Chapter 3

Metrics and Connections

3.1 Riemannian Metrics

Finally we come to the definition of a Riemannian metric, the object that gives this field its name. Let us dispel a common misunderstanding: a Riemannian metric is not a distance function, which goes against modern terminology (a la metric spaces). Instead it is a generalisation of an inner product. As we saw for surfaces, an inner product allows us to define a notion of length, so there is a close relation between distance functions and inner products on manifolds. But a new student to the field must get used to the change in terminology.

Definition 3.1. *A* Riemannian metric *g on a manifold M is a choice of inner product for every tangent space T_pM. If U is a chart of M, then we can express g in charts using the coordinate basis vectors:*

$$g_{ij}(p) = g\left(\left.\frac{\partial}{\partial x^i}\right|_p, \left.\frac{\partial}{\partial x^j}\right|_p\right).$$

A Riemannian metric should be smooth in the sense that the functions g_{ij} are smooth in any chart. A manifold with a Riemannian metric is called a Riemannian manifold. *Length of and angle between vectors $X, Y \in T_pM$ is defined in the usual way*

$$\|X\|_g := \sqrt{g(X,X)}, \qquad \cos\theta = \frac{g(X,Y)}{\|X\|\,\|Y\|}$$

The functions g_{ij} are sufficient to determine the inner product of any two vectors by bilinearity:

$$g\left(X^i\frac{\partial}{\partial x^i}, Y^j\frac{\partial}{\partial x^j}\right) = X^iY^j g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = X^iY^j g_{ij}.$$

The symmetry and positive definiteness of g imply that the matrix (g_{ij}) is symmetric and positive definite.

Example 3.2 (Euclidean Space). We have seen in Example 2.3 that any open subset of euclidean space is a manifold with one chart. It is also a Riemannian manifold with the usual dot product

$$g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \frac{\partial}{\partial x^i} \cdot \frac{\partial}{\partial x^j} = \delta_{ij}.$$

Notice that the matrix of the metric in charts is symmetric and positive definite. This is also called the standard metric on $\mathbb{R}^n$.

Example 3.3 (Helicoid)**.** In fact we have seen Riemannian metrics already, namely the first fundamental form of a surface. For the helicoid, in Example 1.21, in the chart $U = \mathbb{R}^2$ we had coordinates $x^1 = u, x^2 = v$ and

$$g_{1\,1}(u,v) = 1, \qquad g_{1\,2}(u,v) = g_{2\,1}(u,v) = 0, \qquad g_{2\,2}(u,v) = u^2 + b^2.$$

For this example we see that $g_{i\,j}$ are non-constant functions (at least, $g_{2\,2}$ is non-constant). We understand that the length coordinate basis vector

$$\left\| \frac{\partial}{\partial v} \bigg|_{(u,v)} \right\| = \sqrt{u^2 + b^2}$$

is different at different points of the helicoid.

We can ask how the functions $g_{i\,j}$ in a chart U are related to those $\tilde{g}_{i\,j}$ in an overlapping chart $\tilde{U}$. We know that the inner product should be independent of basis, so we compute it in two ways:

$$\tilde{g}_{i\,j} = g\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j} \right) = g\left(\frac{\partial x^k}{\partial y^i} \frac{\partial}{\partial x^k}, \frac{\partial x^l}{\partial y^j} \frac{\partial}{\partial x^l} \right) = \frac{\partial x^k}{\partial y^i} \frac{\partial x^l}{\partial y^j} g\left(\frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^l} \right) = \frac{\partial x^k}{\partial y^i} \frac{\partial x^l}{\partial y^j} g_{k\,l}.$$

Notice the subtle contrast to the equivalence relation for vectors:

$$v^i \frac{\partial}{\partial x^i} = v^i \frac{\partial y^j}{\partial x^i} \frac{\partial}{\partial y^j} = \tilde{v}^j \frac{\partial}{\partial y^j} \qquad \Rightarrow \qquad \tilde{v}^j = v^i \frac{\partial y^j}{\partial x^i}.$$

The term for objects that transform with $\frac{\partial x^k}{\partial y^i}$, like $g_{i\,j}$, is *covariant*, whereas those that transform with $\frac{\partial y^j}{\partial x^i}$, like the coefficients of vectors, are called *contravariant*. The convention is to use lower indices for covariant things, and upper indices for contravariant things. Historically this convention came before the summation convention. Because $\frac{\partial x^i}{\partial y^j} \frac{\partial y^j}{\partial x^i} = 1$ by the chain rule, when covariant and contravariant objects are 'multiplied', as in the above formula for g, then the result is independent of charts. This explains why there are so many sums of upper index with lower index, and was the motivation of the summation convention.

Clearly one can endow a manifold with functions $g_{i\,j}$ that satisfy the necessary properties and thereby make it a Riemannian manifold. But this is not usually how we construct Riemannian manifolds. It is far more common to 'inherit' a metric from a bigger Riemannian manifold. This is how we got a metric on the helicoid. In general, we use the tangent map to move vectors on one manifold into the tangent space of another.

Definition 3.4. *Let M be a manifold, N a Riemannian manifold with metric g. Let $f : M \to N$ be an* immersion*. That means that $T_p f$ is injective at every point. Then we define a metric $f^* g$ on M, called the* pullback metric *or the* induced metric*, by*

$$f^* g(v, w) := g\left(T_p f(v), T_p f(w) \right)$$

for any $v, w \in T_p M$.

Exercise 3.5. The formula for f^*g is well-defined for all smooth functions $f : M \to N$, so why is it necessary that f is an immersion?

Let's go through how the definitions of Section 1.4 fit with the definitions in this section. First we have the definition of a regular parameterised surface $\Phi : U \to \mathbb{R}^3$, Definition 1.19. Φ is a function between euclidean spaces, so the tangent map is just the Jacobian $T_p\Phi = J_p\Phi$. The condition that the Jacobian is rank two is equivalent to it being injective by the rank-nullity theorem of linear algebra. Therefore regular and immersed are equivalent.

The first fundamental form is exactly the standard metric on $\mathbb{R}^3$ pullbacked by Φ. In the coordinate basis vectors, we have

$$g_{ij} = \Phi^* g^{\mathbb{R}^3} \left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right) = g^{\mathbb{R}^3} \left((J_p\Phi)_i^k \frac{\partial}{\partial x^k}, (J_p\Phi)_j^l \frac{\partial}{\partial x^l} \right) = (J_p\Phi)_i^k (J_p\Phi)_j^l g^{\mathbb{R}^3} \left(\frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^l} \right)$$

$$= (J_p\Phi)_i^k (J_p\Phi)_j^l \delta_{kl} = \frac{\partial \Phi^k}{\partial x^i} \frac{\partial \Phi^l}{\partial x^j} \delta_{kl} = \frac{\partial \Phi}{\partial x^i} \cdot \frac{\partial \Phi}{\partial x^j},$$

which is the definition of the first fundamental form.

Example 3.6 (Stereographic Projection). What does the induced metric from $\mathbb{R}^2$ look like in stereographic coordinates on $\mathbb{S}^1$? Well, we need to compute the pushforward of the coordinates vector fields and take the dot product. The pushforward was already computed for the U_N chart in Example 2.22:

$$(J_x \phi_N^{-1})(\frac{\partial}{\partial x}) = \frac{2}{(x^2 + 1)^2} \begin{pmatrix} -x^2 + 1 \\ 2x \end{pmatrix} (1) = \frac{2}{(x^2 + 1)^2} \begin{pmatrix} -x^2 + 1 \\ 2x \end{pmatrix}.$$

Therefore

$$g_{11} = \frac{4}{(x^2 + 1)^4} \left[(-x^2 + 1)^2 + (2x)^2 \right] = \frac{4}{(x^2 + 1)^4} \left[x^4 - 2x^2 + 1 + 4x^2 \right]$$

$$= \frac{4}{(x^2 + 1)^4} \left[x^2 + 1 \right]^2 = \frac{4}{(x^2 + 1)^2}.$$

The matrix of the metric has only one entry because the dimension of the manifold is one.

Using this we can calculate the lengths of vectors. For example $\partial_x|_0$ has length

$$\|\partial_x|_0\|^2 = (1)^T (g_{11}(0)) (1) = 4.$$

This is because we saw in Example 2.22 that it pushes forward to $(2, 0)$.

On the other hand $\partial_x|_1$ has length

$$\|\partial_x|_1\|^2 = (1)^T (g_{11}(1)) (1) = 1.$$

So although the vector field ∂_x appears to be constant in the U_N chart, its length is in fact changing.

Exercise 3.7 (Stereographic Projection). Compute $\tilde{g}_{ij}$ in the chart $U_S \subset \mathbb{S}^1$ and verify the change of chart formula for the metric.

Exercise 3.8 (Stereographic Projection)**.** For the U_N chart of $\mathbb{S}^2$ verify

$$g_{1\,1} = g_{2\,2} = \frac{4}{(\|x\|^2 + 1)^2}, \qquad\qquad g_{1\,2} = g_{2\,1} = 0.$$

Finally, consider the notion of isometry in Definition 1.38. It says that two parameterised surfaces are isometric if their parametrisations induce equal metrics. We give the following more general definition.

Definition 3.9. *Let M, N be Riemannian manifolds and let $f : M \to N$ be an immersion. We call f an* Riemannian immersion *if $g^M = f^*g^N$. In words, if the metric on M induced by the immersion is equal to the existing metric on M. If additionally f is a diffeomorphism (bijective, smooth, smooth inverse) then we call f an* isometry*. Two Riemannian manifolds are* isometric *if there is a isometry between them.*

As above, if M is just a manifold and we have an immersion $f : M \to N$ to a Riemannian manifold, then we can endow M with the pullback metric. Then f becomes a Riemannian immersion by definition.

Example 3.10. Suppose that we have an Riemannian immersion $f : M \to \mathbb{R}^3$ and let $R : \mathbb{R}^3 \to \mathbb{R}^3$ be a rotation. Define $R \circ f : M \to \mathbb{R}^3$; this is also a Riemannian immersion, as we will now prove. The essential step of the calculation is to notice that $TR = R$ because R is a linear transformation, and that R a rotation doesn't change the inner product $g^{\mathbb{R}^3}$. Therefore

$$\begin{aligned}
(R \circ f)^*g^{\mathbb{R}^3}(v, w) &= g^{\mathbb{R}^3}(T(R \circ f)v, T(R \circ f)w) = g^{\mathbb{R}^3}((TR \circ Tf)v, (TR \circ Tf)w) \\
&= g^{\mathbb{R}^3}(R(Tf(v)), R(Tf(w))) = g^{\mathbb{R}^3}(Tf(v), Tf(w)) \\
&= f^*g^{\mathbb{R}^3}(v, w) = g^M(v, w).
\end{aligned}$$

In the last line we used that f is a Riemannian immersion.

Exercise 3.11. Generalise the above example to prove: the composition of two Riemannian immersions is a Riemannian immersion.

A weaker condition to isometry is that of a conformal map.

Definition 3.12. *Let M, N be Riemannian manifolds and let $f : M \to N$ be an immersion. We say that f is* conformal *if there exists a smooth function $\lambda : M \to \mathbb{R}$ such that $g^M = \lambda f^*g^N$.*

A conformal map does not preserve lengths or distances, but it does preserve angles since

$$g^N(Tf(X), Tf(Y)) = f^*g^N(X, Y) = \lambda g^M(X, Y)$$

implies

$$\frac{g^N(Tf(X), Tf(Y))}{\|Tf(X)\|_{g^N} \|Tf(Y)\|_{g^N}} = \frac{\lambda g^M(X, Y)}{\sqrt{\lambda}\|X\|_{g^M} \sqrt{\lambda}\|Y\|_{g^M}} = \frac{g^M(X, Y)}{\|X\|_{g^M} \|Y\|_{g^M}}.$$

Example 3.13 (Stereographic Projection)**.** Consider inverse stereographic projection $\Phi = \phi_N^{-1}$ as a function between $U_N = \mathbb{R}^n$ with the standard metric and the sphere $\mathbb{S}^n$ with the induced metric of $\mathbb{R}^{n+1}$.

For $n = 1$, and indeed on any one-dimensional manifold, all metrics are conformally equivalent because there is only one metric coefficient g_{11}.

