Geometry of geodesics

Joonas Ilmavirta
joonas.ilmavirta@jyu.fi

January 12, 2026

These are lecture notes for the course “MATS4120 Geometry of geodesics” first given at the University of Jyväskylä in Spring 2020. The course describes geodesics and their geometry on Riemannian manifolds. Basic differential geometry or Riemannian geometry is useful background but is not strictly necessary. Exercise problems are included, and at least problems marked important should be solved as you read to ensure that you are able to follow.

Previous feedback has been very useful and new feedback is welcome.

Contents

1 Riemannian manifolds

1.1 A look on geometry

A central concept in Euclidean geometry is the Euclidean inner product, although its importance is somewhat hidden in elementary treatises. We will relax its rigidity to allow for a certain kind of variable inner product. This provides a rich geometrical framework — Riemannian geometry — and shines new light on the nature of Euclidean geometry as well.

There is much to be studied beyond Riemannian geometry, but we will not go there. Neither will we study all of Riemannian geometry; we shall focus on the geometry of geodesics. Gaps will be left, especially early on, and may be filled in by more general courses or textbooks on Riemannian geometry.

Yet another thing we will not be concerned with is regularity. There are interesting phenomena in various spaces of low regularity, but even those are best understood if one has background knowledge of the simplest possible situation. All the structures in this course will be smooth, by which we mean C^∞. Many — but not all — of the resulting functions will be smooth as well, and we will take some care to show how smoothness of structure implies smoothness of derived structure.

We will do local Riemannian geometry in the sense that we will implicitly be working in a single coordinate patch. Even when a more global treatment would be needed using a partition of unity or some such tool, we will pretend that everything is still in a single patch. This promotes the structures essential for this course. A reader with more prior familiarity with manifolds is invited to globalize the proofs presented here in a more honest fashion.

Differential geometry can often be done in a local coordinate formalism or using invariant concepts. We prefer an invariant approach, but the coordinate description will always be given as well so as to give more concrete and calculable definitions.

Some readers may find these notes vague or lacking in detail, but that is entirely purposeful. The goal is to focus on a certain set of phenomena and not to be held back by technicalities. One does not need to manually craft every atom to obtain a coherent big picture, and one might even argue that orientation to details can harm by causing the focus to drift away from the ideas that are important for the present goal.

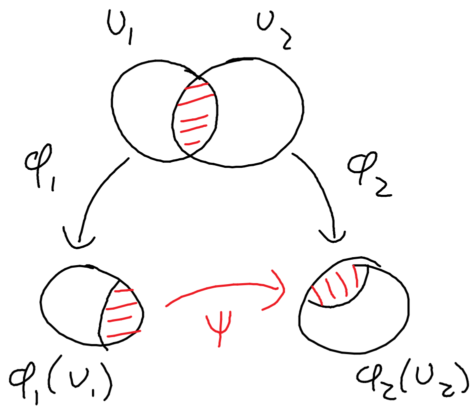


Figure 1: Two charts (U_i, φ_i) with $i = 1, 2$. The transition function ψ maps between the Euclidean domains $\varphi_i(U_1 \cap U_2) \subset \mathbb{R}^n$ corresponding to the intersection set on the manifold M. The diagram commutes. Smooth compatibility is a property of the map ψ.

1.2 Smooth manifolds

Let $n \in \mathbb{N}$. A topological n-dimensional manifold M is a topological space which is second-countable[1], Hausdorff[2] and "looks locally like $\mathbb{R}^n$". The last bit in quotes means that any point $x \in M$ has a neighborhood $U \subset M$ for which there exists a homeomorphism $\varphi: U \to \varphi(U) \subset \mathbb{R}^n$. Such a local homeomorphism is known as a coordinate chart as it gives Euclidean coordinates in an open subset of the manifold.

The conditions above define a topological manifold. To make it smooth, we introduce more structure. As M itself is just an abstract space, there is no way to differentiate on it. All derivatives will have to be considered in local Euclidean coordinates given by a chart, but on a single chart there is nothing to differentiate.

Consider two charts $\varphi_i: U_i \to \varphi_i(U_i)$ with $i = 1, 2$. If the domains U_1 and U_2 intersect, we get a map between the two local coordinate systems. Specifically, if $U := U_1 \cap U_2$, the map $\psi: \varphi_1(U) \to \varphi_2(U)$ defined by $\psi \circ \varphi_1 = \varphi_2$ is a map between two open sets in $\mathbb{R}^n$. This map is called the transition function between the two coordinate charts and it is depicted in figure 1.

Exercise 1.1. Show that the transition function ψ is a homeomorphism. ◯

We say that the two coordinate charts φ_i are smoothly compatible if

[1] A first-countable space has a countable neighborhood base at each point, whereas a second-countable space has a countable base for the whole topology.

[2] The Hausdorff condition is also known as the separation axiom T2. It means that any two distinct points have disjoint neighborhoods.

the map ψ is a diffeomorphism. To either satisfy or irritate the reader, we observe that if the two open sets U_i do not meet, then ψ is the unique map from the empty subset of $\mathbb{R}^n$ to itself and is vacuously smooth; this ensures that checking for compatibility only makes a difference if the two domains meet.

Exercise 1.2. Is smooth compatibility an equivalence relation in the set of coordinate charts on a manifold M? ◯

An atlas is a collection of coordinate charts $(U_\alpha, \varphi_\alpha)_{\alpha\in A}$ so that they cover the whole manifold: $\bigcup_{\alpha\in A} U_\alpha = M$. An atlas is smooth if all pairs of coordinate charts are smoothly compatible. A smooth atlas is maximal if no new coordinate chart can be added to it without breaking smoothness. A maximal smooth atlas is sometimes called a smooth structure.

Exercise 1.3. Show that every atlas is contained in a unique maximal atlas. ◯

Definition 1.1 (Smooth manifold)**.** A smooth n-dimensional manifold is a topological n-manifold with a maximal smooth atlas.

All regularity matters are always defined in terms of the local coordinates given by a fixed atlas. A function $f\colon M \to \mathbb{R}$ on a smooth manifold is defined to be smooth when $f \circ \varphi^{-1}$ is a smooth Euclidean function for any local coordinate map φ.

⋆ *Important exercise* 1.4. Define what it should mean for a function $f\colon M \to N$ between two smooth manifolds of any dimension to be smooth. ◯

The Euclidean space $\mathbb{R}^n$ is an n-dimensional smooth manifold. An atlas is given by any open cover (e.g. the singleton of the space itself) and identity maps.

Remark 1.2. Once we have fixed a smooth structure, a valid coordinate chart is precisely a smooth diffeomorphism $\varphi\colon U \to \varphi(U) \subset \mathbb{R}^n$ from an open set $U \subset M$. This cannot be taken as a starting point, since before the smooth structure and its charts we do not know what smoothness of such a map would mean. This only becomes useful later when deciding whether a given map gives valid coordinates.

1.3 Curves, vectors and differentials

A smooth curve is a smooth map from an interval $I \subset \mathbb{R}$ to our smooth manifold M. The velocity $\dot\gamma(t)$ of a curve $\gamma\colon I \to M$ at any given time $t \in I$ is a tangent vector in the tangent space $T_{\gamma(t)}M$ as depicted in figure 2.

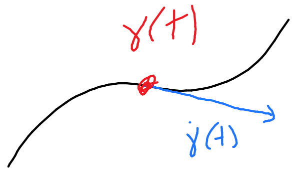


Figure 2: Velocity of a curve γ at time t is a tangent vector $\dot{\gamma}(t)$. It is an element of the tangent space at the point $\gamma(t)$. The picture may be understood either abstractly on a manifold (with the velocity being in the corresponding tangent space) or concretely within a single coordinate chart (with the velocity simply as the time derivative of the Euclidean curve).

Indeed, the tangent space can be defined using velocities of curves[3], but it is not the only possible approach. Different points of view are useful, and we will be free to change perspectives as convenient. It is unimportant for us which approach one chooses to define tangent spaces.

In terms of local coordinates the tangent space T_xM at $x \in M$ can be understood[4] to be just $\mathbb{R}^n$. A typical approach is to define a tangent vector as a derivation, a certain kind of a differential operator. This is related to the curve-based definition as follows: A tangent vector $W \in T_xM$ can be thought of as a differential operator or as the velocity of a curve γ at $t = 0$. A smooth function $f \colon M \to \mathbb{R}$ is differentiated by $Wf = \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))|_{t=0}$.

The same object can function as the velocity of a curve or as a derivation. It would be possible to give different incarnations of tangent vectors different names and introduce canonical isomorphisms between them, but we will leave any such identifications out.

An important feature of a tangent space is that it is a vector space. For any x on an n-dimensional smooth manifold M, the tangent space T_xM is an n-dimensional real vector space. It is therefore isomorphic to $\mathbb{R}^n$, but not in a canonical way. Any local coordinates give a natural way to identify $T_xM \cong \mathbb{R}^n$, but the many possible coordinate charts in neighborhoods of x give different isomorphisms[5].

[3]One says that two curves γ_i are equivalent if in a fixed local coordinate system the Euclidean curves $\varphi \circ \gamma_i$ have the same velocity at the reference point. Then a tangent vector is an equivalence class of curves. To get a coordinate invariant definition, one needs to show that the behavior upon changing coordinates is correct.

[4]It is hopefully evident that any local coordinate chart gives an identification of the tangent space T_xM at x with $\mathbb{R}^n$ with the curve approach of the preceding paragraph.

[5]Indeed, all isomorphisms between the two vector spaces can be realized through a coordinate chart of a maximal atlas.

The dual vector space T_xM is called the cotangent space and denoted by T_x^*M. One could also define T_x^*M first and then define T_xM by duality. The most important example of a covector is the differential of a function $f\colon M \to \mathbb{R}$. The differential at $x \in M$ is $\mathrm{d}f_x \in T_x^*M$ and the duality pairing is defined by

$$\mathrm{d}f_x(W) = Wf \tag{1}$$

for any $W \in T_xM$, considered as a derivation. Be careful to call this the differential, not the gradient, of a function.

We shall study vectors and covectors in more detail later, but the very basics are best learned from introductory material to differential geometry.

1.4 Algebraic constructions on the tangent bundle

All of the tangent spaces of a manifold together make up the tangent bundle. That is, one can define the tangent bundle of our smooth manifold M to be the disjoint union

$$TM = \coprod_{x \in M} T_xM. \tag{2}$$

This is a union of vector spaces, and many operations are done tangent space by tangent space.[6]

In general, a bundle is a disjoint union of spaces of some kind attached to each point. (The tangent bundle is a union of tangent spaces.) These spaces, called the fibers of the bundle, are isomorphic to each other but not necessarily in a canonical way. (Since $T_xM \cong \mathbb{R}^n$ for all $x \in M$, the tangent spaces are indeed isomorphic, but not canonically.)

A section of the tangent bundle TM is a map $W\colon M \to TM$ so that $W(x) \in T_xM$ for all $x \in M$. A section of the tangent bundle is called a vector field. The section of any other bundle is defined in a similar fashion. We will define later what smoothness of a section means. This will be done twice, in local coordinates (section 1.5) and in an invariant fashion (section 9).

Any vector space operation can be performed for the tangent bundle (or any vector bundle for that matter). For example, the dual of the tangent bundle is the cotangent bundle, where the dual is taken fiber by fiber. The cotangent bundle T^*M is the disjoint union of the cotangent spaces T_x^*M.

Remark 1.3. Let us recall (or learn) what tensor products of vector spaces are. Consider three real vector spaces E, F, G. The tensor product of E and F is the space $E \otimes F$ together with a bilinear map $\alpha\colon E \times F \to E \otimes F$

[6]The tangent bundle is also a smooth manifold itself, and we shall make heavy use of that later on. But for now it is merely a collection of tangent spaces. Treating it as a manifold opens new doors, but we will not open them yet.

so that for any bilinear map $b\colon E \times F \to G$ there exists a unique linear[7] map $\hat{b}\colon E \otimes F \to G$ so that the diagram

$$\begin{array}{ccc} E \times F & \xrightarrow{\quad b \quad} & G \\ & {\scriptstyle \alpha} \searrow \quad \nearrow {\scriptstyle \hat{b}} & \\ & E \otimes F & \end{array} \tag{3}$$

commutes. This property defines the space $E \otimes F$ and the map α uniquely up to natural isomorphism.

To get a concrete description (which is then equivalent to the universal description), let E and F have the bases $e_1, \dots, e_k$ and $f_1, \dots, f_l$. We can then declare $E \otimes F$ to be the vector space whose basis consists of the formal products $e_i \otimes f_j$ — so that the space has dimension kl. We define the required map by setting

$$\alpha\left(\sum_{i=1}^{k} v_i e_i, \sum_{j=1}^{k} w_j f_j\right) = \sum_{i=1}^{k} \sum_{j=1}^{k} v_i w_j e_i \otimes f_j. \tag{4}$$

It may be more concrete to think of the tensor product so that it takes the free vector spaces over two (finite) sets into the free vector space over the product set.

Tensor products of several spaces can be defined recursively by tensoring more spaces in or by using the same definitions with multilinear maps with more than two inputs.

The cotangent space was constructed from the tangent space using the linear algebraic construction of dual spaces. Similarly, one can take the tensor product $TM \otimes TM$, which is a bundle whose fiber at x is $T_xM \otimes T_xM$. Tensor products of the tangent and cotangent bundles give rise to many of the bundles one encounters in differential geometry. For example, the Riemann curvature tensor R is a section of the bundle $TM \otimes T^*M \otimes T^*M \otimes T^*M$. In other words, for any $x \in M$ we have a multilinear map

$$R(x)\colon T_x^*M \times T_xM \times T_xM \times T_xM \to \mathbb{R}. \tag{5}$$

It is a 1-contravariant and 3-covariant tensor field, also called a tensor field of type $(1, 3)$.

A vector field is a tensor field of type $(1, 0)$ and covectors have type $(0, 1)$. A scalar has type $(0, 0)$.

[7]Notice that this is indeed the only linear map of the three.

For another example of a tensor field, recall that linear maps $T_xM \to T_xM$ can be thought of as elements of the tensor product $T_xM \otimes T_x^*M$. The bundle with these fibers is $TM \otimes T^*M$. Sections of this bundle are "matrix fields" in the sense that at each point $x \in M$ it provides a linear map $T_xM \to T_xM$. These are tensor fields of type $(1, 1)$. (This is the endomorphism bundle of TM, so called because the value of a section at each point is an endomorphism of the relevant tangent space.)

Tensor products can be understood as spaces of multilinear maps. First, the dual of a real vector space E is the space of linear maps $E \to \mathbb{R}$. We can write $E^* = ML(E; \mathbb{R})$, so it is a multilinear map of one variable — which is a complicated way to say "linear". We also have $E = ML(E^*; \mathbb{R})$ using the natural identification $E = (E^*)^*$ of finite-dimensional spaces. Now we can proceed to tensor products. We have $E^* \otimes E^* = ML(E \times E; \mathbb{R})$ and $E \otimes E \otimes E^* = ML(E^* \times E^* \times E; \mathbb{R})$. Using associativity of tensor products we can also see $E^* \otimes E$ as $ML(E; E)$, and this particular interpretation is studied in exercise 1.5. This allows us to see the Riemann curvature tensor as a multilinear map $(T_xM)^3 \to T_xM$.

Exercise 1.5. Let E and F be two finite-dimensional real vector spaces. There is a natural mapping Φ from the space $L(E; F)$ of linear maps $E \to F$ to the tensor product $F \otimes E^*$. Describe this map (or its inverse) in formulas or in words or in pictures — or a combination thereof. ○

The idea of bundles is necessarily a little vague here as our focus is elsewhere. The hope is that these first impressions make it easier to pick up ideas along the way and make the reader motivated and well equipped to treat general bundles later on. We will return to the structure of bundles in section 9.

1.5 Coordinate representations of tensor fields

Consider now a single coordinate patch $U \subset M$. Identifying U with $\varphi(U) \subset \mathbb{R}^n$, we can use Euclidean coordinates[8] x^i on this subset of M. Let us consider the tangent and cotangent spaces at a point $x \in U$. Both can be identified with $\mathbb{R}^n$, but it is good to choose a specific identification.

A natural basis for the Euclidean space $\mathbb{R}^n$ consists of the standard unit vectors. However, when considering tangent vectors as derivations (first order

[8]The index is up. This is just a convention, but life is much easier when one sticks to it. Lower indices will have a dual meaning.

differential operators), it is most natural to let the basis vectors be[9]

$$\partial_i \coloneqq \left.\frac{\partial}{\partial x^i}\right|_x \in T_xM. \tag{6}$$

Evaluation at the point x and indeed the dependence on x is left implicit in the notation ∂_i. The notation would quickly become unwieldy with everything spelled out, which is why we have chosen to abbreviate the notation of the basis vectors.

The corresponding dual basis consists of the vectors $\mathrm{d}x^i \in T_x^*M$. Just as in regular linear algebra, the dual basis is defined by

$$\mathrm{d}x^i(\partial_j) = \delta^i_j. \tag{7}$$

The Kronecker delta δ^i_j tends to have one index up and another one down. In fact, the ith component of the local coordinates $\varphi\colon U \to \mathbb{R}^n$ can be seen as a map $x^i\colon U \to \mathbb{R}$, and the differential of this map $\mathrm{d}x^i$ is the dual basis element. This justifies the notation.

A vector $W \in T_xM$ and a covector $\alpha \in T_x^*M$ can now be expressed in these bases:

$$\begin{aligned} W &= W^i\partial_i, \qquad \text{and} \\ \alpha &= \alpha_i \mathrm{d}x^i. \end{aligned} \tag{8}$$

Observe that the basis and the components have indices in the opposite places.

Here we have for the first time employed the Einstein summation convention:

$$\begin{aligned} W^i\partial_i &\coloneqq \sum_{i=1}^n W^i\partial_i, \qquad \text{and} \\ \alpha_i \mathrm{d}x^i &\coloneqq \sum_{i=1}^n \alpha_i \mathrm{d}x^i. \end{aligned} \tag{9}$$

That is, when an index appears once up and once down, all possible values are summed over. If an index appears more than twice or both occurrences are up or both down, there is an issue.[10]

★ *Important exercise* 1.6. Show that

[9] When we differentiate with respect to something that has an upper index, we get a lower index. In time this hopefully makes sense.

[10] This is a non-issue in Euclidean geometry.

$$W^i = \mathrm{d}x^i(W) \tag{10}$$

and

$$\alpha_i = \alpha(\partial_i). \tag{11}$$

This gives us a way to find the components of a vector or a covector in a given basis.

As often, dependence on x was left implicit. ◯

These basis elements on the tangent and cotangent spaces are crucial for building the smooth structure of the tangent and cotangent bundles in section 9.

Consider then a tensor field a of type $(1,1)$. As discussed in section 1.4, $a(x)\colon T_xM \to T_xM$ is a linear map. As any linear map, $a(x)$ can be expressed as a matrix once a basis is given. Indeed,

$$a(x) = a^i_j(x)\partial_i \mathrm{d}x^j. \tag{12}$$

The component a^i_j describes how the jth component of the input contributes to the ith component of the output. The component can be extracted from $a(x)$ using

$$a^i_j(x) = \mathrm{d}x^i(a(x)\partial_j). \tag{13}$$

The general method is the same: operate with the tensor field on the basis vector field(s) and then use the basis covector field(s) to evaluate the component(s).

Smoothness of a tensor field means that all component functions are smooth. Given some local coordinates, each component of a tensor field is a real-valued function. The derivative of the component a^i_j with respect to the coordinate x^k is denoted by $a^i_{j,k}$. Such derivatives do not behave well enough under changes of coordinates, so the coordinate derivatives are not generally the components of a tensor field.

Exercise 1.7. Express the components $R^i{}_{jkl}$ of a type $(1,3)$ tensor field R using the basis vectors and covectors. ◯

As we only use a single coordinate system, we need not study how the tensor fields transform when coordinates are changed.

1.6 A new look at Euclidean linear algebra

Consider the manifold $M = \mathbb{R}^n$ and in particular its tangent space $T_0M \cong \mathbb{R}^n$. The basis vectors are given by

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{14}$$

and the other standard basis vectors e_i. In our Riemannian notation $e_i = \partial_i$. A vector is written in terms of the basis as $V = V^i e_i$.

It is natural to think of a vector as a column vector. A row vector corresponds to a covector, $\alpha = \alpha_i e^i$, where e^i are the dual basis vectors to e_i. There is a natural identification of the two bases, given by

$$e^i(W) = \langle e_i, W \rangle \,. \tag{15}$$

If we map $e_i \mapsto e^i$ and extend linearly, we get a linear map $\mathbb{R}^n \to (\mathbb{R}^n)^*$. This identification is based on the inner product. In general, inner products are a way to identify a space with its dual.

The ith component of a vector W is found by

$$W^i = e^i(W) = \langle e_i, W \rangle \tag{16}$$

as familiar.

★ *Important exercise* 1.8. Given a linear map $L\colon \mathbb{R}^n \to \mathbb{R}^m$, how can you find its matrix elements with respect to some bases on the two spaces? Compare to (13). ○

By the identification of the bases we can identify column vectors with row vectors. This corresponds exactly to transposition. The duality pairing $\alpha(W)$ is just the matrix product of a row vector and a column vector. The inner product of two column vectors can be obtained by transposing one of them and then multiplying as matrices. The concept of transpose is based on the inner product and changes if the inner product is changed. And we will change it.

1.7 Riemannian metric

A Riemannian metric is a smooth tensor field g of type $(0, 2)$ that satisfies a positivity condition and a symmetry condition. As a tensor field of this type, $g(x)$ is a bilinear map $T_xM \times T_xM \to \mathbb{R}$. The positivity condition is that

$$g(x)(v, v) > 0 \tag{17}$$

whenever $v \in T_xM$ is non-zero. The symmetry condition is that

$$g(x)(v, w) = g(x)(w, v) \tag{18}$$

for all $v, w \in T_xM$. This gives rise to a rich geometric structure.

The convention in the sequel is as follows: M is always a smooth manifold of dimension n, and it has a fixed Riemannian metric g. In other words, (M, g) is a Riemannian manifold. We assume M to be connected.[11] Unless otherwise mentioned, we will be working in a single coordinate chart so as to avoid unnecessary complications.

The simplest example is the Euclidean metric on $M = \mathbb{R}^n$. All tangent spaces can be canonically identified with the same Euclidean space ($T_xM = \mathbb{R}^n$ for all $x \in M$), and the Euclidean metric $g(x)\colon T_xM \times T_xM \to \mathbb{R}$ is the x-independent quadratic form $g(x)(v, w) = \sum_{i=1}^n v^i w^i$, with the vectors $v, w \in \mathbb{R}^n$ expressed in coordinates in the usual Euclidean manner. Using notation from section 1.6, we can write $g = \sum_{i=1}^n e^i \otimes e^i$, understanding that $e^i \otimes e^i$ acts on a pair of vectors as $(e^i \otimes e^i)(v, w) = v^i w^i$. It is easy to check that this g is a valid Riemannian metric on M. If you are at any point during this course confused, it is advisable to look at the source of confusion from the point of view of this explicit Riemannian manifold.

Many other examples, including spheres and hyperbolic spaces, can be obtained by multiplying the Euclidean metric tensor $g(x)$ with a positive smooth function — a conformal factor — $c(x)$ to get a new Riemannian metric $\tilde{g}(x) = c(x)g(x)$. In fact, all Riemannian metrics in two dimensions are locally of this form upon choosing good coordinates, but this is not true in higher dimensions.[12]

⋆ *Important exercise* 1.9. Do you have any questions or comments regarding section 1? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ◯

2 Distance and geodesics

2.1 An inner product

A Riemannian metric gives an inner product on the tangent space. Namely, the inner product of two vectors $v, w \in T_xM$ is given simply by

$$\langle v, w \rangle := g(v, w). \tag{19}$$

[11] If M is disconnected, the different connected components have completely independent lives. We lose awkward situations but no generality in assuming connectedness.

[12] If you wish to dig further, the keywords would be “conformal flatness” and “uniformization theorem”.

We will often leave the dependence of the metric tensor on the base point x implicit.

Exercise 2.1. Expand objects in terms of their components and show that $\langle v, w\rangle = g_{ij}(x)v^i w^j$. ◯

As described in the Euclidean setting, an inner product gives a canonical way to identify vectors with covectors. In fact, one can consider g as a linear map $T_xM \to T_x^*M$ given by

$$v \mapsto g(v, \cdot). \tag{20}$$

Written in terms of components, the vector with components v^i is mapped to the covector with components $g_{ij}v^j$. This covector is denoted by $v^\flat$ and called "v flat".

★ *Important exercise* 2.2. Show that the map $v \mapsto v^\flat$ is bijective. You will need the positivity condition (17). ◯

The inverse of the map $v \mapsto v^\flat$ maps a covector α to the vector $\alpha^\sharp$, called "α sharp". These are the musical isomorphisms and they satisfy $v = (v^\flat)^\sharp$ and $\alpha = (\alpha^\sharp)^\flat$.

Given the canonical bases on T_xM and T_x^*M, the matrix of the "flat map" is g_{ij} itself. The matrix of the inverse map, the "sharp map", is denoted by g^{ij} and is the inverse of this matrix — it satisfies $g^{ij}g_{jk} = \delta^i_k$. Invariantly, this can be denoted as g^{-1}.

Exercise 2.3. Show that $g^{ij}(v^\flat)_i(w^\flat)_j = \langle v, w\rangle$. ◯

Exercise 2.4. Show that g^{ij} defines an inner product on T_x^*M and the musical isomorphisms preserve the inner product. ◯

The inner products give us natural definitions of norms for the tangent and cotangent spaces: $|v| = \langle v, v\rangle^{1/2}$ and $|\alpha| = \langle \alpha, \alpha\rangle^{1/2}$ using the relevant inner products. The musical isomorphisms are isometries. The (co)tangent space $T_x^{(*)}M$ is also isometric to $\mathbb{R}^n$, as are all n-dimensional real inner product spaces.

Due to the way the musical isomorphisms work in coordinates — $(v^\flat)_i = g_{ij}v^j$ and $(\alpha^\sharp)^i = g^{ij}\alpha_j$ — they are sometimes called lowering and raising indices.

Recall that the differential $\mathrm{d}f$ of a scalar function $f\colon M \to \mathbb{R}$ is a covector field. The corresponding vector field is called its gradient: $\nabla f = (\mathrm{d}f)^\sharp$.

One would obtain much more general structures by taking a norm on the tangent space that does not correspond to an inner product. This would lead to Finsler geometry.

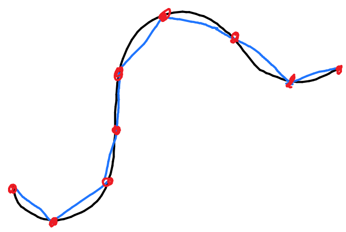


Figure 3: A curve can be approximated by discrete steps between finitely many points on the curve. The length of a step at time t is roughly $|\dot\gamma(t)|\,\Delta t$, where Δt is the time step. Letting the time steps go to zero turns the approximating sum into the Riemann sum of the integral defining length.

2.2 On computations in local coordinates

Let us consider the flat map as an example. If v is a vector field, then $\alpha = v^\flat$ is given in local coordinates as $\alpha_i = g_{ij}v^j$. Including the variable and the sum explicitly, this means

$$\alpha_i(x) = \sum_j g_{ij}(x)v^j(x). \tag{21}$$

If we need to compute a derivative like $\partial_k\alpha_i$ in these local coordinates, it can be helpful to look at (21). Each $\alpha_i(x)$ is a real valued function of $x \in \mathbb{R}^n$ (or rather only in the set $\varphi(U) \subset \mathbb{R}^n$), and so are $g_{ij}(x)$ and $v^j(x)$. Each component is just a real-valued function — coordinate expressions are almost always expressions containing sums and products of real numbers, nothing more elaborate. When you differentiate, the normal product rule applies without any changes.

2.3 Length of curve

Recall that the length of a smooth curve $\gamma\colon [a,b] \to \mathbb{R}^n$ is defined by

$$\ell(\gamma) = \int_a^b |\dot\gamma(t)|\, \mathrm{d}t. \tag{22}$$

We define the length of a smooth curve $\gamma\colon [a,b] \to M$ by the same formula. Figure 3 should make the definition more intuitively sensible.

