Geometric Approximation Theory, Lectures 1-4

P. Grohs

October 16, 2014

1 A Crash Course on Differential Geometry

1.1 What is a Manifold?

Intuitively, when we think of manifolds, we think of subsurfaces of $\mathbb{R}^3$. These are bent surfaces embedded in a linear space, which can be locally flattened. The following definition uses charts to make this property of 'local flattening' of a manifold precise.

Definition 1.1. *Let $\mathcal{M}$ be a topological space which is Hausdorff*[1] *and second-countable*[2]*. A tuple $(\mathcal{U}, \varphi)$ is called a* chart *if $\mathcal{U} \subset \mathcal{M}$ is an open set and $\varphi : \mathcal{U} \to \mathbb{R}^d$ is a homeomorphism*[3]*. An* atlas *of $\mathcal{M}$ is, by definition, a collection $\mathcal{A} = (\mathcal{U}_\alpha, \varphi_\alpha)$ of charts such that*

(i) We have $\bigcup_\alpha \mathcal{U}_\alpha = \mathcal{M}$.

(ii) For any α, β with $\mathcal{U}_\alpha \cap \mathcal{U}_\beta \neq \emptyset$, the mapping $\varphi_\beta^{-1} \circ \varphi_\alpha$ is a smooth mapping[4] *from $\mathbb{R}^d$ to $\mathbb{R}^d$.*

A manifold *of dimension d is, by definition, a pair $(\mathcal{M}, \mathcal{A})$ of a Hausdorff and second-countable topological space $\mathcal{M}$ and an atlas $\mathcal{A}$ as defined above. Two atlases $\mathcal{A}$ and $\mathcal{B}$ are called* equivalent, *if the union $\mathcal{A} \cup \mathcal{B}$ is also an atlas. An equivalence class of atlases defines a* differentiable structure *on $\mathcal{M}$.*

See Figure 1 for an illustration of charts.

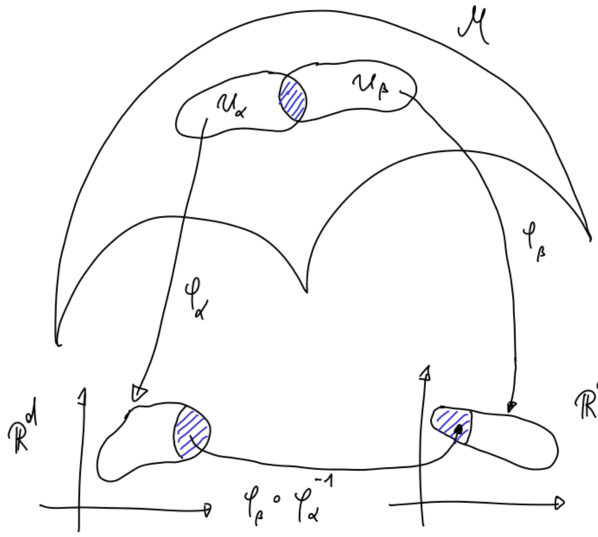

Figure 1: A manifold and two compatible charts

[1] for all $x, y \in \mathcal{M}$ with $x \neq y$ there exist neighborhoods U of x and V of y with $U \cap V = \emptyset$
[2] Every open set can be written as a union of a countable family of open sets
[3] continuous mapping with continuous inverse
[4] In this chapter we always mean 'C^∞' when we say 'smooth'

Example 1.2. *Let $\mathcal{E}$ be a vectorspace with basis $(e_i)_{i=1}^d$. Then the mapping*

$$\varphi : \sum_{i=1}^{d} c_i e_i \mapsto (c_1, \ldots, c_d)^T \in \mathbb{R}^d$$

defines a chart defined on all of $\mathcal{E}$. Hence $(\mathcal{E}, \varphi)$ defines an atlas and $\mathcal{E}$ is a manifold.

The point of this definition is to enable the development of a smooth calculus on nonlinear spaces $\mathcal{M}$ via pullback under charts. To this end we now want to study functions between manifolds and what it means for such a function to be smooth.

Definition 1.3. *Let $\mathcal{M}_1$, $\mathcal{M}_2$ be manifolds of dimension d_1, resp., d_2. Let $F : \mathcal{M}_1 \to \mathcal{M}_2$ be a mapping. Then F is called* smooth *at $x \in \mathcal{M}_1$ if for every (or equivalently one) pair of charts $(\mathcal{U}_1, \varphi_1)$ on $\mathcal{M}_1$ with $x \in \mathcal{U}_1$ and chart $(\mathcal{U}_2, \varphi_2)$ on $\mathcal{M}_2$ with $F(x) \in \mathcal{U}_2$, the associated* chart-representation *$\hat{F} := \varphi_2 \circ F \circ \varphi_1^{-1} : \mathbb{R}^d \to \mathbb{R}^d$ is smooth at $\varphi_1(x) \in \mathbb{R}^d$. F is smooth if it is smooth at every point $x \in \mathcal{M}$. See Figure 2 for an illustration.*