For $n = 2$, Exercise 3.8 shows us that Φ is not a Riemannian immersion, because the pullback metric $\Phi^* g^{\mathbb{S}^2}$ is not equal to the standard metric δ_{ij}. However, Φ is conformal because

$$\Phi^* g^{\mathbb{S}^2} = \frac{4}{(\|x\|^2 + 1)^2} \delta_{ij}.$$

Notice for example, that in stereographic coordinates the lines through the origin are lines of longitude and circles centered at the origin are lines of latitude, and these are always perpendicular to one another.

A calculation similar to the $n = 2$ case shows that sterographic projection is conformal for all n. Therefore stereographic charts have the advantage that the angle between vectors as naively calculated in the chart is the same as in $\mathbb{R}^{n+1}$.

Example 3.14 (Helicoid)**.** We have seen the pullback metric of the helicoid in Example 3.3. It is a metric on $U = \mathbb{R}^2$. On the other hand we could give the plane the standard metric δ_{ij}. With these metrics, the immersion Φ is not conformal.

We could use a different parameterisation of the helicoid $\tilde{\Phi} : \mathbb{R}^2 \to \mathbb{R}^3$

$$\tilde{\Phi}(u, v) = (\sinh u \cos v, \sinh u \sin v, v).$$

The pushforwards of the coordinate vectors are

$$\frac{\partial \tilde{\Phi}}{\partial u} = (\cosh u \cos v, \cosh u \sin v, 0)$$

$$\frac{\partial \tilde{\Phi}}{\partial v} = (-\sinh u \sin v, \sinh u \cos v, 1).$$

The pullback of the standard metric on $\mathbb{R}^3$ by this map is

$$(\tilde{\Phi}^* g^{\mathbb{R}^3})_{11} = g^{\mathbb{R}^3}\left(\frac{\partial \tilde{\Phi}}{\partial u}, \frac{\partial \tilde{\Phi}}{\partial u}\right) = \cosh^2 u,$$

$$(\tilde{\Phi}^* g^{\mathbb{R}^3})_{12} = (\tilde{\Phi}^* g^{\mathbb{R}^3})_{21} = g^{\mathbb{R}^3}\left(\frac{\partial \tilde{\Phi}}{\partial u}, \frac{\partial \tilde{\Phi}}{\partial v}\right) = 0,$$

$$(\tilde{\Phi}^* g^{\mathbb{R}^3})_{22} = \sinh^2 u + 1 = \cosh^2 u.$$

That is to say

$$(\tilde{\Phi}^* g^{\mathbb{R}^3})_{ij} = \cosh^2 u \, \delta_{ij} = \cosh^2 u \, g_{ij}^{\mathbb{R}^2}$$

Therefore $\tilde{\Phi}$ is a conformal map between $\mathbb{R}^2$ and $\mathbb{R}^3$ with the standard metrics.

3.2   Quaternions and $\mathbb{S}^3$

In this section we introduce the quaternions as a means to understand the rotations of $\mathbb{S}^3$. The 3-sphere is a beautiful manifold because it is also a group. A manifold that is also a group is called a *Lie group*. We will not go into the general theory of Lie groups, but they come with a natural way to move vectors around, something we are trying to achieve in this chapter. The example of Lie groups is therefore very instructive for us.

The quaternions are a four dimensional real vector space $\{a_0 + a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}\}$. A quaternion has a real part $\operatorname{Re} a = a_0$ and an imaginary part $\operatorname{Im} a = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$. Unlike for complex numbers, the imaginary part of a quaternion is not real. The quaternionic conjugate is $\bar{a} = \operatorname{Re} a - \operatorname{Im} a$. Clearly $\operatorname{Re} \bar{a} = \operatorname{Re} a$ and $\operatorname{Im} \bar{a} = -\operatorname{Im} a$. Elements of the subspace $\{a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}\}$ are called imaginary.

Famously the quaternions have an associative but non-commutative multiplication, defined by $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$ and 1 is the identity. We also use the notation $\mathbf{e} = 1$ to aid clarity. For example $\mathbf{ij} = \mathbf{k}$ because we multiply $\mathbf{ijk} = -1$ on the right by $\mathbf{k}$ to get $\mathbf{ijk}^2 = -\mathbf{k}$ and use $\mathbf{k}^2 = -1$. On the other hand $\mathbf{ji} = -\mathbf{k}$: from $\mathbf{ijk} = -1$ we get $1 = \mathbf{kji}$ and now multiply on the left by $\mathbf{k}$. This doesn't mean that every multiplication of quaternions is anti-commuting:

$$(1 + \mathbf{i})(1 + \mathbf{j}) = 1 + 1\mathbf{j} + \mathbf{i}1 + \mathbf{ij} = 1 + \mathbf{i} + \mathbf{j} + \mathbf{k},$$
$$(1 + \mathbf{j})(1 + \mathbf{i}) = 1 + 1\mathbf{i} + \mathbf{j}1 + \mathbf{ji} = 1 + \mathbf{i} + \mathbf{j} - \mathbf{k}.$$

According to legend on Monday 16 October 1843, as Hamilton was walking to the Royal Irish Academy, he had the idea that to define a multiplication on $\mathbb{R}^4$ it must be non-commutative, whereupon he carved the above equations into the side of Brougham Bridge. I have been to the bridge but was unable to find the carving, so instead I offer the following simple trick to remember the multiplication rule. Draw $\mathbf{i}, \mathbf{j}, \mathbf{k}$ on a directed circle. Multiplication of two elements gives the third, with a plus sign if they are in the correct direction and a minus sign if they are in the reverse direction. This is of course the same rule as for the cross product in $\mathbb{R}^3$.

A direct computation shows that $a\bar{a} = \bar{a}a = a_0^2 + a_1^2 + a_2^2 + a_3^2$ is always real and non-negative. Thus we can define the norm $|a| = \sqrt{a\bar{a}}$. The norm shows that every non-zero quaternion has a two-sided inverse, namely $a^{-1} = |a|^{-2}\bar{a}$. Therefore the quaternions are a non-commutative field.

Exercise 3.15. Prove the following:

    a. The dot product can be calculated as $a \cdot b = \operatorname{Re}(\bar{a}b)$,

    b. $a\bar{a} = \bar{a}a = a_0^2 + a_1^2 + a_2^2 + a_3^2$,

    c. Conjugation is order reversing $\overline{ab} = \bar{b}\,\bar{a}$,

    d. The norm is multiplicative $|ab| = |a|\,|b|$.

This norm is plainly the same as the usual norm on $\mathbb{R}^4$. The unit quaternions (those with norm 1) are as a set $\mathbb{S}^3 \subset \mathbb{R}^4$. Therefore the 3-sphere is a Lie group, because we can multiply two elements of it together in a way that can be undone. This is rather special, the only spheres that are Lie groups are $\mathbb{S}^0$ ($\mathbb{S}^0 = \{\pm 1\}$ in $\mathbb{R}^1$), $\mathbb{S}^1$ (add the angles), and $\mathbb{S}^3$.

If we choose $a \in \mathbb{S}^3$ we can look at the function $L_a : \mathbb{S}^3 \to \mathbb{S}^3$ defined by $L_a(q) = aq$ This is a bijective function, because the inverse is $L_{a^{-1}}$. And $L_a(\mathbf{e}) = a\mathbf{e} = a$. Therefore the tangent map of L_a takes $T_\mathbf{e}\mathbb{S}^3$ to $T_a\mathbb{S}^3$. Moreover, the tangent map is also bijective: from the chain rule

$$\mathrm{id}_{T_\mathbf{e}\mathbb{S}^3} = T(L_a \circ L_{a^{-1}}) = TL_a \circ TL_{a^{-1}}$$

Indeed, this inverse has the property that it takes a to the identity $L_{a^{-1}}(a) = a^{-1}a = \mathbf{e}$. This gives us a way to move any tangent vector of $\mathbb{S}^3$ to $T_\mathbf{e}\mathbb{S}^3$. Just as in Example 2.28, this shows us that $T\mathbb{S}^3$ is trivial. The function $T_a L_{a^{-1}} : T_a\mathbb{S}^3 \to T_\mathbf{e}\mathbb{S}^3$ is called the *left trivialisation*. Likewise we can define $R_a(q) = qa$ and we have the *right trivialisation* $T_a R_{a^{-1}} : T_a\mathbb{S}^3 \to T_\mathbf{e}\mathbb{S}^3$

Example 3.16 (3-Sphere)**.** Let us compute the trivialisations for the point $a = \mathbf{i} = (0, 1, 0, 0) \in \mathbb{S}^3$. The inverse of a is $a^{-1} = -\mathbf{i}$, since $\mathbf{i}(-\mathbf{i}) = 1$. If we have any point $q = q^0 + q^1\mathbf{i} + q^2\mathbf{j} + q^3\mathbf{k}$ then

$$L_{a^{-1}}(q) = (-\mathbf{i})(q^0 + q^1\mathbf{i} + q^2\mathbf{j} + q^3\mathbf{k}) = q^1 - q^0\mathbf{i} + q^3\mathbf{j} - q^2\mathbf{k}.$$

This does indeed have the property that $L_{a^{-1}}(a) = 1 - 0 + 0 - 0 = 1 = \mathbf{e}$. Next we use some geometry to avoid using charts. We know that the tangent vectors in $T_a\mathbb{S}^3$ are perpendicular to a, because this is a sphere. We write

$$T_a\mathbb{S}^3 = \{v^1\mathbf{e} + v^2\mathbf{j} + v^3\mathbf{k} \mid v^1, v^2, v^3 \in \mathbb{R}\}.$$

Because $L_{a^{-1}}(q)$ is linear in q, we know

$$T_a L_{a^{-1}}(v^1\mathbf{e} + v^2\mathbf{j} + v^3\mathbf{k}) = -v^1\mathbf{i} + v^3\mathbf{j} - v^2\mathbf{k}.$$

For the right trivialisation

$$R_{a^{-1}}(q) = (q^0 + q^1\mathbf{i} + q^2\mathbf{j} + q^3\mathbf{k})(-\mathbf{i}) = q^1 - q^0\mathbf{i} - q^3\mathbf{j} + q^2\mathbf{k},$$
$$T_a R_{a^{-1}}(v^1\mathbf{e} + v^2\mathbf{j} + v^3\mathbf{k}) = -v^1\mathbf{i} - v^3\mathbf{j} + v^2\mathbf{k}.$$

So these two trivialisations on $\mathbb{S}^3$ are different from one another.

Example 3.17. We can generalise the previous example to work for any point $a \in \mathbb{S}^3$. Just like i is a right-angle rotation of the complex plane, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are all right-angle rotations of the quaternions. Therefore $a\mathbf{i}, a\mathbf{j}, a\mathbf{k}$ is an orthonormal basis of $T_a\mathbb{S}^3$. Alternatively, since

$$L_a(q) = a(q^0 + q^1\mathbf{i} + q^2\mathbf{j} + q^3\mathbf{k}) = q^0 a + q^1 a\mathbf{i} + q^2 a\mathbf{j} + q^3 a\mathbf{k}$$

and $\mathbf{i}, \mathbf{j}, \mathbf{k}$ is a basis for $T_\mathbf{e}\mathbb{S}^3$ we know that

$$T_\mathbf{e}L_a(v^1\mathbf{i} + v^2\mathbf{j} + v^3\mathbf{k}) = v^1 a\mathbf{i} + v^2 a\mathbf{j} + v^3 a\mathbf{k}$$

is all of $T_a\mathbb{S}^3$. This shows us that identifying $T_a\mathbb{S}^3$ with $T_\mathbf{e}\mathbb{S}^3$ is the same as writing it with respect to the pushforward of a basis. If $v \in T_a\mathbb{S}^3$ then we get

$$T_a L_{a^{-1}} v = a^{-1} v.$$

We call the vector field on $\mathbb{S}^3$ a *left-invariant* field when it has the form

$$X|_a = av$$

for $v \in T_{\mathbf{e}} \mathbb{S}^3$, because every vector $X|_a$ corresponds to v using the left trivialisation. Ditto we have the *right-invariant* vector fields

$$Y|_a = va.$$

Example 3.18 (3-Sphere)**.**

$$X|_a = a\mathbf{i} = -a_1\mathbf{e} + a_0\mathbf{i} + a_3\mathbf{j} - a_2\mathbf{k}$$

is a left-invariant vector field on $\mathbb{S}^3$. We recognise $X|_{\mathbf{i}} = -\mathbf{e}$ as a vector in $T_{\mathbf{i}}\mathbb{S}^3$.