To properly do so, we must know what $\dot\gamma(t)$ is. As discussed in section 1.3, velocities of curves are one way to define tangent vectors in the first place, so $\dot\gamma(t)$ should be an element of $T_{\gamma(t)}M$.

In local coordinates one can write $\dot\gamma(t) = \dot\gamma^i(t)\partial_i$. The length of $\dot\gamma(t)$ is given by the metric tensor. Notice how the norm used to measure the length of $\dot\gamma(t)$ is different for different values of t.

Everything is defined so that the length of a curve is independent of the choice of coordinates and parametrization.

2.4 Distance between points

Let $p, q \in M$ be any two points. As M is connected, there is a smooth path between the two points[13]. We define the distance between them to be

$$d(p,q) = \inf\{\ell(\gamma); \gamma\colon [0,1] \to M, \gamma(0) = p, \gamma(1) = q\}. \tag{23}$$

It is typical to choose the curve family so that γ is piecewise smooth, but smooth will work just as well.

Exercise 2.5. Explain with a picture or maybe even a proof why minimizing length of piecewise smooth curves will lead to the same infimum as minimizing over smooth curves. ○

This concept of distance defines a metric in the sense of metric spaces. But we will restrict the word "metric" to the metric tensor and call this d the distance.

Exercise 2.6. Give an example of two points in a Euclidean domain where a minimizing curve does not exist within the domain. The same issue can occur on manifolds, so existence of minimizers requires assumptions. (A local result is given in exercise 7.5.) ○

Proposition 2.1. *The manifold M with the distance d satisfies all the axioms of a metric space. Its topology coincides with that of the topological manifold M.*

The proof of coincidence of the two topologies can be found in many introductory treatises of Riemannian geometry. It suffices to prove such equivalence within a chart, and that follows from the distance being bi-Lipschitz to the underlying Euclidean metric where the coordinates live. See exercise 2.8.

★ *Important exercise* 2.7. Explain why d is symmetric and satisfies the triangle inequality. ○

[13] This is a feature of smooth manifolds. In general a connected space is not path connected, and points in a path connected space need not be connectable by a rectifiable (finite length) curve.

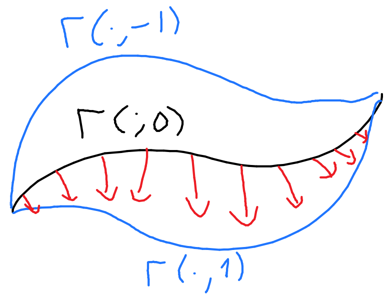


Figure 4: A family of curves. The reference curve is $\Gamma(\,\cdot\,, 0) = \gamma(t)$, and different values of s give neighboring curves. The derivative of the family $\Gamma(t, s)$ in s at $s = 0$ gives rise to the variation field of the family.

Exercise 2.8. Show that if $d(x, y) = 0$, then $x = y$. You can work in local coordinates near x. Argue by continuity that $C^{-1} |v|_{\mathbb{R}^n} \leq |v| \leq C |v|_{\mathbb{R}^n}$ for all $v \in TU$ for a small neighborhood U of x (in those local coordinates) and for some constant $C > 1$. Using that estimate find a lower bound on the length of any smooth curve joining x and y. ○

2.5 First variation of length

We want to find the shortest curve between two points. We do so using smooth calculus of variations. The aim is to find the Euler–Lagrange equation and later show that its solutions are actually minimal.

Let[14] $\Gamma\colon [0, 1] \times (-\varepsilon, \varepsilon) \to M$ be a smooth map. We understand $\Gamma(t, s)$ to be a family of curves so that each $\Gamma(\,\cdot\,, s)$ is a curve. We want to differentiate

$$\ell(\Gamma(\,\cdot\,, s)) = \int_0^1 |\partial_t \Gamma(t, s)| \, \mathrm{d}t \tag{24}$$

at $s = 0$. The family and its variation field are depicted in figure 4. Let us work in local coordinates again.

Exercise 2.9. Show that

$$\begin{aligned} &\partial_s [g(\Gamma(t, s))(\partial_t \Gamma(t, s), \partial_t \Gamma(t, s))]^{1/2} \\ &= \frac{1}{2\,|\partial_t \Gamma|} (g_{ij,k}(\Gamma) \partial_s \Gamma^k \partial_t \Gamma^i \partial_t \Gamma^j + 2 g_{ij}(\Gamma) \partial_t \Gamma^i \partial_t \partial_s \Gamma^j), \end{aligned} \tag{25}$$

where the argument (t, s) of Γ has been left out for clarity. Here we used the derivative notation $g_{ij,k} := \partial_k g_{ij}$ again. ○

[14]Here and henceforth, $\varepsilon > 0$ is a small number and may appear without being quantified.

We are now ready to compute the variation of length of a family of constant speed curves from p to q. Reparametrization does not change length, so we will now parametrize our curves by $[0,1]$. This reparametrization preserves smoothness as long as $\partial_t\Gamma \neq 0$.

Proposition 2.2. *Let $\Gamma\colon [0,1]\times(-\varepsilon,\varepsilon)\to M$ be a smooth map so that*

- *$|\partial_t\Gamma(t,0)|$ is constant,*
- *$\Gamma(0,s)=p$ for all s, and*
- *$\Gamma(1,s)=q$ for all s.*

Denoting[15] $\gamma(t) \coloneqq \Gamma(t,0)$, $\dot\gamma(t) \coloneqq \partial_t\Gamma(t,0)$, $\ddot\gamma(t) \coloneqq \partial_t^2\Gamma(t,0)$, and $V(t) \coloneqq \partial_s\Gamma(t,s)|_{s=0}$, we have

$$\partial_s\ell(\Gamma(\,\cdot\,,s))|_{s=0} = \int_0^1 \frac{1}{|\dot\gamma|}V^k\left[\frac{1}{2}g_{ij,k}\dot\gamma^i\dot\gamma^j - g_{ik,j}\dot\gamma^i\dot\gamma^j - g_{ik}\ddot\gamma^i\right]\mathrm{d}t. \tag{26}$$

Proof. Exercise 2.9 shows that the derivative in question is

$$\int_0^1 \frac{1}{|\dot\gamma|}\left[\frac{1}{2}g_{ij,k}V^k\dot\gamma^i\dot\gamma^j + g_{ik}\dot\gamma^i\partial_tV^k\right]\mathrm{d}t. \tag{27}$$

We integrate by parts in the second term to take the ∂_t away from V^k. As $|\dot\gamma|$ is independent of t and $V(0)=0$ and $V(1)=0$, we find the desired form of the derivative. □

If the curve $\gamma(t)=\Gamma(t,0)$ is to be minimizing within this family, this derivative should vanish for any variation field $V(t)$. This inspires us to define a geodesic to be a constant speed curve which satisfies

$$\frac{1}{2}g_{ij,k}\dot\gamma^i\dot\gamma^j - g_{ik,j}\dot\gamma^i\dot\gamma^j - g_{ik}\ddot\gamma^i = 0. \tag{28}$$

In fact, it turns out that solutions to this equation automatically have constant speed; see corollary 4.3.

It is important to read this result the right way. We have shown that a smooth minimizing curve is a geodesic — which means satisfying the geodesic equation. We have not shown that minimizers exist or that they are smooth. That will come much later.

Remark 2.3. In addition to the length functional $\ell(\gamma)=\int|\dot\gamma|\,\mathrm{d}t$ one can also study the energy functional $E(\gamma)=\frac{1}{2}\int|\dot\gamma|^2\,\mathrm{d}t$. It has the nice property that all critical points are constant speed geodesics, so it leads to the geodesic equation more directly. The reason we do not use it is the lack of a geometric interpretation as clear as that of length. We will study length at length.

[15] Notice that the second order derivatives are computed in local coordinates. We do not yet have proper tools to handle them invariantly. We will later, and the formula simplifies considerably; see (51).

2.6 The Christoffel symbol

The Christoffel symbol is a gadget that looks a bit like a type $(1,2)$ tensor field — but is not due to the the way it changes with coordinates. It is defined in local coordinates as

$$\Gamma^i{}_{jk} = \frac{1}{2} g^{il}(g_{lj,k} + g_{lk,j} - g_{jk,l}). \tag{29}$$

This symbol will appear often in coordinate formulas. We immediately point out the symmetry property:

$$\Gamma^i{}_{jk} = \Gamma^i{}_{kj}. \tag{30}$$

Exercise 2.10. Show that equation (28) is equivalent with

$$\ddot{\gamma}^i + \Gamma^i{}_{jk}\dot{\gamma}^j\dot{\gamma}^k = 0. \tag{31}$$

This is called the geodesic equation. ◯

Observe that in Euclidean geometry where $g_{ij}(x)$ is independent of the base point x the Christoffel symbol vanishes. On more general manifolds its appearance is inevitable, but it will disappear in an invariant treatment. In fact, it is what helps make derivatives invariant.

If one does a non-inertial change of coordinates in classical mechanics, one introduces pseudoforces such as the centrifugal force. The Christoffel symbol can be seen as a pseudoforce term: a geodesic would continue at constant speed ($\ddot{\gamma}^i = 0$) without its effect. A typical Riemannian manifold does not admit "inertial coordinates" and the Christoffel symbol appears. (They can be made vanish at a single point as in exercise 6.8, and even along curves.) We will also find an invariant form of the geodesic equation which in a sense removes the pseudoforces from the picture.

2.7 The geodesic equation

A solution to the geodesic equation is called a geodesic. It follows from standard ODE theory that for any $x \in M$ and any $v \in T_xM$ there is a unique geodesic $\gamma\colon (-\varepsilon, \varepsilon) \to M$ so that $\gamma(0) = x$ and $\dot{\gamma}(0) = v$, as in figure 5. Existence for long times is not guaranteed unless additional structure is introduced.[16]

Exercise 2.11. Use this result:

[16] If you are interested, look up geodesic completeness and the Hopf–Rinow theorem.

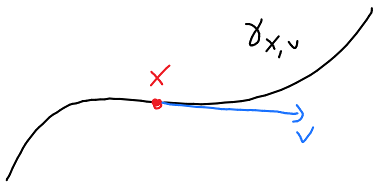


Figure 5: Given any point x and any vector v at it, there is a unique geodesic $\gamma_{x,v}$.

> If $F\colon \mathbb{R}^N \to \mathbb{R}^N$ is Lipschitz, then the ODE $u'(t) = F(u(t))$ has a unique local C^1 solution for any given initial conditions $u(0) = u_0 \in \mathbb{R}^N$.

Prove the local existence and uniqueness result for the geodesic equation. ◯

Exercise 2.12. Consider the quoted ODE result of the previous exercise. Show that if F is smooth, so is u. This proves that geodesics are necessarily smooth. ◯

We stress that we define a geodesic to be a solution to the geodesic equation. (The equation will have a couple of equivalent forms.) That geodesics actually minimize length is not entirely trivial, so we shall prove it later.

Existence of minimizers has not been established yet either. The Arzelà–Ascoli theorem can be used to produce a minimizer, but often of very low regularity. We will use smooth tools instead.

⋆ *Important exercise* 2.13. Do you have any questions or comments regarding section 2? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ◯

3 Connections and covariant differentiation

3.1 Connections in general

It is not always obvious what differentiation should mean. For a function $M \to \mathbb{R}$ we can assign a differential as a covector (a cotangent vector). The derivative of a function $\mathbb{R} \to M$ (a curve) can be treated as a vector (a tangent vector). These behave well under changes of coordinates, and indeed these derivatives can be used to define vectors and covectors in the first place.

Differentiation of vectors does not make sense equally simply. Consider a vector field $W(x)$. What does it mean for $W(x)$ to stay constant as x

changes? Each $W(x)$ belongs to T_xM, so the underlying space changes. We need a way to compare tangent vectors on nearby tangent spaces. We will only compare along curves, not between any pair of points; the comparison will depend on the choice of curve.

The same issue arises with all kinds of bundles. The analogue of a vector field or a tensor field on a general bundle is called a section. A consistent method of differentiating a section of a bundle is called a connection. A connection for vector fields is called an affine connection.

Definition 3.1. An affine connection ∇ on a manifold M is a bilinear map that maps a pair (X, Y) of vector fields into a vector field $\nabla_X Y$ so that the following conditions hold for any smooth function $f\colon M \to \mathbb{R}$:
- $\nabla_{fX} Y = f\nabla_X Y$
- $\nabla_X(fY) = f\nabla_X Y + X(f)Y$.

These conditions describe the linearity when the vector fields are multiplied by a scalar function instead of a single number. (A reader familiar with more abstract linear algebra may enjoy the observation that vector fields constitute a module over the ring $C^\infty(M;\mathbb{R})$ of smooth functions.)

One can read $\nabla_X Y$ as "the derivative of the vector field Y in the direction of the vector field X". If $X, Y\colon \mathbb{R}^n \to \mathbb{R}^n$ are smooth vector fields, the standard affine connection of Euclidean geometry is given by

$$(\nabla_X Y)^j = X^i Y^j_{\ ,i} \tag{32}$$

using the usual coordinates of $\mathbb{R}^n$.

Exercise 3.1. Show that the Euclidean connection defined above is indeed an affine connection on the space $\mathbb{R}^n$. You will see the familiar Leibniz rule take a new form. ◯

3.2 The Levi-Civita connection

There are a great many connections on a smooth manifold. The definition of a connection had nothing to do with a metric tensor. We would of course like the concept of differentiation to be somehow compatible with the metric.

Before giving a definition of such a good connection, we need to recall the concept of a commutator. The commutator of two linear operators A and B is $[A, B] \coloneqq AB - BA$. The commutator of two differential operators of orders k and m is a differential operator of order $k+m-1$. In particular, the commutator of two derivations (first order differential operators) is another derivation.

Therefore the commutator of two vector fields is a vector field. One can define it explicitly as $[X,Y]f = X(Yf) - Y(Xf)$, where the vector fields turn scalar fields to scalar fields.

Exercise 3.2. Let X and Y be two vector fields. Show that their commutator is a vector field with the components

$$[X,Y]^i = X^j Y^i_{\ ,j} - Y^j X^i_{\ ,j}. \tag{33}$$

This shows that the commutator as differential operator has only first order terms and is therefore a vector field. ◯

Definition 3.2. An affine connection ∇ on a Riemannian manifold (M,g) is called a symmetric metric connection if

- $\nabla_X Y - \nabla_Y X = [X,Y]$ and
- $Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$.

The first condition is a Leibniz rule for the inner product; a Leibniz rule of a different nature was included in the definition of an affine connection. The point is that although $g(Y,Z)$ contains three tensor fields (the metric tensor and the two vector fields), there are no derivatives of the metric tensor in the formula. We will see in a moment that indeed the covariant derivative of the metric tensor is zero.

The second condition has nothing to do with the metric. Instead, it states that something called the torsion of the connection vanishes. The torsion measures how the tangent spaces twist as one moves from one base point to another. A rough heuristic way to see the condition is that we want the tangent spaces to rotate but not twist.

Every Riemannian manifold has a unique symmetric metric connection[17], and it is called the Levi-Civita connection[18]. The connection is defined so that for two vector fields $X(x)$ and $Y(x)$ we have

$$(\nabla_X Y)^i = X^j Y^i_{\ ,j} + \Gamma^i_{\ jk} X^j Y^k. \tag{34}$$

It is not apparent as we have not bothered with changing coordinates, but $\nabla_X Y$ is indeed a valid vector field.

Exercise 3.3. Prove that the Levi-Civita connection is an affine connection. ◯

Exercise 3.4. Prove that the Levi-Civita connection is a symmetric metric connection. ◯

[17] We will not prove this theorem.

[18] This is named after Tullio Levi-Civita, a single person. The connection is therefore called the "Levi-Civita connection" instead of the "Levi–Civita connection".

3.3 Covariant differentiation

We would like to be able to differentiate tensor fields of all kinds. We continue to use ∇ for this purpose, but in the sequel we will rarely need to differentiate very complicated tensor fields. For any tensor field T of any type (k, l) and a vector field X, we would like to be able to compute $\nabla_X T$, the covariant derivative of T in the direction of X. This should all be defined so that $\nabla_X T$ is also a tensor field of type (k, l) and thus behaves under coordinate changes as a tensor field should. As $\nabla_X T$ is linear[19] in X, we may regard ∇T as a tensor field of type $(k, l + 1)$.

Any affine connection gives rise to such a way, as long as we require the following:

- On scalar functions the covariant derivative is simply the derivative by a vector field: $\nabla_X f = Xf$.
- On vector fields we have the original connection.
- Tensor products satisfy the Leibniz rule

$$\nabla_X (T \otimes R) = \nabla_X T \otimes R + T \otimes \nabla_X R. \tag{35}$$

- The covariant derivative commutes with any contraction or trace.[20]

The Levi-Civita connection has an additional property that neatly describes the metric compatibility:

$$\nabla g = 0. \tag{36}$$

That is, the concept of differentiation is defined so that the metric tensor g is “constant”. (A more appropriate technical term is “parallel”.)

The interpretation that the metric tensor is “constant” should not be taken too literally. It only concerns covariant differentiation. There is some genuine variation in the metric that cannot be erased by a clever choice of connection, and that is captured by the concept(s) of curvature.

The covariant derivative of a scalar field f in the direction of a vector field X is $\nabla_X f = Xf = \mathrm{d}f(X)$. Therefore, if we think of the covariant derivative ∇f as a tensor field of type $(0, 1)$, we find that $\nabla f = \mathrm{d}f$. We mentioned in section 2.1 that the gradient of a function f can be defined as the vector field $(\mathrm{d}f)^\sharp$ corresponding to the covector field $\mathrm{d}f$. The gradient vector field is usually denoted by ∇f. This is confusing with the covariant derivative, but fortunately the musical isomorphisms send the two objects

[19] Although linearity at a point is more tangible, in some sense the more correct concept of linearity is that over the ring $C^\infty(M)$ of smooth functions. Tensor fields of any given type constitute a module over this ring. In this view, $\nabla_X T$ is not linear in T. It depends not only on the value of T but also its derivatives at a point.

[20] We have not introduced this concept nor will we use it explicitly. This statement is here for completeness.

denoted by ∇f to each other in a canonical way. We shall denote the differential (and therefore the covariant derivative) of a scalar function by $\mathrm{d}f$, although some more consistency with other covariant derivatives would be achieved by different notation.

To get all of this on a more concrete footing, let us see how to covariantly differentiate a tensor field given in terms of components in some local coordinates. For a vector field Y we have directly the formula of the Levi-Civita connection:

$$(\nabla_X Y)^i = X^j Y^i{}_{,j} + \Gamma^i{}_{jk} X^j Y^k. \tag{37}$$

★ *Important exercise* 3.5. The coordinate vector fields ∂_i are of course valid vector fields within their coordinate patch. What is $\mathrm{d}x^i(\nabla_{\partial_j}\partial_k)$? Describe in words what it means and give a formula. ○

We would then like to find a similar expression for $(\nabla_X \alpha)_i$ for a covector field α.

Exercise 3.6. Starting with the covariant derivative of a vector field and the Leibniz rule

$$X(\alpha(Y)) = (\nabla_X \alpha)(Y) + \alpha(\nabla_X Y) \tag{38}$$

(which follows from the tensor product rule and the trace rule stipulated above), show that

$$(\nabla_X \alpha)_i = X^j \alpha_{i,j} - \Gamma^j{}_{ik} X^k \alpha_j. \tag{39}$$

This is the covariant differentiation rule of covector fields. ○

A tensor field of any type can be differentiated in a similar fashion. For every upper index we add a term like we had for vectors and for all lower indices we add a term like for covectors. For example, the covariant derivative of a type $(1,1)$ tensor a is given by

$$(\nabla_X a)^i_j = X^k a^i_{j,k} + \Gamma^i{}_{kl} a^k_j X^l - \Gamma^k{}_{lj} a^i_k X^l. \tag{40}$$

★ *Important exercise* 3.7. What is the coordinate expression for $\nabla_X g$ for a type $(0,2)$-tensor g? ○

Exercise 3.8. Show directly using the formula of the previous exercise that $\nabla_X g = 0$ when g is the metric tensor. ○

Remark 3.3. When plugging in indices, remember what type of object everything is. For example, a vector field X has coordinates X^i. Expressions like X^{ij} or X_i are nonsensical, as this type of object has only one index slot and it is up. Coordinate derivatives can add indices, but only in one way: after a comma and in the subindex. Expressions like $X^i{}_{,j}$ and $X^i{}_{,jklm}$ are valid.

3.4 On notation

There are various different notations in use in differential geometry. Different conventions are convenient in different situations, and the different ways to express the same thing offer new points of view.

For example, the derivative of a scalar function $f\colon M \to \mathbb{R}$ in the directions of a vector field X on M can be written as

$$\nabla_X f = Xf = \mathrm{d}f(X) = \langle \nabla f, X \rangle = \langle \mathrm{d}f, X \rangle, \tag{41}$$

where the last inner product is the duality pairing between T_xM and T_x^*M. And this list is not exhaustive; for example, in some cases it is convenient to denote $\mathrm{d}f$ by f^* and call it the pushforward. The same object can also be expressed in local coordinates as $X^i\partial_i f$ or $X^i f_{,i}$.

Componentwise notations also vary somewhat. It is customary to have all indices "in sequence" whether up or down, so that a gap is left where an index is in the other place. This means writing, for example, $T^{i\ k}_{\ j\ l}$ instead of T^{ik}_{jl}. This only really becomes crucial when raising and lowering indices by the musical isomorphisms (which extends to tensor fields), so this convention is not always followed.

In Riemannian geometry one can naturally identify tangent vectors with cotangent vectors using the musical isomorphisms. It is possible to leave the isomorphisms implicit and just let indices wander around freely. However, it is instructive to keep track at least of vectors and covectors. There are situations where a Riemannian metric is not available for music and often the natural kind of object sits most comfortably in any computation.

We have seen two types of differentiation. The simplest kind is coordinate differentiation. For example, the coordinate derivative of a vector field V^i would be

$$\frac{\partial}{\partial x^j} V^i(x) = \partial_j V^i = V^i_{\ ,j}. \tag{42}$$

This is an object with one index up and another down, but it is not a tensor field of type $(1,1)$ due to the issue of coordinate invariance which we have kept mysterious.

The covariant derivative of V in the direction of the vector field Y is $\nabla_Y V$. Its components are given by (37). One can write this in local coordinates as

$$(\nabla_Y V)^i = Y^j V^i_{\ ;j} \tag{43}$$

by introducing the notation

$$V^i_{\ ;j} = V^i_{\ ,j} + \Gamma^i_{\ jk} V^k. \tag{44}$$

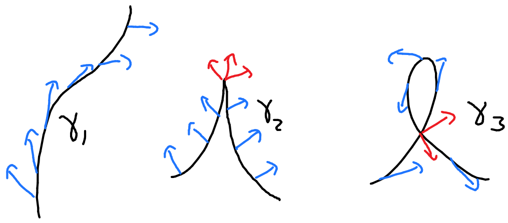


Figure 6: Three curves and vector fields along them. Where the curves are nice, the vector fields can be thought of as functions on the base point. But the curve γ_2 stops at the tip for a bit and the vector changes, and the curve γ_3 intersects itself; both situations forbid the vector field from depending only on the base point. This is why a vector field along a curve must be a function of time, not of point.

These are precisely the components of the $(1,1)$-type tensor field ∇V. The comma is used for coordinate differentiation and semicolon for covariant differentiation.

The Christoffel symbols are used as correction terms to make differentiation behave well.

★ *Important exercise* 3.9. Do you have any questions or comments regarding section 3? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ◯

4 Fields along a curve

4.1 Vector fields along a curve

Let $\gamma\colon I \to M$ be a smooth curve defined on an interval $I \subset \mathbb{R}$. We would like to give a natural space for the velocity vector $\dot{\gamma}(t)$ to live in. Each $\dot{\gamma}(t)$ is in $T_{\gamma(t)}M$, but this is not a vector field as previously described. It is only defined on a subset of the manifold, namely the trace $\gamma(I)$. And what if the curve intersects itself or even stops as in figure 6?

We define[21] a vector field along the curve γ to be a smooth map $V\colon I \to TM$ that satisfies $V(t) \in T_{\gamma(t)}M$ for all $t \in I$. There are two important examples:

- $\dot{\gamma}(t)$ is a vector field along γ.

[21] If one enjoys such language, one should think of vector fields along γ as sections of the pullback bundle γ^*TM. All kinds of tensor fields live along a curve in a similar fashion.

• If V is a vector field on M, then $V(\gamma(t))$ is a vector field along γ.
If $\dot{\gamma} \neq 0$, then at least locally any vector field along γ can be extended to its neighborhood and considered like the second example. But it is best to treat objects so that they require no artificial extensions; a vector field along a curve should only exist on the curve.

It is probably worth pointing out that a vector field along a curve need not point along the curve. It only has to be defined at all points on the curve — or rather at all values of parameters defining the curve.

4.2 Covariant differentiation along a curve

In local coordinates we define the covariant derivative $D_t V(t)$ of $V(t)$ along $\gamma(t)$ with respect to t to be

$$(D_t V(t))^i = \dot{V}^i(t) + \Gamma^i{}_{jk} V^j(t) \dot{\gamma}^k(l). \tag{45}$$

This is a derivative with respect to the time parameter t, but as before, a naive coordinate derivative is invalid.

Exercise 4.1. Suppose that γ is the integral curve of a vector field X on M. This means that $\dot{\gamma}(t) = X(\gamma(t))$ for all t. (We will return to integral curves in section 11.1.) Let V be any vector field on M. Show that[22]

$$D_t V = \nabla_X V. \tag{46}$$

Where does this equation make sense? ◯

The velocity of a curve γ is $\dot{\gamma}$. Its natural time derivative is $D_t \dot{\gamma}$, the "covariant acceleration". In Euclidean geometry it makes sense to say that a curve is straight if its acceleration vanishes. We can now do the same: we can say that a curve is straight when $D_t \dot{\gamma}(t) = 0$ for all t.

★ *Important exercise* 4.2. Show that a smooth curve γ is straight if and only if it is a geodesic. ◯

We have found a familiar fact: The shortest curves are straight. But, unlike in Euclidean geometry, a straight curve is not necessarily the shortest one between its endpoints.

We have found yet another form of the geodesic equation, this time an invariant one:

$$D_t \dot{\gamma}(t) = 0. \tag{47}$$

Compare this to the previous versions (28) and (31).

[22]We defined covariant differentiation along a curve so that this holds. There is only one definition that makes this work.

The first derivative of the curve γ is often denoted by $\dot{\gamma}$. Sometimes it is good to write it as $\partial_t\gamma$ for clarity. And as before, we can define covariant differentiation of the simplest objects to agree with the usual derivative, so that we may well write

$$\dot{\gamma} = \partial_t\gamma = D_t\gamma. \tag{48}$$

This is only a matter of notation, but its benefit will come clear soon. The geodesic equation gets yet another form:

$$D_t^2\gamma = 0. \tag{49}$$

This version is both neat and useful. We will see it soon in section 5 when studying Jacobi fields.

The covariant derivative along a curve is also compatible with the metric as one might expect. The following two rules establish the natural Leibniz rules for vector fields V and W and a scalar field f along γ. (A scalar field along a curve is simply a real-valued function defined on the interval where the curve is parametrized.) The time derivative of a scalar f could be written as $D_t f$ as well, but $\partial_t f$ highlights that we are only differentiating a number.

Exercise 4.3. Show that $D_t(fV) = (\partial_t f)V + fD_tV$. ○

Exercise 4.4. Show that $\partial_t \langle V, W\rangle = \langle D_tV, W\rangle + \langle V, D_tW\rangle$. ○

4.3 Parallel transport

Definition 4.1. A vector field V along a curve γ is said to be parallel if $D_tV = 0$.

A parallel vector field is the closest we can get to a constant vector field. Any vector at any point along a curve can be parallel transported along it.

Exercise 4.5. Let $\gamma\colon I \to M$ be a curve. Given any $t_0 \in I$ and $V_0 \in T_{\gamma(t_0)}M$, show that there is a unique parallel vector field V along γ with $V(t_0) = V_0$.

