COMP760: Geometry and Generative Models Week 2: Geometry Primer II

Notes by: Joey Bose and Prakash Panangaden

References

There are a lot of excellent books on differential geometry and differential topology. I am not going to enumerate and comment on them all. I learned this material from notes of Robert Geroch Geroch [2013], these were quite brief but insightful unpublished notes from 1972 which eventually appeared in print in 2013. The textbook on General Relativity by Wald Wald [1984] gives a more detailed treatment in the same spirit. The most thorough modern treatment is in "Introduction to Smooth Manifolds" by John Lee Lee [2013]. His earlier book on Riemannian manifolds is also very good Lee [1997]. The old classic is by Kobayashi and Nomizu Kobayashi and Nomizu [1963]. I can also recommend Boothby [2003] and the wonderful little undergraduate book, "Calculus on manifolds" by Spivak Spivak [1965]. There is also a 5 volume book by Spivak Spivak [1979] which is just too much to read.

1

DEFINING MANIFOLDS

The point of differential manifolds is to be able to do differential calculus on curved surfaces. Consider the usual formula for the derivative of a function $f : \mathbf{R} \longrightarrow \mathbf{R}$:

$$\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0} = \lim_{\varepsilon \longrightarrow 0} \frac{f(x_0+\varepsilon) - f(x_0)}{\varepsilon}.$$

This makes perfect sense as written, but if we look closely, we see that there is the expression $x_0 + \varepsilon$ which does not make sense if x_0 is a point on a curved surface. All the other occurrences of $+$ or $-$ signs refers to real numbers, but this one refers to "adding two points in a space." Now if we have a *vector space* it is fine; this is the crucial property of vector spaces: they are equipped with a notion of addition. We cannot define addition of points in an arbitrary topological space; we need additional structure. We see that we are working "close to" x_0 so some kind of "local" definition is enough. This is precisely what topology is equipped to formalize.

So the strategy will be to define "patches" of the topological space that "look like" patches of a vector space and then glue them together. The space will locally look like a vector space but globally might be very different. This motivates the definition of a differential or smooth manifold. The word "smooth" is a synonym for *infinitely differentiable* or C^∞.

Definition 1.0.1 *An n-dimensional* ***smooth (or differential) manifold*** *is a topological space M, assumed to be paracompact and Hausdorff, equipped with a family of pairs, called* ***charts****, $\{(U_i, \phi_i) | i \in \mathcal{A}\}$ where:*

- *Each U_i is an open subset of M,*
- *each $\phi_i : U_i \longrightarrow \mathbf{R}^n$ is a* homeomorphism *between U_i and the image $V_i := \phi_i(U_i)$,*
- *the $\{U_i\}$ form an open cover of M.*

In addition, the following compatibility condition *must be satisfied:*
if $U_i \cap U_j \neq \emptyset$ then the map

$$\left.\phi_j \circ \phi_i^{-1}\right|_{\phi_i(U_i \cap U_j)} : \phi_i(U_i \cap U_j) \longrightarrow \phi_j(U_i \cap U_j)$$

is infinitely differentiable, written as C^∞. A collection of compatible charts is called an ***atlas****.*

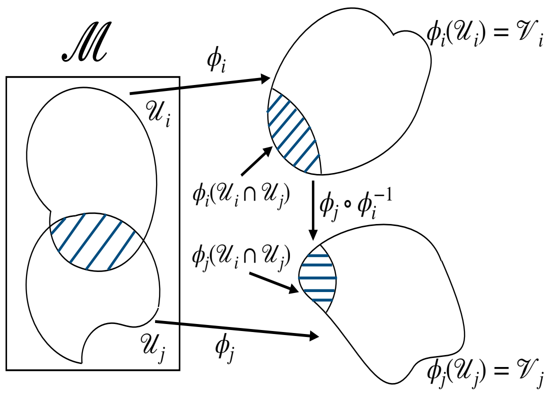

Figure 1.1: The compatibility condition on charts

Note that the very last condition refers to a map between an open set in $\mathbf{R}^n$ and another open set in $\mathbf{R}^n$; hence its meaning is clear since $\mathbf{R}^n$ is a vector space[1]. The picture in Fig. 2.1 illustrates the meaning of this condition.

One can imagine less stringent compatibility conditions: for example, we could require the maps to be k times differentiable or even just continuous. In the latter case we have what is called a *topological manifold*. In the case of complex manifolds we often require that the compatibility condition be that the maps are *analytic* instead of just C^∞. In these notes we will stick to the C^∞ case; recall, the word "smooth" is a synonym for C^∞.

Given an atlas, if there is a chart compatible with the charts of the given atlas we can add it to the atlas and obtain a new bigger atlas. It is always possible to expand an atlas to include all possible compatible charts: this is called a *maximal* atlas. A particular maximal atlas defines a notion of differentiability and is called a differentiable structure. It raises the question: given a topological manifold is there a unique maximal atlas or differentiable structure on it? The answer is "no" and it is easy to give examples even for a simple manifold like $\mathbf{R}$. However, there is a notion of "equivalence" of differential structures called *diffeomorphism* to be defined below. So, while a manifold might have many differential structures perhaps they are all diffeomorphic. So the question really is, "does a given topological manifold have a unique differential structure up to diffeomorphism?"

[1]Familiarity with multivariable calculus is assumed.

For one, two and three dimensional manifolds the answer is "yes" and was known classically. In 1956 John Milnor discovered that the 7-dimensional sphere has 28 different non-diffeomorphic differential structures! This led to intense activity culminating in the 1980's when it was discovered that for $n \neq 4$ the manifolds $\mathbf{R}^n$ have a *unique* differentiable structure but for $n = 4$ there are *uncountably many* inequivalent differential structures.

The collection of charts are used to move between the manifold and open subsets of $\mathbf{R}^n$ so that notions from $\mathbf{R}^n$ can be imported to the manifold. We fix an n-manifold M for the subsequent discussion and always use U_i to denote the open sets and ϕ_i to denote the mappings associated with the chart, the index set is taken to be I. We often call a chart (U, ϕ) a *patch*. Given a point $p \in U_i \subset M$ we have $\phi_i(p) \in \mathbf{R}^n$ so $\phi_i(p) = (x_1, \ldots, x_n)$. We think of this n-tuple of numbers as *coordinates* of p; thus the charts should be thought of as coordinates defined on patches of the manifold. The compatibility conditions means that when two patches overlap we can change coordinates without changing what it means to be differentiable. Of course, we will change the *numerical values* of the derivatives as they will be defined via the charts but the *concept of derivative* does not depend on which chart is used. This should not alarm you: you know very well that if you change the basis of a vector space the numerical values describing the components of a vector will change but the intrinsic vector does not change; we are only changing how we describe it.

Why can't we just have one coordinate system instead of all these patches and have to worry about changing coordinates? It would be great but, except in very special cases (basically in $\mathbf{R}^n$ or open subsets of them), it is simply not possible. There is no one coordinate system that can cover, for example, the two-dimensional sphere S^2. We live on such a manifold so you are well aware that it is two-dimensional. However, the usual coordinate system based on latitude and longitude breaks down at the poles. We need at least two charts to cover S^2. An easy way to see this is to note that S^2 is compact but an open subset of $\mathbf{R}^2$ is not.