3.3   Covariant Derivatives

We have seen numerous examples thus far of how we cannot simply move vectors around in a chart like we can in euclidean space. If you take a tangent vector at one point of the sphere and translate it in $\mathbb{R}^3$ to another point of the sphere, it may not be tangent anymore. As we observed below Example 2.19, a vector might have the same coordinates at different points in one chart, but not in another. And in Example 3.3 we saw that one coordinate basis vector changed its length as you moved around, while the other stayed the same length.

There is also a common thought experiment. Suppose that you are standing on the equator facing east. You walk forward without turning, until you have walked half way around the Earth. Then, still without turning, you begin to sidestep to the north. You sidestep all the way to the north pole, but keep going until you have returned to your original position. The remarkable fact is, even though at no stage did you turn, you are now facing west.

Exercise 3.19. Can you modify the journey so that you end up facing other directions? What is the connection between the area your journey encompasses and the final rotation angle?

However, the naive definition of the derivative of a vector field

$$\lim_{h \to 0} \frac{1}{h}(X|_{p+h} - X|_p)$$

asks us to subtract two vectors at different points. Indeed, any non-trivial definition of a derivative of a vector field is going to require us to compare vectors at different points. Geometrically, thinking about a surface, what we want to do is to 'roll' the tangent plane along the surface to another point. This idea is called development and the relation between two tangent planes was called an *affine connection*, because it was an affine transformation of one plane to another. In modern terminology it is more common to call this a *parallel transport operator*, for reasons that will be explained in Section 3.4. Already from the above thought experiment we see that a parallel transport operator will depend not just on the two start and end points, but on the path between those points.

The modern approach, which we will ultimately take, uses a different point of view. It asks: how much are vector fields are changing? Once we have a basis of vector fields and we know their changes, then we can measure all other vector fields against them. This leads to the definition of a *covariant derivative*, a type of differential operator on vector fields. It is extremely common to call this an *connection*, but we will refrain from doing so, at least until we have made clear the relationship with the parallel transport operator. Though the two approaches are equivalent, the modern approach is the much easier place to begin. On the other hand, some of the definitions and motivations for the modern approach only really make sense from the point of view of the traditional approach.

Definition 3.20. *A* covariant derivative *on a manifold M is a function ∇ that acts on two vector fields to produce a third. We write it as $\nabla_X Y$, with X being the 'direction'. It has the following properties for all smooth functions $f : M \to \mathbb{R}$ and vector fields $X, \tilde{X}, Y, \tilde{Y}$:*

    a. It is C^∞-linear in the direction:

$$\nabla_{fX+\tilde{X}} Y = f\nabla_X Y + \nabla_{\tilde{X}} Y.$$

b. *It is additive in the derivative:*

$$\nabla_X(Y + \tilde{Y}) = \nabla_X Y + \nabla_X \tilde{Y}.$$

c. *It obeys the product (Leibniz) rule:*

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

Example 3.21 (Euclidean Space). Consider euclidean space $\mathbb{R}^n$ and let $X = X^i\partial_i, Y = Y^i\partial_i$ be vector fields in the chart x^i. Then

$$\nabla_X^{\text{euc}} Y := X^i \frac{\partial Y^j}{\partial x^i}\partial_j$$

is a covariant derivative.

You might be confused, because in Example 2.33 we said this formula didn't work. Indeed, this formula is not chart independent. This definition is saying explicitly "use this particular coordinates to do the derivative and not others". If you write this covariant derivative in polar coordinates, then the formula for this covariant derivative will look different. But this is why we say that it is a covariant derivative, we are not claiming uniqueness.

Exercise 3.22. Check the above example has the three properties that are required of a covariant derivative.

The above example suggests that there are many covariant derivatives on a manifold. At least for a manifold that can be covered by a single chart, every set of coordinates gives a covariant derivative. In the following theorem we characterise the set of covariant derivatives.

Theorem 3.23 (Tensorial). *Let ∇^0, ∇^1 be two covariant derivatives derivatives. Define their difference $A(X,Y) := \nabla_X^0 Y - \nabla_X^1 Y$. Then A is C^∞-linear in both X and Y.*

Proof. C^∞-linear in X is immediate from Property a of covariant derivatives. C^∞-linear in Y is not too much harder to show, we use Properties b and c:

$$
\begin{aligned}
A(X, fY + \tilde{Y}) &= \nabla_X^0(fY) + \nabla_X^0 \tilde{Y} - \nabla_X^1(fY) - \nabla_X^1 \tilde{Y} \\
&= X(f)Y + f\nabla_X^0 Y - X(f)Y - \nabla_X^1 Y + A(X, \tilde{Y}) \\
&= fA(X,Y) + A(X, \tilde{Y}).
\end{aligned}
$$

$\square$

Exercise 3.24. Prove the converse of Theorem 3.23: Let ∇ is a covariant derivative on M. For all vector fields X, Y let $A(X,Y)$ be a smooth vector field. Suppose that this function A is C^∞-linear in both X, Y. Then $\tilde{\nabla} := \nabla + A$ is also a covariant derivative.

Corollary 3.25 (Affineness). *The space of covariant derivatives on M is affine in the following sense: if $t \in \mathbb{R}$ is a constant and ∇^0, ∇^1 are two covariant derivatives, so is $\nabla^t := (1-t)\nabla^0 + t\nabla^1$.*

Proof. Observe that $\nabla^t = \nabla^0 + t(\nabla^1 - \nabla^0)$. The corollary now follows from Theorem 3.23 and its converse Exercise 3.24. $\square$

The above theorems give us a way to construct new covariant derivatives from existing ones (and in fact construct every covariant derivative). But we need one to start with. One can prove[1] that every manifold has a covariant derivative, but the proof is technical and not practically useful. We have seen in Example 3.21 that if one chart covers the whole space, then we can declare it is special and use the directional derivative. For manifolds that are a submanifold of a bigger space, the following example is typical.

Example 3.26 (Stereographic Projection,Tangent Connection)**.** Consider the sphere $\mathbb{S}^1$ inside $\mathbb{R}^2$. We can understand any vector field Y on $\mathbb{S}^1$ as a function $\tilde{Y} : \mathbb{S}^1 \to \mathbb{R}^2$ using the pushforward. Therefore we can differentiate $\tilde{Y}$ as an $\mathbb{R}^2$ valued function in the usual way.

For the sake of a numerical example, let us take both X and Y to be the vector field from Example 2.19. The pushforward of the vector field is

$$X = \begin{cases} \frac{-2x^2+2}{(x^2+1)^2} \frac{\partial}{\partial p^1} + \frac{4x}{(x^2+1)^2} \frac{\partial}{\partial p^2} & \text{for } x \in U_N \\ 0 & \text{for } p = N \end{cases}$$

and interpreting this a function to $\mathbb{R}^2$ we have

$$Y = \begin{cases} \left(\frac{-2x^2+2}{(x^2+1)^2}, \frac{4x}{(x^2+1)^2} \right) & \text{for } x \in U_N \\ 0 & \text{for } p = N \end{cases} = \begin{cases} \frac{2}{x^2+1}(-p^2, p^1) & \text{for } p \neq N \\ 0 & \text{for } p = N. \end{cases}$$

If we differentiate $\tilde{Y}$ along X, then using the product rule to avoid some nasty but unimportant terms we get

$$X(\tilde{Y}) = X\left(\frac{2}{x^2+1} \right)(-p^2, p^1) + \frac{2}{x^2+1}\left(-\frac{4x}{(x^2+1)^2}, \frac{-2x^2+2}{(x^2+1)^2} \right)$$

$$= X\left(\frac{2}{x^2+1} \right)(-p^2, p^1) - \frac{4}{(x^2+1)^2}\left(p^1, p^2 \right)$$

The first term is tangent to the circle, but the second is not. So we see the trouble is that the directional derivative $X(\tilde{Y}) = X^i \frac{\partial Y^j}{\partial p^i} \frac{\partial}{\partial p^j}$ is no longer be tangent to $\mathbb{S}^1$. Therefore this does not meet the definition of a covariant derivative on $\mathbb{S}^1$.

What we can do however is to project this directional derivative onto the tangent space. We define the *tangent covariant derivative* as

$$\nabla_X^\top Y = \operatorname*{proj}_{T_p \mathbb{S}^1} X^i \frac{\partial Y^j}{\partial p^i} \frac{\partial}{\partial p^j}.$$

Let's check the three required properties. The two linearity properties just follow from the

---

[1]Lee Proposition 4.5

linearity of the projection

$$\nabla^{\top}_{fX+\tilde{X}}Y = \operatorname*{proj}_{T_p\mathbb{S}^1}\left(fX^i\frac{\partial Y^j}{\partial p^i}\frac{\partial}{\partial p^j} + \tilde{X}^i\frac{\partial Y^j}{\partial p^i}\frac{\partial}{\partial p^j}\right)$$

$$= f\operatorname*{proj}_{T_p\mathbb{S}^1}X^i\frac{\partial Y^j}{\partial p^i}\frac{\partial}{\partial p^j} + \operatorname*{proj}_{T_p\mathbb{S}^1}\tilde{X}^i\frac{\partial Y^j}{\partial p^i}\frac{\partial}{\partial p^j} = f\nabla^{\top}_X Y + \nabla^{\top}_{\tilde{X}}Y,$$

$$\nabla^{\top}_X(Y+\tilde{Y}) = \operatorname*{proj}_{T_p\mathbb{S}^1}\left(X^i\frac{\partial Y^j}{\partial p^i}\frac{\partial}{\partial p^j} + X^i\frac{\partial \tilde{Y}^j}{\partial p^i}\frac{\partial}{\partial p^j}\right) = \nabla^{\top}_X Y + \nabla^{\top}_X \tilde{Y}.$$

For the third property, we need to recognise that $X(f)Y$ is already tangent to $\mathbb{S}^1$, so the projection leaves it unaltered:

$$\nabla^{\top}_X(fY) = \operatorname*{proj}_{T_p\mathbb{S}^1}X^i\frac{\partial(fY^j)}{\partial p^i}\frac{\partial}{\partial p^j} = \operatorname*{proj}_{T_p\mathbb{S}^1}\left(X(f)Y + X^i f\frac{\partial Y^j}{\partial p^i}\frac{\partial}{\partial p^j}\right)$$

$$= X(f)Y + f\nabla^{\top}_X Y.$$

Nothing in the calculation depended on $\mathbb{S}^1$ specifically, so this is a general construction for immersed submanifolds.

Next we examine what type of derivative a covariant derivative is. We will show that it is a directional derivative, in a sense that will be developed. To this end, the first property to notice is that although the direction and the derived vector fields have dramatically different behaviour under scaling by a smooth function, they are both $\mathbb{R}$-linear. If a is a constant then

$$\nabla_{aX}Y = a\nabla_X Y, \qquad \nabla_X(aY) = X(a)Y + a\nabla_X Y = a\nabla_X Y.$$

Consequently, if either field is zero, then so is the covariant derivative. Moreover, using cut-off functions, the covariant derivative only depends on local information.[2] In fact something stronger is true of X:

Lemma 3.27 (Directional Derivative). *The value of $\nabla_X Y$ at $p \in M$ only depends on $X|_p$ and not other values of X.*

Proof. By linearity, it suffices to prove that $X|_p = 0$ implies $(\nabla_X Y)|_p = 0$. Writing X in a chart we have $X = X^i\partial_i$ and $X^i(p) = 0$ for all the coefficients. Then

$$\left(\nabla_{X^i\partial_i}Y\right)|_p = \left(X^i\nabla_{\partial_i}Y\right)\Big|_p = X^i(p)\left(\nabla_{\partial_i}Y\right)\Big|_p = 0. \qquad \square$$

For this reason we sometimes speak of the covariant derivative $\nabla_v Y$ in a direction $v \in T_pM$. The same is not true for Y: the covariant derivative really is a derivative of Y and depends on its values in a neighbourhood of a point. However, to compute $\nabla_v Y$ you don't need to know Y completely on an open neighbourhood of p, it is enough to know Y on a curve whose tangent is v.

Lemma 3.28 (Curve Derivative). *Let $Y, \tilde{Y}$ be two vector fields and let $\alpha : (a,b) \to M$ be a smooth curve with $\alpha(0) = p$ and $\alpha'(0) = v$. Suppose that $Y \circ \alpha = \tilde{Y} \circ \alpha$. Then $\nabla_v Y = \nabla_v \tilde{Y}$.*

---

[2]See Lee Lemma 4.1 for a proof. We prove a stronger statement in Lemma 3.28.

