2(t)\rangle$. Using the Jacobi equation, we compute

$$\begin{aligned} f'(t) &= \langle D_t^2 J_1(t), J_2(t)\rangle + \langle D_t J_1(t), D_t J_2(t)\rangle - \langle D_t J_1(t), D_t J_2(t)\rangle - \langle J_1(t), D_t^2 J_2(t)\rangle \\ &= -R(J_1(t),\gamma'(t),\gamma'(t),J_2(t)) + R(J_2(t),\gamma'(t),\gamma'(t),J_1(t)) = 0, \end{aligned}$$

where the last equality follows from the symmetries of the curvature tensor. □

int conjugate **Theorem 2.4.25.** *Let (M,g) be a Riemannian manifold. Suppose $\gamma : [a,b] \to M$ is a unit-speed geodesic segment without interior conjugate points. If V is any proper normal piecewise smooth vector field along γ, then $I(V,V) \geq 0$, with equality if and only if V is a Jacobi field. In particular, if $\gamma(b)$ is not conjugate to $\gamma(a)$ along γ, then $I(V,V) > 0$ unless $V \equiv 0$.*

Proof. To simplify the notation, we can assume (after replacing t by $t-a$) that $a = 0$. Let $p = \gamma(a)$, and let $(w_1,\dots,w_n)$ be an orthonormal basis for T_pM, chosen so that $w_1 = \gamma'(0)$. For each $\alpha = 2,\dots,n$, let J_α be the unique normal Jacobi field along γ satisfying $J_\alpha(0) = 0$ and $D_t J_\alpha(0) = w_\alpha$.

Our assumption that γ has no interior conjugate points guarantees that no linear combination of the J_α's can vanish for any $t \in (0,b)$, and thus $(J_\alpha(t))$ forms a basis for the orthogonal complement of $\gamma'(t)$ in $T_{\gamma(t)}M$ for each such t. Thus, given V as in the statement of the theorem, for $t \in (0,b)$ we can write

$$V = v^\alpha J_\alpha(t)$$

for some piecewise smooth functions $v^\alpha : (0,b) \to M$. (Here and in the remainder of this proof, the summation convention is in effect, with Greek indices running from 2 to n.)

In fact, each function v_α has a piecewise smooth extension to $[0,b]$. To see why, let (x^i) be the normal coordinates centered at p determined by the basis (w_i). For sufficiently small $t > 0$, we can express $J_\alpha(t)$ and $V(t)$ in normal coordinates as

$$\begin{aligned} J_\alpha(t) &= t\frac{\partial}{\partial x_\alpha}\bigg|_{\gamma(t)}, \quad \alpha = 2,\dots,n, \\ V(t) &= v^\alpha J_\alpha = t v^\alpha(t)\frac{\partial}{\partial x_\alpha}\bigg|_{\gamma(t)}. \end{aligned}$$

Riemann Jacobi vanishing one point coordinate form

(The formula for $J_\alpha(t)$ follows from Prop. 4.7.) Because V is smooth on $[0,\delta)$ for some $\delta > 0$ and $V(0) = 0$, it follows from Taylor's theorem that the components of $V(t)/t$ extend smoothly to $[0,\delta)$, which shows that v^α is smooth there. Because $V(b) = 0$, it follows similarly that v^α extends smoothly to $t = b$ as well. (If $J_\alpha(b) = 0$, the argument is the same as for $t = 0$, while if not, it is even easier.)

Let $(a_0,\dots,a_k)$ be an admissible partition for V. On each subinterval (a_{i-1},a_i) where V is smooth, define vector fields X and Y along γ by

$$X = v^\alpha D_t J_\alpha, \quad Y = \dot v^\alpha J_\alpha.$$

Then $D_t V = X + Y$ on each such interval, and the fact that V is piecewise smooth implies that $D_t V$, X, and Y extend smoothly to $[a_{i-1},a_i]$ for each i, with one-sided derivatives at the endpoints.

To compute $I(V,V)$, we will use the following identity, which holds on each subinterval $[a_{i-1},a_i]$:

$$|D_tV|^2 - R(V,\gamma',\gamma',V) = \frac{d}{dt}\langle V,X\rangle + |Y|^2. \tag{4.25}$$

Granting this for now, we use the fundamental theorem of calculus to compute

$$\begin{aligned} I(V,V) &= \sum_{i=1}^{k}\int_{a_{i-1}}^{a_i}(|D_tV|^2 - R(V,\gamma',\gamma',V))dt \\ &= \sum_{i=1}^{k}\langle V,X\rangle\Big|_{t=a_{i-1}}^{t=a_i} + \int_0^b |Y|^2 dt, \end{aligned}$$

where the boundary terms are to be interpreted as limits from above and below. Because X and V are both continuous on $[0,b]$, the boundary terms at $t=a_1,\dots,a_{k-1}$ all cancel, and because $V(0)=V(b)=0$, the boundary terms at $t=0$ and $t=b$ are both zero. It follows that $I(V,V)=\int_0^b |Y|^2dt \geq 0$. If $I(V,V)=0$, then Y is identically zero on $[0,b]$. Since the J_α's are linearly independent there, this implies that $\dot v^\alpha \equiv 0$ for each α, so each v^α is constant. Thus V is a linear combination of Jacobi fields with constant coefficients, so it is a Jacobi field.

It remains only to prove (4.25). Note that

$$\frac{d}{dt}\langle V,X\rangle = \langle D_tV,X\rangle + \langle V,D_tX\rangle = \langle X+Y,X\rangle + \langle V,D_tX\rangle. \tag{4.26}$$

The Jacobi equation gives

$$D_tX = \dot v^\alpha D_tJ_\alpha + v^\alpha D_t^2J_\alpha = \dot v^\alpha D_tJ_\alpha - v^\alpha R(J_\alpha,\gamma')\gamma' = \dot v^\alpha D_tJ_\alpha - R(V,\gamma')\gamma'.$$

Therefore,

$$\langle D_tX,V\rangle = \langle \dot v^\alpha D_tJ_\alpha, v^\beta J_\beta\rangle - R(V,\gamma',\gamma',V). \tag{4.27}$$

Because $\langle D_tJ_\alpha,J_\beta\rangle - \langle J_\alpha,D_tJ_\beta\rangle = 0$ at $t=0$, it follows from Lemma 2.4.24 that $\langle D_tJ_\alpha,J_\beta\rangle = \langle J_\alpha,D_tJ_\beta\rangle$ all along γ. Thus we can simplify the first term in (4.25) as follows:

$$\begin{aligned} \langle \dot v^\alpha D_tJ_\alpha, v^\beta J_\beta\rangle &= \dot v^\alpha v^\beta\langle D_tJ_\alpha,J_\beta\rangle = \dot v^\alpha v^\beta\langle J_\alpha,D_tJ_\beta\rangle \\ &= \langle \dot v^\alpha J_\alpha, v^\beta D_tJ_\beta\rangle = \langle Y,X\rangle. \end{aligned}$$

Inserting this into (4.27), and then inserting the latter into (4.26) yields

$$\begin{aligned} \frac{d}{dt}\langle V,X\rangle &= \langle X+Y,X\rangle + \langle Y,X\rangle - R(V,\gamma',\gamma',V) \\ &= |X+Y|^2 - |Y|^2 - R(V,\gamma',\gamma',V), \end{aligned}$$

which is equivalent to (4.25). □

The next corollary summarizes the results of Theorems 2.4.23 and 2.4.25.

Corollary 2.4.26. *Let (M,g) be a Riemannian manifold, and let $\gamma:[a,b]\to M$ be a unit-speed geodesic segment.*

(a) If γ has an interior conjugate point, then it is not minimizing.

(b) If $\gamma(a)$ and $\gamma(b)$ are conjugate but γ has no interior conjugate points, then for every proper normal variation Γ of γ, the curve Γ_s is strictly longer than γ for all sufficiently small nonzero

s unless the variation field of Γ is a Jacobi field.

(c) If γ has no conjugate points, then for every proper normal variation Γ of γ, the curve Γ_s is strictly longer than γ for all sufficiently small nonzero s.

There is a far-reaching quantitative generalization of Theorems 10.26 and 10.28, called the Morse index theorem, which we do not treat here. The **index of a geodesic segment** is defined to be the maximum dimension of a linear space of proper normal vector fields along the segment on which I is negative definite. Roughly speaking, the index is the number of independent directions in which γ can be deformed to decrease its length. The Morse index theorem says that the index of every geodesic segment is finite, and is equal to the number of its interior conjugate points counted with multiplicity.

2.4.5 Cut points

Theorem 2.4.23 showed that once a geodesic passes its first conjugate point, it ceases to be minimizing. The converse, however, is not true: a geodesic can cease to be minimizing without reaching a conjugate point. For example, on the cylinder $S^1 \times \mathbb{R}$ with the product metric, there are no conjugate points along any geodesic; but no geodesic segment that wraps more than halfway around the cylinder is minimizing.

Therefore it is useful to make the following definitions. Suppose (M, g) is a complete, connected Riemannian manifold, p is a point of M, and $v \in T_pM$. Define the cut time of (p, v) by

$$t_{cut}(p, v) = \sup\{b > 0 : \text{the restriction of } \gamma_v \text{ to } [0, b] \text{ is minimizing}\}.$$

where γ_v is the maximal geodesic starting at p with initial velocity v. Because γ_v is minimizing as long as its image stays inside a geodesic ball (Prop. 1.4.10), $t_{cut}(p, v)$ is always positive; but it might be $+\infty$.

If $t_{cut}(p, v) < \infty$, the **cut point of p along γ_v** is the point $\gamma_v(t_{cut}(p, v)) \in M$. The **cut locus of p**, denoted by $Cut(p)$, is the set of all $q \in M$ such that q is the cut point of p along some geodesic. Because the question whether a geodesic is minimizing is independent of parametrization, the cut point of p along γ_v is the same as the cut point along γ_v for every positive constant λ, so we may as well restrict attention to unit vectors v. Theorem 2.4.23 says that the cut point (if it exists) occurs at or before the first conjugate point along every geodesic.

The determination of the cut locus of a point is typically very difficult; but the next example gives some special cases in which it is relatively simple.

Example 2.4.27 (Cut Loci). (a) If $(S^n(R), \mathring{g}_R)$ is a sphere with a round metric, the cut locus of every point $p \in S^n(R)$ is the singleton set containing only the antipodal point $-p$.

(b) On a product space $S^n(R) \times \mathbb{R}^m$ with the product metric, the cut locus of every point (p, x) is the set $\{-p\} \times \mathbb{R}^m$.

Proposition 2.4.28 (Properties of Cut Times). *Suppose (M, g) is a complete, connected Riemannian manifold, $p \in M$, and v is a unit vector in T_pM. Let $c = t_{cut}(p, v) \in (0, \infty]$.*

(a) If $0 < b < c$, then $\gamma_v|_{[0,b]}$ has no conjugate points and is the unique unit-speed minimizing curve between its endpoints.

(b) *If $c < +\infty$, then $\gamma_v|_{[0,c]}$ is minimizing, and one or both of the following conditions are true:*

- *$\gamma_v(c)$ is conjugate to p along γ_v.*
- *There are two or more unit-speed minimizing geodesics from p to $\gamma_v(c)$*

Proof. Suppose first that $0 < b < c$. By definition of $t_{cut}(p, v)$, there is a time b' such that $b < b' < c$ and $\gamma_v|_{[0,b']}$ is minimizing. Then $\gamma_v(t)$ cannot be conjugate to p along γ_v for any $0 < t \leq b$, and $\gamma_v|_{[0,b]}$ is minimizing.

To see that $\gamma_v|_{[0,b]}$ is the unique unit-speed minimizing curve between its endpoints, suppose for the sake of contradiction that $\sigma : [a, b] \to M$ is another. Note that $\sigma'(b) \neq \gamma'(b)$, since otherwise σ and γ_v would agree on $[0, b]$ by uniqueness of geodesics. Define a new unit-speed admissible curve $\widetilde{\gamma} : [0, b'] \to M$ that is equal to $\sigma(t)$ for $t \in [0, b]$ and equal to $\gamma_v(t)$ for $t \in [b, b']$. Then $\widetilde{\gamma}$ has length b', so it is also a minimizing curve from p to $\gamma_v(b')$; but it is not smooth at $t = b$, contradicting the fact that minimizing curves are smooth geodesics. This completes the proof of (a).

Now suppose $c < +\infty$. By definition of $t_{cut}(p, v)$, there is a sequence of times $b_i \to c^-$ such that the restriction of γ_v to $[0, b_i]$ is minimizing. By continuity of the distance function, therefore,

$$d_g(p, \gamma_v(c)) = \lim_{i\to\infty} d_g(p, \gamma_v(b_i)) = \lim_{i\to\infty} b_i = c.$$

which shows that γ_v is minimizing on $[0, c]$. To prove that one of the options in (b) must hold, assume that $\gamma_v(c)$ is not conjugate to p along γ_v. We will prove the existence of a second unit-speed minimizing geodesic from p to $\gamma_v(c)$.

Let (b_i) be a sequence of real numbers such that $b_i \to c^+$. By definition of cut time, $\gamma_v|_{[0,b_i]}$ is not minimizing, so for each i there is a unit-speed minimizing geodesic $\sigma_i : [0, a_i] \to M$ such that $\sigma_i(0) = p$, $\sigma_i(a_i) = \gamma_v(b_i)$, and $a_i < b_i$. Set $w_i = \sigma_i'(0) \in T_pM$, so each w_i is a unit vector. By compactness of the unit sphere, after passing to a subsequence we may assume that w_i converges to some unit vector w. Since the a_i's are all positive and bounded above by b_1, by passing to a further subsequence, we may also assume that a_i converges to some number a. Then by continuity of the exponential map, $\sigma_i(a_i) = \exp_p(a_i w_i)$ converges to $\exp_p(aw)$. But we also know that $\sigma_i(a_i) = \gamma_v(b_i)$, which converges to $\gamma_v(c)$, so $\exp_p(aw) = \gamma_v(c)$. Moreover, by continuity of the distance function,

$$c = d_g(p, \gamma_v(c)) = \lim_{i\to\infty} d_g(p, \sigma_i(a_i)) = \lim_{i\to\infty} a_i = a.$$

Thus $\sigma : [0, c] \to M$ given by $\sigma(t) = \exp_p(tw)$ is also a unit-speed minimizing geodesic from p to $\gamma_v(c)$. We need to show that it is not equal to γ_v.

The assumption that $\gamma_v(c)$ is not conjugate to p along γ_v implies that cv is a regular point of $\exp_p$ (Prop. 2.4.17), so $\exp_p$ is injective in some neighborhood V of cv. Note that $\exp_p(a_i w_i) = \exp_p(b_i v)$ for each i, while $a_i w_i \neq b_i v$, since w_i and v are unit vectors and $a_i < b_i$. Since $b_i v$ converges to cv, we conclude that $b_i v \in V$ for sufficiently large i, and thus by injectivity $a_i w_i \notin V$ for these values of i. Therefore $cw = \lim_i a_i w_i \neq cv$, which implies $w \neq v$ and thus $\sigma \neq \gamma_v$, as claimed. □

Next we examine how the cut time varies as the initial point and initial velocity of the geodesic vary. Recall that the unit tangent bundle of a Riemannian manifold (M, g) is the subset $UTM = \{(p, v) : |v|_g = 1\} \subseteq TM$.

Theorem 2.4.29. *Suppose (M, g) is a complete, connected Riemannian manifold. The function $t_{cut} : UTM \to (0, +\infty]$ is continuous.*

Proof. Let $(p, v) \in UTM$ be arbitrary, and let (p_i, v_i) be any sequence in UTM converging to (p, v). Put $c_i = t_{cut}(p_i, v_i)$, and

$$b = \varliminf_{i\to\infty} c_i, \quad c = \varlimsup_{i\to\infty} c_i.$$

We will show that $c \le t_{cut}(p_i, v_i) \le b$, which implies $c_i \to c_{cut}(p, v)$.

To show that $c \le t_{cut}(p, v)$, suppose first that $c < \infty$. By passing to a subsequence, we may assume that c_i is finite for each i and $c_i \to c$. Proposition 2.4.28 shows that γ_{v_i} is minimizing on $[0, c_i]$. By continuity of the exponential map, $\exp(p_i, c_iV_i) \to \exp(p, cv)$ as $i \to \infty$, and therefore by continuity of the distance function we have

$$d_g(p, \exp(p, cv)) = \lim_{i\to\infty} d_g(p, \exp(p_i, c_iv_i)) = \lim_{i\to\infty} c_i = c.$$

This shows that γ_v is minimizing on $[0, c]$, and therefore $t_{cut}(p, v) \ge c$, as claimed.

Now suppose $c = +\infty$. Again, by passing to a subsequence, we may assume $c_i \to +\infty$. It follows that for every positive number c_0, the geodesic γ_{v_i} is minimizing on $[0, c_0]$ for i sufficiently large, and it follows by continuity as above that γ_v is minimizing on $[0, c_0]$. Since c_0 was arbitrary, this means that $t_{cut}(p, v) = +\infty$.

Next we show that $t_{cut}(p, v) \ge b$. If $b = +\infty$, there is nothing to prove, so assume $b < +\infty$. Again by passing to a subsequence, we may assume that c_i is finite for each i and $c_i \to b$. By virtue of Proposition 2.4.28, either there are infinitely many indices i for which $\gamma_{v_i}(c_i)$ is conjugate to p_i along γ_{v_i}, or there are infinitely many i for which there exists a second minimizing unit-speed geodesic σ_i from p_i to $\gamma_{v_i}(c_i)$.

In the first case, because conjugate points are critical values of the restricted exponential map, which can be detected in coordinates by the vanishing of a determinant of a matrix of first derivatives, it follows by continuity that $\gamma_v(b)$ is also a critical value, and thus $\gamma_v(b)$ is conjugate to p along γ_v. Then Theorem 2.4.23 shows that $t_{cut}(p, v) \le b$.

In the second case, let w_i be the unit vector in $T_{p_i}M$ such that $\sigma_i'(0) = w_i$. Because the components of w_i with respect to a local orthonormal frame lie in S^{n-1}, by passing to a subsequence we may assume $(p_i, w_i) \to (p, w)$. If $\gamma_v(b)$ is conjugate to p along γ_v, then $t_{cut}(p, v) \le b$ as above, so we may assume that $\gamma_v(b)$ is not a conjugate point. This means that bv is a regular point of the restricted exponential map $\exp_p$. Since the set of such regular points is an open subset of TM, there is some $\varepsilon > 0$ such that $\exp_{p_i}$ is injective on the ε-neighborhood of c_iw_i for all i sufficiently large. This implies that $|c_iw_i - c_iv_i| \ge \varepsilon$ for all such i, and therefore the limits bw and bv are distinct. Thus $\gamma_w|_{[0,b]}$ is another minimizing geodesic from p to $\gamma_v(b)$, which by Proposition 2.4.28 implies that $t_{cut}(p, v) \le b$. $\square$

Given $p \in M$, we define two subsets of T_pM as follows: the **tangent cut locus** of p is the set

$$\mathrm{TCL}(p) = \{v \in T_pM : |v| = t_{cut}(p, v)\},$$

and the **injectivity domain of p** is

$$\mathrm{ID}(p) = \{v \in T_pM : |v| < t_{cut}(p, v)\}.$$

It is immediate that $\mathrm{TCL}(p) = \partial\mathrm{ID}(p)$, and $\mathrm{Cut}(p) = \exp_p(\mathrm{TCL}(p))$. Further properties of $\mathrm{Cut}(p)$ and $\mathrm{ID}(p)$ are described in the following theorem.

Theorem 2.4.30. *Let (M, g) be a complete, connected Riemannian manifold and $p \in M$.*

(a) The cut locus of p is a closed subset of M of measure zero.

(b) The restriction of $\exp_p$ to $\overline{\mathrm{ID(p)}}$ is surjective.

(c) The restriction of $\exp_p$ to $\mathrm{ID}(p)$ is a diffeomorphism onto $M \setminus \mathrm{Cut}(p)$.

Proof. To prove that the cut locus is closed, suppose (q_i) is a sequence of points in $\mathrm{Cut}(p)$ converging to some $q \in M$. Write $q_i = \exp_p(t_{cut}(p, v_i)v_i)$ for unit vectors v_i. By compactness of the unit sphere, we may assume after passing to a subsequence that v_i converges to some unit vector v, and by Theorem 2.4.29, $t_{cut}(p, v) = \lim_{i\to\infty} t_{cut}(p, v_i)$. Because convergent sequences in a metric space are bounded, the sequence $(t_{cut}(p, v_i))$ is bounded, and therefore $t_{cut}(p, v) < +\infty$. By continuity of the exponential map, therefore, q must be equal to $\exp_p(t_{cut}(p, v)v)$, which shows that $q \in \mathrm{Cut}(p)$, and thus $\mathrm{Cut}(p)$ is closed.

To see that $\mathrm{Cut}(p)$ has measure zero, note first that in any polar coordinates $(\theta^1, \ldots, \theta^{n-1}, r)$ on T_pM, the set $\mathrm{TCL}(p)$ can be expressed locally as the graph of the continuous function $r = t_{cut}(p, (\theta^1, \ldots, \theta^{n-1}))$, using the fact that $(\theta^1, \ldots, \theta^{n-1})$ form smooth local coordinates for the unit sphere in T_pM. Since graphs of continuous functions have measure zero, it follows that $\mathrm{TCL}(p)$ is a union of finitely many sets of measure zero and thus has measure zero in T_pM; and because smooth maps take sets of measure zero to sets of measure zero, $\mathrm{Cut}(p) = \exp_p(\mathrm{TCL}(p))$ has measure zero in M. This proves (a).

Part (b) follows from the fact that every point of M can be connected to p by a minimizing geodesic. To prove (c), note that it follows easily from the definitions that $\exp_p(\mathrm{ID}(p)) = M \setminus \mathrm{Cut}(p)$. Also, the definition of $\mathrm{ID}(p)$ guarantees that no point in $\exp_p(\mathrm{ID}(p))$ can be a cut point of p, and thus no such point can be a conjugate point either. The absence of cut points implies that $\exp_p$ is injective on $\mathrm{ID}(p)$, and the absence of conjugate points implies that it is a local diffeomorphism there. Together these two facts imply that it is a diffeomorphism onto its image. □

The preceding theorem leads to the following intriguing topological result about compact manifolds.

Corollary 2.4.31. *Every compact, connected, smooth n-manifold is homeomorphic to a quotient space of $\overline{B}^n$ by an equivalence relation that identifies only points on the boundary.*

Proof. Let M be a compact, connected, smooth n-manifold, let p be any point of M, and let g be any Riemannian metric on M. Because a compact metric space has finite diameter, every unit vector in T_pM has a finite cut time, no greater than the diameter of M. Let $\overline{B}_1 \subseteq T_pM$ denote the closed unit ball in T_pM, and define a map $f : \overline{B}_1 \to \overline{\mathrm{ID(p)}}$ by

$$f(v) = \begin{cases} t_{cut}(p, v)v, & v \neq 0, \\ 0 & v = 0. \end{cases}$$

It follows from Theorem 2.4.29 that f is continuous, and it is easily seen to be bijective, so it is a homeomorphism by the closed map lemma. Since every orthonormal basis for TpM yields a homeomorphism of $\overline{B}_1$ with $\overline{B}^n$, it follows that $\mathrm{ID}(p)$ is homeomorphic to $\overline{B}^n$ and the homeomorphism takes $\mathrm{TCL}(p) = \partial\mathrm{ID}(p)$ to S^{n-1}.

By Theorem 2.4.30, $\exp_p$ restricts to a surjective map from $\mathrm{ID}(p)$ to M, and it is a quotient map by the closed map lemma. It follows that M is homeomorphic to the quotient of $\mathrm{ID}(p)$

by the equivalence relation $v \sim w$ if and only if $\exp_p(v) = \exp_p(w)$. Since $\exp_p$ is injective on $\mathrm{ID}(p)$ and the images of $\mathrm{ID}(p)$ and $\overline{\mathrm{ID(p)}} = \mathrm{TCL}(p)$ are disjoint, the equivalence relation identifies only points on the boundary of $\mathrm{ID}(p)$. $\square$

Recall that the injectivity radius of M at p, denoted by $\mathrm{inj}(p)$, is the supremum of all positive numbers a such that $\exp_p$ is a diffeomorphism from $B_a(0) \subseteq T_pM$ to its image. The injectivity radius is closely related to the cut locus, as the next proposition shows.

Proposition 2.4.32. *Let (M, g) be a complete, connected Riemannian manifold. For each $p \in M$, the injectivity radius at p is the distance from p to its cut locus if the cut locus is nonempty, and infinite otherwise.*

Proof. Given $p \in M$, let d denote the distance from p to its cut locus, with the convention that $d = +\infty$ if the cut locus is empty. Let $a \in (0, +\infty]$ be arbitrary, and let $B_a \subseteq T_pM$ denote the set of vectors $v \in T_pM$ with $|v| < a$ (so $B_a = T_pM$ if $a = +\infty$). We will show that the restriction of $\exp_p$ to B_a is a diffeomorphism onto its image if and only if $a \leq d$, from which the result follows.

First suppose $a \leq d$. By definition of d, no point of the form $\exp_p(v)$ with $v \in B_a$ can be a cut point of p, so $B_a \subseteq \mathrm{ID}(p)$. It follows from Theorem 2.4.30(c) that $\exp_p$ is a diffeomorphism from B_a onto its image. On the other hand, if $a > d$, then p has a cut point q whose distance from p is less than a. It follows from the definition of cut points that the radial geodesic from p to q is not minimizing past q, so Proposition 1.4.10 shows that there is no geodesic ball of radius greater than $d_g(p, q)$. In particular, the restriction of $\exp_p$ to B_a cannot be a diffeomorphism onto its image. $\square$

Proposition 2.4.33. *Let (M, g) be a complete, connected Riemannian manifold. The function* $\mathrm{inj} : M \to (0, \infty]$ *is continuous.*

Proof. Let $p \in M$ be arbitrary. Proposition 2.4.28(b) shows that for each point $q \in \mathrm{Cut}(p)$, there is a minimizing unit-speed geodesic γ_v from p to q whose length is $t_{cut}(p, v)$, and therefore the distance from p to $\mathrm{Cut}(p)$ is the infimum of the cut times of unit-speed geodesics starting at p. By the previous proposition, therefore,

$$\mathrm{inj}(p) = \inf\{t_{cut}(p, v) : v \in T_pM \text{ with } |v|_g = 1\}.$$

Suppose (p_i) is a sequence in M converging to a point $p \in M$. As in the proof of Theorem 2.4.29, we will prove that $\mathrm{inj}(p_i) \to \mathrm{inj}(p)$ by showing that $c \leq \mathrm{inj}(p) \leq b$, where

$$b = \varliminf_{i \to \infty} \mathrm{inj}(p_i), \quad c = \varlimsup_{i \to \infty} \mathrm{inj}(p_i).$$

First we show that $\mathrm{inj}(p) \leq b$. By passing to a subsequence, we may assume $\mathrm{inj}(p_i) \to b$. By compactness of the unit sphere, for each i there is a unit vector $v_i \in T_{p_i}M$ such that $\mathrm{inj}(p_i) = t_{cut}(p_i, v_i)$, and after passing to a further subsequence, we may assume $(p_i, v_i) \to (p, v)$ for some $v \in T_pM$. By continuity of t_{cut}, we have $t_{cut}(p, v) = \lim_i t_{cut}(p_i, v_i) = b$, so $\mathrm{inj}(p) \leq b$.

Next we show that $\mathrm{inj}(p) \geq c$. Once again, by passing to a subsequence of the original sequence (p_i), we may assume $\mathrm{inj}(p_i) \to c$. Suppose for the sake of contradiction that $\mathrm{inj}(p) < c$, and choose c_0 such that $\mathrm{inj}(p) < c_0 < c$. Let w be a unit vector in T_pM such that $t_{cut}(p, w) = \mathrm{inj}(p)$. We can choose some sequence of unit vectors $w_i \in T_{p_i}M$ such that $(p_i, w_i) \to (p, v)$, so $t_{cut}(p_i, w_i) \to t_{cut}(p, w) = \mathrm{inj}(p)$. For i sufficiently large, this implies $t_{cut}(p_i, w_i) < c_0 < c$, contradicting the facts that $t_{cut}(p_i, w_i) \geq \mathrm{inj}(p_i)$ and $\mathrm{inj}(p_i) \to c$. $\square$

Chapter 3

Additional Topics

3.1 The de Rham cohomology groups

3.1.1 The de Rham cohomology groups

We studied the closed 1-form

$$\omega = \frac{xdy - ydx}{x^2 + y^2},$$

and showed that it is not exact on $\mathbb{R}^2 - \{0\}$, but it is exact on some smaller domains such as the right half-plane $H = \{(x, y) : x > 0\}$, where it is equal to $d\theta$.

As we will see later, this behavior is typical: closed forms are always **locally exact**, so whether a given closed form is exact depends on the global shape of the domain. To capture this dependence, we make the following definitions.

Let M be a smooth manifold with or without boundary, and let p be a nonnegative integer. Because $d : \Omega^p(M) \to \Omega^{p+1}(M)$ is linear, its kernel and image are linear subspaces. We define

$$Z^p(M) = \ker d = \{\text{closed } p\text{-form on } M\},$$
$$B^p(M) = \operatorname{im} d = \{\text{exact } p\text{-form on } M\}.$$

By convention, we consider $\Omega^p(M)$ to be the zero vector space when $p < 0$ or $p > n = \dim M$, so that, for example, $B^0(M) = 0$ and $Z^n = \Omega^n(M)$.

The fact that every exact form is closed implies that $B^p(M) \subseteq Z^p(M)$. Thus, it makes sense to define the **de Rham cohomology group** in degree p of M to be the quotient vector space

$$H^p_{dR}(M) = \frac{Z^p(M)}{B^p(M)}.$$

It is clear that $H^p_{dR}(M) = 0$ for $p < 0$ or $p > \dim M$, because $Z^p(M) = 0$ in those cases. For $0 \le p \le n$, the definition implies that $H^p_{dR}(M) = 0$ if and only if every closed p-form on M is exact.

Example 3.1.1. The fact that there is a closed 1-form on $\mathbb{R}^2 - \{0\}$ that is not exact means that $H^1_{dR}(\mathbb{R}^2 - \{0\}) \neq 0$. On the other hand, the Poincaré lemma for 1-forms implies that $H^1_{dR}(U) = 0$ for any star-shaped open subset $U \subseteq \mathbb{R}^n$.

The first order of business is to show that the de Rham groups are diffeomorphism invariants. For any closed p-form ω on M, we let $[\omega]$ denote the equivalence class of ω in $H^p_{dR}(M)$, called the **cohomology class** of ω. If $[\omega] = [\omega']$, we say that ω and ω' are **cohomologous**.

y induced map

Proposition 3.1.2 (Induced Cohomology Maps). *For any smooth map $F : M \to N$ between smooth manifolds with or without boundary, the pullback $F^* : \Omega^p(N) \to \Omega^p(M)$ carries $Z^p(N)$ into $Z^p(M)$ and $B^p(N)$ into $B^p(M)$. It thus descends to a linear map, still denoted by F^*, from $H^p_{dR}(N)$ to $H^p_{dR}(M)$, called the **induced cohomology map**. It has the following properties:*

(a) If $G : N \to P$ is another smooth map, then $(G \circ F)^* = F^* \circ G^*$.

(b) If id denotes the identity map of M, then id^* is the identity map of $H^p_{dR}(M)$.

Proof. If ω is closed, then $d(F^*\omega) = f^*d\omega = 0$, so $F^*\omega$ is also closed. If $\omega = d\eta$ is exact, then $F^*\omega = F^*(d\eta) = d(F^*\eta)$, which is also exact. Therefore, F^* maps $Z^p(N)$ into $Z^p(M)$ and $B^p(N)$ into $B^p(M)$. The induced cohomology map is defined in the obvious way: for a closed p-form ω, let

$$F^*[\omega] = [F^*\omega].$$

The rest is immediate. □

The next two corollaries are immediate.

Corollary 3.1.3 (Functoriality). *For any integer p, the assignment $M \mapsto H^p_{dR}(M), F \mapsto F^*$ is a contravariant functor from the category of smooth manifolds with boundary to the category of real vector spaces.*

Corollary 3.1.4 (Diffeomorphism Invariance of de Rham Cohomology). *Diffeomorphic smooth manifolds (with or without boundary) have isomorphic de Rham cohomology groups.*

Elementary computations

The direct computation of the de Rham groups is not easy in general. However, there are a number of special cases that can be easily computed by various techniques.

Proposition 3.1.5 (Cohomology of Disjoint Unions). *Let $\{M_j\}$ be a countable collection of smooth n-manifolds with or without boundary, and let $M = \amalg_j M_j$. For each p, the inclusion maps $\iota_j : M_j \hookrightarrow M$ induce an isomorphism from $H^p_{dR}(M)$ to the direct product space $\prod_j H^p_{dR}(M_j)$.*

Because of this proposition, each de Rham group of a disconnected manifold is just the direct product of the corresponding groups of its components. Thus, we can concentrate henceforth on computing the de Rham groups of connected manifolds. Next we give an explicit characterization of de Rham cohomology in degree zero.

logy degree 0

Proposition 3.1.6 (Cohomology in Degree Zero). *If M is a connected smooth manifold with or without boundary, then $H^0_{dR}(M)$ is equal to the space of constant functions and is therefore 1-dimensional.*

Proof. Because there are no (-1)-forms, $B^0(M) = 0$. A closed 0-form is a smooth real-valued function f such that $df = 0$, and since M is connected, this is true if and only if f is constant. Therefore $H^0_{dR}(M) = Z^0(M) = \{\text{constant}\}$. □

Corollary 3.1.7 (Cohomology of Zero-Manifolds). *Suppose M is a manifold of dimension 0. Then $H^0_{dR}(M)$ is a direct product of 1-dimensional vector spaces, one for each point of M, and all other de Rham cohomology groups of M are zero.*

3.1.2 Homotopy invariance

The underlying fact that allows us to prove the homotopy invariance of de Rham cohomology is that homotopic smooth maps induce the same cohomology map. To motivate the proof, suppose $F, G : M \to N$ are smooth maps, and let us think about what it means to prove that $F^* = G^*$. Given a closed p-form ω on N, we need somehow to produce a $(p-1)$-form η on M such that

$$G^*\omega - F^*\omega = d\eta.$$

from which it follows that $[G^*\omega] - [F^*\omega] = 0$. One might hope to construct η in a systematic way, resulting in a map h from closed p-forms on N to $(p-1)$-forms on M that satisfies

$$d(h\omega) = G^*\omega - F^*\omega.$$

Instead of defining $h\omega$ only when ω is closed, it turns out to be far simpler to define a map h from the space of all smooth p-forms on N to the space of smooth $(p-1)$-forms on M. Such a map cannot satisfy our desired formula, but instead we will find a family of such maps, one for each p, satisfying

$$d(h\omega) + h(d\omega) = G^*\omega - F^*\omega. \tag{1.1}$$

In general, if $F, G : M \to N$ are smooth maps, a collection of linear maps $h : \Omega^p(N) \to \Omega^{p-1}(M)$ such that (1.1) is satisfied for all ω is called a homotopy operator between F^* and G^*. (The term cochain homotopy is used frequently in the algebraic topology literature.) The next proposition follows immediately from the argument in the preceding paragraph.

Proposition 3.1.8. *Suppose M and N are smooth manifolds with or without boundary. If $F : M \to N$ are smooth maps and there exists a homotopy operator between the pullback maps F^* and G^*, then the induced cohomology maps F^* and G^* are equal.*

The key to our proof of homotopy invariance is to construct a homotopy operator first in the following special case. Let M be a smooth manifold with or without boundary, and for each $t \in I$, let it $i_t : M \to M \times I$ be the map

$$i_t(x) = (x, t).$$

If M has empty boundary, then $M \times I$ is a smooth manifold with boundary, and all of the results above apply to it. But if $\partial M \neq \varnothing$, then $M \times I$ has to be considered as a smooth manifold with corners. It is straightforward to check that the definitions of the de Rham groups and induced homomorphisms make perfectly good sense on manifolds with corners, and Proposition 3.1.2 is valid in that context as well.

Lemma 3.1.9 (Existence of a Homotopy Operator). *For any smooth manifold M with or without boundary, there exists a homotopy operator between the two maps $i_0^*, i_1^* : \Omega^*(M \times I) \to \Omega^*(M)$.*

Proof. For each p, we need to define a linear map $h : \Omega^p(M) \to \Omega^{p-1}(M)$ such that

$$h(d\omega) + d(h\omega) = i_1^*\omega - i_0^*\omega. \tag{1.2}$$

Let s denote the standard coordinate on $\mathbb{R}$, and let S be the vector field on $M \times \mathbb{R}$ given by $S_{(q,s)} = (0, \partial/\partial s|_s)$ under the usual identification $T_{(q,s)}(M \times \mathbb{R}) \approx T_qM \oplus T_s\mathbb{R}$. Given a smooth p-form ω on $M \times I$, define

$$h\omega = \int_0^1 i_t^*(S \lrcorner\, \omega)\, dt.$$

More specifically, for any $q \in M$, this means

$$(h\omega)_q = \int_0^1 i_t^*\big((S \lrcorner\, \omega)_{(q,t)}\big)\,dt.$$

where the integrand is interpreted as a function of t with values in the vector space $\bigwedge^{p-1}(T_q^*M)$. We can compute $d(h\omega)$ at any point by differentiating under the integral sign in local coordinates, which yields

$$d(h\omega) = \int_0^1 d\big(i_t^*(S \lrcorner\, \omega)\big)\,dt.$$

Therefore, using Cartan's magic formula, we obtain

$$\begin{aligned} d(h\omega) + h(d\omega) &= \int_0^1 d\big(i_t^*(S \lrcorner\, \omega)\big)\,dt + \int_0^1 i_t^*(S \lrcorner\, d\omega)\,dt \\ &= \int_0^1 i_t^*\big(d(S \lrcorner\, \omega) + S \lrcorner\, d\omega\big)\,dt \\ &= \int_0^1 i_t^*(\mathfrak{L}_S\omega)\,dt. \end{aligned}$$

To simplify this last expression, we use the flow of S on $M \times \mathbb{R}$. (If M has nonempty boundary, note that S is tangent to $\partial(M \times \mathbb{R}) = \partial M \times \mathbb{R}$.) The flow is given explicitly by $\theta_t(q,s) = (q, t+s)$, so S is complete. It follows that we can write it $i_t = \theta_t \circ i_0$, and therefore by Proposition Lie der tensor t_0 **??**,

$$i_t^*(\mathfrak{L}_S\omega) = i_0^* \circ \theta_t^*(\mathfrak{L}_S\omega) = i_0^*\Big(\frac{d}{dt}(\theta_t^*\omega)\Big) = \frac{d}{dt} i_0^*(\theta_t^*\omega) = \frac{d}{dt} i_t^*\omega.$$

Applying the fundamental theorem of calculus, we obtain (cochain homotopy-1 1.2). □

homotopy map **Proposition 3.1.10.** *Suppose M and N are smooth manifolds with or without boundary, and $F, G : M \to N$ are homotopic smooth maps. For every p, the induced cohomology maps F^* and G^* are equal.*

Proof. The preceding lemma implies that the two cohomology maps i_0^* and i_1^* are equal. By Theorem homotopy smooth boundary **??**, there is a smooth homotopy $H : M \times I \to N$ from F to G. Because $F = H \circ i_0$ and $G = H \circ i_1$, Proposition cohomology induced map 3.1.2 implies

$$F^* = i_0^* \circ H^* = i_1^* \circ H = G^*.$$

□

Theorem 3.1.11 (Homotopy Invariance of de Rham Cohomology). *If M and N are homotopy equivalent smooth manifolds with or without boundary, then $H^p_{dR}(M) \cong H^p_{dR}(N)$ for each p. The isomorphisms are induced by any smooth homotopy equivalence $F : M \to N$.*

Proof. Suppose $F : M \to N$ is a homotopy equivalence, with homotopy inverse $G : N \to M$. By the Whitney approximation theorem, there are smooth maps $\widetilde{F} : M \to N$ homotopic to F and $\widetilde{G} : N \to M$ homotopic to G. Because homotopy is preserved by composition, it follows that

$$\widetilde{F} \circ \widetilde{G} \simeq F \circ G \simeq \mathrm{id}_M, \quad \widetilde{G} \circ \widetilde{F} \simeq G \circ F \simeq \mathrm{id}_N.$$

so $\widetilde{F}$ and $\widetilde{G}$ are homotopy inverses of each other.

cohomology homotopy map

Proposition 3.1.10 shows that, on cohomology,

$$\widetilde{F}^* \circ \widetilde{G}^* = \mathrm{id}, \quad \widetilde{G}^* \circ \widetilde{F}^* = \mathrm{id}.$$

so $\widetilde{F}^* : H^p_{dR}(M) \to H^p_{dR}(N)$ is an isomorphism. □

Because every homeomorphism is a homotopy equivalence, the next corollary is immediate.

Corollary 3.1.12 (Topological Invariance of de Rham Cohomology). *The de Rham cohomology groups are topological invariants: if M and N are homeomorphic smooth manifolds with or without boundary, then their de Rham cohomology groups are isomorphic.*

This result is remarkable, because the definition of the de Rham groups of M is intimately tied up with its smooth structure, and we had no reason to expect that different differentiable structures on the same topological manifold should give rise to the same de Rham groups.

Computations using homotopy invariance

We can use homotopy invariance to compute a number of de Rham groups. We begin with the simplest case of homotopy equivalence. A topological space X is said to be **contractible** if the identity map of X is homotopic to a constant map.

Theorem 3.1.13 (Cohomology of Contractible Manifolds). *If M is a contractible smooth manifold with or without boundary, then $H^p_{dR}(M) = 0$ for $p \geq 1$.*

Theorem 3.1.14 (The Poincaré Lemma). *If U is a star-shaped open subset of $\mathbb{R}^n$ or $\mathbb{H}^n$, then $\mathbb{H}^p_{dR}(U) = 0$ for $p \geq 1$.*

Proof. If U is star-shaped with respect to c, then it is contractible by the straight-line homotopy $H(x,t) = c + t(x - c)$. □

Corollary 3.1.15 (Local Exactness of Closed Forms). *Let M be a smooth manifold with or without boundary. Each point of M has a neighborhood on which every closed p-form is exact for $p \geq 1$.*

Corollary 3.1.16 (Cohomology of Euclidean Spaces and Half-Spaces). *For any integers $n \geq 0$ and $p \geq 1$, we have $H^p_{dR}(\mathbb{R}^n) = H^p_{dR}(\mathbb{H}^n) = 0$.*

Another case in which we can say quite a lot about de Rham cohomology is in degree 1. Suppose M is a connected smooth manifold and q is any point in M. Let $\mathrm{Hom}(\pi_1(M,q), \mathbb{R})$ denote the set of group homomorphisms from $\pi_1(M,q)$ to the additive group $\mathbb{R}$; it is a vector space under pointwise addition of homomorphisms and multiplication by constants. We define a linear map $\Phi : H^1_{dR}(M) \to \mathrm{Hom}(\pi_1(M,q), \mathbb{R})$ as follows: given a cohomology class $[\omega] \in H^1_{dR}(M)$, define $\Phi[\omega] : \pi_1(M,q) \to \mathbb{R}$ by

$$\Phi[\omega][\gamma] = \int_{\widetilde{\gamma}} \omega.$$

where $[\gamma]$ is any path homotopy class in $\pi_1(M,q)$, and $\widetilde{\gamma}$ is any piecewise smooth curve representing the same path class.

amental group

Theorem 3.1.17 (First Cohomology and the Fundamental Group). *Suppose M is a connected smooth manifold. For each $q \in M$, the linear map $\Phi : H^1_{dR}(M) \to \mathrm{Hom}(\pi_1(M,q), \mathbb{R})$ is an isomorphism.*

Proof. Given $[\gamma] \in \pi_1(M, q)$, it follows from the Whitney approximation theorem that there is some smooth closed curve segment $\widetilde{\gamma}$ in the same path class as γ, and from Theorem int homotopy path **??** that $\int_{\widetilde{\gamma}} \omega$ gives the same result for every piecewise smooth curve $\widetilde{\gamma}$ in the given class. Moreover, if $\widetilde{\omega}$ is another smooth 1-form in the same cohomology class as ω, then $\omega - \widetilde{\omega} = df$ for some smooth function f, which implies

$$\int_{\widetilde{\gamma}} \omega - \int_{\widetilde{\gamma}} \widetilde{\omega} = \int_{\widetilde{\gamma}} df = f(q) - f(q) = 0.$$

Thus Φ is well defined. It follows from Proposition line integral prop **??**(c) that $\Phi[\omega]$ is a group homomorphism from $\pi_1(M, q)$ to $\mathbb{R}$, and from linearity of the line integral that Φ itself is a linear map.

To see that Φ is an isomorphism, we will use the following facts:

$$H_1(M; \mathbb{R}) \cong \mathrm{Ab}(\pi_1(M), q), \quad H^1_{dR}(M) \cong H.$$

The first is the Hurwitz isomorphism, and the second is the de Rham theorem. Now from the observation

$$\mathrm{Hom}(\pi_1(M, q), \mathbb{R}) = \mathrm{Hom}(\mathrm{Ab}(\pi_1(M, q)), \mathbb{R})$$

and that Φ is exactly the isomorphism in the de Rham theorem, the claim follows. □

It follows from Corollary simply con 1 form **??** that $H^1_{dR}(M) = 0$ when M is simply connected. The next corollary generalizes that result.

Proposition 3.1.18. *A group G is called a **torsion group** if for each $g \in G$ there exists an integer k such that $g^k = 1$. If M is a connected smooth manifold whose fundamental group is a torsion group, then $H^1_{dR}(M) = 0$.*

Proof. There are no nontrivial homomorphisms from a torsion group to $\mathbb{R}$. □

mental finite **Corollary 3.1.19.** *If M is connected with finite fundamental group, then $H^1_{dR}(M) = 0$.*

Proof. Any finite group is torsion. □

3.1.3 The Mayer-Vietoris theorem

Suppose M is a smooth manifold with or without boundary, and U, V are open subsets ofM such that $M = U \cup V$. We have four inclusions,

$$\begin{tikzcd} & U \arrow[dr, "j_U"] & \\ U \cap V \arrow[ur, "i_U"] \arrow[dr, "i_V"'] & & M \\ & V \arrow[ur, "j_V"'] & \end{tikzcd}$$

which induce pullback maps on differential forms,

$$\begin{tikzcd} & \Omega^p(U) \arrow[dr, "i_U^*"] & \\ \Omega^p(M) \arrow[ur, "j_U^*"] \arrow[dr, "j_V^*"'] & & \Omega^p(U \cap V) \\ & \Omega^p(V) \arrow[ur, "i_V^*"'] & \end{tikzcd}$$

as well as corresponding induced cohomology maps. Note that these pullback maps are really just restrictions: for example, $j_U^*\omega = \omega|_U$. Consider the following sequence of maps:

$$0 \longrightarrow \Omega^p(M) \xrightarrow{j_U^* \oplus j_V^*} \Omega^p(U) \oplus \Omega^p(V) \xrightarrow{i_U^* - i_V^*} \Omega^p(U \cap V) \longrightarrow 0 \tag{1.3}$$

Mayer-1

Because pullbacks commute with d, these maps descend to linear maps on the corresponding de Rham cohomology groups.

Theorem 3.1.20 (Mayer-Vietoris). *Let M be a smooth manifold with or without boundary, and let U, V be open subsets of M whose union is M. For each p, there is a linear map $\delta : H^p_{dR}(U \cap V) \to H^{p+1}_{dR}(M)$ such that the following sequence, called the **Mayer-Vietoris sequence** for the open cover $\{U, V\}$, is exact:*

$$\cdots \longrightarrow H^p_{dR}(M) \overset{j_U^* \oplus j_V^*}{\longrightarrow} H^p_{dR}(U) \oplus H^p_{dR}(V) \overset{i_U^* - i_V^*}{\longrightarrow} H^p_{dR}(U \cap V) \xrightarrow{\delta} H^{p+1}_{dR}(M) \longrightarrow \cdots$$

Proof. Suppose M is a smooth manifold with or without boundary, and U, V are open subsets of M whose union is M. The heart of the proof is to show that the sequence (1.3) is exact for each p. Because pullback maps commute with the exterior derivative, (1.3) therefore defines a short exact sequence of cochain maps, and the Mayer-Vietoris theorem follows immediately from the zigzag lemma.

We begin by proving exactness at $\Omega^p(M)$, which just means showing that $j_U^* \oplus j_V^*$ is injective. Suppose that $\sigma \in \Omega^p(M)$ satisfies $(j_U^* \oplus j_V^*)\sigma = (\sigma|_U, \sigma|_V) = (0, 0)$. This means that the restrictions of σ to U and V are both zero. Since $\{U, V\}$ is an open cover of M, this implies that σ is zero.

To prove exactness at $\Omega^p(U) \oplus \Omega^p(V)$, first observe that

$$(i_U^* - i_V^*) \circ (j_U^* \oplus j_V^*) = 0,$$

which shows that $\operatorname{im}(j_U^* \oplus j_V^*) \subseteq \ker(i_U^* - i_V^*)$. Conversely, suppose we are given $(\eta, \eta') \in \Omega^p(U) \oplus \Omega^p(V)$ such that $i^*(\eta) - j^*(\eta') = 0$. This means that $\eta|_{U \cap V} = \eta'|_{U \cap V}$, so there is a global smooth p-form σ on M defined by

$$\sigma = \begin{cases} \eta & \text{on } U, \\ \eta' & \text{on } V. \end{cases}$$

so $\ker(i_U^* - i_V^*) \subseteq \operatorname{im}(j_U^* \oplus j_V^*)$.

Exactness at $\Omega^*(U \cap V)$ means that $i_U^* - i_V^*$ is surjective. This is the only nontrivial part of the proof, and the only part that really uses any properties of smooth manifolds and differential forms.