This is what it means to parallel transport V_0 from a single tangent space along the curve. ○

Beware that parallel transport happens along a curve, not just between two points. Even if a curve intersects itself, parallel transport around a loop rarely preserves the vector as depicted in figure 7. But it does preserve something:

Proposition 4.2. *If V and W are parallel vector fields along a curve γ, then their inner product $\langle V, W\rangle$ is constant. In particular, a parallel vector field has constant norm.*

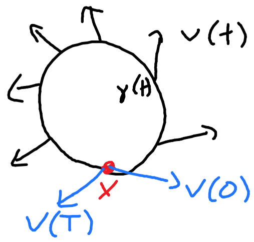


Figure 7: A vector $v(0)$ at a point $x = \gamma(0)$ is parallel transported along the curve γ. The curve loops back to the point x at time T but the parallel transported field $v(T)$ is different from the original. Such a rotation induced by parallel transport around a loop is known as holonomy.

Proof. As $D_t V = D_t W = 0$, exercise 4.4 implies that $\partial_t \langle V, W \rangle = 0$. The second claim is found by letting $V = W$. □

Corollary 4.3. *A geodesic has constant speed.*

Remark 4.4. When we did our calculus of variations to find the geodesic equation, we required that $|\dot{\gamma}|$ is constant. It should therefore be no surprise that a solution to the equation has constant speed. If we are free to reparametrize as we like, geodesics will certainly not be unique anymore. If we drop constant speed parametrization, we can describe geodesics to be those smooth curves $\gamma\colon I \to M$ for which $\dot{\gamma}(t) \neq 0$ and $D_t\dot{\gamma}(t) = f(t)\dot{\gamma}(t)$ for some smooth function $f\colon I \to \mathbb{R}$. This can be interpreted so that the acceleration of the curve must be along the curve. This is similar to describing Euclidean geodesics as $\gamma(t) = x + h(t)v$ for a function h with non-vanishing derivative; in its case $f(t) = h''(t)/h'(t)$.

We have found that a minimizing curve must be a geodesic. Now we know that geodesics are as straight as a curve on a Riemannian manifold can be and that they have constant speed[23]. What we have not discovered yet is whether a geodesic is always minimizing and whether one always exists between any two points. We will prove these statements later, but only locally as they are not generally globally true.

4.4 Orthonormal bases

The Riemannian metric makes each tangent space T_xM into an inner product space of dimension n. Therefore there is an orthonormal basis $e_1, \ldots, e_n$. As

[23] Although the length functional is parametrization independent, we did make use of constant speed parametrization to find the variation of length.

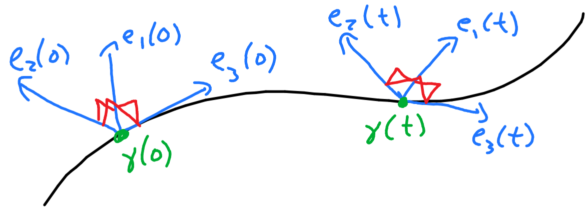


Figure 8: At $t = 0$ we choose an orthonormal basis $e_1(0), e_2(0), e_3(0)$ for the tangent space at the point x of a three-dimensional manifold. Parallel transporting these three vectors gives an orthonormal basis at any point $\gamma(t)$, and these together constitute an orthonormal parallel frame. We often choose $e_3(t) = \dot{\gamma}(t)$ as we have here; this is parallel if γ is a geodesic.

in Euclidean geometry, working within such a basis is convenient.

Now consider a smooth curve γ on M. We can take an orthonormal basis in the tangent space at any point and then parallel transport each[24] e_α along the curve. This gives rise to vector fields $e_\alpha(t)$ along γ.

Such a collection of vectors is called an orthonormal parallel frame along γ. It provides a consistent basis throughout the curve. By proposition 4.2 the vectors $e_\alpha(t) \in T_{\gamma(t)}M$ are orthonormal for all values of t. Figure 8 depicts a frame along a geodesic.

It is common to choose one of the basis vectors to be $\dot{\gamma}(t)$ itself. It is indeed parallel and has unit length if γ is a unit speed geodesic. However, for a general curve $\dot{\gamma}$ is not parallel.

In a parallel frame computations appear more Euclidean.

Exercise 4.6. Any vector field $V(t)$ along γ can be expressed in the orthonormal parallel frame as

$$V(t) = \sum_{\alpha=1}^{n} V_\alpha(t) e_\alpha(t). \tag{50}$$

Show that V is parallel if and only if each $V_\alpha(t)$ is constant. What is the norm of $V(t)$? ○

Parallel frames exist along curves, but not on the whole manifold. It is extremely rare that there would be even one non-zero vector field in a small open subset of the manifold which would be parallel along all curves.

[24] The index of e_α is not a coordinate index, so we try to reduce confusion by using a different kind of letter.

Exercise 4.7. Euclidean geometry is far more rigid than general Riemannian geometry. Give an example of a non-zero vector field on $\mathbb{R}^n$ which is parallel transported along any curve.

Are there n such vectors that could make an orthonormal frame?

Using local coordinates on any Riemannian manifold M makes $U \subset M$ look Euclidean. You can then choose a parallel field of this kind in the local coordinates. Why is it not a parallel field defined in $U \subset M$? ○

Given a basis of a vector space, there is a corresponding dual basis on the dual space. The dual basis of an orthonormal parallel frame is an orthonormal parallel coframe. The same properties of preserved inner products hold with the dual inner product on T_x^*M.

4.5 The variation field of a family of geodesics

We used a family of curves when we studied variations of length. Let us return to studying such a family $\Gamma(t,s)$. Such a family appeared in proposition 2.2. The proposition can be rephrased using our new tools:

> Let $\Gamma\colon [0,1]\times(-\varepsilon,\varepsilon) \to M$ be a smooth map for which $\Gamma(0,s) = p$ and $\Gamma(1,s) = q$ for all s. Denote $\gamma(t) = \Gamma(t,0)$ and $V(t) = \partial_s\Gamma(t,0)$. Then
>
> $$\partial_s \ell(\Gamma(\,\cdot\,,s))|_{s=0} = -\int_0^1 \frac{1}{|\dot\gamma(t)|} \langle V, D_t\dot\gamma\rangle \,\mathrm{d}t. \qquad (51)$$

In this form it is more transparent that the geodesic equation is $D_t\dot\gamma = 0$.

Exercise 4.8. Let us explain the negative sign in (51). Suppose γ is a unit speed curve in $\mathbb{R}^2$. Draw a picture of a non-geodesic curve γ in the plane and draw a nearby shorter curve with the same endpoints. Draw the variation field V and the second derivative $\ddot\gamma$ in a couple of points along the curve. Explain the negative sign in the formula based on this example. ○

The concept of a variation field, depicted in figure 9 already appeared when we minimized length in section 2.5, but we now define it for concreteness, as it will keep reappearing.

Definition 4.5. The variation field $V(t)$ of a family of curves $\Gamma(t,s)$ is defined by $V(t) = \partial_s\Gamma(t,s)|_{s=0}$. It is a vector field along $\gamma(t) = \Gamma(t,0)$.

Previously the curves $\Gamma(\,\cdot\,,s)$ could be anything. Let us now assume that every one of them is a geodesic. We have in fact already used the vector field $V(t) = \partial_s\Gamma(t,s)|_{s=0}$ in our variational calculations. This is a vector

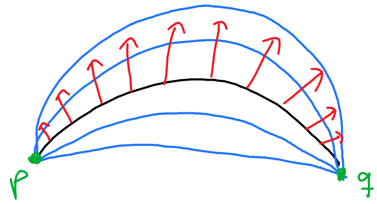


Figure 9: A family of curves joining the points p and q and the corresponding variation field, a vector field along the reference curve. In this family all curves have the same end points, but other families may behave differently in that respect.

field along the reference geodesic $\gamma = \Gamma(\,\cdot\,, 0)$. This field describes first order variations of the curve family, and it is far simpler to study the behaviour of this variation vector field than the whole family of geodesics.

The variation field may be extended to all geodesics in the family by letting $V(t, s) = \partial_s\Gamma(t, s)$. In fact, this is the velocity vector field of the curve $\Gamma(t, \,\cdot\,)$, where now t is fixed. It is important to be able to differentiate with respect to both variables t and s — also covariantly.

Of course one can study variations of any curve family, but more structure emerges when one studies a family of geodesics. Comparison of nearby geodesics is not trivial; geodesics that start nearby can diverge and later converge and maybe even intersect. Nothing similar can happen in Euclidean geometry.

★ *Important exercise* 4.9. Do you have any questions or comments regarding section 4? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ◯

5 Jacobi fields

5.1 Commutators of covariant derivatives

Consider two vector fields X and Y and a scalar field f on M. One can differentiate f with X and Y in two different orders. Their difference is $XYf - YXf = [X, Y]f$. This is the commutator of two vector fields, and it is another vector field; see exercise 3.2.

Consider then three vector fields X, Y, Z on M. Again, one can differentiate Z covariantly with X and Y in the two directions. The difference between the two orders is

$$[\nabla_X, \nabla_Y]Z. \tag{52}$$

Exercise 5.1. Return to the Euclidean connection of exercise 3.1. (This is the Levi-Civita connection of $\mathbb{R}^n$ as a Riemannian manifold.) Show that

$$[\nabla_X, \nabla_Y]Z = \nabla_{[X,Y]}Z. \tag{53}$$

This is exactly what we had for scalar fields on a general Riemannian manifold. ◯

Based on this observation we should ask on a general manifold: What is

$$[\nabla_X, \nabla_Y]Z - \nabla_{[X,Y]}Z? \tag{54}$$

Proposition 5.1. *There is a smooth tensor field R of type*[25] $(1,3)$ *for which*

$$R(X,Y,Z) = [\nabla_X, \nabla_Y]Z - \nabla_{[X,Y]}Z. \tag{55}$$

This tensor is often denoted as $R(X,Y)Z$ instead so that $R(X,Y)$ is seen as a linear map $T_xM \to T_xM$. A tensor field often admits many different ways to view it. This tensor is called the Riemann curvature tensor[26].

Proof of proposition 5.1. It is clear that $R(X,Y)Z$ as given by the formula is linear in the three vector fields. What is not trivial is that it does not depend on any derivatives but only on the values of the three vector fields at a point. This can be verified by calculation. □

Exercise 5.2. Find a local coordinate expression for $[\nabla_X, \nabla_Y]Z - \nabla_{[X,Y]}Z$. If the ith component of the vector $R(X,Y)Z$ is $R^i{}_{jkl}X^jY^kZ^l$, find an expression for the components $R^i{}_{jkl}$ of the Riemann curvature tensor. Second order derivatives of the metric should appear. You may also choose to use first order derivatives of Christoffel symbols. ◯

We will need analogous results for vector fields along curves. First let $\Gamma\colon [0,1]\times(-\varepsilon,\varepsilon) \to M$ be any smooth map. We have the natural vector fields $\partial_s\Gamma$ and $\partial_t\Gamma$ and they are well defined for any values of the two parameters.

Lemma 5.2. *The covariant derivatives of Γ satisfy the commutator relationship*

$$D_t\partial_s\Gamma = D_s\partial_t\Gamma. \tag{56}$$

Exercise 5.3. Prove the lemma. ◯

Lemma 5.3. *If $V(t,s)$ is any smooth vector field depending on the two parameters so that $V(t,s) \in T_{\Gamma(t,s)}M$, then*

$$[D_s, D_t]V = R(\partial_s\Gamma, \partial_t\Gamma)V. \tag{57}$$

The proof of this lemma is a computation similar to that of exercise 5.2.

[25]A multilinear map $T_x^*M \times T_xM \times T_xM \times T_xM \to \mathbb{R}$ can also be seen as a multilinear map $T_xM \times T_xM \times T_xM \to T_xM$. We take this interpretation here.

[26]Much more could be said about the meaning of curvature than is said in these notes. That would be a detour for our purposes.

5.2 Jacobi fields

As mentioned in section 4.5, we will study variation fields of families of geodesics. It is important that all the curves are geodesics; otherwise there is no structure.

Exercise 5.4. Show that for any vector field V along a curve γ there is a family of curves $\Gamma(\,\cdot\,, s)$ so that the variation field of section 4.5 is V. Feel free to work in a single coordinate patch if it helps.[27] ○

When the family consists of geodesics, the variation field has special properties. It will be what we shall call a Jacobi field. We will first define it and then see that it behaves as it should.

Exercise 5.5. A Euclidean geodesic is of the form $\gamma_{x,v}(t) = x+tv$, parametrized by $x, v \in \mathbb{R}^n$. Find all the possible variation fields along a Euclidean geodesic when all curves in the family are geodesics. For any geodesic there should be a $2n$-dimensional space of such fields along it. ○

Definition 5.4. The curvature operator along a geodesic γ is the linear map $T_{\gamma(t)}M \to T_{\gamma(t)}M$ given by

$$R_\gamma V = R(V, \dot{\gamma})\dot{\gamma}. \tag{58}$$

This is in fact a $(1, 1)$-tensor along the geodesic; such concepts can be defined by analogy to our definition of a vector field along a curve.

Lemma 5.5. *We always have* $\langle \dot{\gamma}, R_\gamma V \rangle = 0$.

Proof. This follows from a symmetry property of the Riemann curvature tensor, namely $\langle W, R(X, Y)Z \rangle = - \langle Z, R(X, Y)W \rangle$. □

Lemma 5.6. *The curvature operator along a geodesic from definition 5.4 is symmetric:* $\langle V, R_\gamma W \rangle = \langle R_\gamma V, W \rangle$.

Proof. This follows from a symmetry property of the Riemann curvature tensor, namely $\langle W, R(X, Y)Z \rangle = \langle X, R(W, Z)Y \rangle$. □

The operator R_γ is symmetric, the operator $R(X, Y)$ is antisymmetric.

Definition 5.7. Let γ be a geodesic. A vector field J along γ is called a Jacobi field if it satisfies the Jacobi equation

$$D_t^2 J + R_\gamma J = 0. \tag{59}$$

[27]You have this liberty throughout the course.

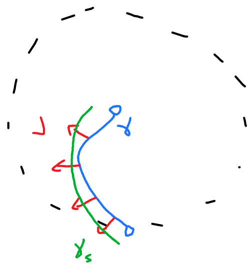


Figure 10: When an end point of a geodesic γ defined on an open interval falls just off the manifold (drawn with dotted boundary), a Jacobi field J might not have a corresponding family of geodesics. A neighboring geodesic γ_s will necessarily fall off the manifold for some amount of time for some sign of s unless γ is normal to the boundary.

Exercise 5.6. Explain why a Jacobi field exists uniquely for all times, given J and $D_t J$ at one time. ◯

Theorem 5.8. *The variation field of a family of geodesics is a Jacobi field. Conversely, for every Jacobi field there is a family of geodesics whose variation field is the Jacobi field.*

Remark 5.9. It is actually important for theorem 5.8 that a family of geodesics is a function $[0,1] \times (-\varepsilon, \varepsilon) \to M$, not $(0,1) \times (-\varepsilon, \varepsilon) \to M$. The open intervals are harmless if the limit points still belong to the manifold, which is always true on a geodesically complete manifold. If an endpoint is just outside the manifold, the family of geodesics might fail to exist as some of the geodesics can be forced to "fall off the manifold" as in figure 10. Feel free to assume geodesic completeness in this course when technical issues seem to arise.

⋆ *Important exercise* 5.7. Prove the first half of the theorem as follows: The fact that each $\Gamma(\,\cdot\,, s)$ is a geodesic can be rewritten as $D_t^2\Gamma = 0$. Take D_s of this equation and commute the derivatives using lemmas 5.2 and 5.3. Evaluate at $s = 0$ to get a vector field along $\gamma = \Gamma(\,\cdot\,, 0)$. ◯

Exercise 5.8. To prove the second half, proceed as follows: You are given a Jacobi field $J(t)$ along a geodesic $\gamma(t)$, and you must find a family $\Gamma(t, s)$ with the correct variation field. Let a be a short curve on M satisfying $a(0) = \gamma(0)$ and $\dot{a}(0) = J(0)$. Argue why such an a exists. Let $b(s)$ be any vector field along $a(s)$ so that $D_s b(s)|_{s=0} = D_t J(0)$ and $b(0) = \dot{\gamma}(0)$. Argue why such a b exists. Now let $\Gamma(\,\cdot\,, s)$ be the geodesic starting at $a(s)$ in the direction $b(s)$. (Smoothness of Γ follows from smoothness of the geodesic

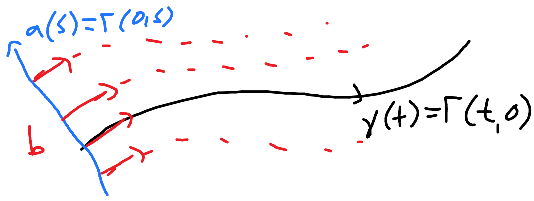


Figure 11: Construction of a family of geodesics $\Gamma(t, s)$ used in exercise 5.8. Through the initial point there are two curves, the reference geodesic $\gamma(t)$ and the new curve $a(s)$. Along that new curve we have a vector field $b(s)$ which we use as the initial data for the family of geodesics (drawn dotted).

flow, to be established later.) Let V be the variation field of this family. Use exercise 5.6 to argue that $J = V$. This procedure is illustrated in figure 11.

○

5.3 Parallel and normal Jacobi fields

Let γ be a geodesic throughout this subsection. There are some special Jacobi fields, and we should understand them and the corresponding families of geodesics.

Reparametrization of geodesics produces more geodesics. Consider the family $\Gamma(t, s) = \gamma(as + (1 + bs)t)$. The parameter a describes the shift in the parametrization and b describes the change in speed. Every geodesic has constant speed, but that speed can vary with s. The corresponding Jacobi field is

$$J(t) = (a + bt)\dot{\gamma}(t). \tag{60}$$

Let us also verify using the Jacobi equation that this is indeed a Jacobi field.

It follows from lemma 5.3 that $R(\dot{\gamma}, \lambda\dot{\gamma}) = 0$ for any $\lambda \in \mathbb{R}$. Therefore $R_\gamma \dot{\gamma} = 0$. The geodesic equation is $D_t\dot{\gamma} = 0$, and so $D_t^2(a+bt)\dot{\gamma}(t) = 0$. Thus the Jacobi equation (59) is satisfied.

Jacobi fields of this form are called parallel Jacobi fields. They are somewhat uninteresting, as they reveal nothing about the behaviour of other geodesics than γ itself.

For a general Jacobi field the inner product $\langle \dot{\gamma}, J \rangle$ measures heuristically how much the varied geodesic gets ahead of $\gamma(t)$. This inner product has a very rigid behaviour:

⋆ *Important exercise* 5.9. Let J be a Jacobi field along a geodesic γ. Show

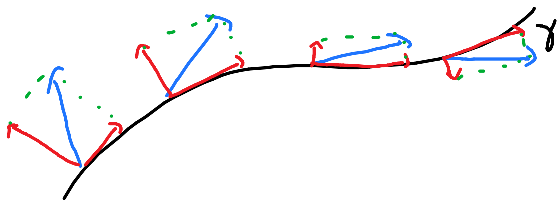


Figure 12: A vector field along a curve γ split orthogonally into parallel and normal components. Both components are vector fields along the curve. It is a special feature of the Jacobi equation that both components of a solution are solutions. That is, the parallel and normal components of the Jacobi equation decouple.

that[28]

$$\langle \dot{\gamma}(t), J(t) \rangle = \langle \dot{\gamma}(0), J(0) \rangle + t \langle \dot{\gamma}(0), D_t J(0) \rangle . \tag{61}$$

The easiest way to do this is to compute the second time derivative of the inner product. ◯

Thus if both J and $D_t J$ are normal to $\dot{\gamma}$ at some point, then they both remain normal at all times. Such Jacobi fields are called normal Jacobi fields.

The parallel component of a Jacobi field is

$$\begin{aligned} J_p(t) &= |\dot{\gamma}|^{-2} \langle \dot{\gamma}(t), J(t) \rangle \dot{\gamma}(t) \\ &= |\dot{\gamma}|^{-2} \langle \dot{\gamma}(0), J(0) \rangle \dot{\gamma}(t) + t |\dot{\gamma}|^{-2} \langle \dot{\gamma}(0), D_t J(0) \rangle \dot{\gamma}(t). \end{aligned} \tag{62}$$

This is indeed a Jacobi field as verified above, and it is clearly parallel to $\dot{\gamma}$ at all times. The normal component is

$$J_n(t) = J(t) - J_p(t). \tag{63}$$

Exercise 5.9 shows that the Jacobi fields J and J_p have the same inner product against $\dot{\gamma}$ at all times. Therefore $J_n(t)$ is indeed normal to $\dot{\gamma}$. As the Jacobi equation is linear, J_n is a Jacobi field.

Figure 12 shows a vector field along a curve split into parallel and normal components. It is not generally true that if a vector field satisfies an equation, then its parallel and normal components will as well. This is a special feature of the Jacobi equation.

The parallel component of a Jacobi field describes how the parametrization of the family of geodesics varies. The normal component describes how

[28] Using $t = 0$ as the reference time is unimportant but convenient.

the geodesics as unparametrized curves or sets vary. If a family of geodesics is reparametrized so that every geodesic has unit speed, then $\langle \dot\gamma, J\rangle$ is constant. The parameters can then be shifted to make this inner product vanish, making the corresponding Jacobi field normal. Therefore it is often reasonable to restrict one's attention to only normal Jacobi fields, as they describe the "true variations" of geodesics.

5.4 Spaces of constant curvature

Let us then take a brief look at Jacobi fields in some example spaces.

A space of constant (sectional) curvature k looks locally like a Euclidean space ($k = 0$), a hyperbolic space ($k < 0$), or a sphere ($k > 0$). On such manifolds the curvature operator along a geodesic is given by

$$R_\gamma V = k(|\dot\gamma|^2 V - \langle V, \dot\gamma\rangle \dot\gamma). \tag{64}$$

The Jacobi equation for a normal Jacobi field along a unit speed geodesic becomes

$$D_t^2 J + kJ = 0. \tag{65}$$

As k is just a constant, this can be solved explicitly.

Let $e_1, \dots, e_{n-1}, \dot\gamma$ be an orthonormal parallel frame along γ. We can write our normal Jacobi field as

$$J(t) = \sum_{\alpha=1}^{n-1} J_\alpha(t) e_\alpha(t). \tag{66}$$

As $D_t e_\alpha = 0$ and the frame is linearly independent at each point, we get the equation

$$J_\alpha''(t) + kJ_\alpha(t) = 0. \tag{67}$$

This is a constant coefficient ODE for a scalar function and can be solved explicitly:

$$J_\alpha(t) = \begin{cases} a\sin(\sqrt{k}\,t) + b\cos(\sqrt{k}\,t) & \text{when } k > 0, \\ at + b & \text{when } k = 0, \\ ae^{\sqrt{-k}\,t} + be^{-\sqrt{-k}\,t} & \text{when } k < 0. \end{cases} \tag{68}$$

The parameters $a, b \in \mathbb{R}$ can of course be different for different indices α.

The flat case ($k = 0$) should be familiar from exercise 5.5. In positive curvature the Jacobi fields oscillate; consider variations of great circles on S^2. In negative curvature the behaviour is exponential; unless very carefully aimed, a Jacobi field grows exponentially when $t \to \pm\infty$.

The basic message is valid even when curvature is not constant: In negative curvature nearby geodesics diverge, in positive curvature they converge. If you want details, look up the Toponogov theorem and the Rauch comparison theorem.

⋆ *Important exercise* 5.10. Do you have any questions or comments regarding section 5? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ◯

6 The exponential map

In this section we will study all geodesics starting from a single point and collect all of them into a single object.

6.1 Definitions

If $x \in M$ and $v \in T_xM$, we denote by $\gamma_{x,v}$ the unique maximal[29] geodesic for which $\gamma_{x,v}(0) = x$ and $\dot{\gamma}_{x,v}(0) = v$. Exercise 2.11 provides the existence and uniqueness of such geodesics.

We would like to define the exponential map at x to be $\exp_x \colon T_xM \to M$,

$$\exp_x(v) = \gamma_{x,v}(1). \tag{69}$$

This can be interpreted geometrically as a "radial wrapping" of the tangent space over the manifold, as illustrated in figure 13.

However, this does not necessarily make sense, as geodesics might not be defined all the way up to time $t = 1$. The definition is sensible as given if all geodesics through x can be parametrized by the whole $\mathbb{R}$. In other cases it needs to be defined on a subset of T_xM as a small enough neighborhood of $0 \in T_xM$ will be mapped nicely to points near x.

A calculation verifies the scaling law $\gamma_{x,\lambda v}(t) = \gamma_{x,v}(\lambda t)$ for any $\lambda \in \mathbb{R}$ for which everything is defined. Therefore when $v \in T_xM$ is not zero, we can write $\exp_x(v) = \gamma_{x,v/|v|}(|v|)$. That is, the norm of the tangent vector gives the travel time along the corresponding unit speed geodesic.

As we can think of T_xM as $\mathbb{R}^n$ upon fixing a basis, it makes sense to ask whether the exponential map is smooth. It is.

Exercise 6.1. Smoothness of the exponential map boils down to a general smoothness result for ODEs (see e.g. [1]):

[29]Defined on as long an interval as possible, containing zero.

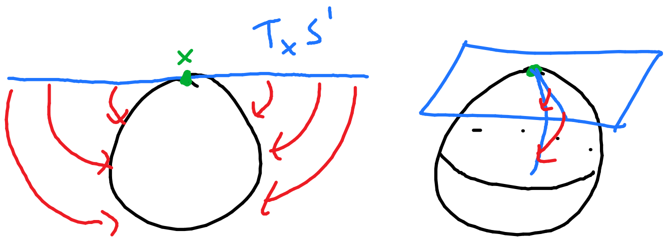


Figure 13: **Left:** The exponential map at a point $x \in S^1$ maps the tangent space at x by wrapping it around the manifold without any stretching. **Right:** The exponential map at a point $x \in S^2$ maps the tangent space at x by wrapping it around the manifold. A radius emanating from the origin of the tangent space is mapped without stretching as a geodesic on the manifold. **In general:** The exponential map may be seen as a radial wrap, akin to gluing a sticker on a curved surface and allowing tears and wrinkles in other directiosn but not radially. If we run out of manifold to wrap over, then the manifold is not geodesically complete and the exponential map is not defined on the whole tangent space.

> Suppose $F\colon \mathbb{R}^N \to \mathbb{R}^N$ is smooth. Let $u(v,t)$ be defined on some open set $\Omega \subset \mathbb{R}^N \times \mathbb{R}$ so that $u(v,\cdot)$ solves the ODE $\partial_t u(v,t) = F(u(v,t))$ and $u(v,0) = v$. Then u is smooth in Ω.

Use this to prove that the exponential map is smooth where it is defined. (Existence and uniqueness of u was proven in exercise 2.11. Smoothness in time was proven in exercise 2.12, but this is not enough.) ◯

There are different versions of the exponential map defined on different spaces. The most immediate example is $\exp\colon TM \to M$ defined by $\exp(v) = \exp_x(v)$ when $v \in T_xM$.

★ *Important exercise* 6.2. Describe all unit speed geodesics through $x \in M$ using the exponential map. ◯

Exercise 6.3. What is the exponential map of the Euclidean space $\mathbb{R}^n$ at a point $x \in \mathbb{R}^n$? ◯

Exercise 6.4. On the smooth manifold $\mathbb{R}$ or a subset thereof a Riemannian metric is just a smooth function $g = g_{11}\colon \mathbb{R} \to (0,\infty)$. The geodesic equation is $\ddot\gamma(t) + \frac{1}{2}g'(\gamma(t))g^{-1}(\gamma(t))\dot\gamma(t)^2 = 0$.

Consider the metric $g(x) = x^{-2}$ on the manifold $M = (0,\infty)$. What is the exponential map $\exp_1\colon T_1M \to M$? ◯

6.2 Normal coordinates

Let us fix $x \in M$. We have learned that there is a neighborhood $\Omega \subset T_xM$ of the origin so that $\exp_x \colon \Omega \to M$ is well defined and smooth. Since it can be differentiated, let us do so.

In general, the differential of a smooth map $f \colon N \to M$ at $y \in N$ is a map $\mathrm{d}f(y) \colon T_yN \to T_{f(y)}M$. Using curves, it can be seen as the unique map for which any smooth curve on N with $\gamma(0) = y$ satisfies $\partial_t(f(\gamma(t))|_{t=0} = \mathrm{d}f(y)\dot{\gamma}(0)$. The curve-based definition is convenient as we may choose any curve with the correct $\dot{\gamma}(0)$.

