Geometry of geodesics

Joonas Ilmavirta
joonas.ilmavirta@jyu.fi

January 2020
Version 2

These are lecture notes for the course "MATS4120 Geometry of geodesics" given at the University of Jyväskylä in Spring 2020. Exercise problems are included.

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 entire 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 important ideas.

1.2 Smooth manifolds

Let $n \in \mathbb{N}$. A topological n-dimensional manifold M is a topoplogical 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 $\varphi\colon U \to \varphi(U) \subset \mathbb{R}^n$. 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 $\varphi_i\colon U_i \to \varphi_i(U_i)$ 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 $U \coloneqq U_1 \cap U_2$, the map $\psi\colon \varphi_1(U) \to \varphi_2(U)$ 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.

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 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 $\bigcup_{\alpha \in A} U_\alpha = M$. 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.

[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.

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 of a curve $\gamma\colon I \to M$ at any given time $t \in I$ is a tangent vector in the tangent space $T_{\gamma(t)}M$. 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.

[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 can also take the equivalence class over systems of coordinates.

[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.

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].

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.

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

[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.

We will define later what smoothness of a section means. This will be done twice, in local coordinates and in an invariant fashion.

Any vector space operation can be perform 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.

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

$$R(x)\colon T_x^*M \times T_xM \times T_xM \times T_xM \to \mathbb{R}. \tag{3}$$

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 $(0, 0)$.

For another example of a tensor field, recall that a 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 $(1, 1)$.

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, 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 in formulas (either for itself or an inverse) 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.

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[7] 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 differential operators), it is most natural to let the basis vectors be[8]

$$\partial_i := \left.\frac{\partial}{\partial x^i}\right|_x \in T_x M. \tag{4}$$

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

$$\mathrm{d}x^i(\partial_j) = \delta^i_j. \tag{5}$$

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

$$\begin{aligned} W &= W^i \partial_i, \qquad \text{and} \\ \alpha &= \alpha_i \mathrm{d}x^i. \end{aligned} \tag{6}$$

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:

$$\begin{aligned} W^i \partial_i &:= \sum_{i=1}^n W^i \partial_i, \qquad \text{and} \\ \alpha_i \mathrm{d}x^i &:= \sum_{i=1}^n \alpha_i \mathrm{d}x^i. \end{aligned} \tag{7}$$

[7] The index is up. This is just a convention, but life is much easier when one sticks to it.

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

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.[9]

★ *Important exercise* 1.6. Show that

$$W^i = \mathrm{d}x^i(W) \tag{8}$$

and

$$\alpha_i = \alpha(\partial_i). \tag{9}$$

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 in the preceding exercise.

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

$$a(x) = a^i_j(x)\partial_i \mathrm{d}x^j. \tag{10}$$

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

$$a^i_j(x) = \mathrm{d}x^i(a(x)\partial_j). \tag{11}$$

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. Find 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.

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

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 $e_1 = (1, 0, \ldots, 0)$ 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

$$e^i(W) = \langle e_i, W \rangle . \tag{12}$$

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.

The ith component of a vector W is found by

$$W^i = e^i(W) = \langle e_i, W \rangle \tag{13}$$

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 (11). ◯

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

$$g(x)(v, v) > 0 \tag{14}$$

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

$$g(x)(v, w) = g(x)(w, v) \tag{15}$$

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.[10] Unless otherwise mentioned, we will be working in a single coordinate chart so as to avoid unnecessary complications.

★ *Important exercise* 1.9. Do you have any questions or comments regarding section 1? Was something confusing or unclear? Were there mistakes? ◯

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

$$\langle v, w\rangle := g(v, w). \tag{16}$$

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

$$v \mapsto g(v, \cdot\,). \tag{17}$$

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 (14). ◯

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}.

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

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.

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

$$\alpha_i(x) = \sum_j g_{ij}(x)v^j(x). \tag{18}$$

If we need to compute a derivative like $\partial_k\alpha_i$ in these local coordinates, it can be helpful to look at (18). 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.

If a specific computation confuses you, please bring it up in the end-of-section exercise or otherwise. Future versions of the notes benefit from all feedback.

2.3 Length of curve

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

$$\ell(\gamma) = \int_a^b |\dot\gamma(t)| \,\mathrm{d}t. \tag{19}$$

We define the length of a smooth curve $\gamma\colon [a,b] \to M$ by the same formula.