Let $\omega \in \Omega^p(U \cap V)$ be arbitrary. We need to show that there exist $\eta \in \Omega^p(U)$ and $\eta' \in \Omega^p(V)$ such that

$$\omega = (i_U^* - i_V^*)(\eta, \eta') = i_U^*\eta - i_V^*\eta' = \eta|_{U \cap V} - \eta'_{U \cap V}.$$

Let $\{\rho_U, \rho_V\}$ be a smooth partition of unity subordinate to the open cover $\{U, V\}$ of M, and define $\eta \in \Omega^p(U)$ by

$$\eta = \begin{cases} \rho_V \omega & \text{on } U \cap V, \\ 0 & \text{on } U \setminus \operatorname{supp}(\rho_V). \end{cases}$$

On the set $(U \cap V) \setminus \operatorname{supp}(\rho_V)$ where these definitions overlap, they both give zero, so this defines η as a smooth p-form on U. Similarly, define $\eta' \in \Omega^p(V)$ by

$$\eta' = \begin{cases} -\rho_U \omega & \text{on } U \cap V, \\ 0 & \text{on } U \setminus \operatorname{supp}(\rho_U). \end{cases}$$

Then we have

$$\eta|_{U\cap V} - \eta'|_{U \cap V} = \rho_V \omega + \rho_U \omega = \omega.$$

which was to be proved. □

Corollary 3.1.21. *The connecting homomorphism $\delta : H^p_{dR}(U \cap V) \to H^{p+1}_{dR}(M)$ in the Mayer-Vietoris sequence is defined as follows. For each $\omega \in Z^p(U \cap V)$, there are p-forms $\eta \in \Omega^p(U)$ and $\eta' \in \Omega^p(V)$ such that $\omega = \eta|_{U\cap V} - \eta'|_{U\cap V}$; and then $\delta[\omega] = [\sigma]$, where*

$$\sigma = \begin{cases} d\eta & \text{on } U, \\ d\eta' & \text{on } V. \end{cases}$$

If $\{\rho_U, \rho_V\}$ is a smooth partition of unity subordinate to $\{U, V\}$, we can take $\eta = \rho_V \omega$ and $\eta' = -\rho_U \omega$, both extended by zero outside the supports of ρ_U and ρ_V.

Proof. The characterization of the connecting homomorphism is given by the following diagram:

$$\begin{tikzcd}
\cdots \arrow[r, dotted] & \sigma \arrow[r] \arrow[d, dotted, no head] & (j_U^* \sigma, j_V^* \sigma) \arrow[r, dotted, no head] & \cdots \arrow[r, dotted, no head] & \cdots \\
\cdots & \cdots & (\eta, \eta') \arrow[u, "d"] \arrow[r, "i_U^* - i_V^*"] & \omega \arrow[u, dotted, no head] & \cdots
\end{tikzcd}$$

We find that $\delta[\omega] = \sigma$, provided there exists $(\eta, \eta') \in \Omega^p(U) \oplus \Omega^p(V)$ such that

$$i_U^* \eta - i_V^* \eta' = \omega, \quad (j_U^* \sigma, j_V^* \sigma) = (d\eta, d\eta').$$

It can be easily checked that our definition of σ satisfies the conditions. □

Computations using the Mayer-Vietoris sequence

Using the Mayer-Vietoris theorem, it is a simple matter to compute all of the de Rham cohomology groups of spheres.

Theorem 3.1.22 (Cohomology of Spheres). *For $n \geq 1$, the de Rham cohomology groups of S^n are*

$$H^p_{dR}(S^n) = \begin{cases} \mathbb{R} & p = 0 \text{ or } p = n, \\ 0 & \text{otherwise}. \end{cases}$$

The cohomology class of any smooth orientation form is a basis for $H^n_{dR}(S^n)$.

Proof. Proposition 3.1.6 shows that $H^0_{dR}(S^n) = \mathbb{R}$, so we need only prove for $p \geq 1$. We do so by induction on n. For $n = 1$, note first that any orientation form on S^1 has nonzero integral, so it is not exact by Corollary **??** thus $\dim H^1_{dR}(S^1) \geq 1$. On the other hand, Theorem 3.1.17 implies that there is an injective linear map from $H^1_{dR}(S^1)$ into $\operatorname{Hom}(\pi_1(S^1, 1), \mathbb{R})$, which is 1-dimensional. Thus, $H^1_{dR}(S^1)$ has dimension exactly 1, and is spanned by the cohomology class of any orientation form.

Next, suppose $n \geq 2$ and assume by induction that the theorem is true for S^{n-1}. Because S^n is simply connected, $H^1_{dR}(S^n) = 0$ by Corollary **??**. For $p > 1$, we use the Mayer-Vietoris theorem as follows. Let N and S be the north and south poles in S^n, respectively, and let $U = S^n - \{S\}, V = S^n - \{V\}$. By stereographic projection, both U and V are diffeomorphic to $\mathbb{R}^n$, and thus $U \cap V$ is diffeomorphic to $\mathbb{R}^n - \{0\}$.

Part of the Mayer-Vietoris sequence for $\{U, V\}$ read

$$H^{p-1}_{dR}(U) \oplus H^{p-1}_{dR}(V) \longrightarrow H^{p-1}_{dR}(U \cap V) \longrightarrow H^p_{dR}(S^n) \longrightarrow H^p_{dR}(U) \oplus H^p_{dR}(V)$$

Because U and V are diffeomorphic to $\mathbb{R}^n$, the groups on both ends are trivial when $p > 1$, which implies that $H^{p-1}_{dR}(U \cap V) \cong H^p_{dR}(S^n)$. Moreover, $U \cap V$ is diffeomorphic to $\mathbb{R}^n - \{0\}$ and therefore homotopy equivalent to S^{n-1}, so in the end we conclude that $H^p_{dR}(S^n) \cong H^{p-1}_{dR}(S^{n-1})$ for $p > 1$. As in the $n = 1$ case, any smooth orientation form on S^n determines a nonzero cohomology class, which therefore spans $H^n_{dR}(S^n)$. □

S^n exact iff **Corollary 3.1.23.** *Let ω be a closed n-form on S^n, then ω is exact if and only if $\int_{S^n} \omega = 0$.*

Proof. Let ω_{S^n} be an orientation form of S^n, then by Theorem 3.1.22 $\omega = a\omega_{S^n} + d\eta$ for some $a \in \mathbb{R}$ and $\eta \in \Omega^{n-1}(S^n)$. Since $\int_{S^n} d\eta = 0$, we get the claim. □

unctured R^n **Corollary 3.1.24 (Cohomology of Punctured Euclidean Space).** *Suppose $n \geq 2$ and $x \in \mathbb{R}^n$, and let $M = \mathbb{R}^n - \{x\}$. The only nontrivial de Rham groups of M are $H^0_{dR}(M)$ and $H^{n-1}_{dR}(M)$, both of which are 1-dimensional. A closed $(n-1)$-form ω on M is exact if and only if $\int_S \omega = 0$ for some (and hence every) $(n-1)$-dimensional sphere $S \subseteq M$ centered at x.*

The same statement remains true if $\mathbb{R}^n - \{x\}$ is replaced by $\mathbb{R}^n - \overline{B}$ for some closed ball $\overline{B} \subseteq \mathbb{R}^n$.

Proof. Let $S \subseteq M$ be any $(n-1)$-dimensional sphere centered at x. Because inclusion $\iota : S \hookrightarrow M$ is a homotopy equivalence, $H^p_{dR}(M) \to H^p_{dR}(S)$ is an isomorphism for each p, so the assertion about the dimension of $H^p_{dR}(M)$ follows from Theorem 3.1.22. If ω is a closed $(n-1)$-form on M, it follows that ω is exact if and only if $\iota^*\omega$ is exact on S, which in turn is true if and only $\int_S \iota^*\omega = 0$ by Corollary 3.1.23. □

ohomology U-x **Corollary 3.1.25.** *Suppose $n \geq 2$, $U \subseteq \mathbb{R}^n$ is any open subset, and $x \in U$. Then $H^{n-1}_{dR}(U - \{x\}) \neq 0$.*

Proof. Because U is open, there is an $(n-1)$-dimensional sphere S centered at x such that $S \subseteq U - \{x\}$. Let $\iota : S \hookrightarrow U - \{x\}$ be inclusion and $r : U - \{x\} \to S$ be the radial projection onto S. Then r and ι are smooth with $r \circ \iota = \mathrm{id}_S$. This implies $\iota^* \circ r^* = \mathrm{id}$, and therefore $r^* : H^{n-1}_{dR}(U - \{x\}) \to H^{n-1}_{dR}(S)$ is injective. Since $H^{n-1}_{dR}(S) \neq 0$ by Theorem 3.1.22, the result follows. □

Here is an important application of the topological invariance of the de Rham cohomology groups. Recall the theorem on invariance of dimension; it is a surprising fact that this purely topological theorem can be proved using de Rham cohomology. Before proving the theorem, we restate it here for convenience.

Proposition 3.1.26 (Topological Invariance of Dimension). *A nonempty n-dimensional topological manifold cannot be homeomorphic to an m-dimensional manifold unless $m = n$.*

Proof. If M is a topological n-manifold that is homeomorphic to an m-manifold, then M is itself both an n-manifold and an m-manifold. The case in which m or n is zero is easily proved, so assume that $m > n \geq 1$. Because M is an m-manifold, there is an open subset $V \subseteq M$ that is homeomorphic to $\mathbb{R}^m$. Because an open subset of an n-manifold is itself an n-manifold, any point $x \in V$ has a neighborhood $U \subseteq V$ that is homeomorphic to $\mathbb{R}^n$. On the one hand, because U is homeomorphic to $\mathbb{R}^n$, we can use the homeomorphism to define a smooth structure on U, and then $H^{m-1}_{dR}(U - \{x\})$ by Corollary cohomology punctured R^n 3.1.24. On the other hand, because U is homeomorphic to an open subset of $\mathbb{R}^m$, we can use that homeomorphism to define another smooth structure on U, and then Corollary cohomology U-x 3.1.25 implies that $H^{m-1}_{dR}(U - \{x\}) \neq 0$. This contradicts the topological invariance of de Rham cohomology. □

3.1.4 The de Rham cohomology with compact supports

For some purposes it is useful to define a generalization of the de Rham cohomology groups using only compactly supported forms. Let M be a smooth manifold with or without boundary and let $\Omega^p_c(M)$ denote the vector space of compactly supported smooth p-forms on M. The p-th **compactly supported de Rham cohomology group** of M is the quotient space

$$H^p_c(M) = \frac{\ker\left(d : \Omega^p_c(M) \to \Omega^{p+1}_c(M)\right)}{\operatorname{im}\left(d : \Omega^{p-1}_c(M) \to \Omega^p_c(M)\right)}.$$

Of course, when M is compact, this just reduces to ordinary de Rham cohomology. But for noncompact manifolds the two groups can be different.

As another application of Corollary cohomology punctured R^n 3.1.24, we prove a generalization of the Poincaré lemma for compactly supported forms. We will use it below to compute top-degree cohomology groups.

compact suup **Lemma 3.1.27 (Poincaré Lemma with Compact Support).** *Let $n \geq p \geq 1$, and suppose ω is a compactly supported closed p-form on $\mathbb{R}^n$. If $p = n$, suppose in addition that $\int_{\mathbb{R}^n} \omega = 0$. Then there exists a compactly supported smooth $(p-1)$-form ω on $\mathbb{R}^n$ such that $d\eta = \omega$.*

Proof. When $n = p = 1$, we can write $\omega = f\,dx$ for some smooth, compactly supported function $f \in C^\infty(\mathbb{R}^n)$. Define $F : \mathbb{R} \to \mathbb{R}$ by

$$F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

By the fundamental theorem of calculus, $dF = F' dx = \omega$. Choose $R > 0$ such that $(f) \subseteq [-R, R]$. When $x < -R$, $F(x) = 0$ by our choice of R. When $x > R$, the fact that $\int_{\mathbb{R}} \omega = 0$ translates to

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{+\infty} f(t)\,dt = 0.$$

so, in fact, $\operatorname{supp}(F) \subseteq [-R, R]$. This completes the proof for the case $n = p = 1$.

Now assume $n \geq 2$, and let $B, B' \subseteq \mathbb{R}^n$ be open balls centered at the origin such that $\operatorname{supp}(\omega) \subseteq B \subseteq \overline{B} \subseteq B'$. By the ordinary Poincaré lemma, there exists a smooth (but not necessarily compactly supported) $(p-1)$-form η_0 on $\mathbb{R}^n$ such that $d\eta_0 = \omega$. This implies, in particular, that $d\eta_0 = 0$ on $\mathbb{R}^n - \overline{B}$. To complete the proof, we consider three cases.

Consider $p = 1$. In this case η_0 is a smooth function. Because $\mathbb{R}^n - \overline{B}$ is connected when $n \geq 2$, it follows that η_0 is equal to a constant c there. Letting $\eta = \eta_0 - c$, we find that η is compactly supported and satisfies $d\eta = \omega$ as claimed.

Let $1 < p < n$. Now the restriction of η_0 to $\mathbb{R}^n - \overline{B}$ is a closed $(p-1)$-form. Because $H^{p-1}_{dR}(\mathbb{R}^n - \overline{B}) = 0$ by Corollary 3.1.24, there is a smooth $(p-2)$-form γ on $\mathbb{R}^n - \overline{B}$ such that $d\gamma = \eta_0$ there. If we let ψ be a smooth bump function that is supported in $\mathbb{R}^n - \overline{B}$ and equal to 1 on $\mathbb{R}^n - B'$, then $\eta = \eta_0 - d(\psi\gamma)$ is smooth on all of $\mathbb{R}^n$ and satisfies $d\eta = d\eta_0 = \omega$. Because $d(\psi\gamma) = d\gamma = \eta_0$ on $\mathbb{R}^n - B'$, η is compactly supported. Finally, let $p = n$. In this case, we cannot use the same argument because $H^{n-1}_{dR}(\mathbb{R}^n - \overline{B}) \neq 0$. However, it follows from Corollary 3.1.24 that the restriction of η_0 to $\mathbb{R}^n - \overline{B}$ is exact provided its integral is zero over some sphere centered at the origin and contained in $\mathbb{R}^n - \overline{B}$. Stokes's theorem implies that

$$0 = \int_{\mathbb{R}^n} \omega = \int_{\overline{B}'} \omega = \int_{\overline{B}'} d\eta_0 = \int_{\partial \overline{B}'} \eta_0.$$

Thus η_0 is exact on $\mathbb{R}^n - \overline{B}$, and the proof proceeds exactly as above. □

Theorem 3.1.28 (Compactly Supported Cohomology of $\mathbb{R}^n$). *For $n \geq 1$, the compactly supported de Rham cohomology groups of $\mathbb{R}^n$ are*

$$H^p_c(M) = \begin{cases} 0 & 0 \leq p < n, \\ \mathbb{R} & p = n. \end{cases}$$

Proof. It follows from Lemma 3.1.27 and Proposition 3.1.6 that $H^p_c(\mathbb{R}^n) = 0$ for $0 \leq p < n$.

For $H^n_c(\mathbb{R}^n)$, define a $\mathbb{R}$-linear map $\Phi : H^n_c(\mathbb{R}^n) \to \mathbb{R}$ by $\Phi[\omega] = \int_{\mathbb{R}^n} \omega$. Then Φ is injective by Lemma 3.1.27, and since there exist a form $\omega \in \Omega^n_c(\mathbb{R}^n)$ satisfying $\int_{\mathbb{R}^n} \omega \neq 0$, Φ is also surjective. □

There is also a Mayer-Vietoris sequence for compactly supported cohomology. But before taking it up we need to discuss the functorial properties of $\Omega^*_c(M)$, the algebra of forms with compact support on the manifold M. In general the pullback by a smooth map of a form with compact support need not have compact support; for example, consider the pullback of functions under the projection $M \times \mathbb{R} \to M$. So Ω^*_c is not a functor on the category of manifolds and smooth maps. However if we consider not all smooth maps, but only an appropriate subset of smooth maps, then Ω^*_c can be made into a functor. There are two ways in which this can be done.

(a) Ω^*_c is a **contravariant functor** under proper maps.

(b) Ω^*_c is a **covariant functor** under inclusions of open sets.

If $i : U \hookrightarrow M$ is the inclusion of the open subset U in the manifold M, then $i_* : \Omega^*_c(U) \to \Omega^*_c(M)$ is the map which extends a form on U by zero to a form on M.

Here we assume that Ω^*_c refers to the covariant functor in (b). There is also a Mayer-Vietoris sequence for this functor. As before, let M be covered by two open sets U and V. The sequence of inclusions

$$\begin{tikzcd} & U \arrow[dr, "j_U"] & \\ U \cap V \arrow[ur, "i_U"] \arrow[dr, "i_V"'] & & M \\ & V \arrow[ur, "j_V"'] & \end{tikzcd}$$

gives rise to

$$\begin{tikzcd} & \Omega_c^p(U) \arrow[dr, "j_*"] & \\ \Omega_c^p(U\cap V) \arrow[ur, "i_*"] \arrow[dr, "i_*"'] & & \Omega_c^p(M) \\ & \Omega_c^p(V) \arrow[ur, "j_*"'] & \end{tikzcd}$$

Consider the following sequence of maps:

$$0 \longrightarrow \Omega_c^p(U\cap V) \xrightarrow{i_*\oplus(-i_*)} \Omega_c^p(U)\oplus\Omega_c^p(V) \xrightarrow{\text{sum}} \Omega_c^p(M) \longrightarrow 0$$

Since these maps are just extension by zero, they descend to linear maps on the cohomological groups.

Proposition 3.1.29. *The Mayer-Vietoris sequence offorms with compact support*

$$0 \longrightarrow \Omega_c^p(U\cap V) \longrightarrow \Omega_c^p(U)\oplus\Omega_c^p(V) \longrightarrow \Omega_c^p(M) \longrightarrow 0$$

is exact. Therefore we have a long exact sequence on the cohomological groups with compact supports:

$$\cdots \longrightarrow H_c^p(U\cap V) \longrightarrow H_c^p(U)\oplus H_c^p(V) \longrightarrow H_c^p(M) \xrightarrow{\delta} H_c^{p+1}(U\cap V) \longrightarrow \cdots$$

Proof. This time exactness is easy to check at every step. Let $(\omega,\omega')\in\Omega_c^p(U)\oplus\Omega_c^p(V)$ be such that $\omega+\omega'\equiv 0$. Then since $\omega\equiv 0$ outside U and $\omega'\equiv 0$ outside V, we find that ω and ω' are both zero outside $U\cap V$, and $\omega=-\omega'$ on $U\cap V$. Therefore $(\omega,\omega')=(i_*\omega,-i_*\omega)$. This gives the exactness at $\Omega_c^p(U)\oplus\Omega_c^p(V)$.

Let ω be a form in $\Omega_c^p(M)$. Then ω is the image of $(\rho_U\omega,\rho_V\omega)$ in $\Omega_c^p(U)\oplus\Omega_c^p(V)$. The form $\rho_U\omega$ has compact support because $\operatorname{supp}(\rho_U\omega)\subseteq\operatorname{supp}(\rho_U)\cap\operatorname{supp}(\omega)$ and $\operatorname{supp}(\rho_U\omega)$ is closed. This shows the surjectivity of the map $\Omega_c^p(U)\oplus\Omega_c^p(V)\to\Omega_c^p(M)$. Note that whereas in the previous Mayer-Vietoris sequence we multiply by ρ_V to get a form on U, here $\rho_U\omega$ is a form on U. $\square$

Similarly, we can write explicitly the connection morphism $\delta: H_c^p(M)\to H_c^p(U\cap V)$. It turns out that it has the same construction as the previous case.

mpact connect **Corollary 3.1.30.** *The connection morphism $\delta: H_c^p(M)\to H_c^p(U\cap V)$ is defined as follows. For $\omega\in Z_c^p(M)$, there are p-forms $\eta\in\Omega_c^p(U)$ and $\eta'\in\Omega_c^p(V)$ such that $\omega=\eta+\eta'$; and then $\delta[\omega]=[\sigma]$ where $\sigma=d\eta=-d\eta'$ on $U\cap V$.*

If $\{\rho_U,\rho_V\}$ is a smooth partition of unity subordinate to $\{U,V\}$, we can take $\eta=\rho_U\omega$ and $\eta'=\rho_V\omega$, both extended by zero outside the supports of ρ_U and ρ_V.

3.1.5 The Poincaré duality

Now we provide a far-reaching generalization of the Poincaré lemma, which asserts that on an orientable manifold M we have a nondegenerate map

$$I: H_{dR}^p(M)\otimes H_c^{n-p}(M)\to\mathbb{R}$$

which induces an isomorphism $H_{dR}^p(M)\cong H_c^{n-p}(M)^*$.

Poincaré duality on an orientable manifold

Let M be a smooth manifold with or without boundary, and let $\omega \in \Omega^p(M), \eta \in \Omega^p(M)$ be closed forms. If $\omega = d\sigma$, then

$$\omega \wedge \eta = d\sigma \wedge \eta = d(\sigma \wedge \eta) - (-1)^{p-1}\sigma \wedge d\eta = d(\sigma \wedge \eta).$$

so $[\omega \wedge \eta] = 0$. Thus we conclude that $[\omega \wedge \eta]$ only depends on $[\omega]$ and $[\eta]$, and thus there is a well-defined bilinear map $\cup : H^p_{dR}(M) \times H^q_{dR}(M) \to H^{p+q}_{dR}(M)$, called the **cup product**, given by $[\omega] \cup [\eta] = [\omega \wedge \eta]$.

If in addition the manifold M is orientable, then we can consider the following map:

$$\int : H^p_{dR}(M) \otimes H^{n-p}_c(M) \to \mathbb{R}, \quad \omega \otimes \eta \mapsto \int_M \omega \wedge \eta.$$

(The compactly supportedness is assumed in order to take the integral.) Our first version of Poincaré duality asserts that this pairing is nondegenerate whenever M is orientable; equivalently,

$$H^p_{dR}(M) \cong (H^{n-p}_c(M))^*.$$

First, let's state a general principle that will use later. Consider a statement $\mathcal{P}(M)$ that meke sense for all manifolds and whose validity does not depend on the diffeomorphism type of M. It is often convenient to only consider the statement for the class of manifolds of a fixed dimension and/or with an orientation. We let $\mathscr{M}^n$ denote that class of all n-manifolds and $\mathscr{M}^n_+$ the class of oriented n-manifolds.

Theorem 3.1.31 (The induction principle). *The statement $\mathcal{P}(M)$ is true for all manifolds in $\mathscr{M}^n$, respectively $\mathscr{M}^n_+$, provided the following conditions hold:*

(a) $\mathcal{P}(\mathbb{R}^n)$ is true.

(b) If $U, V \subseteq M \in \mathscr{M}^n$ are open and $\mathcal{P}(U), \mathcal{P}(V), \mathcal{P}(U \cap V)$ are true, then $\mathcal{P}(U \cup V)$ is true.

(c) If $U_i \subseteq M \in \mathscr{M}^n$ form a countable collection of pairwise disjoint open sets such that and $\mathcal{P}(U_i)$ are true, then $\mathcal{P}(\bigcup_i U_i)$ is true.

Proof. If M is any smooth manifold, let us call an open cover $\{U_i\}$ of M a $\mathcal{P}$**-cover** if $\mathcal{P}$ holds for each subset U_i and every finite intersection $U_{i_1} \cap \cdots \cap U_{i_k}$. A $\mathcal{P}$-cover that is also a basis for the topology of M is called a $\mathcal{P}$**-basis** for M.

Step 1: If M has a finite $\mathcal{P}$-cover, then M satisfies $\mathcal{P}$. Suppose $M = U_1 \cup \cdots \cup U_k$, where the open subsets U_i and their finite intersections satisfy $\mathcal{P}$. We prove the result by induction on k. Assume the claim is true for smooth manifolds admitting a $\mathcal{P}$-cover with $k \geq 2$ sets, and suppose $\{U_1, \dots, U_{k+1}\}$ is a $\mathcal{P}$-cover of M. Define $U = U_1 \cup \cdots \cup U_k$ and $V = U_{k+1}$. The hypothesis implies that U and V satisfy $\mathcal{P}$, and $U \cap V$ also satisfy $\mathcal{P}$ because it has a k-fold $\mathcal{P}$-cover given by $\{U_1 \cap U_{k+1}, \dots, U_k \cap U_{k+1}\}$. Therefore, $M = U \cup V$ also satisfies $\mathcal{P}$ by condition (b).

Step 2: If M has a $\mathcal{P}$-basis, then M satisfies $\mathcal{P}$. Suppose $\{U_\alpha\}$ is a $\mathcal{P}$-basis for M. Let $f : M \to \mathbb{R}$ be an exhaustion function. For each integer i, define subsets A_i and A_i' of M by

$$A_i = f^{-1}([i, i+1]), \quad A_i' = f^{-1}((i - \frac{1}{2}, i + \frac{3}{2})).$$

For each point $p \in A_i$, there is a basis open subset in $\{U_\alpha\}$ containing p and contained in A_i'. The collection of all such basis sets is an open cover of A_i. Since f is an exhaustion function,

A_i is compact, and therefore it is covered by finitely many of these basis sets. Let B_i be the union of this finite collection of sets. By the condition of $\{U_\alpha\}$, this is a finite $\mathcal{P}$-cover of B_i, so by Step 1, B_i satisfies $\mathcal{P}$.

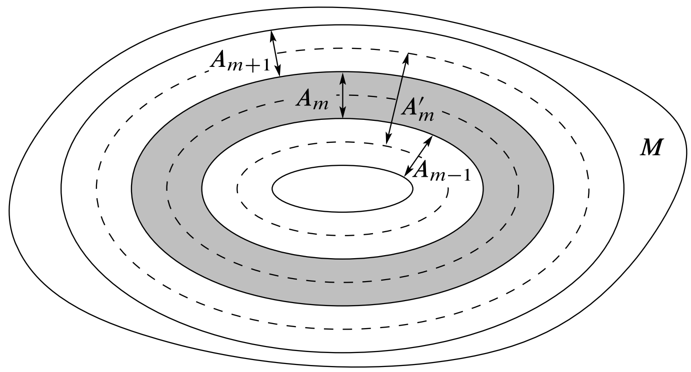

Figure 3.1: Proof of the induction principle, Step 2.

Observe that $B_i \subseteq A'_i$, so B_i can have nonempty intersection with B_j only when $|i-j| \le 1$. Therefore, if we define

$$U = \bigcup_{i \text{ odd}} B_i, \quad V = \bigcup_{i \text{ even}} B_i,$$

then U and V are disjoint unions of manifolds satisfying $\mathcal{P}$, and so they both satisfy $\mathcal{P}$ by condition (c). Finally, $\mathcal{P}$ holds on $U \cap V$ because it is the disjoint union of the sets $B_u \cap B_{i+1}$ for $i \in \mathbb{Z}$, each of which has a finite $\mathcal{P}$-cover consisting of sets of the form $U_\alpha \cap U_\beta$, where U_α and U_β are basis sets used to define B_i and B_{i+1}, respectively. Thus $M = U \cup V$ satisfies $\mathcal{P}$.

Step 3: Every open subset of $\mathbb{R}^n$ satisfies $\mathcal{P}$. If $U \subseteq \mathbb{R}^n$ is such a subset, then U has a basis consisting of Euclidean cubes. Because each cube is diffeomorphic to $\mathbb{R}^n$, and because any finite intersection of cubes is again a cube, finite intersections also satisfies $\mathcal{P}$. Thus, U has a $\mathcal{P}$-basis, so it satisfies $\mathcal{P}$ by Step 2.

Step 4: Every smooth manifold satisfies $\mathcal{P}$. Any smooth manifold has a basis of smooth coordinate domains. Since every smooth coordinate domain is diffeomorphic to an open subset of $\mathbb{R}^n$, as are their finite intersections, this is a $\mathcal{P}$-basis. The claim therefore follows from Step 2. $\square$

The following lemma is needed in the proof of the Poincaré duality.

Lemma 3.1.32. *The two Mayer-Vietoris sequences of the cover $\{U, V\}$ may be paired together to form a sign-commutative diagram*

$$\begin{array}{ccccccccc}
\longrightarrow H^p_{dR}(U \cup V) & \xrightarrow{\text{restriction}} & H^p_{dR}(U) \oplus H^p_{dR}(V) & \xrightarrow{\text{difference}} & H^p_{dR}(U \cap V) & \xrightarrow{\delta^*} & H^{p+1}(U \cup V) \longrightarrow \\
\otimes & & \otimes & & \otimes & & \otimes \\
\longleftarrow H^{n-p}_c(U \cup V) & \xleftarrow{\text{sum}} & H^{n-p}_c(U) \oplus H^{n-p}_c(V) & \longleftarrow & H^{n-p}_c(U \cap V) & \xleftarrow{\delta_*} & H^{n-p-1}_c(U \cup V) \longleftarrow \\
\downarrow \int_{U \cup V} & & \downarrow \int_U + \int_V & & \downarrow \int_{U \cap V} & & \downarrow \int_{U \cup V} \\
\mathbb{R} & & \mathbb{R} & & \mathbb{R} & & \mathbb{R}
\end{array}$$

Here sign-commutativity means, for instance, that

$$\int_{U\cap V} \omega \wedge \delta_* \eta = \pm \int_{U\cup V} \delta^* \omega \wedge \eta.$$

for $\omega \in H^p_{dR}(U\cap V)$, $\eta \in H^{n-p-1}_c(U\cup V)$.

Proof. The first two squares are in fact commutative: for $\omega \in H^p_{dR}(U\cup V)$ and $(\eta,\eta') \in H^{n-p}_c(U)\oplus H^{n-p}_c(V)$ we have

$$\int_{U\cup V} \omega\wedge\eta = \int_U \omega\wedge\eta, \quad \int_{U\cup V}\omega\wedge\eta = \int_V \omega\wedge\eta'$$

for η is zero outside U and η' is zero outside V. Therefore

$$\int_{U\cup V} \omega\wedge(\eta+\eta') = \int_U \omega\wedge\eta + \int_V \omega\wedge\eta'.$$

which is the commutativity of the first squre. Similarly, for $(\omega,\omega') \in H^p_{dR}(U)\oplus H^p_{dR}(V)$ and $\eta\in H^{n-p}_c(U\cap V)$ we have

$$\int_U \omega\wedge\eta = \int_{U\cap V}\omega\wedge\eta, \quad \int_U \omega'\wedge\eta = \int_{U\cap V}\omega'\wedge\eta,$$

and therefore

$$\int_U \omega\wedge(-\eta) + \int_V \omega'\wedge\eta = \int_{U\cap V}(\omega-\omega')\wedge\eta,$$

which is the commutativity of the second squre.

The sign problem emerges on the third squre. Let $\omega\in H^p_{dR}(U\cap V)$ and $\eta\in H^{n-p-1}_c(U\cup V)$. Recall that $\delta^*\omega$ is a form in $H^{p+1}(U\cup V)$ such that

$$\delta^*\omega = \begin{cases} d(\rho_V\omega) & \text{on } U,\\ d(-\rho_U\omega) & \text{on } V.\end{cases}$$

Similaly, $\delta_*\eta$ is a form in $H^{n-p}_c(U\cap V)$ such that $\delta_*\eta = d(\rho_U\eta)$. Note that $d(\rho_U\eta) = (d\rho_U)\wedge\eta$ and $d(\rho_V\omega) = (d\rho_V)\wedge\omega$ because ω and η are closed. Therefore

$$\int_{U\cap V}\omega\wedge\delta_*\eta = \int_{U\cap V}\omega\wedge(d\rho_U)\wedge\eta = (-1)^{\deg\omega}\int_{U\cap V}(d\rho_U)\wedge\omega\wedge\eta.$$

Since ρ_U and ρ_V are both constant outside $U\cap V$, we find that $\delta^*\omega$ has support in $U\cap V$, and therefore

$$\int_{U\cup V}\delta^*\omega\wedge\eta = \int_{U\cap V} d(\rho_V)\omega\wedge\eta = \int_{U\cap V}(d\rho_V)\wedge\omega\wedge\eta = -\int_{U\cap V}(d\rho_U)\wedge\omega\wedge\eta.$$

Thus finally we get

$$\int_{U\cap V}\omega\wedge\delta_*\eta = (-1)^{\det\omega+1}\int_{U\cup V}\delta^*\omega\wedge\eta.$$

This finishes the proof. □

Proposition 3.1.33 (Poincaré Duality). *Let M be an oriented manifold, then the pairing*

$$\int : H^p_{dR}(M)\otimes H^{n-p}_c(M)\to\mathbb{R}$$

is nondegenerate. In particular, the two cohomology groups $H^p_{dR}(M)$ and $H^{n-p}_c(M)$ are dual to each other and therefore have the same dimension provided they are finite-dimensional vector spaces.

Proof. It is easy to see that the pairing is well-defined. Next note that it defines a linear map

$$H^p_{dR}(M) \to H^{n-p}_c(M)^* = \mathrm{Hom}(H^{n-p}_c(M), \mathbb{R}).$$

We claim that this map is an isomorphism for all orientable but not necessarily connected manifolds.

By the Five Lemma if Poincaré duality holds for U, V, and $U \cap V$, then it holds for $U \cup V$. We now proceed by checking the conditions in the induction principle. For M diffeomorphic to $\mathbb{R}^n$, Poincaré duality follows from the two Poincaré lemmas. Next consider an arbitrary union of pairwise disjoint open sets. In this case we have

$$H^p_{dR}(\coprod_i U_i) = \prod_i H^p_{dR}(U_i) \quad H^{n-p}_c(\coprod_i U_i) = \bigoplus_i H^{n-p}_c(U_i).$$

so the claim also follows in this case. Then by Theorem the claim is valid for any oriented manifold. □

Corollary 3.1.34. *Let M be a compact oriented manifold. Then $H^p_{dR}(M)$ and $H^{n-p}_{dR}(M)$ are isomorphic.*

Proof. This requires that we know that $H^p_{dR}(M)$ is finite dimensional for all p. First note that if $O \subseteq \mathbb{R}^n$ is a finite union of open boxes, then the de Rham cohomology groups are finite dimensional by the Mayer-Vietoris sequence.

This will give the result for $M \subseteq \mathbb{R}^k$ as we can find a tubular neighborhood $M \subseteq U \subseteq \mathbb{R}^k$ and a retract $r : U \to M$. Now cover M by open boxes that lie in U and use compactness of M to find $M \subseteq O \subseteq U$ with O being a union of finitely many open boxes. Since $r|_M = \mathrm{id}_M$ the retract $r^* : H^p_{dR}(M) \to H^p_{dR}(O)$ is an injection so it follows that $H^p_{dR}(M)$ is finite dimensional. □

3.1.6 Cohomology computation on the top degree

The Poincaré duality gives an easy way to compute the top cohomology for an manifolds. In this part we state the main results.

compact supp **Theorem 3.1.35** (**Top Cohomology, Orientable Compact Support Case**). *If M is a connected orientable smooth n-manifold, then the integration map $\int : H^n_c(M) \to \mathbb{R}$ is an isomorphism, so $H^n_c(M)$ is 1-dimensional.*

Proof. This follows from the Poincaré duality, since $H^0_{dR}(M)$ consists of constant functions in this case. □

table compact **Theorem 3.1.36** (**Top Cohomology, Orientable Compact Case**). *If M is a compact connected orientable smooth n-manifold, then $H^n_{dR}(M)$ is 1-dimensional, and is spanned by the cohomology class of any smooth orientation form.*

Proof. This follows from the preceding theorem, because $H^p_{dR}(M) = H^p_c(M)$ in that case, and the integral of any orientation form is nonzero. □

le noncompact **Theorem 3.1.37** (**Top Cohomology, Orientable Noncompact Case**). *If M is a noncompact connected orientable smooth n-manifold, then $H^n_{dR}(M) = 0$.*

Proof. This also comes from the Poincaré duality: if M is connected and noncompact, then $H^0_c(M) = 0$. $\square$

Next we consider the nonorientable case. If M is a nonorientable smooth manifold, the key to analyzing its cohomology groups is the orientation covering $\widehat{\pi} : \widehat{M} \to M$. Because a finite-sheeted covering map is a proper map by Exercise covering map proper iff **??**, π induces cohomology maps on both compactly supported and ordinary de Rham cohomology. The next lemma shows that these maps are all injective.

Lemma 3.1.38. *Suppose M is a connected nonorientable smooth manifold and $\widehat{\pi} : \widehat{M} \to M$ is its orientation covering. For each p, the induced cohomology maps $\widehat{\pi}^* : H^p_{dR}(M) \to H^p_{dR}(\widehat{M})$ and $\widehat{\pi}^* : H^p_c(M) \to H^p_c(\widehat{M})$ are injective.*

Proof. First, we prove the lemma for compactly supported cohomology. Suppose ω is a closed, compactly supported p-form on M such that $\widehat{\pi}^*[\omega] = 0 \in H^p_c(\widehat{M})$. Then there exists $\eta \in \Omega^{p-1}_c(\widehat{M})$ such that $d\eta = \widehat{\pi}^*\omega$. Let $\alpha : \widehat{M} \to \widehat{M}$ be the unique nontrivial covering automorphism of $\widehat{M}$, and let $\widetilde{\eta} = \frac{1}{2}(\eta + \alpha^*\eta)$, which is also compactly supported. Using the fact that $\alpha \circ \alpha = \mathrm{id}_{\widehat{M}}$, we compute

$$\alpha^*\widetilde{\eta} = \frac{1}{2}(\alpha^*\eta + \alpha^* \circ \alpha^*\eta) = \widetilde{\eta}.$$

Because $\widehat{\pi} \circ \alpha = \widehat{\pi}$, this implies

$$d\widetilde{\eta} = \frac{1}{2}(d\eta + \alpha^* d\eta) = \frac{1}{2}(\widehat{\pi}^*\omega + \alpha^*\widehat{\pi}^*\omega) = \widehat{\pi}^*\omega.$$

Let $U \subseteq M$ be any evenly covered open subset. There are exactly two smooth local sections $\sigma_1, \sigma_2 : U \to \widehat{M}$ over U, which are related by $\sigma_2 = \alpha \circ \sigma_1$. Observe that

$$\sigma_2^*\widetilde{\eta} = \sigma_1^*\alpha^*\omega = \sigma_1^*\omega.$$

Therefore, we can define a smooth global $(p-1)$-form β on M by setting $\beta|_U := \sigma^*\widetilde{\eta}$ for any smooth local section $\sigma : U \to \widehat{M}$; the argument above guarantees that the various definitions agree where they overlap. Because $\operatorname{supp}(\beta) = \widehat{\pi}(\operatorname{supp}(\widetilde{\eta}))$, it follows that β is compactly supported. To determine the exterior derivative of β, given $p \in M$; choose a smooth local section σ defined on a neighborhood U of p, and compute

$$d\beta = d(\sigma^*\widetilde{\eta}) = \sigma^* d\widetilde{\eta} = \sigma^*\widehat{\pi}^*\omega = \omega.$$

because $\widehat{\pi} \circ \sigma = \mathrm{id}_U$.

The argument for ordinary de Rham cohomology is the same, but with all references to compact support deleted. $\square$

nonorientable **Theorem 3.1.39 (Top Cohomology, Nonorientable Case).** *If M is a connected nonorientable smooth n-manifold, then $H^n_c(M) = 0$ and $H^n_{dR}(M) = 0$.*

Proof. First consider the case of compactly supported cohomology. By the preceding lemma, it suffices to show that $\widehat{\pi}^* : H^p_c(M) \to H^p_c(\widehat{M})$ is the zero map, where $\widehat{\pi} : \widehat{M} \to M$ is the orientation covering of M. Let $\alpha : \widehat{M} \to \widehat{M}$ be the nontrivial covering automorphism as in the preceding proof. Now, α cannot be orientation-preserving: if it were, the entire covering automorphism group $\{\mathrm{id}, \alpha\}$ would be orientation-preserving, and then M would be orientable by Theorem orientation covering thm **??**. By connectedness of $\widehat{M}$ and the fact that α is a diffeomorphism, it follows that α is orientation-reversing.

Suppose ω is any compactly supported smooth n-form on M, and let $\widehat{\omega} = \widehat{\pi}^*\omega$. Because $\widehat{\pi}$ is proper, $\widehat{\omega}$ is compactly supported, and $\widehat{\pi} \circ \alpha = \widehat{\pi}$ implies $\alpha^*\widehat{\omega} = \widehat{\omega}$. Because α is orientation-reversing, we conclude from Proposition **??** that

$$\int_{\widehat{M}} \widehat{\omega} = -\int_{\widehat{M}} \alpha^*\widehat{\omega} = -\int_{\widehat{M}} \widehat{\omega}.$$

This implies that $\int_{\widehat{M}} \widehat{\omega} = 0$, so $[\widehat{\omega}] = 0 \in H_c^p(\widehat{M})$ by Theorem 3.1.35. This completes the proof that $H_c^n(M) = 0$.

It remains only to handle ordinary cohomology. If M is compact, it follows from the argument above that $H_{dR}^n(M) = H_c^n(M) = 0$. On the other hand, if M is noncompact, then so is $\widehat{M}$, and Theorem 3.1.37 shows that $H_{dR}^n(\widehat{M}) = 0$. It follows from the previous lemma that $H_{dR}^n(M) = 0$ as well. $\square$

3.1.7 Degree theory

Now that we know the top-degree cohomology groups of all compact smooth manifolds, we can use them to draw a number of significant conclusions about smooth maps between certain compact manifolds of the same dimension. They all follow from the fact that we can associate an integer to each such map, called its degree, in such a way that homotopic maps have the same degree.

Theorem 3.1.40 (Degree of a Smooth Map). *Suppose M and N are compact, connected, oriented, smooth manifolds of dimension n, and $F : M \to N$ is a smooth map. There exists a unique integer k, called the* ***degree of F****, that satisfies both of the following conditions.*

(a) For every smooth n-form ω on N,

$$\int_M F^*\omega = k \int_N \omega.$$

(b) If $q \in N$ is a regular value of F, then

$$k = \sum_{x \in F^{-1}(q)} \operatorname{sgn}(x),$$

where $\operatorname{sgn}(x) = 1$ if dF_x is orientation-preserving, and -1 if it is orientation-reversing.

Proof. By Theorem 3.1.36, two smooth n-forms on either M or N are cohomologous if and only if they have the same integral. Let θ be any smooth n-form on N such that $\int_N \theta = 1$, and let $k = \int_M F^*\theta$. If $\omega \in \Omega^n(N)$ is arbitrary, then ω is cohomologous to $a\theta$, where $a = \int_N \omega$, and therefore $F^*\omega$ is cohomologous to $aF^*\theta$. It follows that

$$\int_M F^*\omega = a \int_M F^*\theta = ak = k \int_N \omega.$$

Thus k satisfies (a), and is clearly the only number that does so.

Next we show that k also has the characterization given in part (b), from which it follows that it is an integer. Let $q \in N$ be an arbitrary regular value of F. Because $F^{-1}(q)$ is a properly embedded 0-dimensional submanifold of M, it is finite. Suppose first that $F^{-1}(q)$ is not empty—say, $F^{-1}(q) = \{x_1, \ldots, x_r\}$. By the inverse function theorem, for each i there is a neighborhood U_i of x_i such that F is a diffeomorphism from U_i to a neighborhood W_i of q, and by shrinking the U_i's if necessary, we may assume that they are pairwise disjoint.

Then $K := M - \bigcup_{i=1}^{r} U_i$ is closed in M and thus compact, so $F(K)$ is closed in N and disjoint from q. Let W be the connected component of $\bigcap_{i=1}^{r} W_i \cap (N - F(K))$ containing q, and let $V_i = F^{-1}(W) \cap U_i$. It follows that W is a connected neighborhood of q whose preimage under F is the disjoint union $V_1 \cup \cdots \cup V_r$, and F restricts to a diffeomorphism from each V_i to W. Since each V_i is connected, the restriction of F to V_i must be either orientation-preserving or orientation-reversing.

Let ω be a smooth n-form on N that is compactly supported in W and satisfies $\int_N \omega = \int_W \omega = 1$. It follows from part (a) that $\int_M F^*\omega = k$. Since $F^*\omega$ is compactly supported in $F^{-1}(W)$, we have $\int_M F^*\omega = \sum_{i=1}^{r} \int_{V_i} F^*\omega = k$. From Proposition int form prop **??**(d), since F restricts to a diffeomorphism on each V_i, we conclude that $\int_{V_i} \omega = \pm 1$, with the positive sign if F is orientation-preserving on V_i and the negative sign otherwise. This proves (b) when $F^{-1}(q) \neq \varnothing$.

On the other hand, suppose $F^{-1}(q) = \varnothing$. Then q has a neighborhood W contained in $N - F(M)$ (because $F(M)$ is compact and thus closed). If ω is any smooth n-form on N that is compactly supported in W, then $\int_M F^*\omega = 0$, so $k = 0$. This proves (b). □

Corollary 3.1.41. *With the assumptions above, if F is not surjective then $\deg F = 0$.*

Much of the power of degree theory arises from the fact that the two different characterizations of the degree can be played off against each other. For example, it is often easy to compute the degree of a particular map simply by counting the points in the preimage of a regular value, with appropriate signs. On the other hand, the characterization in terms of differential forms makes it easy to prove many important properties, such as the ones given in the next proposition.

degree prop **Proposition 3.1.42 (Properties of the Degree).** *Suppose M, N, and P are compact, connected, oriented, smooth n-manifolds.*

(a) If $F : M \to N$ and $G : N \to P$ are both smooth maps, then $\deg(G \circ F) = (\deg G)(\deg F)$.

(b) If $F : M \to N$ is a diffeomorphism, then $\deg F = +1$ if F is orientation-preserving and -1 if it is orientation-reversing.

(c) If two smooth maps $F_0, F_1 : M \to N$ are homotopic, then they have the same degree.

Proof. Part (b) and (c) follows directly from definition. For (a), we just note that

$$\int_P (G \circ F)^*\omega = \int_P F^*G^*\omega = (\deg F)\int_N G^*\omega = (\deg F)(\deg G)\int_M \omega,$$

which proves $\deg(G \circ F) = (\deg F)(\deg G)$. □

This proposition allows us to define the degree of a continuous map $F : M \to N$ between compact, connected, oriented, smooth n-manifolds, by letting $\deg F$ be the degree of any smooth map that is homotopic to F. The Whitney approximation theorem guarantees that there is such a map, and the preceding proposition guarantees that the degree is the same for every map homotopic to F. Here are some applications of degree theory.

Proposition 3.1.43. *Let $F : M \to N$ be a proper local diffeomorphism of degree ± 1 between oriented connected manifolds, then F is a diffeomorphism.*

Proof. The fact that $\deg F \neq 0$ means F is surjective, and thus a covering map. Now $\deg F = \pm 1$ means F is injective, so it must be a diffeomorphism. $\square$

Proposition 3.1.44. *Even dimensional spheres do not admit non-vanishing smooth vector fields.*

Proof. Let X be a vector field on S^n, we can scale it so that it is a unit vector field. If we consider it as a function $X : S^n \to S^n \subseteq \mathbb{R}^{n+1}$ then it is always perpendicular to its foot point. We can then create a homotopy

$$H(p,t) = p\cos(\pi t) + X_p \sin(\pi t).$$

Since $p \perp X_p$ and both are unit vectors the Pythagorean theorem shows that $H(p,t) \in S^n$ as well. When $t = 0$ the homotopy is the identity, and when $t = 1$ it is the antipodal map. Since the antipodal map reverses orientations on even dimensional spheres it is not possible for the identity map to be homotopic to the antipodal map. $\square$

Theorem 3.1.45. *Suppose N is a compact, connected, oriented, smooth n-manifold, and X is a compact, oriented, smooth $(n+1)$-manifold with connected boundary. If $f : \partial X \to N$ is a continuous map that has a continuous extension to X, then $\deg f = 0$.*

Proof. Suppose f has an extension to a continuous map $F : X \to N$. By the Whitney approximation theorem, there is a smooth map $\widetilde{F} : X \to N$ that is homotopic to F. Replacing F by $\widetilde{F}$ and f by $\widetilde{F}|_{\partial X}$ we may assume that both f and F are smooth.

Let ω be any smooth n-form on N. Then $d\omega = 0$ because it is an $(n+1)$-form on an n-manifold. From Stokes's theorem, we obtain

$$\int_{\partial X} f^*\omega = \int_{\partial X} F^*\omega = \int_X d(F^*\omega) = \int_X F^*(f\omega) = 0.$$

It follows from Theorem 3.1.40 that f has degree zero. $\square$

Theorem 3.1.46 (Brouwer Fixed-Point Theorem). *Every continuous map from $\overline{B}^n$ to itself has a fixed point.*

Proof. Suppose for the sake of contradiction that $F : \overline{B}^n \to \overline{B}^n$ is continuous and has no fixed points. We can define a continuous map

$$G(x) = \frac{x - F(x)}{|x - F(x)|},$$

and let $g = G|_{S^{n-1}} : S^{n-1} \to S^{n-1}$. On the one hand, the previous theorem implies that g has degree zero. On the other hand, consider the map $H : S^{n-1} \times I \to S^{n-1}$ defined by

$$H(x,t) = \frac{x - tF(x)}{|x - tF(x)|}$$

The denominator never vanishes when $t = 1$ because F has no fixed points, and when $t < 1$ it cannot vanish because $|x| = 1$ while $|tF(x)| = t < 1$. Thus H is continuous, so it is a homotopy from the identity to g. It follows from Proposition 3.1.42 that g has degree 1, which is a contradiction. $\square$

3.1.8 The de Rham theorem

The topological invariance of the de Rham groups suggests that there should be some purely topological way of computing them. There is indeed, and the connection between the de

Rham groups and the singular cohomology groups was first proved by Georges de Rham himself in the 1930s. The theorem that bears his name is a major landmark in the development of smooth manifold theory. The purpose of this part is to give a proof of this theorem.

Smooth singular homology

The connection between the singular and de Rham cohomology groups will be established by integrating differential forms over singular chains. More precisely, given a singular p-simplex σ in a manifold M and a p-form ω on M, we would like to pull ω back by σ and integrate the resulting form over Δ_p. However, there is an immediate problem with this approach, because forms can be pulled back only by smooth maps, while singular simplices are in general only continuous. In this part we overcome this problem by showing that singular homology can be computed equally well with smooth simplices.

If M is a smooth manifold, a smooth p-simplex in M is a map $\sigma : \Delta_p \to M$ that is smooth in the sense that it has a smooth extension to a neighborhood of each point. The subgroup of $C_p(M)$ generated by smooth simplices is denoted by $C_p^\infty(M)$ and called the **smooth chain group in degree** $\boldsymbol{p}$. Elements of this group, which are finite formal linear combinations of smooth simplices, are called smooth chains. Because the boundary of a smooth simplex is a smooth chain, we can define the $\boldsymbol{p}$**-th smooth singular homology group** of M to be the quotient group

$$H_p^\infty(M) = \frac{\ker(\partial : C_p^\infty(M) \to C_{p-1}^\infty(M))}{\operatorname{im}(\partial : C_{p+1}^\infty(M) \to C_p^\infty(M))}.$$

The inclusion map $\iota : C_p^\infty(M) \hookrightarrow C_p(M)$ commutes with the boundary operator, and so induces a map on homology: $\iota_* : H_p^\infty(M) \to H_p(M)$ by $\iota_*[c] = [\iota(c)]$.

