Geometry of geodesics

Joonas Ilmavirta
joonas.ilmavirta@jyu.fi

January 2020
Version 1

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 $U := U_1 \cap U_2$, the map $\psi \colon \varphi_1^{-1}(U) \to \varphi_2^{-1}(U)$ 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.

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.

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

[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.

[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.

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. We will define later what smoothness of a section means.

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.

[5]Ineed, 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.

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 $(1,1)$.

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]

$$\partial_i := \frac{\partial}{\partial x^i} \in T_xM. \tag{4}$$

[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.

Evaluation at the point x is left implicit. 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.

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

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

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:

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

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,

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

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

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.

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

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

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.

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

Exercise 2.1. Expand g, v, and w 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.

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.

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.

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 Length of curve

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

$$\ell(\gamma) = \int_a^b |\dot\gamma(t)| \,\mathrm{d}t. \tag{18}$$

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.

2.3 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

$$d(p,q) = \inf\{\ell(\gamma); \gamma\colon [0,1] \to M, \gamma(0) = p, \gamma(1) = q\}. \tag{19}$$

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. ○

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. ○

2.4 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

$$\ell(\Gamma(\,\cdot\,, s)) = \int_0^1 |\partial_t \Gamma(t, s)| \, \mathrm{d}t \tag{20}$$

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

Exercise 2.9. Show that

$$\begin{aligned} &\partial_s [g(\Gamma(t, s))(\partial_t \Gamma(t, s), \partial_t \Gamma(t, s))]^{1/2} \\ &= \frac{1}{2 |\partial_t \Gamma|} (g_{ij,k}(\Gamma) \partial_s \Gamma^k \partial_t \Gamma^i \partial_t \Gamma^k + 2 g_{ij}(\Gamma) \partial_t \Gamma^i \partial_t \partial_s \Gamma^j), \end{aligned} \tag{21}$$

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 satisfying*

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

Denoting[11] $\dot\gamma(t) \coloneqq \partial_t \Gamma(t, 0)$*,* $\ddot\gamma(t) \coloneqq \partial_t^2 \Gamma(t, 0)$*, and* $V(t) \coloneqq \partial_s \Gamma(t, s)$*, we have*

$$\partial_s \ell(\Gamma(\,\cdot\,, s))|_{s=0} = \int_0^1 \frac{1}{|\partial_t \Gamma|} V^k \left[\frac{1}{2} g_{ij,k} \dot\gamma^i \dot\gamma^j - g_{ik,j} \dot\gamma^i \dot\gamma^j - g_{ik} \ddot\gamma^i \right] \mathrm{d}t. \tag{22}$$

[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

$$\int_0^1 \frac{1}{|\dot\gamma|}\left[\frac{1}{2}g_{ij,k}V^k\dot\gamma^i\dot\gamma^j + g_{ij}\dot\gamma^i\partial_t V^k\right] \mathrm{d}t. \tag{23}$$

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

$$\frac{1}{2}g_{ij,k}\dot\gamma^i\dot\gamma^j - g_{ik,j}\dot\gamma^i\dot\gamma^j - g_{ik}\ddot\gamma^i = 0. \tag{24}$$

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

2.5 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

$$\Gamma^i{}_{jk} = \frac{1}{2}g^{il}(g_{lj,k} + g_{lk,j} - g_{jk,l}). \tag{25}$$

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

$$\Gamma^i{}_{jk} = \Gamma^i{}_{kj}. \tag{26}$$

Exercise 2.10. Show that equation (24) is equivalent with

$$\ddot\gamma^i + \Gamma^i{}_{jk}\dot\gamma^j\dot\gamma^k = 0. \tag{27}$$

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. We will also find an invariant form of the geodesic equation which in a sense remove the pseudoforces from the picture.

2.6 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.

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 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 $f\colon M \to \mathbb{R}$:

- $\nabla_{fX} Y = f \nabla_X Y$
- $\nabla_X (fY) = f \nabla_X Y + X(f)Y$.

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

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

$$(\nabla_X Y)^j = X^i Y^j_{,i} \tag{28}$$

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 Leibnitz 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.