To properly do so, we must know what $\dot{\gamma}(t)$ is. As discussed above, 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. We define the distance between them to be

$$d(p,q) = \inf\{\ell(\gamma); \gamma\colon [0,1] \to M, \gamma(0) = p, \gamma(1) = q\}. \tag{20}$$

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

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 $\Gamma\colon [0, 1] \times (-\varepsilon, \varepsilon) \to M$ 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

$$\ell(\Gamma(\,\cdot\,, s)) = \int_0^1 |\partial_t \Gamma(t, s)| \, \mathrm{d}t \tag{21}$$

at $s = 0$. Let us work in local coordinates again.

Exercise 2.9. Show that

$$\begin{aligned} &\partial_s [g(\Gamma(t, s))(\partial_t \Gamma(t, s), \partial_t \Gamma(t, s))]^{1/2} \\ &= \frac{1}{2\,|\partial_t \Gamma|} (g_{ij,k}(\Gamma) \partial_s \Gamma^k \partial_t \Gamma^i \partial_t \Gamma^k + 2 g_{ij}(\Gamma) \partial_t \Gamma^i \partial_t \partial_s \Gamma^j), \end{aligned} \tag{22}$$

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

We are now ready to compute the variation of length of a family of constant speed curves from p to q. Recall that reparametrization does not change length so we are free to do so. 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[11] *$\gamma(t) \coloneqq \Gamma(t, 0)$, $\dot\gamma(t) \coloneqq \partial_t \Gamma(t, 0)$, $\ddot\gamma(t) \coloneqq \partial_t^2 \Gamma(t, 0)$, and $V(t) \coloneqq \partial_s \Gamma(t, s)$, we have*

$$\partial_s \ell(\Gamma(\,\cdot\,, s))|_{s=0} = \int_0^1 \frac{1}{|\dot\gamma|} V^k \left[\frac{1}{2} g_{ij,k} \dot\gamma^i \dot\gamma^j - g_{ik,j} \dot\gamma^i \dot\gamma^j - g_{ik} \ddot\gamma^i \right] \mathrm{d}t. \tag{23}$$

[11] Notice that the second order derivatives are computed in local coordinates. We do not yet have proper tools to handle them invariantly.

Proof. Exercise 2.9 shows that the derivative in question is

$$\int_0^1 \frac{1}{|\dot\gamma|}\left[\frac{1}{2}g_{ij,k}V^k\dot\gamma^i\dot\gamma^j + g_{ij}\dot\gamma^i\partial_t V^k\right]\mathrm{d}t. \tag{24}$$

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

$$\frac{1}{2}g_{ij,k}\dot\gamma^i\dot\gamma^j - g_{ik,j}\dot\gamma^i\dot\gamma^j - g_{ik}\ddot\gamma^i = 0. \tag{25}$$

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.

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 derivatives — is defined in local coordinates as

$$\Gamma^i{}_{jk} = \frac{1}{2}g^{il}(g_{lj,k} + g_{lk,j} - g_{jk,l}). \tag{26}$$

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

$$\Gamma^i{}_{jk} = \Gamma^i{}_{kj}. \tag{27}$$

Exercise 2.10. Show that equation (25) is equivalent with

$$\ddot\gamma^i + \Gamma^i{}_{jk}\dot\gamma^j\dot\gamma^k = 0. \tag{28}$$

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 wouold 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 6.7.) We will also find an invariant form of the geodesic equation which in a sense remove 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 $\dot{\gamma}(0) = v$. Existence for long times is not guaranteed unless additional structure is introduced.[12]

Exercise 2.11. Use this result:

> 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.

⋆ *Important exercise* 2.13. Do you have any questions or comments regarding section 2? Was something confusing or unclear? Were there mistakes? ◯

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

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

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.

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 $f : M \to \mathbb{R}$:
- $\nabla_{fX} Y = f\nabla_X Y$
- $\nabla_X(fY) = f\nabla_X Y + X(f)Y$.

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 $X, Y : \mathbb{R}^n \to \mathbb{R}^n$ are smooth vector fields, the standard affine connection of Euclidean geometry is given by

$$(\nabla_X Y)^j = X^i Y^j_{,i} \tag{29}$$

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. [13]

Definition 3.2. An affine connection ∇ on a Riemannian manifold (M,g) is called a symmetric[14] metric connection if

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

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[15], and it is called the Levi-Civita connection[16]. The connection is defined so that for two vector fields $X(x)$ and $Y(x)$ we have

$$(\nabla_X Y)^i = X^j Y^i_{\ ,j} + \Gamma^i_{\ jk} X^j Y^k. \tag{30}$$

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

Exercise 3.2. Prove that the Levi-Civita connection is an affine connection. ◯

Exercise 3.3. Prove that the Levi-Civita connection is a symmetric metric connection. ◯

[13] There is a new exercise for the formula of the commutator in coordinates. See exercise 3.9.