Theorem 3.1.47 (Smooth Singular vs. Singular Homology). *For any smooth manifold M, the map $\iota_* : H_p^\infty(M) \to H_p(M)$ induced by inclusion is an isomorphism.*

The basic idea of the proof is to construct, with the help of the Whitney approximation theorem, two operators: first, a smoothing operator $s : C_p(M) \to C_p^\infty(M)$ such that $s \circ \partial = \partial \circ s$ and $s \circ \iota$ is the identity on $C_p^\infty(M)$; and second, a homotopy operator that shows that $\iota \circ s$ induces the identity map on $H^p(M)$. Since the details are highly technical, we do not present them here.

The de Rham theorem

Suppose M is a smooth manifold, ω is a closed p-form on M; and σ is a smooth p-simplex in M. We define the integral of ω over σ to be

$$\int_\sigma \omega = \int_{\Delta_p} \sigma^* \omega.$$

This makes sense because Δ_p is a smooth p-submanifold with corners embedded in $\mathbb{R}^p$, and it inherits the orientation of $\mathbb{R}^p$. (Or we could just consider Δ_p as a domain of integration in $\mathbb{R}^p$.) Observe that when $p = 1$, this is the same as the line integral of ω over the smooth curve segment $\sigma : [0,1] \to M$. If $c = \sum_{i=1}^k c_i \sigma_i$ is a smooth p-chain, the integral of ω over c is defined as

$$\int_c \omega = \sum_{i=1}^{k} c_i \int_{\sigma_i} \omega.$$

em for chains

Theorem 3.1.48 (Stokes's Theorem for Chains). *If c is a smooth p-chain in a smooth manifold M, and ω is a smooth $(p-1)$-form on M, then*

$$\int_c d\omega = \int_{\partial c} \omega.$$

Proof. It suffices to prove the theorem when c is just a smooth simplex σ. Since Δ_p is a manifold with corners, Stokes's theorem says that

$$\int_\sigma d\omega = \int_{\Delta_p} \sigma^* d\omega = \int_{\Delta_p} d\sigma^*\omega = \int_{\partial\Delta_p} \sigma^*\omega.$$

The face maps $F_{i,p} : \Delta_{p-1} \to \Delta_p$ are parametrizations of the boundary faces of Δ_p satisfying the conditions of Proposition int mani para ??, except possibly that they might not be orientation-preserving. To check the orientations, note that $F_{i,p}$ is the restriction to $\Delta_p \cap \partial\mathbb{H}^p$ of the affine diffeomorphism sending the simplex $[e_0, \ldots, e_p]$ to $[e_0, \ldots, \widehat{e_i}, \ldots, e_p, e_i]$. This is easily seen to be orientation-preserving if and only if $(e_0, \ldots, \widehat{e_i}, \ldots, e_p, e_i)$ is an even permutation of $(e_0, \ldots, e_p)$, which is the case if and only if $p - i$ is even. Since the standard coordinates on $\partial\mathbb{H}^p$ are positively oriented if and only if p is even, the upshot is that $F_{i,p}$ is orientation-preserving for $\partial\Delta_p$ if and only if i is even. Thus, by Proposition int mani para ??,

$$\int_{\partial\Delta_p} \sigma^*\omega = \sum_{i=0}^{p} (-1)^i \int_{\Delta_{p-1}} F_{i,p}^*\sigma^*\omega = \sum_{i=0}^{p} (-1)^i \int_{\Delta_{p-1}} (\sigma \circ F_{i,p})^*\omega = \sum_{i=0}^{p} (-1)^i \int_{\sigma\circ F_{i,p}} \omega.$$

By definition of the singular boundary operator, this is equal $\int_{\partial\sigma} \omega$. □

Using this theorem, we define a natural linear map $\mathscr{I} : H^p_{dR}(M) \to H^p(M;\mathbb{R})$, called the **de Rham homomorphism**, as follows. For any $[\omega] \in H^p_{dR}(M)$ and $[c] \in H_p(M) \cong H_p^\infty(M)$, we define

$$\mathscr{I}[\omega][c] = \int_{\widetilde{c}} \omega.$$

where $\widetilde{c}$ is any smooth p-cycle representing the homology class $[c]$. This is well defined by Theorem Stokes's theorem for chains 3.1.48.

sm naturality

Proposition 3.1.49 (Naturality of the de Rham Homomorphism). *For a smooth manifold M and nonnegative integer p, let $\mathscr{I} : H^p_{dR}(M) \to H^p(M;\mathbb{R})$ denote the de Rham homomorphism.*

(a) If $F : M \to N$ is a smooth map, then the following diagram commutes:

$$\begin{array}{ccc} H^p_{dR}(N) & \xrightarrow{F^*} & H^p_{dR}(M) \\ \downarrow \mathscr{I} & & \downarrow \mathscr{I} \\ H^p(N;\mathbb{R}) & \xrightarrow{F^*} & H^p(M;\mathbb{R}) \end{array}$$

(b) If M is a smooth manifold and U, V are open subsets of M whose union is M, then the following diagram commutes:

$$\begin{array}{ccc} H^{p-1}_{dR}(U\cap V) & \xrightarrow{\delta} & H^p_{dR}(M) \\ \downarrow \mathscr{I} & & \downarrow \mathscr{I} \\ H^{p-1}(U\cap V;\mathbb{R}) & \xrightarrow{\partial^*} & H^{ppp}(M;\mathbb{R}) \end{array}$$

where δ and ∂^* are the connecting homomorphisms of the Mayer-Vietoris sequences for de Rham and singular cohomology, respectively.

Proof. Directly from the definitions, if σ is a smooth p-simplex in M and ω is a smooth p-form on N,

$$\int_\sigma F^*\omega = \int_{\Delta_p} \sigma^* F^*\omega = \int_{\Delta_p} (F\circ\sigma)^*\omega = \int_{F\circ\sigma} \omega.$$

This implies

$$\mathscr{I}[F^*\omega][\sigma] = \mathscr{I}[\omega][F\circ\sigma] = F^*(\mathscr{I}[\omega])[\sigma].$$

which proves (a).

Now consider (b). Commutativity of this diagram means

$$\mathscr{I}(\delta[\omega])[e] = (\partial^*\mathscr{I}[\omega])[e]$$

for any $[\omega] \in H^{p-1}_{dR}(U\cap V)$ and any $[e] \in H^p_{dR}(M)$. Using our identification of $H^p(M;\mathbb{R})$ with $\operatorname{Hom}(H^p(M),\mathbb{R})$, we can rewrite this as

$$\mathscr{I}(\delta[\omega])[e] = \mathscr{I}[\omega](\partial_*[e]).$$

If σ is a smooth p-form representing $\delta[\omega]$ and c is a smooth $(p-1)$-chain representing $\partial_*[e]$, this is the same as

$$\int_e \sigma = \int_c \omega.$$

By the characterization of ∂_*, we can let $c = \partial f$, where f, f' are smooth p-chains in U and V, respectively, such that $f + f'$ represents the same homology class as e. Similarly, by `Mayer-Vietoris de Rham connect` Corollary 3.1.21, we can choose $\eta \in \Omega^{p-1}(U)$ and $\eta' \in \Omega^{p-1}(V)$ such that $\omega = \eta|_{U\cap V} - \eta'|_{U\cap V}$, and then let σ be the p-form that is equal to $d\eta$ on U and to $d\eta'$ on V. Then, because $\partial f + \partial f' = \partial e = 0$ and $d\eta|_{U\cap V} - d\eta'|_{U\cap V} = d\omega = 0$, we have

$$\begin{aligned}\int_c \omega &= \int_{\partial f}\omega = \int_{\partial f}\eta - \int_{\partial f}\eta' = \int_{\partial f}\eta + \int_{\partial f'}\eta' \\ &= \int_f d\eta + \int_{f'} d\eta' = \int_f \sigma + \int_{f'}\sigma = \int_e \sigma.\end{aligned}$$

Thus the diagram commutes. □

Theorem 3.1.50 (de Rham). *For every smooth manifold M and nonnegative integer p, the de Rham homomorphism $\mathscr{I} : H^p_{dR}(M) \to H^p(M;\mathbb{R})$ is an isomorphism.*

Proof. We will use the induction principle as in the proof of Poincaré duality. Thus we need to check the three conditions.

Consider $\mathbb{R}^n$. The homopoty invariant of H^p implies that the singular cohomology groups of $\mathbb{R}^n$ are also trivial for $p \neq 0$. In the $p = 0$ case, $H^0_{dR}(\mathbb{R}^n)$ is the 1-dimensional space consisting of the constant functions, and $H^0(\mathbb{R}^n;\mathbb{R}) = \operatorname{Hom}(H_0(\mathbb{R}^n),\mathbb{R})$ is also 1-dimensional because $H_0(\mathbb{R}^n)$ is generated by any singular 0-simplex. If $\sigma : \Delta_0 \to M$ is a singular 0-simplex (which is smooth because any map from a 0-manifold is smooth), and f is the constant function equal to 1, then

$$\mathscr{I}[f][\sigma] = \int_{\Delta_0} \sigma^* f = (f\circ\sigma)(0) = 1.$$

Thus $\mathscr{I} : H^0_{dR}(\mathbb{R}^n) \to H^0(\mathbb{R}^n;\mathbb{R})$ is not the zero map, so it is an isomorphism.

If U, V are open subsets of M such that the de Rham theorem is true on U, V and $U \cap V$, then putting together the Mayer-Vietoris sequences for de Rham and singular cohomology, we obtain the following commutative diagram, in which the horizontal rows are exact and the vertical maps are all de Rham homomorphisms:

$$\begin{tikzcd}
H^{p-1}_{dR}(U) \oplus H^{p-1}_{dR}(V) \arrow[r] \arrow[d] & H^{p-1}_{dR}(U \cup V) \arrow[r] \arrow[d] & H^{p}_{dR}(U \cup V) \arrow[r] \arrow[d] & H^{p}_{dR}(U) \oplus H^{p}_{dR}(V) \arrow[d] \\
H^{p-1}(U;\mathbb{R}) \oplus H^{p-1}(V;\mathbb{R}) \arrow[r] & H^{p-1}(U \cup V;\mathbb{R}) \arrow[r] & H^{p}(U \cup V;\mathbb{R}) \arrow[r] & H^{p}(U;\mathbb{R}) \oplus H^{p}(V;\mathbb{R})
\end{tikzcd}$$

The commutativity of the diagram is an immediate consequence of Proposition 3.1.49. Then by the five lemma the de Rham theorem is true on $U \cup V$.

Finally, consider a disjoint union $\amalg_i U_i$. For both de Rham and singular cohomology the inclusions $\iota_i : U_i \hookrightarrow \amalg_i U_i$ induce isomorphisms between the cohomology groups of the disjoint union and the direct product of the cohomology groups of the manifolds U_i. By Proposition 3.1.49, $\mathscr{I}$ commutes with these isomorphisms. Now use the induction principal we get the claim. $\square$

Recall that by Theorem **??** the inclusion $\iota : \operatorname{Int} M \hookrightarrow M$ is a homotopy equivalence, so for manifolds with boundary the de Rham theorem also holds.

Proposition 3.1.51. *For every smooth manifold with boundary M and nonnegative integer p, the de Rham homomorphism $\mathscr{I} : H^p_{dR}(M) \to H^p(M;\mathbb{R})$ is an isomorphism.*

3.1.9 The Thom isomorphism

Compact vertical cohomology and integration along the fiber

For vector bundles there is a third kind of cohomology. Instead of $\Omega^p_c(E)$, the complex of forms with compact support, we consider $\Omega^p_{cv}(E)$, the complex of forms with compact support in the vertical direction, defined as follows: a smooth n-form ω on E is in $\Omega^p_{cv}(E)$ if and only if for every compact set K in M, $\operatorname{supp}(\omega) \cap \pi^{-1}(K)$ is compact. If $\omega \in \Omega^p_{cv}(E)$, then since $\operatorname{supp}(\omega|_{\pi^{-1}(p)}) \subseteq \operatorname{supp}(\omega) \cap \pi^{-1}(p)$ is a closed subset of a compact set, $\operatorname{supp}(\omega|_{\pi^{-1}(p)})$ is compact. Thus, although a form in $\Omega^p_{cv}(E)$ need not have compact support in E, its restriction to each fiber has compact support. The cohomology of this complex, denoted $\Omega^p_{cv}(E)$, is called the **cohomology of E with compact support in the vertical direction**, or **compact vertical cohomology**.

Let E be oriented as a rank k vector bundle. We define the **integration along the fiber**, $\Omega^p_{cv}(E) \to \Omega^{p-k}(M)$, as follows. First consider the case of a trivial bundle $E = M \times \mathbb{R}^n$. Let $(t^1, \ldots, t^k)$ be the coordinates on the fiber $\mathbb{R}^k$. A form on E is a real linear combination of two types of forms: the type (i) forms are those which do not contain as a factor the k-form $dt_1 \wedge \cdots \wedge dt_k$ and the type (ii) forms are those which do. The map π_* is defined by

(i) $(\pi^*\phi) \wedge f(x,t)\, dt^{i_1} \wedge \cdots \wedge dt^{i_r} \mapsto 0,\ r < k.$

(ii) $(\pi^*\phi) \wedge f(x,t)\, dt^{1} \wedge \cdots \wedge dt^{k} \mapsto \phi \int_{\mathbb{R}^k} f(x,t)\, dt^{1} \wedge \cdots \wedge dt^{k}.$

where f has compact support for each fixed x in M and ϕ is a form on M.

Next suppose E is an oriented vector bundle, with oriented trivialization $\{(U_\alpha, \Phi_\alpha)\}$. Let $(x^1, \dots, x^n)$ and $(\widetilde{x}^1, \dots, \widetilde{x}^n)$ be the coordinate functions on U_α and U_β and $(t^1, \dots, t^k)$, $(\widetilde{t}^1, \dots, \widetilde{t}^k)$ the fiber coordinates on $E|_{U_\alpha}$ and $E|_{U_\beta}$, given by Φ_α, Φ_β respectively. Because $\{(U_\alpha, \Phi_\alpha)\}$ is an oriented trivialization for E, the two sets of fiber coordinates $(t^1, \dots, t^k)$ and $(\widetilde{t}^1, \dots, \widetilde{t}^k)$ are related by an element of $\mathrm{GL}_n^+(\mathbb{R})$ at each point of $U_\alpha \cap U_\beta$. Again a form ω in $\Omega^p_{cv}(E)$ is locally of type (i) or (ii). The map π_* is defined to be zero on type (i) forms. To define π_* on type (ii) forms, write ω_α for $\omega|_{U_\alpha}$. Then

$$\omega_\alpha = (\pi^*\phi) \wedge f(x,t)\, dt^1 \wedge \cdots \wedge dt^k, \quad \omega_\beta = (\pi^*\tau) \wedge g(\widetilde{x}, \widetilde{t})\, d\widetilde{t}^1 \wedge \cdots \wedge d\widetilde{t}_k.$$

Define

$$\pi_*\omega_\alpha = \phi \int_{\mathbb{R}^k} f(x,t)\, dt^1 \wedge \cdots \wedge dt^k.$$

By the invariance of the integral under orientation-preserving diffeomorphism, these definitions coincide on their overlap. Hence $\{\omega_\alpha\}$ piece together to give a global form $\pi_*\omega$ on M. Furthermore, this definition is independent of the choice of the oriented trivialization for E.

Proposition 3.1.52. *Integration along the fiber commutes with exterior differentiation.*

Proof. By a partition of unity, we may assume E to be the product bundle $M \times \mathbb{R}^k$. If $\omega = (\pi^*\phi) \wedge f(x,t)\, dt^1 \wedge \cdots \wedge dt^k$, then we have

$$\begin{aligned}
d(\pi_*\omega) &= d(\phi \int_{\mathbb{R}^k} f(x,t)\, dt_1 \wedge \cdots \wedge dt_k) \\
&= (d\phi) \int_{\mathbb{R}^k} f(x,t)\, dt^1 \wedge \cdots \wedge dt^k + (-1)^{\deg\phi} \phi \wedge \Big(d \int_{\mathbb{R}^k} f(x,t)\, dt^1 \wedge \cdots \wedge dt^k \Big) \\
&= (d\phi) \int_{\mathbb{R}^k} f(x,t)\, dt^1 \wedge \cdots \wedge dt^k + (-1)^{\deg\phi} \phi \wedge \Big(\int_{\mathbb{R}^k} \sum_{i=1}^n \frac{\partial f}{\partial x^i}(x,t) dx^i \wedge dt^1 \wedge \cdots \wedge dt^k \\
&\qquad + \int_{\mathbb{R}^k} \sum_{j=1}^k \frac{\partial f}{\partial t^i}(x,t) dt^j \wedge dt^1 \wedge \cdots \wedge dt^k \Big) \\
&= (d\phi) \int_{\mathbb{R}^k} f(x,t)\, dt^1 \wedge \cdots \wedge dt^k + (-1)^{\deg\phi} \phi \wedge dx^i \int_{\mathbb{R}^k} \sum_{i=1}^n \frac{\partial f}{\partial x^i}(x,t)\, dt^1 \wedge \cdots \wedge dt^k.
\end{aligned}$$

And

$$\begin{aligned}
\pi_*(d\omega) &= \pi_* \Big((\pi^* d\phi) \wedge f(x,t)\, dt^1 \wedge \cdots \wedge dt^k + \sum_i (-1)^{\deg\phi} (\pi^*\phi) \wedge \frac{\partial f}{\partial x^i} dx^i \wedge dt^1 \wedge \cdots \wedge dt^k \Big) \\
&= \pi_* \Big((\pi^* d\phi) \wedge f(x,t)\, dt^1 \wedge \cdots \wedge dt^k + \pi^* \Big(\sum_i (-1)^{\deg\phi} \phi \wedge dx^i \Big) \wedge \frac{\partial f}{\partial x^i} dx^i \wedge dt^1 \wedge \cdots \wedge dt^k \Big) \\
&= (d\phi) \int_{\mathbb{R}^k} f(x,t)\, dt^1 \wedge \cdots \wedge dt^k + (-1)^{\deg\phi} \phi \wedge dx^i \int_{\mathbb{R}^k} \sum_{i=1}^n \frac{\partial f}{\partial x^i}(x,t)\, dt^1 \wedge \cdots \wedge dt^k.
\end{aligned}$$

So $d\pi_* = \pi_* d$ for a type (ii) form. Next let $\omega = (\pi^*\phi) \wedge f(x,t)\, dt^{i_1} \wedge \cdots \wedge dt^{i_r}$, $r < k$, be a type (i) form. Then $d\pi_*\omega = 0$, and

$$\begin{aligned}
d\omega = (\pi^* d\phi) \wedge f(x,t)\, dt^{i_1} \wedge \cdots \wedge dt^{i_r} + \pi^* \Big(\sum_{i=1}^k (-1)^{\deg\phi} \phi \wedge dx^i \Big) \wedge \frac{\partial f}{\partial x^i}(x,t)\, dt^{i_1} \wedge \cdots \wedge dt^{i_r} \\
+ (\pi^*\phi) \wedge \Big(\sum_{j=1}^k (-1)^{\deg\phi} \frac{\partial f}{\partial t^j}(x,t)\, dt^j \wedge dt^{i_1} \wedge \cdots \wedge dt^{i_r} \Big)
\end{aligned}$$

and so

$$\begin{aligned}\pi_*(d\omega) &= \sum_{j=1}^{k}(-1)^{\deg\phi}\phi\int_{\mathbb{R}^k}\frac{\partial f}{\partial t^j}(x,t)\,dt^j\wedge dt^{i_1}\wedge\cdots\wedge dt^{i_r}\\ &= 0 \text{ if } dt^j\wedge dt^{i_1}\wedge\cdots dt^{i_r}\neq\pm dt^1\wedge\cdots\wedge dt^k.\end{aligned}$$

If $dt^j\wedge dt^{i_1}\wedge\cdots dt^{i_r}=\pm dt^1\wedge\cdots\wedge dt^k$, then $\int_{\mathbb{R}^k}\partial f/\partial t^j(x,t)\,dt^j\wedge dt^{i_1}\wedge\cdots\wedge dt^{i_r}$ is again 0: because f has compact support,

$$\int_{\mathbb{R}}\frac{\partial f}{\partial t^j}(x,t)dt^j = 0.$$

This completes the proof. □

Note that integration along the fiber lowers the degree of a form by the fiber dimension.

Lemma 3.1.53. *An orientable vector bundle E over an orientable manifold M is an orientable manifold.*

Proof. This follows from the fact that if $\{(U_\alpha,\varphi_\alpha)\}$ is an oriented atlas for M with transition functions $\rho_{\alpha\beta}=\varphi_\alpha\circ\varphi_\beta^{-1}$ and $\Phi_\alpha:\pi^{-1}(U_\alpha)\to U_\alpha\times\mathbb{R}^k$ is a local trivialization for E with transition functions $\{\tau_{\alpha\beta}\}$ then the composition

$$\pi^{-1}(U_\alpha)\xrightarrow{\Phi_\alpha}U_\alpha\times\mathbb{R}^k\xrightarrow{\varphi_\alpha\times\mathrm{id}_{\mathbb{R}^k}}\mathbb{R}^n\times\mathbb{R}^k$$

gives an atlas for E. The typical transition function of this atlas,

$$(\varphi_\alpha\times\mathrm{id}_{\mathbb{R}^k})\circ\Phi_\alpha\circ\Phi_\beta^{-1}\circ(\varphi_\beta\times\mathrm{id}_{\mathbb{R}^k})^{-1}:\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}^n\times\mathbb{R}^k$$

sends (x,y) to $(\rho_{\alpha\beta}(x),\tau_{\alpha\beta}(\varphi_\beta^{-1}(x))y)$ and has Jacobian matrix

$$\begin{pmatrix}\partial(\rho_{\alpha\beta}) & 0\\ * & \tau_{\alpha\beta}(\varphi_\beta^{-1}(x))\end{pmatrix}$$

The determinant of this matrix is clearly positive. □

Remark 3.1.1. The orientation on E described above is called the **local product orientation** on E.

Proposition 3.1.54 (Projection Formula). *Let $\pi:E\to M$ be an oriented rank k vector bundle, $\pi_*:\Omega^p_{cv}(E)\to\Omega^{p-k}(M)$ be the integration along the fiber defined above.*

(a) Let τ be a form on M and ω a form on E with compact support along the fiber. Then

$$\pi_*((\pi^*\tau)\wedge\omega)=\tau\wedge\pi_*\omega.$$

(b) Suppose in addition that M is oriented of dimension n, $\omega\in\Omega^p_{cv}(E)$, and $\tau\in\Omega_c^{n+k-p}(M)$. Then with the local product orientation on E,

$$\int_E(\pi^*\tau)\wedge\omega=\int_M\tau\wedge\pi_*\omega. \tag{1.4}$$

de Rham int c

Proof. By a partition of unity we may assume that E is the product bundle $M\times\mathbb{R}^n$. If ω is a form of type (i), say $\omega = (\pi^*\phi) \wedge f(x,t)\,dt^{i_1} \wedge \cdots \wedge dt^{i_r}$, where $r < k$, then

$$\begin{aligned}\pi_*((\pi^*\tau) \wedge \omega) &= \pi_*((\pi^*\tau) \wedge (\pi^*\phi) \wedge f(x,t)\,dt^{i_1} \wedge \cdots \wedge dt^{i_r}) \\ &= \pi_*(\pi^*(\tau \wedge \phi) \wedge f(x,t)\,dt^{i_1} \wedge \cdots \wedge dt^{i_r}) = 0 \\ &= \tau \wedge \pi_*\omega.\end{aligned}$$

If ω is a form of type (ii), say $\omega = (\pi^*\phi) \wedge f(x,t)\,dt^1 \wedge \cdots \wedge dt^n$, then

$$\begin{aligned}\pi_*((\pi^*\tau) \wedge \omega) &= \pi_*((\pi^*\tau) \wedge (\pi^*\phi) \wedge f(x,t)\,dt^1 \wedge \cdots \wedge dt^n) \\ &= \tau \wedge \phi \int_{\mathbb{R}^k} f(x,t)\,dt^1 \wedge \cdots \wedge dt^n \\ &= \tau \wedge \pi_*\omega.\end{aligned}$$

This proves (a).

For (b), if $\omega = (\pi^*\phi) \wedge f(x,t)\,dt^{i_1} \wedge \cdots \wedge dt^{i_r}$ with $r < k$, then by dimension consideration de Rham int on fiber projection formula-1 we have $(\pi^*\tau) \wedge (\pi^*\phi) = 0$, so both sides of (1.4) are zero. On the other hand, if $\omega = (\pi^*\phi) \wedge f(x,t)\,dt^1 \wedge \cdots \wedge dt^k$, then

$$\begin{aligned}\int_E (\pi^*\tau) \wedge \omega &= \int_{M \times \mathbb{R}^k} \pi^*(\tau \wedge \phi) \wedge f(x,t)\,dt^1 \wedge \cdots \wedge dt^k \\ &= \int_M \tau \wedge \phi \wedge \int_{\mathbb{R}^k} f(x,t)\,dt^1 \wedge \cdots \wedge dt^k \\ &= \int_M \tau \wedge \pi_*\omega.\end{aligned}$$

This gives (b). $\square$

Now we will show that integration along the fiber induces an isomorphism on cohomology groups. To use the induction principle, we first deal with the case $M = \mathbb{R}^n$. But in fact we have the following stronger result.

Proposition 3.1.55 (Poincaré Lemma for Compact Vertical Supports). *Integration along the fiber defines an isomorphism*

$$\pi_* : H^*_{cv}(M \times \mathbb{R}^k) \to H^{*-k}(M).$$

Thom iso **Theorem 3.1.56 (Thom Isomorhism).** *If the vector bundle $\pi : E \to M$ over a manifold M is orientable, then*

$$H^*_{cv}(E) \cong H^{*-k}(M)$$

where k is the rank of E.

Proof. Let U and V be open subsets of M. Using a partition of unity from the base M we see that

$$0 \longrightarrow \Omega^*_{cv}(E|_{U\cup V}) \longrightarrow \Omega^*_{cv}(E|_U) \oplus \Omega^*_{cv}(E|_V) \longrightarrow \Omega^*_{cv}(E|_{U\cap V}) \longrightarrow 0$$

is exact. So we have the diagram of Mayer-Vietoris sequences

$$\begin{tikzcd}
\cdots \arrow[r] & H^p_{cv}(E|_{U\cup V}) \arrow[r] \arrow[d, "\pi_*"] & H^p_{cv}(E|_U) \oplus H^p_{cv}(E|_V) \arrow[r] \arrow[d, "\pi_*"] & H^p_{cv}(E|_{U\cap V}) \arrow[r, "\delta^*"] \arrow[d, "\pi_*"] & H^{p+1}_{cv}(E|_{U\cup V}) \arrow[r] \arrow[d, "\pi_*"] & \cdots \\
\cdots \arrow[r] & H^{p-k}(U\cup V) \arrow[r] & H^{p-k}(U) \oplus H^{p-k}(V) \arrow[r] & H^{p-k}(U\cap V) \arrow[r, "\delta^*"] & H^{p+1-k}(U\cup V) \arrow[r] & \cdots
\end{tikzcd}$$

The commutativity of this diagram is trivial for the first two squares; we will check that of the third. Recalling from Corollary 3.1.21 the explicit formula for the coboundary operator δ^*, we have by the projection formula

Mayer-Vietoris de Rham connect

$$\pi_*\delta^*\omega = \pi_*(d\pi^*\rho_U) \wedge \omega = \pi_*(\pi^*(d\rho_U) \wedge \omega) = (d\rho_U) \wedge \pi_*\omega = \delta^*\pi_*\omega.$$

So the diagram in question is commutative. The theorem now follows from the induction principle. $\square$

Under the Thom isomorphism $\mathscr{T} : H^*(M) \to H^{*+k}_{cv}(E)$, the image of 1 in $H^0(M)$ determines a cohomology class Φ in $H^k_{cv}(E)$, called the **Thom class** of the oriented vector bundle E. Because $\pi_*\Phi = 1$, by the projection formula

$$\pi_*(\pi^*\omega \wedge \Phi) = \omega \wedge \pi_*\Phi = \omega.$$

So the Thom isomorphism, which is inverse to π_*, is given by $\mathscr{T}(\omega) = \pi^*\omega \wedge \Phi$.

ator on fiber

Proposition 3.1.57. *The Thom class Φ on a rank k oriented vector bundle E can be uniquely characterized as the cohomology class in $H^k_{cv}(E)$ which restricts to the generator of $H^k_c(F)$ on each fiber F.*

Proof. Since $\pi_*\Phi = 1$, $\Phi|_{\text{fiber}}$ is a bump form on the fiber with total integral 1. Conversely if Φ' in $H^k_{cv}(E)$ restricts to a generator on each fiber, then

$$\pi_*((\pi^*\omega) \wedge \Phi') = \omega \wedge \pi_*\Phi' = \omega.$$

Hence $\pi^*\omega \wedge \Phi' = \mathscr{T}(\omega)$ and $\Phi' = \mathscr{T}(1)$ is the Thom class. $\square$

s disrect sum

Proposition 3.1.58. *If E and F are two oriented vector bundles over a manifold M, and π_1 and π_2 are the projections*

$$\begin{tikzcd} & E \oplus F \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\ E & & F \end{tikzcd}$$

then the Thom class of $E \oplus F$ is $\pi_1^*\Phi(E) \wedge \pi_2^*\Phi(F)$.

Proof. Let k_1 and k_2 be the rank of E, F. Then $\pi_1^*\Phi(E) \wedge \pi_2^*\Phi(F)$ is a class in $H^{k_1+k_2}(E \oplus F)$ whose restriction to each fiber is a generator of the compact cohomology of the fiber, since the isomorphism

$$H_c^{k_1+k_2}(\mathbb{R}^{k_1} \times \mathbb{R}^{k_2}) \cong H_c^{k_1}(\mathbb{R}^{k_1}) \oplus H_c^{k_2}(H^{k_2})$$

is given by the wedge product of the generators. $\square$

Using the same technique and the Poincaré lemma for compact supports, we can also prove the following result.

Proposition 3.1.59. *If the vector bundle $\pi : E \to M$ over a manifold M is orientable, then*

$$H_c^*(E) \cong H_c^{*-k}(M)$$

where k is the rank of E.

Remark 3.1.2. The result above is not ture if $E \to M$ is not orientable. For example, the Möbious bundle has trivial compact cohomology, but the compact cohomology of S^1 is nontrivial.

Poincaré duality and the Thom class

Let S be a properly embedded oriented submanifold of dimension k in an oriented manifold M of dimension n. The **Poincaré dual of S** is the cohomology class of the closed $(n-k)$-form η_S characterized by the property

$$\int_S \omega = \int_M \omega \wedge \eta_S$$

for any closed k-form with compact support on M. Now we will explain how the Poincaré dual of a submanifold relates to the Thom class of a bundle. To this end we first recall the notion of a tubular neighborhood of S in M; this is by definition an open neighborhood of S in M diffeomorphic to a vector bundle of rank $n-k$ over S such that S is diffeomorphic to the zero section. The tubular neighborhood theorem (Theorem 1.3.25) states that every submanifold S in M has a tubular neighborhood T, and that in fact T is diffeomorphic to the normal bundle of S in M. Let $j : T \hookrightarrow M$ be the inclusion of a tubular neighborhood T of S in M. Since S and M are orientable, the normal bundle NS, being the quotient of $TM|_S$ by TS, is also orientable. By convention it is oriented in such a way that

Riemann tubular neighborhood

$$NS \oplus TS = TM|_S$$

has the direct sum orientation. So the Thom isomorphism theorem applies to the normal bundle $T = NS$ over S. Note that the integral $\int_S \omega$ only depends on the values of ω in a neighborhood of S, thus we can find duals supported in any neighborhood T of S. With this observation, we have the following result.

Poincare dual

Proposition 3.1.60. *Let S be a properly embedded oriented submanifold of an oriented manifold M. Then the Poincaré dual of S is the Thom class of the tube T. More precisely, if η_S is the Poincaré dual of S with support in T, then $j^*\eta_S$ is the Thom class of T.*

Proof. We merely have to show that $j^*\eta_S$ satisfies the defining property of the Thom class of T. Let $i : S \hookrightarrow T$ be the inclusion, so that $j \circ i$ is the inclusion of S into M. Since π is a deformation retraction of T onto S, π^* and i^* are inverse isomorphisms in cohomology. Now by the definition of the dual, for any $\omega \in \Omega_c^p(M)$ we have

$$\begin{aligned}\int_S (j\circ i)^*\omega &= \int_M \omega \wedge \eta_S = \int_T j^*\omega \wedge j^*\eta_S = \int_T \pi^* i^* j^*\omega \wedge j^*\eta_S \\ &= \int_S i^* j^*\omega \wedge \pi_*(j^*\eta_S) = \int_S (j\circ i)^*\omega \wedge \pi_*(j^*\eta_S).\end{aligned}$$

Since ω can be chosen to have support on any open subset of M, this then implies $\pi_*(j^*\eta_S) = 1$, so $j^*\eta_S$ represents the Thom class of T. $\square$

Corollary 3.1.61. *The Poincare dual of S is characterized as a closed form that integrates to 1 along fibers $\pi^{-1}(p)$ for all $p \in S$.*

Now suppose E is an oriented vector bundle over an oriented manifold M. Then M is diffeomorphically embedded as the zero section in E and there is an exact sequence

$$0 \longrightarrow TM \longrightarrow TE|_M \longrightarrow E \longrightarrow 0$$

Thom class Poincare dual

i.e., the normal bundle of M in E is E itself. By Proposition 3.1.60, we have the following.

Corollary 3.1.62. *The Thom class of an oriented vector bundle $\pi : E \to M$ over an oriented manifold M and the Poincaré dual of the zero section of E can be represented by the same form.*

Also, because the normal bundle of the submanifold S in M is diffeomorphic to any tubular neighborhood of S, we have the following proposition.

Proposition 3.1.63 (Localization Principle). *The support of the Poincaré dual of a submanifold S can be shrunk into any given tubular neighborhood of S.*

Example 3.1.64.

(a) **The Poincaré dual of a point p in M.** A tubular neighborhood T of p is simply an open ball around p. A generator of $H^n_{cv}(T)$ is a bump n-form with total integral 1. So the Poincaré dual of a point is a bump n-form on M. The form need not have support at p since all bump n-forms on a connected manifold are cohomologous. Here the dual of p is taken in $H^n_c(M)$, and not in $H^n(M)$.

(b) **The Poincaré dual of M.** Here the tubular neighborhood T is M itself, and $H^*_{cv}(T) = H^*(M)$. So the Poincaré dual of M is the constant function 1.

(c) **The Poincaré dual of a circle on a torus.** The Poincaré dual is a bump 1-form with support in a tubular neighborhood of the circle and with total integral 1 on each fiber of the tubular neighborhood. In the usual representation of the torus as a square, if the circle is a vertical segment, then its Poincaré dual is $\rho(x)dx$ where ρ is a bump function with total integral 1.

Using the explicit construction of the Poincare dual as the Thom class of the normal bundle, we now prove two basic properties of Poincaré duality. recall that two submanifolds R and S in M are said to intersect transversally if

$$T_pR + T_pS = T_pM$$

at all points p in the intersection $R \cap S$. For such a transversal intersection the codimension in M is additive: (Theorem transverse int subm **??**)

$$\operatorname{codim} R \cap S = \operatorname{codim} R + \operatorname{codim} S.$$

This implies that the normal bundle of $R \cap S$ in M is $N(R \cap S) = NR \oplus NS$. Assume M to be an oriented manifold, and R and S to be closed oriented submanifolds. Denoting the Thom class of an oriented vector bundle E by $\Phi(E)$, we have by Proposition Thom class disrect sum 3.1.58

$$\Phi(N(R \cap S)) = \Phi(NR \oplus NS) = \Phi(NR) \wedge \Phi(NR).$$

Therefore $\eta_{R\cap S} = \eta_R \wedge \eta_S$; i.e., under Poincaré duality the transversal intersection of closed oriented submanifolds corresponds to the wedge product of forms.

More generally, a smooth map $F : N \to M$ is said to be transversal to a submanifold $S \subseteq M$ if for every $p \in F^{-1}(S)$, $dF_p(T_pN) + T_{F(p)}S = T_{F(p)}M$. If $F : N \to M$ is an orientation-preserving map of oriented manifolds, T is a sufficiently small tubular neighborhood of the closed oriented submanifold S in M, and F is transversal to S and T, then $F^{-1}(T)$ is a tubular neighborhood of $F^{-1}(S)$ in N. From the commutative diagram

$$\begin{tikzcd}
H^*(S) \arrow[r] \arrow[d, "F^*"] & H^{*+n-k}_{cv}(T) \arrow[d, "F^*"] \\
H^*(F^{-1}(S)) \arrow[r] & H^{*+n-k}_{cv}(F^{-1}(T))
\end{tikzcd}$$

we see that $\Phi(F^{-1}(T)) = F^*\Phi(T)$. This then implies $\eta_{F^{-1}(S)} = F^*\eta_S$; i.e., under Poincaré duality the induced map on cohomology corresponds to the pre-image in geometry. By the Transversality Homotopy Theorem, the transversality hypothesis on F is in fact not necessary.

Relative de Rham theory

Let $f : N \to M$ be a smooth map between two manifolds. We know that the pullback $f^* : \Omega^*(M) \to \Omega^*(N)$ is a cochain map. Now consider the mapping cone of f^*:

$$\Omega^p(f) = \Omega^p(M) \oplus \Omega^{p-1}(N), \quad d(\omega, \eta) = (d\omega, -d\eta + f^*\omega).$$

Note that a cohomology class in $\Omega^*(f)$ is represented by a closed form on M which becomes exact when pulled back to N. By definition we have an exact sequence

$$0 \longrightarrow \Omega^{p-1}(N) \longrightarrow \Omega^p(f) \longrightarrow \Omega^p(M) \longrightarrow 0$$

where $\Omega^{p-1}(N)$ is the shifted complex $\Omega[-1]^*(N)$. Then we have a long exact sequence

$$\cdots \longrightarrow H^{p-1}(N) \longrightarrow H^p(f) \longrightarrow H^p(M) \xrightarrow{\delta} H^p(N) \longrightarrow \cdots$$

where the connection homomorphism is given by f^*. Moreover, if f and g are homotopic maps between N and M, then their mapping cone are isomorphic, so we have $H^*(f) = H^*(g)$. With these observations, we make the following definition.

Definition 3.1.65. *Let $f : N \to M$ be a smooth map between two manifolds. Then the **relative de Rham cohomology** $H^*(M, N)$ is defined to be the cohomology of the mapping cone of f^*. If S is a submanifold of M, then $H^*(M, S)$ is defiend to be $H^*(i)$, where $i : S \hookrightarrow M$ is the inclusion.*

Now that we have a fairly general relative cohomology theory we can establish the well-known excision property.

Proposition 3.1.66. *Let M be a smooth manifold and $\{U, V\}$ be an open cover of M. Then the restriction map*

$$H^p(M, U) \to H^p(V, U \cap V)$$

is an isomorphism.

Proof. First select a partition of unity ρ_U, ρ_V subordinate to $\{U, V\}$. We start with injectivity. Take a class $[(\omega, \psi)] \in H^p(M, U)$; i.e.,

$$d\omega = 0, \quad \omega|_U = d\psi.$$

If the restriction of (ω, ψ) to $(V, U \cap V)$ is exact, then we can find $(\overline{\omega}, \overline{\psi}) \in \Omega^{p-1}(U) \oplus \Omega^{p-2}(U \cap V)$ such that

$$\omega|_V = d\overline{\omega}, \quad \psi|_{U\cap V} = \overline{\omega}|_{U\cap V} - d\overline{\psi}.$$

This then implies

$$(\psi + d(\rho_V \overline{\psi}))|_{U\cap V} = (\overline{\omega} - d(\rho_U \overline{\psi}))|_{U\cap V}.$$

Now we defien a form $\widetilde{\omega}$ on M by gluing ψ and $\overline{\omega}$:

$$\widetilde{\omega} = \begin{cases} \psi + d(\rho_V \overline{\psi}) & \text{on } U, \\ \overline{\omega} - d(\rho_U \overline{\psi}) & \text{on } V. \end{cases}$$

Then we have $\omega = d\widetilde{\omega}$ and $\psi = \widetilde{\omega}|_U - d(\rho_V \overline{\psi})$, therefore (ω, ψ) is exact.

For surjectivity select $(\overline{\omega}, \overline{\psi}) \in \Omega^p(U) \oplus \Omega^{p-1}(U \cap V)$ that is closed:

$$d\overline{\omega} = 0, \quad \overline{\omega}|_{U\cap V} = d\overline{\psi}.$$

We can define a form ω on M by extending $\overline{\omega}$:

$$\omega = \begin{cases} d(\rho_V \overline{\psi}) & \text{on } U, \\ \overline{\omega} - d(\rho_U \overline{\psi}) & \text{on } V. \end{cases}$$

Clearly ω is closed and $\omega|_U = d(\rho_V \overline{\psi})$, so $(\omega, \rho_V \overline{\psi})$ is closed in $\Omega^p(M) \oplus \Omega^{p-1}(U)$. The restriction of this pair to $\Omega^p(V) \oplus \Omega^{p-1}(U \cap V)$ is $(\overline{\omega} - d(\rho_U \overline{\psi}), \rho_V \overline{\psi})$, which is not $(\overline{\omega}, \overline{\psi})$. But their difference is exact:

$$(\overline{\omega}, \overline{\psi}) - (\overline{\omega} - d(\rho_U \overline{\psi}), \rho_V \overline{\psi}) = (d(\rho_U \overline{\psi}), \rho_U \overline{\psi}) = d(\rho_U \overline{\psi}, 0).$$

Therefore $[(\omega, \rho_V \overline{\psi})]$ is mapped to $[(\overline{\omega}, \overline{\psi})]$. □

3.1.10 Exercise

Exercise 3.1.1. For each $n \geq 1$, compute the de Rham cohomology groups of $\mathbb{R}^n - \{e_1, -e_1\}$, and for each nonzero cohomology group, give specific differential forms whose cohomology classes form a basis.

Proof. When $n = 1$, $\mathbb{R} - \{e_1, -e_1\}$ is not connected and noncompact, hence

$$H^0_{dR}(\mathbb{R} - \{e_1, -e_1\}) = \mathbb{R}^3, \quad H^1_{dR}(\mathbb{R} - \{e_1, -e_1\}) = 0.$$

When $n \geq 2$, set $M = \mathbb{R}^n - \{e_1, -e_1\}$. We may assume $e_1 = (0, \ldots, 0, 1)$, and define

$$U = \mathbb{R}^{n-1} \times (-\varepsilon, +\infty), \quad V = \mathbb{R}^{n-1} \times (-\infty, \varepsilon).$$

so that $U \cap V = \mathbb{R}^{n-1} \times (-\varepsilon, \varepsilon) \simeq \mathbb{R}^{n-1}$, $U \simeq V \simeq \mathbb{R}^n - \{0\}$. Then by Mayer-Vietoris sequence we get

$$H^p_{dR}(M) = \begin{cases} \mathbb{R}^2 & p = n-1, \\ \mathbb{R} & p = 0, \\ 0 & \text{otherwise.} \end{cases}$$

□

Exercise 3.1.2. Let M be a connected smooth manifold of dimension $n \geq 3$. For any $x \in M$ and $0 \leq p \leq n-2$, prove that the map $H^p_{dR}(M) \to H^p_{dR}(M - \{x\})$ induced by inclusion $M - \{x\} \hookrightarrow M$ is an isomorphism. Prove that the same is true for $p = n-1$ if M is compact and orientable.

Proof. Let U be a regular coordinate ball around x, note that

$$U \cup (M - \{x\}) = M, \quad U \cap (M - \{x\}) = U - \{x\} \simeq S^{n-1}.$$

Thus by Mayer-Vietoris, for $2 \leq p \leq n-2$ we have an exact sequence

$$H^{p-1}_{dR}(U - \{x\}) \longrightarrow H^p_{dR}(M) \longrightarrow H^p_{dR}(U) \oplus H^p_{dR}(M - \{x\}) \longrightarrow H^p_{dR}(U - \{x\})$$

Thus $H^p_{dR}(M) = H^p_{dR}(M - \{x\})$ for $2 \leq p \leq n-2$.

The case $p = 0$ is immediate. When $p = 1$, since $H^0_{dR}(U) = H^0_{dR}(M) = H^0_{dR}(M - \{x\}) = \mathbb{R}$, the sequence becomes

$$\mathbb{R}^2 \longrightarrow \mathbb{R} \longrightarrow H^1_{dR}(M) \longrightarrow H^1_{dR}(M - \{x\}) \longrightarrow 0$$

hence $H^1_{dR}(M) \to H^1_{dR}(M - \{x\})$ is surjective. To prove injectivity, we need $H^0_{dR}(S^{n-1}) \to H^1_{dR}(M)$ to be trivial. For this, it's enough to show $H^0_{dR}(U) \oplus H^0_{dR}(M - \{x\}) \to H^0_{dR}(S^{n-1})$ is surjective, which is obvious since all these 0-degree cohomology groups are generated by constant functions.

For $p = n - 1$, if M is compact and orientable, then $H^n_{dR}(M) = \mathbb{R}$ and $H^n_{dR}(M - \{x\}) = 0$, by Theorem 3.1.36 and 3.1.37. The sequence becomes

$$0 \longrightarrow H^{n-1}_{dR}(M) \longrightarrow H^{n-1}_{dR}(M - \{x\}) \longrightarrow H^{n-1}_{dR}(S^{n-1}) \longrightarrow H^n_{dR}(M) \longrightarrow 0$$

Since $H^{n-1}_{dR}(S^{n-1}) \cong H^n_{dR}(M) = \mathbb{R}$, the right-side map is infact an isomorphism, and it follows that $H^{n-1}_{dR}(M) \cong H^{n-1}_{dR}(M - \{x\})$. $\square$

Exercise 3.1.3. Let M_1, M_2 be connected smooth manifolds of dimension $n \geq 3$, and let $M_1 \# M_2$ denote their smooth connected sum. Prove that

$$H^p_{dR}(M_1 \# M_2) \cong H^p_{dR}(M_1) \oplus H^p_{dR}(M_2).$$

Prove that the same is true for $p = n - 1$ if M_1 and M_2 are both compact and orientable.

Proof. There are open subsets $U, V \subseteq M_1 \# M_2$ that are diffeomorphic to $M_1 - \{p_1\}$ and $M_2 - \{p_2\}$, respectively, such that $U \cup V = M_1 \# M_2$ and $U \cap V$ is diffeomorphic to $(-1, 1) \times S^{n-1}$. By Mayer-Vietoris, for $0 \leq p \leq n - 2$ we can prove

$$H^p_{dR}(M_1 \# M_2) \cong H^p_{dR}(M_1) \oplus H^p_{dR}(M_2).$$

For $p = n - 1$, if M_1 and M_2 are both compact and orientable, then

$$\begin{cases} H^{n-1}_{dR}(U) \cong H^{n-1}_{dR}(M_1), H^{n-1}_{dR}(V) \cong H^{n-1}_{dR}(M_2), \\ H^n_{dR}(U) = H^n_{dR}(V) = 0, \\ H^n_{dR}(M_1 \# M_2) \cong \mathbb{R}. \end{cases}$$

The sequence becomes

$$0 \longrightarrow H^{n-1}(M_1 \# M_2) \longrightarrow H^{n-1}(U) \oplus H^{n-1}(V) \longrightarrow H^{n-1}(S^{n-1}) \longrightarrow H^n_{dR}(M_1 \# M_2) \longrightarrow 0$$

Since $H^{n-1}(S^{n-1}) \cong H^n_{dR}(M_1 \# M_2) \cong \mathbb{R}$, we get

$$H^{n-1}(M_1 \# M_2) \cong H^{n-1}(U) \oplus H^{n-1}(V) \cong H^{n-1}(M_1) \oplus H^{n-1}(M_2).$$

$\square$

3.2 Spectral sequences and applications

3.2.1 Čech-de Rham complex

Let M be a smooth manifold and $\mathcal{U} = (U_i)_{i\in I}$ be an open cover of M. Since M is second countable, we may assume the index set I to be countable and totally ordered. Now Ω^* is a sheaf on M, so we can formulate the Čech complex:

$$C^p = \prod_{i_0<\cdots<i_p} \Omega^*(U_{i_0\ldots i_p}).$$

In what follows,we will use d and δ tp denote the morphisms of $C^*(\mathcal{U},\Omega^*)$. So d is adjusted by a sign $(-1)^p$, and δ is unchanged. Moreover, we have $d\delta + \delta d = 0$.

The generalized Mayer-Vietoris sequence has the following form.

Proposition 3.2.1 (The Generalized Mayer-Vietoris Sequence). *The sequence*

$$0 \longrightarrow \Omega^*(M) \longrightarrow C^0(\mathcal{U},\Omega^*) \xrightarrow{\delta^0} C^1(\mathcal{U},\Omega^*) \xrightarrow{\delta^1} \cdots$$

is exact.

Proof. Let $\{\rho_i\}$ be a partition of unity subordinate to the open cover $\mathcal{U} = \{U_i\}$. Define a map $h : C^p(\mathcal{U},\Omega^*) \to C^{p-1}(\mathcal{U},\Omega^*)$ by

$$(h\omega)_{i_0\ldots i_{p-1}} = \sum_i \rho_i \omega_{i,i_0\ldots i_{p-1}}. \tag{2.1}$$

Then

$$(d\delta\omega)_{i_0\ldots i_p} = \sum_{j=0}^{p}(-1)^j (h\omega)_{i_0\ldots\widehat{i_j}\ldots i_p} = \sum_{j=0}^{p}(-1)^j \sum_i \rho_i \omega_{i,i_0\ldots\widehat{i_j}\ldots i_p},$$

while

$$\begin{aligned}(h\delta\omega)_{i_0\ldots i_p} = \sum_i \rho_i (\delta\omega)_{i,i_0\ldots i_p} &= \sum_i \rho_i \omega_{i_0\ldots i_p} + \sum_i \rho_i \sum_{j=1}^{p+1}(-1)^j \omega_{i,i_0\ldots\widehat{i_j}\ldots i_p} \\ &= \omega_{i_0\ldots i_p} - (dh\omega)_{i_0\ldots i_p}.\end{aligned}$$

This means h is a homotopy between the identity and the zero map. Therefore the sequence is exact. □

The double complex $C^*(\mathcal{U},\Omega^*)$ is called the **Čech-de Rham complex**, and an element of the Čech-de Rham complex is called a **Čech-de Rham cochain**.