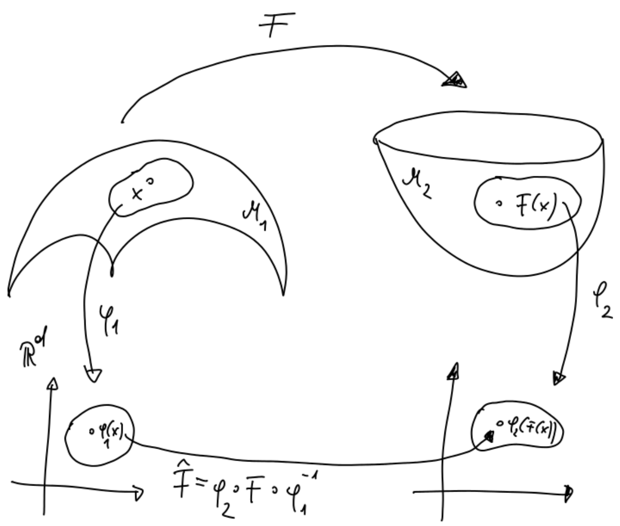

Figure 2: Pullback of a function between two manifolds

Using the pull-back under charts one can also define the rank of mappings between manifolds as follows.

Definition 1.4. *Let F be as above. The* rank *of F at $x \in \mathcal{M}_1$ is defined as the rank of the Jacobian $D\hat{F}|_{\varphi_1(x)}$ (convince yourself that this definition is independent of the choice of charts). F is called an* immersion *if, at every point $x \in \mathcal{M}_1$, F is of rank d_1 (this clearly implies that $d_1 \le d_2$). F is called a* submersion *if, at every point $x \in \mathcal{M}_1$, F is of rank d_2 (this clearly implies that $d_1 \ge d_2$). A point $y \in \mathcal{M}_2$ is called a* regular value of *F if F, restricted to $F^{-1}\{y\}$ is a submersion.*

Up to now we have considered manifolds as an abstract concept. In what follows we will see how to construct submanifolds of a given manifold $\mathcal{M}$, using the concept of submersions. Before we need to define what we mean by a submanifold.

Definition 1.5. *Let $\mathcal{M}$ and $\mathcal{N}$ be manifolds and $\iota : \mathcal{N} \to \mathcal{M}$ an injective, continuous immersion. Then $\mathcal{N}$ is called a* regular submanifold *of $\mathcal{M}$.*

Remark 1.6. *To convince ourselves that this concept of regular submanifold coincides with our intuition (i.e. surfaces/curves in Euclidean space), let's see what happens if we remove one of the assumptions in Definition 1.5. First of all, if we do not require the mapping ι to be globally injective, the embedding will in general possess self-intersections. The importance of* *the continuity assumption can be illustrated with the following example: Consider $\mathcal{M}$ the real line and $\mathcal{N}$ the two-dimensional Torus $\mathbb{R}^2/\sim$, where $(x,y)\sim(\bar{x},\bar{y})$ if and only if $(\bar{x},\bar{y})\in(x,y)+\mathbb{Z}^2$. Given two numbers (a,b) we can define the mapping $f: t\mapsto(ta,tb) \mod \mathbb{Z}^2$. It can be shown that whenever b/a is an irrational number the mapping f is an injective immersion. Moreover the image $f(\mathbb{R})$ is dense in $\mathbb{T}^2$. This is due to the fact that f is not a homeomorphism.*

The following theorem gives a powerful tool to recognize manifolds.

Theorem 1.7. *Let $F:\mathcal{M}_1\to\mathcal{M}_2$ be a smooth mapping between to manifolds of dimensions d_1,d_2 and $d_1>d_2$. Assume that $y\in\mathcal{M}_2$ is a regular value of F. Then $F^{-1}\{y\}$ is a closed regular submanifold of $\mathcal{M}_1$ of dimension d_1-d_2.*

Example 1.8. *The Stiefel manifold $St(p,n)$ is, by definition, the follwing subset of $\mathbb{R}^{n\times p}$:*

$$St(p,n):=\left\{X\in\mathbb{R}^{n\times p}:\ X^TX=I_p\right\}.$$

Special cases include the spheres $\mathbb{S}^{n-1}$ ($p=1$), or the orthogonal group $O(n)$ ($p=n$). We have

$$St(p,n)=F^{-1}\{0\}$$

with

$$F(X):=X^TX-I_p\in Sym(p),$$

where $Sym(p)$ denotes the vector space of symmetric $p\times p$ matrices.

We need to show that 0 is a regular value of $F:\mathbb{R}^{n\times p}\to Sym(p)$.