Exercise 6.5. Given a smooth map $f \colon \mathbb{R}^m \to \mathbb{R}^n$ and a point $y \in \mathbb{R}^m$, show that there exists a unique matrix A for which $\partial_t(f(\gamma(t))|_{t=0} = A\dot{\gamma}(0)$ for any smooth curve γ with $\gamma(0) = y$. What is this A? ◯

The differential of the exponential map at the origin should be a map $\mathrm{d}\exp_x(0) \colon T_0(T_xM) \to T_xM$. But as T_xM is just a vector space (isometric to $\mathbb{R}^n$), we can naturally identify $T_0(T_xM) = T_xM$.

Lemma 6.1. *The differential* $\mathrm{d}\exp_x(0) \colon T_xM \to T_xM$ *of the exponential map is the identity map.*

Proof. We use the curve definition of the differential. Let $v \in T_xM$ be any vector. We need a curve $\gamma \colon (-\varepsilon, \varepsilon) \to T_xM$ with $\gamma(0) = 0$ and $\dot{\gamma}(0) = v$. We choose $\gamma(t) = tv$.

Then we need to know what $\sigma(t) \coloneqq \exp_x(\gamma(t))$ is, because $\mathrm{d}\exp_x(0)v = \dot{\sigma}(0)$. The relationship between the curve γ on the tangent space and σ on the manifold is captured by the following commutative diagram:

$$
\begin{array}{ccc}
 & (-\varepsilon, \varepsilon) & \\
 {\scriptstyle\gamma}\swarrow & & \searrow{\scriptstyle\sigma} \\
T_xM & \xrightarrow[\exp_x]{} & M
\end{array}
\tag{70}
$$

Now $\sigma(t) = \exp_x(tv) = \gamma_{x,tv}(1) = \gamma_{x,v}(t)$. That is, σ coincides with the geodesic $\gamma_{x,v}$. This geodesic satisfies $\dot{\gamma}_{x,v}(0) = v \in T_xM$, so $\dot{\sigma}(0) = v$.

We have thus found that $\mathrm{d}\exp_x(0)v = v$. □

The exponential map maps radial lines in T_xM into geodesics of M. This is not generally true of lines that do not meet the origin.

★ *Important exercise* 6.6. Show that there is a neighborhood $\Omega \subset T_xM$ of the origin and a neighborhood $U \subset M$ of x so that $\exp_x \colon \Omega \to U$ is a diffeomorphism. ◯

If the inverse of the restricted $\exp_x$ of the exercise is called $\varphi\colon U \to \Omega$ and T_xM is identified with $\mathbb{R}^n$ using an orthonormal basis, we have a diffeomorphism $\varphi\colon U \to \varphi(U) \subset \mathbb{R}^n$. In light of remark 1.2 this means that φ is a coordinate chart. These coordinates are called the geodesic normal coordinates or Gaussian normal coordinates or just normal coordinates at x.

Exercise 6.7. Given a point x on a Riemannian manifold, how unique are the normal coordinates at it? ○

Exercise 6.8. Study the geodesic equation (31) in the normal coordinates at x. Consider a geodesic passing through x with velocity $v \in T_xM$. Show that $\Gamma^i{}_{jk}v^jv^k = 0$ at x. Use this information to conclude that $\Gamma^i{}_{jk} = 0$ at x.

In terms of the pseudoforce description of Christoffel symbols, this means that the system of coordinates can be chosen to be inertial (no Christoffel symbol, no pseudoforce) at a single point. The normal coordinates do precisely this, but the symbol cannot be typically made vanish in an open set. ○

6.3 Differential of the exponential map

We saw in lemma 6.1 that the differential of the exponential map $\exp_x$ is the identity map on T_xM. But it is smooth everywhere, so what is the derivative elsewhere?

Consider $0 \neq v \in T_xM$ so that $\exp_x(v)$ is defined. We would like to differentiate $\exp_x$ at v in the direction of any $w \in T_xM$. Therefore we study $\exp_x(v + sw)$ for some parameter $s \in (-\varepsilon, \varepsilon)$.

This gives rise to a family of geodesics defined by $\Gamma(t, s) = \exp_x(t(v+sw))$. The derivative of $\exp_x$ at v in the direction w is

$$\mathrm{d}\exp_x(v)w = \partial_s \exp_x(v + sw) = \partial_s\Gamma(1, s)|_{s=0}. \tag{71}$$

Let us denote $J_w(t) = \partial_s\Gamma(t, 0)$. This is a Jacobi field along $\gamma_{x,v}$. The derivative is the value of this Jacobi field at $t = 1$.

Exercise 6.9. Let us find the initial conditions of the Jacobi field. Verify that $\Gamma(0, s) = x$ and $\partial_t\Gamma(t, s)|_{t=0} = v + sw$ for all s. Show that $J_w(0) = 0$ and $D_tJ_w(0) = w$. ○

We have found that $\mathrm{d}\exp_x(v)$ maps a vector w into the value of a Jacobi field along the geodesic $\gamma_{x,v}$ at $t = 1$ with initial conditions $J_w(0) = 0$ and $D_tJ_w(0) = w$. One can therefore reasonably say that Jacobi fields vanishing at x are the derivative of $\exp_x$.

Exercise 6.10. This description is in fact valid for $v = 0$ as well — a constant curve is a geodesic. Use this description in terms of Jacobi fields to find the differential of the exponential map at the origin. ○

The derivatives have an orthogonality property named after Gauss:

Theorem 6.2 (The Gauss lemma)**.** *Take any* $v, w \in T_xM$ *so that* $\exp_x(v)$ *is defined. Then*

$$\langle \mathrm{d}\exp_x(v)v, \mathrm{d}\exp_x(v)w \rangle = \langle v, w \rangle . \tag{72}$$

Observe that the first inner product is on $T_{\exp_x(v)}M$ and the second one on T_xM. Also notice that one of the two compared vectors has to be the direction of the corresponding geodesic.

Proof. The differential of the exponential is given by Jacobi fields. We have $\mathrm{d}\exp_x(v)v = J_1(1)$ for the Jacobi field J_1 along $\gamma_{x,v}$ with the initial conditions $J_1(0) = 0$ and $D_tJ_1(0) = v$. But this Jacobi field is just $J_1(t) = t\dot{\gamma}_{x,v}(t)$. (Recall that this is a Jacobi field with the correct initial condition and that solutions to the Jacobi equation are unique.) Therefore $\mathrm{d}\exp_x(v)v = \dot{\gamma}_{x,v}(1)$.

Similarly, $\mathrm{d}\exp_x(v)w = J_2(1)$ for the Jacobi field J_2 along $\gamma_{x,v}$ with the initial conditions $J_2(0) = 0$ and $D_tJ_2(0) = w$. Exercise 5.9 gives

$$\langle \mathrm{d}\exp_x(v)v, \mathrm{d}\exp_x(v)w \rangle = \langle \dot{\gamma}_{x,v}(1), J_2(1) \rangle = \langle v, J_2(0) \rangle + 1 \langle v, D_tJ_2(0) \rangle . \tag{73}$$

Using the initial conditions of J_2 gives the claim. □

There is a more geometric version of the lemma, but that requires some setting up. We will do that next and conclude the section with the other version.

Remark 6.3. Take any non-zero $v \in T_xM$ and denote the corresponding unit vector by $\hat{v} = v/|v|$. We can complete $\{\hat{v}\}$ into an orthonormal basis $\{e_1, e_2, \dots, e_{n-1}, e_n = \hat{v}\}$ of T_xM. When we parallel transport these vectors along $\gamma_{x,v}$, we get an orthonormal parallel frame along this geodesic. The differential $\mathrm{d}\exp_x(v)$ of the exponential maps from $T_{\gamma_{x,v}(0)}M$ to $T_{\gamma_{x,v}(1)}M$. Our frame gives a basis for both spaces. Therefore in this frame we can write $\mathrm{d}\exp_x(v)$ as a matrix. Let us write it in block form, separating the last component from the $n-1$ first ones:

$$\mathrm{d}\exp_x(v) = \begin{pmatrix} A & b \\ c^T & d \end{pmatrix}, \tag{74}$$

where A is an $(n-1)\times(n-1)$ matrix, b and c are column vectors of dimension $n-1$, and $d \in \mathbb{R}$.

Exercise 6.11. Use the results obtained so far to argue that

- $b = 0$,

- $d = |v|$,
- $c = 0$, and
- A is given by values of normal Jacobi fields along $\gamma_{x,v}$ that vanish at $t = 0$.

No new proofs should be required here, just recollection and perhaps recontextualization of what has already been done. ○

6.4 Submanifolds

When it comes to submanifolds, geometric intuition serves well for basic concepts and we will not need to go much beyond that. We need to formalize a couple of concepts, but we will not attempt to build a complete theory or give all the details.

A subset $N \subset M$ is submanifold of dimension $k < n$ if near any point $x \in M$ in local coordinates it is a smooth k-dimensional surface in $\mathbb{R}^n$ in the usual sense. A k-dimensional surface $\Sigma \subset \mathbb{R}^n$ can be defined, for example, as the image of a smooth map $\Omega \to \mathbb{R}^n$ from an open $\Omega \subset \mathbb{R}^k$ with an everywhere injective differential. An alternative way is to require that Σ is a level set of a function of a smooth function $\mathbb{R}^n \to \mathbb{R}^{n-k}$ with an everywhere surjective differential. These definitions can be rephrased to work directly on manifolds as well, being careful to work locally.

An important property is that a k-dimensional submanifold $N \subset M$ is also a manifold in its own right. It also inherits a Riemannian structure from the ambient space M.

For any $x \in N \subset M$ the tangent space of N is a subspace of the tangent space of M. That is, $T_xN \subset T_xM$. There is a curve-based way to define this linear subspace: T_xN consists of the velocities $\dot{\gamma}(0)$ of curves $\gamma \colon I \to N \subset M$ for which $\gamma(0) = x$. That is, T_xN consists of velocities of of curves staying in N.

A vector $v \in T_xM$ is said to be normal to a submanifold $N \subset M$ containing x if $\langle v, w \rangle = 0$ for all $w \in T_xN$. A basic argument in linear algebra shows that if N has dimension $n - 1$, then there is a unique unit normal vector to N at x up to sign. One can locally define a smooth normal vector field on N. We can say that a curve γ meets N orthogonally if at the intersection point $\dot{\gamma}$ is normal to N.

6.5 Spheres

The geodesic sphere of radius $r > 0$ centered at $x \in M$ is the set

$$\{\exp_x(v); v \in T_xM, |v| = r\}. \tag{75}$$

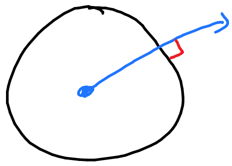


Figure 14: A geodesic starting at the center of a geodesic sphere meets the sphere normally.

This is the image of the sphere $S(0,r) \subset T_xM$ under the exponential map.

The metric sphere of radius $r > 0$ centered $x \in M$ is the set

$$\{y \in M; d(x,y) = r\}. \tag{76}$$

This is the set of points at distance r from x.

These surfaces are closely related as we will soon see[30]. Notice that the geodesic sphere is the image of a smooth $(n-1)$-dimensional surface (a sphere of the tangent space) under a smooth map. Therefore it is smooth at least when $\mathrm{d}\exp_x$ is bijective. This happens at least near the origin by exercise 6.6.

Theorem 6.4 (The Gauss lemma for spheres; see figure 14)**.** *Suppose that the geodesic sphere of radius $|v|$ centered at $x \in M$ is a smooth submanifold near $\exp_x(v)$. Then the geodesic $\gamma_{x,v}$ is normal to the geodesic sphere.*

Proof. Let us take curves staying on the geodesic sphere. These are best described as $\alpha(t) = \exp_x(\sigma(t))$, where $\sigma\colon (-\varepsilon,\varepsilon) \to S(0,|v|) \subset T_xM$ is a smooth curve with $\sigma(0) = v$. Since σ stays on the sphere, we have $0 = \partial_t |\sigma(t)|^2 = 2\langle\sigma(t),\dot\sigma(t)\rangle$ and so $\dot\sigma(0)$ is orthogonal to v. A tangent vector to the geodesic sphere is then $\dot\alpha(0) = \mathrm{d}\exp_x(v)\dot\sigma(0)$, and by theorem 6.2 this is orthogonal to $\dot\gamma_{x,v}(1)$. □

★ *Important exercise* 6.12. Do you have any questions or comments regarding section 6? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ◯

[30]For a quick example, consider the sphere S^2. The metric and the geodesic sphere are exactly the same thing when the radius r is at most π. When $r > \pi$, the metric sphere is empty but the geodesic one is not.

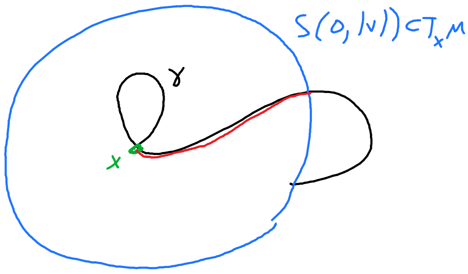


Figure 15: A curve γ on the tangent space, starting at the origin (which is the point x through the local identification through the exponential map) and ending at the sphere of radius $|v|$. If it loops back to the origin and exits the ball, it is enough to show that the relevant segment has at least length $|v|$.

7 Minimization of length

7.1 Short geodesics minimize length

We are now ready to see why geodesics minimize length. Before stating the theorem, we will need to recall the length of a geodesic.

★ *Important exercise* 7.1. Show that the length of the geodesic $\gamma_{x,v}\colon [0,1] \to M$ is $|v|$ whenever the geodesic is defined on the whole interval. ○

Theorem 7.1. *Let $x \in M$ and let $r > 0$ be such that $\exp_x\colon B(0,r) \to U \subset M$ is a diffeomorphism. Then for any $v \in B(0,r) \subset T_xM$ the distance between the endpoints of the corresponding geodesic is*

$$d(x, \exp_x(v)) = |v| \,. \tag{77}$$

In fact, $\gamma_{x,v}|_{[0,1]}$ is the unique shortest curve between its endpoints.

Proof. The result is clear if $v = 0$ so we assume $v \neq 0$. We will show that any curve from x to the geodesic sphere of radius $|v|$ centered at x has at least length $|v|$. Every curve from x to $\exp_x(v)$ will have to meet this sphere. It is enough to show that the segment of the curve until the first intersection with this sphere has at least length r.

We may also assume that the curve we compare to does not meet x again after $t = 0$. Otherwise we could take the segment from a later intersection point to get an even shorter curve. The relevant segment is depicted in figure 15.

That is, we use a segment of the arbitrary curve and show that it has length r or more, whence the original curve will have at least this length.

So, let $\gamma\colon [0,1] \to B(0,r) \subset T_xM$ be a smooth curve with $\gamma(0) = 0$ and $|\gamma(1)| = |v|$. Then $\sigma = \exp_x \circ \gamma$ (cf. diagram (70)) is a curve on M from x to the geodesic sphere of radius $|v|$.

By exercise 7.1 we have $|v| = \ell(\gamma_{x,v}|_{[0,1]})$. On the other hand,

$$
\begin{aligned}
|v| &= |\gamma(1)| \\
&\overset{(a)}{=} \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} |\gamma(t)| \, \mathrm{d}t \\
&\overset{(b)}{=} \int_0^1 |\gamma(t)|^{-1} \langle \gamma(t), \dot\gamma(t) \rangle \, \mathrm{d}t \\
&\overset{(c)}{=} \int_0^1 |\gamma(t)|^{-1} \langle \mathrm{d}\exp_x(\gamma(t))\gamma(t), \mathrm{d}\exp_x(\gamma(t))\dot\gamma(t) \rangle \, \mathrm{d}t \\
&\overset{(d)}{\leq} \int_0^1 |\gamma(t)|^{-1} |\mathrm{d}\exp_x(\gamma(t))\gamma(t)| \, |\mathrm{d}\exp_x(\gamma(t))\dot\gamma(t)| \, \mathrm{d}t \\
&\overset{(e)}{=} \int_0^1 |\mathrm{d}\exp_x(\gamma(t))\dot\gamma(t)| \, \mathrm{d}t \\
&\overset{(f)}{=} \int_0^1 |\dot\sigma(t)| \, \mathrm{d}t \\
&\overset{(g)}{=} \ell(\sigma).
\end{aligned}
\tag{78}
$$

Justifying each step is an exercise.

Therefore

$$
\ell(\gamma_{x,v}|_{[0,1]}) \leq \ell(\sigma). \tag{79}
$$

Thus the geodesic is indeed a shortest curve.

Let us then show that it is the unique one. If equality holds throughout (78), the vectors $\mathrm{d}\exp_x(\gamma(t))\gamma(t)$ and $\mathrm{d}\exp_x(\gamma(t))\dot\gamma(t)$ must be parallel[31] at all times. By exercise 7.4 this means that $\gamma(t)$ and $\dot\gamma(t)$ are parallel.

As we assumed that $\gamma(t) \neq 0$ for $t > 0$, this implies that $\gamma(t) = h(t)w$ for some increasing smooth surjection $h\colon [0,1] \to [0,1]$ and a constant vector $w \in T_xM$ with $|w| = |v|$. Upon choosing constant speed parametrization — which does not change length — we have $\gamma(t) = tw$.

If $\sigma = \exp_x \circ \gamma$ is a shortest path from x to $\exp_x(v)$, then σ must be of the form $\sigma(t) = \exp_x(tw)$. To get the end point right, we must have $\exp_x(w) = \exp_x(v)$. The exponential map is diffeomorphic in the set we are in, so $w = v$.

Thus any minimizing curve between the same endpoints must indeed coincide with our geodesic up to reparamterization. □

[31] This does not refer to parallel transport here, but to one vector being a scalar multiple of the other.

Exercise 7.2. Let us revisit the topological argument used in the proof. We only wanted to work within the ball $B(0,r)$, so we argued that any curve not staying within it will have to meet the sphere.

Let $\gamma\colon [0,1] \to \mathbb{R}^n$ be continuous with $\gamma(0) = 0$ and $|\gamma(1)| > 1$. Show that $|\gamma(t)| = 1$ for some $t \in (0,1)$. ◯

Exercise 7.3. Justify the named steps in (78). ◯

Exercise 7.4. Assume that $\mathrm{d}\exp_x(v)$ is bijective[32]. Show using the Gauss lemma that $\mathrm{d}\exp_x(v)v$ and $\mathrm{d}\exp_x(v)w$ are parallel (so that one is a scalar multiple of the other) if and only if v and w are parallel. ◯

⋆ *Important exercise* 7.5. Show that every point $x \in M$ has a neighborhood U so that for any $y \in U$ there is a unique shortest curve between x and y and it is a geodesic. ◯

Exercise 7.6. Show that for small enough $r > 0$ the metric sphere coincides with the geodesic sphere. ◯

7.2 Conjugate points

We now have a pretty good understanding of what happens when the exponential map is a diffeomorphism. When we go far enough from the base point, it might stop being diffeomorphic. We now turn to studying that.

Proposition 7.2 (cf. figure 16). *The exponential map* $\exp_x\colon T_xM \to M$ *has a bijective differential at* $v \in T_xM \setminus 0$ *if and only if every non-trivial Jacobi field* J *along* $\gamma_{x,v}$ *that vanishes at* $t = 0$ *is non-zero at* $t = 1$.

Proof. In remark 6.3 we write the differential as a matrix using a parallel orthonormal frame along the geodesic $\gamma_{x,v}$. In exercise 6.11 we saw that this matrix is of the form $\begin{pmatrix} A & 0 \\ 0 & d \end{pmatrix}$ for some $d > 0$. Therefore the linear map $\mathrm{d}\exp_x(v)$ is bijective if and only the matrix A is invertible.

The matrix A was defined so that if a Jacobi field J along the geodesic satisfies $J(0) = 0$ and $D_tJ(0) = w$, then $J(1) = Aw$. Notice that $D_tJ(0) \in T_xM$ and $J(1) \in T_{\exp_x(v)}M$, but the parallel frame gives a way to identify these two vector spaces. The matrix A only fails to be invertible when there is $w \neq 0$ so that $Aw = 0$. This is equivalent with the existence of a Jacobi field J for which $J(0) = 0$, $D_tJ(0) \neq 0$, and $J(1) = 0$.

[32]Without this assumption the claim is false, but we only apply this in the neighborhood where the exponential map is a diffeomorphism.

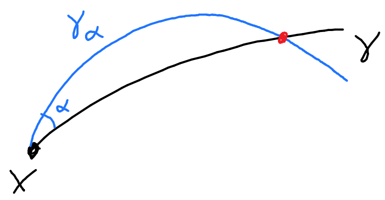


Figure 16: A family γ_α of geodesics depending on an initial angle α with respect to a reference geodesic γ corresponds to a Jacobi field vanishing at $t = 0$. If that Jacobi field vanishes again at $t > 0$, then at that point variations in α are annihilated by $\mathrm{d}\exp_x$. This issue can arise exactly at a conjugate point to x. (It may also help to revisit figure 9 with the added assumption that all curves in the family are geodesics. The points p and q are conjugate because the whole non-trivial family goes through them both.)

By exercise 5.6 a Jacobi field J is uniquely determined by $J(0)$ and $D_t J(0)$. If we require $J(0) = 0$, then the Jacobi field is non-trivial if and only if $D_t J(0) \neq 0$. □

Exercise 7.7. Show that if a non-trivial Jacobi field vanishes at two different points, then it is normal. ◯

Proposition 7.2 inspires us to give a name for the case when a non-trivial Jacobi field vanishes at two points.

Definition 7.3. Let $\gamma \colon I \to M$ be a geodesic and $a, b \in I$. We say that the points $\gamma(a)$ and $\gamma(b)$ are conjugate along γ if there is a non-trivial Jacobi field along γ that vanishes at both a and b.

Just like parallel transport, conjugate points are a concept along a geodesic, not between a pair of points.

Exercise 7.8. Let $\gamma \colon I \to M$ be a geodesic with non-zero speed and $a, b \in I$. Show that the following are equivalent:

- The points $\gamma(a)$ and $\gamma(b)$ are not conjugate along γ.
- The differential $\mathrm{d}\exp_{\gamma(a)}((b-a)\dot\gamma(a))$ is a bijection.
- If a Jacobi field J along γ vanishes at both a and b, it is identically zero.

The last point can be understood as a Jacobi field being uniquely determined by its values at two non-conjugate points. If the two points are conjugate, setting these two values is (somewhat) redundant. ◯

Remark 7.4. Yet another equivalent condition is that the geodesic sphere is smooth at that point. This is very plausible, but it is possible for a smooth map with a non-invertible differential to map a smooth manifold into a smooth manifold. For the exponential map this cannot happen, but studying the details would be a digression.

Exercise 7.9. Give an example of a map $f\colon \mathbb{R}^2 \to \mathbb{R}^3$ for which $f(\mathbb{R}^2)$ is a smooth surface and the derivative matrix of f is injective almost everywhere but not everywhere. ○

7.3 Second variation of length

The way we first found the geodesic equation was to study variations of the length of a curve. We essentially defined geodesics to be critical points of the length functional — with constant speed.

In general there is no guarantee that a critical point is a local minimum. We just showed that short enough geodesics are globally minimal. To study minimality locally, we need to calculate second derivative and see whether it is positive definite.

The second variation is most interesting when the reference curve is a geodesic, a critical point. This will also simplify matters considerably.

We will consider again a family of curves $\Gamma(t,s)$. We now assume that $\Gamma(\,\cdot\,,0)$ is a geodesic and we assume that each $\Gamma(\,\cdot\,,s)$ has constant speed.

Proposition 7.5. *Let* $\Gamma\colon [0,1]\times(-\varepsilon,\varepsilon)\to M$ *be a smooth map so that*

- $\gamma(t)=\Gamma(t,0)$ *is a geodesic,*
- $|\partial_t\Gamma(t,s)|=c_s$*, a constant depending on* s *but not* t*,*
- $\Gamma(0,s)=p$ *for all* s*, and*
- $\Gamma(1,s)=q$ *for all* s*.*

Denoting $V(t)=\partial_s\Gamma(t,s)|_{s=0}$*, we have*

$$\partial_s^2\ell(\Gamma(\,\cdot\,,s))|_{s=0}=\frac{1}{\ell(\gamma)}\int_0^1\left(|D_tV|^2-\langle V,R_\gamma V\rangle\right)\,\mathrm{d}t. \tag{80}$$

Here R_γ is the curvature operator along γ from definition 5.4. Notice that as $\dot\gamma\neq 0$, we have $\partial_t\Gamma(t,s)\neq 0$ everywhere if $\varepsilon>0$ is small enough — therefore constant speed parametrization is legitimate.

Proof. Proposition 2.2 (the first variation) was phrased and proven in local coordinates. Now we will do things invariantly.

Let us denote $\ell(\Gamma(\,\cdot\,,s))=\ell(s)$. First we observe that since each $\Gamma(\,\cdot\,,s)$ has constant speed and is defined on $[0,1]$, we have $\ell(s)=c_s$. In fact, as γ is a geodesic, $\ell'(0)=0$.

To get started, we use the reformulation (51) of the first variation formula. Now that the constant speed condition is satisfied for all s, the formula is valid for all s. We have

$$\ell'(s) = -\int_0^1 \frac{1}{\ell(s)} \langle \partial_s\Gamma, D_t\partial_t\Gamma\rangle \,\mathrm{d}t. \tag{81}$$

We can now simply differentiate under the integral sign and evaluate at $s = 0$ to get

$$\ell''(0) = -\frac{1}{\ell(\gamma)} \int_0^1 \partial_s \langle \partial_s\Gamma, D_t\partial_t\Gamma\rangle |_{s=0} \,\mathrm{d}t. \tag{82}$$

The derivatives ∂_s and D_s are derivatives along the curves $\Gamma(t, \cdot\,)$ for fixed t.

Using exercise 4.4 we get

$$\partial_s \langle \partial_s\Gamma, D_t\partial_t\Gamma\rangle |_{s=0} = \langle D_s\partial_s\Gamma, D_t\partial_t\Gamma\rangle |_{s=0} + \langle \partial_s\Gamma, D_sD_t\partial_t\Gamma\rangle |_{s=0}. \tag{83}$$

The first term vanishes because $D_t\partial_t\Gamma(t, 0) = 0$ — after all, γ is a geodesic. Exercise 7.10 gives that

$$D_sD_t\partial_t\Gamma|_{s=0} = D_t^2V + R_\gamma V. \tag{84}$$

With these ingredients we can simplify our second derivative to

$$\ell''(0) = -\frac{1}{\ell(\gamma)} \int_0^1 \left\langle V, D_t^2V + R_\gamma V\right\rangle \,\mathrm{d}t. \tag{85}$$

Integration by parts in the first term gives the claim since V vanishes at the endpoints. (See exercise 7.11 for details on integration by parts.) □

We will study this formula in more detail in section 8.

★ *Important exercise* 7.10. Commute the derivatives to prove that

$$D_sD_t\partial_t\Gamma = D_t^2\partial_s\Gamma + R(\partial_s\Gamma, \partial_t\Gamma)\partial_t\Gamma. \tag{86}$$

At $s = 0$ this becomes $D_t^2V + R_\gamma V$. ○

Exercise 7.11. Let us justify integration by parts of vector fields. Let V and W be two vector fields along a geodesic $\gamma\colon [a, b] \to M$. Show that

$$\int_a^b \langle V, D_tW\rangle \,\mathrm{d}t = \langle V(b), W(b)\rangle - \langle V(a), W(a)\rangle - \int_a^b \langle D_tV, W\rangle \,\mathrm{d}t. \tag{87}$$

It may help to recall how the integration by parts formula for functions on the real line is proven. ○

Exercise 7.12. Show that it follows from the assumptions of proposition 7.5 that the variation field is normal to the geodesic γ at all times. It can help to show first that $2\partial_t \langle \partial_t \Gamma, \partial_s \Gamma \rangle = \partial_s \langle \partial_t \Gamma, \partial_t \Gamma \rangle$ at $s = 0$ and to recall that $\ell'(0) = 0$. ◯

As was mentioned in section 5.3, only the normal component of the variation field is geometrically meaningful. The parallel component corresponds to reparametrization.

★ *Important exercise* 7.13. Do you have any questions or comments regarding section 7? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ◯

8 The index form

8.1 Second variation of length

Let us denote by $NVF(\gamma)$ the space of normal vector fields along a geodesic $\gamma\colon [a,b] \to M$. Let $NVF_0(\gamma) \subset NVF(\gamma)$ be the subspace of vector fields vanishing at the endpoints. The space $NVF_0(\gamma)$ describes proper first order variations of a geodesic γ with fixed endpoints. Since the first order variation of the length vanishes, the second order variation of length only depends on the first order variation of the curve itself.[33]

Exercise 8.1. Show that if $V \in NVF(\gamma)$, then also $D_t^2 V + R_\gamma V \in NVF(\gamma)$. ◯

We found a formula for the second variation of length in proposition 7.5. Inspired by that, we give a name to the gadget we found.