The fact that all the rows of the augmented complex are exact is the key ingredient in the proof of the following.

Theorem 3.2.2 (Generalized Mayer-Vietoris Principle). *The double complex $C^*(\mathcal{U},\Omega^*)$ computes the de Rham cohomology of M. more precisely, the restriction map $r : \Omega^*(M) \to C^*(\mathcal{U},\Omega^*)$ induces an isomorphism in cohomology:*

$$r^* : H^*_{dR}(M) \to H^*_{TC}(C^*(\mathcal{U},\Omega^*)).$$

Proof. This follows from Proposition 3.2.1 and a spectral sequence argument. □

We can improve a bit this result. For $p \geq 0$ define

$$K : C^p(\mathcal{U}, \Omega^q) \to C^{p-1}(\mathcal{U}, \Omega^q), \quad K^p = h$$

de Rham MV generalized

where h is the homotopy operator constructed in the proof of Proposition 3.2.1. Then

$$K\delta + \delta K = \mathrm{id}.$$

For an element $\omega \in T^p(C^*(\mathcal{U}, \Omega^*))$, we can write

$$\omega = \sum_{i=0}^{p} \omega_i, \quad \omega_i \in C^i(\mathcal{U}, \Omega^{p-i})$$

and

$$D\omega = \sum_{j=0}^{p+1} \eta_j, \quad \eta_j = d_v\omega_j + \delta\omega_{j-1} \in C^j(\mathcal{U}, \Omega^{p+1-j}).$$

where we set $\omega_{p+1} = \omega_{-1} = 0$. We now define a map $f : C^*(\mathcal{U}, \Omega^*) \to C^0(\mathcal{U}, \Omega^*)$ by the formula

$$f(\omega) = \sum_{i=0}^{p} (-dK)^i \omega_i - \sum_{j=1}^{p+1} K(-dK)^{j-1}\eta_j. \tag{2.2}$$

collating for

ating formula

Proposition 3.2.3 (Collating Formula). *The morphism $f : C^*(\mathcal{U}, \Omega^*) \to C^0(\mathcal{U}, \Omega^*)$ commutes with $D = d + \delta$ so it is a morphism of complexes. Moreover, it is chain homotopic to the identity, where the homotopy operator*

$$L : T^p(C^*(\mathcal{U}, \Omega^*)) \to T^{p-1}(C^*(\mathcal{U}, \Omega^*))$$

is given by

$$L\omega_i = \sum_{j=0}^{i-1} K(-dK)^{i-1-j}\omega_i \ \text{ for } \ \omega_i \in C^i(\mathcal{U}, \Omega^{p-i}).$$

To prove this claim, we first need a lemma.

formula lemma

Lemma 3.2.4. *For $i \geq 1$ we have*

$$[\delta, (-dK)^i] := \delta(-dK)^i - (-dK)^i\delta = (dK)^{i-1}d.$$

Proof. In any associative algebra A the commutator $a \mapsto [x, a] := xa - ax$ behaves like a derivation

$$[x, ab] = [x, a]b - a[x, b].$$

We deduce that

$$[\delta, -dK] = -\delta dK + dK\delta = d\delta K + dK\delta = d(\delta K + K\delta) = d.$$

Hence

$$\begin{aligned}[\delta, (-dK)^i] &= [\delta, -dK](-dK)^{i-1} + (-dK)^{i-1}[\delta, -dK] \\ &= d(-dK)^{i-1} + (-dK)^{i-1}d = (-dK)^{i-1}d.\end{aligned}$$

This gives the claim. $\square$

With this, we can now simplify the expression of f. We have

$$\begin{aligned}
f(\omega) &= \sum_{i=0}^{p}(-dK)^i\omega_i - \sum_{j=1}^{p}K(-dK)^{j-1}d\omega_j - \sum_{j=1}^{p+1}K(-dK)^{j-1}\delta\omega_{j-1}\\
&= \sum_{i=0}^{p}(-dK)^i\omega_i - \sum_{j=1}^{p}K[\delta,(-dK)^j]\omega_j - \sum_{j=1}^{p+1}K(-dK)^{j-1}\delta\omega_{j-1}\\
&= \sum_{i=0}^{p}(-dK)^i\omega_i - \sum_{j=1}^{p}K\delta(-dK)^j\omega_j + \sum_{j=1}^{p+1}K(-dK)^j\delta\omega_j - \sum_{j=0}^{p}K(-dK)^j\delta\omega_j\\
&= \omega_0 + \sum_{i=1}^{p}(\mathrm{id}-K\delta)(-dK)^i\omega_i - K\delta\omega_0\\
&= \delta K\omega_0 + \sum_{i=1}^{p}\delta K(-dK)^i\omega_i = \delta K\Big(\sum_{i=0}^{p}(-dK)^i\omega_i\Big).
\end{aligned}$$

Observe that $f(\omega)\in C^0(\mathcal{U},\Omega^q)$. In fact $f(\omega)$ lies in the image of $r:\Omega^q(M)\to C^0(\mathcal{U},\Omega^q)$: we have $\delta f(\omega)=0$, which means that the collection $\{f(\omega)_\alpha\}$ satisfies $f(\omega)_\alpha = f(\omega)_\beta$ on $U_{\alpha\beta}$.

Now we can give the proof of Theorem Cech de Rham collating formula 3.2.3.

Proof of Theorem Cech de Rham collating formula 3.2.3. Let us first show that $fD = df$. Let

$$\omega = \sum_{i=0}^{p}\omega_i,\quad \omega_i\in C^i(\mathcal{U},\Omega^{p-i}),\quad D\omega = \sum_{j=0}^{p+1}\eta_j,\quad \eta_j\in C^j(\mathcal{U},\Omega^{p+1-j}).$$

From the definition (collating formula def 2.2), we deduce

$$\begin{aligned}
df(\omega) &= d\Big(\sum_{i=0}^{p}(-dK)^i\omega_i - \sum_{j=1}^{p+1}K(-dK)^{j-1}\eta_j\Big) = d\omega_0 + \sum_{j=1}^{p+1}(-dK)^j\eta_j\\
&= \eta_0 + \sum_{j=1}^{p+1}(-dK)^j\eta_j = \sum_{j=0}^{p+1}(-dK)^j\eta_j = f(D\omega).
\end{aligned}$$

Let $\omega_i\in C^i(\mathcal{U},\Omega^{p-i})$, then

$$f(\omega_i) = \delta K(-dK)^i\omega_i.$$

Next, we observe that

$$\begin{aligned}
DL\omega_i &= \sum_{j=0}^{i-1}dK(-dK)^{i-1-j}\omega_i + \sum_{j=0}^{i-1}\delta K(-dK)^{i-1-j}\omega_i\\
&= -\sum_{j=0}^{i-1}(-dK)^{i-j}\omega_i + \sum_{j=0}^{i-1}\delta K(-dK)^{i-1-j}\omega_i
\end{aligned}$$

and

$$LD\omega_i = Ld\omega_i + L\delta\omega_i = \sum_{j=0}^{i-1}K(-dK)^{i-1-j}d\omega_i + \sum_{k=0}^{i}K(-dK)^{i-k}\delta\omega_i.$$

Using Lemma collating formula lemma 3.2.4 we deduce

$$(-dK)^{i-k}\delta = \delta(-dK)^{i-k} - (-dK)^{i-k-1}d.$$

On the other hand the homotopy property of K implies

$$K\delta(-dK)^{i-k} = (-dK)^{i-k} - \delta K(-dK)^{i-k},$$

so that

$$\begin{aligned} K(-dK)^{i-k}\delta &= K(\delta(-dK)^{i-k} - (-dK)^{i-k-1}d) = K\delta(-dK)^{i-k} - K(-dK)^{i-k-1}d \\ &= (-dK)^{i-k} - \delta K(-dK)^{i-k} - K(-dK)^{i-k-1}d. \end{aligned}$$

Combine this two equalities, we get

$$\begin{aligned} LD\omega_i &= \sum_{j=0}^{i-1} K(-dK)^{i-1-j}d\omega_i + \sum_{k=0}^{i}(-dK)^{i-k}\omega_i - \sum_{k=0}^{i}\delta K(-dK)^{i-k}\omega_i - \sum_{k=0}^{i} K(-dK)^{i-k-1}d\omega_i \\ &= \sum_{k=0}^{i}(-dK)^{i-k}\omega_i - \sum_{k=0}^{i}\delta K(-dK)^{i-k}\omega_i. \end{aligned}$$

From this, we then deduce that

$$\begin{aligned} DL\omega_i + LD\omega_i &= -\sum_{j=0}^{i-1}(-dK)^{i-j}\omega_i + \sum_{j=0}^{i-1}\delta K(-dK)^{i-1-j}\omega_i + \sum_{k=0}^{i}(-dK)^{i-k}\omega_i - \sum_{k=0}^{i}\delta K(-dK)^{i-k}\omega_i \\ &= \omega_i - \delta K(-dK)^{i}\omega_i = \omega_i - f(\omega_i). \end{aligned}$$

Therefore the claim follows. □

am to de Rham

Corollary 3.2.5. *Suppose that $\omega = \sum_{i=0}^{p}\omega_i$ is a Čech-de Rham cocycle. Then its is cohomologous with the de Rham cocycle*

$$f(\omega) = \sum_{i=0}^{p}(-dK)^{i}\omega_i.$$

It is also natural to augment each column by the kernel of the bottom d, denoted $C^*(\mathcal{U}, \mathbb{R})$. The vector space $C^p(\mathcal{U}, \mathbb{R})$ consists of the locally constant functions on the $(p+1)$-fold intersections $U_{i_0\ldots i_p}$. The cohomology of the bottom row

$$0 \longrightarrow C^0(\mathcal{U}, \mathbb{R}) \longrightarrow C^0(\mathcal{U}, \mathbb{R}) \longrightarrow \cdots$$

is the Čech cohomology of the constant sheaf $\underline{\mathbb{R}}_M$ with respect to the cover $\mathcal{U}$, denoted by $H^*(\mathcal{U}, \mathbb{R})$. This will give us the following diagram

$$\begin{array}{ccccccccccc}
 & & \vdots & & \vdots & & \vdots & & \vdots & & \\
 & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \to & \Omega^2(M) & \to & C^0(\mathcal{U}, \Omega^2) & \to & C^1(\mathcal{U}, \Omega^2) & \to & C^2(\mathcal{U}, \Omega^2) & \to & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \to & \Omega^1(M) & \to & C^0(\mathcal{U}, \Omega^2) & \to & C^1(\mathcal{U}, \Omega^2) & \to & C^2(\mathcal{U}, \Omega^2) & \to & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \to & \Omega^0(M) & \to & C^0(\mathcal{U}, \Omega^2) & \to & C^1(\mathcal{U}, \Omega^2) & \to & C^2(\mathcal{U}, \Omega^2) & \to & \cdots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & C^0(\mathcal{U}, \mathbb{R}) & \longrightarrow & C^1(\mathcal{U}, \mathbb{R}) & \longrightarrow & C^2(\mathcal{U}, \mathbb{R}) & \longrightarrow & \cdots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & 0 & & 0 & & 0 & &
\end{array}$$

If the augmented columns are exact, then by the same method we can prove that $H^*(\mathcal{U}, \mathbb{R}) = H^*_{TC}(C^*(\mathcal{U}, \Omega^*))$. Thus we make the following definition.

Definition 3.2.6. *Let M be a manifold of dimension n. An open cover $\mathcal{U} = \{U_i\}$ of M is called a* ***good cover*** *if all nonempty finite intersections $U_{i_0 \dots i_p}$ are diffeomorphic to $\mathbb{R}^n$.*

Theorem 3.2.7. *Every manifold has a good cover. If the manifold is compact, then the cover may be chosen to be finite.*

Now for a good cover $\mathcal{U}$, the augmented columns are also exact, so we get the following important theorem.

ch good cover

Theorem 3.2.8. *If $\mathcal{U}$ is a good cover of the manifold M, then the de Rham cohomology of M is isomorphic to the Čech cohomology of the good cover*

$$H^*_{dR}(M) \cong H^*(\mathcal{U}, \mathbb{R}).$$

A priori there is no reason why different covers of M should have the same Čech cohomology. However, i.t follows from Theorem 3.2.8 that

de Rham Cech good cover

Corollary 3.2.9. *The Čech cohomology $H^*(\mathcal{U}, \mathbb{R})$ is the same for all good covers $\mathcal{U}$ of M.*

If a manifold has a finite good cover, then the Čech cohomology $H^*(\mathcal{U}, \mathbb{R})$ is clearly finite-dimensional. Thus,

Corollary 3.2.10. *Whenever M has a finite good cover, its de Rham cohomology $H^*_{dR}(M)$ is finite-dimensional.*

We now apply the main theorems to give a proof of the Künneth formula. Before commencing the proof we make some general remarks about a technique for studying maps. Let $\pi : E \to M$ be a map of manifolds. A cover $\mathcal{U}$ on M induces a cover $\pi^{-1}(\mathcal{U})$ on E, and we have the inclusions

$$\begin{tikzcd}
E \arrow[d, "\pi"] & \coprod \pi^{-1}(U_{i_0}) \arrow[l] & \coprod \pi^{-1}(U_{i_0 i_1}) \arrow[l] & \cdots \arrow[l] \\
M & \coprod U_{i_0} \arrow[l] & \coprod U_{i_0 i_1} \arrow[l] & \cdots \arrow[l]
\end{tikzcd}$$

In general $U_i \cap U_j \neq \varnothing$ is not equivalent to $\pi^{-1}(U_i) \cap \pi^{-1}(U_j) \neq \varnothing$. However, if π is surjective, then the two statements are equivalent, so that in this case the combinatorics of the covers $\mathcal{U}$ and $\pi^{-1}(\mathcal{U})$ are the same. The double complex of the inverse cover computes the cohomology of E, which can then be related to the cohomology of M, because the inverse cover comes from a cover on M.

A quick example of how the inverse cover $\pi^{-1}(\mathcal{U})$ may be used to study maps is the following. Note that although the inverse image of a good cover is usually not a good cover, for a vector bundle $\pi : E \to M$ the "goodness" of the cover is preserved. Since the de Rham cohomology is determined by the combinatorics of a good cover, this implies that

$$H^*_{dR}(E) = H^*_{dR}(M).$$

Of course, this also follows from the homotopy axiom for the de Rham cohomology.

Proposition 3.2.11 (Künneth formula). *If M and F are two manifolds and F has finite-dimensional cohomology, then the de Rham cohomology of the product $M \times F$ is*

$$H^*(M \times F) = H^*(M) \otimes H^*(F).$$

Proof. Let $\mathcal{U} = \{U_i\}$ be a good cover for M and $\pi : M \times F \to M$ the projection onto the first factor. Then $\pi^{-1}(\mathcal{U})$ is some sort of a cover for $E = M \times F$, though in general not a good cover. There is a natural map

$$C^*(\mathcal{U}, \Omega^*) \longrightarrow C^*(\pi^{-1}(\mathcal{U}), \Omega^*)$$

which pulls back differential forms on open sets. Choose a basis for $H^*(F)$, say $\{[\omega_\alpha]\}$, and choose differential forms ω_α representing them. These may be used to define a map of double complexes

$$\pi^*_{\mathcal{U}} : H^*(F) \otimes C^*(\mathcal{U}, \Omega^*) \to C^*(\pi^{-1}(\mathcal{U}), \Omega^*), \quad \pi^*_{\mathcal{U}}([\omega_\alpha] \otimes \phi) = \rho^*\omega_\alpha \wedge \pi^*\phi$$

where $\rho : E \to F$ is the the projection on the fiber. Since $H^*(F)$ is a vector space, $H^*(F) \otimes C^*(\mathcal{U}, \Omega^*)$ is a number of copies of $C^*(\mathcal{U}, \Omega^*)$ and the differential operator D on the double complex $C^*(\mathcal{U}, \Omega^*)$ induces an operator on $H^*(F) \otimes C^*(\mathcal{U}, \Omega^*)$ whose cohomology is

$$H^*(F) \otimes H^*_{TC}(C^*(\mathcal{U}, \Omega^*)) = H^*(F) \otimes H^*(M).$$

Since the D-cohomology of $C^*(\pi^{-1}(\mathcal{U}), \Omega^*)$ is $H^*(E)$, if we can show that $\pi^*_{\mathcal{U}}$ induces an isomorphism in D-cohomology, the Künneth formula will follow. Now for a good cover $\mathcal{U}$, the p-th column $C^p(\pi^{-1}(\mathcal{U}), \Omega^*)$ consists of forms on the $(p+1)$-fold intersections $\pi^{-1}(U_{i_0 \dots i_p})$ and $C^p(\mathcal{U}, \Omega^*)$ consists of forms on $U_{i_0 \dots i_p}$. Since each intersection $U_{i_0 \dots i_p}$ is diffeomorphic to $\mathbb{R}^n$, the vertical cohomology of $C^*(\pi^{-1}(\mathcal{U}), \Omega^*)$ is

$$\prod H^*(\pi^{-1}(U_{i_0 \dots i_p})) \cong H^*(F) \otimes \prod H^*(U_{i_0 \dots i_p})$$

the isomorphism being given by the wedge product of pullbacks. So $\pi^*_{\mathcal{U}}$ induces an isomorphism of the vertical cohomology of $C^*(\pi^{-1}(\mathcal{U}), \Omega^*)$ and $H^*(F) \otimes C^*(\mathcal{U}, \Omega^*)$. It follows by a spectral sequence argument that $\pi^*_{\mathcal{U}}$ also induces an isomorphism in D-cohomology. $\square$

The following example shows that some sort of finiteness hypothesis is necessary for the Künneth formula to hold.

Example 3.2.12 (Counterexample to the Künneth formula). Let M and F each be the set $\mathbb{Z}^+$ of all positive integers. Then

$$H^0(M \times F) = \{\text{square matrices of real numbers } (a_{ij}),\, i, j \in \mathbb{Z}^+\}.$$

But $H^0(M) \otimes H^0(F)$ consists of finite sums of matrices (a_{ij}) of rank 1. These two vector spaces are not equal, since a finite sum of matrices of rank 1 has finite rank, but $H^)(M \times F)$ contains matrices of infinite rank.

Now we provide a generalization of the Künneth formula. It is proved in the same manner.

Theorem 3.2.13 (Leray-Hirsch). *Let $\pi : E \to M$ be a smooth fiber bundle on a manifold M with fiber F. If there are global cohomology classes $e_1, \dots, e_r$ on E which when restricted to each fiber freely generate the cohomology of the fiber, then $H^*(E)$ is a free module over $H^*(M)$ with basis $\{e_1, \dots, e_r\}$, i.e.,*

$$H^*(F) \otimes H^*(M) \cong H^*(E)$$

given by pulling back classes from the base space and there forming their wedge product with these generators on the total space:

$$\sum \iota^*(e_i) \otimes c_j \mapsto \sum e_i \wedge \pi^*(c_j)$$

where $\iota : F \to E$ is the inclusion.

Proof. By replacing the ω_α's by e_i's, the proof goes as the previous one, with the observation that any fiber bundle on $\mathbb{R}^n$ is trivial, hence the theorem holds on it. $\square$

Product structures

In this part we define product structures on the Čech-de Rham complex $C^*(\mathcal{U}, \Omega^*)$, the de Rham cohomology, and the Čech cohomology, and show that the isomorphism between de Rham and Čech is an isomorphism of graded algebras. We also discuss the product structures on a spectral sequence.

Let Z be the closed forms and B the exact forms on a manifold M. From the antiderivation property of the exterior derivative

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^{\deg \omega} \omega \wedge d\eta$$

it follows that Z is a subring of $\Omega^*(M)$ and B is an ideal in Z. Hence the wedge product makes the de Rham cohomology $H_{dR}(M) = Z/B$ into a graded algebra. On the double complex $C^*(\mathcal{U}, \Omega^*)$, where $\mathcal{U}$ is any open cover of M, a natural product

$$\cup : C^p(\mathcal{U}, \Omega^q) \otimes C^r(\mathcal{U}, \Omega^s) \to C^{p+r}(\mathcal{U}, \Omega^{q+s})$$

can be defined as follows. If ω is in $C^p(\mathcal{U}, \Omega^q)$ and η is in $C^r(\mathcal{U}, \Omega^s)$, then

$$(\omega \cup \eta)_{i_0 \ldots i_{p+r}} = (-1)^{qr} \omega_{i_0 \ldots i_p} \wedge \eta_{i_p \ldots i_{p+r}}.$$

where on the right-hand side both forms are understood to be restricted to $U_{i_0 \ldots i_{p+r}}$ with the usual convention that $i_0 < \cdots < i_{p+r}$.

cup prod prop **Proposition 3.2.14.** *Let $\omega \in C^p(\mathcal{U}, \Omega^q)$ and $\eta \in C^r(\mathcal{U}, \Omega^s)$. Then*

(a) $\delta(\omega \cup \eta) = (\delta\omega) \cup \eta + (-1)^{p+q} \omega \cup (\delta\eta)$.

(b) $d(\omega \cup \eta) = (d\omega) \cup \eta + (-1)^{p+q} \omega \cup (d\eta)$.

(c) $D(\omega \cup \eta) = (D\omega) \cup \eta + (-1)^{p+q} \omega \cup (D\eta)$.

Proof. Note that (c) follows from (a) and (b). For (a), recall that

$$\begin{aligned}
\delta(\omega \cup \eta)_{i_0 \ldots i_{p+r+1}} &= \sum_{j=0}^{p+r+1} (-1)^j (\omega \cup \eta)_{i_0 \ldots \widehat{i_j} \ldots i_{p+r+1}} \\
&= (-1)^{qr} \sum_{j=0}^{p} (-1)^j \omega_{i_0 \ldots \widehat{i_j} \ldots i_{p+1}} \wedge \eta_{i_{p+1} \ldots i_{p+r+1}} + (-1)^{qr} \sum_{j=p+1}^{p+r+1} (-1)^j \omega_{i_0 \ldots i_p} \wedge \eta_{i_p \ldots \widehat{i_j} \ldots i_{p+r+1}} \\
&= (-1)^{qr} \sum_{j=0}^{p+1} (-1)^j \omega_{i_0 \ldots \widehat{i_j} \ldots i_{p+1}} \wedge \eta_{i_{p+1} \ldots i_{p+r+1}} + (-1)^{qr} \sum_{j=p}^{p+r+1} (-1)^j \omega_{i_0 \ldots i_p} \wedge \eta_{i_p \ldots \widehat{i_j} \ldots i_{p+r+1}} \\
&= (-1)^{qr} \sum_{j=0}^{p+1} (-1)^j \omega_{i_0 \ldots \widehat{i_j} \ldots i_{p+1}} \wedge \eta_{i_{p+1} \ldots i_{p+r+1}} + (-1)^{p+q} (-1)^{q(r+1)} \sum_{j=p}^{p+r+1} (-1)^{j-p} \omega_{i_0 \ldots i_p} \wedge \eta_{i_p \ldots \widehat{i_j} \ldots i_{p+r+1}} \\
&= ((\delta\omega) \cup \eta)_{i_0 \ldots i_{p+r+1}} + (-1)^{p+q} \omega \cup (\delta\eta).
\end{aligned}$$

This gives (a). As for (b), we have

$$\begin{aligned} d(\omega \cup \eta)_{i_0\dots i_{p+r+1}} &= (-1)^{qr} d(\omega_{i_0\dots i_p} \wedge \eta_{i_p\dots i_{p+r}}) \\ &= (-1)^{qr} (d\omega)_{i_0\dots i_p} \wedge \eta_{i_p\dots i_{p+r}} + (-1)^{p+qr} \omega_{i_0\dots i_p} \wedge d(\eta_{i_p\dots i_{p+r}}) \\ &= (d\omega) \cup \eta + (-1)^{p+q} \omega \cup (d\eta). \end{aligned}$$

This gives the claim. □

The inclusion of the Cech complex $C^*(\mathcal{U}, \mathbb{R})$ in the Čech-de Rham complex induces a product structure on $C^*(\mathcal{U}, \mathbb{R})$: if ω is a p-cochain and η an r-cochain, then

$$(\omega \cup \eta)_{i_0\dots i_{p+r}} = \omega_{i_0\dots i_p} \eta_{i_p\dots i_{p+r}}. \tag{2.3}$$

By Proposition 3.2.14, δ is an antiderivation relative to this product. So just as in the case of de Rham cohomology this makes the Čech cohomology $H^*(\mathcal{U}, \mathbb{R})$ into a graded algebra. If $\mathcal{V}$ is a refinement of $\mathcal{U}$, then the restriction map $H^*(\mathcal{U}, \mathbb{R}) \to H^*(\mathcal{V}, \mathbb{R})$ is a homomorphism of algebras. Hence the direct limit $H^*(M, \mathbb{R})$ is also a graded algebra. Note that (2.3) also makes sense for the Čech complex $C^*(\mathcal{U}, \mathbb{R})$ on a topological space X; this gives a product structure on the Čech cohomology $H^*(X, \mathbb{R})$ of any topological space X.

With the product structures just defined, both inclusions

$$r^* : \Omega^*(M) \to C^*(\mathcal{U}, \Omega^*), \quad i^* : C^*(\mathcal{U}, \mathbb{R}) \to C^*(\mathcal{U}, \Omega^*)$$

are algebra homomorphisms. Since as we saw in Proposition 3.2.8, for a good cover these homomorphisms induce bijective maps in cohomology

$$H^*_{dR}(M) \cong H^*_{TC}(C^*(\mathcal{U}, \Omega^*)), \quad H^*(\mathcal{U}, \mathbb{R}) \cong H^*_{TC}(C^*(\mathcal{U}, \Omega^*)),$$

the isomorphism between $H^*_{dR}(M)$ and $H^*(\mathcal{U}, \mathbb{R})$ is an algebra isomorphism. Because $H^*(M, \mathbb{R}) = H^*(\mathcal{U}, \mathbb{R})$ for a good cover $\mathcal{U}$, we have the following

Theorem 3.2.15. *The isomorphism between de Rham and Čech*

$$H^*_{dR}(M) \cong H^*(M, \mathbb{R})$$

is an isomorphism of graded algebras.

If a double complex K has a product structure relative to which its differential D is an antiderivation, the same is true of all the groups E_r, and their operators d_r since E_r is the homology of E_{r-1} and d_r is induced from D.

Proposition 3.2.16. *Let K be a double complex with a product structure relative to which D is an antiderivation. There exists a spectral sequence $\{(E_r, d_r)\}$ converging to $H_{TC}(K)$ with the following properties: Each E_r being the homology of its predecessor E_{r-1}, inherits a product structure from E_{r-1}. Relative to this product, d_r is an antiderivation.*

3.2.2 Sphere bundle

Let $\pi : E \to M$ be a fiber bundle with fiber the sphere S^k. As the structure group we normally take the largest group possible, namely the diffeomorphism group $\mathrm{Diff}(S^k)$, but sometimes we also consider sphere bundles with structure group $\mathrm{O}(k+1)$. These two notions are not equivalent: there are examples of sphere bundles whose structure groups cannot be reduced to the orthogonal group. Thus, every vector bundle defines a sphere bundle, but not conversely. By the Leray-Hirsch theorem if there is a closed global k-form on E whose restriction to each fiber generates the cohomology of the fiber, then the cohomology of E is

$$H^*(E) = H^*(M) \otimes H^*(S^k).$$

It is therefore of interest to know when such a global form exists. We will see that the existence of a global form as above entails overcoming two obstructions: orientability and the Euler class.

Orientability

In this part the base space of the bundle is assumed to be connected. A sphere bundle with fiber S^k ($k \geq 1$) is said to be orientable if for each fiber E_p it is possible to choose a generator $[\sigma_p]$ of $H^k(E_p)$ satisfying the **local compatibility condition**: around any point there is a neighborhood U and a generator $[\sigma_U]$ of $H^k(E|_U)$ such that for any p in U, $[\sigma_U]$ restricted to the fiber E_p is the chosen generator $[\sigma_p]$; equivalently, there is an trivialization $\{(U_\alpha, \Phi_\alpha)\}$ of M and generators $[\sigma_\alpha]$ of $H^k(E|_{U_\alpha})$ so that $[\sigma_\alpha] = [\sigma_\beta]$ in $H^k(E|_{U_\alpha \cap U_\beta})$.

Since a generator of the top cohomology of a fiber is an k-form with total integral 1, there are two possible generators, depending on the orientation of the fiber. A priori all that one could say is that $[\sigma_\alpha] = \pm[\sigma_\beta]$ on $U_\alpha \cap U_\beta$. For an orientable sphere bundle either choice of a consistent system of generators is called an **orientation of the sphere bundle**. A bundle with a given orientation is said to be **oriented**. An S^0-bundle over a manifold M is a double cover of M; such a bundle over a connected base space is said to be **orientable** if and only if the total space has two connected components.

Recall that we called a vector bundle of rank $k+1$ orientable if and only if it can be given by transition functions with values in $\mathrm{SO}(k+1)$. We now study the relation between the orientability of a sphere bundle and the orientability of a vector bundle.

Let E be a vector bundle of rank $k+1$ endowed with a Riemannian metric so that its structure group is reduced to $\mathrm{O}(k+1)$. Its **unit sphere bundle** $S(E)$ is the fiber bundle whose fiber at p consists of all the unit vectors in E_p and whose transition functions are the same as those of E. $S(E)$ is an S^k-bundle with structure group $O(k+1)$.

Remark 3.2.1. Fix an orientation on the sphere S^k. If a linear transformation T is in the special orthogonal group $\mathrm{SO}(k+1)$ and $[\sigma]$ is a generator of $H^k(S^k)$ with $\int_{S^k} \sigma = 1$, then the image $T(S^k)$ is the sphere S^k with the same orientation, so that

$$\int_{S^k} T^*\sigma = \int_{T(S^k)} \sigma = \int_{S^k} \sigma = 1.$$

Thus for an orthogonal transformation T, $T^*\sigma$ and σ represent the same cohomology class if and only if T has positive determinant.

Proposition 3.2.17. *A vector bundle E is orientable if and only if its sphere bundle $S(E)$ is orientable.*

Proof. Let E be an orientable vector bundle of rank $k+1$. Fix a generator σ on S^k and fix a trivialization $\{(U_\alpha, \Phi_\alpha)\}$ for E so that the transition functions $\tau_{\alpha\beta}$ assume values in $\mathrm{SO}(k+1)$. Let $\rho_\alpha : U_\alpha \times S^k \to S^k$ be the projection and let $\pi^{-1}(p)$ be the fiber of the sphere bundle $\pi : S(E) \to M$ at p. Define $[\sigma_\alpha]$ in $H^k(S(E)|_{U_\alpha})$ by

$$[\sigma_\alpha] = \Phi_\alpha^* \circ \rho_\alpha^*[\sigma] = (\rho_\alpha \circ \Phi_\alpha)^*[\sigma].$$

Write $\Phi_{\alpha,p}$ for $\Phi_\alpha|_{\pi^{-1}(p)}$, then for $p \in U_\alpha \cap U_\beta$ we have

$$[\sigma_\alpha]|_{\pi^{-1}(p)} = [\sigma_\beta]|_{\pi^{-1}(p)} \iff \Phi^*_{\alpha,p} \circ \rho^*_\alpha[\sigma] = \Phi^*_{\beta,p} \circ \rho^*_\beta[\sigma].$$

Now consider the following diagram:

$$\begin{tikzcd}
 & \{p\} \times S^k \arrow[r, "\rho_\alpha"] & S^k \\
\pi^{-1}(p) \arrow[ur, "\Phi_{\alpha,p}"] \arrow[dr, "\Phi_{\beta,p}"'] & & \\
 & \{p\} \times S^k \arrow[r, "\rho_\beta"] \arrow[uu, "\mathrm{id}\times\tau_{\alpha\beta}"] & S^k \arrow[uu, "\tau_{\alpha\beta}"']
\end{tikzcd}$$

This then implies that

$$\Phi^*_{\beta,p} \circ \rho^*_\beta[\sigma] = \Phi^*_{\alpha,p} \circ \rho^*_\alpha \circ \tau_{\alpha\beta}(p)^*[\sigma].$$

Since $\tau_{\alpha\beta}(p) \in \mathrm{SO}(k+1)$, we see that $\tau_{\alpha\beta}(p)^*[\sigma] = [\sigma]$, and $[\sigma_\alpha] = [\sigma_\beta]$. Thus $S(E)$ is orientable.

Conversely, let $\{U_\alpha, [\sigma_\alpha]\}$ be an orientation on the sphere bundle $S(E)$ and let (S^n, σ) be an oriented sphere in $\mathbb{R}^{k+1}$, where σ is a nontrivial top form on S^k. By shrinking U_α, we can also assume that $E|_{U_\alpha}$ is trivial for each α with a trivialization Φ_α. We may assume that Φ_α preserves the metric, so that the transition functions are in $\mathrm{O}(k+1)$. Now since $S(E)$ is orientable, we can adjust the trivializations Φ_α further such that

$$\Phi^*_\alpha \circ \rho^*_\alpha[\sigma] = [\sigma_\alpha].$$

Then at any point p in $U_\alpha \cap U_\beta$ the transition function $\tau_{\alpha\beta}(p)$ pulls $[\sigma]$ to itself and so $\tau_{\alpha\beta}(p)$ must be in $\mathrm{SO}(k+1)$. $\square$

Remark 3.2.2. Since $\mathrm{SO}(1) = \{1\}$, a line bundle L over a connected base space is orientable if and only if it is trivial. In this case the sphere bundle $S(L)$ consists of two connected components.

et orientable **Proposition 3.2.18.** *A vector bundle $E \to M$ is orientable if and only if its determinant bundle $\det E$ is orientable.*

Proof. The determinant bundle of E is a line bundle on M with transition functions $\det \tau_{\alpha\beta}$, where $\tau_{\alpha\beta}$ is the transition function of a trivialization $\{(U_\alpha, \Psi_\alpha)\}$ of E. For another trivilization $\{(U_\alpha, \Psi_\alpha)\}$, by Lemma vector bundle equivalent transition lem **??** there exist maps $\sigma_\alpha : U_\alpha \to \mathrm{GL}_k(\mathbb{R})$ such that

$$\tau'_{\alpha\beta}(p) = \sigma_\alpha(p)\tau_{\alpha\beta}(p)\sigma_\beta^{-1}(p), \quad p \in U_\alpha \cap U_\beta.$$

By taking determinant we see that $\{\det \tau_{\alpha\beta}\}$ is equivalent to $\{\det \tau'_{\alpha\beta}\}$, so they define the same bundle.

It is clear that this definition does not depends on the trivialization. An orthogonal matrix $\tau_{\alpha\beta}$ assumes values in $\mathrm{SO}(k+1)$ if and only if $\det \tau_{\alpha\beta}$ is positive, so the proposition follows. $\square$

Whether E is orientable or not, the 0-sphere bundle $S(\det E)$ is always a 2-sheeted covering of M. Combining Proposition vector bundle otientable iff det orientable 3.2.18, we see that over a connected base space a vector bundle E is orientable if and only if $S(\det E)$ is disconnected. Since a simply connected base space cannot have any connected covering space of more than one sheet, we have proven the following.

ed orientable **Proposition 3.2.19.** *Every vector bundle over a simply connected base space is orientable.*

In particular, the tangent bundle of a simply connected manifold is orientable. Since a manifold is orientable if and only if its tangent bundle is, this gives

Corollary 3.2.20. *Every simply connected manifold is orientable.*

The Euler class of an oriented sphere bundle

We first consider the case of a circle bundle $\pi : E \to M$ with structure group $\mathrm{Diff}(S^1)$. Our problem is to find a closed global 1-form on E which restricts to a generator of the cohomology on each fiber. As a first approximation, in each U_α of a good cover $\{U_\alpha\}$ of M we choose a generator $[\sigma_\alpha]$ of $H^1(E|_{U_\alpha})$. The collection $\{\sigma_\alpha\}$ is an element $\sigma^{1,0}$ in the double complex $C^*(\pi^{-1}(\mathcal{U}), \Omega^*)$. From the isomorphism between the cohomology of E and the cohomology of this double complex

$$H^*_{dR}(E) \cong H^*_{TC}(C^*(\pi^{-1}(\mathcal{U}), \Omega^*))$$

which is induced by the restriction, we see that to find a global form which restricts to the d-cohomology class of $\sigma^{0,1}$ it suffices to extend $\sigma^{0,1}$ to a D-cocycle. The first step of the extension requires that $\delta(\sigma_\alpha)$ be exact, i.e., $[\sigma_\alpha] = [\sigma_\beta]$ for all α, β. This is precisely the orientability condition. Assume the bundle E to be oriented with orientation $\sigma^{0,1}$, so that $\delta\sigma^{0,1} = d\sigma^{1,0}$ for some $\sigma^{1,0}$ in $C^1(\pi^{-1}(\mathcal{U}), \Omega^0)$. Then $\sigma^{0,1} + \sigma^{1,0}$ is a D-cocycle if and only if $\delta\sigma^{1,0} = 0$. Since

$$d(\delta\sigma^{1,0}) = \delta(d\sigma^{1,0}) = \delta\delta\sigma^{0,1} = 0,$$

$\delta\sigma^{1,0}$ actually comes from an element $-\varepsilon$ of the cochain group $C^2(\pi^{-1}(\mathcal{U}), \mathbb{R})$. Now since the open covers $\mathcal{U}$ and $\pi^{-1}(\mathcal{U})$ have the same combinatorics, $C^*(\pi^{-1}(\mathcal{U}), \mathbb{R}) = C^*(\mathcal{U}, \mathbb{R})$ and we may regard ε as an element of $C^2(\mathcal{U}, \mathbb{R})$. In fact, because $\delta\varepsilon = 0$, ε defines a Čech cohomology class in $H^2(\mathcal{U}, \mathbb{R})$. By the isomorphism between the Čech cohomology of a good cover and de Rham cohomology, ε corresponds to a cohomology class $e(E)$ in $H^2(M)$, which is called the **Euler class**.

The discussion above generalizes immediately to any sphere bundle with fiber S^k, $k \geq 1$. Such a sphere bundle is orientable if and only if it is possible to find an element $\sigma^{0,k}$ in $C^0(\pi^{-1}(\mathcal{U}), \Omega^n)$ which extends one step down toward being a D-cocycle:

$$\delta\sigma^{0,k} = d\sigma^{1,k-1}.$$

There is no obstruction to extending $\sigma^{1,k-1}$ one step further, since $E|_{U_{i_0 i_1 i_2}}$ is trivial and thus every closed $(k-1)$-form on it is exact. In general, extension is possible until we hit a nontrivial cohomology of the fiber. Thus for an oriented sphere bundle E we can extend all the way down to $\sigma^{k,0}$ as in the diagram:

$$\begin{array}{ccccccccccc}
\sigma^{0,k} & \xrightarrow{\delta} & \delta\sigma^{0,k} & & & & & & & & \\
 & & d\uparrow & & & & & & & & \\
 & & \sigma^{1,k-1} & \xrightarrow{\delta} & \delta\sigma^{1,k-1} & & & & & & \\
 & & & & d\uparrow & & & & & & \\
 & & & & \sigma^{2,k-2} & \longrightarrow & \cdots & & & & \\
 & & & & & & \uparrow & & & & \\
 & & & & & & \cdots & \rightarrow & \delta\sigma^{k-1,1} & & \\
 & & & & & & & & d\uparrow & & \\
 & & & & & & & & \sigma^{k,0} & \xrightarrow{\delta} & \delta\sigma^{k,0} \\
 & & & & & & & & & & i\uparrow \\
 & & & & & & & & & & -\varepsilon
\end{array}$$

If we set

$$\sigma = \sigma^{1,k} + \sigma^{1,k-1} + \cdots + \sigma^{k,0},$$

then by our construction,

$$D\sigma = \delta\sigma^{k,0}.$$

Since $d(\delta\sigma^{k,0}) = \delta(d\sigma^{k,0}) = \delta\delta(\sigma^{k-1,1}) = 0$, we find $D\sigma = \delta\sigma^{k,0} = i(-\varepsilon)$ for some $\varepsilon \in C^{k+1}(\pi^{-1}(U), \mathbb{R}) = C^{k+1}(\mathcal{U}, \mathbb{R})$, where i is the inclusion $C^{k+1}(\pi^{-1}(\mathcal{U}), \mathbb{R}) \to C^{n+1}(\pi^{-1}(\mathcal{U}), \Omega^0)$. Clearly $\delta\varepsilon = 0$, so ε defines a cohomology class $e(E)$ in $H^{k+1}(\mathcal{U}, \mathbb{R}) \cong H^{k+1}(M)$, the Euler class of the oriented S^k-bundle E. The Euler class of an oriented S^0-bundle is defined to be 0. Note that the Euler class depends on the orientation $\{[\sigma_\alpha]\}$ on E; the opposite orientation would give $-e(E)$ instead.

If E is an oriented vector bundle, the complement E^0 of its zero section has the homotopy type of an oriented sphere bundle. The Euler class of E is defined to be that of E^0. Equivalently, if E is endowed with a Riemannian metric, then the unit sphere bundle $S(E)$ of E makes sense and we may define the Euler class of E to be that of its unit sphere bundle. This latter definition is independent of the metric and in fact agrees with the definition in terms of E^0, since for any metric on E, the unit sphere bundle $S(E)$ has the homotopy type of E^0.

In the next two propositions we show that the Euler class is well defined.

Proposition 3.2.21. *For a given orientation $\{[\sigma_\alpha]\}$ the Euler class is independent of the choice of $\sigma^{i,k-i}$, $j = 0, \ldots, k$.*

Proof. This follows by a diagram chasing. □

Proposition 3.2.22. *The Euler class $e(E)$ is independent of the choice of the good cover.*

Proof. Write $\varepsilon_{\mathcal{U}}$ for the cocycle in $H^{k+1}(\mathcal{U}, \mathbb{R})$ which defines the Euler class in terms of the good cover $\mathcal{U}$. If a good cover $\mathcal{V}$ is a refinement of $\mathcal{U}$, then there is a commutative diagram

$$\begin{tikzcd}
H^{k+1}(\mathcal{U}, \mathbb{R}) \arrow[rr] \arrow[dr, "\cong"'] & & H^{k+1}(\mathcal{V}, \mathbb{R}) \arrow[dl, "\cong"] \\
 & H^{k+1}_{dR}(M) &
\end{tikzcd}$$

Note that $\varepsilon_{\mathcal{U}}$ and $\varepsilon_{\mathcal{V}}$ give the same element in $H^{k+1}_{dR}(M)$, because if we choose the $\sigma^{0,k}$ on $\pi^{-1}(\mathcal{V})$ to be the restriction of the $\sigma^{0,k}$ on $\pi^{-1}(\mathcal{U})$, the cocycle $\varepsilon_{\mathcal{V}}$ in $C^{k+1}(\mathcal{V},\mathbb{R})$ will be the restriction of the cocycle $\varepsilon_{\mathcal{U}}$ in $C^{k+1}(\mathcal{U},\mathbb{R})$, so that as elements of the Čech cohomology $H^{k+1}(M,\mathbb{R})$ they are equal. Given two arbitrary good covers $\mathcal{U}$ and $\mathcal{B}$, we can take a common refinement $\mathcal{V}$; then $\varepsilon_{\mathcal{U}} = \varepsilon_{\mathcal{V}} = \varepsilon_{\mathcal{B}}$ in $H^{k+1}_{dR}(M)$. So the Euler class is independent of the cover. □

Now we prove the promised theorem about the existence of a global form that restricts to a generator on each fiber.

class vanish **Proposition 3.2.23.** *Let $\pi : E \to M$ be a oriented sphere bundle with an orientation $\{\sigma_\alpha\}$. Then there exists a global form on E that restricts to each $[\sigma_\alpha]$ on U_α if and only if the Euler class of E vanishes.*

Proof. If the Euler class $e(E) \in H^{k+1}(M)$ vanishes, its representative $\varepsilon \in C^{k+1}(\mathcal{U},\mathbb{R})$ is a δ-coboundary; this permits one to alter $\sigma^{k,0}$ so that $D\sigma = 0$. The D-cocycle σ then corresponds to a global form $\psi \in H^k(E)$ via the isomorphism

$$r^* : H^*(E) \to H^*_{TC}(C^*(\pi^{-1}(\mathcal{U},\Omega^*))), \quad \omega \mapsto (\omega|_{U_\alpha}).$$

This then implies that

$$(\psi|_{U_\alpha}) - \sigma = D\eta \ \text{ for } \ \eta = \eta^1 + \cdots + \eta^{k-1} \in \bigoplus_{i+j=k-1} C^i(\pi^{-1}(\mathcal{U}),\Omega^j).$$

Expanding the definition and concentrate on the summand $C^0(\pi^{-1}(\mathcal{U},\Omega^k))$, we find that

$$(\psi|_{U_\alpha}) - \sigma^{0,k} = d\eta^1.$$

This then implies that $[\psi|_{U_\alpha}] = [\sigma_\alpha]$ for all α, so ψ is a global form that restricts to each σ_α, and hence a generator on each fiber. Conversely, if we can write $[\sigma_\alpha] = [\psi|_{U_\alpha}]$ for some global form $\psi \in H^k(E)$, then we can choose $\sigma^{0,k} = (\psi|_{U_\alpha})$ and hence $\sigma^{1,k-1} = \cdots = \sigma^{k,0} = 0$. This then implies $e(E) = 0$. □

For a product bundle E, we can easily find orientations $\{[\sigma_\alpha]\}$ such that $\sigma_\alpha = \sigma_\beta$ on $U_\alpha \cap U_\beta$, rather than the equality on d-cohomology class. Then it follows by our construction that the Euler class for E vanishes. In this sense the Euler class is a measure of the twisting of an oriented sphere bundle. However, as we will see in the proposition below, E need not be a product bundle for its Euler class to vanish.

Proposition 3.2.24. *If the oriented sphere bundle E has a global section, then its Euler class vanishes.*

Proof. Let s be a section of E. It follows from $\pi \circ s = \mathrm{id}$ that $s^* \circ \pi^* = \mathrm{id}$. We saw in the construction of the Euler class that

$$D\sigma = -\pi^*\varepsilon$$

for some D-cochain σ. Applying s^* to both sides gives $s^*D\sigma = Ds^*\sigma = \varepsilon$. By the same diagram chasing as in the proof of Proposition 3.2.23, we can show that ε is a coboundary in $H^*(\mathcal{U},\mathbb{R})$. Therefore the Eucler class $e(E)$ vanishes. □

sphere bundle global generator iff Euler class vanish

The global angular form

Using the collating formula we will now construct a form ψ on any oriented S^k-bundle such that

- its restriction to each fiber is a generator of the cohomology of the fiber.
- $d\psi = -\pi^* e$, where e represents the Euler class of the circle bundle.

Let $U = \{U_\alpha\}$ be an open cover of M. Recall that the Euler class of E is defined by the following diagram:

$$
\begin{array}{cccccccccc}
\sigma^{0,k} & \longrightarrow & \delta\sigma^{0,k} & & & & & & & \\
 & & \uparrow & & & & & & & \\
 & & \sigma^{1,k-1} & \to & \delta\sigma^{1,k-1} & & & & & \\
 & & & & \uparrow & & & & & \\
 & & & & \sigma^{2,k-2} & \longrightarrow & \cdots & & & \\
 & & & & & & \uparrow & & & \\
 & & & & & & \cdots \to & \delta\sigma^{k-1,1} & & \\
 & & & & & & & \uparrow & & \\
 & & & & & & & \sigma^{k,0} & \longrightarrow & \delta\sigma^{k,0} \\
 & & & & & & & & & \uparrow \\
 & & & & & & & & & -\varepsilon
\end{array}
$$

where $\alpha_0 \in C^0(\mathcal{U}, \Omega^*)$ is the orientation of E and

$$\delta\alpha_i = d\alpha_{i+1}, \quad \delta\alpha_n = -\pi^*\varepsilon.$$

Hence $D(\alpha_0 + \cdots + \alpha_n) = -\pi^*\varepsilon$, where α_i is what we formerly wrote as $\sigma^{i,k-i}$.

If $\{\rho_i\}$ is a partition of unity subordinate to the open cover $U = \{U_i\}$, then $\{\pi^*\rho_i\}$ is a partition of unity subordinate to the inverse cover $\pi^{-1}(\mathcal{U})$. Using these data we can define a homotopy operator K on the double complex $C^*(\mathcal{U}, \Omega^*)$ and also one on $C^*(\pi^{-1}(\mathcal{U}), \Omega^*)$ as (2.1). We denote both operators by K. Both K satisfy

$$\delta K + K\delta = \text{id}.$$

Moreover, K commutes with π^* since

$$(K\pi^* i)_{i_0 \ldots i_{p-1}} = \sum_i (\pi^*\rho_i)(\pi^*\omega)_{i, i_0 \ldots i_{p-1}} = \pi^* \sum_i \rho_i \omega_{i, i_0 \ldots i_{p-1}} = \pi^* K\omega.$$

By the collating formula (2.2), the form

$$\psi := f(\alpha) = \sum_{i=0}^{k} (-dK)^i \alpha_i + (-1)^{k+1} K(-dK)^k(-\pi^*\varepsilon)$$

is a global form on E. Furthermore, we have

$$d\psi = df(\alpha) = f(D\alpha) = f(-\pi^*\varepsilon) = -\pi^*(-dK)^{k+1}\varepsilon = -\pi^* e.$$

By Corollary 3.2.5 the global form ψ to each fiber is d-cohomologous to α_0, hence is a generator of the cohomology of the fiber. The global n-form ψ on the sphere bundle E is called the **global angular form** on the sphere bundle.

Proposition 3.2.25. *Let $\{U_i\}$ be an open cover of M which trivializes the k-sphere bundle E and let ψ and e be the global angular form and the Euler class. Then*

$$\operatorname{supp}(e) \subseteq \bigcup U_{i_0 \ldots i_{k+1}}, \quad \operatorname{supp}(d\psi) \subseteq \bigcup \pi^{-1}(U_{i_0 \ldots i_{k+1}}).$$

de Rham MV generalized homotopy
Proof. By (2.1) we have

$$\operatorname{supp}(K\omega)_{i_0\ldots i_{p-1}} \subseteq \bigcup_i \omega_{i,i_0\ldots i_p} \subseteq \bigcup_i U_{i,i_0\ldots i_p}.$$

Since $\varepsilon \in C^{k+1}(\mathcal{U},\mathbb{R})$, we must have $\operatorname{supp}(\varepsilon) \subseteq \bigcup U_{i_0\ldots i_{k+1}}$. Then the claim follows from $d\psi = -\pi^* e$ and $e = f(\varepsilon)$. $\square$

The Euler class of an oriented rank 2 vector bundle

In this part we will construct explicitly the Euler class of an oriented rank 2 vector bundle $\pi : E \to M$, using such data as a partition of unity on M and the transition functions of E.