1.1 Examples of manifolds

The most basic example of a manifold is $\mathbf{R}^n$. It requires just one chart, the "patch" is all of $\mathbf{R}^n$ and the the map of the chart is the identity function. This is not surprising, a manifold is locally like a vector space but if the manifold is already a vector space the structure of the charts is trivial; this is called the *standard* structure. But note that we can define other smooth structures on $\mathbf{R}^n$. For example, we can take our manifold to be $\mathbf{R}$ with one patch, as before, but we take the map to be $x \mapsto x^3$. This is a *different* differentiable structure. However, as we shall see in the next section, it is in a strong sense, *equivalent* to the standard structure.

Let us define the n-sphere S^n as a smooth manifold. As a point set

$$S^n := \{(x_1, x_2, \ldots, x_{n+1}) \in \mathbf{R}^{n+1} | \sum_{i=1}^{n+1} x_i^2 = 1\}.$$

As patches we take $U_N = S^n \setminus (0, 0, \ldots, 0, 1)$, this is the sphere with the "North pole" removed and $U_S = S^n \setminus (0, 0, \ldots, 0, -1)$, this is the sphere with the "South pole" removed. We need to map U_N and U_S to $\mathbf{R}^n$[2]. Geometrically imagine a line from the North pole through the point $\mathbf{x} = (x_1, \ldots, x_{n+1})$ in $U_N \subset S^n$ and extended until it hits the equatorial plane, *i.e.* the set of points with $x_{n+1} = 0$. We map $\mathbf{x}$ to the n-tuple of numbers corresponding to the point in the equatorial plane, the explicit formula is

$$(x_1, x_2, \ldots, x_{n+1}) \mapsto (\frac{x_1}{1 - x_{n+1}}, \frac{x_2}{1 - x_{n+1}}, \ldots, \frac{x_n}{1 - x_{n+1}}).$$

There is a similar formula for the projection from the South Pole which gives a homeomorphism from U_S to an open subset of $\mathbf{R}^n$.

Another important example is the hyperbolic space of constant negative curvature. We define the points to be

$$H := \{(x_1, x_2, \ldots, x_{n+1}) \in \mathbf{R}^{n+1} | \sum_{i=1}^{n} x_i^2 - x_{n+1}^2 = -1\}.$$

This manifold only requires one chart, unlike the sphere it is non-compact. We perform projection from the point $(0, 0, \ldots, 0, -1)$ onto the plane $x_{n+1} = 0$ which can be explicitly written as

$$(x_1, x_2, \ldots, x_{n+1}) \mapsto (\frac{x_1}{1 + x_{n+1}}, \frac{x_2}{1 + x_{n+1}}, \ldots, \frac{x_n}{1 + x_{n+1}}).$$

These examples suggest that manifolds live as submanifolds (which we have not formally defined yet) in $\mathbf{R}^n$. This is true and is a deep theorem called the *Nash embedding theorem*. It is absolutely *false* that any n-dimensional manifold embeds in $\mathbf{R}^{n+1}$ but it does embed in some $\mathbf{R}^N$ for some $N > n$. Why then deal with charts and not the embedding space? Thinking of the manifold as we have described it is the *intrinsic view*, whereas thinking in terms of embeddings is the *extrinsic view*. The intrinsic view is much better for studying geometry; there is nothing unique about the embedding in the extrinsic view and it becomes awkward to argue that some fact is true of the manifold independently of how it is embedded. For this reason, all geometry books use the intrinsic view. For specific calculations, for

[2]Not to $\mathbf{R}^{n+1}$.

example with the sphere or the hyperboloid, the use of an embedding may be convenient for programming purposes.

1.2 Smooth functions and mappings

Let us consider functions of various kinds.

Definition 1.2.1 *A function $f : M \longrightarrow \mathbf{R}$ is said to be* ***smooth*** *or C^∞ if for every chart (U_i, ϕ_i) the map $f \circ \phi_i^{-1} : \phi_i(U_i) \longrightarrow \mathbf{R}$ is smooth.*

Note that the map $f \circ \phi_i^{-1} : \phi_i(U_i) \longrightarrow \mathbf{R}$ is between an open set in $\mathbf{R}^n$ and $\mathbf{R}$ so we know what it means to differentiate such a map and hence what it means to be infinitely differentiable (*i.e.* smooth). Note the strategy: we use the charts to move things from M to $\mathbf{R}^n$ where we know what we are talking about. Now suppose that $\mathcal{A}$ and $\mathcal{A}'$ are two maximal atlases on M. Suppose that the set of smooth functions on M with respect to the atlas $\mathcal{A}$ is exactly the same as the set of smooth functions on M with respect to the atlas $\mathcal{A}'$, then the two atlases are the same. If they are not maximal atlases, then they are the "same" in the sense that they are contained in the same maximal atlas. So the collection of smooth functions effectively determines the manifold structure.

So now we know what it means for a real-valued function on a manifold to be smooth. What about functions between manifolds? Let M and N be two manifolds not necessarily of the same dimension, say that M has dimension m and N has dimension n. We say that a mapping $\psi : M \longrightarrow N$ is smooth if for every smooth real-valued function f on N the composite $f \circ \psi$ is a smooth function on M. We used the fact that the smooth functions determine the differentiable structure. We could have given the definition of smooth mapping in terms of charts directly but the above definition is more slick.

Exercise 1.2.2 *Give the definition of smooth mapping in terms of charts.*

The obvious expected property of smooth mappings holds: the composite of smooth mappings is smooth. We can define a notion of *product* of two manifolds as follows.

Definition 1.2.3 *Let M be an m-dimensional manifold and N be an n-dimensional manifold. The* ***product****, written $M \times N$ is an $(m+n)$-dimensional manifold with underlying point set the cartesian product. The charts are obtained by taking products $U_i \times V_j$ where U_i is a chart of M and V_j is a chart of N. The associated mapping is*

$$p \in M, q \in N, \;\; \lambda_{i,j} : U_i \times V_j \longrightarrow \mathbf{R}^{m+n}, \;\; \lambda((p,q)) = (\phi_i(p), \theta_j(q)),$$

where (U_i, ϕ_i) is a chart of M containing p and (V_j, θ_j) is a chart of N containing q.

Exercise 1.2.4 *Verify that the maps $\pi_{1,2} : M \times N \longrightarrow M, N$ respectively, given by $\pi_1((p,q)) = p$ and $\pi_2((p,q)) = q$ are smooth mappings.*

Finally, we consider when two manifolds are "essentially the same."

Definition 1.2.5 *Let M and M' be two manifolds. We say that they are* ***diffeomorphic*** *if there is a smooth map $\psi : M \longrightarrow M'$ which is a bijection and whose inverse $\psi^{-1} : M' \longrightarrow M$ is also smooth. The map ψ is called a* ***diffeomorphism****.*

If two manifolds are diffeomorphic they must have the same dimension; intuitively this is clear, to prove it one uses the fact that if an open subset U of $\mathbf{R}^n$ is homeomorphic to an open subset V of $\mathbf{R}^m$ then $n = m$. Even if the mapping is just a homeomorphism the dimensions must be the same, this requires a technical fact called *invariance of domain.*

The composition of two diffeomorphisms is a diffeomorphism and the inverse of a diffeomorphism is also a diffeomorphism. Thus, being diffeomorphic is clearly an equivalence relation.