Proof. Let us consider the situation in a chart, writing $v = v^i \partial_i|_p$, $Y = Y^i \partial_i$ and $\tilde{Y} = \tilde{Y}^i \partial_i$. Then by the properties of covariant derivatives,

$$\nabla_v Y = \nabla_{v^i \partial_i|_p}(Y^j \partial_j) = v^i \nabla_{\partial_i|_p}(Y^j \partial_j) = v^i \left.\frac{\partial Y^j}{\partial x^i}\right|_p \partial_j + v^i Y^j(p) \nabla_{\partial_i|_p} \partial_j,$$

and likewise for $\tilde{Y}$. Now, Y and $\tilde{Y}$ agree on α, so $Y(p) = \tilde{Y}(p)$. Moreover, by the chain rule

$$v^i \left.\frac{\partial Y^j}{\partial x^i}\right|_p = \left.\frac{d}{dt}(Y^j \circ \alpha)\right|_p = \left.\frac{d}{dt}(\tilde{Y}^j \circ \alpha)\right|_p = v^i \left.\frac{\partial \tilde{Y}^j}{\partial x^i}\right|_p.$$

Hence

$$\nabla_v Y = v^i \left.\frac{\partial Y^j}{\partial x^i}\right|_p \partial_j + v^i Y^j(p) \nabla_{\partial_i|_p} \partial_j = v^i \left.\frac{\partial \tilde{Y}^j}{\partial x^i}\right|_p \partial_j + v^i \tilde{Y}^j(p) \nabla_{\partial_i|_p} \partial_j = \nabla_v \tilde{Y} \qquad \square$$

This lemma tells us that we can really view the covariant derivative as a generalisation of a directional derivative. This is in contrast to other derivatives of vector fields. Recall Example 2.37. Now consider the vector fields from that example along the curve $\alpha(t) = (t, 0)$, the x-axis. We have $X \circ \alpha = \partial_1$, $Y \circ \alpha = \partial_2$, and $V \circ \alpha = \partial_1$. But $[X, Y] = 0$ while $[V, Y] = -\partial_1$. This shows that the Lie bracket is not a covariant derivative.

To break up all this theory, let's do another example.

Example 3.29 (3-Sphere). We define a covariant derivative ∇^L on $\mathbb{S}^3$ in the following way. Given any vector field Y on $\mathbb{S}^3$, use left trivialisation to write it as a function $\tilde{Y} : \mathbb{S}^3 \to T_{\mathbf{e}}\mathbb{S}^3$. From Example 3.17 we know this has the formula $p \mapsto p^{-1}Y|_p$ using quaternions. Now that we have a function to the same vector space, there is no problem differentiating. This gives us a function $X(\tilde{Y}) : M \to T_{\mathbf{e}}\mathbb{S}^3$. Use the left trivialisation again to move the result back to $T_p\mathbb{S}^3$.

Putting this all in one formula gives

$$(\nabla^L_X Y)|_p := (T_{\mathbf{e}}L_p \circ X \circ T_p L_{p^{-1}})Y.$$

This covariant derivative has the property that the derivative of a left-invariant vector field is always zero. This is because, by definition, after you bring its vectors to $\mathbf{e}$ they are all the same. In other words $\tilde{Y}$ is constant and thus has zero derivative.

So to see an interesting example, we need to use a non-left-invariant vector field. Consider $Y|_p = \mathbf{i}p$. We know that $\tilde{Y}(p) = p^{-1}\mathbf{i}p$. To proceed we need to choose a direction field X. We know that the value of the covariant derivative at any point only depends on the value of X at that point. So for simplicity let us calculate for the point $\mathbf{i}$ in the direction $\mathbf{j} = \frac{\partial}{\partial p^2}$:

$$X|_{\mathbf{i}}\tilde{Y} = \left.\frac{\partial}{\partial p^2}p^{-1}\mathbf{i}p\right|_{\mathbf{i}} = -p^{-1}\frac{\partial p}{\partial p^2}p^{-1}\mathbf{i}p + p^{-1}\mathbf{i}\left.\frac{\partial p}{\partial p^2}\right|_{\mathbf{i}} = -p^{-1}\mathbf{j}p^{-1}\mathbf{i}p + p^{-1}\mathbf{i}\mathbf{j}|_{\mathbf{i}}$$

$$= -\mathbf{i}^{-1}\mathbf{j}\mathbf{i}^{-1}\mathbf{i}\mathbf{i} + \mathbf{i}^{-1}\mathbf{i}\mathbf{j} = \mathbf{j} + \mathbf{j} = 2\mathbf{j}.$$

Finally, we move this back to $T_{\mathbf{i}}\mathbb{S}^3$

$$(\nabla_X Y)|_{\mathbf{i}} = T_{\mathbf{e}}L_{\mathbf{i}}(2\mathbf{j}) = \mathbf{i}\,2\mathbf{j} = 2\mathbf{k}.$$

In the same manner, we can define a covariant derivative ∇^R using the right trivialisation.

In the examples above, to define a covariant derivative we really gave a directional derivative. But what is the minimal information required to specify a covariant derivative? Because covariant derivatives are local, we give the answer in a chart. Let ∂_i be the coordinate vector fields. Then for each pair i, j we have a vector field $\nabla_{\partial_i}\partial_j$. This vector field must be able to be written

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

for some coefficients Γ_{ij}^k. These coefficients are called *Christoffel coefficients*, though be aware that some authors reserve this name for a special case. This is sufficient information to determine ∇ because

$$\nabla_X Y = X^i \nabla_{\partial_i}(Y^j \partial_j) = X^i \frac{\partial Y^j}{\partial x^i}\partial_j + X^i Y^j \nabla_{\partial_i}\partial_j = \left(X^i \frac{\partial Y^k}{\partial x^i} + X^i Y^j \Gamma_{ij}^k \right)\partial_k.$$

Example 3.30 (Polar Coordinates)**.** Let us consider $\mathbb{R}^2$ with ∇^{euc}. We see by comparison of its definition in Example 3.21 with the formula above that Γ_{ij}^k is zero for all points and all indices in the standard chart.

But let us compute it with respect to polar coordinates. By the definition of ∇^{euc}, we have to calculate in the x^1, x^2 coordinates. We have

$$\frac{\partial}{\partial r} = \cos\theta\frac{\partial}{\partial x^1} + \sin\theta\frac{\partial}{\partial x^2} = \frac{x^1}{\sqrt{(x^1)^2 + (x^2)^2}}\frac{\partial}{\partial x^1} + \frac{x^2}{\sqrt{(x^1)^2 + (x^2)^2}}\frac{\partial}{\partial x^2}$$

$$\frac{\partial}{\partial\theta} = -r\sin\theta\frac{\partial}{\partial x^1} + r\cos\theta\frac{\partial}{\partial x^2} = -x^2\frac{\partial}{\partial x^1} + x^1\frac{\partial}{\partial x^2}.$$

Hence we can calculate

$$\begin{aligned}
\nabla_{\frac{\partial}{\partial r}}\frac{\partial}{\partial\theta} &= \frac{x^1}{\sqrt{(x^1)^2 + (x^2)^2}}\nabla_{\frac{\partial}{\partial x^1}}\frac{\partial}{\partial\theta} + \frac{x^2}{\sqrt{(x^1)^2 + (x^2)^2}}\nabla_{\frac{\partial}{\partial x^2}}\frac{\partial}{\partial\theta} \\
&= \frac{x^1}{\sqrt{(x^1)^2 + (x^2)^2}}\left(\frac{\partial(-x^2)}{\partial x^1}\frac{\partial}{\partial x^1} + \frac{\partial x^1}{\partial x^1}\frac{\partial}{\partial x^2} \right) \\
&\quad + \frac{x^2}{\sqrt{(x^1)^2 + (x^2)^2}}\left(\frac{\partial(-x^2)}{\partial x^2}\frac{\partial}{\partial x^1} + \frac{\partial x^1}{\partial x^2}\frac{\partial}{\partial x^2} \right) \\
&= \frac{x^1}{\sqrt{(x^1)^2 + (x^2)^2}}\frac{\partial}{\partial x^2} - \frac{x^2}{\sqrt{(x^1)^2 + (x^2)^2}}\frac{\partial}{\partial x^1} \\
&= -\sin\theta\frac{\partial}{\partial x^1} + \cos\theta\frac{\partial}{\partial x^2} = \frac{1}{r}\frac{\partial}{\partial\theta},
\end{aligned}$$

and hence in polar coordinates

$$\Gamma_{r,\theta}^r = 0, \qquad \Gamma_{r,\theta}^\theta = \frac{1}{r}.$$

The other six coefficients are calculated similarly.

Example 3.31 (Tangent Connection)**.** Let's calculate the Christoffel coefficients for a submanifold $f : M \to \mathbb{R}^n$ with the connection $\nabla^\top$ from Example 3.26 in some chart U. Let $\Phi = f \circ \phi^{-1}$ be a parameterisation, a map from a chart U to $\mathbb{R}^n$. Because the definition of $\nabla^\top$ uses the geometry of $\mathbb{R}^n$ we need the pushforwards of the coordinate basis vectors. We use the notation $E_i = T_x(f)\frac{\partial}{\partial x^i} = J_x(\Phi)^j_i \frac{\partial}{\partial p^j}$. From the directional derivative definition of the pushforward map

$$E_i(Y^j) = T\Phi\left(\frac{\partial}{\partial x^i}\right)(Y^j) = \frac{\partial}{\partial x^i}(Y^j \circ \Phi).$$

Therefore the covariant derivative is

$$\nabla^\top_{\partial_i}\partial_j = \operatorname*{proj}_{T_pM} \frac{\partial}{\partial x^i}(E^k_j \circ \Phi)\frac{\partial}{\partial p^k}.$$

Finally to give the Christoffel coefficients, we write this vector in the coordinate basis E_i. This requires solving some linear algebra problem.

Example 3.32 (Stereographic Projection)**.** Let's calculate the Christoffel coefficients for $\mathbb{S}^1$ with the connection $\nabla^\top$ from Example 3.26 in the chart U_N. This is a special case of the previous example. The immersion f is the identity map, so the parameterisation is $\Phi = \phi_N^{-1}$. There is only one coordinate vector field

$$E \circ \Phi = -2\frac{x^2 - 1}{(x^2 + 1)^2}\frac{\partial}{\partial p^1} + \frac{4x}{(x^2 + 1)^2}\frac{\partial}{\partial p^2} = \frac{2}{(x^2 + 1)^2}\left((1 - x^2)\frac{\partial}{\partial p^1} + 2x\frac{\partial}{\partial p^2}\right).$$

The composition with Φ is simply saying that we should express the coefficients in the variables of the chart. We prepare some calculations

$$\frac{\partial}{\partial x}(E^1 \circ \Phi) = \frac{-8x}{(x^2 + 1)^3}(1 - x^2) + \frac{2}{(x^2 + 1)^2}(-2x)$$

$$\frac{\partial}{\partial x}(E^2 \circ \Phi) = \frac{-8x}{(x^2 + 1)^3}(2x) + \frac{2}{(x^2 + 1)^2}(2)$$

You can do the orthogonal projection in the standard linear algebra way, but because this is the plane it's easy to write down a vector perpendicular to E. This leads to

$$\frac{\partial}{\partial x^i}(E^k \circ \Phi)\frac{\partial}{\partial p^k} = \frac{-4x}{x^2 + 1}E + \frac{2}{(x^2 + 1)^2}\left(-2x\frac{\partial}{\partial p^1} + 2\frac{\partial}{\partial p^2}\right)$$

$$= \frac{-4x}{x^2 + 1}E + \frac{4}{(x^2 + 1)^3}\left[x\left((1 - x^2)\frac{\partial}{\partial p^1} + 2x\frac{\partial}{\partial p^2}\right) + \left(-2x\frac{\partial}{\partial p^1} + (1 - x^2)\frac{\partial}{\partial p^2}\right)\right]$$

$$= \frac{-2x}{x^2 + 1}E + \frac{4}{(x^2 + 1)^3}\left(-2x\frac{\partial}{\partial p^1} + (1 - x^2)\frac{\partial}{\partial p^2}\right).$$

Hence

$$\nabla_{\partial_1}^{\top} \partial_1 = \underset{T_p \mathbb{S}^1}{\text{proj}} \frac{\partial}{\partial x^i}(E_j^k \circ \Phi)\frac{\partial}{\partial p^k} = \frac{-2x}{x^2+1}E \qquad\qquad \Rightarrow \Gamma_{11}^1 = \frac{-2x}{x^2+1}.$$

Exercise 3.33 (Lee Lemma 4.4). Suppose that M is a manifold covered by a single chart U. Show that the set of covariant derivatives on M is in one-to-one correspondence with the set of Christoffel coefficients. That is, show that every choice of n^3 functions Γ_{ij}^k gives a covariant derivative.