Now let $\pi : E \to M$ be a rank 2 vector bundle on a manifold M and $\{U_\alpha\}$ be a coordinate open cover of M that trivializes E. Since E has a Riemannian structure, over each U_α we can choose an orthonormal frame. This defines on $E^0|_{U_\alpha}$ polar coordinates r_α and θ_α; if $(x^1,\ldots,x^n)$ are coordinates on U_α, then $(\pi^*x^1,\ldots,\pi^*x^n, r_\alpha, \theta_\alpha)$ are coordinates on $E^0|_{U_\alpha}$. On the overlap $U_\alpha \cap U_\beta$, the radii r_α and r_β are equal but the angular coordinates θ_α and θ_β differ by a rotation. Since E is orientable, we can define unambiguously $\varphi_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathbb{R}$ (up to a constant multiple of 2π) as the angle of rotation in the counterclockwise direction from the α-coordinate system to the β-coordinate system:

$$\theta_\beta - \theta_\alpha = \pi^*\varphi_{\alpha\beta}.$$

Moreover, from the observation

$$\varphi_{\alpha\beta} + \varphi_{\beta\gamma} - \varphi_{\alpha\gamma} \in 2\pi\mathbb{Z},$$

it follows that the forms $\{d\theta_\alpha/2\pi\}$ give an orientation for the sphere bundle $S(E)$. Therefore, in view of the definition of the Euler class, we consider the following diagram:

$$\begin{array}{ccccc}
\dfrac{d\theta_\alpha}{2\pi} & \xrightarrow{\ \delta\ } & \dfrac{\pi^* d\varphi_{\alpha\beta}}{2\pi} & & \\
 & & \Big\uparrow{\scriptstyle d} & & \\
 & & \dfrac{\pi^*\varphi_{\alpha\beta}}{2\pi} & \xrightarrow{\ \delta\ } & -\pi^*\varepsilon
\end{array}$$

Cech de Rham to de Rham
By Corollary 3.2.5, the Euler class e is given by the cohomology class of $(-d_v K)^2\varepsilon$. Since $-\varepsilon = \delta(\varphi/2\pi)$, we have

$$\begin{aligned}
(-d_vK)^2\varepsilon &= -d_vKd_vK\delta(\frac{\varphi}{2\pi}) = dKdK\delta(\frac{\varphi}{2\pi}) = dKd(\frac{\varphi}{2\pi} - \delta K(\frac{\varphi}{2\pi})) \\
&= dK(\frac{d\varphi}{2\pi}) - dKd\delta K(\frac{\varphi}{2\pi}).
\end{aligned}$$

(Here we use d_v to denote the differential in the double complex, that is, $d_v = (-1)^p d$.) If we set $\xi := K(d\varphi/2\pi)$, then

$$d\xi - dKd\delta\xi = d\xi - dK\delta d\xi = d\xi - d(\mathrm{id} - \delta K)d\xi = d\xi + d\delta K d\xi = d\xi + \delta dKd\xi.$$

Note that by definition $Kd\xi \in \Omega^1(M)$, therefore $dKd\xi$ is a global exact form. This then implies $[\delta dKd\xi] = 0$, and so the cohomology class of $(-d_vK)^2\varepsilon$ equals $d\xi$. In summary, we get the following simple formula for the Euler class:

$$e = d\xi.$$

The Euler class of an oriented rank 2 vector bundle may be given in terms of the transition functions, as follows. Let $\tau_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{SO}(2)$ be the transition functions of E. By identifying $\mathrm{SO}(2)$ with the unit circle in the complex plane via the map

$$e^{i\theta} \mapsto \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

the map $\tau_{\alpha\beta}$ may be thought of as complex-valued functions with values in $S^1 \subseteq \mathbb{C}^2$. In this context the angle from the β-coordinate system to the α-coordinate system is $(1/i)\log\tau_{\alpha\beta}$. Thus

$$\theta_\alpha - \theta_\beta = \frac{1}{i}\pi^* \log\tau_{\alpha\beta}.$$

This means we can choose $\varphi_{\alpha\beta}$ to be

$$\varphi_{\alpha\beta} = -\frac{1}{i}\log\tau_{\alpha\beta}.$$

Let $\{\rho_\alpha\}$ be a partition of unity subordinate to $\{U_\alpha\}$. Then the form ξ is defined to be

$$\xi_\alpha = \frac{1}{2\pi}\sum_\gamma \rho_\gamma d\varphi_{\gamma\alpha}.$$

Therefore on U_α we have

$$e(E) = d\xi = -\frac{1}{2\pi i}\sum_\gamma d(\rho_\gamma d\log\tau_{\gamma\alpha}). \tag{2.4}$$

Euler class r

Euler number and the isolated singularities of a section

Let $\pi : E \to M$ be an oriented $(n-1)$-sphere bundle over a compact oriented manifold of dimension n. Since $H^n(M) = \mathbb{R}$, the Euler class of E may be identified with the number $\int_M e(E)$, which is by definition the **Euler number** of E. The Euler number of the manifold M is defined to be that of its unit tangent bundle $S(TM)$ relative to some Riemannian structure on M. While the Euler number of an orientable sphere bundle is defined only up to sign, depending on the orientations of both E and M, the Euler number of the orientable manifold M is unambiguous, since reversing the orientation of M also reverses that of the tangent bundle.

In general the sphere bundle E will not have a global section; however, there may be a section s over the complement of a finite number of points $p_1, \dots, p_r$ in M. In fact, as we will show, if the sphere bundle has structure group $\mathrm{O}(n)$, then such a "partial" section s always exists. In this part we will explain how one may compute the Euler class of E in terms of the behavior of the section s near the singularities $p_1, \dots, p_k$.

Proposition 3.2.26. *Let $\pi : E \to M$ be a $(n-1)$-sphere bundle over a compact manifold of dimension n. Suppose the structure group of E can be reduced to $\mathrm{O}(n)$. Then E has a section over $M \setminus \{p_1, \dots, p_r\}$ for some finite number of points in M.*

Proof. Since the structure group of E is $\mathrm{O}(n)$, we can form a Riemannian vector bundle E' of rank n whose unit sphere bundle is E. A section s' of E' over M gives rise to a partial section s of E: $s(x) = s'(x)/|s'(x)|$, where $|\cdot|$ denotes the length of a vector in E'. Let Z be the zero locus of s'; s is only a partial section in the sense that it is not defined over Z. Thus to prove the proposition, we only have to show that the vector bundle E' has a section that vanishes over a finite number of points. $\square$

Suppose s is a section over a punctured neighborhood of a point p in M. Choose this neighborhood sufficiently small so that it is diffeomorphic to a punctured disc in $\mathbb{R}^n$ and E is trivial over it. Let D_r be the open neighborhood of p corresponding to the ball of radius r in $\mathbb{R}^n$ under the diffeomorphism above. As an open subset of the oriented manifold M, D_r is also oriented. Choose the orientation on the sphere S^{n-1} in such a way that the isomorphism $E|_{D_r} \cong D_r \times S^{n-1}$ is orientation-preserving, where $D_r \times S^{n-1}$ is given the product orientation. The **local degree** of the section s at x is defined to be the degree of the composite map

$$\partial \overline{D}_r \xrightarrow{\ s\ } E|_{\overline{D}_r} \xrightarrow{\ \cong\ } \overline{D}_r \times S^{n-1} \xrightarrow{\ \rho\ } S^{n-1}$$

where ρ is the projection.

Now we will show that the Euler number of E can be computed using these local degrees. First we show that it is possible to move the support of the Euler class away from finitely many points.

Lemma 3.2.27. *Let M be a manifold and $\{U_\alpha\}_{\alpha\in A}$ an open cover of M. Given finitely many points $p_1, \dots, p_r$ on M, there is a refinement $\{V_\alpha\}_{\alpha\in A}$ of $\{U_\alpha\}$ such that $V_\alpha \subseteq U_\alpha$ and each p_i has a neighborhood W_i which is disjoint from all but one of the V_α's.*

Proof. Suppose $p_1 \in U_1$. Let W_1 be an open neighborhood of p_1 such that $p_1 \in W_1 \subseteq \overline{W_1} \subseteq U_1$. We define a new open cover $\{U'_\alpha\}_{\alpha\in A}$ by setting $U'_1 = U_1$ and $U'_\alpha = U_\alpha \setminus \overline{W}_1$ for $\alpha \neq 1$. The neighborhood W_1 of p_1 is contained in U_1 but disjoint from all U'_α except U'_1.

Next suppose $p_2 \in U_2$. Let W_2 be an open neighborhood of p_2 such that $p_2 \in W_2 \subseteq \overline{W_2} \subseteq U'_2$. As before define a new open cover $\{U''_\alpha\}$ by setting $U''_2 = U'_2$ and $U''_\alpha = U'_\alpha - W_2$ for $\alpha \neq 2$. Since $U''_\alpha \subseteq U'_\alpha$ the open neighborhood W_1 of p_1 is disjoint from all U''_α except U''_1. By definition, the open neighborhood W_2 of p_2 is disjoint from all U''_α but U''_2. Repeating this process to $p_3, \dots, p_r$ in succession yields the open cover $\{V_\alpha\}$ of the lemma. □

local degree **Theorem 3.2.28.** *Let $\pi : E \to M$ be an oriented $(n-1)$-sphere bundle over a compact oriented manifold of dimension n. If E has a section over $M \setminus \{p_1, \dots, p_r\}$, then the Euler number of E is the sum of the local degrees of s at $p_1, \dots, p_r$.*

Proof. Let $\{U_\alpha\}$ be an open cover of M which trivializes E. By the lemma we may assume that each p_i has a neighborhood W_i which is contained in exactly one U_α. Construct the global angular form ψ; and the form e relative to $\{U_\alpha\}$. By Proposition 3.2.25, the form e sphere bundle Euler class supp must vanish on W_i for all $i = 1, \dots, r$. So e is supported away from the points p_i.

For each i choose an open ball D_i around the point p_i so that $\overline{D}_i \subseteq W_i$. Then

$$\int_M e = \int_{M - \bigcup_i D_i} e = \int_{M-\bigcup_i D_i} s^*\pi^* e = -\int_{M-\bigcup_i D_i} s^* d\psi = \sum_i \int_{\partial \overline{D}_i} s^*\psi.$$

where the last step is obtained by Stokes' theorem and the fact that $\partial\overline{D_i}$ has the opposite orientation as $\partial(M - \bigcup_i \overline{D}_i)$. Although the global angular form is not closed, by our construction $d\psi = -\pi^* e = 0$ on $E|_{W_i}$, so ψ defines a cohomology class in $H^{n-1}(E|_{W_i})$, which is in fact the generator. Let σ be the generator of S^{n-1}. Then $\rho^*\sigma$ restricts to the generator on each fiber of $E|_{W_i}$. So $\rho^*\sigma$ and the angular form ψ define the same cohomology class in $H^{k-1}(E|_{W_i})$. This then implies

$$\int_{\partial\overline{D}_i} s^*\psi = \int_{\partial\overline{D}_i} s^*\rho^*\sigma.$$

By definition, the right hand is the local degree of p_i, so we get the claim. $\square$

This theorem can also be phrased in terms of vector bundles. Let $\pi : E \to M$ be an oriented rank n vector bundle over a manifold of dimension n and s a section of E with a finite number of zeros. The **multiplicity** of a zero p of s is defined to be the local degree of p as a singularity of the section $s/|s|$ of the unit sphere bundle of E relative to some Riemannian structure on E. (This definition of the index is independent of the Riemannian structure because the local degree is a homotopy invariant.) Since the Euler class $e(E)$ of E is a n-form on M. it is Poincare dual to nP, where $n = \int_M e(E)$ and P is a point on M. Thus we have the following.

Theorem 3.2.29. *Let $\pi : E \to M$ be an oriented rank n vector bundle over a compact oriented manifold of dimension n. Let s be a section of E with a finite number of zeros. The Euler class of E is Poincare dual to the zeros of s, counted with the appropriate multiplicities.*

Example 3.2.30 (The Euler class of the unit tangent bundle to S^2). Let $S(TS^2)$ be the unit tangent bundle to S^2. It is a circle bundle over S^2:

$$S^1 \to S(TS^2) \to S^2.$$

Fix a unit tangent vector v at the north pole N. We can define a smooth vector field on $S^2 \backslash \{S\}$ by parallel translating v along the great circles from the north pole to the south pole, where S is the south pole. This gives a section s of $S(TS^2)$ over $S^2 \setminus \{S\}$. On a small circle around the south pole, as we go around the circle 90, the vectors rotate through 180; therefore, the local degree of s at the south pole is 2. By Theorem 3.2.28, the Euler number of the unit tangent bundle to S^2 is 2.

More generally, the same technique can be used to compute the Euler number of $S(TS^n)$. Since there exists nonvanishing vector fields on S^{k+1}, we find that

$$e(S(TS^n)) = \begin{cases} 0, & n \text{ is odd}, \\ 2 & n \text{ is even}. \end{cases}$$

Euler characteristic and the Hopf index theorem

In this part we show that the Euler number $\int_M e(TM)$ is the same as the Euler characteristic $\chi(M) = \sum_p (-1)^p \dim H^p(M)$ and deduce as a corollary the Hopf index theorem. The manifold M is assumed to be compact and oriented. Let $\{\omega_i\}$ be a basis of the vector space $H^*(M)$, $\{\tau_j\}$ be the dual basis under Poincare duality, i.e., $\int_M \omega_i \wedge \tau_j = \delta_{ij}$, and let π_1 and π_2 be the two projections of $M \times M$ to M:

$$\begin{tikzcd} & M \times M \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\ M & & M \end{tikzcd}$$

By the Künneth formula, $H^*(M \times M) = H^*(M) \otimes H^*(M)$ with $\{\pi_1^*\omega_i \wedge \pi_2^*\tau_j\}$ as an additive basis. So the Poincare dual η_Δ of the diagonal Δ in $M \times M$ is some linear combination $\eta_\Delta = \sum c_{ij}\pi_1^*\omega_i \wedge \pi_2^*\tau_j$.

dual diagonal **Lemma 3.2.31.** *We have*

$$\eta_\Delta = \sum_i (-1)^{\deg \omega_i} \pi_1^* \omega_i \wedge \pi_2^* \tau_i.$$

Proof. We compute $\int_\Delta \pi_1^*\tau_k \wedge \pi_2^*\omega_l$ in two ways. On the one hand, we can pull this integral back to M via the diagonal map $i : M \hookrightarrow \Delta \subseteq M \times M$:

$$\int_\Delta \pi_1^*\tau_k \wedge \pi_2^*\omega_l = \int_M i^*\pi_1^*\tau_k \wedge i^*\pi_2^*\omega_l = \int_M \tau_k \wedge \omega_l = (-1)^{(\deg \tau_k)(\deg \omega_l)} \delta_{kl}.$$

On the other hand, by the definition of the Poincare dual of a closed oriented submanifold,

$$\begin{aligned}
\int_\Delta \pi_1^*\tau_k \wedge \pi_2^*\omega_l &= \int_{M\times M} \pi_1^*\tau_k \wedge \pi_2^*\omega_l \wedge \eta_\Delta \\
&= \sum c_{ij} \int_{M\times M} \pi_1^*\tau_k \wedge \pi_2^*\omega_l \wedge \pi_1^*\tau_i \wedge \pi_2^*\omega_j \\
&= \sum c_{ij} (-1)^{(\deg \tau_k + \deg \omega_l)(\deg \omega_i)} \int_{M\times M} \pi_1^*(\omega_i \wedge \tau_k) \wedge \pi_2^*(\omega_l \wedge \tau_j) \\
&= (-1)^{(\deg \tau_k + \deg \omega_l)(\deg \omega_i)} c_{kl}.
\end{aligned}$$

Therefore

$$c_{kl} = \begin{cases} 0 & \text{if } k \neq l, \\ (-1)^{\deg \omega_i} & \text{if } k = l. \end{cases}$$

□

Lemma 3.2.32. *The normal bundle $N\Delta$ of the diagonal Δ in $M \times M$ is isomorphic to the tangent bundle.*

Proof. This follows from the diagram

$$\begin{tikzcd}
0 \arrow[r] & T\Delta \arrow[r] \arrow[d, "\cong"] & T(M\times M)|_\Delta \arrow[r] \arrow[d, "\cong"] & N\Delta \arrow[r] & 0 \\
0 \arrow[r] & TM \arrow[r] & TM \oplus TM \arrow[r] & TM \arrow[r] & 0
\end{tikzcd}$$

□

er Euler char **Theorem 3.2.33.** *Let M be a compact orientable manifold, then the Euler number of TM equals to the Euler characteristic of M.*

Proof. Recall that the Poincare dual of a closed oriented submanifold S is represented by the same form as the Thom class of a tubular neighborhood of S. Thus

$$\int_\Delta \eta_\Delta = \int_\Delta \Phi(N\Delta) = \int_\Delta e(N\Delta) = \int_\Delta e(T\Delta) = \int_M e(TM).$$

Here we use the fact that the Thom class restricted to the zero section of the bundle is the Euler class. We will prove later. Now the right-hand side of Lemma Poincare dual diagonal 3.2.31 evaluated on the diagonal Δ is

$$\begin{aligned}\int_\Delta \eta_\Delta &= \sum_i (-1)^{\deg \omega_i} \int_\Delta \pi_1^* \omega_i \wedge \pi_2^* \tau_i = \sum_i (-1)^{\deg \omega_i} \int_M \omega_i \wedge \tau_i \\ &= \sum_i (-1)^{\deg \omega_i} = \sum_p (-1)^p \dim H^p(M) \\ &= \chi(M).\end{aligned}$$

□

It is now a simple matter to derive the Hopf index theorem. Let V be a vector field with isolated zeros on M. The **index** of V at a zero p is defined to be the local degree at p of $V/|V|$ as a section of the unit tangent bundle of M relative to some Riemannian metric on M. By Theorem sphere bundle Euler number local degree 3.2.28 the sum of the indices of V is the Euler number of M. The equality of the Euler number and the Euler characteristic then yields the following.

Theorem 3.2.34 (Hopf Index Theorem). *The sum of the indices of a vector field on a compact oriented manifold M is the Euler characteristic of M.*

Now we would like to give a generalization of Theorem Euler number Euler char 3.2.33. Consider a smooth map $f : M \to M$ of a compact orientable manifold M and the induced map $H(f)$ on $H^*(M)$. Let $\Gamma \subseteq M \times M$ be the graph of f. Let $F : M \to \Gamma$ be the map $F(p) = (p, f(p))$, then we have

$$\begin{aligned}\int_\Delta \eta_\Gamma &= \int_{M\times M} \eta_\Gamma \wedge \eta_\Delta = (-1)^{n^2} \int_{M\times M} \eta_\Delta \wedge \eta_\Gamma = (-1)^n \int_\Gamma \eta_\Delta \\ &= (-1)^n \sum_i (-1)^{\deg \omega_i} \int_\Gamma \pi_1^* \omega_i \wedge \pi_2^* \tau_i \\ &= (-1)^n \sum_i (-1)^{\deg \omega_i} \int_M F^* \pi_1^* \omega_i \wedge F^* \pi_2^* \tau_i \\ &= (-1)^n \sum_i (-1)^{\deg \omega_i} \int_M \omega_i \wedge f^* \tau_i \\ &= (-1)^n \sum_p (-1)^p \mathrm{tr} H^{n-p}(f) \\ &= \sum_p (-1)^p \mathrm{tr} H^p(f).\end{aligned}$$

Therefore, we make the following definition.

Definition 3.2.35. *Let $f : M \to M$ be a smooth map of a compact oriented manifold into itself. Denote by $H^p(f)$ the induced map on the cohomology $H^p(M)$. The Lefschetz number of f is defined to be*

$$L(f) = \sum_p (-1)^p \mathrm{tr} H^p(f).$$

Theorem 3.2.36 (Lefschetz fixed point theorem). *Let $f : M \to M$ be a smooth map of a compact oriented manifold into itself. If $L(f) \neq 0$, then f has a fixed point.*

Proof. By our computation above, we have

$$L(f) = \int_{M\times M} \eta_\Gamma \wedge \eta_\Delta.$$

If f has no fixed point, then $\Gamma \cap \Delta = \varnothing$. Since they are both closed, there exists disjoint neighborhoods of Γ and Δ. Now the forms η_Γ and η_Δ can be choosen to have supports in these neighborhoods, so $\eta_\Gamma \wedge \eta_\Delta = 0$ and $L(f) = 0$. $\square$

At a fixed point P of f the derivative df_p is an endomorphism of the tangent space T_pM. We define the multiplicity of the fixed point p to be

$$m_p = \operatorname{sgn}(\det(df_p - \mathrm{id})).$$

Now assume that the graph Γ is transversal to the diagonal Δ in $M \times M$, then $\eta_\Gamma \wedge \eta_\Delta$ is in fact the dual $\eta_{\Gamma\cap\Delta}$ and thus we have

$$\int_{M\times M} \eta_\Gamma \wedge \eta_\Delta = \int_{\Gamma\cap\Delta} 1.$$

Note that $\Gamma \cap \Delta$ is the fixed points of f and since Γ is transversal to Δ, it is also an orientable manifold. Now M is compact implies $\Gamma \cap \Delta$ is finite, and its orientation is determined by the matrix

$$\begin{pmatrix} df_p & I \\ I & I \end{pmatrix}$$

The determinant of this matrix is $\det(df_p - \mathrm{id})$, therefore the number m_p gives the orientation of $\Gamma \cap \Delta$. With this, we then conclude

$$L(f) = \sum_p m_p.$$

3.2.3 Thom isomorphism and Poincaré duality revisited

Thom isomorphism

Let $\pi : E \to M$ be a rank k vector bundle. E is not assumed to be orientable. We are interested in the cohomology of E with compact support in the vertical direction $H^*_{cv}(E)$. Recall that

$$H^*_c(\mathbb{R}^n) = \begin{cases} \mathbb{R} & p = 0, \\ 0 & \text{otherwise} \end{cases} \qquad H^*_{cv}(M \times \mathbb{R}^k) = H^{*-k}(M).$$

Let $\mathcal{U}$ be a good cover of the base manifold M. We augment the double complex $C^*(\pi^{-1}(\mathcal{U}), \Omega^*_{cv})$ by adding a column consisting of the kernels of the first δ:

$$0 \longrightarrow \Omega^*_{cv}(E) \longrightarrow C^0(\pi^{-1}(\mathcal{U}), \Omega^*_{cv}) \longrightarrow \cdots$$

Using a partition of unity from the base, it can be shown that this agumented complex are exact. It follows that the cohomology of the initial column is the total cohomology of the double complex, i.e.,

$$H^*_{cv}(E) \cong H^*_{TC}(C^*(\pi^{-1}(\mathcal{U}), \Omega^*_{cv})).$$

On the other hand, if $E_i^{p,q}$ is the spectral sequence induced by $C^*(\pi^{-1}(\mathcal{U}), \Omega^*_{cv})$ with vertical filtration, then

$$E_1^{p,q} = H^q(C^p(\pi^{-1}(\mathcal{U}), \Omega^*_{cv})) = H^q_{cv}(\prod \pi^{-1}(U_{i_0\dots i_p})) = \prod H^q_{cv}(\pi^{-1}(U_{i_0\dots i_p})) = C^p(\mathcal{U}, \mathscr{H}^q_{cv}),$$

where $\mathscr{H}^q$ is the presheaf given by

$$\mathscr{H}^q(U) = H^q(\pi^{-1}(U)).$$

By the Poincare lemma for compactly supported cohomology, if U is contractible, then $\pi^{-1}(U) \cong U \times \mathbb{R}^k$ and

$$\mathscr{H}^q_{cv}(\pi^{-1}(U)) = \begin{cases} \mathbb{R} & p = k \\ 0 & \text{otherwise} \end{cases}$$

Therefore E_1 and hence E_2 have entries only in the k-th row. This then implies

$$H^*_{cv}(E) = H^*_{TC}(C^*(\pi^{-1}(\mathcal{U}), \Omega^*_{cv})) = H^{*-k}(\mathcal{U}, \mathscr{H}^k_{cv}).$$

This is the Thom isomorphism for a not necessarily orientable vector bundle.

Theorem 3.2.37 (Thom Isomorphism). *For $\pi : E \to M$ any vector bundle of rank k over M and $\mathcal{U}$ a good cover of M,*

$$H^*_{cv}(E) \cong H^{*-k}(\mathcal{U}, \mathscr{H}^k_{cv}).$$

where $\mathscr{H}^k_{cv}$ is the presheaf $\mathscr{H}^k_{cv}(U) = H^k_{cv}(\pi^{-1}(U))$.

We now deduce the orientable version of the Thom isomorphism from this. So suppose $\pi : E \to M$ is an orientable vector bundle of rank k over M. This means there exist forms σ_α on the sphere bundles $S(E)|_{U_\alpha}$ which restrict to a generator on each fiber and such that on overlaps $U_\alpha \cap U_\beta$, their cohomology classes agree: $[\sigma_\alpha] = [\sigma_\beta]$. Now choose a Riemannian metric on E so that the radius r is well-defined on each fiber. Let $\rho(r)$ be the function shown in Figure 3.1. Then $(d\rho) \wedge \sigma_\alpha$ is a form on $E|_{U_\alpha}$, where we regard σ_α as a form on the complement of the zero section. Furthermore, $[(d\rho) \wedge \sigma_\alpha] \in H^k_{cv}(E|_{U_\alpha})$ restricts to a generator of the compactly supported cohomology of the fiber and $[(d\rho) \wedge \sigma_\alpha] = [(d\rho) \wedge \sigma_\beta]$ on $U_\alpha \cap U_\beta$. Recall that

$$\mathscr{H}^p_{cv}(\pi^{-1}(U)) = \begin{cases} \mathbb{R} & p = k \\ 0 & \text{otherwise} \end{cases}$$

so we can extend $\sigma^{0,k} = \{d(\rho) \wedge \sigma_\alpha\}$ to a chain in $C^*(\mathcal{U}, \Omega^*_{cv})$, and this chain is closed. This D-cocycle corresponds to a global closed form Φ on E, the **Thom class** of E, which restricts to a generator on each fiber. Now $\mathscr{H}^k_{cv}(U)$ is generated by $\Phi|_U$ and for $V \subseteq U$ the restriction map from $\mathscr{H}^k_{cv}(U)$ to $\mathscr{H}^k_{cv}(V)$ sends $\Phi|_U$ to $\Phi|_V$. Hence, via the map which sends $\Phi|_U$ for every open set U to the generator 1 of the constant presheaf $\mathbb{R}$, the presheaf $\mathscr{H}^k_{cv}$ is isomorphic to $\mathbb{R}$. The Thom isomorphism theorem then assumes the form

$$H^*_{cv}(E) \cong H^{*-k}(\mathcal{U}, \mathscr{H}^k_{cv}) \cong H^{*-k}(\mathcal{U}, \mathbb{R}) \cong H^{*-k}(M).$$

for an orientable rank k vector bundle E. This agrees with Proposition 3.1.56. It holds in particular when M is simply connected, since by Corollary 3.2.19, every vector bundle over a simply connected manifold is orientable.

Let $f : E^0 \to S(E)$ be a deformation retraction of the complement of the zero section in E onto the unit sphere bundle. If ψ_S is the global angular form on $S(E)$, then $\psi = f^*\psi_S \in H^{k-1}(E^0)$ is the global angular form on E^0. It has the property that

$$d\psi = -\pi^* e.$$

where e represents the Euler class of the bundle E.

Proposition 3.2.38. *The cohomology class of*

$$\Phi = d(\rho(r) \wedge \psi) \in \Omega^k_{cv}(E)$$

is the Thom class of the oriented vector bundle E.

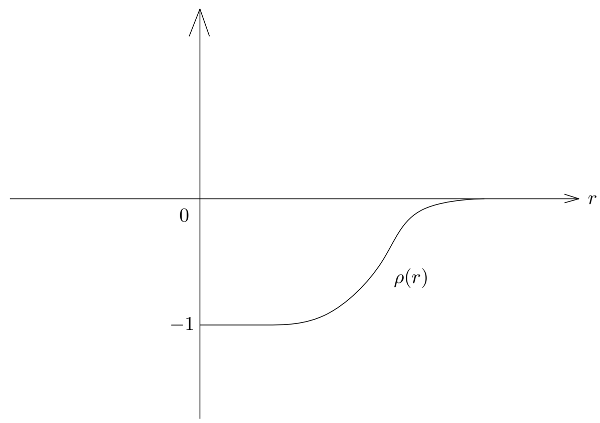

Figure 3.1: The function $\rho(r)$. Thom iso func

Proof. Note that

$$\Phi = d\rho(r) \wedge \psi + \rho(r) \wedge d\psi = d\rho(r) \wedge \psi - \rho(r) \wedge \pi^* e.$$

By our choice of $\rho(r)$, it is easy to see $\Phi \in \Omega_{cv}(E)$ and it is closed. Its restriction to the fiber at p is

$$\iota_p^*(d(\rho(r) \wedge \psi)) = d(\iota_p^* \rho(r) \wedge \iota_p^* \psi) = (d\rho(r)) \wedge \iota_p^* \psi - \rho(r) \wedge \iota_p^* \pi^* e,$$

where $i_p : E_p \to E$ is the inclusion and ι_p^* gives a generator of $H^{k-1}(\mathbb{R}^k \setminus \{0\}) = H^{k-1}(S^{k-1})$. From the diagram

$$\begin{tikzcd} E_p \arrow[r] \arrow[d] & \{p\} \arrow[d] \\ E \arrow[r, "\pi"] & M \end{tikzcd}$$

we see that $\iota_p^* \pi^* e = 0$, therefore

$$\int_{\mathbb{R}^k} \iota_p^* \Phi = \int_{\mathbb{R}^k} (d\rho(r)) \wedge \iota_p^* \psi = \int_{\mathbb{R}} d\rho(r) \int_{S^{k-1}} \iota_p^* \psi = 1.$$

Thus by Proposition Thom iso 3.1.56 Φ is the Thom class of E. ☐

We now have our promised fact about the Thom class and the Euler class.

Euler class **Proposition 3.2.39.** *The pullback of the Thom class of an oriented rank k vector undle via the zero section to the base manifold is the Euler class.*

Proof. Let s be the zero section of E. Use the formula of Φ, we find that

$$s^* \Phi = s^* d(\rho(r) \wedge \psi) = s^*((d\rho(r)) \wedge \psi - \rho(r) \pi^* e) = s^* d\rho(r) \wedge s^* \psi - s^* \rho(r) s^* \pi^* e = e.$$

Thus we get the claim. ☐

Remark 3.2.3. From the formula for the Thom class, it is clear that by making the support of $\rho(r)$ sufficiently close to 0, the Thom class Φ can be made to have support arbitrarily close to the zero section of the vector bundle.

Remark 3.2.4. In fact, any section will pull the Thom class back to the Euler class. Let s be a section of the oriented vector bundle E and $s^* : H^*_{cv}(E) \to H^*(M)$ the induced map in cohomology. Note that s^* can be written as the composition of the natural maps $i : H^*_{cv}(E) \to H^*(E)$ and $\bar{s}^* : H^*(E) \to H^*(M)$. As a map from M into E, the section s is homotopic to the zero section s_0. By the homotopy axiom for de Rham cohomology $\bar{s}^* = \bar{s}_0^*$. Hence, $s^* = s_0^*$.

Using the description of the Euler class as the pullback of the Thom class, it is easy to prove the Whitney product formula.

Proposition 3.2.40 (Whitney Product Formula for the Euler Class). *If E and F are two oriented vector bundles, then $e(E \oplus F) = e(E) \wedge e(F)$.*

Thom class disrect sum

Proof. By Proposition 3.1.58, the Thorn class of $E \oplus F$ is

$$\Phi(E \oplus F) = \pi_1^*\Phi(E) \wedge \pi_2^*\Phi(F).$$

Let s be the zero section of $E \oplus F$. Then $\pi_1 \circ s$ sand $\pi_2 \circ s$ are the zero sections of E and F.

Thom class pullback Euler class

By Proposition 3.2.39,

$$e(E \oplus F) = s^*\Phi(E \oplus F) = s^*\pi_1^*\Phi(E) \wedge s^*\pi_2^*\Phi(F) = e(E) \wedge e(F).$$

□

Euler class and the zero locus of a section

Let $\pi : E \to M$ be a vector bundle and S_0 the image of the zero section in E. A section s of E is **transversal** if its image $S = s(M)$ intersects S_0 transversally.

ie zero locus

Proposition 3.2.41. *Let $\pi : E \to M$ be any vector bundle and Z the zero locus of a transversal section. Then Z is a submanifold of M and its normal bundle in M is $NZ = E|_Z$.*

Proof. Write $S = s(M)$ for the image of the section s. Because S intersects S_0 transversally, $S \cap S_0$ is a submanifold of S by the transversality theorem. Under the diffeomorphism $s : M \to S$, Z is mapped homeomorphically to $S \cap S_0$. So Z can be made into a submanifold of M.

To compute the normal bundle of Z, we first note that because E is locally trivial, its tangent bundle on S_0 has the following canonical decomposition

$$TE|_{S_0} = E|_{S_0} \oplus TS_0.$$

By the transversality of $S \cap S_0$,

$$TS + TS_0 = TE = E \oplus TS_0.$$

Therefore we obtain

$$E = \frac{E \oplus TS_0}{TS_0} \cong \frac{TS + TS_0}{TS_0} \cong \frac{TS}{TS \cap TS_0}.$$

This then gives an exact sequence

$$0 \longrightarrow (TS \cap TS_0)|_Z \longrightarrow TS|_Z \longrightarrow E|_Z \longrightarrow 0$$

Since $(TS \cap TS_0)|_Z = TZ$, we find that $NZ \cong E|_Z$. □

In the proposition above, if E and M are both oriented, then the zero locus Z of a transversal section is naturally an oriented manifold, oriented in such a way that

$$E|_Z \oplus TZ = TM|_Z.$$

has the direct sum orientation.

Proposition 3.2.42. *Let $\pi : E \to M$ be an oriented vector bundle over an oriented manifold M. Then the Euler class $e(E)$ is Poincaré dual to the zero locus of a transversal section.*

Proof. We will identify M with the image S_0 of the zero section. If S is the image in E of the transversal section $s : M \to E$, then the zero locus of s is $Z = S \cap S_0$. Z is a closed oriented submanifold of M and by Proposition tran section normal bundle zero locus 3.2.41, its normal bundle in M is $NZ \cong E|_Z$. Since S is diffeomorphic to M, the normal bundle of Z in S is also isomorphic to $E|_Z$. The normal bundles of Z in M and S, denoted by $N_M Z$ and $N_S Z$, will be identified with the tubular neighborhoods of Z in M and in S, respectively.

Choose the Thom class Φ of E to have support so close to the zero section that Φ restricted to the tubular neighborhood $N_S Z$ in S has compact support in the vertical direction. We will now show that $s^*\Phi$ is the Thom class of the tubular neighborhood $N_M Z$ in M.

Let E_p, S_p, and M_p be the fibers of $E|_Z$, $N_S Z$, $N_M Z$ respectively above the point p in Z. Because Φ has compact support in S_p, $s^*\Phi$ has compact support in M_p. Furthermore, after identifying S_p and M_p with their conterpart in the tubular neighborhoods, we have the following diffeomorphisms:

$$\begin{tikzcd} M_p \arrow[r, "s"] \arrow[dr, "\beta"'] & S_p \arrow[d, "\alpha"] \\ & E_p \end{tikzcd}$$

Therefore we get

$$\int_{M_p} s^*\Phi = \int_{M_p} s^*\alpha^*\Phi = \int_{S_p} \alpha^*\Phi = \int_{E_p} \Phi = 1.$$

So $s^*\Phi$ is the Thom class of $N_M Z$. By Proposition Thom class pullback Euler class 3.2.39, $s^*\Phi = e(E)$. Since by Proposition Thom class Poincare dual 3.1.60 the Thom class of $N_M Z$ is Poincare dual to Z in M, the Euler class $e(E)$ is Poincare dual to Z in M. □

Poincaré duality

Let M be a manifold of dimension n and $\mathcal{U} = \{U_i\}$ any open cover of M. Define the coboundary operator

$$\delta : \bigoplus \Omega_c^*(U_{i_0 \ldots i_p}) \to \bigoplus \Omega_c^*(U_{i_0 \ldots i_{p-1}})$$

by the formula

$$(\delta\omega)_{i_0 \ldots i_{p-1}} = \sum_i \omega_{i, i_0 \ldots i_{p-1}}.$$

where on the right-hand side we mean the extension by zero of $\omega_{i,i_0 \ldots i_{p-1}}$ to a form on $U_{i_0 \ldots i_{p-1}}$. To ensure that each component of $\delta\omega$ has compact support, the groups here are direct sums rather than direct products, so that ω by definition has only a finite number of nonzero components.

compact supp **Proposition 3.2.43 (Generalized Mayer-Vietoris Sequence for Compact Supports).** *Suppose the open cover $\mathcal{U}$ of the manifold M is locally finite. Then the sequence*

$$\cdots \longrightarrow \bigoplus \Omega_c^*(U_{i_0 i_1}) \longrightarrow \bigoplus \Omega_c^*(U_{i_0}) \xrightarrow{\text{sum}} \Omega_c^*(M) \longrightarrow 0$$

is exact.

Proof. We first show $\delta^2 = 0$. Let ω be in $\bigoplus \Omega_c^*(U_{i_0\dots i_p})$. Then

$$(\delta^2\omega)_{i_0\dots i_{p-2}} = \sum_i (\delta\omega)_{i,i_0\dots i_{p-2}} = \sum_i \sum_j \omega_{j,i,i_0\dots i_{p-2}} = 0.$$

(Here we use the alternating condition $\omega_{i_0\dots i\dots j\dots i_p} = -\omega_{i_0\dots j\dots i\dots i_p}$.) Now we define a homotopy from the identity to zero. Let $\{\rho_i\}$ be a partition of unity subordinate to the cover $\{U_i\}$. Define a map

$$K : \Omega_c^*(U_{i_0\dots i_p}) \to \Omega_c^*(U_{i_0\dots i_{p+1}})$$

by the formula

$$(K\omega)_{i_0\dots i_{p+1}} = \sum_{j=0}^{p+1} (-1)^i \rho_{i_j} \omega_{i_0\dots \widehat{i_j}\dots i_{p+1}}.$$

Note that $(K\omega)_{i_0\dots i_{p+1}}$ has compact support. Moreover, there are only finitely many $(j, i_0, \dots, i_p)$ for which $\rho_j \omega_{i_0\dots i_p} \neq 0$, since $\omega_{i_0\dots i_p} \neq 0$ for finitely many $(i_0, \dots, i_p)$ and by the locally finiteness each $U_{i_0\dots i_p} \subseteq U_{i_0}$ intersects only finitely many U_j. With this definition, we have

$$\begin{aligned}
(\delta K\omega)_{i_0\dots i_p} &= \sum_i (K\omega)_{i,i_0\dots i_p} = \sum_i \rho_i \omega_{i_0\dots i_p} + \sum_i \sum_{j=0}^{p} (-1)^{j+1} \rho_{i_j} \omega_{i,i_0\dots \widehat{i_j}\dots i_p} \\
&= \omega_{i_0\dots i_p} - \sum_{j=0}^{p} (-1)^j \rho_{i_j} \sum_i \omega_{i,i_0\dots \widehat{i_j}\dots i_p} \\
&= \omega_{i_0\dots i_p} - \sum_{j=0}^{p} (-1)^j \rho_{i_j} \sum_i \omega_{i,i_0\dots \widehat{i_j}\dots i_p} \\
&= \omega_{i_0\dots i_p} - \sum_{j=0}^{p} (-1)^j \rho_{i_j} (\delta\omega)_{i_0\dots \widehat{i_j}\dots i_p} \\
&= \omega_{i_0\dots i_p} - (K\delta\omega)_{i_0\dots i_p}.
\end{aligned}$$

Therefore $\delta K + K\delta = \mathrm{id}$ and we conclude the claim. □

Consider the double complex $C^*(\mathcal{U}, \Omega_c^*)$, where $\mathcal{U}$ is a locally finite cover. In this double complex the δ-operator goes in the wrong direction, so we define a new complex

$$K^{-p,q} = C^p(\mathcal{U}, \Omega_c^q).$$

Generalized MV compact supp

Then by Proposition 3.2.43 and a spectral sequence argument, we get

$$H_{TC}^*(K) = H_c^*(M).$$

On the other hand, by definition we have

$${}_vE_1^{-p,q} = C^p(\mathcal{U}, \mathscr{H}_c^q)$$

where $\mathscr{H}_c^q$ is the covariant functor which associates to every open set U the compact cohomology $H_c^q(U)$ and to every inclusion i, the extension by zero i_*; moreover, is $\mathcal{U}$ is a good cover, then

$$\mathscr{H}_c^q(U) = \begin{cases} \mathbb{R}, & q = n; \\ 0, & q \neq n. \end{cases}$$

Therefore in this case the ${}_vE_1$ page has only one row, and therefore

$$H^*_{TC}(K) = {}_vE_2^{*-n,n} = H_{n-*}(\mathcal{U}, \mathscr{H}^n_c).$$

Here $H(\mathcal{U}, \mathscr{H}^n_c)$ is the tech homology of the cover $\mathcal{U}$ with coefficients in the covariant functor $\mathscr{H}^n_c$.

Proposition 3.2.44. *Let M be a manifold of dimension n and $\mathcal{U}$ any locally finite good cover of M. Here M is not assumed to be orientable. Then*

$$H^*_c(M) \cong H_{n-*}(\mathcal{U}, \mathscr{H}^n_c),$$

where $\mathscr{H}^n_c$ is the covariant functor $\mathscr{H}^n_c(U) = H^n_c(U)$.

3.2.4 Monodromy

When is a locally constant presheaf constant

In the preceding part we saw that the compact vertical cohomology $H^*_{cv}(E)$ of a vector bundle E may be computed as the cohomology of the base with coefficients in the presheaf $\mathscr{H}^n_{cv}$. When the presheaf $\mathscr{H}^n_{cv}$ is the constant presheaf $\mathbb{R}^n$, $H^*_{cv}(E)$ is expressible in terms of the de Rham cohomology of the base manifold. In this case the problem of computing $H^*_{cv}(E)$ is greatly simplified. It is therefore important to determine the conditions under which a presheaf such as $\mathscr{H}_{cv}$ is constant.

We first introduce the notion of a presheaf on a cover. Let X be a topological space and $\mathcal{U} = \{U_\alpha\}$ a cover of X. The presheaf $\mathscr{F}$ on $\mathcal{U}$ is defined to be a functor $\mathscr{F}$ on the subcategory of $\mathbf{Open}(X)$ consisting of all finite intersections $U_{i_0\dots i_p}$ of open sets in $\mathcal{U}$. Equivalently, if $N(\mathcal{U})$ is the nerve of $\mathcal{U}$, the presheaf $\mathscr{F}$ on $\mathcal{U}$ is the assignment of an appropriate group to the barycenter of each simplex in $N(\mathcal{U})$; for example, the group attached to the barycenter of the 2-simplex representing $U \cap V \cap W$ is $\mathscr{F}(U \cap V \cap W)$. Each inclusion, say $U \cap V \hookrightarrow U$, becomes an arrow in the picture, $\mathscr{F}(U) \to \mathscr{F}(U \cap V)$, and the transitivity of the arrows says that Figure is a commutative diagram.

Two presheaves $\mathscr{F}$ and $\mathscr{G}$ are isomorphic relative to a cover $\mathcal{U} = (U_i)$ if for each $U_{i_0\dots i_p}$ there is an isomorphism

$$h_{i_0\dots i_p} : \mathscr{F}(U_{i_0\dots i_p}) \to \mathscr{G}(U_{i_0\dots i_p})$$

compatible with all arrows. In other words, there is a natural equivalence of functors $\mathscr{F} \to \mathscr{G}$ where $\mathscr{F}$ and $\mathscr{G}$ are regarded as functors on the subcategory of $\mathbf{Open}(X)$ consisting of all finite intersections of open sets in $\mathcal{U}$. The constant presheaf with group G on a good cover $\mathcal{U}$ is defined as usual, it associates to every open set the group of locally constant and hence constant functions: $U_{i_0\dots i_p} \to G$. Thus, for a constant presheaf on a cover, all the groups are G and all the arrows are the identity map. We say that a presheaf $\mathscr{F}$ is **locally constant** on a cover $\mathcal{U}$ if all the groups are isomorphic and all the arrows are isomorphisms.

Of course, if two presheaves $\mathscr{F}$ and $\mathscr{G}$ are isomorphic on a cover $\mathcal{U}$, then the Čech cohomology groups $H^*(\mathcal{U}, \mathscr{F})$ and $H^*(\mathcal{U}, \mathscr{G})$ are isomorphic.

Let $\mathscr{F}$ be a locally constant presheaf with group G on a cover $\mathcal{U} = (U_i)$. Fix isomorphisms

$\phi_i : \mathscr{F}(U_i) \to G$. If U_i and U_j intersect, then from the diagram

$$\begin{tikzcd}
\mathscr{F}(U_i) \arrow[r, "\rho^i_{ij}"] \arrow[d, "\phi_i"] & \mathscr{F}(U_{ij}) & \mathscr{F}(U_j) \arrow[l, "\rho^j_{ij}"'] \arrow[d, "\phi_j"] \\
G \arrow[rr, dashed] & & G
\end{tikzcd}$$

we obtain an automorphism of G, namely $\phi_j \circ (\rho^j_{ij})^{-1} \circ \rho^i_{ij} \circ \phi_i^{-1}$. Write $\rho^i_j : \mathscr{F}(U_i) \to \mathscr{F}(U_j)$ for the isomorphism $(\rho^j_{ij})^{-1} \circ \rho^i_{ij}$. Choose some vertex U_0 as the base point of the nerve $N(\mathcal{U})$. For $U_0 U_1 \cdots U_r U_0$ a loop based at U_0 we get an automorphism of G by following along the edges

$$\begin{tikzcd}
\mathscr{F}(U_0) \arrow[r] \arrow[d, "\phi_0"] & \mathscr{F}(U_1) \arrow[r] \arrow[d, "\phi_1"] & \cdots \arrow[r] & \mathscr{F}(U_r) \arrow[r] \arrow[d, "\phi_r"] & \mathscr{F}(U_0) \arrow[d, "\phi_0"] \\
G \arrow[r] & G \arrow[r] & \cdots \arrow[r] & G \arrow[r] & G
\end{tikzcd}$$

This gives a map from loops at U_0 to $\mathrm{Aut}(G)$. By the transitivity of the transition map, we claim that if a loop bounds a 2-chain in $N(\mathcal{U})$, then the associated automorphism of G is the identity. Hence there is a Homomorphism

$$\rho : \pi_1(N(\mathcal{U})) = \frac{\text{loops}}{\text{bounding loops}} \to \mathrm{Aut}(G).$$

called the **monodromy representation** of the presheaf $\mathscr{F}$. Here $\pi_1(N(\mathcal{U}))$ denotes the edge path group of the simplicial complex $N(\mathcal{U})$.

Theorem 3.2.45. *Let $\mathcal{U}$ be a cover on a connected topological space X and $N(\mathcal{U})$ its nerve. If $\pi_1(N(\mathcal{U})) = 0$, then every locally constant presheaf on $\mathcal{U}$ is constant.*

Proof. For each open set U_i, choose a path from U_0 to U_i, say $U_0 U_{i_1} \cdots U_{i_r} U_i$ and define

$$\psi_i = \phi_0 \circ (\rho^{i_r}_i \circ \cdots \circ \rho^{i_1}_{i_2} \circ \rho^0_{i_1})^{-1} : \mathscr{F}(U_i) \to G.$$

ψ_i is well-defined independent of the chosen path, because ρ is trivial. Now carry out the barycentric subdivision of the nerve $N(\mathcal{U})$ to get a new simplictal complex K so that every open set $U_{i_0 \ldots i_p}$ corresponds to a vertex of K. Clearl $\pi_1(K) = \pi_1(N(\mathcal{U})) = 0$. By the same procedure as in the preceding paragraph we can define isomorphisms $\psi_{i_0 \ldots i_p} : \mathscr{F}(U_{i_0 \ldots i_p}) \to G$. The maps $\psi_{i_0 \ldots i_p}$ give an isomorphism of the presheaf $\mathscr{F}$ to the constant presheaf G on the cover $\mathcal{U}$. □

By a good cover on a topological space we shall mean an open cover for which all finite intersections are contractible.

Theorem 3.2.46. *Suppose the topological space X has a good cover $\mathcal{U}$. Then the fundamental group of X is isomorphic to the fundamental group $\pi_1(N(\mathcal{U}))$ of the nerve of the good cover.*

Corollary 3.2.47. *Let $\mathcal{U}$ be a good cover on a simply connected topological space X. Then any locally constant sheaf on $\mathcal{U}$ is constant.*

Corollary 3.2.48. *Any locally constant sheaf on a simply connected manifold is constant.*

3.2.5 The spectral sequence of a fiber bundle

Let $\pi : E \to M$ be a fiber bundle with fiber F over a manifold M. Applying spectral sequence gives a general method for computing the cohomology of E from that of F and M. Indeed, given a good cover $\mathcal{U}$ of M, $\pi^{-1}(\mathcal{U})$ is a cover on E and we can form the double complex $C^*(\pi^{-1}(\mathcal{U}), \Omega^*)$, whose E_1 term is

$$E_1^{p,q} = \prod H^q(\pi^{-1}(U_{i_0 \ldots i_p})) = C^p(\pi^{-1}(\mathcal{U}), \mathscr{H}^q).$$

where $\mathscr{H}^q$ is the presheaf $\mathscr{H}^q(U) = H^q(\pi^{-1}(U))$ on M. For emphasis we sometimes write the presheaf $\mathscr{H}^q$ as $\mathscr{H}^q(F)$. Since $\mathcal{U}$ is a good cover, $E|_{U_i}$ is trivial and thus $\mathscr{H}^q$ is a locally constant presheaf on $\mathcal{U}$ with group $H^q(F)$. Since $d_1 = \delta$ on E_1, the E_2 term is

$$E_2^{p,q} = H^p(\mathcal{U}, \mathscr{H}^q).$$

By the generalized Mayer-Vietoris principle $H^*_{TC}(C^*(\pi^{-1}(\mathcal{U}), \Omega^*))$ is equal to $H^*(E)$, because $\pi^{-1}(\mathcal{U})$ is a cover on E. In the case that $\mathscr{H}^q$ is constant and $H^q(F)$ is finite-dimensional, the E_2 term is isomorphic as a vector space to the tensor product $H^p(M) \otimes H^q(F)$, since

$$E_2^{p,q} = H^p(\mathcal{U}, \mathbb{R}^{\dim H^q(F)}) = H^p(\mathcal{U}, \mathbb{R}) \otimes H^q(F) = H^p(M) \otimes H^q(F).$$

Therefore we have proved the following.