[14] This adjective was missing from the first version. Usually the metric condition only refers to the first property. Symmetry or lack of torsion is the second one.

[15] We will not prove this theorem.

[16] This is named after Tullio Levi-Civita, a single person. Therefore the connection is 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 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

$$\nabla_X(T \otimes R) = \nabla_X T \otimes R + T \otimes \nabla_X R. \tag{31}$$

- The covariant derivative commutes with any contraction or trace.[17]

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

$$\nabla g = 0. \tag{32}$$

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

Recall the differential of a smooth function $f\colon M \to \mathbb{R}$ as a cotangent vector. If tangent vectors are seen as derivations, then $\mathrm{d}f(X) = Xf$. The covariant derivative of f in the direction of a vector field X was just defined so that $\nabla_X f = Xf$. Therefore $\mathrm{d}f(X) = \nabla_X f$. As f is a tensor field of type $(0,0)$, its covariant derivative ∇f as defined above is a tensor field of type $(0,1)$ — a covector field. This covector field should satisfy $(\nabla f)(X) = \nabla_X f$ for any vector field X, so we conclude that the covariant derivative ∇f is exactly $\mathrm{d}f$, the differential of f.

We mentioned earlier 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 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 $\mathrm{d}f$,

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

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:

$$(\nabla_X Y)^i = X^j Y^i_{,j} + \Gamma^i{}_{jk} X^j Y^k. \tag{33}$$

★ *Important exercise* 3.4. 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 3.5. Starting with the covariant derivative of a vector field and the Leibniz rule

$$X(\alpha(Y)) = (\nabla_X \alpha)(Y) + \alpha(\nabla_X Y) \tag{34}$$

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

$$(\nabla_X \alpha)_i = X^j \alpha_{i,j} - \Gamma^j{}_{ik} X^j \alpha_k. \tag{35}$$

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

$$(\nabla_X a)^i_j = X^k a^i_{j,k} + \Gamma^i{}_{kl} a^k_j X^l - \Gamma^k{}_{lj} a^i_k X^l. \tag{36}$$

★ *Important exercise* 3.6. What is the coordinate expression for $\nabla_X g$ for a type $(0,2)$-tensor g? ○

Exercise 3.7. Show directly using the formula of the previous exercise that $\nabla_X g = 0$ when g is the metric tensor. ○

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

$$\nabla_X f = Xf = \mathrm{d}f(X) = \langle \nabla f, X \rangle = \langle \mathrm{d}f, X \rangle, \tag{37}$$

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

$$\frac{\partial}{\partial x^j}V^i(x) = \partial_j V^i = V^i_{\ ,j}. \tag{38}$$

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 (33). One can write this in local coordinates as

$$(\nabla_Y V)^i = Y^j V^i_{\ ;j} \tag{39}$$

by introducing the notation

$$V^i_{\ ;j} = V^i_{\ ,j} + \Gamma^i_{\ jk}V^k. \tag{40}$$

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* 3.8. Do you have any questions or comments regarding section 3? Was something confusing or unclear? Were there mistakes? ◯

Exercise 3.9. Let X and Y be two vector fields. Their commutator as differential operator has only first order terms and is therefore a vector field. Show that this commutator vector field has the components

$$[X,Y]^i = X^j Y^i_{\ j} - Y^j X^i_{\ j}. \tag{41}$$

○

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?

We define 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.
- 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 along the curve.

4.2 Covariant differentiation along a curve

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

$$(D_t V(t))^i = \dot V^i(t) + \Gamma^i_{\ jk} l V^j(t)\dot\gamma(l). \tag{42}$$

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. Let V be any vector field on M. Show that[18]

$$D_t V = \nabla_X V. \tag{43}$$

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

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:

$$D_t\dot{\gamma}(t) = 0. \tag{44}$$

Compare this to the previous versions (25) and (28).

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

$$\dot{\gamma} = \partial_t\gamma = D_t\gamma. \tag{45}$$

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

$$D_t^2\gamma = 0. \tag{46}$$

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_tf as well, but ∂_tf highlights that we are only differentiating a number.