We have introduced two notions of two manifolds being the same and this may be confusing. One notion is that they have the same underlying topological space and the same maximal atlas: this implies that the set of smooth real-valued functions are the same. The second notion is that they are diffeomorphic. The latter has the advantage that one can compare manifolds with different underlying sets. Importantly the latter is coarser than the former. To see this via an explicit example, consider $\mathbf{R}$ as a manifold M_1, with the chart $(U := \mathbf{R}, \phi : \mathbf{R} \longrightarrow \mathbf{R})$ given by $\phi(x) = x$ whereas M_2 is another manifold based on $\mathbf{R}$ with the chart $(V := \mathbf{R}, \psi : \mathbf{R} \longrightarrow \mathbf{R})$ given by $\psi(x) = x^3$. I claim that the manifolds M_1 and M_2 are diffeomorphic. Indeed, the map $F : M_1 \longrightarrow M_2 : x \mapsto x^{1/3}$ is a diffeomorphism. Let us check this by following the definition, that is by computing the local coordinate expressions of F and its inverse F^{-1} in the charts (U, ϕ), (V, ψ). We have

$$(\psi \circ F \circ \phi^{-1})(s) = (s^{1/3})^3 = s,$$

which is a smooth bijection. Likewise

$$(\phi \circ F^{-1} \circ \psi^{-1})(t) = (t^{1/3})^3 = t,$$

which is again smooth bijection. Thus the two manifolds are diffeomorphic but they have different maximal atlases.

1.3 Multilinear algebra

Define tensor product, dual space, change of basis, natural iso of double dual.

We assume that the concepts of linear algebra are familiar and, in particular the definition of vector space is well known. We recapitulate it here for convenience.

Definition 1.3.1 *A real vector space is a set V equipped with two operations: $+$ and $\cdot$, the latter often omitted and indicated by juxtaposition. These operations are of type: $+ : V \times V \longrightarrow V$ and $\cdot : \mathbf{R} \times V \longrightarrow V$. They satisfy the following conditions:*

1. *V with the $+$ operation forms an abelian group with identity element denoted $\mathbf{0}$ and the inverse of $v \in V$ written as $-v$.*
2. *If $r, s \in \mathbf{R}$ and $v \in V$ then $r \cdot (s \cdot v) = (rs) \cdot v$ and $(r+s) \cdot v = r \cdot v + s \cdot v$. If $u \in V$ then $r \cdot (u + v) = r \cdot u + r \cdot v$.*
3. *$-1 \cdot v = -v$.*
4. *We have $0 \cdot v = \mathbf{0}$ and $1 \cdot v = v$.*

I am not going to write $\cdot$ explicitly henceforth.

Vector spaces can be defined over any field; it does not have to be the reals. For these notes we will stick to real vector spaces. I will take for granted that you know the following notions: linear independence, basis, dimension. I assume that you know many examples of vector spaces including infinite dimensional ones. I assume that you do **not** think of a vector as a "tuple of numbers"! Tuples of numbers *represent a vector once a basis has been chosen.*

Similarly, I assume that you know what is a linear map, *i.e.* a linear function from V to $\mathbf{R}$, and a linear transformation, which is a linear map from one vector space V to another (possibly the same) vector space U. Once again, do **not** think of linear algebra in terms of matrices; matrices arise when one explicitly describes a linear transformation in terms of chosen bases.

Definition 1.3.2 *Given a vector space V, the **dual space**, written V^* is the space of linear maps: $V \longrightarrow \mathbf{R}$ made into a vector space in its own right by defining $+$ as*

$$\forall \lambda, \sigma \in V^*,\ \lambda + \sigma := v \mapsto \lambda(v) + \sigma(v).$$

A similar definition is used for scalar multiplication.

The following fact is basic.

Proposition 1.3.3 *If V is an n-dimensional vector space, then V^* is also an n-dimensional vector space.*

Proof. Let $\{v_i\}_{i=1}^n$ be a basis for V. Thus any vector $v \in V$ can be written as $v = \sum_{i=1}^n r_i v_i$, where the r_i are real numbers. We will define a basis for V^* in terms of the v_i, as follows. Note that, by linearity, any member of V^* is completely specified by saying what it does to the basis vectors; accordingly, let $\sigma_i(v_j) = 0$ if $i \neq j$ and 1 if $i = j$. Now given any vector v and any linear map σ we can write, by linearity:

$$\sigma(v) = \sum_i r_i \sigma(v_i).$$

I claim that we can write σ as $\sum_i \sigma(v_i)\sigma_i$. To see that note

$$\sum_i \sigma(v_i)\sigma_i(v) = \sum_i \sigma(v_i)\sigma_i(\sum_j r_j v_j) = \sum_{i,j} r_j \sigma(v_i)\sigma_i(v_j) = \sum_i r_i \sigma(v_i).$$

This is exactly the expression above for $\sigma(v)$. So we have shown that any linear map can be expressed as a linear combination of the σ_i's. ∎

Elements of the space V^* are often called *covectors* or dual vectors. The basis that we constructed for V^* is called the *dual basis* to the given one.

Definition 1.3.4 *Given a linear transformation $\lambda : U \longrightarrow V$ between an n-dimensional vector space U and an M-dimensional vector space V, and given bases $\{u_i\}_{i=1}^n$ for U and $\{v_j\}_{j=1}^m$ for V, the* ***matrix*** *representing λ in these chosen bases, is the rectangular array of numbers given by expressing $\lambda(u_i)$ in the basis v_j. Thus, we have*

$$\lambda(u_i) = \sum_j L[i,j] v_j.$$

The action of λ on a generic vector $u = \sum_i r_i u_i$ is given by

$$\lambda(u) = \lambda(\sum_{i=1}^n r_i u_i) = \sum_{i=1}^n \sum_{j=1}^m r_i L[i,j] v_j.$$

The notion of dual applies to transformations as well as spaces.

Definition 1.3.5 *Given a linear transformation* $\lambda : U \longrightarrow V$*, we obtain a linear transformation* $\lambda^* : V^* \longrightarrow U^*$ *given by* $\forall \sigma \in V^*,\ \lambda^*(\sigma) = u \mapsto \sigma(\lambda(u))$ *or* $\lambda^*(\sigma) = \sigma \circ \lambda$.

Notice that this is a basis-independent description. A good exercise it to verify that if you have given bases for U and V then the matrix of λ^* in the dual bases of U^*, V^* is exactly the transpose of the matrix for λ in the bases for U, V. This transposing matrices is not some ad-hoc fiddling with matrices but a very fundamental part of the structure. Defining it as "changing rows and columns" obscures this fact.

Given that a finite-dimensional vector space and its dual space have the same dimension there must be a linear isomorphism between them. In fact, it is easy to construct such isomorphisms. Pick a basis $\{v_i\}$for V and construct the dual basis $\{\sigma_i\}$ for V^*. Now map v_i to σ_i; it is obviously a linear isomorphism. We had to choose a basis to define it however.