Definition 8.1. Let $\gamma\colon [a,b] \to M$ be a geodesic. The index form $I = I_\gamma$ of γ is a quadratic form on $NVF(\gamma)$ defined by

$$I(V,W) = \int_a^b \left(\langle D_t V, D_t W \rangle - \langle V, R_\gamma W \rangle \right) \mathrm{d}t. \tag{88}$$

It follows from lemma 5.6 that the index form is symmetric.

Definition 8.2. Let E be a real vector space and $Q\colon E \times E \to \mathbb{R}$ a quadratic form[34]. We say that

[33] Maybe this point is clearer for smooth functions $\mathbb{R} \to \mathbb{R}$. Consider $f = \ell \circ \alpha$. If $f'(0) = 0$, then $f''(0)$ does not depend on $\alpha''(0)$ but only $\alpha'(0)$. In our setting ℓ would be the length functional, α a parametrized family of geodesics, and f the lengths of that family.

[34] That, is Q is a symmetric element of $E^* \otimes E^*$.

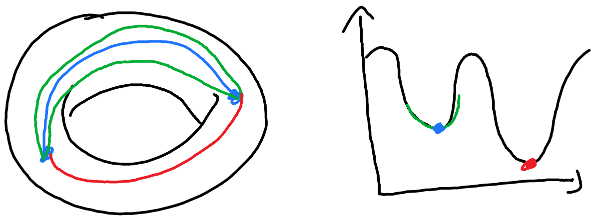


Figure 17: Analogous situations. **Left:** Two points connected by a locally minimizing geodesic. Any nearby geodesic is longer if the index form is positive definite. But there may well be an entirely different, shorter geodesic, so local minimization does not imply global minimization. **Right:** For real valued functions a local minimum is below nearby values but there may well be a different global minimum.

- Q is positive definite if $Q(v, v) > 0$ for all $v \in E \setminus 0$.
- Q is positive semidefinite if $Q(v, v) \geq 0$ for all $v \in E$.
- Q is negative (semi)definite if $-Q$ is positive (semi)definite.
- Q is indefinite if $Q(v, v) > 0$ and $Q(w, w) < 0$ for some $v, w \in E$.

Exercise 8.2. Let $\gamma\colon [a, b] \to M$ be a unit speed geodesic. Show that the second variation of its length corresponding to a family of curves with a normal variation field $V \in NVF(\gamma)$ is $I(V, V)$. You only need to rescale proposition 7.5 to unit speed and a general interval. ◯

One should therefore think of the index form as the Hessian of the length functional. Any geodesic can be made longer by adding wiggles, so the index form cannot be negative definite or semidefinite. All other options are possible as we will see.

Exercise 8.3. Show that if I_γ is not positive semidefinite on $NVF_0(\gamma)$, then γ is not the shortest curve between its endpoints. This together with theorem 7.1 implies that for any x and $v \in T_xM$ there is $\delta > 0$ so that $I_{\gamma_{x,v}|_{[0,\delta]}}$ is positive semidefinite on $NVF_0(\gamma)$. ◯

A local minimum need not be a global one as depicted in figure 17. Even if the index form is positive definite, the geodesic can fail to be minimizing. There can be a curve taking an entirely different route between the two endpoints. No amount of local analysis along a curve can rule this out.

Remark 8.3. Minimization can fail quickly, and the index form will not help predicting or estimating it. For example, let $M \subset \mathbb{R}^2$ be a circle of radius

$\varepsilon > 0$. There is no intrinsic curvature on a one-dimensional manifold, so there are no conjugate points and the index form is always positive definite. The index form is independent of ε. On the other hand, geodesics only minimize length up to distance $\varepsilon\pi$.

8.2 Jacobi fields, conjugate points, and definiteness

Integration by parts (exercise 7.11) reveals a connection between the index form and Jacobi fields.

★ *Important exercise* 8.4. Let $V \in NVF(\gamma)$. Show that the following are equivalent:

1. V is a Jacobi field.
2. $I(V, W) = 0$ for all $W \in NVF_0(\gamma)$.

Why is it important that W vanishes at the endpoints? ◯

Remark 8.4. Exercise 8.4 has an interesting implication if the endpoints of the geodesic are conjugate. Then there is a Jacobi field $J \in NVF_0(\gamma) \setminus \{0\}$, and by the exercise $I(J, J) = 0$. Therefore positive definiteness is impossible in this case. This connection between conjugate points and the definiteness of the index form goes much further as we will see next.

Lemma 8.5. *Let $\gamma\colon [a, b] \to M$ be a geodesic. If there are conjugate points $\gamma(a')$ and $\gamma(b')$ along γ so that $0 < b' - a' < b - a$, then there is $V \in NVF_0(\gamma)$ so that $I(V, V) < 0$.*

Proof. There is a non-trivial Jacobi field along γ satisfying $J(a') = 0 = J(b')$. The piecewise smooth vector field $\bar{J}$ defined by

$$\bar{J}(t) = \begin{cases} J(t), & a' < t < b' \\ 0, & \text{otherwise} \end{cases} \tag{89}$$

describes, roughly, a piecewise geodesic curve with the same length as γ and with corners at a' and b'. Once we cut the corners, we should get a curve shorter than γ, as illustrated in figure 18.

We assume that $a < a'$ and $b' < b$. At least one has to be true, and if the other is replaced by an equality, the analysis we will do can be restricted to the other point. It is enough to find a normal C^1 vector field V with the desired property; see exercise 8.5.

Let us denote $\zeta = D_t J(a')$. We can then parallel transport it as a vector field $\zeta(t)$ with $\zeta(a') = \zeta$. This vector is normal to $\dot{\gamma}$ at all times. Notice that since $J(a') = 0$ but J is not identically zero, $\zeta \neq 0$. For small $\varepsilon > 0$ we define

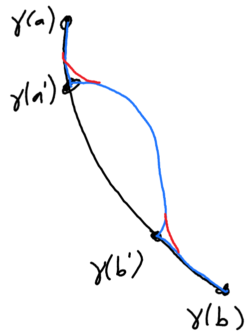


Figure 18: The construction of the variation field V in the proof of lemma 8.5. The first step is the curve corresponding to $\bar{J}$. This is, to leading order, a curve of the same length as the reference curve. We then cut the corners on the scale ε by adding Z and H to make the curve simultaneously both C^1 and shorter than the reference curve. The only contribution to the index form is from these corner regions.

a normal vector field Z along γ as

$$Z(t) = \begin{cases} C\varepsilon^{-1}(|t-a'|-\varepsilon)^2\zeta(t), & |t-a'|<\varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{90}$$

with some positive constant $C > 0$.

Similarly, if $\eta = D_t J(b')$, we define a parallel transport $\eta(t)$ and let[35]

$$H(t) = \begin{cases} -C\varepsilon^{-1}(|t-b'|-\varepsilon)^2\eta(t), & |t-b'|<\varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{91}$$

with the same constant $C > 0$. These two vector fields "cut the corners" as explained above.

We define $V(t) = \bar{J}(t) + Z(t) + H(t)$. As a sum of three normal vector fields it is a normal vector field. With a suitable choice of $C > 0$ this vector field is C^1; see exercise 8.6. Now it remains to show that $I(V, V) < 0$ when

[35]Capital ζ is Z, capital η is H.

$\varepsilon > 0$ is small enough. We have

$$\begin{aligned}
I(V,V) &= \int_a^b \left(|D_tV|^2 - \langle R_\gamma V, V\rangle\right) \mathrm{d}t \\
&= \int_a^b \Big(\left|D_t\bar{J}\right|^2 + 2\left\langle D_t\bar{J}, D_t(Z+H)\right\rangle + |D_t(Z+H)|^2 \\
&\quad - \left\langle R_\gamma\bar{J}, \bar{J}\right\rangle - 2\left\langle R_\gamma\bar{J}, Z+H\right\rangle - \langle R_\gamma(Z+H), Z+H\rangle \Big)\,\mathrm{d}t \\
&= \int_{a'}^{b'} \left(\left|D_t\bar{J}\right|^2 - \left\langle R_\gamma\bar{J}, \bar{J}\right\rangle\right) \mathrm{d}t \\
&\quad + \int_{a'-\varepsilon}^{a'+\varepsilon} \left(2\left\langle D_t\bar{J}, D_tZ\right\rangle + |D_tZ|^2 - 2\left\langle R_\gamma\bar{J}, Z\right\rangle - \langle R_\gamma Z, Z\rangle\right) \mathrm{d}t \\
&\quad + \int_{b'-\varepsilon}^{b'+\varepsilon} \left(2\left\langle D_t\bar{J}, D_tH\right\rangle + |D_tH|^2 - 2\left\langle R_\gamma\bar{J}, H\right\rangle - \langle R_\gamma H, H\rangle\right) \mathrm{d}t.
\end{aligned} \tag{92}$$

If we use exercise 8.4 or remark 8.4 on the geodesic segment $\gamma|_{a',b'}$, we see that

$$\int_{a'}^{b'} \left(\left|D_t\bar{J}\right|^2 - \left\langle R_\gamma\bar{J}, \bar{J}\right\rangle\right) \mathrm{d}t = 0. \tag{93}$$

Since $\bar{J}$ is Lipschitz and vanishes at a' and b', we have $\left|\bar{J}\right| = \mathcal{O}(\varepsilon)$ in the last two integrals of (92). We also have $|Z| = \mathcal{O}(\varepsilon)$ and $|H| = \mathcal{O}(\varepsilon)$. Exercise 8.7 gives the other two integrals which contain only Z and H. As the intervals of integration have length 2ε, we have

$$\begin{aligned}
I(V,V) = 2\int_{a'}^{a'+\varepsilon} \langle D_tJ, D_tZ\rangle \,\mathrm{d}t + 2\int_{b'-\varepsilon}^{b'} \langle D_tJ, D_tH\rangle \,\mathrm{d}t \\
+ \frac{8}{3}C^2 |\zeta|^2 \varepsilon + \frac{8}{3}C^2 |\eta|^2 \varepsilon + \mathcal{O}(\varepsilon^3).
\end{aligned} \tag{94}$$

Let us study the first remaining integral. In it $D_tJ(t) = \zeta(t) + \mathcal{O}(\varepsilon)$. Using exercise 8.7 gives thus

$$2\int_{a'}^{a'+\varepsilon} \langle D_tJ, D_tZ\rangle \,\mathrm{d}t = -4C|\zeta|^2 \varepsilon + \mathcal{O}(\varepsilon^2). \tag{95}$$

The other integral gives a similar negative leading order term.

We have arrived at

$$I(V,V) = -4C|\zeta|^2\varepsilon - 4C|\eta|^2\varepsilon + \frac{8}{3}C^2|\zeta|^2\varepsilon + \frac{8}{3}C^2|\eta|^2\varepsilon + \mathcal{O}(\varepsilon^2). \tag{96}$$

With our $C > 0$ we have $4C > 8C^2/3$, whence

$$I(V,V) = -|\zeta|^2\left(4C - \frac{8}{3}C^2\right)\varepsilon - |\eta|^2\left(4C - \frac{8}{3}C^2\right)\varepsilon + \mathcal{O}(\varepsilon^2) \tag{97}$$

is indeed negative for $\varepsilon > 0$ small enough. □

Exercise 8.5. Polish the proof by showing that if there is a compactly supported normal vector field V with C^1 regularity so that $I_\gamma(V,V) < 0$, then there is a smooth one as well. ○

Exercise 8.6. Choose $C > 0$ so that the vector field $V(t)$ of the proof is actually C^1. What is the value of the constant and why is the resulting vector field C^1? Verify that $4C > 8C^2/3$. ○

Exercise 8.7. Show that

$$\int_{a'-\varepsilon}^{a+\varepsilon} |D_t Z|^2 \, \mathrm{d}t = \frac{8}{3}C^2 |\zeta|^2 \varepsilon \tag{98}$$

and

$$\int_{a'}^{a'+\varepsilon} \langle \zeta(t), D_t Z(t)\rangle \, \mathrm{d}t = -2C |\zeta|^2 \varepsilon. \tag{99}$$

Similar formulas hold for H with the norm of η. ○

Lemma 8.6. *Let $\gamma\colon [a,b] \to M$ be a geodesic. If there are no conjugate points along γ, then $I(V,V) > 0$ for all $V \in NVF_0(\gamma) \setminus 0$.*

Proof. Let $\zeta_1, \dots, \zeta_{n-1}, \dot\gamma(a)$ be an orthonormal basis of $T_{\gamma(a)}M$. We can extend these into an orthonormal parallel frame with the transported vectors $\zeta_\alpha(t)$. For $\alpha \in \{1, \dots, n-1\}$ let J_α be the Jacobi field with $J_\alpha(a) = 0$ and $D_t J_\alpha(a) = \zeta_\alpha$. Near the initial point we have $J_\alpha(t) = t\zeta_\alpha(t) + \mathcal{O}(t^2)$.

When $t_0 \in (a,b]$, the vectors $J_\alpha(t_0)$ are linearly independent. To see this, suppose that there are coefficients λ_α so that

$$\sum_\alpha \lambda_\alpha J_\alpha(t_0) = 0. \tag{100}$$

Then the $J = \sum_\alpha \lambda_\alpha J_\alpha$ is a Jacobi field which vanishes at $t = a$ and $t = t_0$. As there are no conjugate points by assumption, J must vanish identically. Therefore

$$0 = D_t J(a) = \sum_\alpha \lambda_\alpha \zeta_\alpha. \tag{101}$$

The vectors ζ_α are linearly independent, so every λ_α vanishes. This proves the linear independence.[36] The Jacobi fields $J_\alpha(t)$ therefore constitute a basis for the orthogonal complement of $\dot\gamma(t)$ in $T_{\gamma(t)}M$ for any $t > a$.

We can thus write our normal vector field $V \in NVF_0(\gamma)$ in this basis:

$$V(t) = \sum_\alpha V_\alpha(t) J_\alpha(t). \tag{102}$$

Here $V_\alpha(t)$ are real-valued functions. As $V(a) = 0$, the functions $V_\alpha(t)$ are smooth up to $t = a$; see exercise 8.9.

Let us denote

$$A(t) = \sum_\alpha \dot V_\alpha(t) J_\alpha(t) \tag{103}$$

and

$$B(t) = \sum_\alpha V_\alpha(t) D_t J_\alpha(t). \tag{104}$$

With this notation we have $D_t V = A + B$.

Let us compute $\partial_t \langle V, B\rangle$ — this turns out to simplify matters greatly. At first we get

$$\partial_t \langle V, B\rangle = \langle D_t V, B\rangle + \langle V, D_t B\rangle . \tag{105}$$

We already know that $D_t V = A + B$, so let us find $D_t B$. The Leibniz rule and the Jacobi equation give

$$\begin{aligned} D_t B &= \sum_\alpha \left[\dot V_\alpha(t) D_t J_\alpha(t) + V_\alpha(t) D_t^2 J_\alpha(t)\right] \\ &= \sum_\alpha \left[\dot V_\alpha(t) D_t J_\alpha(t) - V_\alpha(t) R_\gamma J_\alpha(t)\right] \\ &= -R_\gamma V(t) + \sum_\alpha \dot V_\alpha(t) D_t J_\alpha(t). \end{aligned} \tag{106}$$

[36] One could say that the vectors $J_\alpha(t)$ form a "Jacobi frame" along γ. This provides a valid basis in every tangent space due to the lack of conjugate points.

Using this with exercise 8.8 leads to

$$\begin{aligned}
\langle V, D_t B\rangle &= -\langle V, R_\gamma V\rangle + \sum_\alpha \left\langle V, \dot{V}_\alpha D_t J_\alpha \right\rangle \\
&= -\langle V, R_\gamma V\rangle + \sum_{\alpha,\beta} \left\langle V_\beta J_\beta, \dot{V}_\alpha D_t J_\alpha \right\rangle \\
&= -\langle V, R_\gamma V\rangle + \sum_{\alpha,\beta} V_\beta \dot{V}_\alpha \langle J_\beta, D_t J_\alpha\rangle \\
&= -\langle V, R_\gamma V\rangle + \sum_{\alpha,\beta} V_\beta \dot{V}_\alpha \langle D_t J_\beta, J_\alpha\rangle \\
&= -\langle V, R_\gamma V\rangle + \sum_{\alpha,\beta} \left\langle V_\beta D_t J_\beta, \dot{V}_\alpha J_\alpha \right\rangle \\
&= -\langle V, R_\gamma V\rangle + \langle B, A\rangle .
\end{aligned} \tag{107}$$

Putting all of this together gives

$$\begin{aligned}
\partial_t \langle V, B\rangle &= \langle A + B, B\rangle - \langle V, R_\gamma V\rangle + \langle B, A\rangle \\
&= |D_t V|^2 - |A|^2 - \langle V, R_\gamma V\rangle .
\end{aligned} \tag{108}$$

Now we can finally turn to the index form. With these preparations it becomes easy to analyze.

Because $V(a) = 0 = V(b)$, we have

$$\begin{aligned}
I(V, V) &= \int_a^b \left(|D_t V|^2 - \langle R_\gamma V, V\rangle\right) \mathrm{d}t \\
&= \int_a^b \left(\partial_t \langle V, B\rangle + |A|^2\right) \mathrm{d}t \\
&= \int_a^b |A|^2 \,\mathrm{d}t \geq 0.
\end{aligned} \tag{109}$$

If equality holds, then $A = 0$, which means that $\dot{V}_\alpha = 0$ and thus each coefficient $V_\alpha(t)$ is constant. But every $V_\alpha(t)$ vanishes at $t = b$, so $V_\alpha = 0$. This means that $V = 0$, so $I(V, V) = 0$ is only possible when $V = 0$. □

Remark 8.7. If there are conjugate points, the "Jacobi frame" used above only fails to be a frame at conjugate points. This makes one think that perhaps the Hessian only has very few negative eigenvalues and that they should correspond to conjugate points. This is indeed true but is beyond the scope of this course. The maximal dimension of a subspace of $NVF_0(\gamma)$ on which the index form is negative definite is called the index of the geodesic. This index is finite and is indeed equal to the number of interior conjugate points, as long as one counts with multiplicity.

★ *Important exercise* 8.8. Let J_1 and J_2 be two Jacobi fields along the same geodesic. Show that

$$\partial_t \left(\langle D_t J_1, J_2 \rangle - \langle J_1, D_t J_2 \rangle \right) = 0. \tag{110}$$

Conclude that if J_1 and J_2 both vanish at the same point, then $\langle D_t J_1, J_2 \rangle = \langle J_1, D_t J_2 \rangle$ at all times. ◯

Exercise 8.9. Little Bézout's theorem concerns polynomials: If r is a root of a polynomial p, then $p(x) = (x - r)q(x)$ for some polynomial q.

Show that a similar result holds for smooth functions. That is, show that if $f \in C^\infty(\mathbb{R})$ and $f(0) = 0$, then $f(t) = tg(t)$ for some smooth function g. A neat way to do this is to compute $\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} f(tx)\,\mathrm{d}t$ in two ways. This gives an explicit formula for g as an integral, and smoothness is far easier to see than by studying $g(t) = f(t)/t$. ◯

Theorem 8.8. *Let $\gamma \colon [a, b] \to M$ be a geodesic. Consider the index form I_γ along it on NVF_0.*

1. *If there are no conjugate points along γ, then it is positive definite.*
2. *If the endpoints are conjugate but there are no other conjugate points, then it is positive semidefinite.*
3. *If an interior point is conjugate to another point, then it is indefinite.*

Proof. This follows from remark 8.4, lemma 8.5, and lemma 8.6. Recall that there are always vector fields $V \in NVF_0(\gamma)$ with positive index form. □

8.3 The index form in constant curvature

For a somewhat concrete example, let us take another look at space of constant curvature. See section 5.4. In this setting the index form on normal vector fields takes the form

$$I(V, W) = \int_a^b \left(\langle D_t V, D_t W \rangle - k \langle V, W \rangle \right) \mathrm{d}t. \tag{111}$$

When $k \leq 0$, this is positive definite, and more strongly so when $k < 0$.

Indeed, if one studies the forms of Jacobi fields in constant curvature as given in section 5.4, one sees that there are no conjugate points when $k \leq 0$. Theorem 8.8 predicts exactly this behaviour.

If $k > 0$, definiteness depends on length. As we saw in exercise 8.3, the index form is positive semidefinite (and in fact positive definite) when the geodesic is short enough. Conjugate points in constant curvature $k > 0$ are distance $\pi/\sqrt{k}$ apart. If the geodesic is longer, then the index form becomes indefinite.

One way to interpret this is to consider the Poincaré inequality

$$\int_a^b |V|^2 \, \mathrm{d}t \leq C \int_a^b |D_t V|^2 \, \mathrm{d}t, \tag{112}$$

valid for all $V \in NVF_0(\gamma)$. If C is small enough, this ensures that the index form is positive. The constant C becomes bigger when the interval $[a, b]$ gets longer. At $b - a = \pi/\sqrt{k}$ the optimal Poincaré constant C becomes exactly $1/k$, making the index form barely positive semidefinite.

★ *Important exercise* 8.10. Do you have any questions or comments regarding section 8? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ◯

9 The tangent bundle

9.1 The tangent bundle as a manifold

Previously, we have considered the tangent bundle as the disjoint union of tangent spaces:

$$TM = \coprod_{x \in M} T_x M. \tag{113}$$

While this is correct as a set, there is more structure. The tangent bundle is a manifold.

It is often convenient to write a tangent vector as a pair (x, v), where $x \in M$ and $v \in T_x M$. The tangent bundle is the set of all such pairs. Sometimes the base point x is left implicit. When $U \subset M$ is open, we denote $TU = \{(x, v) \in TM; x \in U\}$.

Consider an open subset $U \subset M$ and a diffeomorphism $\varphi\colon U \to \varphi(U) \subset \mathbb{R}^n$. The coordinate maps of a coordinate chart are often denoted by x^i, so that each $x^i\colon U \to \mathbb{R}$ is a smooth function and its differential is the familiar basis covector field $\mathrm{d}x^i$. That is, at any point x the differential $\mathrm{d}x^i\colon T_x M \to \mathbb{R}$ is a linear map.

Combining the components together, we have the map $\mathrm{d}\varphi(x)\colon T_x M \to \mathbb{R}^n$ given by

$$\mathrm{d}\varphi(x)v = (\mathrm{d}x^1(x)v, \mathrm{d}x^2(x)v, \ldots, \mathrm{d}x^n(x)v) = (v^1, v^2, \ldots, v^n). \tag{114}$$

This map is a linear bijection since $\mathrm{d}\varphi(x)v$ expresses v in a basis.

We have a map on each tangent space, and we can promote it to a map $\mathrm{d}\varphi$ on the whole bundle. We define $\mathrm{d}\varphi\colon TU \to \mathbb{R}^n \times \mathbb{R}^n$ so that

$$\mathrm{d}\varphi(x, v) = (\varphi(x), \mathrm{d}\varphi(x)v). \tag{115}$$

The base point x is mapped with the coordinate map φ itself, whereas the tangent vector v is mapped by its differential $\mathrm{d}\varphi(x)$.

We want to use $\mathrm{d}\varphi$ as a coordinate chart on TM to make it into a manifold. This chart makes TU look like the product $U \times \mathbb{R}^n$. However, the tangent bundle is not always a product globally although; this only works for open sets U diffeomorphic to an open Euclidean set.

Exercise 9.1. Let $(U_\alpha, \varphi_\alpha)_{\alpha\in A}$ be a smooth atlas of M. We defined a topology on TM by saying that $V \subset TM$ is open if and only if $\mathrm{d}\varphi_\alpha(TU_\alpha \cap V) \subset \mathbb{R}^{2n}$ is open for all $\alpha \in A$. Show that this is a topology. ◯

Exercise 9.2. A chart $\varphi_\alpha \colon U_\alpha \to \varphi_\alpha(U_\alpha) \subset \mathbb{R}^n$ induces a map $\mathrm{d}\varphi_\alpha \colon TU_\alpha \to V_\alpha \subset \mathbb{R}^{2n}$ as described above. Consider two of these, $\alpha = 1, 2$. Given the diffeomorphic transition function ψ between φ_1 and φ_2, write down the transition function Ψ between $\mathrm{d}\varphi_1$ and $\mathrm{d}\varphi_2$. Prove that it is a diffeomorphism.

This shows that a smooth atlas $(U_\alpha, \varphi_\alpha)_{\alpha\in A}$ on M induces a smooth atlas $(TU_\alpha, \mathrm{d}\varphi_\alpha)_{\alpha\in A}$ on TM. In particular, TM is a smooth manifold of dimension $2n$. ◯

Exercise 9.3. Is the smooth atlas induced by a maximal smooth atlas maximal? ◯

A chart $\varphi \colon U \to \mathbb{R}^n$ gives local coordinates on M. The map $\mathrm{d}\varphi \colon TU \to \mathbb{R}^{2n}$ gives the induced coordinates on TM.

⋆ *Important exercise* 9.4. There is a canonical projection $\pi \colon TM \to M$ given by $\pi(x, v) = x$. Show that this is a smooth map between the smooth manifolds TM and M. ◯

⋆ *Important exercise* 9.5. Draw a picture of the tangent bundle so that M is horizontal and the fibers are vertical. Indicate M, a point x, and a fiber T_xM on it. It is important to draw the picture in this orientation. ◯

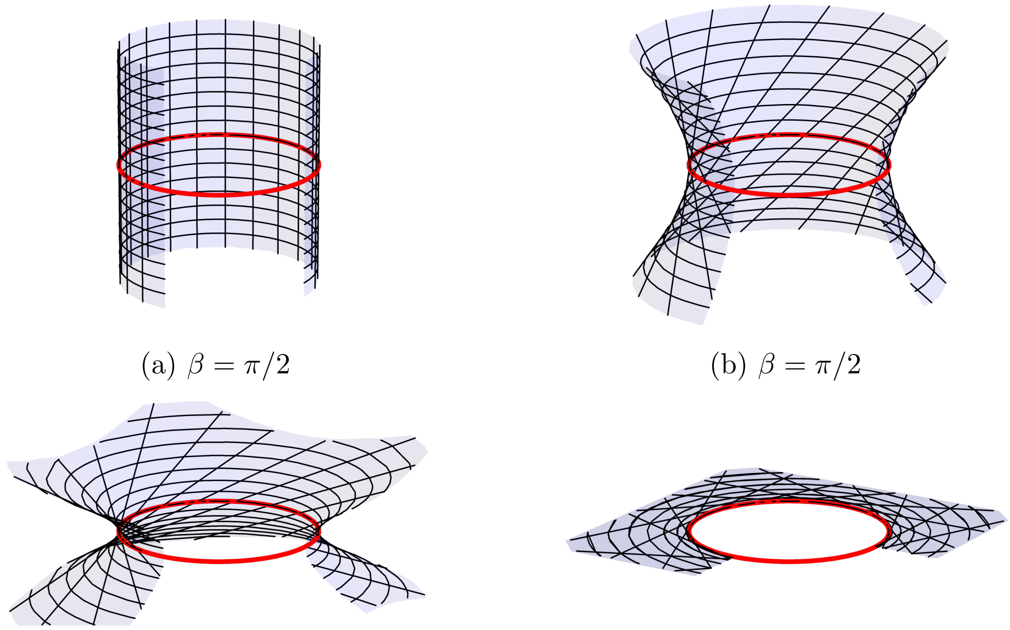
One way to visualize tangent bundles, especially by comparison to our general picture of bundles, is given in figure 19.

9.2 Tensor bundles

Fix some local coordinates on $U \subset M$. We saw above that the linear maps $\mathrm{d}x^i \colon T_xM \to \mathbb{R}$ produced a map $T_xM \to \mathbb{R}^n$ and thus local coordinates $TU \to \mathbb{R}^{2n}$.

Recall that T_xM is the dual of T_x^*M. We can use the linear maps $\partial_i \colon T_x^*M \to \mathbb{R}$ to produce a map $T_x^*M \to \mathbb{R}^n$ and thus coordinates on T^*U. A similar construction turns T^*M into a smooth manifold of dimension $2n$.