Exercise 3.34. Derive the transformation formula for Γ_{ij}^k between two charts. Observe that it is neither covariant nor contravariant.

Exercise 3.35 (Stereographic Projection). Repeat the calculation of the Christoffel coefficients from Example 3.32 for $\mathbb{S}^2$ in the chart U_N. The following formulas may prove useful. Here we have the pushforwards of the coordinate vector fields and combinations that align with longitude and latitude:

$$E_1 = \frac{2}{(\|x\|^2+1)^2}\begin{pmatrix}-(x^1)^2+(x^2)^2+1\\-2x^1x^2\\2x^1\end{pmatrix} \qquad x^1E_1+x^2E_2 = \frac{2}{(\|x\|^2+1)^2}\begin{pmatrix}-x^1(\|x\|^2-1)\\-x^2(\|x\|^2-1)\\2\|x\|^2\end{pmatrix}$$

$$E_2 = \frac{2}{(\|x\|^2+1)^2}\begin{pmatrix}-2x^1x^2\\(x^1)^2-(x^2)^2+1\\2x^2\end{pmatrix} \qquad x^2E_1-x^1E_2 = \frac{2}{\|x\|^2+1}\begin{pmatrix}x^2\\-x^1\\0\end{pmatrix}.$$

The derivatives are

$$\frac{\partial}{\partial x^1}E_1 = \frac{-4x^1}{\|x\|^2+1}E_1 + \frac{4}{(\|x\|^2+1)^2}\left[-p+\frac{1}{2}(\|x\|^2+1)(x^1E_1+x^2E_2)\right]$$

$$\frac{\partial}{\partial x^2}E_1 = \frac{-4x^2}{\|x\|^2+1}E_1 + \frac{2}{\|x\|^2+1}\left[x^2E_1-x^1E_2\right],$$

and

$$\frac{\partial}{\partial x^1}E_2 = \frac{-4x^1}{\|x\|^2+1}E_2 - \frac{2}{\|x\|^2+1}\left[x^2E_1-x^1E_2\right]$$

$$\frac{\partial}{\partial x^2}E_2 = \frac{-4x^2}{\|x\|^2+1}E_2 + \frac{4}{(\|x\|^2+1)^2}\left[-p+\frac{1}{2}(\|x\|^2+1)(x^1E_1+x^2E_2)\right].$$

With the derivatives in this form, you should be able to calculate the Christoffel coefficients easily. For example, from

$$\nabla_{\partial_1}^{\top} \partial_1 = \frac{-2x^1}{\|x\|^2+1}E_1 + \frac{2x^2}{\|x\|^2+1}E_2$$

we read that

$$\Gamma_{11}^1 = \frac{-2x^1}{\|x\|^2+1}, \qquad\qquad \Gamma_{11}^2 = \frac{2x^2}{\|x\|^2+1}.$$

For the other derivative of E_1, the projection is trivial, and

$$\Gamma_{21}^1 = \frac{-2x^2}{\|x\|^2+1}, \qquad\qquad \Gamma_{21}^2 = \frac{-2x^1}{\|x\|^2+1}.$$

And from the derivatives of E_2 we obtain:

$$\Gamma_{12}^1 = \frac{-2x^2}{\|x\|^2 + 1}, \qquad \Gamma_{12}^2 = \frac{-2x^1}{\|x\|^2 + 1},$$
$$\Gamma_{22}^1 = \frac{2x^1}{\|x\|^2 + 1}, \qquad \Gamma_{22}^2 = \frac{-2x^2}{\|x\|^2 + 1}.$$

3.4   Parallel Transport

We began Section 3.3 with the motivation that we want to compare different tangent spaces to one another and a thought experiment about walking around the Earth. Then we went on to define covariant derivatives. Now it is time to connect the two (pardon the pun).

Definition 3.36. *Let M be a manifold with a covariant derivative ∇, $\alpha : (a, b) \to M$ a smooth curve and Y a vector field. We say that Y is* parallel *along α (with respect to ∇) if $\nabla_{\alpha'} Y = 0$ at all points on the curve.*

The inspiration of the name parallel is that the vectors of the vector field at different points are meant to be (in some sense) parallel to one another. Phrased different: we have a field of parallel vectors. Even though α' is not a vector field on M, this is well-defined due to Lemma 3.27. Similarly, we really only need to values of Y along the curve α to compute this condition, due to Lemma 3.28. Therefore many books build a theory of 'vector fields on curves'. We will avoid this extra theory by assuming the main result: so long as the curve α is injective and not pathological, every vector field on α can be extended to a vector field on M.

In a chart we have $\alpha'(t) = \frac{d\alpha^i}{dt} \partial_i$, so the condition becomes

$$(3.37) \qquad 0 = \left(\frac{d\alpha^i}{dt} \frac{\partial Y^k}{\partial x^i} + \Gamma^k_{ij} \frac{d\alpha^i}{dt} Y^j \right) \partial_j = \left(\frac{dY^k}{dt} + \Gamma^k_{ij} \frac{d\alpha^i}{dt} Y^j \right) \partial_j,$$

where we treat the vector field as a function of t, i.e. $Y(\alpha(t))$. Since Γ^k_{ij} and $\frac{d\alpha^i}{dt}$ are specified, we treat this as a system of ODEs for the functions $Y^i(t) : (a, b) \to \mathbb{R}$. By the uniqueness of solutions to ODEs, a parallel vector field is uniquely determined by its value at one point of the curve. On the other hand the existence of solutions to ODEs ensures that given a vector $v \in T_{\alpha(t_0)}$ there exists a unique parallel field Y along α with $Y(t_0) = v$.

Let us make our thought experiment rigorous by using the tangent covariant derivative. We can expand the thought experiment in the following way: while we are walking around the world without turning, we are holding a stick. The stick represents a vector field along the curve of our journey. Suppose at the start of our journey, the stick is pointing south (recall we are facing east). As we walk east around the world, our stick will continue to point south. Thus we ask whether the vector field $Y|_p = (0, 0, -1) \in T_p \mathbb{S}^2$ is parallel with respect to $\nabla^\top$ along the equator $\alpha(t) = (\cos t, \sin t, 0)$. Indeed it is, since Y is constant with respect to p,

$$\nabla^\top_{\alpha'} Y = \underset{T_p \mathbb{S}^2}{\operatorname{proj}} \left(-\sin t \, \frac{\partial Y^j}{\partial p^1} \frac{\partial}{\partial p^j} + \cos t \, \frac{\partial Y^j}{\partial p^2} \frac{\partial}{\partial p^j} + 0 \, \frac{\partial Y^j}{\partial p^3} \frac{\partial}{\partial p^j} \right) = 0.$$

Now what about the original thought experiment? This time as we walk around the world, let the stick point forward. Clearly, if we don't turn, it should continue to point forward. In other words

$$Y = \alpha'(t) = (-\sin t, \cos t, 0). = (-p^2, p^1, 0).$$

This is not constant as a function into $\mathbb{R}^3$. Now when we compute

$$\nabla^\top_{\alpha'} Y = \underset{T_p \mathbb{S}^2}{\operatorname{proj}} \left(-\sin t \, (1) \frac{\partial}{\partial p^2} + \cos t \, (-1) \frac{\partial}{\partial p^1} + 0 \right) = \underset{T_p \mathbb{S}^2}{\operatorname{proj}} (-p) = 0.$$

We see from the calculation that the derivative of Y along the curve points towards the center of the sphere, so when projected to the tangent plane it becomes zero. In summary, parallel transport by $\nabla^\top$ on the sphere matches our intuition of 'walking without turning'. Of course there are many other covariant derivatives on the sphere, and with respect to them perhaps these two vector fields are not parallel.

Example 3.38 (3-Sphere)**.** Let us consider the covariant derivative ∇^L on $\mathbb{S}^3$ from Example 3.29. We noted there that left-invariant vector fields have ∇^L-derivative zero at any point and in any direction. Hence left-invariant fields are parallel along every curve in $\mathbb{S}^3$.

Conversely, suppose Y is parallel along α. It follows from the definition of ∇^L that $t \mapsto T_{\alpha(t)}L_{\alpha(t)^{-1}}Y(\alpha(t))$ is constant. In words, if we consider Y as a function of t, ie $Y(\alpha(t))$ and move the vectors to $\mathbf{e}$ using the tangent map of the left action, ie $T_{\alpha(t)}L_{\alpha(t)^{-1}}$, then this function is constant. Though we don't have a formal definition, it is fair to say that Y is left-invariant along the curve.

The final observation for this example is that given any vector $w \in T_p\mathbb{S}^3$ there is a unique left-invariant vector field Y with $Y|_p = w$. Let $v = T_pLp^{-1}w$. Then $Y|_p = pv$ is the field. Therefore there is a unique way to parallel transport any vector to any other point of $\mathbb{S}^3$. Manifolds with this property are called *parallelisable*. It is equivalent to having a trivial tangent bundle.

In the above example, we encountered the idea of taking a vector v at one point $\alpha(t_0)$, finding a vector field Y with $Y|_{\alpha(t_0)}$ that is parallel along α, and in particular calculating the parallel vector at another point $w = Y|_{\alpha(t_1)}$. We call w the *parallel transport* of v along α. This is a function $P(\alpha)_t^s : T_{\alpha(t)}M \to T_{\alpha(s)}M$ called the *parallel transport operator*. Because the ODE is linear in Y, the parallel transport operator is linear: If Y is the parallel vector field with $Y|_{\alpha(t)} = v$ and $\tilde{Y}$ is the parallel vector field with $\tilde{Y}|_{\alpha(t)} = \tilde{v}$, then $Y + \tilde{Y}$ is also parallel and $(Y + \tilde{Y})|_{\alpha(t)} = v + \tilde{v}$. The same idea works with scaling v.

Some other properties of $P(\alpha)_t^s$ follow easily from its definition as the solution of an ODE. We have semi-group properties $P(\alpha)_t^t = \mathrm{id}$ and $P(\alpha)_s^u \circ P(\alpha)_t^s = P(\alpha)_t^u$. By the uniqueness of the solutions to ODEs, we have that $P(\alpha)_t^s$ is injective, and therefore an isomorphism of vector spaces. And so on.

Conversely, if one has the parallel transport operator for a curve α, the we can recover the covariant derivative in the direction α' through the formula

$$\nabla_{\alpha'(0)}Y = \lim_{h\to 0} \frac{1}{h}\left[P(\alpha)_h^0 Y|_{\alpha(h)} - Y|_{\alpha(0)}\right] \in T_{\alpha(0)}M.$$

Exercise 3.39. Prove the above formula. Hint: Take a basis of $T_{\alpha(0)}M$ and parallel transport it along α. As a reward for solving this exercise, you may now use the word *connection* for a covariant derivative.

Exercise 3.40. Argue that parallel transport with respect to ∇^L on $\mathbb{S}^3$ is $P(\alpha)_t^s = \alpha(t)\alpha(s)^{-1}$. If we insert this into the above equation we obtain

$$\nabla_{\alpha'(0)}^L Y = \alpha(0)\lim_{h\to 0}\frac{1}{h}\left[\alpha(h)^{-1}Y|_{\alpha(h)} - \alpha(0)^{-1}Y|_{\alpha(0)}\right].$$

Explain why this is the same formula as Example 3.29.

So intuitively the two approaches, covariant derivatives and parallel transport operators, are equivalent. The reason that it is difficult to start with parallel transport operators is that is tricky to characterise exactly when a set of linear functions between tangent spaces, one for every curve, correspond to a covariant derivative. Note our logic above: if we begin with a covariant derivative, then we have a parallel transport operator, and taking a limit we can recover the covariant derivative. But if you begin with a arbitrary set of operators, there is no guarantee that the limit will exist. You need to have some type of smooth dependence of $P(\alpha)_t^s$ on t and s. Further, what conditions should you impose on the dependence of $P(\alpha)$ on α such if two curves are tangent at a point, the above limit produces the same result. Hopefully, these questions give you an appreciation of the difficulty involved.