Exercise 4.3. Show that $D_t(fV) = (\partial_tf)V + fD_tV$. ◯

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 $\gamma\colon I \to \mathbb{R}$ 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. 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.*

Proof. As $D_tV = D_tW = 0$, exercise 4.4 implies that $\partial_t \langle V, W\rangle$. 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[19]. 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.

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

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 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[20] 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.

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

$$V(t) = \sum_{\alpha=1}^{n} V_\alpha(t) e_\alpha(t). \tag{47}$$

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.

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.

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

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
>
> $$\partial_s \ell(\Gamma(\,\cdot\,,s))|_{s=0} = -\int_0^1 \frac{1}{|\dot\gamma(t)|} \langle V, D_t\dot\gamma\rangle \, \mathrm{d}t. \tag{48}$$

In this form it is more transparent that the geodesic equation is $D_t\dot\gamma = 0$.
[21]

The variation field of the family, $V(t)$, is a vector field along γ.

Every $\Gamma(\,\cdot\,,s)$ is assumed to be 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 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* 4.8. Do you have any questions or comments regarding section 4? Was something confusing or unclear? Were there mistakes? ○

Exercise 4.9. Let us explain the negative sign in (48). 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. ○

[21] See exercise 4.9 for the sign in front of the integral.

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.

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

$$[\nabla_X, \nabla_Y]Z. \tag{49}$$

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

$$[\nabla_X, \nabla_Y]Z = \nabla_{[X,Y]}Z. \tag{50}$$

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

Based on this observation we rephrase our question: What is

$$[\nabla_X, \nabla_Y]Z - \nabla_{[X,Y]}Z? \tag{51}$$

Proposition 5.1. *There is a smooth tensor field R of type*[22] $(1,3)$ *for which*

$$R(X,Y,Z) = [\nabla_X, \nabla_Y]Z - \nabla_{[X,Y]}Z. \tag{52}$$

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.

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. ○

[22] 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.

We will need analogous results for vector fields along curves. First let $\Gamma\colon [0,1] \to (-\varepsilon, \varepsilon) \to M$ 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*

$$D_t\partial_s\Gamma = D_s\partial_t\Gamma. \tag{53}$$

Exercise 5.3. Prove the lemma. ○

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

$$[D_s, D_t]V = R(\partial_s\Gamma, \partial_t\Gamma)V. \tag{54}$$

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

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.[23] ○

When the family consists of geodesics, the variation field has special properties. It will be what we shall call a Jacobi field.

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 a linear map $T_{\gamma(t)}M \to T_{\gamma(t)}M$ given by

$$R_\gamma V = R(V, \dot\gamma)\dot\gamma. \tag{55}$$

This is in fact a $(1,1)$-tensor along the geodesic; such concepts can be defined by analogy to what we have done.

[23] You have this liberty throughout the course.

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 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

$$D_t^2 J + R_\gamma J = 0. \tag{56}$$

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.*

★ *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 $b(s)$. Let V be the variation field of this family. Use exercise 5.6 to argue that $J = V$. ○

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

$$J(t) = (a + bt)\dot{\gamma}(t). \tag{57}$$

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 (56) 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 that[24]

$$\langle \dot{\gamma}(t), J(t) \rangle = \langle \dot{\gamma}(0), J(0) \rangle + t \langle \dot{\gamma}(0), D_t J(0) \rangle . \tag{58}$$

The easiest way to do this is to compute the second covariant derivative of the inner product. ○

Thus if both J and D_tJ 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

$$J_p(t) = \langle \dot{\gamma}(0), J(0) \rangle \dot{\gamma}(t) + t \langle \dot{\gamma}(0), D_t J(0) \rangle \dot{\gamma}(t). \tag{59}$$

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

$$J_n(t) = J(t) - J_p(t). \tag{60}$$

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.

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

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

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

$$R_\gamma V = k(|\dot{\gamma}|^2 V - \langle V, \dot{\gamma}\rangle \dot{\gamma}). \tag{61}$$

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

$$D_t^2 J + kJ = 0. \tag{62}$$

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

$$J(t) = \sum_{\alpha=1}^{n-1} J_\alpha(t) e_\alpha(t). \tag{63}$$

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

$$J''_\alpha(t) + kJ_\alpha(t) = 0. \tag{64}$$

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

$$J_\alpha(t) = \begin{cases} J_\alpha(t) = a\sin(\sqrt{k}\,t) + b\cos(\sqrt{k}\,t) & \text{when } k > 0, \\ J_\alpha(t) = at + b & \text{when } k = 0, \\ J_\alpha(t) = ae^{\sqrt{-k}\,t} + be^{-\sqrt{-k}\,t} & \text{when } k < 0. \end{cases} \tag{65}$$

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

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 geodesics diverge, in positive curvature they converge.