Remark 9.1. The musical isomorphisms of a Riemannian manifold are diffeomorphisms between TM and T^*M.



(a) $\beta = \pi/2$

(b) $\beta = \pi/2$

(c) $\beta = \pi/8$

(d) $\beta = 0$

Figure 19: The tangent bundle of the circle S^1 can be visualised in many ways. First of all, abstractly $TS^1 = S^1 \times \mathbb{R}$. We have used here the maps $f_\beta(\alpha, t) = (\cos\alpha, \sin\alpha, 0) + t(\sin\alpha, -\cos\alpha, 0)\cos\beta + t(0, 0, 1)\sin\beta$ from TS^1 to $\mathbb{R}^3$, parametrized by an angle β. The value $t = 0$ corresponds to the base point and it is mapped to the same circle for all β, but the realization of the fibers varies. If $\beta = 0$, the tangent spaces are actually visually tangent to the circle. The map f_β is not injective in this case and the union of the tangent lines is not a manifold. If $\beta > 0$, the lines are slightly twisted from the plane of the circle so that they do not intersect. Now f_β is a diffeomorphism from TS^1 to its image, the hyperboloid $x^2 + y^2 - (\cot(\beta)z)^2 = 1$. At the other extreme $\beta = \pi/2$ the fibers are orthogonal to the base, matching our intuition of a bundle in general but depriving points on the fibers from the geometric meaning of pointing along the base. A quarter of the bundle here has been cut off for better visibility. The mesh lines correspond to horizontal (circles parallel to the manifold itself) and vertical (along the fibers, straight lines at the angle β). The parameter β is an adjustable trade-off between two desired properties: (1) horizontal and vertical directions being orthogonal, and (2) the tangent spaces being tangent to the manifold.

Any tensor bundles can be treated in a similar fashion. For example, consider $TM \otimes TM$. The basis elements of $T_xM \otimes T_xM$ are $\partial_i \otimes \partial_j$. The dual basis consists of $\mathrm{d}x^i \otimes \mathrm{d}x^j$ given by

$$\mathrm{d}x^i \otimes \mathrm{d}x^j(a) = a^{ij} \tag{116}$$

for $a \in T_xM \otimes T_xM$. Equivalently, if we expand a in terms of basis elements as

$$a = a^{ij}\partial_i \otimes \partial_j, \tag{117}$$

we can describe the property as

$$\mathrm{d}x^i \otimes \mathrm{d}x^j(\partial_k \otimes \partial_l) = \delta_{ik}\delta_{jl}. \tag{118}$$

Using the maps $\mathrm{d}x^i \otimes \mathrm{d}x^j \colon TU \otimes TU \to \mathbb{R}$ we get coordinate charts $TU \otimes TU \to \mathbb{R}^{3n}$ on $TM \otimes TM$. These make the tensor bundle $TM \otimes TM$ into a smooth manifold.

If E is any tensor bundle (like TM or $T^*M \otimes TM$), we denote the projection $\pi \colon E \to M$ by the same symbol. In general, a bundle is a local product that comes with a global projection.

The preimage $\pi^{-1}(x)$ of a singleton is called a fiber of the bundle. The fibers of the tangent bundle are the tangent spaces $T_xM = \pi^{-1}(x)$.

9.3 Tensor fields

Definition 9.2. A smooth section of a tensor bundle E is a smooth map $f \colon M \to E$ for which $\pi(f(x)) = x$ for all $x \in M$. (In other words, it is a smooth right inverse of the projection π.)

A smooth section of the tangent bundle is also called a smooth vector field. We defined this concept earlier in a different fashion. Sections of general tensor bundles are called tensor fields.

Exercise 9.6. Show that a vector field is smooth if and only if all its components are smooth real-valued functions in any local coordinate system. This shows that our two definitions of a smooth vector field agree. The same holds true for tensor fields of any type. ○

9.4 The sphere bundle

In all of our examples so far the fiber of a bundle is a vector space. Such bundles are called vector bundles. There are other kinds of bundles as well, and many interesting ones are obtained by subbundles of vector bundles.

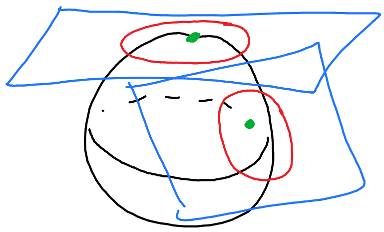


Figure 20: The tangent bundle TS^2 consists of the tangent plane attached to each point. The sphere bundle SS^2 consists of the unit sphere of the tangent plane attached to each point.

A subbundle is, informally, a subset of a bundle that looks locally like a product. A subbundle is a submanifold of the bundle.

The most important example to us is the sphere bundle of a Riemannian manifold

$$SM = \{(x, v) \in TM; |v| = 1\}. \tag{119}$$

The fibers S_xM of SM are unit spheres in the tangent spaces T_xM as depicted in 20.

Exercise 9.7. If $f\colon N \to \mathbb{R}$ is a smooth function on a smooth manifold, then the level set $f^{-1}(0)$ is a smooth submanifold if $\mathrm{d}f \neq 0$ on this set. This smoothness follows from the implicit function theorem. Use this to show that the sphere bundle is a smooth submanifold of TM. ○

The tensor bundles work on a smooth manifold, but the sphere bundle requires a metric.

9.5 Directions and iterated bundles

We can think that T_xM is, informally, the set of all directions one could move on M from x. Thinking of TM as the set of all possible directions of motion is sometimes useful.

The tangent bundle TM is a smooth manifold. The possible directions on it are described by its tangent bundle, the double tangent bundle $TTM = T(TM) = T^2M$.

The fiber at $(x, y) \in TM$, the space $T_{(x,y)}TM$, describes all the possible directions one can move in from (x, y). Heuristically, one should be able to move in two kinds of directions: on the base or on the fiber. This is indeed true invariantly and usefully, but formalizing it is postponed to the next section.

We can, however, describe the tangent vectors in local coordinates. A local coordinate chart $\varphi\colon U \to \mathbb{R}^n$ induces local coordinates $\mathrm{d}\varphi\colon TU \to \mathbb{R}^n \times \mathbb{R}^n$ as described above. Let us denote these coordinates by x^i and y^i — it makes sense to divide coordinates in two halves for base and fiber. The natural basis of $T_{(x,y)}TM$ is given by the vectors

$$\partial_{x^1}, \dots, \partial_{x^n}, \partial_{y^1}, \dots, \partial_{y^n}. \tag{120}$$

The dual basis on $T^*_{(x,y)}TM$ is given by $\mathrm{d}x^i$ and $\mathrm{d}y^i$.

One can take the tangent bundle of any smooth manifold whatsoever. A very natural space for us will be TSM.

★ *Important exercise* 9.8. Let M have dimension n as always. What are the dimensions of the smooth manifolds TM, T^*M, SM, TTM, and TSM? ◯

Manifolds can be embedded in Euclidean spaces and this can give a way to visualize matters. But when it comes to the tangent bundle or especially the double tangent bundle, it is far more transparent to work with abstract manifolds.

9.6 Lifts and geodesics

Many things can be lifted from manifolds to their tangent bundles.

Exercise 9.9. Promoting a smooth function into a function between the bundles is often useful. We defined earlier the differential of a smooth function $f\colon M \to N$ at $x \in M$ as a linear map $\mathrm{d}f(x)\colon T_xM \to T_{f(x)}N$. This induces a map $\mathrm{d}f\colon TM \to TN$. Show that $\mathrm{d}f$ is a bijection if and only if f is a diffeomorphism. ◯

The lift of a smooth curve $\gamma\colon \mathbb{R} \to M$ is the curve $\sigma\colon \mathbb{R} \to TM$ given by $\sigma(t) = (\gamma(t), \dot\gamma(t))$. The second order geodesic equation for γ is a first order equation for the lift σ. We used this to prove existence, uniqueness, and smoothness of geodesics; see exercises 2.11 and 6.1.

Writing a curve σ on TM in terms of the local coordinates on TM gives $x^i(\sigma(t)) = \gamma^i(t)$ and $y^i(\sigma(t)) = \dot\gamma^i(t)$. If σ is the lift of a geodesic γ, then $\partial_t\gamma^i = \dot\gamma^i$ and $\partial_t\dot\gamma^i = -\Gamma^i{}_{jk}\dot\gamma^j\dot\gamma^k$. In other words,

$$\begin{aligned}\partial_t x^i(\sigma) &= \partial_t\gamma^i \\ &= \dot\gamma^i \\ &= y^i(\sigma).\end{aligned} \tag{121}$$

Similarly,

$$\begin{aligned}\partial_t y^i(\sigma) &= \partial_t \dot{\gamma}^i \\ &= -\Gamma^i{}_{jk}\dot{\gamma}^j\dot{\gamma}^k \\ &= -\Gamma^i{}_{jk} y^j(\sigma) y^k(\sigma).\end{aligned} \tag{122}$$

That is, σ satisfies

$$\partial_t \sigma(t) = X(\sigma(t)), \tag{123}$$

for all t, where X is a vector field on TM given in local coordinates by

$$X = y^i \partial_{x^i} - \Gamma^i{}_{jk} y^j y^k \partial_{y^i}. \tag{124}$$

This is called the geodesic vector field.

This should be interpreted so that if $\sigma = (x, v) \in TM$ is some initial data for a geodesic, then $X(\sigma) \in T_{(x,v)}TM$ tells which way the lift of the geodesic $\gamma_{x,v}$ will start moving. The x-component of σ moves in the direction of y and the y-component moves in a direction depending on the Christoffel symbol.

Let us recall that an integral γ of a vector field V on a smooth manifold N is a smooth curve on N satisfying $\dot{\gamma}(t) = V(\gamma(t))$.

Exercise 9.10. Show that if a smooth curve $\sigma\colon \mathbb{R} \to TM$ is the integral curve of the geodesic vector field, then it is a lift of a geodesic. The opposite conclusion was obtained above.

We have found a new description of geodesics: A curve is a geodesic if and only if its lift is an integral curve of the geodesic vector field. Another way to phrase it is that a geodesic is a projection of an integral curve of the geodesic vector field. ◯

We will study this idea further, but we will first need to split $T_{(x,y)}TM$ into "base directions" and "fiber directions" invariantly. The span of the vectors ∂_{y^i} depends on the choice of coordinates.

⋆ *Important exercise* 9.11. Do you have any questions or comments regarding section 9? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ◯

10 Horizontal and vertical subbundles

As we discussed above, the double tangent bundle TTM describes the directions of motion on TM. There are two basic ways to move: along a fiber or in the base. As per exercise 9.5, directions along the fiber are called vertical and those in the base horizontal.

In order to describe the horizontal and vertical subbundles H and V of TTM, we need to describe their fibers $H(\theta)$ and $V(\theta)$ atop each $\theta \in TM$.

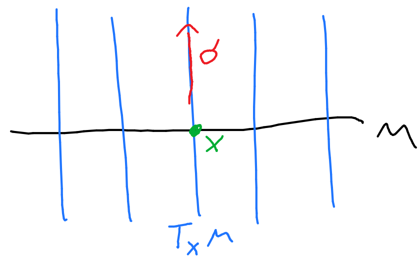


Figure 21: A vertical curve σ on the tangent bundle is such that it stays on a single fiber T_xM atop a single point x.

10.1 The vertical fiber

Consider a curve $\sigma\colon(-\varepsilon,\varepsilon)\to TM$ with $\sigma(0)=\theta=(x,y)\in TM$. If $\sigma(t)$ stays on the fiber T_xM as in figure 21, it makes sense to consider $\dot\sigma(0)\in T_\theta TM$ vertical. We can describe σ staying on the same fiber by saying that $\pi(\sigma(t))$ stays constant. Differentiating this with respect to t at $t=0$ leads to $\mathrm{d}\pi(\theta)\dot\sigma(0)=0$.

Definition 10.1. The vertical fiber at $\theta\in TM$ is

$$V(\theta)=\ker(\mathrm{d}\pi(\theta))\subset T_\theta TM. \tag{125}$$

Observe that this definition does not depend on the Riemannian metric g and can thus be defined on any smooth manifold.

10.2 The horizontal fiber

Consider again a curve σ through $\theta\in TM$. The velocity of the curve should be considered horizontal if only the base point moves but the tangent vector does not as in figure 22. But that does not directly make sense; as x changes, we cannot keep $v\in T_xM$ constant. Fortunately, there is a way to make sense of this through covariant derivatives.

Consider the curve $\gamma=\pi\circ\sigma\colon(-\varepsilon,\varepsilon)\to M$ projected to the base manifold M. Now $\sigma(t)\in T_{\gamma(t)}M$ for all t, so σ can be regarded as a vector field along the curve γ. To make this more explicit, we write $\sigma(t)=(\gamma(t),\Sigma(t))\in TM$. It makes sense to say that the curve σ goes in a horizontal direction if the covariant derivative of Σ along γ vanishes.

To make this more precise, we define a map $K_\theta\colon T_\theta TM\to T_xM$ by requiring that each curve σ on TM with $\sigma(0)=\theta$ satisfies

$$K_\theta\dot\sigma(0)=D_t\Sigma(0), \tag{126}$$

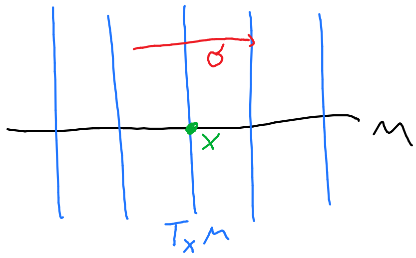


Figure 22: A horizontal curve σ on the tangent bundle is such that the base point moves but there is no vertical motion. As the curve goes through different fibers, lack of vertical motion has to mean parallel transport.

where D_t is the covariant derivative along the curve γ.

To make K_θ a well-defined map, we need to check two properties:

1. The map is defined everywhere: For every $\xi \in T_\theta TM$ there is a curve σ through θ with $\dot{\sigma}(0) = \xi$.
2. The map has a unique value everywhere: If σ_1 and σ_2 are two curves through θ in the direction $\xi \in T_\theta TM$, then $D_t\Sigma_1(0) = D_t\Sigma_2(0)$.

Exercise 10.1. Explain why these properties hold and why K_θ is linear. ◯

Informally, we can think of a vector $\xi \in T_\theta TM$ as (A, B), where A points along the base and B along the fiber. In this view $K_\theta\xi$ is the covariant derivative of B in the direction A.

We can promote these maps K_θ into a global connection map

$$K\colon TTM \to TM \tag{127}$$

given by

$$K(\theta, \xi) = (\pi(\theta), K_\theta\xi). \tag{128}$$

Definition 10.2. The horizontal fiber at $\theta \in TM$ is

$$H(\theta) = \ker(K_\theta) \subset T_\theta TM. \tag{129}$$

An alternative way to describe horizontal directions is to require that parallel transported objects are horizontal. To achieve this, we define a horizontal lift $L_\theta\colon T_xM \to T_\theta TM$ at $\theta = (x, v)$ which describes the ways parallel transports of v evolve in different directions. Given any $w \in T_xM$, let γ_w be a curve through x with $\dot{\gamma}_w(0) = w$. Let $P_w^v(t)$ be the parallel transport of v along γ_w. We get a curve $\sigma_w^v(t) = (\gamma_w(t), P_w^v(t))$ on TM. We now define

$$L_\theta w = \dot{\sigma}_w^v(0). \tag{130}$$

Checking that this map is well defined (independent of the choice of the curve γ_w) and linear is similar to the check of K_θ.

Exercise 10.2. Show that $\ker(K_\theta) = \operatorname{im}(L_\theta)$. This gives a different way to view $H(\theta)$. ◯

Notice that $H(\theta)$ does depend on the notion of parallel transport and therefore on g. Vertical directions are smooth concept, horizontal ones are a metric one.

10.3 Properties of the vertical and horizontal bundles

The vertical subbundle of TTM is the has the fiber $V(\theta)$ at $\theta \in TM$. Similarly, the fibers of the horizontal subbundle are $H(\theta)$. The vertical subbundle gives all the directions along the fibers and the horizontal ones all the directions "along the base".

The various maps we have seen so far have several interesting properties.

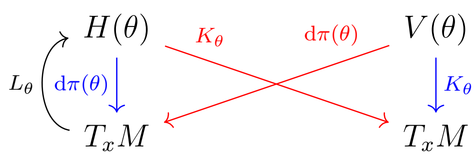
★ *Important exercise* 10.3. Show that $\mathrm{d}\pi(\theta) \circ L_\theta = \mathrm{id}$ on T_xM. ◯

Exercise 10.4. Show that $\mathrm{d}\pi(\theta)|_{H(\theta)} \colon H(\theta) \to T_xM$ is a linear bijection. ◯

Exercise 10.5. Show that $K_\theta|_{V(\theta)} \colon V(\theta) \to T_xM$ is a linear bijection. ◯

★ *Important exercise* 10.6. Show that[37] $T_\theta TM = H(\theta) \oplus V(\theta)$. That is, show that the horizontal and vertical fibers together span $T_\theta TM$ and they only intersect at the origin. ◯

The properties of these maps restricted to the horizontal and vertical fibers can be collected in the following diagram, where blue arrows (straight down) are bijective and red arrows (diagonally) are the zero maps:



(131)

In conclusion, $T_\theta TM$ is can be seen as a product of the horizonal and the vertical fiber. Both $H(\theta)$ and $V(\theta)$ can be identified with T_xM. The projections from $T_\theta TM$ to these two components are $\mathrm{d}\pi(\theta)$ and K_θ. Indeed, the map

$$j_\theta \colon T_\theta TM \to T_xM \times T_xM \tag{132}$$

[37] The symbol $\oplus$ stands for the direct sum, and in this course all direct sums are internal. For a vector space E and two linear subspaces $A, B \subset E$ the equation "$E = A \oplus B$" means that $E = A + B$ and $A \cap B = \{0\}$. In perhaps more intuitive terms, it means that E can be canonically identified with $A \times B$ so that $0 \times B$ and $A \times 0$ map to the corresponding subspaces.

given by

$$j_\theta(\xi) = (\mathrm{d}\pi(\theta)\xi, K_\theta\xi) \tag{133}$$

is a linear bijection.

It is useful to denote the horizontal and vertical parts of $\xi \in T_\theta TM$ as $\xi_h = \mathrm{d}\pi(\theta)\xi \in T_xM$ and $\xi_v = K_\theta\xi \in T_xM$. Identifying with j_θ, we can write $\xi = (\xi_h, \xi_v)$.

10.4 The Sasaki metric

The space T_xM has an inner product given by the metric tensor. The product of two inner product spaces A and B is an inner product space in a natural way:

$$\langle (a,b),(a',b')\rangle_{A\times B} = \langle a,a'\rangle_A + \langle b,b'\rangle_B \,. \tag{134}$$

Definition 10.3. The Sasaki metric on TM is defined so that for each $\theta \in TM$ the map $j_\theta \colon T_\theta TM \to T_xM \times T_xM$ of (132) is a linear isometry.

In other words, the Sasaki metric is a metric tensor on TM — a section of $T^*(TM) \otimes T^*(TM)$ — defined so that

- $H(\theta)$ is orthogonal to $V(\theta)$,
- $\mathrm{d}\pi(\theta)|_{H(\theta)} \colon H(\theta) \to T_xM$ is isometric, and
- $K_\theta|_{V(\theta)} \colon V(\theta) \to T_xM$ is isometric.

For any $\xi, \xi' \in T_\theta TM$ we have

$$\begin{aligned}\langle \xi,\xi'\rangle &= \langle \mathrm{d}\pi(\theta)\xi, \mathrm{d}\pi(\theta)\xi'\rangle + \langle K_\theta\xi, K_\theta\xi'\rangle \\ &= \langle \xi_h,\xi'_h\rangle + \langle \xi_v,\xi'_v\rangle \,.\end{aligned} \tag{135}$$

10.5 Coordinate expressions

Suppose we are given some coordinates x on an open set $U \subset M$. That is, we have a map $x \colon U \to \mathbb{R}^n$ whose coordinates are x^i. We freely identify the point with its coordinates, so we have dropped the chart φ altogether from notation.

The local coordinates x on M induce local coordinates (x, y) on TM. Informally, "$y = \mathrm{d}x$" since the induced coordinates are given by differentials of the original coordinates. That is, a vector $v \in T_xM$ can be written as

$$v = y^i\partial_{x^i}. \tag{136}$$

If we stay on the same fiber, only the variable y changes. The coordinates y^i on the fiber are simply the components of the tangent vector in the coordinates x.

Similarly, the local coordinates (x, y) on TM induce local coordinates (x, y, X, Y) on TTM. A vector $\xi \in T_\theta TM$ at $\theta = (x, y) \in TM$ can be written as

$$\xi = X^i \partial_{x^i} + Y^i \partial_{y^i}. \tag{137}$$

The vectors ∂_{x^i} and ∂_{y^i} form a basis for $T_\theta TM$, but this basis does not go well together with the decomposition to horizontal and vertical directions. The vertical part behaves better.

Lemma 10.4. *The vector fields ∂_{y^i} are a basis for $V(\theta)$.*

Proof. Recall definition 10.1 and exercise 10.4. The claim of the lemma follows from $\mathrm{d}\pi(\theta)\partial_{y^i} = 0$ (so that the vectors are in the right space) and $K_\theta \partial_{y^i} = \partial_{x^i}$ (so that the isomorphism maps them to a known basis).

To show the first property, consider $\mathrm{d}\pi(\theta)\partial_{y^i} \in T_x M$ as a derivation. To that end, let $f : M \to \mathbb{R}$ be smooth. We have

$$\begin{aligned} \mathrm{d}\pi(\theta)\partial_{y^i} f|_x &= \mathrm{d}f(x)[\mathrm{d}\pi(\theta)\partial_{y^i}] \\ &= \mathrm{d}(f \circ \pi)(\theta)\partial_{y^i} \\ &= \partial_{y^i}(f \circ \pi)|_\theta. \end{aligned} \tag{138}$$

The function $f \circ \pi : TM \to \mathbb{R}$ is constant on fibers, so $\partial_{y^i}(f \circ \pi) = 0$.

Let us then move to the second claim. To use the definition (or defining property) of K_θ, we need a curve $\sigma(t)$ on TM for which $\sigma(0) = \theta$ and $\dot\sigma(0) = \partial_{y^i}$. In local coordinates this can be achieved with $\sigma(t) = (\gamma(t), \Sigma(t)) = (x, v + t\partial_{x^i})$. This curve stays on the fiber $T_x M$ and its time derivative is the ith basis vector on $T_x M$. Since $\dot\gamma = 0$, the covariant derivative is simply

$$D_t \Sigma(t)|_{t=0} = \partial_t(v + t\partial_{x^i})|_{t=0} = \partial_{x^i} \tag{139}$$

as required. □

Let us define new vector fields $\delta_{x^i} = \partial_{x^i} - \Gamma^j{}_{ik} y^k \partial_{y^j}$.

Lemma 10.5. *The vector fields δ_{x^i} are a basis for $H(\theta)$.*

The proof consists of two steps:

Exercise 10.7. Prove that $K_\theta \delta_{x^i} = 0$. ○

Exercise 10.8. Prove that $\mathrm{d}\pi(\theta)\delta_{x^i} = \partial_{x^i}$. ○

Using the bases given above, any vector $\xi \in T_\theta TM$ can be written as

$$\xi = X^i \delta_{x^i} + Y^i \partial_{y^i} \tag{140}$$

and the horizontal and vertical components are

$$\xi_h = \mathrm{d}\pi(\theta)\xi = X^i \partial_{x^i} \tag{141}$$

and

$$\xi_v = K_\theta \xi = Y^i \partial_{x^i}. \tag{142}$$

The components stay the same but the basis changes, as one might expect of a natural isomorphism.

The inner product in the Sasaki metric between two vectors $\xi, \tilde{\xi} \in T_\theta TM$ expressed like so is given by

$$\begin{aligned} \left\langle \xi, \tilde{\xi} \right\rangle &= \left\langle \mathrm{d}\pi(\theta)\xi, \mathrm{d}\pi(\theta)\tilde{\xi} \right\rangle + \left\langle K_\theta \xi, K_\theta \tilde{\xi} \right\rangle \\ &= \left\langle X^i \partial_{x^i}, \tilde{X}^j \partial_{x^j} \right\rangle + \left\langle Y^i \partial_{x^i}, \tilde{Y}^j \partial_{x^j} \right\rangle \\ &= X^i \tilde{X}^j \left\langle \partial_{x^i}, \partial_{x^j} \right\rangle + Y^i \tilde{Y}^j \left\langle \partial_{x^i}, \partial_{x^j} \right\rangle \\ &= X^i \tilde{X}^j g_{ij} + Y^i \tilde{Y}^j g_{ij}. \end{aligned} \tag{143}$$

This basis makes the structure of the Sasaki metric more transparent.

Let us then consider what happens on the dual side.

Exercise 10.9. Let $e_1, \ldots, e_k$ be a basis of a vector space. Suppose another basis is given by $f_i = \sum_j A_{ij} e_j$. If the dual basis of the original one is given by e_i^*, the new dual basis is of the form $f_i^* = \sum_j B_{ij} e_j^*$. Show that the matrix B is the inverse transpose of A. ○

The change of basis from ∂_{x^i} and ∂_{y^i} to δ_{x^i} and ∂_{y^i} is given by a matrix of the form

$$A = \begin{pmatrix} I & -G \\ 0 & I \end{pmatrix}, \tag{144}$$

where $G_i^{\ j} = \Gamma^j_{\ ik} y^k$. Therefore the change of basis for the dual basis is given by the matrix

$$B = \begin{pmatrix} I & 0 \\ G^T & I \end{pmatrix}. \tag{145}$$

That is, the dual basis is given by $\mathrm{d}x^i$ and $\delta y^i = \mathrm{d}y^i + \Gamma^i_{\ jk} y^k \mathrm{d}x^j$.

Exercise 10.10. Check that

$$\begin{aligned} &\mathrm{d}x^i(\delta_{x^j}) = \delta^i_j, \\ &\mathrm{d}x^i(\partial_{y^j}) = 0, \\ &\delta y^i(\delta_{x^j}) = 0, \quad \text{and} \\ &\delta y^i(\partial_{y^j}) = \delta^i_j. \end{aligned} \tag{146}$$

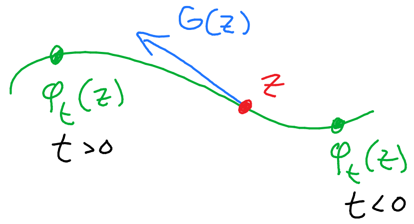


Figure 23: A point $z \in N$ and its associated trajectory obtained by mapping z with φ_t. The generator $G(z)$ of the flow at z is the velocity of the trajectory.

This ensures that we have indeed a dual basis to δ_{x^i} and ∂_{y^i}. This follows from the general observation of exercise 10.9 and the considerations after it, but it is worthwhile to verify by hand. ◯

Observe that we needed to fix the base components ∂_{x^i} to get a nice basis for $T_\theta TM$ and the fiber components $\mathrm{d}y^i$ to get a nice basis on $T^*_\theta TM$. One could say that the vertical and cohorizontal directions do not depend on the metric but the horizontal and covertical ones do.

If needed, these new basis vectors can be used to span the cohorizontal and covertical subspaces of $T^*_\theta TM$. This will rarely be needed, as the Sasaki metric gives a way to identify $T^*_\theta TM$ with $T_\theta TM$.

★ *Important exercise* 10.11. Do you have any questions or comments regarding section 10? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ◯

11 The geodesic flow

11.1 Smooth dynamical systems

A smooth dynamical system or a flow on a smooth manifold N is a smooth action of the group $(\mathbb{R}, +)$ on the diffeomorphism group of N. More concretely, it is a family of smooth maps $\varphi_t \colon N \to N$ so that

- φ_t depends smoothly on t,
- $\varphi_0 = \mathrm{id}$, and
- $\varphi_t \circ \varphi_s = \varphi_{t+s}$.

We often speak of such systems so that a point $z \in N$ flows to the point $\varphi_t(z) \in N$ in time t. The curves $t \mapsto \varphi_t(z)$ are called trajectories. A flow is illustrated in figure 23.