Now, since every finite-dimensional vector space is isomorphic to its dual space, we should have V^* isomorphic to V^{**}; the double dual. Then it follows that V is isomorphic to V^{**}. That is true, but a remarkable thing happens: the isomorphism from V to V^{**} can be described in a *basis independent way.* Specifically

$$v \in V \mapsto \Lambda_v \in V^{**} \text{ where } \forall \sigma \in V^*,\ \Lambda_v(\sigma) = \sigma(v).$$

This is called a *natural isomorphism* and is the starting point of a general theory of natural transformations that gave rise to category theory.

Finally, we have the notion of tensor product. This is a way of internalizing the notion of multiargument linear functions into the framework of linear algebra. Given two vector spaces, U of dimension n and V of dimension m, we define the tensor product as follows. It is a new vector space of dimension nm. Let $\{u_i\}_{i=1}^n$ and $\{v_j\}_{j=1}^m$ be bases for U and V respectively. We define $U \otimes V$ as the vector space of all formal finite linear combinations of the form

$$\sum_{i=1}^{n} \sum_{j=1}^{m} r_{i,j} u_i \otimes v_j.$$

The notation $u_i \otimes v_j$ is purely formal. It is easy to see that this is a vector space of the stated dimension.

What is the point of this construction? First, it does not actually depend on the choice of basis; one needs the bases to describe it explicitly, but the actual space one gets in the end is the same whatever bases one chooses. Second, it has the following very important property.

$$\begin{tikzcd} U \times V \arrow[r, "!"] \arrow[dr, "L"'] & U \otimes V \arrow[d, "\lambda"] \\ & W \end{tikzcd}$$

Suppose L is two-argument function from U, V to W that is *bilinear*: this means that it is linear is each argument *separately*. There is a unique function, called ! from $U \times V$ to $U \otimes V$, it takes (u, v) to $u \otimes v$, such that for every bilinear function L, there is a unique linear transformation $\lambda : U \otimes V \longrightarrow W$ with the property that $\lambda(!(u, v)) = L(u, v)$. Thus, the bilinearity can now be treated as ordinary linearity.

It is important to understand that not every element of $V \otimes V$ is of the form $u \otimes v$ for some $u, v \in V$. For example, if v_1, v_2 are two linearly independent vectors in V there is no way that $v_1 \otimes v_1 + v_2 \otimes v_2$ can be written as the tensor of a pair of vectors from V[3]. In quantum mechanics the state space is a vector space and the state spaces of composite systems are constructed as tensor products. A state that cannot be decomposed into a direct tensor product is said to be *entangled*.

1.4 Abstract and concrete indices

One of the complaints about differential geometry is the proliferation of indices on top of symbols. In this section, I will introduce Penrose's *abstract index notation* which was inspired by Einstein's summation convention. It is a handy way to do some calculations in differential geometry. One of the important things that computer scientists realized is the importance of *type information*. Mathematicians are aware of this but they tend to trivialize it and be sloppy with it, whereas physicists make an absolute mess of it! Nevertheless, it was mathematical physicists who introduced this notation which is basically a way of tracking type information.

One can iterate the construction and produce more complicated spaces like $V \otimes V \otimes V^* \otimes V \otimes V^*$. Elements of such spaces are called *tensors*. To keep track of such products of V and V^* we record how many copies of V and how many copies of V^* are present. If there are k copies of V and l copies of V^* we say that an element of the space has type (k, l); this does not say everything, it does not tell us in what order these factors occur. Of course, even this notation gets out of hand if there are many different types of vector spaces occurring. However, in differential geometry it suffices to work with just one vector space, the *tangent* space to be defined in the next section, its dual space, called the *cotangent* space, and tensor products of these spaces. So we will develop the notation to handle just and single V and its dual V^* and repeated tensor products of these such as

[3] If this is not clear, and it is not immediately obvious to me, please prove it.

$V \otimes V \otimes V^* \otimes V \otimes V^*$ or $V \otimes V \otimes V \otimes V^*$. In view of the *natural* isomorphism between V and V^{**}, we will identify them so there will be no need to think about things like V^{*****}. A tensor of type $(1,0)$ is just a vector and a tensor of type $(0,1)$ is a dual vector or covector.

First we introduce what might be called the *concrete index* notation. Let V be our fundamental vector space and let $\{e_i\}_{i=1}^n$ be a basis for it fixed once and for all. A given vector v can be written as $v = \sum_{i=1}^n v^i e_i$ where the v^i are an n-tuple of real numbers: they are the *components of the vector in the given basis*. If we have a covector $\omega \in V^*$ we can use the dual basis, which I will call $\{\sigma^i\}_{i=1}^n$, to write $\omega = \sum_{i=1}^n \omega_i \sigma^i$. Note that for the vector I have written the index for the components as a superscript whereas for the covector I have written the index for the components as a subscript; this is a convention I am going to adopt henceforth.

It will be convenient to introduce the Kronecker symbol δ^i_j; here i, j are indices that run from 1 to n. This is written as

$$\delta^i_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Thus, if $\{e_i\}_{i=1}^n$ is a basis for V and $\{\sigma^i\}_{i=1}^n$ is the dual basis we can write the defining condition for the σ^i as $\sigma^i(e_j) = \delta^i_j$. Note the following obvious but useful formula

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \beta^j \delta^i_j = \sum_i \alpha_i \beta^i.$$

When we apply a covector (linear map) to a vector we get

$$\omega(v) = \sum_i^n \omega_i \sigma^i (\sum_j v^j e_j) = \sum_{i,j=1}^n v^j \omega_i \sigma^i(e_j) = \sum_{i,j=1}^n v^j \omega_i \delta^i_j = \sum_{i=1}^n v^i \omega_i.$$

We will adopt the convention that when there is a pair of repeated indices, *one up and one down*, a summation is implied and we will not write the summation or its limits. Thus our formula becomes

$$\omega(v) = w_i v^i.$$

Now this formula means something in a specific basis only. However, one can free oneself from this basis dependence by using the abstract index notation. Here we attach indices not to any numbers running from 1 to n but just view them as labels telling us to what kind of space a tensor or vector belongs. The up indices indicate vectors and the lower indices indicate a covector.

We can extend this to more complex tensors as follows. Let us consider, for example, an element ξ of $V \otimes V \otimes V^* \otimes V$. In order to keep track of which space ξ comes from we write[4] $\xi^{ab}{}_{c}{}^{d}$. The third index is lower which indicates that the third component of the tensor product is a dual space, all others are products of V's; the complete tensor product can be read off from the index structure.

There is a very important operation called *contraction*. Consider the vector space $V \otimes V^*$. There is a *canonical* map written $\langle , \rangle : V \otimes V^* \longrightarrow \mathbf{R}$ given by

$$\langle v, \omega \rangle = \omega(v).$$

One should read this as the contraction map action on $v \otimes \omega$. Of course, not every element of $V \otimes V^*$ is of the form $v \otimes \omega$ as noted above. But such elements do form a basis for $V \otimes V^*$ and the action on a generic member of $V \otimes V^*$ is given by linearity. We can write the explicit expression but it is a bit messy:

$$\langle , \rangle (\sum_k c_k v^k \omega_k) = \sum c_k \omega_k (v^k).$$

We will always use linear extensions when we can instead of writing explicit formulas for generic elements.