⋆ *Important exercise* 5.10. Do you have any questions or comments regarding section 5? Was something confusing or unclear? Were there mistakes? ○

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[25] 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 $\exp_x\colon T_xM \to M$,

$$\exp_x(v) = \gamma_{x,v}(1). \tag{66}$$

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.

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:

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

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

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 differentiate, 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 6.4. 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 $\mathrm{d}\exp_x(0)v = \dot{\sigma}(0)$. 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* 6.5. 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 6.6. Given a point x on a Riemannian manifold, how unique are the normal coordinates at it? ◯

Exercise 6.7. Study the geodesic equation (28) 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

$$\mathrm{d}\exp_x(v)w = \partial_s \exp_x(v + sw) = \partial_s\Gamma(1, s)|_{s=0}. \tag{67}$$

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 6.8. 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. Find $J(0)$ and $D_tJ(0)$. ○

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 $J(0) = 0$ and $D_tJ(0) = w$. One can therefore reasonably say that Jacobi fields vanishing at x are the derivative of $\exp_x$.

Exercise 6.9. This description is in fact valid for $v = 0$ as well. Use this description in terms of Jacobi fields to find the differential of the exponential map at the origin. ○

The derivatives satisfy an orthogonality condition named after Gauss:

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

$$\langle \mathrm{d}\exp_x(v)v, \mathrm{d}\exp_x(v)w\rangle = \langle v, w\rangle . \tag{68}$$

Observe that the first inner product is on $T_{\exp_x(v)}M$ and the first 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 $D_t J_1(0) = v$. 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 $D_t J_2(0) = w$. Exercise 5.9 gives

$$\langle \mathrm{d}\exp_x(v)v, \mathrm{d}\exp_x(v)w\rangle = \langle \dot{\gamma}_{x,v}(1), J_2(1)\rangle = \langle v, J_2(0)\rangle + 1\,\langle v, D_t J_2(0)\rangle\,. \tag{69}$$

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.

Remark 6.3. Take any non-zero $v \in T_x M$ and denote the corresponding unit vector by $\hat{v} = v/|v|$. We can complete $\{v\}$ into an orthonormal basis $\{e_1, e_2, \ldots, e_{n-1}, e_n = \hat{v}\}$ of $T_x M$. 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:

$$\mathrm{d}\exp_x(v) = \begin{pmatrix} A & b \\ c^T & d \end{pmatrix}, \tag{70}$$

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 6.10. 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 been done. ○

6.4 Submanifolds

When it comes to submanifolds, geometric intuition serves well for basics 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

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

$$\{\exp_x(v); v \in T_xM, |v| = r\}. \tag{71}$$

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

$$\{y \in M; d(x, y) = r\}. \tag{72}$$

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

These surfaces are closely related as we will soon see. 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 6.5.

Theorem 6.4 (The Gauss lemma for spheres). *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 $\sigma\colon (-\varepsilon, \varepsilon) \to S(0, |v|) \subset T_x M$ 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* 6.11. Do you have any questions or comments regarding section 6? Was something confusing or unclear? Were there mistakes? ○

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 know 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 $v \in B(0,r) \subset T_x M$ the distance between the endpoints of the corresponding geodesic is*

$$d(x, \exp_x(v)) = |v|. \tag{73}$$

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 r. 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.

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 $|\gamma(1)| = |v|$. Then $\sigma = \exp_x \circ \gamma$ is a curve on M from x to the geodesic sphere of radius $|v|$. We have

$$
\begin{aligned}
|v| &= |\gamma(1)| \\
&\overset{(a)}{=} \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} |\gamma(t)| \,\mathrm{d}t \\
&\overset{(b)}{=} \int_0^1 |\gamma(t)|^{-1} \langle \gamma(t), \dot\gamma(t) \rangle \,\mathrm{d}t \\
&\overset{(c)}{=} \int_0^1 |\gamma(t)|^{-1} \langle \mathrm{d}\exp_x(\gamma(t))\gamma(t), \mathrm{d}\exp_x(\gamma(t))\dot\gamma(t) \rangle \,\mathrm{d}t \\
&\overset{(d)}{\leq} \int_0^1 |\gamma(t)|^{-1} |\mathrm{d}\exp_x(\gamma(t))\gamma(t)| \, |\mathrm{d}\exp_x(\gamma(t))\dot\gamma(t)| \,\mathrm{d}t \\
&\overset{(e)}{=} \int_0^1 |\mathrm{d}\exp_x(\gamma(t))\dot\gamma(t)| \,\mathrm{d}t \\
&\overset{(f)}{=} \int_0^1 |\dot\sigma(t)| \,\mathrm{d}t \\
&\overset{(g)}{=} \ell(\sigma).
\end{aligned}
\tag{74}
$$

Justifying each step is an exercise.