Exercise 11.1. Show that N is a disjoint union of trajectories. ◯

Exercise 11.2. Show that each $\varphi_t \colon N \to N$ is a diffeomorphism. ◯

The dynamical system gives rise to a vector field G on N. It can be defined as a velocity of a trajectory (vector as the velocity of a curve) or as a differential operator (vector as a derivation). The first point of view says that

$$G(z) = \partial_t \varphi(z)|_{t=0}. \tag{147}$$

Taking the second point of view, we can differentiate $f \colon N \to \mathbb{R}$ along the flow by

$$Gf(z) = \partial_t f(\varphi_t(x))|_{t=0}. \tag{148}$$

As the derivative can be written as $\mathrm{d}f(z)(G)$, the two descriptions agree.

Exercise 11.3. Show that a trajectory of the flow φ_t is an integral curve of G. That is, show that a trajectory $\sigma(t) = \varphi_t(z)$ satisfies $\dot\sigma(t) = G(\sigma(t))$. ◯

The exercise shows that the vector field G determines the flow uniquely. Therefore G is called the generator of the flow.

11.2 The geodesic flow

We defined the geodesic vector field in (124). It is a vector field on TM and therefore a section of TTM.

Definition 11.1. The geodesic flow is the flow on the tangent bundle TM of a Riemannian manifold M generated by the geodesic vector field X.

We saw in exercise 9.10 that trajectories of the geodesic flow are exactly the lifts of geodesics.

We should hurry to mention that this definition only makes sense as is if M is geodesically complete. Otherwise some geodesics are not defined for all times. If M is incomplete, the geodesic flow $\varphi \colon \mathbb{R} \times TM \to TM$ is only defined on some open subset of $\mathbb{R} \times TM$. All our considerations will be local, so it does not matter whether the flow is globally defined or not. To keep things simple, we assume M to be geodesically complete, but the assumption is unimportant.

If $(x, v) \in TM$ and $t \in \mathbb{R}$, then $\varphi_t(x, v) = (\gamma_{x,v}(t), \dot\gamma_{x,v}(t))$ gives the position and direction of the geodesic starting at (x, v) after time t. Exercise 6.1 proves the smoothness of the geodesic flow, although in that context we only argued that exponential maps are smooth.

On $T\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n$ the geodesic flow is given simply by $\varphi_t(x, v) = (x + tv, v)$.

Exercise 11.4. Like any vector field on TM, the geodesic vector field can be decomposed to horizontal and vertical components. Verify that at (x, v) we have $X_h = v$ and $X_v = 0$.

That is, the geodesic flow heuristically changes the point on the base but keeps the direction fixed. This corresponds to geodesics parallel transporting their velocity. ◯

⋆ *Important exercise* 11.5. If we want to encode all geodesics on M into a single dynamical system, why does it have to be a system over TM instead of just M? Why cannot a flow on the base encode all geodesics but a flow on the bundle can? ◯

11.3 The differential of the geodesic flow

The geodesic flow is a smooth map $\varphi\colon \mathbb{R} \times TM \to TM$, and for each $t \in \mathbb{R}$ the map $\varphi_t\colon TM \to TM$ is a diffeomorphism. The time derivative is given by the geodesic vector field. Let us therefore study the derivative of φ_t for a fixed $t \in \mathbb{R}$.

To this end, consider a smooth curve $\sigma\colon (-\varepsilon, \varepsilon) \to TM$ through $\theta \in TM$. We would like to find $\partial_s \varphi_t(\sigma(s))|_{s=0}$ in terms of $\sigma'(0)$. The mapping from the latter to the former is $\mathrm{d}\varphi_t(\theta)\colon T_\theta TM \to T_{\varphi_t(\theta)}TM$.

For each $s \in (-\varepsilon, \varepsilon)$ the curve $t \mapsto \varphi_t(\sigma(s))$ is the lift of a geodesic. Therefore $\Gamma(t, s) = \pi(\varphi_t(\sigma(s)))$ is a family of geodesics. Thus we are led to study the Jacobi field $J(t) = \partial_s \Gamma(t, s)|_{s=0}$ along $\gamma(t) = \Gamma(t, 0)$.

If we denote $\sigma'(s) = \partial_s \sigma(s)$ (so that the dot refers to a derivative in t but not in s), we have

$$\begin{aligned} J(t) &= \partial_s \Gamma(t, s)|_{s=0} \\ &= \partial_s \pi(\varphi_t(\sigma(s)))|_{s=0} \\ &= \mathrm{d}\pi(\varphi_t \sigma(0)) \mathrm{d}\varphi_t(\sigma(0))\sigma'(0) \\ &= [\mathrm{d}\varphi_t(\theta)\sigma'(0)]_h. \end{aligned} \tag{149}$$

The Jacobi field gives the horizontal part of the differential.

Let us then find the covariant derivative of this Jacobi field. To that end, we write $\varphi_t(\sigma(s)) = (\alpha(t,s), \beta(t,s)) \in TM$. We find

$$\begin{aligned} D_t J(t) &= D_t \partial_s \Gamma(t,s)|_{s=0} \\ &\overset{(a)}{=} D_s \partial_t \Gamma(t,s)|_{s=0} \\ &\overset{(b)}{=} D_s \partial_t \pi(\varphi_t(\sigma(s)))|_{s=0} \\ &\overset{(c)}{=} D_s \partial_t \alpha(t,s)|_{s=0} \\ &\overset{(d)}{=} D_s \beta(t,s)|_{s=0} \\ &\overset{(e)}{=} K_{\varphi_t(\sigma(0))} \partial_s \varphi_t(\sigma(s))|_{s=0} \\ &\overset{(f)}{=} K_{\varphi_t(\theta)} \mathrm{d}\varphi_t(\sigma(0))\sigma'(0) \\ &\overset{(g)}{=} [\mathrm{d}\varphi_t(\theta)\sigma'(0)]_v. \end{aligned} \tag{150}$$

That is, the covariant derivative of the Jacobi field gives the vertical part of the differential.

Exercise 11.6. Explain the named steps in (150). ○

To get the initial conditions of the Jacobi field, we study what happens at $t = 0$. There $\varphi_0 = \mathrm{id}$ and $\mathrm{d}\varphi_0 = \mathrm{id}$, so $J(0) = [\sigma'(0)]_h$ and $D_t J(0) = [\sigma'(0)]_v$.

Theorem 11.2. *Consider the differential of φ_t at $\theta \in TM$. Choose any $\xi \in T_\theta TM$ and denote $\eta := \mathrm{d}\varphi_t(\theta)\xi \in T_{\varphi_t(\theta)}TM$. If these are decomposed in horizontal and vertical parts as $\xi = (\xi_h, \xi_v)$ and $\eta = (\eta_h, \eta_v)$, then*

$$\begin{aligned} \eta_h &= J_\xi(t) \quad \textit{and} \\ \eta_v &= D_t J_\xi(t), \end{aligned} \tag{151}$$

where J_ξ is the Jacobi field along the geodesic $t \mapsto \pi(\varphi_t(\theta))$ with initial conditions

$$\begin{aligned} J_\xi(0) &= \xi_h \quad \textit{and} \\ D_t J_\xi(0) &= \xi_v. \end{aligned} \tag{152}$$

In words: Through the identifications coming with the horizontal–vertical decomposition, the differential of the geodesic flow is exactly the solution operator of the Jacobi equation.

Exercise 11.7. Prove theorem 11.2. ○

Jacobi fields describe perturbations in position (horizontal), whereas their covariant derivatives describe perturbations in direction (vertical).

If we write the tangent space as $H \oplus V$ at both θ and $\varphi_t(\theta)$, the differential $\mathrm{d}\varphi_t \colon T_\theta TM \to T_{\varphi_t(\theta)}TM$ of theorem 11.2 can be written in block form as

$$\mathrm{d}\varphi_t(\theta) = \begin{pmatrix} A_{hh} & A_{hv} \\ A_{vh} & A_{vv} \end{pmatrix}, \tag{153}$$

where

$$\begin{aligned} &A_{hh} \colon H(\theta) \to H(\varphi_t(\theta)), \\ &A_{hv} \colon V(\theta) \to H(\varphi_t(\theta)), \\ &A_{vh} \colon H(\theta) \to V(\varphi_t(\theta)), \quad \text{and} \\ &A_{vv} \colon V(\theta) \to V(\varphi_t(\theta)) \end{aligned} \tag{154}$$

are linear maps. That is,

$$\begin{pmatrix} \eta_h \\ \eta_v \end{pmatrix} = \begin{pmatrix} A_{hh} & A_{hv} \\ A_{vh} & A_{vv} \end{pmatrix} \begin{pmatrix} \xi_h \\ \xi_v \end{pmatrix}. \tag{155}$$

Each linear map $A(t, \theta)$ depends on t and θ, but we left that dependence implicit.

11.4 The exponential map

Let us return to the exponential map from section 6 and see it from the point of view of the geodesic flow. The geodesic flow contains the lifts of all geodesics for all times. The exponential map only contains the geodesics starting from a single point.

⋆ *Important exercise* 11.8. Show that $\exp_x = \pi \circ \varphi_1|_{T_xM}$. ◯

One could say that the exponential map maps directions to points, as it maps from one fiber (restriction) to the base (projection). Indeed, its differential is indeed a vertical-to-horizontal map.

Exercise 11.9. Consider the block structure of $\mathrm{d}\varphi_t(\theta)$ at $\theta = (x, v)$ given in (153). Show that $\mathrm{d}\exp_x(v) = A_{hv}(1, (x, v))$ when one identifies the horizontal and vertical fibers with tangent spaces of M in the canonical way. ◯

In light of exercise 7.8, the points $\pi(\theta)$ and $\pi(\varphi_T(\theta))$ are conjugate along the geodesic $t \mapsto \pi(\varphi_t(\theta))$ if and only if $A_{hv}(t = T)$ is singular. The whole block matrix of (153) is always invertible because φ_t is a diffeomorphism, but the individual blocks can fail to be invertible.

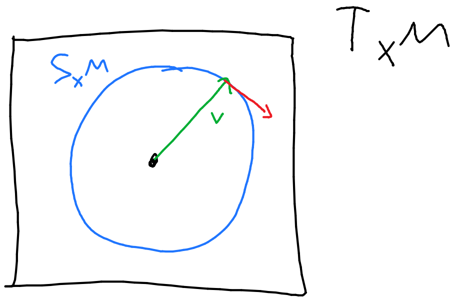


Figure 24: The sphere S_xM sits inside the tangent space T_xM. To stay on the sphere, an admissible direction of motion is orthogonal to the tangent vector.

11.5 The flow on the sphere bundle

Recall that the unit sphere bundle is the set of those $(x, v) \in TM$ for which $|v| = 1$.

Because the speed of a geodesic is constant, the geodesic flow preserves the norm of a vector. Therefore we may restrict the diffeomorphism

$$\varphi_t\colon TM \to TM \tag{156}$$

to

$$\varphi_t\colon SM \to SM. \tag{157}$$

This is a dynamical system on the sphere bundle SM, and its generator is still called the geodesic vector field although it is a slightly different object due to the different ambient manifold. If clarity is required, we will decorate objects with "TM" or "SM".

Since all geodesics have unit speed in the flow on SM, we miss some directions on the bundle. The only direction missing on $T_{(x,v)}TM$ is the vertical direction of v as illustrated in figure 24. Indeed, we have

$$TSM = \{((x, v), \xi) \in TTM; |v| = 1, \xi_v \perp v\}. \tag{158}$$

The missing direction corresponds to reparametrizations of geodesics, so no geometric information is lost in studying the flow on SM.

Consequently, the Jacobi field $t\dot{\gamma}(t)$ does not appear in the differential of the geodesic flow on SM. We can also further restrict directions so that $\dot{\gamma}(t)$ does not appear either. After we have done this in the next section, all Jacobi fields are normal.

★ *Important exercise* 11.10. Do you have any questions or comments regarding section 11? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ○

12 Derivatives on the unit sphere bundle

12.1 Horizontal and vertical bundles on SM

The sphere bundle is a level set of the function $f\colon TM \to \mathbb{R}$, $f(x,y) = g_{ij}(x)y^iy^j$. Given $\theta = (x,y) \in TM$ and $\eta \in T_\theta TM$, let us compute $\mathrm{d}f(\theta)\eta$. Take any curve $\alpha(t) = (a(t), A(t))$ on TM so that $\alpha(0) = \theta$ and $\dot\alpha(0) = \eta$. Now $f(\alpha(t)) = g_{ij}(a(t))A^i(t)A^j(t) = \langle A(t), A(t)\rangle$ and using the covariant derivative along a gives $\partial_t f(\alpha(t)) = 2\langle A(t), D_tA(t)\rangle$.

Let us write η in the new basis we found in section 10.5:

Lemma 12.1. *Let $\alpha(t) = (a(t), A(t))$ be a curve on TM. Its velocity is*

$$\dot\alpha = \dot a^i\partial_{x^i} + \dot A^i\partial_{y^i} = \dot a^i\delta_{x^i} + (D_tA)^i\partial_{y^i}, \tag{159}$$

where D_t is the covariant derivative along a.

Exercise 12.1. Prove the lemma. ○

If we write our $\eta \in T_\theta TM$ as

$$\eta = X^i\delta_{x^i} + Y^i\partial_{y^i}, \tag{160}$$

lemma 12.1 gives us $\mathrm{d}f(\theta)\eta = \partial_t f(\alpha(t))|_{t=0} = 2g_{ij}(x)y^iY^j$.

Because $SM = f^{-1}(1) \subset TM$ is a level set of f, we have

$$TSM = \{X^i\delta_{x^i} + Y^i\partial_{y^i} \in TTM; (x,y) \in SM, g_{ij}(x)y^iY^j = 0\}. \tag{161}$$

There is no restriction in X but Y cannot have anything in the direction of y. We remedy this asymmetry with the following definition using the decomposition on TTM.

Definition 12.2. Let $\theta = (x,y) \in SM$. We define the horizontal and vertical fibers of TSM at θ to be

$$H^{SM}(\theta) = \{X^i\delta_{x^i} \in TSM; g_{ij}(x)y^iX^j = 0\} \tag{162}$$

and

$$V^{SM}(\theta) = T_\theta SM \cap V^{TM}(\theta). \tag{163}$$

In the definition of $H^{SM}(\theta)$ we set $Y^i = 0$. This ensures that the vector is horizontal; see lemmas 10.4 and 10.5. Similarly, in the vertical fiber we have $X^i = 0$, so the point on TSM can be written as $Y^i\partial_{y^i}$ with the constraint $g_{ij}(x)y^iY^j = 0$.

In words:

- At $\theta = (x, v) \in SM$ the horizontal subspace $H^{SM}(\theta) \subset T_\theta SM$ consists of vectors that are purely horizontal and whose horizontal component (a vector on $T_x M$) is orthogonal to v.
- At $\theta = (x, v) \in SM$ the vertical subspace $V^{SM}(\theta) \subset T_\theta SM$ consists of vectors that are purely vertical and whose vertical component (a vector on $T_x M$) is orthogonal to v.

Exercise 12.2. We took out one direction from H^{SM}, and that is the direction of the geodesic vector field X. Show that this one-dimensional subspace is all that is not horizontal or vertical, that is,

$$T_\theta SM = H^{SM}(\theta) \oplus V^{SM}(\theta) \oplus \mathbb{R}X(\theta). \tag{164}$$

This gives a decomposition of TSM into horizontal, vertical, and geodesic directions. ◯

Let N be a bundle over SM whose fiber at $(x, y) \in SM$ is

$$N_{(x,y)} = \{v \in T_x M; \langle v, y \rangle = 0\}. \tag{165}$$

This bundle gives us a way to formalize tangent spaces of $T_x M$ with the direction of y removed.

The sphere $S_x M$ is a manifold, and so it has a tangent space at every point. It is well justified to think that $N_{(x,y)} = T_y S_x M$.

Exercise 12.3. Let $\theta \in SM$. Recall that $T_\theta SM \subset T_\theta TM$. Show that $K_\theta \colon V^{SM}(\theta) \to N_\theta$ is a linear bijection. ◯

Because $T_\theta SM \subset T_\theta TM$, the Sasaki metric gives an inner product on $T_\theta SM$. This is the Riemannian metric the submanifold SM inherits from TM.

Exercise 12.4. We defined things so that $H^{TM}(\theta) = H^{SM}(\theta) \oplus \mathbb{R}X(\theta)$. Show that this direct sum is orthogonal. Is the decomposition (164) orthogonal as well? ◯

As in exercise 12.3, $\mathrm{d}\pi(\theta) \colon H^{SM}(\theta) \to N_\theta$ is a linear bijection. The Sasaki metric was defined so that our maps $V^{SM}(\theta) \to N_\theta$ and $H^{SM}(\theta) \to N_\theta$ are isometries.

Remark 12.3. If we have a section W of the bundle N and a unit speed geodesic γ, we get a natural normal vector field along γ as follows. The lift of γ is the curve $(\gamma, \dot\gamma)$ on SM. At every point $W_\gamma(t) := W(\gamma(t), \dot\gamma(t)) \in T_{\gamma(t)}M$. Because the fiber $N_{\gamma(t),\dot\gamma(t)}$ is the orthogonal complement of $\dot\gamma(t)$, the vector field $W_\gamma(t)$ is indeed orthogonal to $\dot\gamma(t)$.

12.2 Horizontal and vertical gradients on SM

As we will now work mostly on the sphere bundle, let us drop the decorations and write $H^{SM}(\theta) = H(\theta)$ and $V^{SM}(\theta) = V(\theta)$.

Now that we have a handle on different directions on SM, let us differentiate functions. Consider a function $u\colon SM \to \mathbb{R}$. As the Sasaki metric makes SM into a Riemannian manifold, u has a gradient $\nabla u(\theta) \in T_\theta SM$ at $\theta \in SM$. Using the decomposition (see (164))

$$T_\theta SM = H(\theta) \oplus V(\theta) \oplus \mathbb{R}X(\theta), \tag{166}$$

we may decompose the gradient into these three components. The last element in the decomposition is one-dimensional, so it makes sense to treat the third component of the full gradient as a scalar.

Once we identify $H(\theta)$ and $V(\theta)$ with N_θ, we have

$$T_\theta SM = N_\theta \times N_\theta \times \mathbb{R}. \tag{167}$$

In this decomposition

$$\nabla u(\theta) = (\overset{\mathtt{h}}{\nabla} u(\theta), \overset{\mathtt{v}}{\nabla} u(\theta), Xu(\theta)), \tag{168}$$

where $\overset{\mathtt{h}}{\nabla} u$ and $\overset{\mathtt{v}}{\nabla} u$ are the horizontal and vertical gradients of u. Both $\overset{\mathtt{h}}{\nabla} u$ and $\overset{\mathtt{v}}{\nabla} u$ are sections of the bundle N because at every point $\theta \in SM$ they take values in N_θ.

Let us write these derivatives in terms of coordinates. On TM we have the basic derivatives δ_{x^i} and ∂_{y^i}. If we want to express derivatives on SM using these, we need to extend functions from SM to TM to differentiate there. A natural extension is given by the scaling map[38] $s\colon TM \setminus 0 \to SM$, $s(x, y) = (x, y/|y|)$. Now that u is a function on SM, the scaled map $u \circ s$ is a smooth function in a neighborhood of SM on TM. We thus define basic derivatives of u as

$$\begin{aligned} \delta_i u &= \delta_{x^i}(u \circ s)|_{SM} \quad \text{and} \\ \partial_i u &= \partial_{y^i}(u \circ s)|_{SM}. \end{aligned} \tag{169}$$

These operators can be used to write the differential of a function. To get the gradient, we use the musical isomorphisms to write $\delta^i = g^{ij}\delta_j$ and $\partial^i = g^{ij}\partial_j$.

Let us consider the restriction $u_x = u|_{S_xM}$. The gradient[39] of u_x should correspond to $\overset{\mathtt{v}}{\nabla} u$. If we extend u_x to a neighborhood of S_xM in T_xM as $u_x \circ s$ (but evaluate everything on SM), then the differential is

$$\mathrm{d}(u_x \circ s) = \partial_i u \mathrm{d}y^i. \tag{170}$$

[38]Here $0 \subset TM$ stands for the set of the origins of the fibers.

[39]The gradient in the inner product space T_xM or its subset S_xM. That is, this is a Euclidean gradient.

The gradient is obtained by musical isomorphism:

$$\nabla(u_x \circ s) = \partial^i u \partial_{y^i}. \tag{171}$$

The scaling ensures that the radial derivative of $u_x \circ s$ vanishes, and so the gradient is orthogonal to the radial vector on TTM and the gradient belongs to $V(\theta)$. Identifying $V(\theta)$ with N_θ via K_θ, we find the vertical gradient of u to be

$$\overset{\mathtt{v}}{\nabla} u = \partial^i u \partial_{x^i}. \tag{172}$$

Notice that the natural isomorphism K_θ changes the basis, not the components.

Exercise 12.5. Show that the geodesic vector field operates on u in local coordinates as $Xu(x,v) = v^i \delta_i u(x,v)$ at any $(x,v) \in SM$. This justifies thinking of the geodesic vector field as "$X = v \cdot \nabla_x$", the x-derivative in the direction of v. In $\mathbb{R}^n$ we have $S\mathbb{R}^n = \mathbb{R}^n \times S^{n-1}$ and indeed $X = v \cdot \nabla_x$. ◯

To find the horizontal gradient at $\theta = (x,v)$, we can proceed similarly and differentiate $u \circ s$ on TM using the basis elements δ_{x^i}. The full horizontal gradient on $H^{TM}(\theta)$ is

$$\delta^i u \delta_{x^i}. \tag{173}$$

The component in the direction of v should be projected out, as that is already contained in Xu. Recall exercise 12.4. Once we project this geodesic direction out and apply the isomorphism $\mathrm{d}\pi(\theta)\colon H^{SM}(\theta) \to N_\theta$, we find that the horizontal gradient is

$$\overset{\mathtt{h}}{\nabla} u = (\delta^i u - (Xu)v^i)\partial_{x^i}. \tag{174}$$

Now we have found coordinate expressions for the decomposition (168).

12.3 Derivatives of sections of N

There is a natural way to integrate on a Riemannian manifold M. The divergence $\operatorname{div} V$ of a vector field V is defined so that

$$\int_M \langle V, \nabla f \rangle = -\int_M f \operatorname{div} V \tag{175}$$

for all smooth compactly supported $f\colon M \to \mathbb{R}$. In other words, the divergence is the negative formal transpose of the gradient: "div $= -\nabla^T$". The divergence is a first order differential operator given in local coordinates as $\operatorname{div} V = V^i_{;i}$. It is the trace of the covariant derivative ∇V.

Similarly, we may integrate over the Riemannian manifold SM. The horizontal and vertical divergences $\overset{\mathtt{h}}{\mathrm{div}} u$ and $\overset{\mathtt{v}}{\mathrm{div}} u$ of u are defined similarly through transposes by requiring that

$$\int_{SM} \left\langle V, \overset{\mathtt{h}}{\nabla} u \right\rangle = -\int_{SM} u \overset{\mathtt{h}}{\mathrm{div}} V \tag{176}$$

and similarly for $\overset{\mathtt{v}}{\mathrm{div}}$. The geodesic vector field X is skew-adjoint: $X^T = -X$.

The horizontal and vertical divergences map smooth sections of N into smooth functions on SM.

⋆ *Important exercise* 12.6. The geodesic vector field also operates on sections of N. If V is a section, we define

$$XV(\theta) = D_t V(\varphi_t(\theta))|_{t=0}. \tag{177}$$

This is the same formula as for scalar differentiation, but the derivative is now covariant. Show that XV is a section of N.

It follows that if we restrict V to a normal vector field V_γ along γ as in remark 12.3, then the geodesic vector field corresponds to the covariant derivative along the geodesics. That is, $(XV)_\gamma = D_t V_\gamma$. ○

Unfortunately more complete details are beyond the scope of this course.

12.4 Commutator relations

Now that we can differentiate, the question arises whether the various differential operators commute. This is easiest to study on TM first. The coordinate derivatives ∂_{x^i} and ∂_{y^i} all commute with each other.

⋆ *Important exercise* 12.7. Show that $[\delta_{x^i}, \partial_{y^j}] = \Gamma^k{}_{ij}\partial_{y^k}$. ○

The commutator $[\delta_{x^i}, \delta_{x^j}]$ will involve derivatives of Christoffel symbols. Alternatively, it can be seen as a commutator of covariant derivatives. Either way, it should be no surprise that the commutator contains the curvature operator. Using the commutator relations for the basis elements on TTM allows one to compute the commutators for the various derivatives on SM.

Recall the curvature operator along a geodesic γ from definition 5.4. It is an operator depending on $\gamma(t)$ and $\dot\gamma(t)$ and maps $T_{\gamma(t)}M \to T_{\gamma(t)}M$. By lemma 5.5 it also maps $N_{(\gamma(t),\dot\gamma(t))}$ to itself. Therefore the curvature operator induces a map R that maps sections of N to sections of N.

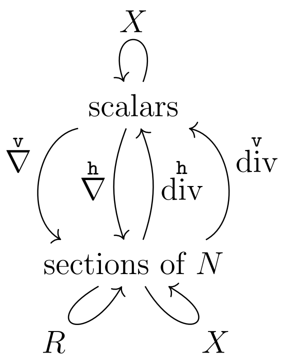
Proposition 12.4. *The differential operators on SM satisfy the following commutator relations, with X interpreted as operating on scalars or sections* *of N as appropriate:*

$$
\begin{aligned}
[X, \overset{\mathtt{v}}{\nabla}] &= -\overset{\mathtt{h}}{\nabla}, \\
[X, \overset{\mathtt{h}}{\nabla}] &= R\overset{\mathtt{v}}{\nabla}, \\
\overset{\mathtt{h}}{\operatorname{div}}\overset{\mathtt{v}}{\nabla} - \overset{\mathtt{v}}{\operatorname{div}}\overset{\mathtt{h}}{\nabla} &= (n-1)X, \\
[X, \overset{\mathtt{v}}{\operatorname{div}}] &= -\overset{\mathtt{h}}{\operatorname{div}}, \\
[X, \overset{\mathtt{h}}{\operatorname{div}}] &= \overset{\mathtt{v}}{\operatorname{div}}R.
\end{aligned}
\tag{178}
$$

We will not prove this proposition[40], but we observe a redundancy.

Exercise 12.8. Prove the formula $[X, \overset{\mathtt{v}}{\operatorname{div}}] = -\overset{\mathtt{h}}{\operatorname{div}}$ assuming $[X, \overset{\mathtt{v}}{\nabla}] = -\overset{\mathtt{h}}{\nabla}$ using the definitions by formal transposes. (A similar argument works for the two commutator relations involving the curvature operator.) ◯

These maps, including both copies of X, can be summarized in a diagram:



(179)

12.5 The Santaló formula

Let us return briefly to integration over SM. While the exact proofs would consume too much time, there is an important idea that we need to discuss: a change of variables associated with the geodesic flow.

To make everything well defined, we have to impose restrictions on the geometry. First of all, we assume M to be a compact Riemannian manifold with boundary. One can define manifolds with boundary abstractly, but one can also think of M as a compact subset with a smooth boundary inside a Riemannian manifold without boundary.

The manifold M has a boundary, and so does its sphere bundle SM:

$$\partial(SM) = \{(x, v) \in SM; x \in \partial M\}. \tag{180}$$

[40] See [6] for a proof and more details.

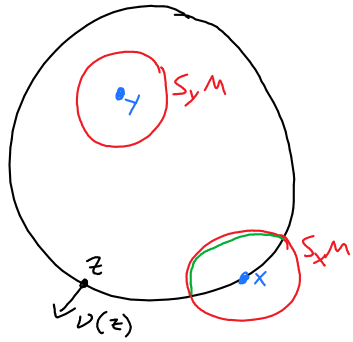


Figure 25: The sphere bundle and its boundary. At an interior point y the sphere $S_y M$ is just the unit tangent sphere of the tangent space and can be thought of as an infinitesimal circle around y. At a boundary point x the sphere $S_x M$ is again a similar sphere, but only part of it points inward. These directions on top of boundary points make up the bundle $\partial_{\text{in}}(SM)$ and serve as initial data for maximal geodesics. We choose the normal vector ν to point outward.

A vector at the boundary can point in three kinds of directions: inwards, tangentially to ∂M, or outwards.

Let $\nu(x)$ be the outer unit normal vector at $x \in \partial M$. Then the tangential vectors at x are precisely those that are normal to $\nu(x)$. The inward pointing boundary is

$$\partial_{\text{in}}(SM) = \{(x,v) \in \partial(SM); \langle v, \nu(x)\rangle < 0.\} \tag{181}$$

This set parametrizes all geodesics that start at the boundary and go inwards and is depicted in figure 25.