What is the contraction in terms of matrices? Consider a square matrix $M[i,j]$ which is the array of numbers used to represent a linear transformation from V to V. Now here is an absolutely fundamental fact, it is the heart of linear algebra: the spaces $V^* \otimes V$ and $v \longrightarrow V$ where by the latter we mean the linear transformations, are canonically isomorphic. This is so basic that it is overlooked and one only realizes its importance when it is not true. This happens in *infinite-dimensional* spaces. In fancy terminology one says that "the category of finite dimensional vector spaces is compact closed." It is what allows one to construct transposes and change the direction of maps. Coming back to matrices, we will view them as arrays of coefficients describing an element of $V^* \otimes V$, thus if $M \in V^* \otimes V$ and we use our standard notation for bases

$$M = \sum_{i,j=1}^{n} M[i,j] \sigma_i e^j = M^i_j \sigma_i e^j,$$

where we have rewritten things in index notation at the end with the summation implicit. Now contraction applied to σ_i and e^j just gives δ_i^j and hence the sum reduces to $\sum_{i=1}^{n} = M^i_i$. In short it gives the trace! One can readily extend the contraction map to arbitrary tensors. Given a tensor of type (k,l) contraction produces a tensor of type $(k-1, l-1)$ and is

[4] You will need the tensor package to typeset index structures like this.

indicated by repeating a letter in an upper and a lower index as in the following: $T^{abc}{}_{dce}{}^{fg}{}_{hij}$ which contracts on the 3rd and 5th indices. In terms of components one is taking a partial trace.

The point of the abstract index notation is that one is writing true tensor equations that hold in a basis independent way whereas when writes in terms of concrete indices and bases one cannot be sure that the equation only holds in the given basis or not. There are many times, especially in computational settings, where one wants to write down concrete components in a basis and use the concrete index notation. In order to avoid confusion we will adopt the convention from Wald [1984] of using Greek letters for concrete indices and Latin letters for abstract indices. Wald's book gives the clearest exposition of the abstract index notation.

2

TANGENT VECTORS AND TENSORS

We have the apparatus in place, we can finally do some elementary differential calculus. In a manifold we have a structure that "locally looks like a vector space" so we have to try and do everything infinitesimally. Vectors will emerge as "tangent vectors" to curves. Of course, this is exactly where we need to define derivatives. But first we need to define curves.

Definition 2.0.1 *A curve in a manifold M is a map $\gamma : [0,1] \longrightarrow M$ or $\gamma : \mathbf{R} \longrightarrow M$. A curve γ is* ***smooth*** *at a point $p = \gamma(t_0)$ of M, if for any chart (U, ϕ) with $p \in U$, we have $\phi \circ \gamma$ is smooth at t_0. If γ is smooth everywhere we say γ is a smooth curve.*

Note that the curve is not just the image of γ; the parameter is an essential part of the curve. So the curve γ' given by $\gamma'(t) = \gamma(2t)$ defines the same point set but is a different curve. One should think of the parameter as time, then a curve is actually a motion through the manifold. Here γ' follows the same path as γ but twice as fast, so it is a different curve.

One can think of a manifold as a "surface" embedded in some $\mathbf{R}^m$ and then one can imagine tangent vectors to this surface. But this extrinsic view, while useful for some calculational purposes, suffers from the drawback that there is absolutely nothing unique about how a manifold may be embedded and the intrinsic geometric properties are obscured in this view. So we will **not** be doing this. We will think of tangent vectors as living in a vector space attached to each point of M but they will be constructed as tangent vectors to curves in M: a tangent vector will acquire meaning through directional derivatives.

In $\mathbf{R}^n$ the correspondence between vectors and directional derivatives is clear. Given a vector v in $\mathbf{R}^n$ we can think of it as an n-tuple $(v^1, v^2, \ldots, v^n)$ in the canonical basis of $reals^n$. If we let $x^\mu = (x^1, \ldots, x^n)$ stand for the basis vectors of $\mathbf{R}^n$ we have the corresponding directional derivative operator $\sum_\mu v^\mu (\frac{\partial\,(\cdot)}{\partial x^\mu})$. We will treat the derivative as basic and axiomatize it through its basic properties. Let $\mathcal{F}$ denote the collection of smooth functions on M.

Definition 2.0.2 *Given a point p of M we define a* ***tangent vector at*** *p to be a map $v : \mathcal{F} \longrightarrow \mathbf{R}$ such that, $\forall a, b \in \mathbf{R}$ and $\forall f, g \in \mathcal{F}$:*

1. $v(af + bg) = av(f) + bv(g)$,
2. $v(fg) = f(p)v(g) + g(p)f(v)$.

It follows immediately that if f is a constant function then $v(f) = 0$. Note how the second condition makes specific reference to the point p.

This just defines a set. In order to make it a vector space we need to have addition and scalar multiplication: but these are readily defined pointwise. We define these as follows:

$$(v_1 + v_2)(f) = v_1(f) + v_2(f); \qquad (av)(f) = av(f).$$

It is clear that we now have a vector space; all the required properties follow from the fact that $\mathbf{R}$ are a vector space. However, it may not look like the notion of directional derivatives that I promised you. That will emerge in the proof of the next theorem.

Theorem 2.0.3 *Let M be an n-dimensional manifold. Let $p \in M$ and let T_p denote the tangent space at p. Then $\dim(T_p) = n$.*

I will not give the proof in detail but I will construct a basis of n vectors. The full proof requires some basic notions of multivariable calculus; see Wald [1984] for details.

Proof. Let (U, ϕ) be a chart of M with $p \in U$ and $\phi : U \longrightarrow O \subset \mathbf{R}^n$ is a homeomorphism between U and O. Given $f \in \mathcal{F}$, we have $f \circ \phi^{-1} : O \longrightarrow \mathbf{R}$ is a smooth map by definition of $\mathcal{F}$. Let x^μ, $\mu = 1, \ldots, n$ be the coordinate functions of $\mathbf{R}^n$. We define, for each μ, the map $X_\mu : \mathcal{F} \longrightarrow \mathbf{R}$ by

$$X_\mu = \left.\frac{\partial\, f \circ \phi^{-1}}{\partial x^\mu}\right|_{\phi(p)}.$$

These define vectors at p; the required properties follow from the corresponding properties of derivatives. It is easy to construct functions that are constant along all the axes except one: this shows that these vectors are in fact linearly independent.

Any tangent vector v at p can be expressed as a linear combination of the X_μ; this is the part that requires some facts from multivariable calculus:

$$v(f) = \sum_{\mu=1}^{n} v^\mu X_\mu(f); \text{ where } v^\mu = v(x^\mu \circ \phi).$$

∎

This establishes the link between tangent vectors defined abstractly and directional derivatives. Note, all this depended on choosing charts so we were writing it all using *concrete indices*.

Given the tangent space T_p at p we consider the dual space T_p^*; also called the *covector* space at p. This is also n-dimensional. We typically use Greek letters for elements of the covector space. Covectors are sometimes called 1-forms.

The tensor algebra that will be relevant for us will consist of tensor products of T_p and T_p^* possibly repeated. Products of 4 spaces are common, for example the Riemann curvature is described by an element of $T_p \otimes T_p \otimes T_P \otimes T_P^*$.