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

$$\ell(\gamma_{x,v}|_{[0,1]}) \leq \ell(\sigma). \tag{75}$$

Thus the geodesic is indeed the shortest curve.

Let us then show that it is the unique one. If equality holds throughout (74), the vectors $\mathrm{d}\exp_x(\gamma(t))\gamma(t)$ and $\mathrm{d}\exp_x(\gamma(t))\dot\gamma(t)$ must be parallel[26] 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. □

[26] 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 (74). ◯

Exercise 7.4. 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 will now turn to studying that.

Proposition 7.2. *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 for any 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 6.10 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 $J(1) = 0$.

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

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

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

$$\partial_s^2\ell(\Gamma(\,\cdot\,,s))|_{s=0} = \frac{1}{\ell(\gamma)}\int_0^1 \left(|D_tV|^2 - \langle V, R_\gamma V\rangle\right)\,\mathrm{d}t. \tag{76}$$

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 (48) 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

$$\ell'(s) = -\int_0^1 \frac{1}{\ell(s)}\langle\partial_s\Gamma, D_t\partial_t\Gamma\rangle\,\mathrm{d}t. \tag{77}$$

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

$$\ell''(0) = -\frac{1}{\ell(\gamma)}\int_0^1 \partial_s\langle\partial_s\Gamma, D_t\partial_t\Gamma\rangle|_{s=0}\,\mathrm{d}t. \tag{78}$$

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

Using exercise 4.4 we get

$$\partial_s\langle\partial_s\Gamma, D_t\partial_t\Gamma\rangle|_{s=0} = \langle D_s\partial_s\Gamma, D_t\partial_t\Gamma\rangle|_{s=0} + \langle\partial_s\Gamma, D_sD_t\partial_t\Gamma\rangle|_{s=0}. \tag{79}$$

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

$$D_sD_t\partial_t\Gamma|_{s=0} = D_t^2V + R_\gamma V. \tag{80}$$

With these ingredients we can simplify our second derivative to

$$\ell''(0) = -\frac{1}{\ell(\gamma)}\int_0^1 \left\langle V, D_t^2V + R_\gamma V\right\rangle\,\mathrm{d}t. \tag{81}$$

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

$$D_s D_t \partial_t \Gamma = D_t^2 \partial_s \Gamma + R(\partial_s \Gamma, \partial_t \Gamma)\partial_t \Gamma. \tag{82}$$

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

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

$$\int_a^b \langle V, D_t W \rangle \,\mathrm{d}t = \langle V(b), W(b) \rangle - \langle V(a), W(a) \rangle - \int_a^b \langle D_t V, W \rangle \,\mathrm{d}t. \tag{83}$$

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? ○

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.

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

$$I(V,W) = \int_a^b \left(\langle D_t V, D_t W \rangle - \langle V, R_\gamma W \rangle \right) \mathrm{d}t. \tag{84}$$

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[27]. We say that

- 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 8.1. 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 8.2. 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 $v \in T_x M$ there is $\delta > 0$ so that $I_{\gamma_{x,v}|_{[0,\varepsilon]}}$ is positive semidefinite on $NVF_0(\gamma)$. ◯

A local minimum need not be a global one. 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.

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* 8.3. Let $V \in NVF(\gamma)$. Show that the following are equivalent:

1. V is a Jacobi field.

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

2. $I(V, W) = 0$ for all $W \in NVF_0(\gamma)$.

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

Remark 8.3. Exercise 8.3 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 8.4. *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

$$\bar{J}(t) = \begin{cases} J(t), & a' < t < b' \\ 0, & \text{otherwise} \end{cases} \tag{85}$$

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.

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 8.4.

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 a normal vector field Z along γ as

$$Z(t) = \begin{cases} C\varepsilon^{-1}(|t - a'| - \varepsilon)^2 \zeta(t), & |t - a'| < \varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{86}$$

with some positive constant $C > 0$.

Similarly, if $\eta = D_t J(b')$, we define a parallel transport $\eta(t)$ and let[28]

$$H(t) = \begin{cases} -C\varepsilon^{-1}(|t - b'| - \varepsilon)^2 \eta(t), & |t - b'| < \varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{87}$$

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

[28] Capital ζ is Z, capital η is H.