Special mention should go to Appendix B in Sharpe, which does start with the classical idea of rolling a plane (or another space) around on a surface and shows how that gives various modern structures on the manifold.

3.5 Torsion

In this section we discuss a quantity called torsion that is derived from a covariant derivative. There is a relation between the torsion of a connection and the torsion of a space curve, but we will not be explore it in this course[3]. Ultimately we will only be interested in covariant derivatives with zero torsion, so in a sense we are introducing it only to rule it out. Which brings us to the point: how should we motivate the definitions in this section without going deep into theory we will not use? We ask some natural questions and give some reasonable answers.

In euclidean space we have Schwarz' theorem, also known as Clairaut's theorem, that the partial derivatives with respect to different variables commute (for smooth functions among others). This result is embedded in the definition of the Lie bracket, where it was necessary to have the second order terms cancel. In fact sometimes the theorem is expressed as $[\partial_i, \partial_j] = 0$. So naturally we ask this question of the covariant derivative, but the answer is negative in general:

$$\nabla_{\partial_i}\partial_j - \nabla_{\partial_j}\partial_i = \Gamma_{ij}^k \partial_k - \Gamma_{ji}^k \partial_k = \left(\Gamma_{ij}^k - \Gamma_{ji}^k\right)\partial_k.$$

This leads to the following definition

Definition 3.41. *We say that a covariant derivative is* torsion-free *(in some chart) if* $\nabla_{\partial_i}\partial_j - \nabla_{\partial_j}\partial_i = 0$. *Equivalently in terms of Christoffel coefficients, if* $\Gamma_{ij}^k = \Gamma_{ji}^k$ *at every point.*

In this first definition, torsion of a covariant derivative is a measure of the non-commutativity of coordinate vector fields. It seems natural therefore that this should depend on the choice of chart as much as the covariant derivative. But if you have done Exercise 3.34, you may already know that if $\Gamma_{ij}^k = \Gamma_{ji}^k$ at a point in one chart then it also holds at that point in any overlapping chart. We will return to this idea shortly.

Example 3.42 (Euclidean Space). We have $\mathbb{R}^n$ with one chart, and ∇^{euc} from Example 3.21. In Example 3.30 computed that the Christoffel coefficients are all zero. Thus this covariant derivative is torsion-free in this chart.

Example 3.43 (Stereographic Projection). In Exercise 3.35 you found all the Christoffel coefficients. Observe that they are symmetric in the lower two indices

$$\Gamma_{21}^1 = \frac{-2x^2}{\|x\|^2 + 1} = \Gamma_{12}^1, \qquad \Gamma_{21}^2 = \frac{-2x^1}{\|x\|^2 + 1} = \Gamma_{12}^2.$$

This shows that the covariant derivative $\nabla^\top$ of $\mathbb{S}^2$ is torsion-free on U_N. Since the torsion is a continuous function, it must also be zero at the north pole.

We have the expectation that the coordinate vector fields should commute, or that this is a

---

[3]See the 'American football example' Lee Problem 6-1

desirable property, but we do not have that expectation for general vector fields X, Y. We find

$$
\begin{aligned}
\nabla_X Y - \nabla_Y X &= X^i \nabla_{\partial_i}(Y^j \partial_j) - Y^j \nabla_{\partial_j}(X^i \partial_i) \\
&= \left(X^i \frac{\partial Y^k}{\partial x^i} + X^i Y^j \Gamma^k_{ij} \right) \partial_k - \left(Y^j \frac{\partial X^k}{\partial x^i} + Y^j X^i \Gamma^k_{ji} \right) \partial_k \\
&= \left(X^i \frac{\partial Y^k}{\partial x^i} - Y^j \frac{\partial X^k}{\partial x^i} \right) \partial_k + X^i Y^j \left(\Gamma^k_{ij} - \Gamma^k_{ji} \right) \partial_k \\
&= [X, Y] + X^i Y^j \left(\Gamma^k_{ij} - \Gamma^k_{ji} \right) \partial_k.
\end{aligned}
$$

The meaning of this equation is that the 'covariant derivative commutator' of two vector fields is their Lie bracket plus a factor coming from the fact that the coordinate vector fields do not 'covariantly commute'.

Definition 3.44. *Given a covariant derivative ∇, we define the torsion of two vector fields X, Y to be a third vector field*

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

Remarkably the value of $T(X, Y)$ at any point p only depends on $X|_p, Y|_p$, with the formula

$$
T(X, Y) = X^i Y^j T^k_{ij} \partial_k \qquad \text{for} \qquad T^k_{ij} = \Gamma^k_{ij} - \Gamma^k_{ji}.
$$

The definition of $T(X, Y)$ is in terms of three vector fields $\nabla_X Y$, $\nabla_Y X$, and $[X, Y]$, so clearly is independent of charts. A covariant derivative is torsion-free if $T^k_{ij} = 0$, and so this too is independent of charts. The second formula is just a rearrangement of the calculation preceding the definition. We say that the second formula is remarkable because although T is defined using derivatives both of which depend on the local behaviour of vector fields, the torsion only depends on the pointwise values of the vector fields. Because the Lie bracket is an antisymmetric function of X, Y, so too is the torsion $T(X, Y) = -T(Y, X)$.

Example 3.45 (3-Sphere)**.** In this example we show that the torsion of the covariant derivative ∇^L on $\mathbb{S}^3$ from Example 3.29 is non-zero. The trick is to not work with coordinate vector fields, but rather work with left-invariant vector fields. Let $E_1|_p = p\mathbf{i}$ and likewise $E_2|_p = p\mathbf{j}, E_3|_p = p\mathbf{k}$ denote the left-invariant vector fields that are obtained by pushing forward $\mathbf{i}, \mathbf{j}, \mathbf{k} \in T_e\mathbb{S}^3$. We have already noted in Example 3.38 that $\nabla^L_v E_i = 0$ for any vector $v \in T_p M$.

Further at any point $E_1|_p, E_2|_p, E_3|_p$ is a basis for $T_p\mathbb{S}^3$. This means that every vector field X on $\mathbb{S}^3$ can be written as

$$
X = X^1 E_1 + X^2 E_2 + X^3 E_3.
$$

Thus E_i have similar properties to the coordinate vector basis field, except that they do not come from coordinates. A set of vector fields with this basis property is called a *frame*, but we will not explore this concept in generality. In this frame, the covariant derivative can be reckoned with

$$
\nabla^L_X Y = \nabla^L_X \left(Y^j E_j \right) = X(Y^j) E_j + Y^j \nabla^L_X E_j = X(Y^j) E_j.
$$

Similarly the Lie bracket simplifies

$$[E_i, Y] = [E_i, Y^j E_j] = E_i(Y^j) E_j + Y^j [E_i, E_j]$$
$$[X, Y] = [X^i E_i, Y] = X^i [E_i, Y] - Y(X^i) E_i$$
$$= X^i E_i(Y^j) E_j + X^i Y^j [E_i, E_j] - Y^j E_j(X^i) E_i$$
$$= X(Y^j) E_j + X^i Y^j [E_i, E_j] - Y(X^i) E_i.$$

Together this yields

$$T^L(X, Y) = \nabla_X^L Y - \nabla_Y^L X - [X, Y]$$
$$= X(Y^j) E_j - Y(X^j) E_j - X(Y^j) E_j - X^i Y^j [E_i, E_j] + Y(X^i) E_i$$
$$= -X^i Y^j [E_i, E_j].$$

Thus the torsion comes down to the Lie brackets of this frame.

For this example we will evaluate $[E_1, E_2]$:

$$[E_1, E_2] = [p\mathbf{i}, p\mathbf{j}] = [-p^1 + p^0 \mathbf{i} + p^3 \mathbf{j} - p^2 \mathbf{k}, -p^2 - p^3 \mathbf{i} + p^0 \mathbf{j} + p^1 \mathbf{k}]$$
$$= [-p^1 \partial_0 + p^0 \partial_1 + p^3 \partial_2 - p^2 \partial_3, -p^2 \partial_0 - p^3 \partial_1 + p^0 \partial_2 + p^1 \partial_3]$$
$$= -p^1 \partial_2 + p^0 \partial_3 + p^3(-\partial_0) - p^2(-\partial_1) - \left[-p^2 \partial_1 - p^3(-\partial_0) + p^0(-\partial_3) + p^1 \partial_2 \right]$$
$$= -2p^3 \partial_0 + 2p^2 \partial_1 - 2p^1 \partial_2 + 2p^0 \partial_3 = 2E_3.$$

We can generalise this argument; set $\mathbf{i}_1 = \mathbf{i}, \mathbf{i}_2 = \mathbf{j}, \mathbf{i}_3 = \mathbf{k}$ so that we can use index notation.

$$[E_i, E_j] = [p\partial_i, p\partial_j] = p\mathbf{i}_i \partial_j - p\mathbf{i}_j \partial_i = p(\mathbf{i}_i \mathbf{i}_j - \mathbf{i}_j \mathbf{i}_i).$$

When $i = j$, the quaternions commute and the bracket is zero (as expected). If they are not equal then the quaternions anti-commute. This gives $[E_2, E_3] = 2E_1$ and $[E_3, E_1] = 2E_2$. (There is in fact a close relationship between the Lie bracket of $\mathbb{S}^3$ and the cross product of $\mathbb{R}^3$).

Example 3.46 (3-Sphere). We can also ask for the torsion of ∇^R on $\mathbb{S}^3$. Of course we could do the same as the previous example, except using a right-invariant frame, and get a similar answer. But to make the two examples comparable, let us compute the torsion of ∇^R using the left-invariant frame E_i.

What changes about the calculation is that $\nabla_{E_i}^R E_j \neq 0$. Instead we must generalise the calculation from Example 3.29:

$$E_i(p) = p\partial_i p = p\mathbf{i}_i,$$
$$\nabla_{E_i}^R E_j = \left(E_i(p\mathbf{i}_j p^{-1}) \right) p = \left(E_i(p)\mathbf{i}_j p^{-1} - p\mathbf{i}_j p^{-1} E_i(p) p^{-1} \right) p = p\mathbf{i}_i \mathbf{i}_j - p\mathbf{i}_j \mathbf{i}_i = [E_i, E_j].$$

The covariant derivative of an arbitrary vector field is

$$\nabla_X^R Y = X(Y^j) E_j + X^i Y^j \nabla_{E_i}^R E_j = X(Y^j) E_j + X^i Y^j [E_i, E_j].$$

Hence

$$
\begin{aligned}
T^R(X,Y) &= \nabla^R_X Y - \nabla^R_Y X - [X,Y] \\
&= X(Y^j)E_j + X^i Y^j [E_i, E_j] - Y(X^j)E_j - X^i Y^j [E_j, E_i] \\
&\quad - X(Y^j)E_j - X^i Y^j [E_i, E_j] + Y(X^i)E_i \\
&= X^i Y^j [E_i, E_j].
\end{aligned}
$$

Thus the torsion of ∇^R is the negative of the torsion of ∇^L.

Recall Exercise 3.24 that given one connection we can create another by the addition of a vector valued function $A(X,Y)$. We can ask how the torsion of the new covariant derivative related to the torsion of the original. This follows easily, for $\tilde{\nabla} = \nabla + A$,

$$
\begin{aligned}
T^{\tilde{\nabla}}(X,Y) &= \tilde{\nabla}_X Y - \tilde{\nabla}_Y X - [X,Y] = \nabla_X Y + A(X,Y) - \nabla_Y X - A(Y,X) - [X,Y] \\
&= T^{\nabla}(X,Y) + A(X,Y) - A(Y,X).
\end{aligned}
$$

Purely algebraically, for any function of two variables we can split it into a symmetric and antisymmetric parts

$$
A(X,Y) = \frac{1}{2}\Big(A(X,Y) + A(Y,X)\Big) + \frac{1}{2}\Big(A(X,Y) - A(Y,X)\Big).
$$

If A is already symmetric or antisymmetric, then it is just equal to its symmetric or antisymmetric part respectively and the other part is zero. Thus we can express the relationship of the torsions by the dictum "adding A to a covariant derivative adds twice the antisymmetric part of A to its torsion". In particular, for any covariant derivative, we can *absorb the torsion*. This means we construct a new torsion-free covariant derivative $\tilde{\nabla} := \nabla - \frac{1}{2}T$.