We worked with a fixed choice of basis in $\mathbf{R}^n$. What if we had used another basis $\{x'^\mu\}_{\mu=1}^n$ for $\mathbf{R}^n$? We would get another basis for T_p which we could call X'_μ. The question then is how to relate X_μ and X'_μ. Since X_μ is defined in terms of $\frac{\partial}{\partial x^\mu}$ and X'^μ is defined in terms of $\frac{\partial}{\partial x'^\mu}$ we can just use the chain rule to deduce that

$$X_\mu = \sum_{\nu=1}^{n} \frac{\partial x'^\nu}{\partial x^\mu}\bigg|_{\phi(p)} X'_\nu.$$

What about the dual space? If the dual basis is denoted by $\{\omega^\mu\}_{\mu=1}^n$, *i.e.* $\omega^\mu(X_\nu) = \delta^\mu_\nu$ then the transformation law is inverted

$$\omega^\mu = \sum_{\nu=1}^{n} \frac{\partial x^\mu}{\partial x'^\nu}\bigg|_{\phi(p)} \omega'^\nu.$$

Tangent to a curve at a point. Let $\gamma : [-1,1] \longrightarrow M$ be a smooth curve in M and let $\gamma(0) = p \in M$. Thus, γ is a smooth curve passing through p. Let $f : M \longrightarrow \mathbf{R}$ be a smooth function on M. We define a tangent vector to γ at p as follows. We have a function $f \circ \gamma : [-1,1] \longrightarrow \mathbf{R}$. For such a function we can readily define the derivative at 0, here t is the parameter taking values in $[-1,1]$:

$$\frac{\mathrm{d}(f \circ \gamma)}{\mathrm{d}t}\bigg|_{t=0}.$$

This can be readily verified to be a mapping from smooth functions on M to $\mathbf{R}$ and to satisfy the requirements of a tangent vector at p. This is thus an element of T_p and is called the tangent vector to the curve γ at p.

One can think of vectors in two ways: as directional derivatives as we did, or as tangent vectors: these two ways are equivalent, see Lee [2013] for a careful exposition of this equivalence.

2.1 Vector and tensor fields

So far we have considered vectors, covectors and tensors at a point. But, of course, the more important structure is that of a *vector field*: a family of vectors, one for each point of the manifold. The crucial question is how does the vector vary as one moves around the manifold? Recall that a vector at p maps a smooth function to a number. A vector field, therefore, will act on a function at every point to produce a number in every point: this is just a long-winded way of saying that a vector field acts on a smooth function to produce another function.

Definition 2.1.1 *A* ***smooth vector field*** *on a manifold*[5] *M is an assignment of a vector v_p to every point p of M in such a way that the function f is mapped to the* smooth *function $p \mapsto v_p(f)$.*

From here to tensor fields one can continue in the same way. Thus, a smooth covector field ω maps smooth vector fields $\mathbf{v}$ to smooth functions $p \mapsto \omega_p(\mathbf{v}_p)$, where ω_p is the value of ω at p and $\mathbf{v}_p$ is the value of $\mathbf{v}$ at p. A smooth tensor field of type (k, l) maps k-tuples of covector fields and l-tuples of vector fields to a smooth function in the evident way.

The assembly of all the tangent vector spaces is a clumsy structure so it is worth organizing it into something more manageable and conceptually clearer. Given an n-manifold M, the collection of all the tangent vector spaces can be organized into a structure called the *tangent bundle* of M and is written TM. As a set TM is just the disjoint union of the vector spaces T_p: $TM = \coprod_{p \in M} T_p$. For those of you who have studied advanced type theory in programming languages, this is an excellent example of a dependent-product type. The elements of TM are usually written (p, v), where $p \in M$ and $v \in T_p$. There is a canonical projection map $\pi : TM \to M$ with $\pi((p, v)) = p$. But the tangent bundle is much more than a set. The following theorem from Lee [2013] pp. 66-67, spells this out carefully.

Proposition 2.1.2 *For any smooth n-manifold M, the tangent bundle has a natural topology and differentiable structure making it a smooth $2n$-dimensional manifold. With respect to this differentiable structure the projection map π is smooth.*

I will not give the proof here but I do recommend reading this from the book just cited.

[5] I am not going to bother to say “smooth manifold” any longer.

This is an example of what is called a vector bundle. Locally a vector bundle looks like a product but globally it may not be a product at all. One can similarly organize the covector spaces into a vector bundle and indeed all the tensors of a fixed type can be made into vector bundles.

A smooth vector field can now be regarded as a smooth map v from M to TM such that $\pi \circ v = \mathrm{Id}_M$. Such maps are called *sections* of the bundle.

Integral curves and commutators.

Definition 2.1.3 *A **one-parameter group of diffeomorphisms** ϕ_t of an n-manifold M is a smooth map $\mathbf{R} \times M \longrightarrow M$ such that for each fixed $t \in \mathbf{R}$ we have that $\phi_t : M \longrightarrow M$ is a diffeomorphism and for all $t, s \in \mathbf{R}$ we have $\phi_t \circ \phi_s = \phi_{t+s}$.*

We can associate to ϕ_t a vector field as follows. For a fixed $p \in M$, $\gamma(t) = \phi_t(p) : \mathbf{R} \longrightarrow M$ is a smooth curve passing through p when $t = 0$. Define v_p to be the tangent to this curve at p, *i.e.* at $t = 0$. This defines a smooth vector field on M. In summary: the one-parameter group of diffeomorphisms defines a family of non-intersecting curves that fill all of M. The tangents to these curves defines the vector field. Conversely, given a smooth vector field $\mathbf{v}$ we can define a family of curves such that all the curves are smooth, through every point there is one *and only one* curve, so the curves never intersect. These curves are called the *integral* curves of the vector field. Constructing these curves amounts to solving a system of first-order ordinary differential equations. In fact, one should think of first-order ODE's as vector fields!

Given two vector fields $\mathbf{u}$ and $\mathbf{v}$ we have an important operation called the **commutator** which constructs a new vector field from $\mathbf{u}$ and $\mathbf{v}$. Recall that a vector field is a map $\mathcal{F}(M) \longrightarrow \mathcal{F}(M)$ which is linear and satisfies the Leibnitz rule. Can we try to define a product? Let us try to define $\mathbf{uv}$ by $\mathbf{uv}(f) = \mathbf{u}(\mathbf{v}(f))$, *i.e.* by $\mathbf{uv} = \mathbf{u} \circ \mathbf{v}$; this is clearly not a commutative product. It is definitely linear but does it satisfy Leibnitz?

$$\mathbf{uv}(fg) = \mathbf{u}(\mathbf{v}(fg)) = \mathbf{u}(g\mathbf{v}(f)+f\mathbf{v}(g)) = \mathbf{u}(g)\mathbf{v}(f)+g\mathbf{uv}(f)+\mathbf{u}(f)\mathbf{v}(g)+f\mathbf{uv}(g).$$

This does not look like Leibnitz' rule. But if we define $[\mathbf{u}, \mathbf{v}] = (\mathbf{uv} - \mathbf{vu})$ then an explicit calculation[6] shows that

$$[\mathbf{u}, \mathbf{v}](fg) = g[\mathbf{u}, \mathbf{v}](f) + f[\mathbf{u}, \mathbf{v}](g).$$

This operation is called the commutator of $\mathbf{u}$ and $\mathbf{v}$.