Proposition 3.2.49 (Leray's Theorem for de Rham Cohomology). *Given a fiber bundle $\pi : E \to M$ with fiber F over a manifold M and a good cover $\mathcal{U}$ of M, there is a spectral sequence $\{E_r\}$ converging to the cohomology of the total space $H^*(E)$ with*

$$E_2^{p,q} = H^p(\mathcal{U}, \mathscr{H}^q).$$

where $\mathscr{H}^q$ is the locally constant presheaf $\mathscr{H}^q(U) = H^q(\pi^{-1}(U))$ on $\mathcal{U}$. If $\mathscr{H}^q$ is constant (for example, if M is simply connected) and $H^q(F)$ is finite-dimensional, then

$$E_2^{p,q} = H^p(M) \otimes H^q(F).$$

Example 3.2.50 (The Künneth formula and the Leray-Hirsch theorem). We now give a spectral sequence proof of the Künneth formula. Let M and F be two manifolds and $\mathcal{U}$ a good cover of M. Suppose F has finite-dimensional cohomology. By Leray's theorem, the spectral sequence of the trivial bundle $F \to M \times F \to M$ has E_2 term

$$E_2^{p,q} = H^p(\mathcal{U}, \mathscr{H}^q).$$

Because $M \times F$ is a trivial bundle over M, the presheaf $\mathscr{H}^q(F)$ is constant, so that

$$E_2^{p,q} = H^p(\mathcal{U}, \mathbb{R}) \otimes H^q(F) = H^p(M) \otimes H^q(F).$$

The differential d_r measures the extent to which an element of $C^*(\pi^{-1}(\mathcal{U}), \Omega^*)$ that lives to E_r fails to be extended one step further to a D-cocycle. Since every element of the E_2 term is already a global form on $M \times F$. So $E_2 = E_\infty$. Therefore we get the K"unneth formula

$$H^*(M \times F) = H^*(M) \otimes H^*(F).$$

The proof of the Leray-Hirsch theorem is analogous.

spectral seq

Example 3.2.51 (Orientability and the Euler class of a sphere bundle). Let $\pi : E \to M$ be an S^k-bundle over a manifold M and let $\mathcal{U}$ be a good cover of M. The spectral sequence of this fiber bundle has

$$E_1^{p,q} = C^p(\mathcal{U}, \mathscr{H}^q(S^k)) = \begin{cases} C^p(\mathcal{U}, \mathbb{R}), & q = 0, k; \\ 0, & \text{otherwise} \end{cases}.$$

Let σ be the element of $E_1^{0,k}$ corresponding to the local angular forms on the sphere bundle E. From the description of the differential d_r as the δ of the tail of a zig-zag, we see that E is orientable if and only if $d_1\sigma = 0$. For an orientable S^k-bundle then, such a σ lives to E_k:

$$E_k = E_2 = H^*(\mathcal{U}, \mathscr{H}^*(S^k)).$$

Up to a sign $d_k\sigma$ in $H^{k+1}(\mathcal{U}, \mathscr{H}^0(S^k)) = H^{k+1}(\mathcal{U}, \mathbb{R}) = H^{k+1}(M)$ is the Euler class of the sphere bundle. It measures the extent to which σ fails to be extended to a D-cocycle, i.e., a global closed k-form on the sphere bundle.

Example 3.2.52 (Orientability of a simply connected manifold). Let M be a simply connected manifold of dimension n and $S(TM)$ its unit tangent bundle. The spectral sequence of the fiber bundle

$$S^{k-1} \longrightarrow S(TM) \longrightarrow M$$

has E_2 term

$$E_2 = H^*(\mathcal{U}, \mathscr{H}^*(S^k)).$$

Since each restricted bundle on $U_{i_0 \ldots i_p}$ is trivial, $\mathscr{H}^0(S^k)$ and $\mathscr{H}^k(S^k)$ is contant on it. Therefore

$$E_2^{p,q} = H^p(\mathcal{U}, \mathscr{H}^q(S^{n-1})) = \begin{cases} H^p(\mathcal{U}, \mathbb{R}) & q = 0, n-1 \\ 0 & \text{otherwise} \end{cases}$$

Moreover, by the isomorphism $H^*(M) = H^*(\mathcal{U}, \mathbb{R})$ and the simply connected condition, we further have

$$E_2^{1,0} = E_2^{1,n-1} = 0.$$

This, together with Example 3.2.51, shows that there is an element in $C^0(\pi^{-1}(\mathcal{U}), \mathscr{H}^{n-1})$ which can be extended one step down toward being a D-cocycle. Therefore $S(TM)$ and also M are orientable. This gives an alternative proof of the orientability of a simply connected manifold.

Example 3.2.53 (The cohomology of the complex projective space). Consider the sphere S^{2n+1} in $\mathbb{C}^{n+1}$. Let S^1 act on S^{2n+1} by

$$\lambda \cdot (z_0, \ldots, z_n) = (\lambda z_0, \ldots, \lambda z_n).$$

where λ in S^1 is a complex number of absolute value 1. The quotient of S^{2n+1} by this action is the complex projective space $\mathbb{CP}^n$. This gives S^{2n+1} the structure of a circle bundle over $\mathbb{CP}^n$:

$$S^1 \longrightarrow S^{2n+1} \longrightarrow \mathbb{CP}^n$$

As we will see from the homotopy exact sequence to be discussed later, $\mathbb{CP}^n$ is simply connected. Thus

$$E_2^{p,q} = H^p(\mathbb{CP}^n) \otimes H^q(S^1).$$

So E_2 has only two nonzero rows, $q = 0, 1$, and the two rows are identical, both being $H^*(\mathbb{CP}^n)$. Because $\mathbb{CP}^n$ has dimension $2n$, we have $H^p(\mathbb{CP}^n) = 0$ for $p > 2n$. In the spectral seuqnece we have $d_r = 0$ for $r > 2$, therefore $E_3 = E_4 = \cdots = E_\infty = H^*(S^{2n+1})$. Therefore E_3 is

$$\begin{array}{ccccc} 0 & 0 & 0 & \cdots & \mathbb{R} \\ \\ \mathbb{R} & 0 & 0 & \cdots & 0 \end{array}$$

From this we can chase the maps in E_2. Since $E_2^{0,0} = E_2^{0,1} = \mathbb{R}$ and $E_2^{p,0} = E_2^{p,1}$, we finally conclude that E_2 has the form

$$\begin{tikzcd} \mathbb{R} \arrow[drr] & 0 & \mathbb{R} \arrow[drr] & 0 & \cdots & \mathbb{R} \arrow[drr] & 0 & \mathbb{R} \\ \mathbb{R} & 0 & \mathbb{R} & 0 & \cdots & \mathbb{R} & 0 & \mathbb{R} \end{tikzcd}$$

Further, from this spectral sequence we can get the product structure of $H^*(\mathbb{CP}^n)$:

$$\begin{tikzcd} a \arrow[dr] & ax \arrow[dr] & ax^2 & \cdots & ax^{n-1} \arrow[dr] & ax^n \\ 1 & x & x^2 & \cdots & x^{n-1} & x^n \end{tikzcd}$$

Therefore

$$H^*(\mathbb{CP}^n) = \mathbb{R}[x]/(x^{n+1}), \quad \dim x = 2.$$

Note that S^{2n-1} is a circle bundle on $\mathbb{CP}^n$, so we can compute its Euler class. By definition, the Euler class is given by the diagram

$$\begin{tikzcd} \sigma \arrow[r] & {} \\ & {} \arrow[u] \arrow[r, "\delta"] & d_2(\sigma) = -\pi^*\varepsilon \end{tikzcd}$$

where σ is a generator of $E_2^{0,1}$. By our choice of a and x, it is clear that $x = -e$, therefore the Euler class of S^{2n-1} generates the cohomology ring of $\mathbb{CP}^n$.

The Gysin sequence

The spectral sequence of a fiber bundle is essentially a way of describing the complicated algebraic relations among the cohomology of the base space, the fiber, and the total space of the bundle, In certain special situations the spectral sequence simplifies to a long exact sequence, One such special case is the cohomology of a sphere bundle. The resulting sequence is called the **Gysin sequence**, which we now derive. Let $\pi : E \to M$ be an oriented sphere bundle with fiber S^k. By the orientability assumption, for any good cover $\mathcal{U}$ on M, the locally constant presheaf $\mathscr{H}^k$ has no monodromy and is the constant presheaf $\mathbb{R}$. Therefore the E_2 term of the spectral sequence only has two nonzero rows:

$$E_2^{p,q} = H^p(M) \otimes H^q(S^k).$$

This then means $E_2 = E_3 = \cdots = E_{k+1}$, and we get an exact seuqnece:

$$\begin{tikzcd} 0 \arrow[r] & E_\infty^{p-k,k} \arrow[r] & E_2^{p-k,k} \arrow[r, "d_{k+1}"] \arrow[d, equal] & E_2^{p+1,0} \arrow[r] \arrow[d, equal] & E_\infty^{p+1,0} \arrow[r] & 0 \\ & & H^{p-k}(M) & H^{p+1}(M) & & \end{tikzcd} \tag{2.5}$$

Gysin seq-1

Because of the shape of the E_2 term, the filtration on $H^*(E)$ becomes

$$H^p(E) \supseteq E_\infty^{p,0} \supseteq 0, \quad H^p(E)/E^{p,0} \cong E_\infty^{p-k,k}.$$

In other words, there is an exact sequence

$$0 \longrightarrow E_\infty^{p,0} \longrightarrow H^p(E) \longrightarrow E_\infty^{p-k,k} \longrightarrow 0 \tag{2.6}$$

Gysin seq-2

The two sequences (2.5) and (2.6) may be combined into a single long exact sequence

$$\cdots \longrightarrow H^p(E) \xrightarrow{\alpha} H^{p-k}(M) \xrightarrow{d_{k+1}} H^{p+1}(M) \xrightarrow{\beta} H^{p+1}(E) \longrightarrow \cdots$$

This is the **Gysin sequence** of the sphere bundle.

It remains to identify the maps in the Gysin sequence. Let $\mathcal{U}$ be a good cover on M. The map α is the composition

$$H^p(E) \longrightarrow E_\infty^{p-k,k} \hookrightarrow E_2^{p-k,k} = H^{p-k}(\pi^{-1}(\mathcal{U}), \mathscr{H}^k) \cong H^{p-k}(M) \otimes H^k(S^k) \cong H^{p-k}(M)$$

In this sequence of maps the first three are the identity on the level of forms and the last one sends a generator of $H^k(S^k)$ to 1 by integration. Therefore α is integration along the fiber.

Next consider d_{k+1}. By representing an element of $E_2^{p-k,k} = H^{p-k}(M) \otimes H^k(S^k)$ by $(\pi^*\omega) \wedge (-\psi)$, where ω is a closed form on M and ψ is the global angular form on E (note that ψ is locally closed), the map d_{k+1} is defined to be

$$\begin{aligned} d_{k+1}(\omega) &:= d_{k+1}((\pi^*\omega) \wedge (-\psi)) \\ &= d((\pi^*\omega) \wedge (-\psi)) = d(\pi^*\omega) \wedge (-\psi) + (-1)^{p-k}(\pi^*\omega) \wedge d(-\psi) \\ &= (-1)^{pxs-k}\pi^*\omega \wedge \pi^* e = (-1)^{p-k}\pi^*(\omega \wedge e). \end{aligned}$$

Hence, up to a sign $d_{k+1} : H^{p-k}(M) \to H^{p+1}(M)$ is multiplication by the Euler class e.

Finally the map β is the composition

$$H^{p+1}(M) = H^{p+1}(\mathcal{U}, \mathscr{H}^0) \underset{\cong}{\xrightarrow{\pi^*}} H^{p+1}(\pi^{-1}(\mathcal{U}), \mathscr{H}^0) = E_2^{p+1,0} \longrightarrow E_\infty^{p+1,0} \hookrightarrow H^{p+1}(E)$$

So $\beta : H^{p+1}(M) \to H^{p+1}(E)$ is the natural pullback map π^*.

We summarize this discussion as follows.

Theorem 3.2.54. *Let $\pi : E \to M$ be an oriented sphere bundle with fiber S^k. Then there is a long exact sequence*

$$\cdots \longrightarrow H^p(E) \xrightarrow{\pi_*} H^{p-k}(M) \xrightarrow{\wedge e} H^{p+1}(M) \xrightarrow{\pi^*} H^{p+1}(E) \longrightarrow \cdots$$

where e is the Euler class of E.

Leray's construction

We consider now more generally not a fiber bundle but any map $f : X \to Y$ from one manifold to another, and study how the cohomology groups of X relate to those of Y. Let $\mathcal{U}$ be any cover for Y, not necessarily a good cover. Then $\pi^{-1}(\mathcal{U})$ is a cover for X. By the Mayer-Vietoris principle

$$H^*(X) = H^*_{TC}(C^*(\pi^{-1}(\mathcal{U}), \Omega^*).$$

The spectral sequence of this double complex has

$$E_2^{p,q} = H^p(\mathcal{U}, \mathscr{H}^q),$$

where $\mathscr{H}^q$ is the presheaf on Y defined by $\mathscr{H}^q(U) = H^q(f^{-1}(U))$. The main difference between this situation and that of a fiber bundle is that the presheaf $\mathscr{H}^q$ is no longer locally constant on U; indeed the groups $H^q(f^{-1}(U))$ will in general be different for different contractible open sets U.

Example 3.2.55. Let $f : X \to Y$ be any smooth map between manifolds and $\mathcal{U}$ a finite good cover of Y. Then from the spectral sequence we have

$$\begin{aligned}\dim H^n(X) &= \sum_{p+q=n} \dim E_\infty^{p,q} = \cdots = \sum_{p+q=n} \dim E_1^{p,q} = \sum_{p+q=n} \dim C^p(\mathcal{U}, \mathscr{H}^q) \\ &= \sum_{p+q=n} \sum_{i_0\dots i_p} \dim H^q(f^{-1}(U_{i_0\dots i_p})).\end{aligned}$$

This then implies

$$\chi(X) = \sum_n (-1)^n \sum_{p+q=n} \sum_{i_0\dots i_p} \dim H^q(f^{-1}(U_{i_0\dots i_p})).$$

If in particular $\pi : X \to Y$ is a fiber bundle with fiber F, Y admits a finite good cover and F has finite-dimensional cohomology, then

$$\begin{aligned}\chi(X) &= \sum_n (-1)^n \sum_{p+q=n} \sum_{i_0\dots i_p} \dim H^q(\pi^{-1}(U_{i_0\dots i_p})) \\ &= \sum_n (-1)^n \sum_{p+q=n} \sum_{U_{i_0\cdots i_p}\neq\varnothing} \dim H^q \\ &= \sum_{p+q=n} (-1)^n \dim C^p(\mathcal{U}, Y) \cdot \dim H^q(F) \\ &= \chi(Y)\chi(F).\end{aligned}$$

3.2.6 Spectral sequence for singular cohomology

The Mayer-Vietoris sequence for singular chains

Let $U = \{U_i\}$ be an open cover of the topological space X. Just as for differential forms on a manifold, the sequence of inclusions

$$X \longleftarrow \textstyle\coprod_i U_i \longleftarrow \coprod_{i,j} U_{ij} \longleftarrow \cdots$$

induces a Mayer-Vietoris sequence. However, we must consider here the group $C^{\mathcal{U}}(X)$ of $\mathcal{U}$-small chains in X; these are chains made up of simplices each of which lies in some open set of the cover $\mathcal{U}$. The inclusion

$$i : C_*^{\mathcal{U}}(X) \to C_*(X)$$

is a homotopy equivalence, hence induces an isomorphism on homology groups.

Now we define a boundary operator

$$\delta : \bigoplus C_q(U_{i_0 \dots i_p}) \to \bigoplus C_q(U_{i_0 \dots i_{p-1}})$$

by the alternating sum formula

$$(\delta c)_{i_0 \dots i_p} = \sum_i c_{i, i_0 \dots i_{p-1}}.$$

Here, as always, we adopt the alternating convention for chains. The fact that $\delta^2 = 0$ is proved as in Proposition Generalized MV compact supp 3.2.43. The boundary operator δ on $\bigoplus C_q(U_i) \to C_q(X)$ is simply the sum; we denote this by ε.

Proposition 3.2.56. *The following sequence is exact*

$$0 \longleftarrow C_*^{\mathcal{U}} \longleftarrow \bigoplus C_*(U_i) \longleftarrow \bigoplus C_*(U_{ij}) \longleftarrow \cdots$$

Applying the functor $\mathrm{Hom}(-, \mathbb{Z})$ to the Mayer-Vietoris sequence for singular chains we obtain the Mayer-Vietoris sequence for singular cochains

$$0 \longrightarrow C^*_{\mathcal{U}}(X) \longrightarrow \prod C^*(U_i) \longrightarrow \prod C^*(U_{i_0 i_1}) \longrightarrow \cdots$$

Since the functor $\mathrm{Hom}(-, \mathbb{Z})$ preserves the exactness of a sequence of free $\mathbb{Z}$-modules, the Mayer-Vietoris sequence for singular cochains is exact.

Once we have the Mayer-Vietoris sequence we can set up the double complex $C^*(\mathcal{U}, C^*)$. Just as in the de Rham theory the double complex computes the singular cohomology of X:

$$H^*_{TC}(C^*(\mathcal{U}, C^*)) = H^*(X).$$

If $\mathcal{U}$ is a good cover of a topological space X, then by the same argument as in the Čech-de Rham case, we get

$$H^*_{TC}(C^*(\mathcal{U}, C^*)) = H^*(\mathcal{U}, \mathbb{Z}).$$

Suppose X triangularizable. Since the good covers are cofinal in the set of all covers of X, we have

$$H^*(X, \mathbb{Z}) = H^*(\mathcal{U}, \mathbb{Z}),$$

where $H^*(X, \mathbb{Z})$ is the Čech cohomology of X with coefficients in the constant presheaf $\mathbb{Z}$. Therefore,

Theorem 3.2.57. *The singular cohomology of a triangularizable space X is isomorphic to its Čech cohomology with coefficients in the constant presheaf $\mathbb{Z}$. If $\mathcal{U}$ is a good cover of X, then*

$$H^*(X) \cong H^*(X, \mathbb{Z}) \cong H * (\mathcal{U}, \mathbb{Z}).$$

Let $\pi : E \to X$ be a fiber bundle with fiber F over a triangularizable topological space X. Just as in Theorem 14.18, from the double complex $C^*(\pi^{-1}(\mathcal{U}), C^*)$ on E we obtain a spectral sequence converging to the singular cohomology $H^*(E)$ whose E_2 term is

$$E_2^{p,q} = H^p(\mathcal{U}, \mathscr{H}^q(F)).$$

If $\mathscr{H}^q(F)$ happens to be the constant presheaf $\mathbb{Z}^n$ on $\mathcal{U}$, then

$$E_2^{p,q} = H^p(\mathcal{U}, \mathbb{Z}^n) = H^p(X) \otimes H^q(F).$$

By a similar arguments there is also a Mayer-Vietoris sequence for singular cochains with coefficients in a commutative ring R. Using the cup product in place of the wedge product, the spectral sequence of the Cech-singular complex $C^*(\mathcal{U}, C^*)$ can be given a product structure. The arguments before carry over mutatis mutandis. Hence the results on spectral sequences remain true for singular cohomology with coefficients in R. However, note that the E_2 term of a fiber bundle $\pi : E \to X$ with fiber F over a simply connected base space M is the tensor product $H^*(X; \mathbb{R}) \otimes H^*(F; R)$ only if the cohomology of F is a free R-module. In summary we have the following.

Theorem 3.2.58 (Leray's Theorem for Singular Cohomology with Coefficients $\boldsymbol{R}$). *Let $\pi : E \to X$ be a fiber bundle with fiber F over a topological space X and $\mathcal{U}$ an open cover of X. Then there is a spectral sequence converging to $H^*(E; R)$ with E_2 term*

$$E_2^{p,q} = H^p(\mathcal{U}, \mathscr{H}^q(F; R)).$$

Each E_r in the spectral sequence can be given a product structure relative to which the differential d_r is an antiderivation. If X is simply connected and has a good cover, then

$$E_2^{p,q} = H^p(X, H^q(F; R)).$$

Ifin addition $H^*(F; R)$ is a finitely generated free R-module, then

$$E_2 = H^*(X; R) \otimes H^*(F; R)$$

as R-algebras.

3.3 Characteristic classes

3.3.1 Chern classes of a complex vector bundle

In this part we will study the characteristic classes of a complex vector bundle. To begin with we define the first Chern class of a complex line bundle as the Euler class of its underlying real bundle. Applying the Leray-Hirsch theorem, we then compute the cohomology ring of the projectivization $P(E)$ of a complex vector bundle E and define the Chern classes of E in terms of the ring structure of $H^*(P(E))$. We conclude with a list of the main properties of the Chern classes.

The first Chern class of a complex line bundle

Recall that a complex vector bundle of rank n is a fiber bundle with fiber $\mathbb{C}^n$ and structure group $\mathrm{GL}(n, \mathbb{C})$. A complex vector bundle of rank 1 is also called a complex line bundle. Just as the structure group of a real vector bundle can be reduced to the orthogonal group $\mathrm{O}(n)$, by the Hermitian analogue the structure group of a rank n complex vector bundle can be reduced to the unitary group $\mathrm{U}(n)$. Every complex vector bundle E of rank n has an underlying real vector bundle $E_{\mathbb{R}}$ of rank $2n$, obtained by discarding the complex structure on each fiber. By the isomorphism of $\mathrm{U}(1)$ with $\mathrm{SO}(2)$, this sets up a one-to-one correspondence between the complex line bundles and the oriented rank 2 real bundles. We define the first Chern class of a complex line bundle L over a manifold M to be the Euler class of its underlying real bundle $L_{\mathbb{R}}$: $c_1(L) = e(L_{\mathbb{R}}) \in H^2(M)$.

If $L$and L' are complex line bundles with transition functions $\{\tau_{\alpha\beta}\}$ and $\{\tau'_{\alpha\beta}\}$,

$$\tau_{\alpha\beta}, \tau'_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathbb{C}^*.$$

then their tensor product $L \otimes L'$ is the complex line bundle with transition functions $\{\tau_{\alpha\beta} \cdot \tau'_{\alpha\beta}\}$. By the formula (2.4) which gives the Euler class in terms of the transition functions, we have

$$c_1(L \otimes L') = c_1(L) + c_1(L'). \tag{3.1}$$

Let L^* be the dual of L. Since the line bundle $L \otimes L^* = \mathrm{Hom}(L, L)$ has a nowhere vanishing section given by the identity map, $L \otimes L^*$ is a trivial bundle. By (3.1), $c_1(L) + c_1(L^*) = c_1(L \otimes L^*) = 0$. Therefore, $c_1(L^*) = -c_1(L)$.

Example 3.3.1 (Tautological bundles on a projective space). Let V be a complex vector space of dimension n and $\mathbb{P}(V)$ its projectivization:

$$\mathbb{P}(V) = \{\text{1-dimensional subspaces of } V\}.$$

On $\mathbb{P}(V)$ there are several God-given vector bundles: the product bundle $\widehat{V} = \mathbb{P}(V) \times V$, the universal subbundle S, which is the subbundle of V defined by

$$S = \{(\ell, v) \in \mathbb{P}(V) \times V : v \in \ell\}.$$

and the **universal quotient bundle** Q, defined by the exact sequence

$$0 \longrightarrow S \longrightarrow \widehat{V} \longrightarrow Q \longrightarrow 0$$

The fiber of S above each point ℓ in $\mathbb{P}(V)$ consists of all the points in ℓ, where ℓ is viewed as a line in the vector space V. The sequence above is called the **tautological exact sequence** over $\mathbb{P}(V)$, and S^* the **hyperplane bundle**.

We now compute the cohomology of $\mathbb{P}(V)$. Endow V with a Hermitian metric and let E be the unit sphere bundle of the universal subbundle S:

$$E = \{(\ell, v) : v \in \ell, \|v\| = 1\}.$$

Since the complement of the zero section S^0 is diffeomorphic to $V \setminus \{0\}$, we see the map $\pi : E \to \mathbb{P}(V)$ gives a fibering

$$S^1 \to S^{2n-1} \to \mathbb{P}(V).$$

By a computation similar to Example 3.2.53, the cohomology ring $H^*(\mathbb{P}(V))$ is seen to be generated by the Euler class of the circle bundle E, i.e., the first Chern class of the universal subbundle S. It is customary to take $x = c_1(S^*) = -c_1(S)$ to be the generator and write

$$H^*(\mathbb{P}(V)) = \mathbb{R}[x]/(x^n), \quad n = \dim_{\mathbb{C}} V.$$

We then see that the Poincaré series of the projective space $\mathbb{P}(V)$ is

$$P_t(\mathbb{P}(V)) = 1 + t^2 + \cdots + t^{2(n-1)} = \frac{1 - t^{2n}}{1 - t^2}.$$

The projectivization of a vector bundle

Let $\rho : E \to M$ be a complex vector bundle with transition functions $\tau_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(n, \mathbb{C})$. We write E_p for the fiber over p and $\mathrm{PGL}(n, \mathbb{C})$ for the projective general linear group $\mathrm{GL}(n, \mathbb{C})/\{\text{scalar matrices}\}$. The projectivization of E, $\pi : \mathbb{P}(E) \to M$, is by definition the fiber bundle whose fiber at a point p in M is the projective space $\mathbb{P}(E_p)$ and whose transition functions $\tau_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{PGL}(n, \mathbb{C})$ are induced from $\tau_{\alpha\beta}$. Thus a point of $\mathbb{P}(E)$ is a line ℓ_p in the fiber E_p.

As on the projectivization of a vector space, on $\mathbb{P}(E)$ there are several tautological bundles: the pullback π^*E, the universal subbundle S, and the universal quotient bundle Q.

$$\begin{tikzcd}
0 \arrow[r] & S \arrow[r] & \pi^*E \arrow[r] \arrow[d] & Q \arrow[r] & 0 & E \arrow[d, "\rho"] \\
 & & \mathbb{P}(E) \arrow[rrr, "\pi"] & & & M
\end{tikzcd}$$

The pullback bundle π^*E is the vector bundle over $\mathbb{P}(E)$ whose fiber at ℓ_p is E_p. When restricted to the fiber $\pi^{-1}(p)$ it becomes the trivial bundle,

$$\pi^*E|_{\pi^{-1}(p)} = \mathbb{P}(E)_p \times E_p.$$

since $\rho : E_p \to \{p\}$ is a trivial bundle. The universal subbundle S over $\mathbb{P}(E)$ is defined by

$$S = \{(\ell_p, v) \in \pi^*E : v \in \ell_p\}.$$

Its fiber at ℓ_p consists of all the points in ℓ_p. The universal quotient bundle Q is determined by the tautological exact sequence

$$0 \longrightarrow S \longrightarrow \pi^*E \longrightarrow Q \longrightarrow 0$$

Set $x = c_1(S^*)$, then x is a cohomology class in $H^2(\mathbb{P}(E))$. Since the restriction of the universal subbundle S on $\mathbb{P}(E)$ to a fiber $\mathbb{P}(E_p)$ is the universal subbundle S of the projective space $\mathbb{P}(E)$, by the naturality property of the first Chern class, it follows that $c_1(S)$ is the restriction of $-x$ to $\mathbb{P}(E_p)$. Hence the cohomology classes $1, x, \ldots, x^{n-1}$ are global classes on $\mathbb{P}(E)$ whose restrictions to each fiber $\mathbb{P}(E_p)$ freely generate the cohomology of the fiber. By the Leray-Hirsch theorem the cohomology $H^*(\mathbb{P}(E))$ is a free module over $H^*(M)$ with basis $\{1, x, \ldots, x^{n-1}\}$. So x^n can be written uniquely as a linear combination of $1, x, \ldots, x^{n-1}$ with coefficients in $H^*(M)$; these coefficients are by definition the **Chern classes** of the complex vector bundle E:

$$x^n + c_1(E)x^{n-1} + \cdots + c_n(E) = 0, \quad c_i(E) \in H^{2i}(M). \tag{3.2}$$

Chern class c

In this equation by $c_i(E)$ we really mean $\pi^*c_i(E)$. We call $c_i(E)$ the i-th Chern class of E and

$$c(E) = 1 + c_1(E) + \cdots + c_n(E) \in H^*(M)$$

its **total Chern class**. With this definition of the Chern classes, we see that the ring structure of the cohomology of $\mathbb{P}(E)$ is given by

$$H^*(\mathbb{P}(E)) = H^*(M)[x]/(x^n + c_1(E)x^{n-1} + \ldots c_n(E)),$$

where $x = c_1(S^*)$ and n is the rank of E. Since additively

$$H^*(\mathbb{P}(E)) = H^*(M) \otimes H^*(\mathbb{CP}^{n-1})$$

the Poincaré series of $\mathbb{P}(E)$ is

$$P_t(\mathbb{P}(E)) = P_t(M)\frac{1 - t^{2n}}{1 - t^2}.$$

We now have two definitions of the first Chern class of a line bundle L: as the Euler class of $L_{\mathbb{R}}$, and as a coefficient in (3.2). To check that these two definitions agree we will temporarily reserve the notation c_1 for the second definition. What must be shown is that $e(L_{\mathbb{R}}) = c_1(L)$.

For a line bundle L, $\mathbb{P}(L) = M$, $\pi^*L = L$ and the universal subbundle S on $\mathbb{P}(L)$ is L itself. Therefore, $x = e(S^*) = -e(S_{\mathbb{R}}) = -e(L_{\mathbb{R}})$. So the relation (3.2) is $x + e(L_{\mathbb{R}}) = 0$, which proves that $c_1(L) = e(L_{\mathbb{R}})$.

If E is the trivial bundle $M \times V$ over M, then $\mathbb{P}(E) = M \times \mathbb{P}(V)$, so $x^n = 0$. Hence all the Chern classes of a trivial bundle are zero. In this sense the Chern classes measure the twisting of a complex vector bundle. There is an alternate description of the ring structure $H^*(\mathbb{P}(E))$ which is sometimes very useful. We write $H^*(M)[c(S), c(Q)]$ for $H^*(M)[c_1(S), c_1(Q), \ldots, c_{n-1}(Q)]$, where S and Q are the universal subbundle and quotient bundle on $\mathbb{P}(E)$.

$$\begin{tikzcd}
0 \arrow[r] & S \arrow[r] & \pi^*E \arrow[r] \arrow[d] & Q \arrow[r] & 0 & E \arrow[d, "\rho"] \\
 & & \mathbb{P}(E) \arrow[rrr, "\pi"] & & & M
\end{tikzcd}$$

Proposition 3.3.2. $H^*(\mathbb{P}(E)) = H^*(M)[c(S), c(Q)]/(c(S)c(Q) = \pi^*c(E))$.

Proof. The idea is to eliminate the generators $c_1(Q), \ldots, c_{n-1}(Q)$ by using the relation $c(S)c(Q) = \pi^*(E)$. Let $x = c_1(S^*)$, $y_i = c_i(Q)$ and $c_i = \pi^*c_i(E)$. Equating the terms of equal degrees in

$$(1 - x)(1 + y_1 + \cdots + y_n)$$

we get

$$\begin{aligned}
y_1 - x &= c_1, \\
y_2 - xy_1 &= c_2, \\
&\vdots \\
y_{n-1} - xy_{n-2} &= c_{n-1}, \\
-xy_{n-1} &= c_n.
\end{aligned}$$

By the first $n-1$ equations, $y_1, \ldots, y_{n-1}$ can be expressed in terms of x and elements of $H^*(M)$, and so can be eliminated as generators of $H^*(M)[c(S), c(Q)]/(c(S)c(Q) = \pi^*c(E))$. The last equation $-xy_{n-1} = c_n$ translates into

$$x^n + c_1 x^{n-1} + \cdots + x_n = 0.$$

Hence $H^*(M)[c(S), c(Q)]/(c(S)c(Q) = \pi^*c(E))$ is isomorphic to the polynomial ring over $H^*(M)$ with the single generator x and the single relation above. $\square$

Main properties of the Chern classes

Now we collect together some basic properties of the Chern classes.

Proposition 3.3.3. *If f is a map from N to M and E is a complex vector bundle over M, then $c(f^*E) = f^*c(E)$.*

Proof. Basically this property follows from the functoriality of all the constructions in the definition of the Chern class. To be precise, the first Chern class of a line bundle is functorial.

Write S_E for the universal subbundle over $\mathbb{P}(E)$. Now $f^*\mathbb{P}(E) = \mathbb{P}(f^*E)$ and $f^*(S_E^*) = S_{f^*E}^*$ so if $x_E = c_1(S_E^*)$, then

$$x_{f^*E} = c_1(S_{f^*E}^*) = c_1(f^*S_E^*) = f^*c_1(E) = f^*x_E.$$

Applying f^* to the relation (Chern class def 3.2), we get

$$x_{f^*E}^n + f^*c_1(E)x_{f^*E}^{n-1} + \cdots + f^*c_n(E) = 0.$$

Hence $c_i(f^*E) = f^*c_i(E)$. □

It follows from the naturality of the Chern class that if E and F are isomorphic vector bundles over X, then $c(E) = c(F)$.

Proposition 3.3.4. *Let V be a complex vector space. If S^* is the hyperplane bundle over $\mathbb{P}(V)$, then $c_1(S^*)$ generates the algebra $H^*(\mathbb{P}(V))$.*

Proposition 3.3.5. *If E has rank n as a complex vector bundle, then $c_1(E) = 0$ for $i > n$.*

Proposition 3.3.6. *If E has a nonvanishing section, then the top Chern class $c_n(E)$ is zero.*

Proof. Such a section s induces a section $\widetilde{s}$ of $\mathbb{P}(E)$ as follows. At a point p in M, the value of $\widetilde{s}$ is the line in E_p through the origin and $s(p)$.

$$\begin{tikzcd} \mathbb{P}(E) \arrow[d, shift left, "\pi"] \\ M \arrow[u, shift left, "\widetilde{s}"] \end{tikzcd}$$

Then $\widetilde{s}^*S_E$ is a line bundle over M whose fiber at p is the line in E_p spanned by $s(p)$. Since every line bundle with a nonvanishing section is isomorphic to the trivial bundle, we conclude that $\widetilde{s}^*S_E$ is trivial. It follows from the naturality of the Chern class that

$$\widetilde{s}^*c_1(S_E) = 0$$

which implies that $\widetilde{s}^*x = 0$. Applying $\widetilde{s}^*$ to the relation (Chern class def 3.2), we get

$$\widetilde{s}^*c_n = 0.$$

By our abuse of notation this really means $\widetilde{s}^*\pi^*c_n = 0$. Therefore $c_n = 0$. □

3.3.2 The splitting principle and flag manifolds

In this part we prove the Whitney product formula and compute a few Chern classes. The proof and the computations are based on the splitting principle, which, roughly speaking, states that if a polynomial identity in the Chern classes holds for direct sums of line bundles, then it holds for general vector bundles. In the course of establishing the splitting principle we introduce the flag manifolds. We conclude by computing the cohomology ring of a flag manifold.

The splitting principle

Let $\rho : E \to M$ be a smooth complex vector bundle of rank n over a manifold M. Our goal is to construct a space $F(E)$ and a map $\sigma : F(E) \to M$ such that:

- the pullback of E to $F(E)$ splits into a direct sum of line bundles: $\sigma^*E = L_1 \oplus \cdots \oplus L_n$.
- σ^* embeds $H^*(M)$ in $H^*(F(E))$.

Such a space $F(E)$, which is in fact a manifold by construction, is called a **split manifold** of E.

If E has rank 1, there is nothing to prove. If E has rank 2, we can take as a split manifold $F(E)$ the projective bundle $\mathbb{P}(E)$, for on $\mathbb{P}(E)$ there is the exact sequence

$$0 \longrightarrow S_E \longrightarrow \pi^*E \longrightarrow Q_E \longrightarrow 0$$

and therefore $\pi^*E = S_E \oplus Q_E$, which is a direct sum of line bundles.

Now suppose E has rank 3. Over $\mathbb{P}(E)$ the line bundle S_E splits off as before. The quotient bundle Q_E over $\mathbb{P}(E)$ has rank 2 and so can be split into a direct sum of line bundles when pulled back to $\mathbb{P}(Q_E)$.

$$\begin{tikzcd}
E \arrow[d] & S_E \oplus Q_E \arrow[d] & \beta^*S_E \oplus S_{Q_E} \oplus Q_{Q_E} \arrow[d] \\
M & \mathbb{P}(E) \arrow[l, "\alpha"'] & \mathbb{P}(Q_E) \arrow[l, "\beta"']
\end{tikzcd}$$

Thus we may take $\mathbb{P}(Q_E)$ to be a split manifold $F(E)$. Let $x_1 = \beta^*c_1(S_E^*)$ and $x_2 = c_1(S_{Q_E}^*)$. By the result on the cohomology of a projective bundle,

$$H^*(F(E)) = H^*(M)[x_1, x_2]/(x_1^3 + c_1(E)x_1^2 + c_2(E)x_1 + c_3(E), x_2^2 + c_1(Q_E)x_2 + c_2(Q_E))$$

Therefore $H^*(M)$ is embedded in $H^*(F(E))$.

The pattern is now clear; we split off one subbundle at a time by pulling back to the projectivization of a quotient bundle.

$$\begin{tikzcd}
E \arrow[d] & S_1 \oplus Q_1 \arrow[d] & S_1 \oplus S_2 \oplus Q_2 \arrow[d] & \cdots \arrow[d] & S_1 \oplus \cdots \oplus S_{n-1} \oplus Q_{n-1} \arrow[d] \\
M & \mathbb{P}(E) \arrow[l] & \mathbb{P}(Q_1) \arrow[l] & \cdots \arrow[l] & \mathbb{P}(Q_{n-2}) = F(E) \arrow[l]
\end{tikzcd} \tag{3.3}$$

split manifol

split manifold const

So for a bundle E of any rank n, a split manifold $F(E)$ exists and is given explicitly by (3.3). Its cohomology $H^*(F(E))$ is a free $H^*(M)$-module having as a basis all monomials of the form

$$x_1^{\alpha_1}x_2^{\alpha_2}\cdots x_{n-1}^{\alpha_{n-1}}, \quad 0 \le \alpha_i \le n-i$$

cohomology P(E) description

where $x_i = c_1(S_i^*)$ in the notation of the diagram. Using Proposition 3.3.2 and an induction, we see the cohomology ring of $H^*(F(E))$ can also be written into

$$H^*(F(E)) = H^*(M)[x_1, \ldots, x_n]/(\prod_{i=1}^{n}(1 + x_i) = \pi^*c(E)).$$

More generally, by iterating the construction above we see that given any number of vector bundles $E_1, \ldots, E_r$ over M, there is a manifold N and a map $\sigma : N \to M$ such that the pullbacks of $E_1, \ldots, E_r$ to N are all direct sums of line bundles and that $H^*(M)$ injects into $H^*(N)$ under σ^*. The manifold N is a **split manifold** for $E_1, \ldots, E_r$.

Because of the existence of the split manifolds we can formulate the following general principle.

Proposition 3.3.7 (The Splitting Principle). *To prove a polynomial identity in the Chern classes of complex vector bundles, it suffices to prove it under the assumption that the vector bundles are direct sums of line bundles.*

For example, suppose we want to prove a certain polynomial relation $P(c(E), c(F), c(E\otimes F)) = 0$ for vector bundles E and F over a manifold M. Let $\sigma : N \to M$ be a split manifold for the pair E, F. By the naturality of the Chern classes

$$\sigma^* P(c(E), c(F), c(E \otimes F)) = P(c(\sigma^* E), c(\sigma^* F), c((\sigma^* E) \otimes (\sigma^* F))),$$

where $\sigma^* E$ and $\sigma^* F$ are direct sums of line bundles. So if the identity holds for direct sums of line bundles, then

$$\sigma^* P(c(E), c(F), c(E \otimes F)) = 0,$$

and the injectivity of σ^*,

$$P(c(E), c(F), c(E \otimes F)) = 0.$$

Whitney product formula and the top Chern class

Proposition 3.3.8 (Whitney Product Formula). *Let E and E' be smooth complex vector bundles, then $c(E \oplus E') = c(E)c(E')$.*

Proof. We consider first the case of a direct sum of line bundles:

$$E = L_1 \oplus \cdots \oplus L_n.$$

By abuse of notation we write $\pi^* E = L_1 \oplus \cdots \oplus L_n$ for the pullback of E to the projectivization $\mathbb{P}(E)$. Over $\mathbb{P}(E)$, the universal subbundle S splits off from $\pi^* E$.

$$\begin{array}{ccc} E & & S \subseteq \pi^* E \\ \big\downarrow & & \big\downarrow \\ M & \xleftarrow[\pi]{} & \mathbb{P}(E) \end{array}$$

For each i, let $s_i : S \to L_i$ be the projection map, then s_i is a section of $\mathrm{Hom}(S, L_i) = S^* \otimes L_i$. Since at every point y of $\mathbb{P}(E)$, the fiber S_y is a 1-dimensional subspace of $(\pi^* E)_y$, the projections $s_1, \ldots, s_n$ cannot be simultaneously zero. It follows that the open sets

$$U_i = \{y \in \mathbb{P}(E) : s_i(y) \neq 0\}$$

form an open cover of $\mathbb{P}(E)$. Over each U_i the bundle $(S^* \otimes L_i)|_{U_i}$ has a nowhere-vanishing section, namely s_i; so $(S^* \otimes L_i)|_{U_i}$ is trivial. Let ξ_i be a closed global 2-form on $\mathbb{P}(E)$ representing $c_1(S^* \otimes L_i)$. Then $\xi_i|_{U_i} = d\omega_i$ for some 1-form ω_i on U_i. The crux of the proof is to find a global form on $\mathbb{P}(E)$ that represents $c_1(S^* \otimes L_i)$ and that vanishes on U_i; because ω_i is not a global form on $\mathbb{P}(E)$, $e_i - d\omega_i$ won't do. However. by shrinking the open cover $\{U_i\}$ slightly we can extend $e_i - d\omega_i$ to a global form. To be precise we will need the following lemma.

Lemma 3.3.9 (The Shrinking Lemma). *Let X be a normal topological space and $\{U_i\}$ a finite open cover of X. Then there is an open cover $\{V_i\}$ with $\overline{V}_i \subseteq U_i$.*

It follows from that on $\mathbb{P}(E)$ there exists an open cover $\{V_i\}$ and smooth functions ρ_i satisfying

- $\overline{V}_i \subseteq U_i$.
- ρ_i is 1 on $\overline{V}_i$ and is 0 outside U_i.

Now $\rho_i\omega_i$ is a global form which agrees with ω_i on V_i so that

$$\xi_i - d(\rho_i\omega_i)$$

is a global form representing $c_1(S^* \otimes L_i)$ and vanishing on V_i. In summary, there is an open cover $\{V_i\}$ of $\mathbb{P}(E)$ such that $c_i(S^* \otimes L_i)$ may be represented by a global form which vanishes on V_i.

Since $\{V_i\}$ covers $\mathbb{P}(E)$, $\prod_{i=1}^n c_1(S^* \otimes L_i) = 0$. Writing $x = c_1(S^*)$, this gives

$$\prod_{i=1}^n (x + c_1(L_i)) = x^n + \sigma_1 x^{n-1} + \cdots + \sigma_n = 0$$

where σ_i is the i-th elementary symmetric polynomial of $c_1(L_1), \ldots, c_1(L_n)$. But this equation is precisely the defining equation of $c(E)$. Thus $\sigma_i = c_i(E)$ and

$$c(E) = \prod_{i=1}^n (1 + c_1(L_i)) = \prod_{i=1}^n c(L_i).$$

So the Whitney product formula holds for a direct sum of line bundles.

By the splitting principle it holds for any complex vector bundle. As an illustration of the splitting principle we will go through the argument in detail. Let E and E' be two complex vector bundles of rank n and m respectively and let $\pi : F(E) \to M$ and $\pi' : F(\pi^*E') \to F(E)$ be the splitting constructions. Both bundles split completely when pulled back to $F(\pi^*E')$ as indicated in the diagram below.

$$\begin{tikzcd}
E \otimes E' \arrow[d] & L_1 \oplus \cdots \oplus L_n \oplus \pi^*E' \arrow[d] & L_1 \oplus \cdots \oplus L_n \oplus L'_1 \oplus \cdots \oplus L'_m \arrow[d] \\
M & F(E) \arrow[l, "\pi"'] & F(\pi^*E') \arrow[l, "\pi'"']
\end{tikzcd}$$

Let $\sigma = \pi' \circ \pi$. Then

$$\begin{aligned}
\sigma^* c(E \oplus E') = c(\sigma^*(E \oplus E')) &= c(L_1 \oplus \cdots \oplus L_n \oplus L'_1 \oplus \cdots \oplus L'_n) \\
&= \prod c(L_i)c(L'_i) = \sigma^* c(E)\sigma^* c(E') = \sigma^* c(E)c(E').
\end{aligned}$$

Since σ^* is injective, $c(E \oplus E') = c(E)c(E')$. This concludes the proof. $\square$

Corollary 3.3.10. *Let*

$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0$$

be an exact sequence of smooth complex vector bundles, then $c(B) = c(A)c(C)$.

As an application of the existence of the split manifold and the Whitney product formula, we will prove now the relation between the top Chern class and the Euler class.

Proposition 3.3.11. *The top Chern class of a complex vector bundle E is the Euler class of its realization:*

$$c_n(E) = e(E_{\mathbb{R}}), \quad n = \mathrm{rank} E.$$

Proof. Let E be a rank n complex vector bundle and $\sigma : F(E) \to E$ its split manifold. Write $\sigma^* E = L_1 \oplus \cdots \oplus L_n$, where the L_i's are line bundles on the split manifold $F(E)$. Then

$$\begin{aligned}\sigma^* c_n(E) &= c_n(\sigma^* E) = c_1(L_1) \cdots c_1(L_n) = e((L_1)_{\mathbb{R}}) \cdots e((L_n)_{\mathbb{R}}) \\ &= e((L_1)_{\mathbb{R}} \oplus \cdots \oplus (L_n)_{\mathbb{R}}) = e((\sigma^* E)_{\mathbb{R}}) \\ &= \sigma^* e(E_{\mathbb{R}}).\end{aligned}$$

By the injectivity of σ^* on cohomology, $c_n(E) = e(E_{\mathbb{R}})$. □

Computation of some Chern classes

Given a rank n complex vector bundle E we may write formally

$$c(E) = \prod_{i=1}^{n} (1 + x_i)$$

where the x_i's may be thought of as the first Chern class of the line bundles into which E splits when pulled back to the splitting manifold $F(E)$. Since the Chern classes $c_1(E), \ldots, c_n(E)$ are the elementary symmetric functions of $x_1, \ldots, x_n$, by the symmetric function theorem any symmetric polynomial in $x_1, \ldots, x_n$ is a polynomial in $c_1(E), \ldots, c_n(E)$; a similar result holds for power series.

Example 3.3.12 (Exterior powers, symmetric powers, and tensor products). Recall that if V is a vector space with basis $\{v_1, \ldots, v_n\}$, then the exterior power $\bigwedge^p V$ is the vector space with basis $\{v_{i_1} \wedge \cdots \wedge v_{i_p} : 1 \le i_1 < \cdots < i_p \le n\}$. So if E is the direct sum of line bundles $E = L_1 \oplus \cdots \oplus L_n$, then

$$\bigwedge^p E = \bigoplus_{1 \le i_1 < \cdots < i_p \le n} (L_{i_1} \oplus \cdots \oplus L_{i_p}).$$

Hence

$$c(\bigwedge^p E) = \prod_{1 \le i_1 < \cdots < i_p \le n} (1 + c_1(L_{i_1} \oplus \cdots \oplus L_{i_n})) = \prod_{1 \le i_1 < \cdots < i_p \le n} (1 + x_{i_1} + \cdots + x_{i_n}).$$

Since the right-hand side is symmetric in $x_1, \ldots, x_n$, it is expressible as a polynomial of $c_1(E), \ldots, c_n(E)$, so

$$c(\bigwedge^p E) = Q(c_1(E), \ldots, c_n(E)).$$

By the splitting principle this formula holds for every rank n vector bundle, whether it is a direct sum or not.

Similarly, if V and W are vector spaces with bases $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$ respectively, then the p-th symmetric power $\Sigma^p V$ of V is the vector space with basis $\{v_{i_1} \otimes \cdots \otimes v_{i_p} : 1 \le i_1 \le \cdots \le i_p \le n\}$ and the tensor product $V \otimes W$ is the vector space with basis $\{v_i \otimes v_j\}$. By the same discussion as above, if E is a rank n vector bundle with $c(E) = \prod_{i=1}^n (1 + x_i)$ and F is a rank m vector bundle with $c(F) = \prod_{j=1}^m (1 + y_j)$, then

$$c(\Sigma^p V) = \bigoplus_{1 \le i_1 \le \cdots \le i_p \le n} (1 + x_{i_1} + \cdots + x_{i_p})$$

and

$$c(E \otimes F) = \prod (1 + x_i + y_j).$$

In particular if L is a complex line bundle with first Chern class y, then

$$c(E \otimes L) = \prod_{i=1}^{n}(1 + y + x_i) = \sum_{i=0}^{n} c_i(E)(1+y)^{n-i}$$

where by convention we set $c_0(E) = 1$.

nd Todd class **Example 3.3.13 (The L-class and the Todd class).** In the notation of the preceding example the power series

$$\prod_{i=1}^{n} \frac{\sqrt{x_i}}{\tanh \sqrt{x_i}}$$

is symmetric in $x_1, \ldots, x_i$, hence is some power series L in $c_i(E), \ldots, c_n(E)$. This power series $L(E) = L(c_1(E), \ldots, c_n(E))$ is called the L-class of E. By the splitting principle the L-class automatically satisfies the product formula

$$L(E \oplus F) = L(E)L(F).$$

Similarly,

$$\prod_{i=1}^{n} \frac{x_i}{1 - e^{-x_i}} = \mathrm{Td}(c_1(E), \ldots, c_n(E) = \mathrm{Td}(E)$$

defines the Todd class of E. By the splitting principle the Todd class also automatically satisfies the product formula. The L-class and the Todd class turn out to be of fundamental importance in the Hirzebruch signature formula and the Riemann-Roch theorem.

n dual bundle **Example 3.3.14.** Consider a direct sum of line bundles

$$E = L_1 \oplus \cdots \oplus L_n$$

By the Whitney product formula

$$c(E) = c(L_1) \cdots c(L_n) = (1 + c_1(L_1)) \cdots (1 + c_1(L_n)).$$

On the other hand

$$E^* = L^* \otimes \cdots \otimes L_n^*$$

and so

$$c(E^*) = c(L_1^*) \cdots c(L_n^*) = (1 - c_1(L_1)) \cdots (1 - c_1(L_n)).$$

Therefore

$$c_i(E^*) = (-1)^i c_i(E).$$

By the splitting principle this result holds for all complex vector bundles E.

rn class CP^n **Example 3.3.15 (The Chern classes of the complex projective space).** By analogy with the definition of a differentiable manifold, we say that a second countable, Hausdorff space M is a complex manifold of dimension n if every point has a neighborhood U homeomorphic to some open ball in $\mathbb{C}^n$ such that the transition functions are holomorphic. Smooth maps and smooth vector bundles have obvious analogues in the holomorphic category. The holomorphic tangent bundle of M is a complex vector bundle of rank n. The **Chern class of a complex manifold** is defined to be the Chern class of its holomorphic tangent bundle.