To this end consider the Jacobian

$$DF|_X(Z):=\frac{d}{dt}|_{t=0}F(X+tZ)=X^TZ+Z^TX.$$

We need to show that for $X\in St(p,n)$ the Jacobian $DF|_X$ defines a surjective mapping, so for any $\hat{Z}\in Sym(p)$ we have to find $Z\in\mathbb{R}^{n\times p}$ with $DF|_X(Z)=\hat{Z}$. Define $Z:=\frac{1}{2}X\hat{Z}$. A direct calculation yields that Z maps to $\hat{Z}$ and we are done. Moreover we have that $d_1=np$ and $d_2=\frac{1}{2}p(p+1)$ and therefore $St(p,n)$ is a closed regular submanifold of $\mathbb{R}^{n\times p}$ of dimension $np-\frac{1}{2}p(p+1)$.

1.2 Tangent Vectors

How can we differentiate functions between manifolds? What is their differential? Consider first a curve $\gamma:\mathbb{R}\to\mathcal{M}$. What is $\gamma'(0)$? If $\mathcal{M}=\mathbb{R}^d$, this would be easy:

$$\gamma'(0)=\lim_{h\to 0}\frac{1}{h}(\gamma(h)-\gamma(0)).$$

On a general manifold (think of the sphere), this expression makes no sense. We have to interpret tangent vectors in a slightly more abstract sense.

Observe first that for a general curve γ and any smooth mapping $f:M\to\mathbb{R}$ (defined around $x:=\gamma(0)$) we can calculate the derivative of $f\circ\gamma$ as above and get as a derivative $\frac{d}{dt}f\circ\gamma(0)$.

If still we had $\mathcal{M}=\mathbb{R}^d$, then we would simply get $\frac{d}{dt}f\circ\gamma(0)=\nabla f(\gamma(0))\cdot\gamma'(0)$. We can then interpret the vector $\gamma'(0)$ as a linear functional on smooth functions $f:\mathcal{M}\to\mathbb{R}$ which acts as $f\mapsto\nabla f(\gamma(0))\cdot\gamma'(0)$. Conversely, any functional of the form $f\mapsto\nabla f(\gamma(0))\cdot v$ belongs to a tangent vector generated by the curve $\gamma(t):=x+tv$. So we can equivalently interpret a tangent vector at x as a tangent vector to a curve or as a derivation operator, acting on functions f.

Lets go back to the general case.

Definition 1.9. *Let $x\in\mathcal{M}$. Define $\mathcal{F}_x(\mathcal{M})$ as the set of all smooth functions $f:\mathcal{U}\to\mathbb{R}$ with $\mathcal{U}$ a neighborhood of x. For a curve $\gamma:\mathbb{R}\to\mathcal{M}$ with $\gamma(0)=x$, define the linear functional $\dot{\gamma}(0):\mathcal{F}_x(\mathcal{M})\to\mathbb{R}$ via*

$$\dot{\gamma}(0)f:=\frac{d}{dt}f\circ\gamma(0).$$

The set of all such operators is called the tangent space *at x and denoted $T_x\mathcal{M}$. The* tangent bundle *of $\mathcal{M}$ is given by the disjoint union $T\mathcal{M}:=\bigcup_{x\in\mathcal{M}}T_x\mathcal{M}$.*

Lemma 1.10. *The tangent space $T_x\mathcal{M}$ is a vector space.*

Proof. Consider two linear functionals $\dot\gamma_1(0), \dot\gamma_2(0)$, as above. We need to show that there exists a curve γ which generates the linear functional $a\dot\gamma_1(0) + b\dot\gamma_2(0)$. Take a chart φ around x with $\varphi(x) = 0$. Define

$$\gamma(t) := \varphi^{-1}(a\varphi(\gamma_1(t)) + b\varphi(\gamma_2(t))).$$

Then we have

$$\gamma(0) = \varphi^{-1}(a\varphi(\gamma_1(0)) + b\varphi(\gamma_2(0))) = \varphi^{-1}(0) = x.$$

Furthermore,

$$\begin{aligned}
\dot\gamma(0)f &= D(f\circ\varphi^{-1})|_0 \left(a\frac{d}{dt}\varphi\circ\gamma_1(0) + b\frac{d}{dt}\varphi\circ\gamma_2(0)\right) \\
&= aD(f\circ\varphi^{-1})|_0 \frac{d}{dt}\varphi\circ\gamma_1(0) + bD(f\circ\varphi^{-1})|_0\frac{d}{dt}\varphi\circ\gamma_2(0) \\
&= a\frac{d}{dt}f\circ\gamma_1(0) + b\frac{d}{dt}f\circ\gamma_2(0) = a\dot\gamma_1(0)f + b\dot\gamma_2(0)f
\end{aligned}$$

□

In general there are infinitely many curves which realize the same tangent vector $\dot\gamma(0)$. The following lemma shows that two curves generate the same tangent vector if and only if their derivatives coincide in a chart.