[6] Do it, if you don't believe me!

The commutator, also called the **Lie bracket** is anti-commutative, $[\mathbf{u}, \mathbf{v}] = -[\mathbf{v}, \mathbf{u}]$, rather than commutative. It is also nonassociative but obeys the following rule

$$[\mathbf{u}, [\mathbf{v}, \mathbf{w}]] + [\mathbf{w}, [\mathbf{u}, \mathbf{v}]] + [\mathbf{v}, [\mathbf{w}, \mathbf{u}]] = 0.$$

This is called the *Jacobi identity*.

There is a beautiful geometric interpretation of the commutator shown in the picture below.

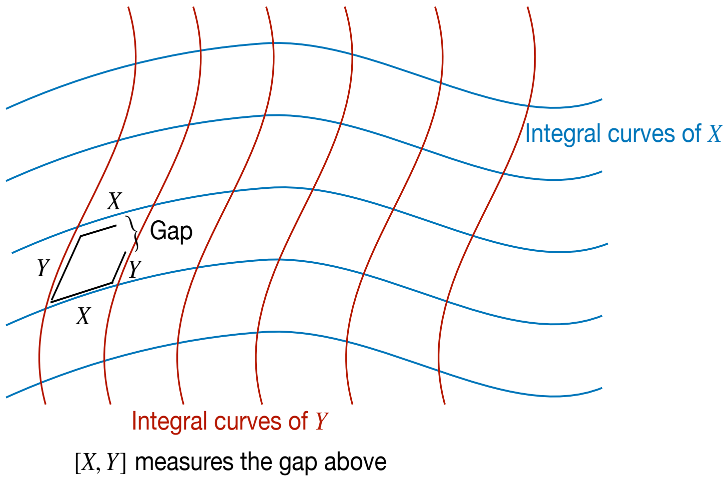

Figure 2.1: The meaning of the commutator of two vector fields X, Y

2.2 The differential of a map

Let M, N be smooth manifolds and let $\psi : M \longrightarrow N$ be a smooth map. Let p be a point of of M. We want to understand how vectors map under the action of ψ.

The **differential** of the map ψ at p is the linear map $\mathrm{d}\psi : T_pM \longrightarrow T_{\psi(p)}N$ defined as follows. Let g be a smooth function on a neighbourhood of $\psi(p)$ and let v be a tangent vector at p. We have to define a tangent vector at $\psi(p)$ so it should be able to act on g. We define

$$\mathrm{d}\psi(v)(g) := v(g \circ \psi).$$

The action of ψ pushes covectors in the opposite direction. Let ω be a covector at $\psi(p)$, hence a linear map on $T_{\psi(p)}N$. We define $\delta\psi : T^*_{\psi(p)}N$

$\to T_p^* M$ by

$$\forall v \in T_p M,\ \delta\psi(\omega)(v) := \omega(\mathrm{d}\psi(v)).$$

2.3 Derivative operators and connections

Vectors at different points cannot be compared without additional structure being defined. The spaces T_p and T_q are different, though both are isomorphic to $\mathbf{R}^n$, when p and q are distinct points of an n-manifold M. What one needs is a way of “moving vectors from one place to another”. But we need to move them in a way that keeps them “essentially the same”. This is called *parallel transport* and the mathematical structure is called an *affine connection.*

Since the issue is working out how vectors change as we move around it is natural to define them in terms of derivatives. Accordingly we begin with an axiomatic treatment of derivatives on manifolds. Recall that we write $\mathcal{T}(k,l)$ for the space of (k,l) tensor fields.

Definition 2.3.1 *A* ***derivative operator****, written ∇ and also called a* ***covariant derivative****, on an n-manifold M is a map that takes a smooth tensor* field *of type (k,l) to a smooth tensor field of type $(k,l+1)$ and satisfies the* 5 *conditions enumerated below. If $T^{a_1\ldots a_k}{}_{b_1\ldots b_l}$ is a type (k,l) tensor we will write $\nabla_c T^{a_1\ldots a_k}{}_{b_1\ldots b_l}$ for the action of ∇ on T. In stating the conditions we will suppress most of the indices.*

1. *$\forall\, T, S \in \mathcal{T}(k,l)$ and $\alpha, \beta \in \mathbf{R}$ we have:*

$$\nabla_a(\alpha T^{\cdots}_{\cdots} + \beta S^{\cdots}_{\cdots}) = \alpha \nabla_a T^{\cdots}_{\cdots} + \beta \nabla_a S^{\cdots}_{\cdots}$$

2. *$\forall\, T \in \mathcal{T}(k,l)$ and $\forall\, S \in \mathcal{T}(k',l')$ we have*

$$\nabla_a[T^{\cdots}_{\cdots} S^{\cdots}_{\cdots}] = (\nabla_a T^{\cdots}_{\cdots}) S^{\cdots}_{\cdots} + T^{\cdots}_{\cdots}(\nabla_a S^{\cdots}_{\cdots})$$

3.
$$\nabla_a[T^{a_1\ldots c\ldots a_k}_{b_1\ldots c\ldots b_l}] = \nabla_a T^{a_1\ldots c\ldots a_k}_{b_1\ldots c\ldots b_l}$$

   This appears trivial but it says that the action of ∇ commutes with contraction.

4. *If $f \in \mathcal{F}(M)$ and $t^a \in T_p$ for some $p \in M$ then $t(f) = t^a \nabla_a f$.*
5. *$\forall\, f \in \mathcal{F}(M)$ we have $\nabla_a \nabla_b f = \nabla_b \nabla_a f$.*

An explicit calculation shows that

$$[\mathbf{u}, \mathbf{v}]^b = \mathbf{u}^a \nabla_a \mathbf{v}^b - \mathbf{v}^a \nabla_a \mathbf{u}^a.$$

How do we know that there are *any* derivative operators? We can always construct derivative operators using local charts (aka coordinate systems on "patches"). Let ψ be a coordinate system on a local patch U of M. Let $\{\frac{\partial\cdot}{\partial x^\mu}\}$ and $\{\mathrm{d}x^\mu\}$ be coordinate bases for T_p and T^*_P where $p \in U$. For any smooth tensor field $T^{a_1\ldots a_k}_{b_1\ldots b_l}$ the components in these bases are $T^{\mu_1\ldots\mu_k}_{\nu_1\ldots\nu_l}$. We define $\partial_c T^{a_1\ldots a_k}_{b_1\ldots b_l}$ to be the tensor whose components are

$$\frac{\partial(T^{\mu_1\ldots\mu_k}_{\nu_1\ldots\nu_l})}{\partial x^\sigma}.$$

The fact that this satisfies the requirements of a derivative operator follows from standard properties of partial derivatives. These are called coordinate derivatives and, of course, depend on the choice of coordinate system. There are many other derivatives that can be defined.