Example 3.47 (3-Sphere). We have just seen in Examples 3.45 and 3.46 that with respect to the left-invariant fields E_i the covariant derivatives are

$$
\begin{aligned}
\nabla^L_{E_i} E_j &= 0 & \nabla^R_{E_i} E_j &= [E_i, E_j] \\
T^L(E_i, E_j) &= -[E_i, E_j] & T^R(E_i, E_j) &= [E_i, E_j].
\end{aligned}
$$

(Aside: the formula on the right makes it seem as if ∇^R and T^R are equal. They are not in general, only for left-invariant vector fields. Remember: a covariant derivative has the product rule in Y, whereas the torsion is C^∞-linear.)

If we absorb the torsion on these two connections we get the torsion-free connection

$$
\nabla^{LC}_{E_i} E_j = \frac{1}{2}[E_i, E_j] = \nabla^L_{E_i} E_j + \frac{1}{2}[E_i, E_j] = \nabla^R_{E_i} E_j - \frac{1}{2}[E_i, E_j].
$$

This fits nicely with Corollary 3.25, because ∇^{LC} can also be understood as the average of the left and right covariant derivatives: $\nabla^{LC} = \frac{1}{2}\nabla^L + \frac{1}{2}\nabla^R$. I'll give you one guess what the LC stands for!

We have seen now that for a torsion-free connection that the coordinate vector fields will 'covariant commute' but general vector fields will not.

Definition 3.48. *A smooth family of curves is a function* $\alpha_s(t) : (-\varepsilon, \varepsilon) \times (a, b) \to M$. *By smooth family we mean that it is smooth in both variables* s *and* t. *We typically think of the main* curves of the family $t \mapsto \alpha_s(t)$ *for fixed* s. *But we also have the* transverse *curves, where we fix* t *and allow* s *to vary. We can write* $\alpha(s, t)$ *to emphasise this duality.*

Therefore we have two vector fields: the tangents in the main direction and the tangents in the transverse direction. Well, this is not completely true as we do not really have vector fields because the curves may cross each other, giving multiple vectors at the same point. (Technically what we have is the pushforwards of two vector fields.) Regardless, for each value of (s, t) it makes sense to ask how the derivative $\partial_s \alpha$ is changing in comparison to $\partial_t \alpha$.

Lemma 3.49 (Mixed Derivatives)**.** *Let* ∇ *be a torsion-free covariant derivative and* $\alpha(s, t) : (-\varepsilon, \varepsilon) \times (a, b) \to M$ *a smooth family of curves. Then* $\nabla_{\partial_s \alpha} \partial_t \alpha = \nabla_{\partial_t \alpha} \partial_s \alpha$.

Proof. This is a purely computational proof. In a chart, the tangent vectors are

$$\partial_s \alpha = \frac{\partial \alpha^k}{\partial s} \partial_k, \qquad \partial_t \alpha = \frac{\partial \alpha^k}{\partial t} \partial_k.$$

Then

$$\nabla_{\partial_s \alpha} \partial_t \alpha = \left(\frac{\partial^2 \alpha^k}{\partial s \partial t} + \Gamma_{ij}^k \frac{\partial \alpha^i}{\partial s} \frac{\partial \alpha^j}{\partial t} \right) \partial_k,$$

$$\nabla_{\partial_t \alpha} \partial_s \alpha = \left(\frac{\partial^2 \alpha^k}{\partial t \partial s} + \Gamma_{ij}^k \frac{\partial \alpha^i}{\partial t} \frac{\partial \alpha^j}{\partial s} \right) \partial_k.$$

By the symmetry of the Christoffel coefficients for torsion-free covariant derivatives, these are equal. □

We should comment about why the expression $\nabla_{\partial_s \alpha} \partial_t \alpha$ is well-defined even though the tangents do not necessarily form a vector field. We know that the direction of ∇ depends only on the pointwise value, so this is no issue. And for $\partial_t \alpha$ we need to know its values along a curve in the direction of $\partial_s \alpha$, but this is exactly the meaning of partial derivative. So understood correctly, these expressions are valid. This is an instance where a fleshed out notion of 'vector field on a curve' would have been more precise, but hopefully you see that not much has been lost by skipping this concept.

3.6   The Levi-Civita connection

Let us once more return to the thought experiment of walking along the equator $\alpha(t) = (\cos t, \sin t, 0)$ with our stick. We now understand that we are parallel transporting our stick. But consider the vector field $Z(\alpha(t)) = (0, 0, -\cos^2 t)$. To push the metaphor into silliness, it is an telescoping selfie stick that is lengthening and shortening. The vector field Z always points south, but it is not parallel according to definition. If we write $Z = \cos^2 t\, Y$ for $Y(\alpha(t)) = (0, 0, -1)$, a known parallel vector field, then

$$\nabla_{\alpha'}^{\top} Z = \nabla_{\alpha'}^{\top}(\cos^2 t\, Y) = \frac{d}{dt}\Big(\cos^2 t \Big) Y + \cos^2 t\, \nabla_{\alpha'}^{\top} Y = -2\sin t \cos t\, Y \not\equiv 0.$$

This illustrates the point that parallel is about more than just direction, it also concerns length (which is unlike how we use the term in elementary geometry and linear algebra). Therefore, among the many covariant derivatives that exists on a Riemannian manifold, we are interested in those whose parallel transport preserves length and angle.

Let us now turn this intuition into a definition. Suppose M is a Riemannian manifold with metric g and that ∇ is a connection that preserves the lengths and angles of parallel transport vectors. For any curve γ, let X, Y be parallel fields along γ with respect to ∇. This means that $g(X, Y)$ is a constant function along γ. For all smooth functions a, b, we must have

$$\frac{d}{dt} g(aX, bY) = \frac{d}{dt}\Big(ab\, g(X, Y) \Big) = \frac{da}{dt} bg(X, Y) + a\frac{db}{dt} g(X, Y) + ab\frac{d}{dt} g(X, Y)$$
$$= g(a'X, bY) + g(aX, b'Y) + 0.$$

On the other hand

$$\nabla_{\gamma'}(aX) = a'X + a\nabla_{\gamma'}X = a'X.$$

Therefore we make the definition

Definition 3.50. *A covariant derivative ∇ is called* metric-compatible *or a* metric connection *if for all vector fields X, Y, Z*

$$Z\big(g(X, Y)\big) = g\big(\nabla_Z X, Y\big) + g\big(X, \nabla_Z Y\big).$$

The choice to define this property using a third vector field Z instead of the tangent vector γ' is purely a matter of style. The converse of the above argument is immediate: if X, Y are parallel along a curve γ then the right hand side is zero and thus $g(X, Y)$ is constant on the curve.

> **Example 3.51** (3-Sphere)**.** We can show that the left and right covariant derivatives are compatible with the metric on $\mathbb{S}^3$ coming from $\mathbb{R}^4$. Write vector fields $X = X^i E_i$ and $Y = Y^i E_i$ with respect to the left-invariant basis fields from Example 3.45. By the property of quaternions that $a \cdot b = \operatorname{Re} \bar{a}b$ we see that
>
> $$E_i \cdot E_j = \operatorname{Re} \overline{p\mathbf{i}_i} p\mathbf{i}_j = \operatorname{Re} \bar{\mathbf{i}}_i \bar{p} p \mathbf{i}_j = \operatorname{Re} \bar{\mathbf{i}}_i \mathbf{i}_j = \mathbf{i}_i \cdot \mathbf{i}_j,$$
>
> since $p \in \mathbb{S}^3$ has unit length. In particular it is constant on all of $\mathbb{S}^3$. Additionally, the covariant derivatives of the E_i are zero in every direction. Therefore, similar to the calculation before

the definition, we have

$$\begin{aligned}
Z(X \cdot Y) = Z(X^i Y^j \, E_i \cdot E_j) &= Z(X^i) Y^j \, E_i \cdot E_j + X^i Z(Y^j) \, E_i \cdot E_j \\
&= \big(Z(X^i) E_i\big) \cdot Y + X \cdot \big(Z(Y^j) E_j\big) \\
&= \big(Z(X^i) E_i + X^i \nabla_Z^L E_i\big) \cdot Y + X \cdot \big(Z(Y^j) E_j + Y^j \nabla_Z^L E_j\big) \\
&= \big(\nabla_Z^L X\big) \cdot Y + X \cdot \big(\nabla_Z^L Y\big).
\end{aligned}$$

This shows that ∇^L is metric-compatible.

For ∇^R we can reuse some of this calculation. What changes is that $\nabla_Z^R E_i$ may not be zero. Instead $\nabla_Z^R E_i = Z^k [E_k, E_i]$. We need to prove a version of the cyclic property for the triple product (for vectors in $\mathbb{R}^3$ we have $a \cdot (b \times c) = b \cdot (c \times a)$):

$$\begin{aligned}
[E_k, E_i] \cdot E_j + E_i \cdot [E_k, E_j] &= \mathrm{Re}\,\overline{(\mathbf{i}_k \mathbf{i}_i - \mathbf{i}_i \mathbf{i}_k)} \mathbf{i}_j + \mathrm{Re}\,\overline{\mathbf{i}}_i (\mathbf{i}_k \mathbf{i}_j - \mathbf{i}_j \mathbf{i}_k) \\
&= \mathrm{Re}(\mathbf{i}_i \mathbf{i}_k \mathbf{i}_j - \mathbf{i}_k \mathbf{i}_i \mathbf{i}_j - \mathbf{i}_i \mathbf{i}_k \mathbf{i}_j + \mathbf{i}_i \mathbf{i}_j \mathbf{i}_k) \\
&= \mathrm{Re}(-\mathbf{i}_k \mathbf{i}_i \mathbf{i}_j + \mathbf{i}_i \mathbf{i}_j \mathbf{i}_k) = 0.
\end{aligned}$$

This allows us to write

$$\begin{aligned}
Z(X \cdot Y) &= \big(Z(X^i) E_i\big) \cdot Y + X \cdot \big(Z(Y^j) E_j\big) + X^i Y^j Z^k ([E_k, E_i] \cdot E_j + E_i \cdot [E_k, E_j]) \\
&= \big(Z(X^i) E_i\big) \cdot Y + X \cdot \big(Z(Y^j) E_j\big) + X^i \nabla_Z^R E_i \cdot Y + Y^j X \cdot \nabla_Z^R E_j \\
&= \big(\nabla_Z^R X\big) \cdot Y + X \cdot \big(\nabla_Z^R Y\big).
\end{aligned}$$

This proves that ∇^R is also metric-compatible.

It is useful to reduce the metric-compatibility condition to a condition on the Christoffel coefficients in some chart.

Lemma 3.52. *Let a connection ∇ in some chart be described by the Christoffel coefficients Γ_{ij}^k. It is compatible with the metric if and only if*

$$(3.53) \qquad\qquad \partial_k g_{ij} = \Gamma_{ki}^l g_{lj} + \Gamma_{kj}^l g_{il}$$

Proof. Notice that the formula for metric-compatibility is C^∞-linear in Z, so it enough to show it holds for each coordinate basis vector. The following calculation is a set of equivalences:

$$\begin{aligned}
\partial_k \big(g(X, Y)\big) &= g\big(\nabla_{\partial_k} X, Y\big) + g\big(X, \nabla_{\partial_k} Y\big) \\
\partial_k \big(X^i Y^j g_{ij}\big) &= g\big((\partial_k X^i + X^l \Gamma_{kl}^i) \partial_i, Y^j \partial_j\big) + g\big(X^i \partial_i, (\partial_k Y^j + Y^l \Gamma_{kl}^j) \partial_j\big) \\
\partial_k X^i Y^j g_{ij} + X^i \partial_k Y^j g_{ij} + X^i Y^j \partial_k g_{ij} &= (\partial_k X^i + X^l \Gamma_{kl}^i) Y^j g_{ij} + X^i (\partial_k Y^j + Y^l \Gamma_{kl}^j) g_{ij} \\
X^i Y^j \partial_k g_{ij} &= X^l \Gamma_{kl}^i Y^j g_{ij} + X^i Y^l \Gamma_{kl}^j g_{ij} \\
\partial_k g_{ij} &= \Gamma_{ki}^l g_{lj} + \Gamma_{kj}^l g_{il}.
\end{aligned}$$

In other words, a covariant derivative is metric-compatible if and only if its Christoffel coefficients satisfy (3.53). $\qquad\square$

The above equation seems to say that metric-compatibility is a rather strong condition. We know that there are n^3 choices of smooth functions for the Christoffel coefficients, and counting

the possible values for i, j, k gives n^3 conditions. It is almost enough to guarantee uniqueness, but not quite, because the we get the same condition if we swap i and j. However, metric-compatibility and torsion-free are enough to ensure uniqueness. This result is given a rather impressive sounding name, though sometimes it is called a theorem and other times a lemma. We have our cake and eat it too:

Theorem 3.54 (Fundamental Lemma of Riemannian Geometry). *On every Riemannian manifold there exists a unique metric-compatible torsion-free covariant derivative.*

Proof. Our strategy for the proof is as follows. First we will establish the so-called Koszul formula. Uniqueness is then a direct consequence. To prove existence we will show that the Koszul formula defines a torsion-free metric-compatible covariant derivative in every chart. Since we already have uniqueness, we can conclude that these give a well-defined covariant derivative on the whole manifold.