Lemma 1.11. *Two curves γ_1, γ_2 with $\gamma_1(0) = \gamma_2(0) = x$ satisfy $\dot\gamma_1(0) = \dot\gamma_2(0)$ if and only if for some (then any) chart φ around x we have*

$$\frac{d}{dt}\varphi\circ\gamma_1(0) = \frac{d}{dt}\varphi\circ\gamma_2(0).$$

Proof. Assume that

$$\frac{d}{dt}\varphi\circ\gamma_1(0) = \frac{d}{dt}\varphi\circ\gamma_2(0).$$

Then

$$\dot\gamma_i(0)f = \frac{d}{dt}(f\circ\varphi^{-1}\circ\varphi\circ\gamma)(0).$$

Using the chain rule gives the desired statement. □

The previous lemma leads to another, perhaps slightly more intuitive notion of tangent space, namely the set of all curves $\gamma : \mathbb{R} \to \mathcal{M}$, modulo the equivalence relation $\sim_x$ between curves, defined as follows:

$$\gamma_1 \sim_x \gamma_2 :\Leftrightarrow \frac{d}{dt}\varphi\circ\gamma_1(0) = \frac{d}{dt}\varphi\circ\gamma_2(0)$$

for one (then any) chart φ around x. So we can equivalently define

$$T_x\mathcal{M} = \{\gamma : \mathbb{R} \to \mathcal{M} :\ \gamma(0) = x\}/_{\sim_x}$$

and interpret a tangent vector as a representative $[\gamma]_{\sim_x}$ of curves with equal derivative vector at 0.

We next show that $T_x\mathcal{M}$ is a vector space of dimension d and exhibit a basis for it.

Lemma 1.12. *The tangent space at $x \in \mathcal{M}$ is a vector space of dimension d.*

Proof. Consider a chart $\varphi = (\varphi_1, \ldots, \varphi_d)$ around x. Define the curves

$$\gamma_i(t) = \varphi^{-1}(\varphi(x) + te_i),$$

see Figure 3 We have that

$$\dot\gamma_i(0)f = \partial_i(f\circ\varphi^{-1})(\varphi(x)).$$

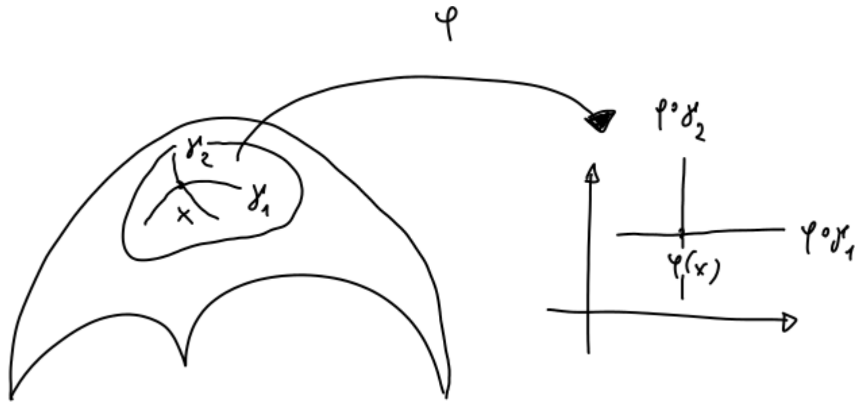

Figure 3: Basis curves γ_1, γ_2 which span the tangent space $T_x\mathcal{M}$

Moreover, we have that

$$\dot{\gamma}(0)f = \frac{d}{dt}(f \circ \varphi^{-1} \circ \varphi \circ \gamma)(0) = \sum_{i=1}^{d} \partial_i (f \circ \varphi^{-1})(\varphi(x)) \frac{d}{dt}(\varphi \circ \gamma)(0)$$

Hence we have

$$\dot{\gamma}(0) = \sum_{i=1}^{d} (\dot{\gamma}_i(0)\varphi_i)\dot{\gamma}_i(0).$$

□

Proposition 1.13. *Suppose that $\mathcal{M}$ is (locally or globally) defined as a level set of a full-rank function $F : \mathcal{E} \to \mathcal{R}^n$ for a vector space $\mathcal{E}$. Then $T_x\mathcal{M}$ is isomorphic to $ker(DF|_x)$.*

Proof. A dimension count shows that both are vector spaces of equal dimension. Let $\gamma(t)$ be a curve with $\gamma(0) = x$. Since $\mathcal{M}$ is a submanifold of a vector space we may define $\gamma'(0) := \lim_{h\to 0} \frac{1}{h}(\gamma(h) - \gamma(0))$. Let $f \in \mathcal{F}_x(\mathcal{M})$. Then we can extend f to a function $\bar{f}$, defined on a neighborhood of x in $\mathcal{E}$. We have

$$\dot{\gamma}(0)f = \nabla \bar{f}(x) \cdot \gamma'(0)$$

Hence we get the identification

$$T_x\mathcal{M} \cong \{\gamma'(0) : \ \gamma : \mathbb{R} \to \mathcal{M}, \ \gamma(0) = x\}$$