Suppose ∇ and ∇' are two different derivative operators, how could they differ? They must agree on scalar fields because of (4). Let ω_a be a covector field, then using the Leibnitz rule (2) we compute:

$$\nabla'_a(f\omega_b) - \nabla_a(f\omega_b) = (\nabla'_a f)\omega_b + f\nabla'_a\omega_b - (\nabla_a f)\omega_b - f(\nabla_a\omega_b)$$

but the first and third terms cancel since ∇ and ∇' agree on f so we finally get $f(\nabla'_a\omega_b - \nabla_a\omega_b)$. This equation can be used to show that the quantity $(\nabla'_a\omega_b - \nabla_a\omega_b)$ only depends on the value of ω_b at p and not on how it varies near p; this is not completely obvious, see page 32 of Wald [1984] for the argument. Thus $\nabla' - \nabla$ defines a linear map from type $(0,1)$ tensors to $(0,2)$ tensors; in short it must be a type $(1,2)$ tensor. Thus the difference between two derivative operators is captured by a tensor field C^c_{ab}. Thus we have

$$\nabla'_a\omega_b = \nabla_a\omega_b + C^c_{ab}\omega_c.$$

From (5) it follows that C is symmetric in its lower indices: $C^c_{ab} = C^c_{ba}$. Using the Leibnitz rule one can show how derivative operators differ on vector fields

$$\nabla'_a t^b = \nabla_a t^b - C^b_{ac}t^c.$$

In a similar way one can compute the action of ∇' on higher tensor fields in terms of ∇ and C. In the special case where we have the coordinate derivative we get

$$\nabla_a t^b = \partial_a t^b + \Gamma^b_{ac}t^c$$

where it is conventional to write Γ rather than C in this situation; Γ is called the *Christoffel symbol*.

What is the geometric meaning of a derivative operator? Vectors at different points cannot be compared just based on the manifold structure. Given a curve γ and a vector v at a point p of the curve, what does it mean to "move v parallel to itself along the curve?" The idea is that if one takes the derivative of v along the curve one should get 0.

Definition 2.3.2 *We say a vector v^a at p, is parallel transported along a curve γ through p if $t^a \nabla_a v^b = 0$, where t^a is the tangent vector to the curve.*

Of course this depends on a choice of derivative operator. *Thus choosing a derivative operator amounts to choosing a way of parallel transporting vectors.* This structure is called an *affine connection.* It is up to us to choose an affine connection. In the next section we will see that there is a special affine connection which is to be preferred given a metric.

3

METRICS

First let me straighten out a point of possible terminological confusion: the word "metric" is not used in the sense of metric spaces. What we are going to define is a way of measuring the *length of vectors* and also the *angle between vectors*. So it is *not* a function from $M \times M$ to $\mathbf{R}$. It is called the Riemannian metric but that is a mouthful to say (and write in LaTeX) so we will say "metric" here meaning Riemannian metrics.

Definition 3.0.1 *A Riemannian metric on a manifold M is a choice of scalar product at each tangent space T_p of M in such a way that it varies smoothly as p varies.*

To see what this really means we examine what it means in a chart. In the coordinate basis a vector is represented by an n-tuple of numbers. A scalar product, therefore, is an $n \times n$ symmetric matrix. Freeing ourselves from the shackles of bases and coordinates, we see that a Riemannian metric is simply a smooth tensor field of type $(0, 2)$ which is symmetric and non-degenerate. We usually denote a metric by g_{ab}. The "inner product" of two vectors is $g_{ab}u^a v^b$ and $\sqrt{g_{ab}u^a u^b}$ defines the *norm* or length of the vector u.

If u^a, v^a are parallel transported their angle should not change so if t^a is the tangent to the curve along which u and v are being parallel transported we have

$$t^a \nabla_a (g_{bc} u^b v^c) = 0.$$

Now, since u and v are being parallel transported we have $t^a \nabla_a u^b = 0$ ns $t^a \nabla_a v^b = 0$. Using Leibnitz and killing these two terms we are left with $u^b v^c t^a \nabla_a g_{bc} = 0$. If this holds for *any* choice of vectors and curves we must have $\nabla_a g_{bc} = 0$. So the derivative operator has a special relationship to the metric if the notion of parallel transport is to be coherent. This is captured in the next theorem.

Theorem 3.0.2 *Let g_{ab} be a Riemannian metric. There exists a unique derivative operator ∇_a satisfying $\nabla_a g_{bc} = 0$.*

If the metric is fixed then the unique compatible derivative operator is called the *covariant derivative*. There is an explicit formula for the covariant derivative in terms of the coordinate derivative. Recall that two derivatives are connected by the Christoffel symbol so the explicit formula gives the Christoffel symbols as follows:

$$\Gamma^c_{ab} = \frac{1}{2} g^{cd}[\partial_a g_{bd} + \partial_b g_{ad} - \partial_d g_{ab}],$$

where g^{ab} is the inverse $g^{ac}g_{cb} = \delta^a_b$.

How do we know we can define a Riemannian metric on a manifold? Well we can certainly define a Riemannian metric on $\mathbf{R}^n$ in the usual way $g_{ab}u^a v^b = \sum_{\mu=1}^n u^\mu v^\mu$ where we have written concrete indices to indicate that we are using the standard chart and written the vectors in terms of their components. The explicit matrix representing this metric is simply the identity matrix.

Theorem 3.0.3 *Every differentiable manifold can be equipped with a Riemannian metric.*

Proof. Let $\{(U_\alpha, \phi_\alpha) \mid \alpha \in \mathcal{A}\}$ be an atlas. Let ψ_α be a partition of unity subordinate to the covering defined by the atlas. Given two vector fields u and v let $(u^1_\alpha, \ldots, u^n_\alpha)$ and $(v^1_\alpha, \ldots, v^n_\alpha)$ be the components of u and v in the bases induced by the coordinates in each chart in the tangent space T_p. Then we define the following metric

$$(g_{ab}u^a v^b)(p) = \sum_{\alpha \in \mathcal{A}} \psi_\alpha(p)[\sum_{\mu=1}^n u^\mu_\alpha v^\mu_\alpha.$$

∎

One has to verify smoothness but it is pretty clear from the construction. What we have done is used a partition of unity to glue together than basic metric of $\mathbf{R}^n$ on each patch. This is the fundamental reason why we assumed paracompactness for manifolds, so that we can construct partitions of unity at will and hence Riemannian metrics.

Bibliography

W. M. Boothby. *An introduction to differential geometry and Riemannian geometry.* Academic Press, second revised edition edition, 2003.

R. Geroch. *Differential Geometry.* Minkowski Institute Press, 2013. Available in 1972 as unpublished notes from the University of Chicago.

S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry, Volume 1.* John Wiley and Sons, wiley classics edition published 1991 edition, 1963.

J. M. Lee. *Riemannian Manifolds: an introduction to curvature.* Number 176 in Graduate Texts in Mathematics. Springer-Verlag, 1997.

J. M. Lee. *Introduction to Smooth Manifolds.* Number 218 in Graduate Texts in Mathematics. Springer-Verlag, 2nd edition, 2013.

M. Spivak. *Calculus on manifolds.* WA Benjamin Inc., 1965.

M. Spivak. *A Comprehensive Introduction to Differential Geometry vols 1-5.* Publish or Perish Inc., 1979.

R. Wald. *General relativity.* The University of Chicago Press, 1984.