The idea of the Koszul formula is to use the symmetries of the metric and the Lie bracket to get an expression with exact one covariant derivative. Begin with the metric-compatibility property and then use the fact that torsion is zero:

$$\begin{aligned} Z\big(g(X,Y)\big) &= g\big(\nabla_Z X, Y\big) + g\big(X, \nabla_Z Y\big) \\ &= g\big(\nabla_Z X, Y\big) + g\big(X, T(Z,Y) + \nabla_Y Z + [Z,Y]\big) \\ &= g\big(\nabla_Z X, Y\big) + g\big(X, \nabla_Y Z\big) + g\big(X, [Z,Y]\big). \end{aligned}$$

Now write this equation two more times with the vector fields permuted

$$\begin{aligned} Y\big(g(Z,X)\big) &= g\big(\nabla_Y Z, X\big) + g\big(Z, \nabla_X Y\big) + g\big(Z, [Y,X]\big) \\ X\big(g(Y,Z)\big) &= g\big(\nabla_X Y, Z\big) + g\big(Y, \nabla_Z X\big) + g\big(Y, [X,Z]\big). \end{aligned}$$

Notice that of the six possible permutations, only $\nabla_Z X$, $\nabla_Y Z$ and $\nabla_X Y$ occur. This is a result of using the torsion-free property. Each of the three covariant derivatives occurs twice. Now, add any two equations and subtract the other. We will add the second and third and subtract the first, but it's not important which you choose.

$$\begin{aligned} X\big(g(Y,Z)\big) &+ Y\big(g(Z,X)\big) - Z\big(g(X,Y)\big) \\ &= 2g\big(Z, \nabla_X Y\big) + g\big(Z, [Y,X]\big) + g\big(Y, [X,Z]\big) - g\big(X, [Z,Y]\big). \end{aligned}$$

If you like, you can clean this up a little, though the role each of the vector fields play in $g\big(Z, \nabla_X Y\big)$ is different, so there cannot be perfect symmetry in the formula. Here is a version I like:

$$\begin{aligned} 2g\big(\nabla_X Y, Z\big) = \; &X\big(g(Y,Z)\big) - g\big(X, [Y,Z]\big) \\ &+ Y\big(g(X,Z)\big) - g\big(Y, [X,Z]\big) \\ &- Z\big(g(X,Y)\big) + g\big(Z, [X,Y]\big) \end{aligned}$$

This is the Koszul formula. Since the metric is non-degenerate, what it shows is that if there is a metric-compatible torsion-free covariant derivative then it can be calculated purely in terms of Lie brackets and inner products. Therefore we have established uniqueness.

For existence, it is possible to take the Koszul formula as the definition and directly check all the required properties. It is easier however to first reduce the Koszul formula to an expression

in charts. Choose any chart and suppose that $X = \partial_i, Y = \partial_j, Z = \partial_k$ are coordinate vector fields. The Lie brackets are zero. We get

$$2g(\Gamma^l_{ij}\partial_l, \partial_k) = 2\Gamma^l_{ij}g_{lk} = \partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}.$$

If we view the left hand side in matrix notation rather than index notation, we see that to solve for Γ we need to invert the matrix $G = (g_{ij})$. There is a sneaky convention that the components of the inverse matrix use upper indices $(g^{ij}) = G^{-1}$. With this convention, the fact that these matrices are inverse can be written $g^{ij}g_{jk} = \delta^i_k$. In index notation, multiplying by the inverse matrix looks like $\Gamma^l_{ij}g_{lk}g^{km} = \Gamma^l_{ij}\delta^m_l = \Gamma^m_{ij}$. Thus we can write

(3.55) $$\Gamma^m_{ij} = \frac{1}{2}g^{km}\big(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}\big).$$

So given a metric g, define a covariant derivative on this chart using this formula for the Christoffel coefficients. Exercise 3.33 tells us that this does indeed define a covariant derivative on this chart, but it remains to show that it is metric-compatible and torsion-free. Torsion-free is an easy because the above formula is symmetric in i and j. Using the Christoffel coefficients defined through the Koszul formula, we see that the condition of Lemma 3.52 satisfied:

$$2\Gamma^l_{ki}g_{lj} + 2\Gamma^l_{kj}g_{il} = \big(\partial_k g_{ij} + \partial_i g_{kj} - \partial_j g_{ki}\big) + \big(\partial_k g_{ji} + \partial_j g_{ki} - \partial_i g_{kj}\big) = 2\partial_k g_{ij}.$$

Therefore the covariant derivative that we have defined in each chart is metric-compatible and torsion-free. As mentioned at the outset of the proof, it only remains to show that this definition in each chart agrees, but this follows due to uniqueness. $\qquad\square$

We celebrate this result with more terminology. It honours the Italian mathematician Tullio Levi-Civita, who developed much of the 'tensor calculus' (covariant, contravariant, indices, etc). His name tricks many students (myself included) into thinking there are two mathematicians Levi and Civita. In response to being asked what he liked best about Italy, Einstein once said "spaghetti and Levi-Civita".

Definition 3.56. *The unique metric-compatible torsion-free covariant derivative on a Riemannian manifold is called the* Levi-Civita connection *or the* Riemannian connection.

> **Example 3.57** (Euclidean Space)**.** On any open subset of $\mathbb{R}^n$ with the dot product as metric, the Levi-Civita connection is ∇^{euc} from Example 3.21. Because its Christoffel coefficients are identically zero, obviously it is torsion-free and satisfies Equation (3.53) so is metric-compatible.

> **Example 3.58** (3-Sphere)**.** A corollary of Lemma 3.52 is that the set of metric-compatible covariant derivatives has a affine structure. Corollary 3.25 shows us that the affine combination of covariant derivatives is again a covariant derivative. The Christoffel coefficients of the new covariant derivative is the same affine combination of Christoffel coefficients $(\Gamma^t)^k_{ij} = (1-t)(\Gamma^0)^k_{ij} + t(\Gamma^1)^k_{ij}$. Inserting this into Equation (3.53) shows that such an affine combination is also metric-compatible.
>
> We saw in Example 3.51 that both ∇^L and ∇^R were metric-compatible. Therefore $\nabla^{LC} = \frac{1}{2}\nabla^L + \frac{1}{2}\nabla^R$ is metric-compatible. Additionally, we proved in Example 3.47 that it is torsion-

free. Therefore ∇^{LC} really is the Levi-Civita connection for $\mathbb{S}^3$.

Another obvious example of a Levi-Civita connection would be $\nabla^\top$ on $\mathbb{S}^2$. Instead of proving it for the specific case, instead we generalise the construction to any Riemannian immersed submanifold.

Definition 3.59 (Tangent Connection)**.** *Let $M \to N$ be an Riemannian immersed submanifold. We identify the manifold M with its image under the immersion to simplify the statement. Let ∇^N be the Levi-Civita connection of N. We define the* tangent connection *on M to be the covariant derivative*

$$(\nabla_X^\top Y)|_p = \operatorname*{proj}_{T_pM} \nabla_X^N Y.$$

This is definition extends the previous definition from Example 3.26 because ∇^{euc} is the Levi-Civita connection of $\mathbb{R}^n$.

Theorem 3.60 (Gauss Formula)**.** *The tangent connection is the Levi-Civita connection of M.*

Proof. The proof that it is in fact a covariant derivative is entirely similar to the corresponding statement in Example 3.26. We check the three properties of a covariant derivative:

$$\nabla_{fX+\tilde{X}}^\top Y = \operatorname*{proj}_{T_pM} \nabla_{fX+\tilde{X}}^N Y = \operatorname*{proj}_{T_pM} \left(f\nabla_X^N Y + \nabla_{\tilde{X}}^N Y \right) = f\nabla_X^\top Y + \nabla_{\tilde{X}}^\top Y,$$

$$\nabla_X^\top (Y + \tilde{Y}) = \operatorname*{proj}_{T_pM} \left(\nabla_X^N Y + \nabla_X^N \tilde{Y} \right) = \nabla_X^\top Y + \nabla_X^\top \tilde{Y},$$

$$\nabla_X^\top (fY) = \operatorname*{proj}_{T_pM} \left(X(f)Y + \nabla_X^N Y \right) = X(f)Y + f\nabla_X^\top Y,$$

using that Y is already tangent to M. You might observe that this part of the proof works for any covariant derivative and that we have not yet used the metric-compatibility or torsion-free of ∇^N.

For torsion-free, we need to know that if X, Y are tangent to M that $[X, Y]$ is too, even when we consider them as vector fields on N. To prove this fact requires a proper investigation of submanifolds, and the construction of a special chart on N that aligns with a chart on M. This is beyond the scope of this course, which has tried to avoid manifold theory as much as possible. We have seen an example of this phenomenon though: in Example 3.45 the Lie bracket of the E_i fields was again an E_i field. Assuming this result,

$$T^\top(X, Y) = \nabla_X^\top Y - \nabla_Y^\top X - [X, Y] = \operatorname*{proj}_{T_pM} \left(\nabla_X^N Y - \nabla_Y^N X - [X, Y] \right) = \operatorname*{proj}_{T_pM} T^N(X, Y)$$

is zero. The generalised statement for arbitrary connections would be that the tangent connection is torsion-free iff the torsion of ∇^N is perpendicular to T_pM at every point of M. Though we hadn't defined it, one could also say iff T^N lies in the normal bundle of M.

Lastly, we need to show that the tangent connection is metric-compatible. This is where we need to use that M is Riemannian immersed, so that the metric on M and the metric on N

agree for tangent vectors to M.

$$
\begin{aligned}
Z\big(g^M(X,Y)\big) = Z\big(g^N(X,Y)\big) &= g^N\big(\nabla_Z^N X, Y\big) + g^N\big(X, \nabla_Z^N Y\big) \\
&= g^N\big(\operatorname*{proj}_{T_p M} \nabla_Z^N X, Y\big) + g^N\big(X, \operatorname*{proj}_{T_p M} \nabla_Z^N Y\big) \\
&= g^N\big(\nabla_Z^\top X, Y\big) + g^N\big(X, \nabla_Z^\top Y\big) \\
&= g^M\big(\nabla_Z^\top X, Y\big) + g^M\big(X, \nabla_Z^\top Y\big).
\end{aligned}
$$

To explain the working here a little, for any vector in $T_p N$ we can split it into a part in $T_p M$ and a part perpendicular to $T_p M$. Because $Y \in T_p M$, the inner product of Y with a vector perpendicular to $T_p M$ is zero. Thus we can go from the first to the second line.

Now we know that $\nabla^\top$ is a metric-compatible torsion-free covariant derivative on M. By the uniqueness in Theorem 3.54, it is the Levi-Civita connection. $\qquad\square$

To close the chapter, we revisit the question "why torsion-free"? Our first answer was that it is a natural expectation, based on the commutativity of partial derivatives in the euclidean setting. Our second answer is that torsion-free is a matter of convenience:

- The most common connections, namely the euclidean connection ∇^{euc} and the tangent connection $\nabla^\top$ are torsion-free.

- The Levi-Civita connection is unique and the Koszul formula allows it to be calculated.

- For many applications, whether or not a connection has torsion makes no difference. So absorbing the torsion doesn't lead to a loss of generality. We will see this in action in the next chapter.