To describe how far the geodesic can be extended before falling out of the manifold, we define $\tau\colon SM \to \mathbb{R}$ to be the travel time function so that a geodesic starting at $(x,v) \in SM$ can be maximally extended to the future to be defined on $[0, \tau(x,v)]$.

We want to rule out two problems, depicted in figure 26:

1. There might be geodesics that do not meet the boundary and are thus not parametrized by $\partial_{\text{in}}(SM)$.
2. Some geodesics might start tangentially but still go inside the manifold.

To rule out the first one, we assume that every maximal geodesic has finite length. In other words, given any point and direction, the geodesic comes out in finite time. To rule out the second one, we assume that the boundary is strictly convex in the sense that the second fundamental form of the boundary is positive definite.

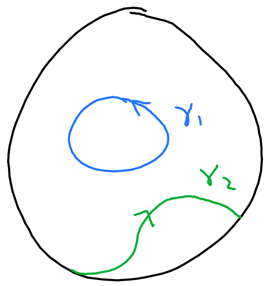


Figure 26: Two kinds of problematic geodesics. The geodesic loop γ_1 cannot be parametrized by an initial condition on the boundary. The geodesic γ_2 can but it starts tangent to the boundary, allowing which is technically awkward.

As $\partial_{\text{in}}(SM) \subset \partial(SM) \subset SM$ is a submanifold, it inherits a Riemannian metric. Therefore one may integrate over it. Let μ be the natural measure on SM and λ the one one $\partial(SM)$. A measure more compatible with the geodesic flow is obtained by $\tilde{\lambda} = |\langle v, \nu \rangle| \, \lambda$.

Proposition 12.5 (The Santaló formula)**.** *Let M be a compact Riemannian manifold with boundary so that every geodesic has finite length and the boundary is strictly convex. Then for any smooth $u \colon SM \to \mathbb{R}$ we have*

$$\begin{aligned} &\int_{(x,v)\in SM} u(x,v)\,\mathrm{d}\mu(x,v) = \\ &\qquad \int_{(x,v)\in\partial_{in}(SM)} \int_0^{\tau(x,v)} u(\varphi_t(x,v))\,\mathrm{d}t\,\mathrm{d}\tilde{\lambda}(x,v). \end{aligned} \tag{182}$$

We omit the proof.[41]

So, to integrate over SM, one can integrate over the space of all geodesics ($\partial_{\text{in}}(SM)$) and then over each geodesic. Think of this as a Fubini-type theorem. In the usual Fubini theorem, one can write the plane as a disjoint union of parallel lines and integrate first over each line and then integrate all those integrals together. Now we just write SM as a union of trajectories of the geodesic flow; see exercise 11.1.

★ *Important exercise* 12.9. Do you have any questions or comments regarding section 12? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ○

[41] See [8, lemma 3.3.2].

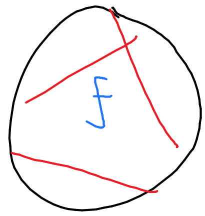


Figure 27: In geodesic X-ray tomography one aims to reconstruct a scalar function f on the manifold M from its integrals over all maximal geodesics.

13 Geodesic X-ray tomography

This section is devoted to a problem whose solution serves as a recap of the course and shows how to apply the tools. The question, illustrated in figure 27, is: Is a function $f\colon M \to \mathbb{R}$ on a Riemannian manifold uniquely determined by its integrals over all geodesics?

13.1 The geodesic X-ray transform

To formalize the question, we define the geodesic X-ray transform. If Γ is the set of all maximal unit speed geodesics on M, the geodesic X-ray transform of $f\colon M \to \mathbb{R}$ is the function $\mathcal{I}f\colon \Gamma \to \mathbb{R}$ given by

$$\mathcal{I}f(\gamma) = \int_a^b f(\gamma(t))\,\mathrm{d}t \tag{183}$$

for a maximal geodesic $\gamma\colon (a,b) \to M$.

Even if f is smooth and compactly supported, the integral might not exist over all geodesics. Therefore we need to impose some restrictions on the geometry of M. We assume M to be compact and all maximal geodesics to have finite length. Then the operator $\mathcal{I}$ is well defined on $C^\infty(M;\mathbb{R})$.

The problem is easiest to study when the space Γ of all geodesics has a good structure. To that end we require that the boundary is strictly convex. This ensures that all geodesics are parametrized by the submanifold $\partial_{\text{in}}(SM) \subset SM$.

Exercise 13.1. We can always take Γ to be the quotient of SM by the geodesic flow. That is, we can define an equivalence relation on SM so that $\theta \sim \theta'$ if and only if $\theta' = \varphi_t(\theta)$ for some $t \in \mathbb{R}$. There are manifolds for which the geodesic flow has a dense trajectory on SM. Show that in this case the quotient $SM/{\sim}$ is not a topological manifold. ◯

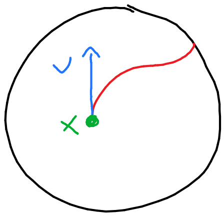


Figure 28: A point x and a unit vector v at it define a forward geodesic segment $\gamma_{x,v}$. The value of $u^f(x,v)$ is the integral of f over this segment starting at (x,v). The travel time function is the integral of the constant function: $\tau = u^1$.

Exercise 13.2. How would you parametrize geodesics in $\mathbb{R}^n$? The parametrization can be redundant. Give a formula for $\mathcal{I}f$ in $\mathbb{R}^n$, when f is smooth and compactly supported. ○

Furthermore, to avoid problems near the boundary, we only study functions that are compactly supported in the interior of the manifold M. That is, there is a positive distance between ∂M and $\operatorname{spt}(f)$. In this case we obtain an operator $\mathcal{I}\colon C_c^\infty(M;\mathbb{R}) \to C_c^\infty(\partial_{\text{in}}(SM);\mathbb{R})$.

The question is: Is this operator injective? That is, do the integrals of f over all geodesics $\gamma \in \Gamma$ determine f uniquely?

We shall show that the operator is indeed injective. To do so, we need to show that if $f \in \ker(\mathcal{I})$, then $f = 0$.

13.2 The transport equation

Take any smooth $f\colon M \to \mathbb{R}$. We define its integral function $u^f\colon SM \to \mathbb{R}$ to be

$$u^f(x,v) = \int_0^{\tau(x,v)} f(\gamma_{x,v}(t))\,\mathrm{d}t. \tag{184}$$

Recall that $\gamma_{x,v}$ is the geodesic starting at (x,v) and $\tau(x,v)$ is the travel time function. The integral is taken from any point all the way up to the boundary. The functions u^f and τ are illustrated in figure 28.

As geodesics are parametrized by their starting points at $\partial_{\text{in}}(SM)$, we may actually write $\mathcal{I}f = u^f|_{\partial_{\text{in}}(SM)}$. The restriction $u^f|_{\partial(SM)\setminus\partial_{\text{in}}(SM)}$ is always zero because the geodesics to be integrated over have zero length.

Exercise 13.3. The manifold M with boundary ∂M can be thought of as follows. Consider a Riemannian manifold $\tilde{M}$ without boundary and a smooth function $\rho\colon \tilde{M} \to \mathbb{R}$. Suppose $M = \rho^{-1}([0,\infty))$, $\partial M = \rho^{-1}(0)$, and $\mathrm{d}\rho \neq 0$

at ∂M. (Smooth domains can always be defined in terms of a smooth defining function ρ like this.)

Take any $(x, v) \in SM \setminus \partial(SM)$ so that the maximal geodesic starting there meets boundary in finite time and is not tangent to it at the exit point. Use the implicit function theorem to show that the travel time function τ is smooth in a neighborhood of (x, v). It follows then from our assumptions that τ is smooth in all of $SM \setminus \partial(SM)$. ◯

Exercise 13.4. Let $\rho\colon M \to \mathbb{R}$ be a boundary defining function; see the previous exercise for a definition. Show that for all $x \in \partial M$ we have $\ker(\mathrm{d}\rho(x)) = T_x\partial M$. ◯

Lemma 13.1. *If f is a smooth compactly supported function in the kernel of $\mathcal{I}$, then u^f is smooth and compactly supported.*

Proof. Everything appearing in the defining integral (184) is smooth in the interior $SM \setminus \partial(SM)$, so u^f is smooth in this set.

If x is close enough to ∂M, then for any $v \in S_xM$ either $\gamma_{x,v}$ or $\gamma_{x,-v}$ will avoid the support of f for all future times and so $u^f(x, \pm v) = 0$ for at least one sign.[42] As $f \in \ker(\mathcal{I})$, we have $u^f(x, v) + u^f(x, -v) = 0$. Thus $u^f(x, v) = 0$ when x is close enough to ∂M. □

Exercise 13.5. Now that we have established that u^f is regular, it remains to establish its crucial property. Show that $Xu^f = -\pi^* f$.

Here the pullback $\pi^* f$ means the composition $f \circ \pi$. As f is a function of $x \in M$ only, and $\pi^* f$ promotes it into a function of $(x, v) \in SM$ which does not depend on v. ◯

We now know that if $\mathcal{I}f = 0$, then u^f is smooth (and compactly supported) and satisfies the transport equation

$$\begin{cases} Xu^f = -\pi^* f & \text{in } SM, \\ u^f = 0 & \text{on } \partial(SM). \end{cases} \tag{185}$$

We will show that this boundary value problem for a partial differential equation has the unique solution $u^f = 0$. It then follows that $\pi^* f = -Xu^f = 0$ and so $f = 0$. This shows that $\mathcal{I}$ is injective.

To show the uniqueness of the solution of the transport equation, observe that the right-hand side of the transport equation $Xu^f = -\pi^* f$ is independent of direction. Therefore its derivative with respect to the direction v vanishes. In other words,

$$0 = \overset{\mathrm{v}}{\nabla}(-\pi^* f) = \overset{\mathrm{v}}{\nabla} Xu^f. \tag{186}$$

[42]This is a little tricky to prove precisely, but the geometric intuition is hopefully clear enough.

Now we have found a homogeneous second order equation for u^f.

13.3 The Pestov identity

To show uniqueness of solutions to the PDE $\overset{\mathtt{v}}{\nabla} Xu = 0$, we will use an energy identity known as a Pestov identity or Mukhometov–Pestov identity. The identity is not hard to prove using our tools, but it can be hard to guess.

Proposition 13.2 (Pestov identity)**.** *If $u \colon SM \to \mathbb{R}$ is smooth and compactly supported, then*

$$\int_{SM} \left| \overset{\mathtt{v}}{\nabla} Xu \right|^2 = \int_{SM} \left| X \overset{\mathtt{v}}{\nabla} u \right|^2 - \int_{SM} \left\langle \overset{\mathtt{v}}{\nabla} u, R \overset{\mathtt{v}}{\nabla} u \right\rangle + (n-1) \int_{SM} \left| Xu \right|^2 . \quad (187)$$

Proof. We will write the various integrals as norms and inner products in $L^2(SM)$. Compact support allows us to integrate by parts without boundary terms. We want to compute

$$\begin{aligned} \left\| \overset{\mathtt{v}}{\nabla} Xu \right\|^2 - \left\| X \overset{\mathtt{v}}{\nabla} u \right\|^2 &= \left(\overset{\mathtt{v}}{\nabla} Xu, \overset{\mathtt{v}}{\nabla} Xu \right) - \left(X \overset{\mathtt{v}}{\nabla} u, X \overset{\mathtt{v}}{\nabla} u \right) \\ &= -\left(\overset{\mathtt{v}}{\operatorname{div}} \overset{\mathtt{v}}{\nabla} Xu, Xu \right) + \left(XX \overset{\mathtt{v}}{\nabla} u, \overset{\mathtt{v}}{\nabla} u \right) \\ &= \left(X \overset{\mathtt{v}}{\operatorname{div}} \overset{\mathtt{v}}{\nabla} Xu, u \right) - \left(\overset{\mathtt{v}}{\operatorname{div}} XX \overset{\mathtt{v}}{\nabla} u, u \right) \\ &= \left((X \overset{\mathtt{v}}{\operatorname{div}} \overset{\mathtt{v}}{\nabla} X - \overset{\mathtt{v}}{\operatorname{div}} XX \overset{\mathtt{v}}{\nabla}) u, u \right) . \end{aligned} \quad (188)$$

To simplify this, we apply the commutator rules of proposition 12.4 to find

$$\begin{aligned} X \overset{\mathtt{v}}{\operatorname{div}} \overset{\mathtt{v}}{\nabla} X - \overset{\mathtt{v}}{\operatorname{div}} XX \overset{\mathtt{v}}{\nabla} = &= (\overset{\mathtt{v}}{\operatorname{div}} X - \overset{\mathtt{h}}{\operatorname{div}}) \overset{\mathtt{v}}{\nabla} X - \overset{\mathtt{v}}{\operatorname{div}} X (\overset{\mathtt{v}}{\nabla} X - \overset{\mathtt{h}}{\nabla}) \\ &= -\overset{\mathtt{h}}{\operatorname{div}} \overset{\mathtt{v}}{\nabla} X + \overset{\mathtt{v}}{\operatorname{div}} X \overset{\mathtt{h}}{\nabla} \\ &= -\overset{\mathtt{h}}{\operatorname{div}} \overset{\mathtt{v}}{\nabla} X + \overset{\mathtt{v}}{\operatorname{div}} (\overset{\mathtt{h}}{\nabla} X + R \overset{\mathtt{v}}{\nabla}) \\ &= (\overset{\mathtt{v}}{\operatorname{div}} \overset{\mathtt{h}}{\nabla} - \overset{\mathtt{h}}{\operatorname{div}} \overset{\mathtt{v}}{\nabla}) X + \overset{\mathtt{v}}{\operatorname{div}} R \overset{\mathtt{v}}{\nabla} \\ &= -(n-1) XX + \overset{\mathtt{v}}{\operatorname{div}} R \overset{\mathtt{v}}{\nabla} . \end{aligned} \quad (189)$$

Therefore

$$\begin{aligned} \left\| \overset{\mathtt{v}}{\nabla} Xu \right\|^2 - \left\| X \overset{\mathtt{v}}{\nabla} u \right\|^2 &= \left((X \overset{\mathtt{v}}{\operatorname{div}} \overset{\mathtt{v}}{\nabla} X - \overset{\mathtt{v}}{\operatorname{div}} XX \overset{\mathtt{v}}{\nabla}) u, u \right) \\ &= -(n-1) \left(XXu, u \right) + \left(\overset{\mathtt{v}}{\operatorname{div}} R \overset{\mathtt{v}}{\nabla} u, u \right) \\ &= (n-1) \left(Xu, Xu \right) - \left(R \overset{\mathtt{v}}{\nabla} u, \overset{\mathtt{v}}{\nabla} u \right) . \end{aligned} \quad (190)$$

This is the claimed identity. ∎

Exercise 13.6. What is the commutator $[\overset{\mathrm{v}}{\mathrm{div}}\overset{\mathrm{v}}{\nabla}, X]$? ○

The Pestov identity is easy to use when $\overset{\mathrm{v}}{\nabla}Xu = 0$ as in our case. Let us try to understand the structure of the right-hand side better. The first two terms only depend on u through $V \coloneqq \overset{\mathrm{v}}{\nabla}u$, which is a smooth section of the bundle N. Per remark 12.3 this section gives rise to a normal vector field $V_\gamma \in NVF_0(\gamma)$ along any maximal geodesic γ. The zero boundary values are due to compact support.

Lemma 13.3. *If V is a smooth section of the bundle N, then*

$$\begin{aligned}\int_{SM} &\left(|XV|^2 - \langle V, RV\rangle\right) \\ &= \int_{(x,v)\in\partial_{in}(SM)} I_{\gamma_{x,v}}(V_{\gamma_{x,v}}, V_{\gamma_{x,v}})\,\mathrm{d}\tilde{\lambda}(x,v),\end{aligned} \tag{191}$$

where λ is the Riemannian volume measure on $\partial_{in}(SM)$, I_γ is the index form along γ, and V_γ is the normal vector field along γ arising from the section V of N.

Proof. We begin by applying the Santaló formula of proposition 12.5 to our integral over SM. In the notation of the proposition, $u(x,v) = |XV|^2 - \langle V, RV\rangle\,|_{(x,v)}$. Santaló gives an integral over the inward pointing boundary $\partial_{\mathrm{in}}(SM)$ and over each geodesic $\gamma = \gamma_{x,v}$ we end up with the integral

$$\begin{aligned}\int_0^{\tau(x,v)} u(\varphi_t(x,v))\,\mathrm{d}t &= \int_0^{\tau(x,v)} \left(|(XV)_\gamma|^2 - \langle V_\gamma, RV_\gamma\rangle\right)\,\mathrm{d}t \\ &= \int_0^{\tau(x,v)} \left(|D_tV_\gamma|^2 - \langle V_\gamma, RV_\gamma\rangle\right)\,\mathrm{d}t \\ &= I_\gamma(V_\gamma, V_\gamma).\end{aligned} \tag{192}$$

This completes the proof. □

To prove uniqueness, we want the right-hand side of our Pestov identity to be positive. We shall see how to do so soon, but we need to make the right assumption to guarantee positivity.

13.4 Injectivity on simple manifolds

Definition 13.4. A simple Riemannian manifold is a compact Riemannian manifold with strictly convex boundary so that each maximal geodesic has finite length and there are no conjugate points.

For example, the closed Euclidean ball is simple. We can think of simple manifolds as "ball-like", but they are quite a bit more flexible.

Theorem 13.5. *The geodesic X-ray transform is injective on smooth compactly supported functions on a simple Riemannian manifold of any dimension $n \geq 2$.*

Proof. Let us take a smooth and compactly supported $f\colon M \to \mathbb{R}$. We assume that $\mathcal{I}f = 0$ and aim to show that $f = 0$.

The integral function $u^f\colon SM \to \mathbb{R}$ is also smooth and compactly supported by lemma 13.1. As we found, this function satisfies $\overset{\mathrm{v}}{\nabla} X u^f = 0$.

Let us then turn to the Pestov identity of proposition 13.2:

$$\left\|\overset{\mathrm{v}}{\nabla} X u^f\right\|^2 = \left\|X\overset{\mathrm{v}}{\nabla} u^f\right\|^2 - \left(\overset{\mathrm{v}}{\nabla} u^f, R\overset{\mathrm{v}}{\nabla} u^f\right) + (n-1)\left\|Xu^f\right\|^2. \tag{193}$$

The left-hand side vanishes.

By theorem 8.8 the index form is positive definite in the absence of conjugate points. Combining this with lemma 13.3 shows that

$$\left\|X\overset{\mathrm{v}}{\nabla} u^f\right\|^2 - \left(\overset{\mathrm{v}}{\nabla} u^f, R\overset{\mathrm{v}}{\nabla} u^f\right) \geq 0. \tag{194}$$

Therefore our energy identity reduces to[43]

$$0 \geq (n-1)\left\|Xu^f\right\|^2. \tag{195}$$

This can only hold if $Xu^f = 0$. Therefore $\pi^* f = -Xu^f = 0$ and so $f = 0$. □

Remark 13.6. One would obtain positivity in the Pestov identity more directly if each of the three terms on the right-hand side is positive. This is the case if $R \leq 0$ in the appropriate sense. This brings us back to sections 5.4 and 8.3, where we saw that there are no conjugate points in non-positive curvature.

★ *Important exercise* 13.7. To summarize, list which tools developed through this course were used to prove theorem 13.5. ○

13.5 Applications

The geodesic X-ray transform appears frequently in the theory of inverse problems. It arises in the study of many inverse boundary value problems for PDEs and as a linearization of non-linear geometric problems. For example, the derivative of the distance between two points with respect to the

[43]This estimate hints at things failing when $n = 1$. And injectivity does indeed fail.

Riemannian metric is an X-ray transform of the variation of the metric tensor. This makes the geodesic X-ray transform appear in linearized travel time tomography in non-Euclidean geometry, which is useful for global seismology and ultrasound imaging. In $\mathbb{R}^3$ and $\mathbb{R}^2$ the transform has direct medical applications, as computerized tomography (CT) is based on it.

⋆ *Important exercise* 13.8. Do you have any questions or comments regarding section 13? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ◯

14 Looking back and forward

14.1 Ways to view geodesics

We found a number of different ways to see geodesics, as

1. critical points of the length functional,
2. minimizers of length (at least locally),
3. solutions to the geodesic equation,
4. as curves that parallel transport their velocity,
5. as projections of trajectories of the geodesic flow, and
6. as curves that lift to integral curves of the geodesic vector field.

Exercise 14.1. Where were these different aspects discussed in the notes? Give, briefly and in your own words, each definition of a geodesic from the list above. ◯

Exercise 14.2. How are the different definitions linked to each other? After all, they define the same concept. ◯

One way to view geodesics that we ignored is to realize the geodesic flow as a Hamiltonian flow on the cotangent bundle with its natural symplectic structure. This topic is highly recommended to readers with any familiarity with Hamiltonian mechanics — and those without any.

14.2 Families of geodesics

The course had two main goals: to understand individual geodesics and families of geodesics. There were several different objects that collected or compared various geodesics:

1. Jacobi fields,
2. the exponential map, and
3. the geodesic flow.

Exercise 14.3. Summarize what geodesics are described by each of the three objects above. ○

Section 11.4 compared the exponential map to the geodesic flow. Differentiating either one leads to Jacobi fields.

The flow lives on SM, so its differential lives on TSM. This was split in three directions: geodesic, horizontal, and vertical. The horizontal and vertical components on TSM correspond to Jacobi fields and their covariant derivatives.

The geodesic flow is always a diffeomorphism, but the exponential map as its restriction can fail to be so. This failure happens locally at conjugate points. Points are conjugate along a geodesic if a Jacobi field vanishes at both points but not identically.

14.3 More aspects of geodesics

We studied local minimization of length and its connections to the index form and conjugate points. All conjugate points to a given point can be collected in a so-called conjugate locus of that point. Theorem 8.8 can be rephrased so that geodesics are locally minimal only up to the conjugate locus but not beyond.

The corresponding global minimization works up to the so-called cut locus — this is not a theorem but a definition. The conjugate locus is further away than the cut locus by theorem 8.8. We did not study global minimization properties of geodesics.

In addition to distances between points, one can study distance between general submanifolds. Zero-dimensional submanifolds are points. Existence and uniqueness of minimizing curves between two submanifolds depends on the geometry of the submanifolds in addition to that of the whole manifold. A minimizing curve is always a geodesic, but the boundary conditions are different.

The endpoint is not fixed, but the direction must be normal to the submanifold. The nature of this condition depends on the codimension of the submanifold. The first variation formula has a boundary term that forces this. The second variation formula has a more complicated boundary term depending on the curvature of the submanifold.

When we studied minimization between points, we used Jacobi fields that correspond to families of geodesics between the two points. Now those have to be replaced by families of geodesics that are normal to the submanifolds at the endpoints. This leads to conditions on vanishing Jacobi fields but the initial conditions are different. Something critical happened when the two points were conjugate. Similarly, something critical happens with the distance from a hypersurface to a point when the point is a focal point. Focal points are analogous to conjugate points, but one endpoint has to be replaced by a hypersurface.

We briefly touched upon geodesic spheres in section 6.5. The shapes of these spheres have interesting properties as one varies the radius and the center. There is an evolution equation for the shape operator of the geodesic sphere that corresponds to the Jacobi equation.

In general, our endeavors have been very local in nature, but there is a substantial amount of global geometry of geodesics to be studied.

14.4 General geometry

If you have not read Riemannian geometry before this course, perhaps you have now found a reason to look into the fundamentals of the theory. Lee's book [4] is highly recommended for that purpose.

The theory of Riemannian geometry branches out quickly, and we have only focused on the branch along a geodesic. Matters like integration, curvature, submanifolds, general fiber bundles, and global geometry deserve a look.

Differential geometry does not end with Riemannian geometry. Pseudo-Riemannian manifolds are very similar to Riemannian manifolds. The metric tensor is not assumed to be a positive definite and symmetric matrix in local coordinates, but only invertible and symmetric. As we lose positivity, we lose a sense of distance, but many of the considerations do not really rely on distance. The geodesic equation, parallel transport, the exponential map, Jacobi fields, and the flow work just as well. If one wants to introduce geodesics as critical points of a functional, energy is better than length. Pseudo-Riemannian, especially Lorentzian, manifolds are heavily used in general relativity.

A step in a different direction can be taken by throwing away not positivity but the existence of a quadratic form. If we only require that every tangent space has a (smooth and strictly convex) norm, we end up with Finsler geometry. Many of our considerations generalize to Finsler geometry, but the details are more technical. A Finsler manifold has a natural Riemannian metric on each tangent space, but this metric depends on a reference direction. Therefore Finsler geometry can be seen as "anisotropic Riemannian geometry".

It turns out that the (co)tangent bundle has a canonical symplectic structure, which in turn induces a contact structure on the (co)sphere bundle. To maintain a narrower focus, we avoided the theory of symplectic manifolds and contact manifolds, but one should be aware of the existence of a broader theory of them. There is a huge variety of different structures one could put on manifolds and bundles, and they easily fill books.

Of course, one can drop all metric properties altogether and only study a smooth manifold or perhaps introduce another kind of additional structure. Or one can keep a metric structure but lose the smooth one and study metric geometry.

The rabbit hole is deep and branches indefinitely. Nevertheless, the reader is invited to enter.

14.5 Geodesic flows

We have only scratched the surface of the theory of geodesic flows. The global and local geometry of the manifold influence the behavior of the flow. For example, curvature has an effect on ergodicity. For a detailed and deep exposition of geodesic flow, see the book [5] by Paternain.

14.6 Integral geometry

In section 13 we studied whether a function is determined by its integrals over geodesics. This is an example of an inverse problem in integral geometry. There are a number of different problems in this spirit. The object to be determined can be a tensor field or a connection, for example. The problems can also be non-linear, and the task can be to determine the whole manifold from some kind of data.

For integral geometry on manifolds, we recommend the books by Sharafutdinov [8] and Paternain–Salo–Uhlmann [7]. For the details we omitted in section 13, the article [6] and its appendices are a good reference. If you want a big picture of the current state of research on such problems, the review [3] and references therein can get you started.

These problems are interesting and highly non-trivial already in Euclidean geometry. For an overview of the various tools and ideas in Euclidean X-ray tomography, we refer the reader to [2].

Another reference on these topics is the email address given on the cover page of these notes.

14.7 Feedback

⋆ *Important exercise* 14.4. Which results or ideas did you find most interesting in this course? ◯

⋆ *Important exercise* 14.5. At times, this course had more focus on ideas than technical details than usual. How did you find this kind of a course? ◯

⋆ *Important exercise* 14.6. Do you feel that something was left out? Is there something — perhaps some of the further study directions mentioned above — that you would like to have seen covered? ◯

⋆ *Important exercise* 14.7. Do you have any questions or comments regarding section 14? Was something confusing or unclear? Were there mistakes? Are there pictures that should be included in this section? ◯

Previous feedback has been of great help in improving these notes. Many thanks for all students who contributed!

References

[1] Philip Hartman, *Ordinary Differential Equations*, SIAM Classics in Applied Mathematics, 2002.

[2] Joonas Ilmavirta, *Analysis and X-ray tomography*, lecture notes, 2017. arXiv:1711.06557

[3] Joonas Ilmavirta & François Monard, *Integral geometry on manifolds with boundary and applications*, chapter 4 in "The Radon Transform: The First 100 Years and Beyond" (Ronny Ramlau, Otmar Scherzer, eds.), de Gruyter, 2019.
arXiv:1806.06088

[4] John Lee, *Introduction to Riemannian Manifolds*, Springer International Publishing, 2nd edition, 2018.

[5] Gabriel P. Paternain, *Geodesic Flows*, Birkhäuser Basel, 1999.

[6] Gabriel P. Paternain & Mikko Salo & Gunther Uhlmann, *Invariant distributions, Beurling transforms and tensor tomography in higher dimensions*, Mathematische Annalen 363 (2015) 305–362.
arXiv:1404.7009

[7] Gabriel P. Paternain & Mikko Salo & Gunther Uhlmann, *Geometric inverse problems in two dimension*, book in preparation.

[8] Vladimir Sharafutdinov, *Ray Transform on Riemannian Manifolds. Eight Lectures on Integral Geometry*, lecture notes, 1999.
`http://www.math.nsc.ru/~sharafutdinov/files/Lectures.pdf`