Definition 3.2. An affine connection ∇ on a Riemannian manifold (M, g) is called a 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 Leibnitz rule for the inner product; a Leibnitz 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 metric connect, and it is called the Levi-Civita connection[12]. The connection is defined so that for two vector fields $X(x)$ and $Y(x)$ we have

$$(\nabla_X Y)^i = X^j Y^i_{\ ,j} + \Gamma^i_{\ jk} X^j Y^k. \tag{29}$$

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 metric 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: $\nabla_X f = Xf$.

[12]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.

- On vector fields we have the original connection.
- Tensor products satisfy the Leibniz rule

$$\nabla_X(T \otimes R) = \nabla_X T \otimes R + T \otimes \nabla_X R. \tag{30}$$

- The covariant derivative commutes with any contraction or trace.[13]

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

$$\nabla g = 0. \tag{31}$$

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 $\mathrm{d}f$, 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:

$$(\nabla_X Y)^i = X^j Y^i{}_{,j} + \Gamma^i{}_{jk} X^j Y^k. \tag{32}$$

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.

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

Exercise 3.5. Starting with the covariant derivative of a vector field and the Leibnitz rule
$$X(\alpha(Y)) = (\nabla_X \alpha)(Y) + \alpha(\nabla_X Y) \tag{33}$$
(which follows from the tensor product rule and the trace rule stipulated above), show that
$$(\nabla_X \alpha)_i = X^j \alpha_{i,j} - \Gamma^j{}_{ik} X^j \alpha_k. \tag{34}$$
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
$$(\nabla_X a)^i_j = X^k a^i_{j,k} + \Gamma^i{}_{kl} a^i_j X^l - \Gamma^k{}_{lj} a^l_i X^k. \tag{35}$$

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
$$\nabla_X f = Xf = \mathrm{d}f(X) = \langle \nabla f, X \rangle = \langle \mathrm{d}f, X \rangle\,, \tag{36}$$
where the last inner product is the duality pairing between $T_x M$ 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, e.g., $T^{i\ k}_{\ j\ l}$ instead of T^{ik}_{jl}. This only really becomes crucial when raising and lowering indices by the musical isomorphisms (which extend 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

$$\frac{\partial}{\partial x^j} V^i(x) = \partial_j V^i = V^i_{\ ,j}. \tag{37}$$

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

$$(\nabla_Y V)^i = Y^j V^i_{\ ;j} \tag{38}$$

by introducing the notation

$$V^i_{\ ;j} = V^i_{\ ,j} + \Gamma^i_{\ jk} V^k. \tag{39}$$

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? ○

4 Fields along a curve

4.1 Vector fields along a curve

Let $\gamma\colon I \to M$ be a smooth curve. 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 $V(t) \in T_{\gamma(t)}M$. 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

$$D_t V^i(t) = \dot V^i(t) + \Gamma^i{}_{jk} l V^j(t)\dot\gamma(l). \tag{40}$$

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

$$D_t V = \nabla_X V. \tag{41}$$

Where does this equation make sense? $\bigcirc$

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. $\bigcirc$

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:

$$D_t\dot\gamma(t) = 0. \tag{42}$$

Compare this to the previous versions (24) and (27).

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

$$\dot\gamma = \partial_t\gamma = D_t\gamma. \tag{43}$$

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

$$D_t^2\gamma = 0. \tag{44}$$

The covariant derivative along a curve is also compatible with the metric as one might expect. The following two rules establish the natural Leibnitz 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.)

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

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

4.3 Parallel transport

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

A parallel vector field is the closest we can get to a constant vector field.

Any vector at any point along a curve can be parallel transported along it.

Exercise 4.5. Let $\gamma\colon I \to \mathbb{R}$ be a geodesic. 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 $V(t_0) = V_0$. ◯

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.

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.*

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[14]. 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.

[14] 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 $e_1, \dots, e_n$. 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[15] 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

$$V(t) = \sum_{\alpha=1}^{n} V_\alpha(t) e_\alpha(t). \tag{45}$$

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.

[15] 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 $\Gamma(t, s)$.

Every $\Gamma(\,\cdot\,, s)$ is assumed to be a geodesic. We have infact 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? ◯