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 8.5. Now it remains to show that $I(V, V) < 0$ when $\varepsilon > 0$ is small enough. We have

$$
\begin{aligned}
I(V,V) &= \int_a^b \left(|D_t V|^2 - \langle R_\gamma V, V\rangle\right) \mathrm{d}t \\
&= \int_a^b \Big(\left|D_t \bar{J}\right|^2 + 2\left\langle D_t \bar{J}, D_t(Z+H)\right\rangle + |D_t(Z+H)|^2 \\
&\quad - \left\langle R_\gamma \bar{J}, \bar{J}\right\rangle - 2\left\langle R_\gamma \bar{J}, Z+H\right\rangle - \langle R_\gamma(Z+H), Z+H\rangle \Big) \mathrm{d}t \\
&= \int_{a'}^{b'} \left(\left|D_t \bar{J}\right|^2 - \left\langle R_\gamma \bar{J}, \bar{J}\right\rangle\right) \mathrm{d}t \\
&\quad + \int_{a'-\varepsilon}^{a'+\varepsilon} \left(2\left\langle D_t \bar{J}, D_t Z\right\rangle + |D_t Z|^2 - 2\left\langle R_\gamma \bar{J}, Z\right\rangle - \langle R_\gamma Z, Z\rangle\right) \mathrm{d}t \\
&\quad + \int_{b'-\varepsilon}^{b'+\varepsilon} \left(2\left\langle D_t \bar{J}, D_t H\right\rangle + |D_t H|^2 - 2\left\langle R_\gamma \bar{J}, H\right\rangle - \langle R_\gamma H, H\rangle\right) \mathrm{d}t.
\end{aligned}
\tag{88}
$$

If we use exercise 8.3 or remark 8.3 on the geodesic segment $\gamma|_{a',b'}$, we see that

$$
\int_{a'}^{b'} \left(\left|D_t \bar{J}\right|^2 - \left\langle R_\gamma \bar{J}, \bar{J}\right\rangle\right) \mathrm{d}t = 0. \tag{89}
$$

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 (88). We also have $|Z| = \mathcal{O}(\varepsilon)$ and $|H| = \mathcal{O}(\varepsilon)$. Exercise 8.6 gives the other two integrals with contain only Z and H. As the intervals of integration have length 2ε, we have

$$
\begin{aligned}
I(V,V) = 2\int_{a'}^{a'+\varepsilon} \langle D_t J, D_t Z\rangle \,\mathrm{d}t + 2\int_{b'-\varepsilon}^{b'} \langle D_t J, D_t H\rangle \,\mathrm{d}t \\
+ \frac{8}{3} C^2 |\zeta|^2 \varepsilon + \frac{8}{3} C^2 |\eta|^2 \varepsilon + \mathcal{O}(\varepsilon^3).
\end{aligned}
\tag{90}
$$

Let us study the first remaining integral. In it $D_t J(t) = \zeta(t) + \mathcal{O}(\varepsilon)$. Using exercise 8.6 gives thus

$$
2\int_{a'}^{a'+\varepsilon} \langle D_t J, D_t Z\rangle \,\mathrm{d}t = -4C |\zeta|^2 \varepsilon + \mathcal{O}(\varepsilon^2). \tag{91}
$$

The other integral gives a similar negative leading order term.

We have arrived at

$$I(V,V) = -4C\,|\zeta|^2\,\varepsilon - 4C\,|\eta|^2\,\varepsilon + \frac{8}{3}C^2\,|\zeta|^2\,\varepsilon + \frac{8}{3}C^2\,|\eta|^2\,\varepsilon + \mathcal{O}(\varepsilon^2). \qquad (92)$$

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

$$I(V,V) = -\,|\zeta|^2\left(4C - \frac{8}{3}C^2\right)\varepsilon - |\eta|^2\left(4C - \frac{8}{3}C^2\right)\varepsilon + \mathcal{O}(\varepsilon^2) \qquad (93)$$

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

Exercise 8.4. 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 8.5. 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 8.6. Show that $\int_{a'-\varepsilon}^{a+\varepsilon} |D_t Z|\ \mathrm{d}t = \frac{8}{3}C^2\,|\zeta|^2\,\varepsilon$ and $\int_{a'}^{a'+\varepsilon} \langle \zeta(t), D_t Z(t)\rangle\ \mathrm{d}t = -2C\,|\zeta|^2\,\varepsilon$. Similar formulas hold for H with the norm of η. ○

Lemma 8.5. *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. Too see this, suppose that there are coefficients λ_α so that

$$\sum_\alpha \lambda_\alpha J_\alpha(t_0) = 0. \qquad (94)$$

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

$$0 = D_t J(a) = \sum_\alpha \lambda_\alpha \zeta_\alpha. \qquad (95)$$

The vectors ζ_α are linearly independent, so every λ_α vanishes. This proves the linear independence.[29] 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 $t > a$.