Let us now take such a tangent vector $\gamma'(0)$. Clearly, we have

$$DF|_x(\gamma'(0)) = 0.$$

This proves the statement. □

Example 1.14. *We would like to understand the tangent space of $St(p, n)$. $St(p, n)$ is given as the level-set at 0 of the full-rank function $F(X) = X^TX - I_p$, mapping from $\mathbb{R}^{n\times p}$ to $Sym(p)$. We have*

$$DF|_X(Z) = X^TZ + Z^TX.$$

Therefore

$$T_XSt(p, n) \cong \left\{Z \in \mathbb{R}^{n\times p} : X^TZ + Z^TX = 0\right\}.$$

For $p = 1$ (sphere) we get

$$T_X\mathbb{S}^{n-1} = \left\{Z \in \mathbb{R}^n : \ X^TZ = 0\right\}.$$

For $p = n$ (orthogonal group) we get

$$T_XO(n) = \left\{Z = X\Omega \in \mathbb{R}^{n\times n} : \ \Omega^T = -\Omega\right\} \cong Skew(n).$$

We have now established the abstract concept of a tangent space of a manifold. The benefit that we get from this hard work is that now we are able to generalize concepts from (linear) calculus to manifolds, starting below with the differential of a map.

Definition 1.15. *Let $F : \mathcal{M}_1 \to \mathcal{M}_2$ be a smooth mapping. Then the differential $DF|_x : T_x\mathcal{M}_1 \to T_{F(x)}\mathcal{M}_1$ is given by the mapping*

$$DF|_x(\dot\gamma(0)) := (F \dot\circ \gamma)(0).$$

Think carefully how this notion of differential generalizes the notion you are familiar with from calculus. Unfortunately, most calculus courses obscure the distinction between configuration space $\mathbb{R}^d$ and tangent space and identify the tangent spaces with $\mathbb{R}^d$. But also in the linear case the picture you should have in mind is that the differential of a mapping F maps (derivative vectors to) curves to (derivative vectors to) curves.

The first application of our definition of the differential is that it enables us to put a differential structure on the tangent bundle $T\mathcal{M}$.

Lemma 1.16. *Let $\mathcal{M}$ be a differentiable manifold of dimension d. Then the tangent bundle $T\mathcal{M}$ is a differential manifold of dimension $2d$.*

Proof. Let $(\mathcal{U}_\alpha, \varphi_\alpha)$ be an atlas of $\mathcal{M}$. Then the collection $(T\mathcal{U}_\alpha, D\varphi_\alpha)$ constitutes an atlas of $T\mathcal{M}$, as you can easily verify. $\square$

In particular it now makes sense to speak of smooth mappings from and into the tangent bundle of a manifold and it enables us to define the following important notion of *vector fields*.

Definition 1.17. *A vector field on $\mathcal{M}$ is, by definition, a smooth mapping $\xi : \mathcal{M} \to T\mathcal{M}$ which satisfies $\xi(x) \in T_x\mathcal{M}$ for all $x \in \mathcal{M}$. The set of all vector fields on $\mathcal{M}$ is denoted $\mathcal{X}(\mathcal{M})$.*

Remark 1.18. *Structurally, the set $\mathcal{X}(\mathcal{M})$ can be viewed as a $C^\infty(\mathcal{M}, \mathbb{R})$-module. In particular linear combinations of two vector fields are clearly well-defined pointwise, and additionally one can define a $C^\infty(\mathcal{M}, \mathbb{R})$-multiplication by $(f, \xi) \in C^\infty(\mathcal{M}, \mathbb{R}) \times \mathcal{X}(\mathcal{M}) \mapsto f\xi(x) := f(x)\xi(x)$.*

Remark 1.19. *Any vector field can be locally represented in a chart in a natural way. Consider a chart $(\mathcal{U}, \varphi)$. Then for all $x \in \mathcal{U}$ and any vector field ξ we can write*

$$\xi(x) = \sum_{i=1}^{d} (\xi(x)\varphi_i) E_i(x),$$

where $E_i f(x) := \partial_i(f \circ \varphi^{-1})|_{\varphi(x)}$.

1.3 Riemannian Metric

How can we measure distances on a manifold? Suppose again that $\mathcal{M} = \mathbb{R}^d$. Then, for a curve $\gamma : [a, b] \to \mathcal{M}$ we can measure its length by

$$L(\gamma) := \int_a^b (\dot\gamma(t) \cdot \dot\gamma(t))^{1/2}\, dt.$$

So all we really need to do is to compute inner products of tangent vectors.