The complex projective space $\mathbb{C}^n$ is an example of a complex manifold, since the transition functions g_{ji} relative to the standard open cover are given by multiplication by z_i/z_j, which are holomorphic functions on $U_i \cap U_j$. Recall that there is a tautological exact sequence on $\mathbb{CP}^n$

$$0 \longrightarrow S \longrightarrow \mathbb{C}^{n+1} \longrightarrow Q \longrightarrow 0$$

where $\mathbb{C}^{n+1}$ denotes the trivial bundle of rank $n+1$ over $\mathbb{CP}^n$. A tangent vector to $\mathbb{CP}^n$ at a line ℓ in $\mathbb{C}^{n+1}$ may be regarded as an infinitesimal motion of the line. Such a motion corresponds to a linear map from ℓ to the quotient space $\mathbb{C}^{n+1}/\ell$, which may be represented by the complementary subspace of ℓ in $\mathbb{C}^{n+1}$ (relative to some metric). Thus, denoting the holomorphic tangent bundle by T, we have

$$T \cong \mathrm{Hom}(S, Q) = S^* \otimes Q.$$

We will compute the Chern class of T in two ways.

(1) Tensoring the tautological sequence with S^*, we get

$$0 \longrightarrow \mathbb{C} \longrightarrow S^* \otimes \mathbb{C}^{n+1} \longrightarrow S^* \otimes Q \longrightarrow 0$$

By the Whitney product formula

$$c(T) = c(S^* \otimes Q) = c(S^* \otimes \mathbb{C}^{n+1}) = c(S^* \otimes \cdots \otimes S^*) = (1+x)^{n+1}.$$

where $x = c_1(S^*)$.

(2) From the tautological exact sequence and the Whitney product formula

$$Q = \frac{1}{c(S)} = \frac{1}{1-x} = 1 + x + \cdots + x^n$$

since $x^{n+1} = 0$ in $H^*(\mathbb{CP}^n)$. By

$$\begin{aligned} c(T) = c(S^* \otimes Q) &= \sum_{i=0}^{n} c_i(Q)(1+x)^{n-i} = \sum_{i=0}^{n} x^i (1+x)^{n-i} \\ &= (1+x)^n \sum_{i=0}^{n} \Big(\frac{x}{1+x}\Big)^i \\ &= (1+x)^{n+1}\Big[1 - \Big(\frac{x}{1+x}\Big)^{n+1}\Big] \\ &= (1+x)^{n+1} - x^{n+1} \\ &= (1+x)^{n+1}. \end{aligned}$$

3.3.3 Pontrjagin classes

Although the Chern classes are invariants of a complex bundle, they can be used to define invariants of a real vector bundle, called the Pontrjagin classes. In this part we define the Pontrjagin classes, compute a few examples, and as an application obtain an embedding criterion for differentiable manifolds.

Conjugate bundles

Let V be a complex vector space. If $z \in \mathbb{C}$ and $v \in V$, the formula $z \cdot v := \bar{z} v$ defines an action of $\mathbb{C}$ on V. The underlying additive group of V with this action as scalar multiplication is called the conjugate vector space of V, denoted $\overline{V}$. The conjugate space $\overline{V}$ may be thought of as V with the opposite complex structure; as a vector space, V is anti-isomorphic to V. A linear map $T : V \to W$ of two complex vector spaces V and W is also a linear map of the conjugate vector spaces $T : V \to W$; we denote both by T as they are represented by the same matrix.

Given a complex vector bundle E with trivialization $\Phi_\alpha : E|_{U_\alpha} \to U_\alpha \times \mathbb{C}^n$, we construct the conjugate vector bundle E by replacing each fiber of E by its conjugate. The trivialization of E is given by

$$\bar{\Phi}_\alpha : \overline{E}|_{U_\alpha} \to U_\alpha \times \overline{C}^n,$$

which is the composition

$$\overline{E}|_{U_\alpha} \xrightarrow{\Phi_\alpha} U_\alpha \times \overline{\mathbb{C}}^n \xrightarrow{\text{conjugate}} U_\alpha \times \mathbb{C}^n$$

In terms of transition functions, if the cocycle $\{\tau_{\alpha\beta}\}$ defines E, then

$$\bar{\Phi}_\alpha \circ \bar{\Phi}_\beta^{-1}(p, v) = \bar{\Phi}_\alpha \circ \Phi_\beta^{-1}(p, \bar{v}) = (p, \overline{\tau_{\alpha\beta}\bar{v}}) = (p, \overline{\tau}_{\alpha\beta} v).$$

Therefore $\{\overline{\tau}_{\alpha\beta}\}$ defines $\overline{E}$.

By endowing a complex vector bundle on a manifold with a Hermitian metric, we can reduce its structure group to the unitary group. Since unitary matrices $\tau_{\alpha\beta}$ satisfy $\overline{\tau}_{\alpha\beta} = (\tau_{\alpha\beta}^T)^{-1}$, we see that the conjugate bundle $\overline{E}$ and the dual bundle E^* have the same transition functions and hence are isomorphic. So by Example 3.3.14, if $c(E) = \prod(1 + x_i)$, then $c(\overline{E}) = \prod(1 - x_i)$.

Realization and complexification

By simply forgetting the complex structure, we can regard a linear map of complex vector spaces $L : \mathbb{C}^n \to \mathbb{C}^n$ with coordinates $z_1, \ldots, z_n$ as a linear map of the underlying real vector spaces $L_\mathbb{R} : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ with coordinates $x_1, \ldots, x_2$ where $z_k = x_{2k-1} + iy_{2k}$. Conversely, via the natural embedding of $\mathbb{R}^n$ in $\mathbb{C}^n$, a linear map of real vector spaces $L : \mathbb{R}^n \to \mathbb{R}^n$ gives rise to a map $L \otimes \mathbb{C} : \mathbb{C}^n \to \mathbb{C}^n$. The first operation is called **realization** and the second, **complexification**.

The complexification of a real matrix is the matrix itself, but with the entries viewed as complex numbers. The realization of a complex matrix is given by

$$\begin{pmatrix} a_1^1 + ib_1^1 & \cdots & a_n^1 + ib_n^1 \\ \vdots & & \vdots \\ a_1^n + ib_1^n & \cdots & a_n^n + ib_n^n \end{pmatrix}_\mathbb{R} := \begin{pmatrix} a_1^1 & -b_1^1 & \cdots & a_n^1 & -b_n^1 \\ b_1^1 & a_1^1 & & b_n^1 & a_n^1 \\ & \vdots & & & \vdots \\ a_1^n & -b_1^n & \cdots & a_n^n & -b_n^n \\ b_1^n & a_1^n & & b_n^n & a_n^n \end{pmatrix}$$

Lemma 3.3.16. *Let A be an $n \times n$ complex matrix. There is a $2n \times 2n$ matrix B, independent of A, such that $A_\mathbb{R} \otimes \mathbb{C}$ is similar to $\left(\begin{smallmatrix} A & 0 \\ 0 & \overline{A} \end{smallmatrix}\right)$ via B.*

Proof. In the case $n = 1$, we need to diagonalize the matrix

$$\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$$

Corresponding to the eigenvalues $\alpha + i\beta$ and $\alpha - i\beta$ are the eigenvectors $(1, -i)^T$ and $(1, i)^T$. Therefore, $\left(\begin{smallmatrix} 1 & 1 \\ -i & i \end{smallmatrix}\right)$.

For the general case, we can take B to be

$$\begin{pmatrix}
1 & & & & 1 & & & \\
-i & & & & i & & & \\
 & 1 & & & & 1 & & \\
 & -i & & & & i & & \\
 & & \ddots & & & & \ddots & \\
 & & & 1 & & & & 1 \\
 & & & -i & & & & i
\end{pmatrix}$$

□

If E is a complex vector bundle of rank n with transition functions $\{\tau_{\alpha\beta}\}$, then $E_{\mathbb{R}} \otimes \mathbb{C}$ is the complex vector bundle of rank $2n$ with transition functions $\{(\tau_{\alpha\beta})_{\mathbb{R}} \otimes \mathbb{C}\}$. By Lemma 3.3.16, complexification matr

$$E_{\mathbb{R}} \otimes \mathbb{C} \cong E \oplus \overline{E}$$

This result may be seen alternatively as follows. On the complex vector space $E_{\mathbb{R}} \otimes \mathbb{C}$, multiplication by i is a linear transformation J satisfying $J^2 = -\mathrm{id}$. Therefore, the eigenvalues of J are $\pm i$ and $E_{\mathbb{R}} \otimes \mathbb{C}$ accordingly decomposes into a direct sum

$$E_{\mathbb{R}} \otimes \mathbb{C} = E_i \oplus E_{-i}.$$

On the i-eigenspace, J acts as multiplication by i, hence $E \subseteq E_i$. Similarly, $\overline{E} \subseteq E_{-i}$. It follows by reasons of dimension that

$$E_{\mathbb{R}} \otimes \mathbb{C} = E \oplus \overline{E}.$$

The Pontrjagin classes of a real vector bundle

By their naturality property the Chern classes of a smooth complex vector bundle are invariants of the bundle. For a real vector bundle E similar invariants may be obtained by considering the Chern classes of its complexification $E \otimes_{\mathbb{R}} \mathbb{C}$; these are the **Pontrjagin classes** of E. More precisely, if E is a rank n real vector bundle over M, then its total Pontrjagin class is

$$p(E) = 1 + p_1(E) + \cdots + p_n(E) = 1 + c_1(E \otimes \mathbb{C}) + \cdots + c_n(E \otimes \mathbb{C}) \in H^*(M).$$

The **Pontrjagin class of a manifold** is defined to be that of its tangent bundle.

Remark 3.3.1. Let E be a real vector bundle. Because the transition functions of $E \otimes \mathbb{C}$ are the same as those of E, they are real-valued, and therefore $E \otimes \mathbb{C}$ is isomorphic to its conjugate $E \otimes \mathbb{C}$. It follows that $c_i(E \otimes \mathbb{C}) = c_i(\overline{E \otimes \mathbb{C}}) = (-1)^i c_i(E \otimes \mathbb{C})$. For an odd i, then, $2c_i(E \otimes \mathbb{C}) = 0$. Thus the odd Pontrjagin classes, as we have defined them, are zero in the de Rham cohomology, and torsion of order 2 in the integral cohomology. The usual definition of the Pontrjagin classes in the literature ignores these odd Chern classes and defines $p(E)$ to be $(-1)^i c_{2i}(E \otimes \mathbb{C})$

Example 3.3.17 (The Pontrjagin class of the sphere). Since the sphere S^n is orientable, its normal bundle N in $\mathbb{R}^{n+1}$ is trivial. From the exact sequence

$$0 \longrightarrow TS^n \longrightarrow T\mathbb{R}^{n+1}|_{S^n} \longrightarrow NS^n \longrightarrow 0$$

we see by the Whitney product formula that

$$p(S^n)p(NS^n) = p(T\mathbb{R}^{n+1}|_{S^n}).$$

Therefore $p(S^n) = 1$.

complex mani

Example 3.3.18 (The Pontrjagin class of a complex manifold). The Pontrjagin class of a complex manifold M is defined to be that of the underlying real manifold $M_{\mathbb{R}}$. Let T be the holomorphic tangent bundle to M. Then the tangent bundle to $M_{\mathbb{R}}$ is the realization of T and

$$p(M) = p(T_{\mathbb{R}}) = c(T_{\mathbb{R}} \otimes \mathbb{C}) = c(T \oplus \overline{T}) = c(T)c(\overline{T}).$$

So if the total Chern class of the complex manifold M is $c(M) = \prod(1+x_i)$, then the Pontrjagin class is $p(M) = \prod(1 - x_i^2)$.

Remark 3.3.2. If we had followed the usual sign convention for the Pontrjagin classes, the Pontrjagin class of a complex manifold would be $p(M) = \prod(1 + x_i)$, where the x_i's are defined as above. To have only positive terms in this formula is one of the reasons for the sign in $(-1)^i c_{2i}(E \otimes \mathbb{C})$ in the usual definition of the Pontrjagin class.

Remark 3.3.3. Let M be a compact oriented manifold of dimension $4n$. By Poincaré duality the wedge product $\wedge : H^{2n}(M) \otimes H^{2n}(M) \to \mathbb{R}$ is a nondegenerate symmetric bilinear form and hence has a signature; this is called the signature of M. Hirzebruch proved that the signature is expressible in terms of the Pontrjagin classes.

$$\operatorname{sgn}(M) = (-1)^n \int_M L(p_1(M), \ldots, p_n(M)).$$

where L is the polynomial defined in Example 3.3.13.

Application to the embedding of a manifold in a Euclidean space

Example 3.3.19 (Decide if $\mathbb{CP}^n$ can be differentiably embedded in $\mathbb{R}^m$). By Example 3.3.15 and Example 3.3.18 the Pontrjagin class of $\mathbb{CP}^n$ is

$$p(\mathbb{CP}^n) = c(T\mathbb{CP}^n)c(\overline{T\mathbb{CP}^n}) = (1+x)^{n+1}(1-x)^{n+1} = (1-x^2)^{n+1}.$$

If $\mathbb{CP}^n$ can be differentiably embedded in $\mathbb{R}^m$, then there is an exact sequence

$$0 \longrightarrow (T\mathbb{CP}^n)_{\mathbb{R}} \longrightarrow T\mathbb{R}^m \longrightarrow N \longrightarrow 0$$

where $(T\mathbb{CP}^n)_{\mathbb{R}}$ is the realization of the holomorphic tangent bundle $T\mathbb{CP}^n$ and N is the normal bundle of $\mathbb{CP}^n$ in $\mathbb{R}^m$. By the Whitney product formula

$$1 = p(T\mathbb{R}^m) = p((T\mathbb{CP}^n)_{\mathbb{R}})p(N).$$

Therefore

$$p(N) = \frac{1}{p((T\mathbb{CP}^n)_{\mathbb{R}})} = \frac{1}{(1-x^2)^{n+1}} = 1 + a_1x^2 + \cdots + a_{[\frac{n}{2}]}x^{2[\frac{n}{2}]}.$$

Since N is a real vector bundle of rank $m - 2n$, the top component of $p(N)$ should be in $H^{2(m-2n)}(\mathbb{CP}^n)$. From the fact that $x^{2[\frac{n}{2}]}$ is nonzero in $H^*(\mathbb{CP}^n)$, we conclude that

$$2(m-2n) \geq \deg(x^{2[\frac{n}{2}]}) = 4[\frac{n}{2}],$$

which is

$$m \geq 2n + 2[\frac{n}{2}].$$

Therefore there is no embedding of $\mathbb{CP}^n$ into $\mathbb{R}^m$ provided $m < 2n + 2[n/2]$.

3.4 Symplectic manifolds

3.4.1 Symplectic tensors

We begin with linear algebra. A 2-covector ω on a finite-dimensional vector space V is said to be **nondegenerate** if the linear map $\widehat{\omega} : V \to V^*$ defined by $\widehat{\omega}(v) = v \lrcorner \omega$ is invertible for every nonzero $v \in V$.

Proposition 3.4.1. *The following are equivalent for 2-covector ω on a finite-dimensional vector space V:*

(a) ω is nondegenerate.

(b) For each nonzero $v \in V$, there exists $w \in V$ such that $\omega(v, w) \neq 0$.

(c) In terms of some (hence every) basis, the matrix (ω_{ij}) representing ω is nonsingular.

Proof. The nondegenerate condition is to say: for each nonzero $v \in V$, $v \lrcorner \omega \neq 0$, which is exactly (b). Now let $(E_1, \ldots, E_n)$ be a basis of V, and $W = (\omega_{ij})$ be the matrix. Then for $v = v^i E_i$ and $w = w^j E_j$,

$$\omega(v, w) = \omega_{ij} v^i w^j.$$

Thus

$$\omega(v, w) = 0 \text{ for all } w \in V \iff \omega_{ij} v^i = 0 \text{ for all } j \iff v \text{ is a solution of } W^T X = 0.$$

so ω is nondegenerate if and only if W is nonsingular. □

A nondegenerate 2-covector is called a **symplectic tensor**. A vector space V endowed with a specific symplectic tensor is called a **symplectic vector space**.

ic tensor eg **Example 3.4.2.** Let V be a $2n$-dimensional vector space. Let $(A_1, \ldots, A_n, B_1, \ldots, B_n)$ be a basis of V and $(\alpha^1, \ldots, \alpha^n, \beta^1, \ldots, \beta^n)$ denote the corresponding dual basis for V^*, and let $\omega \in \bigwedge^2(V^*)$ be the 2-covector defined by

$$\omega = \sum_{i=1}^{n} \alpha^i \wedge \beta^i. \tag{4.1}$$ symplectic te

Note that the action of ω on basis vectors is given by

$$\omega(A_i, A_j) = \omega(B_i, B_j) = 0, \quad \omega(A_i, B_j) = -\omega(B_j, A_i) = \delta_{ij}. \tag{4.2}$$ symplectic te

and the matrix (ω_{ij}) representing ω is

$$\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$$

Thus ω is nondegenerate, and so is a symplectic tensor.

If (V, ω) is a symplectic vector space and $S \subseteq V$ is any linear subspace, we define the **symplectic complement** of S, denoted by $S^\perp$, to be the subspace

$$S^\perp = \{v \in V : \omega(v, w) = 0 \text{ for all } w \in S\}.$$

As the notation suggests, the symplectic complement is analogous to the orthogonal complement in an inner product space. Just as in the inner product case, the dimension of $S^\perp$ is the codimension of S, as the next lemma shows.

omplement dim **Lemma 3.4.3.** *Let (V, ω) be a symplectic vector space. For any linear subspace $S \subseteq V$, we have $\dim S + \dim S^\perp = \dim V$.*

Proof. Let $S \subseteq V$ be a subspace, and define a linear map $\Phi : V \to S^*$ by $\Phi(v) = (v \lrcorner \omega)|_S$, or equivalently

$$\Phi(v)(w) = \omega(v, w).$$

Suppose φ is an arbitrary element of S^*, and let $\widetilde{\varphi} \in V^*$ be any extension of φ to a linear functional on all of V. Since the map $\widehat{\omega} : V \to V^*$ is an isomorphism, there exists $v \in V$ such that $v \lrcorner \omega = \widetilde{\varphi}$. It follows that $\Phi(v) = \varphi$, and therefore Φ is surjective. By the rank-nullity law, $S^\perp = \ker \Phi$ has dimension equal to $\dim V - \dim S^* = \dim V - \dim S$. □

Proposition 3.4.4. *Let (V, ω) be a symplectic vector space and $S \subseteq V$ be a linear subspace. Then $(S^\perp)^\perp = S$.*

Proof. By definition, since $\omega(v, w) = -\omega(w, v)$, we have $S \subseteq (S^\perp)^\perp$. Now note that

$$\dim(S^\perp)^\perp = \dim V - \dim S^\perp = \dim S,$$

thus $(S^\perp)^\perp = S$. □

Symplectic complements differ from orthogonal complements in one important respect: although it is always true that $S \cap S^\perp = 0$ in an inner product space, this need not be true in a symplectic vector space. Indeed, if S is 1-dimensional, the fact that ω is alternating forces $\omega(v, v) = 0$ for every $v \in S$, so $S = S^\perp$. Carrying this idea a little further, a linear subspace $S \subseteq V$ is said to be

- **symplectic** if $S \cap S^\perp = \{0\}$.
- **isotropic** if $S \subseteq S^\perp$.
- **coisotropic** if $S \supseteq S^\perp$.
- **Lagrangian** if $S = S^\perp$.

subspace iff **Proposition 3.4.5.** *Let (V, ω) be a symplectic vector space and $S \subseteq V$ be a linear subspace. Then*

(a) S is symplectic if and only if $S^\perp$ is symplectic.

(b) S is symplectic if and only if $\omega|_S$ is nondegenerate.

(c) S is isotropic if and only if $\omega|_S = 0$.

(d) S is coisotropic if and only if $S^\perp$ is isotropic.

(e) S is Lagrangian if and only if $\omega|_S = 0$ and $\dim S = \dfrac{1}{2} \dim V$.

Proof. Since $(S^\perp)^\perp = S$, part (a) and (d) are immediate. Next, we note that

$$v \in S \cap S^\perp \iff \omega(v, w) = 0 \text{ for all } w \in S.$$

Thus S is symplectic if and only if $\omega|_S$ is nondegenerate, and S is isotropic if and only if $\omega|_S = 0$.

Finally, S is Lagrangian means S and $S^\perp$ are both isotropic, which implies $\omega|_S = 0$ and $\dim S = \frac{1}{2} \dim V$. Conversely, if S is isotropic and $\dim S = \frac{1}{2} \dim V$, then $S \subseteq S^\perp$ with $\dim S = \dim S^\perp$. Thus $S = S^\perp$ and S is Lagrangian. □

The symplectic tensor ω defined in Example 3.4.2 turns out to be the prototype of all symplectic tensors, as the next proposition shows. This can be viewed as a symplectic version of the GramSchmidt algorithm.

Proposition 3.4.6 (Canonical Form for a Symplectic Tensor). *Let ω be a symplectic tensor on an m-dimensional vector space V. Then V has even dimension $m = 2n$, and there exists a basis for V in which ω has the form (4.1).*

Proof. The tensor ω has the form (4.1) with respect to a basis $(A_1, \ldots, A_n, B_1, \ldots, B_n)$ if and only if its action on basis vectors is given by (4.2).We prove the theorem by induction on $m = \dim V$ by showing that there is a basis with this property.

For $m = 0$ there is nothing to prove. Suppose (V, ω) is a symplectic vector space of dimension $m = 1$, and assume that the proposition is true for all symplectic vector spaces of dimension less than m. Let A_1 be any nonzero vector in V. Since ω is nondegenerate, there exists $B_1 \in V$ such that $\omega(A_1, B_1) \neq 0$. Multiplying B_1 by a constant if necessary, we may assume that $\omega(A_1, B_1) = 1$. Because ω is alternating, B_1 cannot be a multiple of A_1, so the set $\{A_1, B_1\}$ is linearly independent, and hence $\dim V \geq 2$.

Let $S \subseteq V$ be the span of $\{A_1, B_1\}$. Then $\dim S^\perp = \dim m - 2$ by Lemma 3.4.3. Since $\omega|_S$ is nondegenerate, by Proposition 3.4.5 it follows that S is symplectic, and thus $S^\perp$ is also symplectic. By induction, $S^\perp$ is even-dimensional and there is a basis $(A_2, \ldots, A_n, B_2, \ldots, B_n)$ for $S^\perp$ such that (4.2) is satisfied for $2 \leq i, j \leq n$. It follows easily that $(A_1, \ldots, A_n, B_1, \ldots, B_n)$ is the required basis for V. □

Because of this, if (V, ω) is a symplectic vector space, a basis $(A_1, \ldots, A_n, B_1, \ldots, B_n)$ for V is called a **symplectic basis** if (4.2) holds, which is equivalent to ω being given by (4.1) in terms of the dual basis. The proposition then says that every symplectic vector space has a symplectic basis.

This leads to another useful criterion for 2-covector to be nondegenerate. For an alternating tensor ω, the notation ω^k denotes the k-fold wedge product $\omega \wedge \cdots \wedge \omega$.

Proposition 3.4.7. *Suppose V is a $2n$-dimensional vector space and $\omega \in \bigwedge^2(V^*)$. Then ω is a symplectic tensor if and only if $\omega^n \neq 0$.*

Proof. Suppose first that ω is a symplectic tensor. Let (A_i, B_i) be a symplectic basis for V, and write $\omega = \sum_{i=1}^n \alpha^i \wedge \beta^i$ in terms of the dual coframe. Then we compute

$$\omega^n = \sum_I{}' \alpha^{i_1} \wedge \beta^{i_1} \wedge \cdots \wedge \alpha^{i_n} \wedge \beta^{i_n} = n!(\alpha^1 \wedge \beta^1 \wedge \cdots \wedge \alpha^n \wedge \beta^n) \neq 0.$$

Conversely, suppose ω is degenerate. Then there is a nonzero vector $v \in V$ such that $v \lrcorner \omega = \widehat{\omega}(v) = 0$. Since interior multiplication by v is an antiderivation, by induction we can show

$$v \lrcorner (\omega^n) = n(v \lrcorner \omega) \wedge \omega^{n-1} = 0.$$

We can extend v to a basis $(E_1, \ldots, E_{2n})$ for V with $E_1 = v$, and then $\omega^n(E_1, \ldots, E_{2n}) = 0$, which implies $\omega^n = 0$. □

3.4.2 Symplectic structures on manifolds

Now let us turn to a smooth manifold M. A **nondegenerate 2-form** on M is a 2-form ω such that ω_p is a nondegenerate 2-covector for each $p \in M$. A **symplectic form** on M is a closed nondegenerate 2-form. A smooth manifold endowed with a specific choice of symplectic form is called a **symplectic manifold**. A choice of symplectic form is also sometimes called a **symplectic structure**.

If (M_1, ω_1) and (M_2, ω_2) are symplectic manifolds, a diffeomorphism $F : M_1 \to M_2$ satisfying $F^*\omega_2 = \omega_1$ is called a **symplectomorphism**. The study of properties of symplectic manifolds that are invariant under symplectomorphisms is known as symplectic geometry or symplectic topology.

Example 3.4.8 (Symplectic Manifolds).

(a) With standard coordinates on $\mathbb{R}^{2n}$ denoted by $(x^1, \dots, x^n, y^1, \dots, y^n)$, the 2-form

$$\omega = \sum_{i=1}^{n} dx^i \wedge dy^i$$

is symplectic: it is obviously closed, and it is nondegenerate because its value at each point is the symplectic tensor of Example 3.4.2. This is called the **standard symplectic form** on $\mathbb{R}^{2n}$.

(b) Suppose Σ is any orientable smooth 2-manifold and ω is a nonvanishing smooth 2-form on ω. Then ω is closed because $d\omega$ is a 3-form on a 2-manifold. Moreover, in two dimensions every nonvanishing 2-form is nondegenerate (by Proposition 3.4.7), so (Σ, ω) is a symplectic manifold.

Suppose (M, ω) is a symplectic manifold. An (immersed or embedded) submanifold $S \subseteq M$ is said to be a symplectic, isotropic, coisotropic, or Lagrangian submanifold if T_pS (thought of as a subspace of T_pM) has the corresponding property at each point $p \in S$. More generally, a smooth immersion (or embedding) $F : N \to M$ is said to have one of these properties if the subspace $dF_p(T_pN) \subseteq T_{F(p)}M$ has the corresponding property for every $p \in N$. Thus a submanifold is symplectic (isotropic, etc.) if and only if its inclusion map has the same property.

Proposition 3.4.9. *Suppose (M, ω) is a symplectic manifold and $F : N \to M$ is a smooth immersion. Then F is isotropic if and only if $F^*\omega = 0$, and F is symplectic if and only if $F^*\omega$ is a symplectic form.*

Proof. By Proposition 3.4.5, F is isotropic means $\omega|_{dF(TN)} = 0$, which means for all $v, w \in T_pN$ we have

$$0 = \omega(dF_p(v), dF_p(w)) = F^*\omega(v, w).$$

So Therefore F is isotropic if and only if $F^*\omega = 0$ on N. Similarly, by Proposition 3.4.5 we can show that F is symplectic if and only if $F^*\omega$ is nondegenerate on N, which is to say $F^*\omega$ is a symplectic form. $\square$

The canonical symplectic form on the cotangent bundle

The most important symplectic manifolds are total spaces of cotangent bundles, which carry canonical symplectic structures that we now define. First, there is a natural 1-form τ on the total space of T^*M, called the **tautological 1-form**, defined as follows. A point in T^*M is a covector $\varphi \in T_q^*M$ for some $q \in M$, we denote such a point by the notation (q, φ). We define $\tau \in \Omega^1(T^*M)$ (a 1-form on the total space of T^*M) by

$$\tau_{(q,\varphi)} = d\pi^*_{(q,\varphi)}\varphi.$$

where $\pi : T^*M \to M$ is the projection. In other words, the value of τ at (q, φ) is the pullback with respect to π of the covector φ itself. If v is a tangent vector in $T_{(q,\varphi)}(T^*M)$, then

$$\tau_{(q,\varphi)}(v) = \varphi\big(d\pi_{(q,\varphi)}(v)\big). \tag{4.3}$$ tautological

rm symplectic **Proposition 3.4.10.** *Let M be a smooth manifold. The tautological 1-form τ is smooth, and $\omega = -d\tau$ is a symplectic form on the total space of T^*M.*

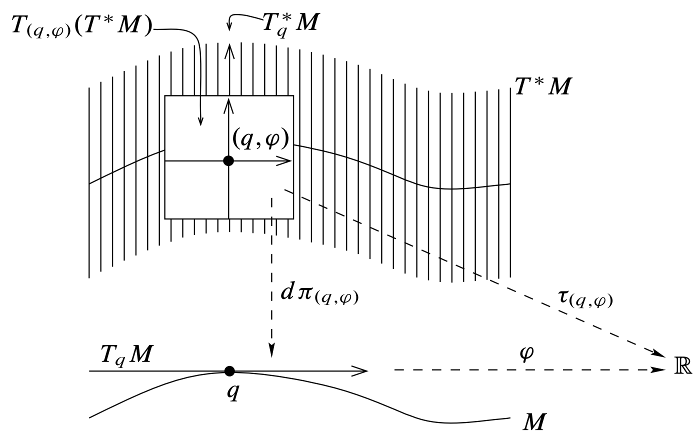

Figure 3.1: The tautological 1-form on T^*M.

Proof. Let (x^i) be smooth coordinates on M, and let (x^i, ξ_i) denote the corresponding natural coordinates on T^*M. Recall that the coordinates of $(q, \varphi) \in T^*M$ are defined to be (x^i, ξ_i), where (x^i) is the coordinate representation of q, and $\xi_i dx^i$ is the coordinate representation of φ. In terms of these coordinates, the projection $\pi : T^*M \to M$ has the coordinate expression $\pi(x, \xi) = x$. This implies that $d\pi^*(dx^i) = dx^i$, so the coordinate expression for τ is

$$\tau_{(x,\xi)} = d\pi^*_{(x,\xi)}(\xi_i dx^i) = \xi_i dx^i.$$

It follows immediately that τ is smooth, because its component functions in these coordinates are linear.

Let $\omega = -d\tau \in \Omega^2(T^*M)$. Clearly, ω is closed, because it is exact. Moreover, in natural coordinates,

$$\omega = \sum_{i=1}^n dx^i \wedge d\xi_i.$$

Under the identification of an open subset of T^*M with an open subset of $\mathbb{R}^{2n}$ by means of these coordinates, ω corresponds to the standard symplectic form on $\mathbb{R}^{2n}$. It follows that ω is symplectic. $\square$

The symplectic form defined in this proposition is called the canonical symplectic form on T^*M. One of its many uses is in giving the following somewhat more geometric interpretation of what it means for a 1-form to be closed.

is embedding **Proposition 3.4.11.** *Let M be a smooth manifold, and let σ be a smooth 1-form on M. Thought of as a smooth map from M to T^*M, σ is a smooth embedding, and σ is closed if and only if its image $\sigma(M)$ is a Lagrangian submanifold of T^*M.*

Proof. Throughout this proof we need to remember that $\sigma : M \to T^*M$ is playing two roles: on the one hand, it is a 1-form on M, and on the other hand, it is a smooth map between manifolds. Since they are literally the same map, we do not use different notations to distinguish between them; but you should be careful to think about which role σ is playing at each step of the argument.

In terms of smooth local coordinates (x^i) for M and corresponding natural coordinates (x^i, ξ_i) for T^*M, the map $\sigma : M \to T^*M$ has the coordinate representation

$$\sigma(x^1, \ldots, x^n) = (x^1, \ldots, x^n, \sigma_1(x), \ldots, \sigma_n(x)),$$

where $\sigma_i dx^i$ is the coordinate representation of σ as a 1-form. It follows immediately that σ is a smooth immersion, and it is injective because $\pi \circ \sigma = \mathrm{id}_M$. To show that it is an embedding, it suffices by Proposition smooth embedd if **??** to show that it is a proper map. This in turn follows from the fact that π is a left inverse for σ, by Proposition proper map if **??**.

Because $\sigma(M)$ is n-dimensional, it is Lagrangian if and only if it is isotropic, which is the case if and only if $\sigma^*\omega = 0$ (Proposition symplectic isotropic map iff pullback 3.4.9). The pullback of the tautological form τ under σ is

$$\sigma^*\tau = \sigma^*(\xi_i dx^i) = \sigma_i dx^i = \sigma.$$

This can also be seen somewhat more invariantly from the computation

$$(\sigma^*\tau)_p(v) = \tau_{\sigma(p)}\big(d\sigma_p(v)\big) = \sigma_p\big(d\pi_{\sigma(p)} \circ d\sigma_p(v)\big) = \sigma_p(d(\pi \circ \sigma)_p(v)) = \sigma_p(v).$$

which follows from the definition of τ and the fact that $\pi \circ \sigma = \mathrm{id}_M$. Therefore,

$$\sigma^*\omega = -\sigma^* d\tau = -d(\sigma^*\tau) = -d\sigma.$$

It follows that σ is a Lagrangian embedding if and only if $d\sigma = 0$. □

s section iff **Proposition 3.4.12.** *Let M be a smooth manifold, and let S be an embedded Lagrangian submanifold of the total space of T^*M.*

(a) If S is transverse to the fiber of T^*M at a point $q \in T^*M$, then there exist a neighborhood V of q in S and a neighborhood U of $\pi(q)$ in M such that V is the image of a smooth closed 1-form defined on U.

(b) S is the image of a globally defined smooth closed 1-form on M if and only if S intersects each fiber transversely in exactly one point.

Proof. If S is transverse to the fiber of T^*M at a point $q \in T^*M$, then $d(\pi)_q : T_qS \to T_{\pi(q)}M$ is an isomorphism. Therefore $\pi|_S$ restricts to a diffeomorphism from a neighborhood V of q in S to a neighborhood U of $\pi(q)$. Then V is the graph of $\sigma = \rho \circ (\pi)^{-1}$, where ρ is the projection to the fiber.

If S intersects each fiber transversely in exactly one point, then the projection $\pi|_S$ is bijective and a submersion, hence a diffeomorphism. Therefore $(\pi|_S)^{-1}$ is a closed 1-form whose image is S. □

3.4.3 The Darboux theorem

Our next theorem is one of the most fundamental results in symplectic geometry. It is a nonlinear analogue of the canonical form for a symplectic tensor given in Proposition 3.4.6. It illustrates the most dramatic difference between symplectic structures and Riemannian metrics: unlike the Riemannian case, there is no local obstruction to a symplectic structure being locally equivalent to the standard flat model.

Theorem 3.4.13 (Darboux). *Let (M,ω)be a $2n$-dimensional symplectic manifold. For any $p \in M$, there are smooth coordinates $(x^1,\dots,x^n,y^1,\dots,y^n)$ centered at p in which ω has the coordinate representation*

$$\omega = \sum_{i=1}^{n} x^i \wedge y^i. \tag{4.4}$$

We will prove the theorem below. Any coordinates satisfying (4.4) are called Darboux coordinates, symplectic coordinates, or canonical coordinates. Obviously, the standard coordinates $(x^1,\dots,^n,y^1,\dots,y^n)$ on $\mathbb{R}^{2n}$ are Darboux coordinates. The proof of Proposition 3.4.10 showed that the natural coordinates (x^i,ξ_i) are Darboux coordinates for T^*M with its canonical symplectic structure.

First, recall that Proposition **??** shows how to use Lie derivatives to compute the derivative of a tensor field under a flow.We need the following generalization of that result to the case of time-dependent flows.

Proposition 3.4.14. *Let M be a smooth manifold. Suppose $V : J \times M \to TM$ is a smooth time-dependent vector field and $\psi : \mathcal{E} \to M$ is its time-dependent flow. For any smooth covariant tensor field A on M and any $(t_1,t_0,p) \in \mathcal{E}$,*

$$\frac{d}{dt}\Big|_{t=t_1} (\psi^*_{t,t_0} A)_p = \big(\psi^*_{t_1,t_0}(\mathfrak{L}_{V_{t_1}} A)\big)_p. \tag{4.5}$$

Proof. First, assume $t_1 = t_0$. In this case, ψ_{t_0,t_1} is the identity map of M. so we need to prove

$$\frac{d}{dt}\Big|_{t=t_0} (\psi^*_{t,t_0} A)_p = (\mathfrak{L}_{V_{t_1}} A)_p. \tag{4.6}$$

We begin with the special case in which $A = f$ is a smooth 0-tensor field:

$$\frac{d}{dt}\Big|_{t=t_0} (\psi^*_{t,t_0} f)(p) = \frac{\partial}{\partial t}\Big|_{t=t_0} (f(\psi(t,t_0,p))) = V(t_0,\psi(t_0,t_0,p))f = (\mathfrak{L}_{V_{t_0}} f)(p).$$

Next consider an exact 1-form $A = df$. In any smooth local coordinates (x^i), the function $\psi^*_{t,t_0} f(x) = f(\psi(t,t_0,x))$ depends smoothly on all $n+1$ variables $(t,x^1,\dots,x^n)$. Thus, the operator ∂/∂ commutes with each of the partial derivatives $\partial/\partial x^i$ when applied to $\psi^*_{t,t_0} f$. In particular, this means that the exterior derivative operator d commutes with $\partial/\partial t$, and so by Corollary **??**

$$\frac{d}{dt}\Big|_{t=t_0} (\psi^*_{t,t_0} df)(p) = \frac{\partial}{\partial t}\Big|_{t=t_0} d(\psi^*_{t,t_0} f)_p = d\Big(\frac{\partial}{\partial t}\Big|_{t=t_0} (\psi^*_{t,t_0} f)\Big)_p = d(\mathfrak{L}_{V_{t_0}} f)_p = (\mathfrak{L}_{V_{t_0}} df)_p.$$

Thus, the result is proved for 0-tensors and for exact 1-forms.

Now suppose that $A = B \otimes C$ for some smooth covariant tensor fields B and C, and assume that the proposition is true for B and C. (We include the possibility that B or C has rank 0, in which case the tensor product is just ordinary multiplication.) By the product rule for Lie derivatives (Proposition **??**(c)), the right-hand side of (4.6) satisfies

$$(\mathfrak{L}_{V_{t_0}}(B\otimes C))_p = (\mathfrak{L}_{V_{t_0}}B)_p \otimes C_p + B_p \otimes (\mathfrak{L}_{V_{t_0}}C)_p.$$

On the other hand, by an argument entirely analogous to that in the proof of Proposition **??**, the left-hand side satisfies a similar product rule:

$$\frac{d}{dt}\bigg|_{t=t_0}(\psi^*_{t,t_0}(B\otimes C))_p = \left(\frac{d}{dt}\bigg|_{t=t_0}(\psi^*_{t,t_0}B)\right)_p \otimes C_p + B_p \otimes \left(\frac{d}{dt}\bigg|_{t=t_0}(\psi^*_{t,t_0}C)\right)_p.$$

This shows that (4.6) holds for $A = B\otimes C$, provided it holds for B and C. The case of arbitrary tensor fields now follows by induction, using the fact that any smooth covariant tensor field can be written locally as a sum of tensor fields of the form $A = f\,dx^{i_1}\otimes\cdots\otimes dx^{i_k}$.

To handle arbitrary t_1, we use Theorem **??**(d), which shows that $\psi_{t,t_0} = \psi_{t,t_1}\circ\psi_{t_1,t_0}$ wherever the right-hand side is defined. Therefore, because the linear map $d(\psi_{t_1,t_0})^*_p$ does not depend on t,

$$\begin{aligned}\frac{d}{dt}\bigg|_{t=t_1}(\psi^*_{t,t_0}A)_p &= \frac{d}{dt}\bigg|_{t=t_1} d(\psi_{t,t_0})^*_p(A_{\psi_{t,t_0}(p)}) = \frac{d}{dt}\bigg|_{t=t_1} d(\psi_{t_1,t_0})^*_p \circ d(\psi_{t,t_1})^*_{\psi_{t_1,t_0}(p)}(A_{\psi_{t,t_0}(p)})\\ &= d(\psi_{t_1,t_0})^*_p\left(\frac{d}{dt}\bigg|_{t=t_1} d(\psi_{t,t_1})^*_{\psi_{t_1,t_0}(p)}(A_{\psi_{t,t_1}\circ\psi_{t_1,t_0}(p)})\right) = \big(\psi^*_{t_1,t_0}(\mathfrak{L}_{V_{t_1}}A)\big)_p.\end{aligned}$$

This finishes the proof. □

A **smooth time-dependent tensor field** on a smooth manifold M is a smooth map $A : J\times M \to T^kT^*M$, where $J\subseteq\mathbb{R}$ is an interval, satisfying $A(t,p)\in T^k(T^*_pM)$ for each $(t,p)\in J\times M$.

Proposition 3.4.15. *Let M be a smooth manifold and $J\subseteq\mathbb{R}$ be an open interval. Suppose $V : J\times M\to TM$ is a smooth time-dependent vector field on M, $\psi:\mathcal{E}\to M$ is its time-dependent flow, and $A : J\times M\to T^kT^*M$ is a smooth time-dependent tensor field on M. Then for any $(t_1,t_0,p)\in\mathcal{E}$,*

$$\frac{d}{dt}\bigg|_{t=t_1}(\psi^*_{t,t_0}A_t)_p = \left(\psi^*_{t_1,t_0}\left(\mathfrak{L}_{V_{t_1}}A_{t_1} + \frac{d}{dt}\bigg|_{t=t_1}A_t\right)\right)_p. \tag{4.7}$$

Proof. For sufficiently small $\varepsilon>0$, consider the smooth map $F : (t_1-\varepsilon,t_1+\varepsilon)\times(t_1-\varepsilon,t_1+\varepsilon)\to T^k(T^*_pM)$ defined by

$$F(u,v) = (\psi^*_{u,t_0}A_v)_p = d(\psi_{u,t_0})^*_p(A_v|_{\psi_{u,t_0}(p)}).$$

Since F takes its values in the finite-dimensional vector space $T^k(T^*_pM)$, we can apply the chain rule together with Proposition 3.4.14 to conclude that

$$\frac{d}{dt}\bigg|_{t=t_1}F(t,t) = \frac{\partial F}{\partial u}(t_1,t_1) + \frac{\partial F}{\partial v}(t_1,t_1) = (\psi^*_{t_1,t_0}(\mathfrak{L}_{V_{t_1}}A_{t_1}))_p + \frac{\partial}{\partial v}\bigg|_{t=t_1} d(\psi^*_{t_1,t_0})_p(A_v|_{\psi_{u,t_0}(p)}).$$

Just as in the proof of Proposition 3.4.14, the linear map $d(\psi^*_{t_1,t_0})_p$ commutes past $\partial/\partial v$, yielding (4.7). □

Proof of the Darboux theorem. Let ω_0 denote the given symplectic form on M, and let $p_0\in M$ be arbitrary. The theorem will be proved if we can find a smooth coordinate chart (U_0,φ) centered at p_0 such that $\varphi^*\omega_0=\omega_1$, where $\omega_1=\sum_{i=1}^n dx^i\wedge dy^i$ is the standard symplectic form on $\mathbb{R}^{2n}$. Since this is a local question, by choosing smooth coordinates $(x^1,\dots,x^n,y^1,\dots,y^n)$ in a neighborhood of p_0, we may replace M with an open ball $U\subseteq\mathbb{R}^{2n}$. Proposition 3.4.6 shows that we can arrange by a linear change of coordinates that $\omega_0|_{p_0}=\omega_1|_{p_0}$.

Let $\eta = \omega_1 - \omega_0$. Because ω is closed, the Poincaré lemma shows that we can find a smooth 1-form α on U such that $d\alpha = -\eta$. By subtracting a constant-coefficient (and thus closed) 1-form from α, we may assume without loss of generality that $\alpha_{p_0} = 0$.

For each $t \in \mathbb{R}$, define a closed 2-form ω_t on U by

$$\omega_t = \omega_0 + t\eta = (1-t)\omega_0 + t\omega_1.$$

Let J be a bounded open interval containing $[0,1]$. Because $\omega_t|_{p_0} = \omega_0|_{p_0}$ is nondegenerate for all t, a simple compactness argument shows that there is some neighborhood $U_1 \subseteq U$ of p_0 such that ω_t is nondegenerate on U_1 for all $t \in \bar{J}$. Because of this nondegeneracy, the smooth bundle homomorphism $\widehat{\omega}_t : TU_1 \to T^*U_1$ defined by $\widehat{\omega}_t(X) = X \lrcorner\, \omega_t$ is an isomorphism for each $t \in \bar{J}$.

Define a smooth time-dependent vector field $V : J \times U_1 \to TU_1$ by $V_t = \widehat{\omega}_t^{-1}\alpha$, or equivalently

$$V_t \lrcorner\, \omega_t = \alpha.$$

Our assumption that $\alpha_{p_0} = 0$ implies that $V_t|_{p_0} = 0$ for each t. If $\psi : \mathcal{E} \to U_1$ denotes the time-dependent flow of V, it follows that $\psi(t, 0, p_0) = p_0$ for all $t \in J$, so $J \times \{0\} \times \{p_0\} \subseteq \mathcal{E}$. Because E is open in $J \times J \times M$ and $[0,1]$ is compact, there is some neighborhood U_0 of p_0 such that $[0,1] \times \{0\} \times U_0 \subseteq \mathcal{E}$.

For each $t_1 \in [0,1]$, it follows from Proposition 3.4.15 that `Lie der tensor time-dependent`

$$\begin{aligned}
\frac{d}{dt}\bigg|_{t=t_1} \psi_{t,0}^*\omega_t &= \psi_{t_1,0}^*\Big(\mathfrak{L}_{V_{t_1}\omega_{t_1}} + \frac{d}{dt}\bigg|_{t=t_1}\omega_t\Big) \\
&= \psi_{t_1,0}^*\big(V_{t_1} \lrcorner\, d\omega_{t_1} + d(V_{t_1} \lrcorner\, \omega_t) + \eta\big) \\
&= \psi_{t_1,0}^*(V_{t_1} \lrcorner\, 0 + d\alpha + \eta) = 0.
\end{aligned}$$

Therefore, $\psi_{t,0}^*\omega_t = \psi_{0,0}^*\omega_0 = \omega_0$ for all t. In particular, $\psi_{1,0}^*\omega_1 = \omega_0$. It follows from Theorem **??**(c) `flow fundamental thm time-dependent` that $\psi_{1,0}^*$ is a diffeomorphism onto its image, so it is a coordinate map. Because $\psi_{1,0}(p_0) = p_0$, these coordinate are centered at p_0. □

3.4.4 Hamiltonian vector fields

One of the most useful constructions on symplectic manifolds is a symplectic analogue of the gradient, defined as follows. Suppose (M, ω) is a symplectic manifold. For any smooth function $f \in C^\infty(M)$, we define the Hamiltonian vector field of f to be the smooth vector field X_f defined by

$$X_f = \widehat{\omega}^{-1}(df),$$

where $\widehat{\omega} : TM \to T^*M$ is the bundle isomorphism determined by ω. Equivalently,

$$X_f \lrcorner\, \omega = df.$$

or for any vector field Y,

$$\omega(X_f, Y) = df(Y) = Yf.$$

In any Darboux coordinates, X_f can be computed explicitly as follows. Writing

$$X_f = \sum_{i=1}^n \Big(a^i \frac{\partial}{\partial x^i} - b^i \frac{\partial}{\partial y^i}\Big)$$

for some coefficient functions (a^i, b^i) to be determined, we compute

$$X_f \lrcorner\, \omega = \sum_{j=1}^{n} \Big(a^j \frac{\partial}{\partial x^j} + b^j \frac{\partial}{\partial y^j} \Big) \lrcorner \sum_{i=1}^{n} dx^i \wedge dy^i = \sum_{i=1}^{n} (a^i dy^i - b^i dx^i).$$

On the other hand,

$$df = \sum_{i=1}^{n} \Big(\frac{\partial f}{\partial x^i} dx^i + \frac{\partial f}{\partial x^i} dy^i \Big).$$

Setting these two expressions equal to each other, we find that $a^i = \partial f / \partial y^i$ and $b^i = -\partial f / \partial x^i$, which yields the following formula for the Hamiltonian vector field of f in Darboux coordinates:

$$X_f = \sum_{i=1}^{n} \Big(\frac{\partial f}{\partial y^i} \frac{\partial}{\partial x^i} - \frac{\partial f}{\partial x^i} \frac{\partial}{\partial y^i} \Big). \tag{4.8}$$

Hamiltonian v

This formula holds, in particular, on $\mathbb{R}^{2n}$ with its standard symplectic form.

Although the definition of the Hamiltonian vector field is formally analogous to that of the gradient on a Riemannian manifold, Hamiltonian vector fields differ from gradients in some very significant ways, as the next lemma shows.

or field prop

Proposition 3.4.16 (Properties of Hamiltonian Vector Fields). *Let (M, ω) be a symplectic manifold and let $f \in C^\infty(M)$.*

(a) f is constant along each integral curve of X_f.