[29] 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.

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

$$V(t) = \sum_\alpha V_\alpha(t) J_\alpha(t). \tag{96}$$

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 8.8.

Let us denote

$$A(t) = \sum_\alpha \dot{V}_\alpha(t) J_\alpha(t) \tag{97}$$

and

$$B(t) = \sum_\alpha V_\alpha(t) D_t J_\alpha(t). \tag{98}$$

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

$$\partial_t \langle V, B \rangle = \langle D_t V, B \rangle + \langle V, D_t B \rangle . \tag{99}$$

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

$$\begin{aligned} D_t B &= \sum_\alpha \left[\dot{V}_\alpha(t) D_t J_\alpha(t) + V_\alpha(t) D_t^2 J_\alpha(t) \right] \\ &= \sum_\alpha \left[\dot{V}_\alpha(t) D_t J_\alpha(t) - V_\alpha(t) R_\gamma J_\alpha(t) \right] \\ &= -R_\gamma V(t) + \sum_\alpha \dot{V}_\alpha(t) D_t J_\alpha(t). \end{aligned} \tag{100}$$

Using this with exercise 8.7 leads to

$$\begin{aligned} \langle V, D_t B \rangle &= - \langle V, R_\gamma V \rangle + \sum_\alpha \left\langle V, \dot{V}_\alpha D_t J_\alpha \right\rangle \\ &= - \langle V, R_\gamma V \rangle + \sum_{\alpha,\beta} \left\langle V_\beta J_\beta, \dot{V}_\alpha D_t J_\alpha \right\rangle \\ &= - \langle V, R_\gamma V \rangle + \sum_{\alpha,\beta} V_\beta \dot{V}_\alpha \langle J_\beta, D_t J_\alpha \rangle \\ &= - \langle V, R_\gamma V \rangle + \sum_{\alpha,\beta} V_\beta \dot{V}_\alpha \langle D_t J_\beta, J_\alpha \rangle \\ &= - \langle V, R_\gamma V \rangle + \sum_{\alpha,\beta} \left\langle V_\beta D_t J_\beta, \dot{V}_\alpha J_\alpha \right\rangle \\ &= - \langle V, R_\gamma V \rangle + \langle B, A \rangle . \end{aligned} \tag{101}$$

Putting all of this together gives

$$\begin{aligned}\partial_t \langle V, B\rangle &= \langle A + B, B\rangle - \langle V, R_\gamma V\rangle + \langle B, A\rangle \\ &= |D_t V|^2 - |A|^2 - \langle V, R_\gamma V\rangle .\end{aligned} \tag{102}$$

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

$$\begin{aligned} I(V, V) &= \int_a^b \left(|D_t V|^2 - \langle R_\gamma V, V\rangle\right) \mathrm{d}t \\ &= \int_a^b \left(\partial_t \langle V, B\rangle + |A|^2\right) \mathrm{d}t \\ &= \int_a^b |A|^2 \,\mathrm{d}t \geq 0. \end{aligned} \tag{103}$$

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 8.6. 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* 8.7. Let J_1 and J_2 be two Jacobi fields along the same geodesic. Show that

$$\partial_t \left(\langle D_t J_1, J_2\rangle - \langle J_1, D_t J_2\rangle\right). \tag{104}$$

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 8.8. 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 8.7. *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 8.3, lemma 8.4, and lemma 8.5. 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

$$I(V,W) = \int_a^b \left(\langle D_t V, D_t W \rangle - k \langle V, W \rangle \right) \mathrm{d}t. \tag{105}$$

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 8.7 predicts exactly this behaviour.

If $k > 0$ definiteness depends on length. As we saw in exercise 8.2, 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

$$\int_a^b |V|^2 \,\mathrm{d}t \leq C \int_a^b |D_t V|^2 \,\mathrm{d}t, \tag{106}$$

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* 8.9. Do you have any questions or comments regarding section 8? Was something confusing or unclear? Were there mistakes? ◯