Definition 1.20. *A* Riemannian structure *on a manifold $\mathcal{M}$ is a smoothly varying family $(g_x(\cdot, \cdot))_{x \in \mathcal{M}}$ of inner products $g_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$. A manifold $\mathcal{M}$, together with a Riemannian structure is called a* Riemannian manifold. *For a curve $\gamma : [a, b] \to \mathcal{M}$ we define its* length *as*

$$L(\gamma) := \int_a^b g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t))^{1/2} dt.$$

Any Riemannian metric induces a distance on $\mathcal{M}$ by

$$\text{dist}(x,y) = \min_{\gamma \text{ curve connecting } x,y} L(\gamma).$$

Suppose now that $\mathcal{N}$ is a submanifold of the Riemannian manifold $\mathcal{M}$ with $\mathcal{N} \subset \mathcal{M}$. Then for any $x \in \mathcal{M}$ we have that $T_x\mathcal{N} \subset T_x\mathcal{M}$. So we can define a Riemannian structure on $\mathcal{N}$ simply by restricting the Riemannian inner product of $\mathcal{M}$ to $T\mathcal{N}$ and call the resulting manifold a Riemannian submanifold of $\mathcal{N}$. Naturally, since $\mathcal{M}$ is regularly embedded into $\mathcal{N}$ we can identify $T_x\mathcal{M}$ with a subspace of the inner product space $T_x\mathcal{N}$ with inner product g_x. We call P_x the (g_x-) orthogonal projection from $T_x\mathcal{N}$ to $T_x\mathcal{M}$ and P_x its orthogonal complement.

Example 1.21. *A natural inner product on $\mathbb{R}^{p\times n}$ is given by*

$$g_X(Z,W) := \textit{trace}(Z^TW).$$

Since $St(p,n)$ is a submanifold of $\mathbb{R}^{p\times n}$, we can define a Riemannian structure on $St(p,n)$ simply by restriction. In the case $p = n$ it is easy to see that we have $P_X(Z) = X\textit{skew}(X^TZ)$, where $\textit{skew}(A) := \frac{1}{2}(A - A^T)$.

Above, we defined the distance between two points $x, y \in \mathcal{M}$ as the length of the shortest curve connecting x and y. This is clearly not very practical; do we really need to check *all* curves connecting the two points? Certainly, if $\mathcal{M} = \mathbb{R}^d$ we don't; we know that the shortest curve connecting x and y is given by a straight line, or equivalently, the unique curve $\gamma : [0,1] \to \mathcal{M}$ with $\gamma(0) = x$, $\gamma(1) = y$ and $\frac{d^2}{dt^2}\gamma(t) = 0$ for all $t \in [0,1]$. Does there exist a similar characterization for general manifolds $\mathcal{M}$?

Example 1.22. *Let's look at a very simple case, the unit circle $\mathbb{S}^1$. Consider two points $x = (\sin(\theta_1), \cos(\theta_1))^T$ and $y = (\sin(\theta_2), \cos(\theta_2))^T$ for two angles $\theta_1, \theta_2 \in [0, 2\pi)$, $\theta_2 \geq \theta_1$. The shortest curve connecting x and y is given by*

$$\gamma(t) = (\sin(\theta), \cos(\theta))^T, \quad \theta \in [\theta_1, \theta_2].$$

We can compute the lenght of γ:

$$L(\gamma) = \int_{\theta_1}^{\theta_2} (\gamma'(\theta) \cdot \gamma'(\theta))^{1/2}\, d\theta = \theta_2 - \theta_1,$$

and this is the distance between x and y. The second derivative of the shortest curve is given by

$$\gamma''(t) = -(\sin(\theta), \cos(\theta))^T \neq 0.$$

Even though, the acceleration of γ is nonzero we observe that the acceleration vector $\gamma''(t)$ is always orthogonal to the tangent space $T_{\gamma(t)}\mathbb{S}^1$, or in other words,

$$P_{\gamma(t)}(\gamma''(t)) = 0 \quad \textit{for all } t \in [\theta_1, \theta_2].$$

So the tangent vector of $\gamma(t)$ does not change at all tangentially. We can interpret this property geometrically in the following sense: An observer who lives on the circle and moves along with the curve γ does not see any change in γ'.

In the previous example we have seen that the shortest curves are exactly those which satisfy $P_{\gamma(t)}(\gamma''(t)) = 0$. For Riemannian submanifolds of Euclidean space this is actually always true as we shall see. First we need to define the notion of covariant derivative on Riemannian submanifolds of Euclidean space. The general definition (for non-embedded Riemannian manifolds) is more abstract and will be omitted at this point.

Definition 1.23. *Let $\mathcal{M}$ be a Riemannian submanifold of Euclidean space $\mathbb{R}^n$. Let $\eta, \xi \in \mathcal{X}(\mathcal{M})$ be two vector fields. Then the* covariant derivative *$\nabla_\eta\xi \in \mathcal{X}(\mathcal{M})$ of ξ along η is defined as*

$$\nabla_\eta\xi(x) := P_x\left(D\xi|_x(\eta(x))\right).$$

If $\xi(t)$ is a vector field along a curve $\gamma(t)$, i.e., $\xi(t) \in T_{\gamma(t)}\mathcal{M}$, then we define the covariant derivative $\frac{D}{dt}\xi$ *as*

$$\frac{D}{dt}\xi(x) := P_x\left(\frac{d}{dt}\xi(t)\right).$$

A curve $\gamma : \mathbb{R} \to \mathcal{M}$ is called geodesic *if*

$$\frac{D}{dt}(\dot{\gamma}(t)) = 0 \quad \textit{for all } t.$$

Remark 1.24. *Of course the notions described above can also be defined intrinsically, that is, without reference to an Euclidean embedding. There we require the notion of* affine connection *an a manifold. An affine connection $\nabla : \mathcal{X}(\mathcal{M}) \times \mathcal{X}(\mathcal{M}) \to \mathcal{X}(\mathcal{M})$ satisfies for all $\eta, \tau, \xi \in \mathcal{X}(\mathcal{M})$, $f, g : \mathcal{M} \to \mathbb{R}$ smooth, and $a, b \in \mathbb{R}$,*