(b) At each regular point of f, the Hamiltonian vector field X_f is tangent to the level set of f.

Proof. Both assertions follow from the fact that

$$X_f f = df(X_f) = \omega(X_f, X_f) = 0$$

because ω is alternating. □

A smooth vector field X on M is said to be **symplectic** if ω is invariant under the flow of X. It is said to be **Hamiltonian** (or **globally Hamiltonian**) if there exists a function $f \in C^\infty(M)$ such that $X = X_f$, and **locally Hamiltonian** if each point p has a neighborhood on which X is Hamiltonian. Clearly, every globally Hamiltonian vector field is locally Hamiltonian.

vector field

Proposition 3.4.17 (Hamiltonian and Symplectic Vector Fields). *Let (M, ω) be a symplectic manifold. A smooth vector field on M is symplectic if and only if it is locally Hamiltonian. Every locally Hamiltonian vector field on M is globally Hamiltonian if and only if $H^1_{dR}(M) = 0$.*

tensor filed invariant flow iff

Proof. By Theorem **??**, a smooth vector field X is symplectic if and only if $\mathfrak{L}_X \omega = 0$. Using Cartans magic formula, we compute

$$\mathfrak{L}_X \omega = d(X \lrcorner\, \omega) + X \lrcorner\, d\omega = d(X \lrcorner\, \omega). \tag{4.9}$$

Hamiltonian a

Therefore, X is symplectic if and only if the 1-form $X \lrcorner\, \omega$ is closed. On the one hand, if X is locally Hamiltonian, then in a neighborhood of each point there is a real-valued function f such that $X = X_f$, so $X \lrcorner\, \omega = df$, which is certainly closed. Conversely, if X is symplectic, then by the Poincaré lemma each point $p \in M$ has a neighborhood U on which the closed

1-form $X \lrcorner \omega$ is exact. This means there is a smooth real-valued function f defined on U such that $X \lrcorner \omega = df$; because ω is nondegenerate, this implies that $X = X_f$ on U.

Now suppose M is a smooth manifold with $H^1_{dR}(M) = 0$. If X is a locally Hamiltonian vector field, then it is symplectic, so (4.9) shows that $X \lrcorner \omega$ is closed. The hypothesis then implies that there is a function $f \in C^\infty(M)$ such that $X \lrcorner \omega = df$. This means that $X = X_f$, so X is globally Hamiltonian. Conversely, suppose every locally Hamiltonian vector field is globally Hamiltonian. Let η be a closed 1-form, and let X be the vector field $X = \widehat{\omega}^{-1}\eta$. Then (4.9) shows that $\mathfrak{L}_X \omega = 0$, so X is symplectic and therefore locally Hamiltonian. By hypothesis, there is a global smooth real-valued function f such that $X = X_f$, and then unwinding the definitions, we find that $\eta = df$. □

A symplectic manifold (M, ω) together with a smooth function $H \in C^\infty(M)$ is called a **Hamiltonian system**. The function H is called the **Hamiltonian** of the system; the flow of the Hamiltonian vector field X_H is called its **Hamiltonian flow**, and the integral curves of X_H are called the **trajectories** or **orbits** of the system. In Darboux coordinates, formula (4.4) implies that the orbits are those curves $\gamma(t) = (x^i(t), y^i(t))$ that satisfy

$$\begin{cases} \dot{x}^i(t) = \dfrac{\partial H}{\partial y^i}(x(t), y(t)), \\ \dot{y}^i(t) = -\dfrac{\partial H}{\partial y^i}(x(t), y(t)). \end{cases} \tag{4.10}$$

These are called Hamilton's equations.

Hamiltonian systems play a central role in classical mechanics. We illustrate how they arise with a simple example.

Example 3.4.18 (The n-Body Problem). Consider n physical particles moving in space, and suppose their masses are $m_1, \ldots, m_n$. For many purposes, an effective model of such a system is obtained by idealizing the particles as points in $\mathbb{R}^3$, which we denote by $\mathbf{q}_1, \ldots, \mathbf{q}_n$. Writing the coordinates of $\mathbf{q}_k$ at time t as $(q^1_k(t), q^2_k(t), q^3_k(t))$, we can represent the evolution of the system over time by a curve in $\mathbb{R}^{3n}$:

$$q(t) = (q^1_1(t), q^2_1(t), q^3_1(t), \ldots, q^1_n(t), q^2_n(t), q^3_n(t)).$$

The **collision set** is the subset $\mathcal{C} \subseteq \mathbb{R}^{3n}$ where two or more particles occupy the same position in space:

$$\mathcal{C} = \{q \in \mathbb{R}^{3n} : \mathbf{q}_k = \mathbf{q}_l \text{ for some } k \neq l\}.$$

We consider only motions with no collisions, so we are interested in curves in the open subset $Q = \mathbb{R}^{3n} \setminus \mathcal{C}$.

Suppose the particles are acted upon by forces that depend only on the positions of all the particles in the system. (A typical example is gravitational forces.) If we denote the components of the net force on the kth particle by $(F^1_k(q), F^2_k(q), F^3_k(q))$, then Newton's second law of motion asserts that the particles' motion satisfies $m_k \ddot{\mathbf{q}}_k = \mathbf{F}_k(\mathbf{q}(t))$ for each k, which translates into the $3n \times 3n$ system of second-order ODEs

$$\begin{cases} m_k \ddot{q}^1_k(t) = F^1_k(q(t)), \\ m_k \ddot{q}^2_k(t) = F^2_k(q(t)), \\ m_k \ddot{q}^3_k(t) = F^3_k(q(t)), \end{cases}$$

for $k = 1, \ldots, n$.

This can be written in a more compact form if we relabel the $3n$ position coordinates as $q(t) = (q^1(t), \ldots, q^{3n}(t))$ and the $3n$ components of the forces as $F(q) = (F_1(q), \ldots, F_{3n}(q))$, and let $M = (M_{ij})$ denote the $3n \times 3n$ diagonal matrix $\operatorname{diag}(m_1, m_1, m_1, \ldots, m_n, m_n, m_n)$. Then Newton's second law can be written

$$M_{ij}\ddot{q}^j(t) = F_i(q(t)). \tag{4.11}$$

We assume that the forces depend smoothly on q, so we can interpret $F(q) = (F_1(q), \ldots, F_{3n}(q))$ as the components of a smooth covector field on Q. We assume further that the forces are conservative, which by Proposition **??** is equivalent to the existence of a smooth function $V \in C^\infty(Q)$ (called the potential energy of the system) such that $F = -dV$.

Under the physically reasonable assumption that all of the masses are positive, the matrix M is positive definite, and thus can be interpreted as a (constant-coefficient) Riemannian metric on Q. It therefore defines a smooth bundle isomorphism $\widehat{M} : TQ \to T^*Q$. If we denote the natural coordinates on TQ by (q^i, v^i) and those on T^*Q by (q^i, p_i), then $M(v, w) = M_{ij}v^i w^j$, and $\widehat{M}$ has the coordinate representation

$$(q^i, p_i) = \widehat{M}(q^i, v^i) = (q^i, M_{ij}v^j).$$

If $q'(t) = (\dot{q}^1(t), \ldots, \dot{q}^{3n}(t))$ is the velocity vector of the system of particles at time t, then the covector $p(t) = \widehat{M}(q'(t))$ is given by the formula

$$p_i(t) = M_{ij}\dot{q}^j(t). \tag{4.12}$$

To give this equation a physical interpretation, we can revert to our original labeling of the coordinates and write

$$p(t) = (p_1^1(t), p_1^2(t), p_1^3(t), \ldots, p_n^1(t), p_n^2(t), p_n^3(t)),$$

and then $\mathbf{p}_k(t) = (p_k^1(t), p_k^2(t), p_k^3(t)) = m_k\dot{\mathbf{q}}_k(t)$ is interpreted as the momentum of the k-th particle at time t.

Using (4.12), we see that a curve $q(t)$ in Q satisfies Newton's second law (4.11) if and only if the curve $\gamma(t) = (q(t), p(t))$ in T^*Q satisfies the first-order system of ODEs

$$\begin{cases} \dot{q}^i(t) = M^{ij}p_j(t), \\[1ex] \dot{p}^i(t) = -\dfrac{\partial V}{\partial q^i}(q(t)), \end{cases} \tag{4.13}$$

where (M^{ij}) is the inverse of the matrix of (M_{ij}). Define a function $E \in C^\infty(T^*Q)$, called the **total energy** of the system, by

$$E(q, p) = V(q) + K(p),$$

where V is the potential energy introduced above, and K is the **kinetic energy**, defined by

$$K(p) = \frac{1}{2}M^{ij}p_i p_j.$$

Since (q^i, p_j) are Darboux coordinates on T^*Q, a comparison of (4.13) with (4.10) shows that (4.13) is precisely Hamilton's equations for the Hamiltonian flow of E. The fact that E is constant along the trajectories of its own Hamiltonian flow is known as the **law of conservation of energy**.

Poisson brackets

Hamiltonian vector fields allow us to define an operation on real-valued functions on a symplectic manifoldM similar to the Lie bracket of vector fields. Given $f, g \in C^\infty(M)$, we define their Poisson bracket $\{f, g\} \in C^\infty(M)$ by any of the following equivalent formulas:

$$\{f,g\} = \omega(X_f, X_g) = df(X_g) = X_g f. \tag{4.14}$$ Possion brack

Two functions are said to Poisson commute if their Poisson bracket is zero.

The geometric interpretation of the Poisson bracket is evident from the characterization $\{f, g\} = X_g f$: it is a measure of the rate of change of f along the Hamiltonian flow of g. In particular, f and g Poisson commute if and only if f is constant along the Hamiltonian flow of g.

Hamiltonian vector field
Using (4.8), we can readily compute the Poisson bracket of two functions f, g in Darboux coordinates:

$$\{f,g\} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\frac{\partial g}{\partial y^i} - \frac{\partial f}{\partial y^i}\frac{\partial g}{\partial x^i}.$$

Proposition 3.4.19 (Properties of the Poisson Bracket). *Suppose (M, ω) is a symplectic manifold, and $f, g \in C^\infty(M)$.*

(a) Bilinearity: $\{f, g\}$ is linear over $\mathbb{R}$ in f and g.

(b) Antisymmetry: $\{f, g\} = -\{g, f\}$.

(c) Jacobi identity: $\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0$.

(d) $X_{\{f,g\}} = -[X_f, X_g]$.

Proof. Parts (a) and (b) are obvious from the characterization $\{f, g\} = \omega(X_f, X_g)$ together with the fact that $X_f = \widehat{\omega}^{-1}(df)$ depends linearly on f. Because of the nondegeneracy of ω, to prove (d), it suffices to show that the following holds for every vector field Y:

$$\omega(X_{\{f,g\}}, Y) + \omega([X_f, X_g], Y) = 0. \tag{4.15}$$ Possion brack

On the one hand, note that $\omega(X_{\{f,g\}}, Y) = d(\{f, g\})(Y) = Y\{f, g\} = YX_g f$. On the other hand, because Hamiltonian vector fields are symplectic, the Lie derivative formula of Corollary
Lie derivative tensor
?? yields

$$\begin{aligned}
0 = (\mathfrak{L}_{X_g}\omega)(X_f, Y) &= X_g(\omega(X_f, Y)) - \omega([X_g, X_f], Y) - \omega(X_f, [X_g, Y]) \\
&= X_g(df(Y)) + \omega([X_f, X_g], Y) - df([X_g, Y]) \\
&= X_g Y f + \omega([X_f, X_g], Y) - [X_g, Y]f \\
&= \omega([X_f, X_g], Y) + YX_g f.
\end{aligned}$$

Therefore (d) follows.

Possion bracket
Finally, (c) follows from (d), (b), and (4.14):

$$\begin{aligned}
\{f, \{g, h\}\} = X_{\{g,h\}} f = -[X_g, X_h]f &= -X_g X_h f + X_h X_g f \\
&= -X_g\{f, h\} + X_h\{f, g\} = -\{\{f, h\}, g\} + \{\{f, g\}, h\} \\
&= -\{g, \{h, f\}\} - \{h, \{f, g\}\}.
\end{aligned}$$

□

The following corollary is immediate.

Corollary 3.4.20. *If (M, ω) is a symplectic manifold, the vector space $C^\infty(M)$ is a Lie algebra under the Poisson bracket.*

If (M, ω, H) is a Hamiltonian system, any function $f \in C^\infty(M)$ that is constant on every integral curve of X_H is called a **conserved quantity** of the system. Conserved quantities turn out to be deeply related to symmetries, as we now show.

A smooth vector field V on M is called an **infinitesimal symmetry** of (M, ω, H) if both ω and H are invariant under the flow of V.

Proposition 3.4.21. *Let (M, ω, H) be a Hamiltonian system.*

(a) A function $f \in C^\infty(M)$ is a conserved quantity if and only if $\{f, H\} = 0$.

(b) The infinitesimal symmetries of (M, ω, H) are precisely the symplectic vector fields V that satisfy $VH = 0$.

(c) If θ is the flow of an infinitesimal symmetry and γ is a trajectory of the system, then for each $s \in \mathbb{R}$, $\theta_s \circ \gamma$ is also a trajectory on its domain of definition.

Proof. Since the Possion bracket $\{f, H\}$ measures the rate of change of f along the Hamiltonian flow of H, it is clear that f is conserved if and only if $\{f, H\} = 0$. This proves (a). For (b), ω is invairant under the flow of V if and only if V is symplectic, and H is invairant under the flow of V if and only if $VH = 0$. Therefore (b) is clear.

Finally, for (c), γ is a trajectory of the system means $\gamma'(t) = X_H|_{\gamma(t)}$. Since θ is a flow of an infinitesimal symmetry, ω and H are invairant under V. This then implies

$$(\theta_s \circ \gamma)'(t) = d(\theta_s)_{\gamma(t)}(\gamma'(t)) = d(\theta_s)_{\gamma(t)}(V_{\gamma(t)}) = V_{\theta_s \circ \gamma(t)}.$$

Therefore $\theta_s \circ \gamma$ is still a trajectory of the system. □

The following theorem, first proved (in a somewhat different form) by Emmy Noether in 1918, has had a profound influence on both physics and mathematics. It shows that for many Hamiltonian systems, there is a one-to-one correspondence between conserved quantities (modulo constants) and infinitesimal symmetries.

Theorem 3.4.22 (Noether's Theorem). *Let (M, ω, H) be a Hamiltonian system. If f is any conserved quantity, then its Hamiltonian vector field is an infinitesimal symmetry. Conversely, if $H^1_{dR}(M) = 0$, then each infinitesimal symmetry is the Hamiltonian vector field of a conserved quantity, which is unique up to addition of a function that is constant on each component of M.*

Proof. Suppose f is a conserved quantity. Proposition 22.21 shows that $\{f, H\} = 0$. This in turn implies that $X_f H = \{H, f\} = 0$, so H is constant along the flow of X_f. Since ω is invariant along the flow of any Hamiltonian vector field by Proposition 3.4.17, this shows that X_f is an infinitesimal symmetry.

Now suppose that M is a smooth manifold with $H^1_{dR}(M) = 0$. Let V be an infinitesimal symmetry of (M, ω, H). Then V is symplectic by definition, and globally Hamiltonian by Proposition 3.4.17. Writing $V = X_f$, the fact that H is constant along the flow of V implies that $\{H, f\} = X_f H = VH = 0$, so f is a conserved quantity. If $\widetilde{f}$ is any other function that satisfies $X_{\widetilde{f}} = X_f$, then $d(\widetilde{f} - f) = \widehat{\omega}(X_{\widetilde{f}} - X_f) = 0$, so $\widetilde{f} - f$ must be constant on each component of M. □

There is one conserved quantity that every Hamiltonian system possesses: the Hamiltonian H itself. The infinitesimal symmetry corresponding to it, of course, generates the Hamiltonian flow of the system, which describes how the system evolves over time. Since H is typically interpreted as the total energy of the system, one usually says that the symmetry corresponding to conservation of energy is "translation in the time variable".

Hamiltonian Flowouts

Hamiltonian vector fields are powerful tools for constructing isotropic and Lagrangian submanifolds. Because Lagrangian submanifolds of T^*M correspond to closed 1-forms (Proposition 3.4.11), which in turn correspond locally to differentials of functions, such constructions have numerous applications in PDE theory.

Theorem 3.4.23 (Hamiltonian Flowout Theorem). *Suppose (M, ω) is a symplectic manifold, $H \in C^\infty(M)$, Γ is an embedded isotropic submanifold of M that is contained in a single level set of H, and the Hamiltonian vector field X_H is nowhere tangent to Γ. If $\mathcal{S}$ is a flowout from Γ along X_H, then $\mathcal{S}$ is also isotropic and contained in the same level set of H.*

Proof. Let θ be the flow of X_H. Recall from Theorem **??** that the flowout is parametrized by the restriction of θ to a neighborhood $\mathcal{O}_\delta$ of $\{0\} \times \Gamma$ in $\mathbb{R} \times \Gamma$. First consider a point $p \in \Gamma \subseteq \mathcal{S}$. If we choose a basis $E_1, \dots, E_k$ for $T_p\Gamma$, then $T_p\mathcal{S}$ is spanned by $(X_H|_p, E_1, \dots, E_k)$. The assumption that Γ is isotropic implies that $\omega_p(E_i, E_j) = 0$ for all i and j. On the other hand, by definition of the Hamiltonian vector field,

$$\omega_p(X_H|_p, E_j) = dH_p(E_j) = 0.$$

because E_j is tangent to Γ, which is contained in a level set of H. This shows that the restriction of ω to $T_p\mathcal{S}$ is zero when $p \in \Gamma$.

Any other point $p;' \in \mathcal{S}$ is of the form $p' = \theta_t(p)$ for some $(t, p) \in \mathcal{O}_\delta$. Because θ_t is a local diffeomorphism that maps a neighborhood of p in $\mathcal{S}$ to a neighborhood of p' in $\mathcal{S}$, its differential takes $T_p\mathcal{S}$ isomorphically onto $T_{p'}\mathcal{S}$. Thus, for any vectors $v, w \in T_{p'}\mathcal{S}$, there are vectors $\widehat{v}, \widehat{w}$ such that $d(\theta_t)_p(\widehat{v}) = v$ and $d(\theta_t)_p(\widehat{w}) = w$. Moreover, because X_H is a symplectic vector field, its flow preserves ω. Therefore,

$$\omega_{p'}(v, w) = \omega_{p'}(d(\theta_t)_p(\widehat{v}), d(\theta_t)_p(\widehat{w})) = (\theta_t^*\omega)_p(\widehat{v}, \widehat{w}) = \omega_p(\widehat{v}, \widehat{w}) = 0.$$

It follows that $\mathcal{S}$ is isotropic. By Proposition 3.4.16, H is constant along each integral curve of X_H, so $\mathcal{S}$ is contained in the same level set of H as Γ. □

3.4.5 Contact structures

As we have seen, symplectic manifolds must be even-dimensional; but there is a closely related structure called a contact structure that one can define on odd-dimensional manifolds. It also has important applications in geometry and analysis. In this part, we introduce the main elements of contact geometry.

Suppose M is a smooth manifold of odd dimension $2n+1$. A contact form on M is a nonvanishing smooth 1-form θ with the property that for each $p \in M$, the restriction of $d\theta_p$ to the subspace $\ker \theta_p \subseteq T_pM$ is nondegenerate, which is to say it is a symplectic tensor. A **contact structure** on M is a smooth distribution $H \subseteq TM$ of rank $2n$ with the property that any smooth local defining form θ for H is a contact form (see Lemma **??**). A contact manifold is a smooth manifold M together with a contact structure on M. If (M, H) is a contact manifold, any (local or global) defining form for H is called a **contact form** for H.

Proposition 3.4.24. *A smooth 1-form θ on a $(2n+1)$-manifold is a contact form if and only if $\theta \wedge (d\theta)^n$ is nonzero everywhere on M, where $(d\theta)^n$ represents the n-fold wedge product of $d\theta$.*

Proof. By Lemma **??**, $(d\theta)^n|_p$ annihilates $\ker \theta_p$ if and only if there is a $(2n-1)$-form α_p such that

$$(d\theta)^n|_p = \theta|_p \wedge \alpha_p.$$

This is, in turn, equivalent to $\theta_p \wedge (d\theta)^n|_p = 0$. □

Proposition 3.4.25. *Suppose H is a contact structure on a smooth manifold M. Show that if θ_1 and θ_2 are any two local contact forms for H, then on their common domain there is a smooth nonvanishing function f such that $\theta_2 = f\theta_1$.*

Proof. This follows from the fact that two 1-forms defines the same kernel if and only if they differ by a constant. □

It follows from the result of Proposition **??** that a codimension-1 distribution $H \subseteq TM$ is integrable if and only if any local defining form θ satisfies $\theta \wedge d\theta = 0$. If H is a contact structure, by contrast, not only is nonzero everywhere on M, but it remains nonzero after taking $n-1$ more wedge products with $d\theta$. Thus, a contact structure is, in a sense, a "maximally nonintegrable distribution".

Example 3.4.26 (Contact Forms).

(a) On $\mathbb{R}^{2n+1}$ with coordinates $(x^1, \ldots, x^n, y^1, \ldots, y^n, z)$, define a 1-form θ by

$$\theta = dz - \sum_{i=1}^{n} y^i dx^i. \tag{4.16}$$

and let $H \subseteq T\mathbb{R}^{2n+1}$ be the rank-$2n$ distribution annihilated by θ. Then $d\theta = \sum_{i=1}^{n} dx^i \wedge dy^i$. If we define vector fields X_i, Y_i by

$$X_i = \frac{\partial}{\partial x^i} + y^i \frac{\partial}{\partial z}, \quad Y^i = \frac{\partial}{\partial y^i}.$$

then (X_i, Y_i) is a smooth frame for H, and it is straightforward to check that it satisfies $d\theta(X_i, X_j) = d\theta(Y_i, Y_j) = 0$ and $d\theta(X_i, Y_j) = \delta_{ij}$. It follows just as in Example 3.4.2 that $d\theta|_H$ is nondegenerate, so θ is a contact form.

(b) Let M be a smooth n-manifold, and define a smooth 1-form θ on the $(2n+1)$-manifold $\mathbb{R} \times T^*M$ by $\theta = dz - \tau$, where z is the standard coordinate on $\mathbb{R}$ and τ is the tautological 1-form on T^*M. In terms of natural coordinates (x^i, ξ_i) for T^*M; the form θ has the coordinate representation

$$\theta = dz - \sum_{i=1}^{n} \xi_i dx^i.$$

so it is a contact form by the same argument as in part (a).

(c) On $\mathbb{R}^{2n+2}$ with coordinates $(x^1, \dots, x^{n+1}, y^1, \dots, y^n)$, consider the 1-form

$$\Theta = \sum_{i=1}^{n+1} (x^i dy^i - y^i dx^i).$$

The **standard contact form on S^{2n+1}** is the smooth 1-form $\theta = \iota^*\Theta$, where $\iota : S^{2n+1} \to \mathbb{R}^{2n+2}$ is inclusion. To see that θ is indeed a contact form, note first that $d\theta = \sum_{i=1}^{n+1}(dx^i \wedge dy^i)$ is a symplectic form on $\mathbb{R}^{2n+2}$. Consider the following vector fields on $\mathbb{R}^{2n+2} \setminus \{0\}$:

$$N = x^i \frac{\partial}{\partial x^i} + y^i \frac{\partial}{\partial y^i}, \quad T = x^i \frac{\partial}{\partial y^i} - y^i \frac{\partial}{\partial x^i}.$$

A computation shows that N is normal to S^{2n+1} (with respect to the Euclidean metric) and T is tangent to it. Let $S \subseteq T(\mathbb{R}^{2n+2} \setminus \{0\})$ denote the subbundle spanned by N and T, and let $S^\perp$ denote its symplectic complement with respect to $d\Theta$. For each $p \in S^{2n+1}$, $S_p^\perp$ is the set of vectors $X \in T_p\mathbb{R}^{2n+2}$ such that $d\Theta_p(N_p, X_p) = d\Theta_p(T_p, X_p) = 0$. We compute

$$N \lrcorner\, d\Theta = 2\sum_{i=1}^{n+1}(x^i dy^i - y^i dx^i) = 2\Theta,$$

$$T \lrcorner\, d\Theta = 2\sum_{i=1}^{n+1}(x^i dx^i + y^i dy^i) = d(|x|^2 + |y|^2).$$

It follows that $S_p^\perp$ is the common kernel of Θ_p and $d(|x|^2 + |y|^2)_p$, which is $\ker \Theta_p \cap T_pS^{2n+1} = \ker \theta_p$. Because $d\Theta(N, T) = |x|^2 + |y|^2 \neq 0$, S_p is a symplectic subspace of $T_p\mathbb{R}^{2n+2}$, and thus $\ker \theta_p = S_p^\perp$ is also a symplectic subspace by Proposition 3.4.5(a). Because the restriction of $d\theta_p$ to $\ker \theta_p$ is the same as the restriction of $d\Theta_p$, it is nondegenerate, so θ is a contact form.

Theorem 3.4.27 (The Reeb Field). *Let (M, H) be a contact manifold, and suppose θ is a contact form for H. There is a unique vector field $T \in \mathfrak{X}(M)$, called the Reeb field of θ, that satisfies the following two conditions:*

$$T \lrcorner\, d\theta = 0, \quad \theta(T) = 1. \tag{4.17}$$

Proof. Define a smooth bundle homomorphism $\Phi : TM \to T^*M$ by $\Phi(X) = X \lrcorner\, d\theta$, and for each $p \in M$, let Φ_p denote the linear map $\Phi|_{T_pM} : T_pM \to T_p^*M$. The fact that $d\theta_p$ restricts to a nondegenerate 2-tensor on H_p implies that $\Phi_p|_{H_p}$ is injective, so Φ_p has rank at least $2n$ (where $2n+1$ is the dimension of M). On the other hand, Φ_p cannot have rank $2n+1$, because then $d\theta_p$ would be nondegenerate, which is impossible because T_pM is odd-dimensional. Thus Φ_p has rank exactly $2n$, so $\dim \ker \Phi_p = 1$. Since $\ker \Phi_p$ is not contained in $H_p = \ker \theta_p$, there is a unique vector $T_p \in \ker \Phi_p$ satisfying $\theta_p(T_p) = 1$. This shows that there is a unique rough vector field T satisfying (4.17).

To see that T is smooth, note that $\ker \Phi$ is a smooth rank-1 subbundle of TM by Theorem **??**. Given $p \in M$, let X be any smooth nonvanishing section of $\ker \Phi$ on a neighborhood of p. Because $\theta(X) \neq 0$, we can write the Reeb field locally as $T = X/\theta(X)$, which is also smooth. □

Remark 3.4.1. Let θ be a contact form and T be its Reeb field. Note that

$$\mathfrak{L}_T\theta = d(T \lrcorner\, \theta) + T \lrcorner\, d\theta = d(\theta(T)) + T \lrcorner\, d\theta = 0.$$

Therefore θ is invairant under the flow of T.

contact forms eq

Example 3.4.28. The Reeb fields of the contact forms of Example 3.4.26(a),(b) are given by $\partial/\partial z$, and that of (c) is given by

$$T = \left(x^i \frac{\partial}{\partial y^i} - y^i \frac{\partial}{\partial x^i} \right)\Big|_{S^{2n+1}}.$$

Many of the constructs that we described for symplectic manifolds have analogues in contact geometry. We begin with an analogue of the Darboux theorem.

Theorem 3.4.29 (Contact Darboux Theorem). *Suppose θ is a contact form on a $(2n+1)$-dimensional manifold M. For each $p \in M$, there are smooth coordinates $(x^1, \ldots, x^n, y^1, \ldots, y^n, z)$ centered at p in which θ has the form*

$$\theta = dz - \sum_{i=1}^{n} y^i dx^i.$$

Proof. Let $p \in M$ be arbitrary. Let $(U, (u^i))$ be a smooth coordinate cube centered at p in which the Reeb field of θ has the form $T = \partial/\partial u^1$, and let $Y \subseteq U$ be the slice defined by $u^1 = 0$. Because T is nowhere tangent to Y, it follows that the pullback of $d\theta$ to Y is a symplectic form. After shrinking U and Y if necessary, we can find Darboux coordinates $(x^1, \ldots, x^n, y^1, \ldots, y^n)$ for Y centered at p, and extend them to U by requiring them to be independent of u^1 (or equivalently, to be constant on the integral curves of T). Let α be the 1-form $\sum_i y^i dx^i$ on U, so the pullbacks of $d\theta$ and $-d\alpha$ to Y agree. Because $T \lrcorner\, d\theta = T \lrcorner\, d\alpha = 0$, it follows that $d\theta + d\alpha = 0$ at points of Y. Then $\mathfrak{L}_T\theta = \mathfrak{L}_T\alpha = 0$ implies that $d(\theta + \alpha)$ is invariant under the flow of T, so in fact $d(\theta + \alpha) = 0$ on all of U. By the Poincaré lemma, there is a smooth function z on U such that $dz = \theta + \alpha$ by subtracting a constant, we may arrange that $z(p) = 0$. Because $dz_p(T_p) = \theta_p(T_p) = 1$, it follows that $(dx^i|_p, dy^i|_p, dz|_p)$ are linearly independent, so there is a neighborhood of p on which $(x^1, \ldots, x^n, y^1, \ldots, y^n, z)$ are the coordinates we seek. $\square$

The next proposition describes a contact analogue of Hamiltonian vector fields.

vector field

Proposition 3.4.30. *Suppose (M, H) is a contact manifold and θ is a contact form for H. For any function $f \in C^\infty(M)$, there is a unique vector field X_f, called the **contact Hamiltonian vector field** of f, that satisfies $\theta(X_f) = f$ and $(X_f \lrcorner\, d\theta)|_H = -df|_H$.*

Proof. Suppose $f \in C^\infty(M)$. Because the restriction of $d\theta$ to H is nondegenerate, there is a unique smooth vector field $B \in \Gamma(H)$ such that $B \lrcorner\, d\theta|_H = -df|_H$. If we set $X_f = fT + B$, where T is the Reeb field for θ, then it is easy to check that the required conditions are satisfied. $\square$

Suppose (M, H) is a contact manifold. A smooth vector field $X \in \mathfrak{X}(M)$ is called a **contact vector field** if its flow φ preserves the contact structure, in the sense that $d(\varphi_t)_p(H_p) = H_{\varphi_t(p)}$ for all (t, p) in the domain of φ.

Theorem 3.4.31. *If (M, H) is a contact manifold and θ is a contact form for H, then a smooth vector field on M is a contact vector field if and only if it is a contact Hamiltonian vector field.*

Proof. Since H is characterized as the kernel of θ, the equality $d(\varphi_t)_p(H_p) = H_{\varphi_t(p)}$ is equivalent to

$$\theta_{\varphi_t(p)}(d(\varphi_t)_p(H_p)) = d(\varphi_t)_p^*\theta_{\varphi_t(p)}(H_p) = 0.$$

Since θ_p is zero on H_p, this is equivalent to $\mathfrak{L}_X\theta|_H = 0$. That is,

$$(X \lrcorner\, d\theta)|_H = -d(X \lrcorner\, \theta)|_H = -d(\theta(X))|_H.$$

It is clear that this condition is satisfied if X is a contact Hamiltonian vector field.

Conversely, if X is a contact vector field, then we have

$$0 = \mathfrak{L}_X\theta|_H = (X \lrcorner\, d\theta)|_H + d(X \lrcorner\, \theta)|_H = (X \lrcorner\, d\theta)|_H + d(\theta(X))|_H.$$

contact Hamiltonian vector field

Then, from the proof of Proposition 3.4.30, we see that

$$X_{\theta(X)} = \theta(X)T + X|_H$$

where $X|_H$ is the component of X in H. Since TM is generated by T and H, the right-hand side $\theta(X)T + X|_H$ is exactly X. Therefore $X = X_{\theta(X)}$ is a contact Hamiltonian vector field. □

If (M, H) is a contact manifold, a smooth submanifold $S \subseteq M$ is said to be **isotropic** if $TS \subseteq H$, or equivalently if $\iota^*\theta = 0$ for any contact form θ, where $\iota : S \hookrightarrow M$ is inclusion. If $S \subseteq M$ is isotropic, then $\iota^* d\theta = d\iota^*\theta = 0$. This implies that for each $p \in S$, the tangent space T_pS is an isotropic subspace of the symplectic vector space H_p, and thus its dimension cannot be any larger than n, where $2n+1$ is the dimension of M. An isotropic submanifold of the maximum possible dimension n is called a **Legendrian submanifold**.

The next theorem is a contact analogue of the Hamiltonian flowout theorem, and is proved in much the same way. It is the main tool for constructing solutions of fully nonlinear PDEs.

ntact flowout

Theorem 3.4.32 (Contact Flowout Theorem). *Suppose that (M, H) is a contact manifold, $f \in C^\infty(M)$, Γ is an embedded isotropic submanifold of M that is contained in a single level set of f, and the contact Hamiltonian vector field X_F is nowhere tangent to Γ. If $\mathcal{S}$ is a flowout from Γ along X_f, then $\mathcal{S}$ is also isotropic and contained in the same level set of f.*

Proof. In this case, $\iota^* T\Gamma \subseteq H$. Let $p \in \Gamma$. If $v \in T_p\Gamma$ then by defintion we have

$$d\theta_p(X_f|_p, v) = (X_f \lrcorner\, d\theta)_p(v) = -df_p(v) = 0.$$

This shows the restriction of $d\theta$ is zero on $T_p\mathcal{S}$ when $p \in \Gamma$. Since H is preserved by the flow

Hamiltonian flowout

of X_f, we can apply the method used in Theorem 3.4.23 to get the claim. □

3.4.6 Nonlinear first-order PDEs

flow section

In Section **??**, we discussed first-order partial differential equations, and showed how to use the theory of flows to solve them in the linear and quasilinear cases. In this part, we show how to use symplectic and contact geometry to solve fully nonlinear first-order equations (i.e., equations that are not quasilinear).

We begin with a somewhat special case. A first-order partial differential equation that involves only the first derivatives of the unknown function but not the values of the function itself is called a **Hamilton-Jacobi equation**. Such an equation for an unknown function

$u(x^1, \ldots, x^n)$ can be written in the form

$$F\Big(x^1, \ldots, x^n, \frac{\partial u}{\partial x^1}(x), \ldots, \frac{\partial u}{\partial x^n}(x)\Big) = 0. \tag{4.18}$$

where F is a smooth function defined on an open subset of $\mathbb{R}^{2n}$.

More generally, if M is a smooth manifold, a Hamilton-Jacobi equation on M is given by a smooth real-valued function F defined on an open subset $W \subseteq T^*M$ and a solution to the equation is a smooth real-valued function u defined on an open subset $U \subseteq M$ such that the image of du lies in the zero set of F:

$$F(x, du(x)) = 0 \text{ for all } x \in U. \tag{4.19}$$

(We write the covector $du_x \in T_xM$ as $(x, du(x))$, in order to be more consistent with the coordinate representation (4.18) of the equation.) We are interested in solving a Cauchy problem for (4.19): given an embedded hypersurface $S \subseteq M$ and a smooth function $\varphi : S \to \mathbb{R}$, we wish to find a smooth function u defined on a neighborhood of S in M and satisfying (4.19) together with the initial condition

$$u|_S = \varphi. \tag{4.20}$$

Just as in Section **??**, in order to obtain solutions we need to assume that the problem is of a type called noncharacteristic; we will describe what this means below.

Because Equation (4.19) involves only du, not u itself, we look first for a closed 1-form α satisfying $F(x, \alpha(x)) \equiv 0$; then the Poincaré lemma guarantees that locally $\alpha = du$ for some function u, which then satisfies (4.19). By Proposition 3.4.11, it suffices to construct a Lagrangian submanifold of T^*M that is the image of a 1-form and is contained in $F^{-1}(0)$. The key to finding such a submanifold is the Hamiltonian flowout theorem (Theorem 3.4.23): after identifying an appropriate isotropic embedded initial submanifold $\Gamma \subseteq T^*M$, we will construct the image of α as the flowout from Γ along the Hamiltonian field of F.

The first challenge is to construct an appropriate initial submanifold $\Gamma \subseteq T^*M$. The image of $d\varphi$ will not do, because it lies in T^*S, not T^*M (and there is no canonical way to identify T^*S as a subset of T^*M). Thus, we must first look for an appropriate section of the restricted bundle $TM|_S$, that is, a smooth map $\sigma : S \to T^*M$ such that $\sigma(x) \in T^*_xM$ for each $x \in S$. This will be the value of du along S for our eventual solution u. Thus, we should expect that it matches $d\varphi$ when restricted to vectors tangent to S, and that it satisfies the PDE at points of S. In summary, we require σ to satisfy the following conditions:

$$\sigma(x)|_{T_xS} = d\varphi(x) \qquad \text{for all } x \in S, \tag{4.21}$$

$$F(x, \sigma(x)) = 0 \qquad \text{for all } x \in S. \tag{4.22}$$

To find such a σ, at least locally, begin by extending φ to a smooth function $\widetilde{\varphi}$ in a neighborhood of S and choosing a smooth local defining function ψ for S. Since σ must agree with $d\varphi$ when restricted to TS, and the annihilator of TS at each point is spanned by $d\psi$, the only possibility for σ is a section of the form $\sigma = d\widetilde{\varphi} + f\,d\psi$ for some unknown real-valued function f defined in a neighborhood of S. You can then insert this into the equation $F(x, \sigma(x)) = 0$ and attempt to solve for the values of f along S.

The Cauchy problem (4.19)–(4.20) is said to be noncharacteristic if there exists a smooth section $\sigma \in \Gamma(T^*M|_S)$ satisfying (4.21)–(4.22), with the additional property that if (x^i) are any local coordinates on M and $(x^1, \ldots, x^n, \xi_1, \ldots, \xi_n)$ are the corresponding natural coordinates on T^*M, the following vector field along S is nowhere tangent to S:

$$A^\sigma|_x = \frac{\partial F}{\partial \xi_1}(x,\sigma(x))\frac{\partial}{\partial x^1} + \cdots + \frac{\partial F}{\partial \xi_n}(x,\sigma(x))\frac{\partial}{\partial x^n}. \tag{4.23}$$

(As we will see in the proof of the next theorem, A^σ is actually globally defined as a vector field along S, and does not depend on the choice of coordinates.) When this condition is satisfied, we can solve the Cauchy problem.

Theorem 3.4.33 (The Cauchy Problem for a Hamilton-Jacobi Equation). *Suppose that M is a smooth manifold, $W \subseteq T^*M$ is an open subset, $F : W \to \mathbb{R}$ is a smooth function, $S \subseteq M$ is an embedded hypersurface, and $\varphi : S \to \mathbb{R}$ is a smooth function. If the Cauchy problem (4.19)–(4.20) is noncharacteristic, then for each $p \in S$ there is a smooth solution defined on some neighborhood of p in M.*

Proof. Given $\sigma : S \to T^*M|_S$ satisfying (4.21)–(4.22), let $\Gamma \subseteq W$ be the image of σ. Then Γ is an embedded submanifold of dimension $n-1$, where $n = \dim M$. In order to apply the Hamiltonian flowout theorem, we need to check first that Γ is isotropic with respect to the canonical symplectic structure ω on T^*M. Since $\sigma : S \to T^*M$ is a smooth embedding whose image is Γ, this is equivalent to showing that $\sigma^*\omega = 0$. Let $\pi : T^*M \to M$ be the projection; then $\pi \circ \sigma$ is equal to the inclusion $\iota : S \hookrightarrow M$. If τ denotes the tautological 1-form on T^*M, the defining equation (4.3) for τ implies

$$(\sigma^*\tau)(p) = d\sigma_p^*(d\pi^*_{\sigma(p)}\sigma(p)) = d(\pi\circ\sigma)^*\sigma(p) = d\iota_p^*\sigma(p) = d\varphi(p).$$

Thus $\sigma^*\tau = d\varphi$, and it follows that $\sigma^*\tau = -d^2\varphi = 0$. Therefore Γ is isotropic.

Next we need to check that the Hamiltonian vector field X_F is nowhere tangent to Γ. This follows from the noncharacteristic condition just as in the proof of Theorem ??: because $\pi : T^*M \to M$ restricts to a diffeomorphism from Γ to S, if X_F were tangent to Γ at some point $(p,\sigma(p)) \in \Gamma$, then $d\pi(X_F|_{(p,\sigma(p))})$ would be tangent to S at p. Using (4.8) in natural coordinates (x^i, ξ_i) on T^*M (which are Darboux coordinates for the canonical symplectic form), we find that

$$X_F = \frac{\partial F}{\partial \xi_1}\frac{\partial}{\partial x^1} + \cdots + \frac{\partial F}{\partial \xi_n}\frac{\partial}{\partial x^n} - \frac{\partial F}{\partial x^1}\frac{\partial}{\partial \xi_1} + \cdots + \frac{\partial F}{\partial x^n}\frac{\partial}{\partial \xi_n}$$

Thus $d\pi(X_F|_{(p,\sigma(p))}) = A^\sigma|_p$, so the assumption that the Cauchy problem is noncharacteristic guarantees that X_F is nowhere tangent to Γ. (This calculation also shows that A^σ is well defined independently of coordinates, because it is the pushforward of X_F from points of Γ.)

Let $\mathcal{S}$ be a flowout from Γ along X_F. The Hamiltonian flowout theorem guarantees that $\mathcal{S}$ is an n-dimensional isotropic—and therefore Lagrangian—submanifold of T^*M contained in $F^{-1}(0)$. Using the result of Proposition 3.4.12, we conclude that it will be the image of a closed 1-form on a neighborhood of p provided that it is transverse to the fiber of π at $(p,\sigma(p))$. Once again, we use the fact that $T_{(p,\sigma(p))}\mathcal{S}$ is spanned by $T_{(p,\sigma(p))}\Gamma$ and $X_F|_{(p,\sigma(p))}$. Because $d\pi$ maps $X_F|_{(p,\sigma(p))}$ to $A^\sigma|_p$ and maps $T_{(p,\sigma(p))}\Gamma$ isomorphically onto T_pS, the noncharacteristic assumption guarantees that $T_{(p,\sigma(p))}T^*M = T_{(p,\sigma(p))}\mathcal{S} \oplus \ker d\pi_{(p,\sigma(p))}$, and thus $\mathcal{S}$ intersects the fiber transversely at $(p,\sigma(p))$. By Proposition 3.4.12, there is a closed 1-form α defined on a neighborhood U of p whose graph is an open subset of $\mathcal{S}$. Because the image of α is contained in $\mathcal{S}$, it follows that

$$\alpha(x) = \sigma(x) \text{ for all } x \in S \cap U. \tag{4.24}$$

By the Poincaré lemma, after shrinking U further if necessary, we can find a smooth function $u : U \to \mathbb{R}$ such that $du = \alpha$. Because $\mathcal{S} \subseteq F^{-1}(0)$, we conclude that u satisfies (4.19). To ensure that the initial condition is also satisfied, shrink U further so that $S \cap U$ is connected. By adding a constant to u, we may arrange that $u(p) = \varphi(p)$. Then for any $x \in S \cap U$, it follows from (4.21) and (4.24) that

$$du(x)|_{T_xS} = \alpha(x)|_{T_xS} = \sigma(x)|_{T_xS} = d\varphi(x).$$

Because $S \cap U$ is connected, this means that $u - \varphi$ is constant on $S \cap U$. Since this difference vanishes at p, it vanishes identically, so (4.20) is satisfied on $S \cap U$. □

Note that we did not claim any uniqueness in this theorem. In Cauchy problems for fully nonlinear equations, even local uniqueness can fail. For example, consider the following Cauchy problem in the plane:

$$\left(\frac{\partial u}{\partial x}\right)^2 = 1, \quad u(0, y) = 0.$$

This is noncharacteristic, as you can check. Both $u(x, y) = x$ and $u(x, y) = -x$ are solutions to this problem, but they are not equal in any open subset. The problem here is that there are two possible choices for the initial 1-form σ (namely, $\sigma = dx$ and $\sigma = -dx$), and they yield different initial manifolds σ and therefore different solutions to the Cauchy problem. However, by a similar argument as in the quasilinear case, we can see that once σ is chosen, the local uniqueness still holds.

Example 3.4.34 (A Hamilton-Jacobi Equation). Consider the following Cauchy problem in the plane:

$$\frac{\partial u}{\partial x} - \left(\frac{\partial u}{\partial y}\right)^2 = 0, \quad u(0, y) = y^2.$$

The corresponding function on $T^*\mathbb{R}^2$ is $F(x, y, \xi, \eta) = \xi - \eta^2$, where we use (x, y, ξ, η) to denote natural coordinates on $T^*\mathbb{R}^2$ associated with (x, y). To check that the problem is noncharacteristic, we need to find a suitable 1-form σ along the initial manifold $S = \{(x, y) : x = 0\}$. Since x is a defining function for S, we can write $\sigma = d(y^2) + f(y)dx = f(y)dx + 2ydy$ and solve the equation

$$F(0, y, f(y), 2y) = f(y) - (2y)^2$$

to obtain $f(y) = 4y^2$. Thus we can set $\sigma(y) = 4y^2dx + 2ydy$. The vector field A^σ is given by

$$A^\sigma = \frac{\partial}{\partial x} - 4y\frac{\partial}{\partial y}$$

which is nowhere tangent to S.

The initial curve S can be parametrized by $X(s) = (0, s)$, and therefore the initial curve $\Gamma \subseteq T^*\mathbb{R}^2$ (the image of σ) can be parametrized by $\widetilde{X}(s) = (0, s, 4s^2, 2s)$. The Hamiltonian field of F is

$$X_F|_{(x,y,\xi,\eta)} = \frac{\partial}{\partial x} - 2\eta\frac{\partial}{\partial y}.$$

and it is an easy matter to solve the corresponding system of ODEs with initial conditions $(x, y, \xi, \eta) = (0, s, 4s^2, 2s)$ to obtain the following parametrization of $\mathcal{S}$:

$$\Psi(t, s) = (t, s - 4st, 4s^2, 2s).$$

Solving (x, y) from $(x, y) = (t, s - 4st)$ and inserting into the formulas for (ξ, η), we find that

$\mathcal{S}$ is the image of the following 1-form:

$$\alpha = \frac{4y^2}{(1-4x)^2}dx + \frac{2y}{1-4x}dy.$$

This is indeed a closed 1-form, and we find that $\alpha = du$ on the set $\{(x,y) : x < 1/4\}$, where

$$u(x,y) = \frac{y^2}{1-4x}.$$

In principle, we might have to add a constant to u to satisfy the initial condition, but in this case $u(0,y) = y^2$ already, so this is the solution to our Cauchy problem.

General nonlinear equations

Finally, we show how the preceding method can be adapted to solve arbitrary first-order PDEs by using contact geometry in place of symplectic geometry. For this purpose, we introduce one last geometric construction. If M is a smooth manifold, the 1-jet bundle of M is the smooth vector bundle $J^1M = \mathbb{R} \times T^*M \to M$, whose fiber at $x \in M$ is $R \times T_x^*M$. (It is the Whitney sum of a trivial $\mathbb{R}$-bundle with T^*M.) If $u : M \to \mathbb{R}$ is a smooth function, the 1-jet of u is the section $j^1u : M \to J^1M$ defined by $j^1u(x) = (u(x), du(x))$. A point in the fiber of J^1M over $x \in M$ can be viewed as a first-order Taylor polynomial at x of a smooth function on M, represented invariantly as the values of the function and its differential at x. (One can also define higher-order jet bundles that give invariant representations of higher-order Taylor polynomials. They are useful for studying higher-order PDEs, but we do not pursue them here.)

The canonical contact form on J^1M is the 1-form $\theta = dz - \tau$ defined in Example 3.4.26(b). A smooth (local or global) section $\eta : M \to J^1M$ is said to be **Legendrian** if its image is a Legendrian submanifold of J^1M, or equivalently if $\eta^*\theta = 0$. The next proposition is a contact analogue of Proposition 3.4.11.

Proposition 3.4.35. *Let M be a smooth manifold. A smooth local section of J^1M is the 1-jet of a smooth function if and only if it is Legendrian.*

Proof. Let $\eta : M \to J^1M$ be a smooth local section. Then we can write $\eta(x) = (u(x), \sigma_x)$, then by Proposition 3.4.11 we have

$$\eta^*\theta = (u,\sigma)^*(dz - \tau) = u^*(dz) - (\sigma)^*\tau = du - \sigma.$$

Therefore η is Legendrian if and only if it is a 1-jet. □

The 1-jet bundle provides the most general setting in which to consider first-order partial differential equations. If M is a smooth manifold, a first-order PDE for a function $u : M \to \mathbb{R}$ can be viewed as a real-valued function F on the 1-jet bundle of M, and a solution is a function whose 1-jet takes its values in the zero set of F.

Let M be a smooth manifold, and suppose we are given a function $F \in C^\infty(W)$ on some open subset $W \subseteq J^1M$, a smooth hypersurface $S \subseteq M$, and a smooth function $\varphi : S \to \mathbb{R}$. We wish to solve the following Cauchy problem for u:

$$F(x, u(x), du(x)) = 0, \qquad (4.25)$$

$$u|_S = \varphi. \qquad (4.26)$$

This problem is said to be noncharacteristic if there exists a smooth section $\sigma \in \Gamma(T^*M|_S)$ taking its values in W and satisfying

$$\sigma(x)|_{T_xS} = d\varphi(x) \qquad \text{for all } x \in S, \tag{4.27}$$

$$F(x, \varphi(x), \sigma(x)) = 0 \qquad \text{for all } x \in S. \tag{4.28}$$

and such that the following vector field along S is nowhere tangent to S:

$$A^\sigma|_x = \frac{\partial F}{\partial \xi_1}(x, \varphi(x), \sigma(x))\frac{\partial}{\partial x^1} + \cdots + \frac{\partial F}{\partial \xi_n}(x, \varphi(x), \sigma(x))\frac{\partial}{\partial x^n}. \tag{4.29}$$

The proof of the next theorem is very similar to that of Theorem 3.4.33, but uses the contact flowout theorem instead of the Hamiltonian one.

Theorem 3.4.36 (The General First-Order Cauchy Problem). *Suppose M is a smooth manifold, $W \subseteq J^1M$ is an open subset, $F : W \to \mathbb{R}$ is a smooth function, $S \subseteq M$ is an embedded hypersurface, and $\varphi : S \to \mathbb{R}$ is a smooth function. If the Cauchy problem (4.25)–(4.26) is noncharacteristic, then for each $p \in S$ there is a smooth solution on some neighborhood of p in M.*