(i) $\nabla_{f\eta+g\tau}\xi = f\nabla_\eta\xi + g\nabla_\tau\xi$,

(ii) $\nabla_\eta(a\xi + b\tau) = a\nabla_\eta\xi + b\nabla_\tau\xi$,

(iii) $\nabla_\eta(f\xi) = (\eta f)\nabla_\eta\xi + f\nabla_\eta\xi$.

An affine connection $\nabla_\eta\xi$ mimicks a directional derivative of a vector field ξ along another vector field η. Given two vector fields $\eta, \xi \in \mathcal{X}(\mathcal{M})$ we can define its Lie-bracket $[\eta, \xi] \in \mathcal{X}(\mathcal{M})$, which acts on smooth functions $f : \mathcal{M} \to \mathbb{R}$ via $\eta(\xi f) - \xi(\eta f) : \mathcal{M} \to \mathbb{R}$. Check that this is indeed a vector field! An affine connection is called symmetric *if for all vector fields $\eta, \xi \in \mathcal{X}(\mathcal{M})$ we have that*

$$\nabla_\eta\xi - \nabla_\xi\eta = [\eta, \xi].$$

An affine connection on a Riemannian manifold is called compatible with the Riemannian metric, *if for any vector fields $\eta, \xi, \tau \in \mathcal{X}(\mathcal{M})$ we have*

$$\tau g(\eta, \xi) = g(\nabla_\tau\eta, \xi) + g(\eta, \nabla_\tau\xi).$$

A famous theorem by Levi-Civita states that, for a given Riemannian manifold there exists a unique affine connection which is symmetric and compatible. This connection is called the Riemannian connection and its associated derivative operator is called the covariant derivative on $\mathcal{M}$. Check that the the covariant derivative defined above for embedded Riemannian manifolds coincides with the Riemannian connection.

Lemma 1.25. *If γ is a geodesic with $\gamma(0) = p$ and $\dot{\gamma}(0) = \xi$, then the curve $\gamma_a(t) := \gamma(at)$ is also a geodesic with $\gamma_a(0) = p$ and $\dot{\gamma}_a(0) = a\xi \in T_p\mathcal{M}$. For every $p \in \mathcal{M}$ there exists $\varepsilon > 0$ such that for all $\xi \in T_p\mathcal{M}$ with $g_p(\xi, \xi)^{1/2} < \varepsilon$ there exists a unique geodesic γ with $\gamma(0) = p$ and $\dot{\gamma}(0) = \xi$ which is well-defined for all parameters $t \in [-2, 2]$.*

Proof. The first homogeneity assertion follows directly from the definition of the covariant derivative (second derivative + linear projection, but you can also verify this for non-embedded manifolds, using Remark 1.24). The second assertion follows from Picard's theorem together with the homogeneity assertion. More precisely, by Picard's theorem we know that there exist $\delta_1, \delta_2 > 0$ such that for all $\xi \in T_p\mathcal{M}$ with $g_p(\xi, \xi)^{1/2} < \delta_1$ the geodesic γ with $\dot{\gamma}(0) = \xi$ is well-defined for $t \in [-\delta_2, \delta_2]$. By the homogeneity property it now suffices to pick $\varepsilon < \delta_1\delta_2/2$. □

Definition 1.26. *Let p, ε as in Lemma 1.25. Then the* exponential map

$$\exp_p : B_\varepsilon(T_p\mathcal{M}) := \{\xi \in T_p\mathcal{M} : \; g_p(\xi, \xi)^{1/2} < \varepsilon\} \to \mathcal{M}$$

is defined as

$$\exp_p(\xi) = \gamma(1),$$

where γ is the unique geodesic with $\gamma(0) = p$ and $\dot{\gamma}(0) = \xi$.

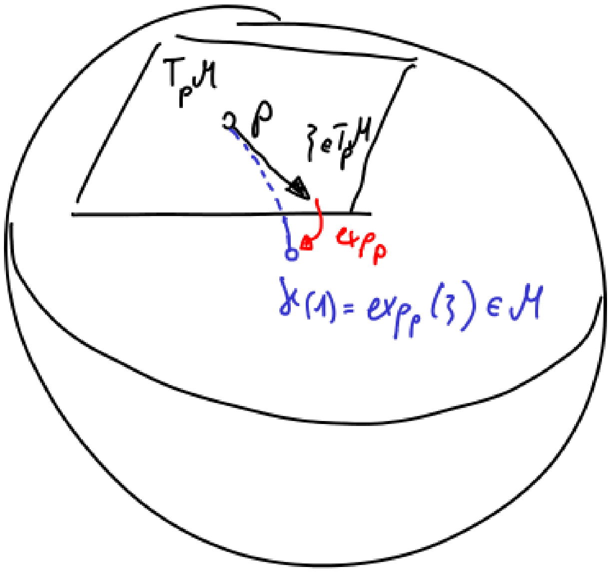

Figure 4: The exponential map. For $\xi \in T_p\mathbb{M}$, $\exp_p(\xi)$ is defined as $\gamma(1)$ where γ is the unique geodesic through p with tangent ξ at p. The exponential map can be regarded as a geometric version of 'point-vector-addition'.

Lemma 1.27. *For all $p \in \mathcal{M}$ there exists an open neighborhood U of $0 \in T_p\mathcal{M}$ and a neighborhood V of $p \in \mathcal{M}$ such that $\exp_p$ is a diffeomorphism from U to V. Its local inverse is denoted $\log_p : V \to U$ and called the* logarithm map *of $\mathcal{M}$.*

The exponential map is the right analogue of the notion of 'point-vector-addition' that we know from $\mathbb{R}^N$. Adding a vector $\xi \in T_p\mathcal{M}$ to a point $p \in \mathcal{M}$ amounts to computing $\exp_p(\xi)$. Its inverse is a geometric version of 'point-point-difference': the vector $\log_p(q) \in T_p\mathcal{M}$ is the vector 'pointing from p to q, see also Figure 4.

Proof. We will show that $D\exp_p|_0 = I$, where the differential D is defined as in Definition 1.15. The assertion then follows from the inverse function theorem. So let's look at $D\exp_p|0$. Since $\exp_p$ is defined on a linear space $T_p\mathcal{M}$ we can identify $T_0T_p\mathcal{M}$ with $T_p\mathcal{M}$ Let $s(t) := t\xi$ with $\xi \in T_p\mathcal{M}$. Then

$$D\exp_p|_0(\xi) = \frac{d}{dt}|_0 \exp(s(t)) = \frac{d}{dt}|_0 \exp(t\xi) = \xi,$$

by the definition of the exponential map. This shows the statement. □

We will now show that geodesics are locally length minimizing curves, as straight lines are in Euclidean space. More precisely:

Theorem 1.28. *Any shortest curve is a geodesic. Conversely, for each $p \in \mathcal{M}$ there exists $\varepsilon > 0$ such that for all $q \in \mathcal{M}$ with dist$(p,q) < \varepsilon$, the unique shortest geodesic between p and q is a length-minimizing. Moreover we have in this case that*

$$\textit{dist}(p,q) = g_p(\log_p(q), \log_p(q))^{1/2}. \tag{1}$$

Proof. We only sketch the argument. To compute the shortest curve γ connecting p, q we need to minimize $L(\gamma)$. However, L possesses too many symmetries. It is invariant under reparamatrization of γ, which makes the level sets of L noncompact. To remove these invariances (and make the functional L amemable to minimization), we need to break this symmetry. We do this by only considering curves $\gamma : [0,1] \to \mathcal{M}$ which are parametrized by unit-length, e.g.

$$g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t)) \equiv \text{const},$$

in which case we get for the minimizing curve that

$$\text{dist}(p, q) = L(\gamma) = g_p(\dot{\gamma}(0), \dot{\gamma}(0))^{1/2}.$$

Now observe that for unit-length parametrized curves as above we have

$$L(\gamma) = E(\gamma)^{1/2},$$

where

$$E(\gamma)^2 := \int_0^1 g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))dt.$$

Using calculus of variations (details TBA, but you can also do it; it is quite straightforward!) one sees that stationary points of E^2 are precisely geodesics, parameterized at unit length. This shows that any length-minimizing curve is a geodesic. To show the local converse we need to examine the second variation of E^2 which turns out to be positive definite if p, q are sufficiently close (TBA). This shows the statement. $\square$

Example 1.29. *For $\mathcal{M} = St(n, n) = O(n)$ we have that for $X \in \mathcal{M}$ and $X\Omega \in T_X\mathcal{M}$ with $\Omega^T = -\Omega$, that*

$$\exp_X(X\Omega) = X\exp(t\Omega),$$

where $\exp(A) := \sum_{i\geq 0} A^i/i!$ denotes the matrix exponential. For $X, Y \in \mathcal{M}$ we have that

$$\log_X(Y) = X\log(X^TY),$$

where $\log(A) := \sum_{i\geq 0}(I - A)^i/i$ denotes the matrix logarithm. Hence we have

$$\mathit{dist}(X, Y) = \|\log(X^TY)\|_F$$

with $\|\cdot\|_F$ the Frobenius norm.