Lecture 21

9.2 Smooth manifolds

Adapted from [11, Chapters 2.2.1, 2.2.2]

Most data sets are embedded in $\mathbb{R}^N$ for some N. Hence we restrict our treatment of smooth manifolds to those that are embedded in $\mathbb{R}^N$.

First we recall some definitions. Let $d, N \in \mathbb{N}$; for now they are arbitrary, although later $d < N$. Let $f : \mathbb{R}^d \to \mathbb{R}^N$ and let $a \in \mathbb{R}^d$. The function f is *differentiable at* $a \in \mathbb{R}^d$ if there is a linear continuous map, $\ell : \mathbb{R}^d \to \mathbb{R}^N$, and a function $\epsilon(h)$, such that

$$\forall\, a + h \in \mathbb{R}^d, \quad f(a+h) = f(a) + \ell(h) + \epsilon(h)\|h\|,$$

where

$$\lim_{h \to 0} \epsilon(h) = 0.$$

The linear map ℓ is denoted by $Df(a)$, Df_a, $df(a)$, df_a, or $f'(a)$, and is called the Fréchet derivative, the total derivative, or derivative, or total differential, or differential, of f at a. We will just call it the derivative of f, and denote it by $f'(a)$ or $Df(a)$.

The derivative $Df(a)$ is an abstract linear map between $\mathbb{R}^d$ and $\mathbb{R}^N$. However, in the standard bases of $\mathbb{R}^d$ and $\mathbb{R}^N$ we may represent it by the $N \times d$ Jacobian matrix $J(f)(a)$. Let

$$f(x) = (f_1(x), \ldots, f_N(x)), \quad x \in \mathbb{R}^d.$$

The Jacobian matrix is:

$$J(f)(a)_{ij} = \frac{\partial f_i}{\partial x_j}(a), \quad i = 1, \ldots, N, \;\; j = 1, \ldots, d.$$

The function f is *smooth* if for each of the component functions f_i, all partial derivatives of every order exist.

Let $A \subset \mathbb{R}^d$ and $B \subset \mathbb{R}^N$. A function $f : A \to B$ is a *homeomorphism* if:

1. f is a bijection

2. f is continous

3. f^{-1} is continuous

The function f is a C^∞-*diffeomorphism* if f is a smooth homeomorphism.

Definition 1. We are at last ready to define a smooth manifold. Now let $N \geq d \geq 1$. A *d-dimensional smooth manifold in* $\mathbb{R}^N$, or for short a *manifold*, is a nonempty subset $\mathcal{M} \subset \mathbb{R}^N$ such that for every point $p \in \mathcal{M}$ there are two open subsets $\Omega \subset \mathbb{R}^d$ and $U \subset \mathcal{M}$, with $p \in U$, and a function $\varphi : \Omega \to \mathbb{R}^N$ such that φ is a diffeomorphism between Ω and $U = \varphi(\Omega)$, and $D\varphi(a)$ is injective, where $a = \varphi^{-1}(p)$.

The function $\varphi : \Omega \to U$ is called a *(local) parameterization of* $\mathcal{M}$ *at* p. If $0 \in \Omega$ and $\varphi(0) = p$, we say that φ is centered at p.

Since $\varphi : \Omega \to U$ is a homeomorphism, it has an inverse $\varphi^{-1} : U \to \Omega$ that is also a homeomorphism. The map φ^{-1} is called a *local chart*. Since $\Omega \subset \mathbb{R}^d$, we have $\varphi^{-1}(p) = (z_1, \ldots, z_d) \in \mathbb{R}^d$, which we call the local coordinates of p. Intuitively a chart provides a "flattened" local map of a region on a manifold. For instance, in the case of surfaces (2-dimensional manifolds), a chart is analogous to a planar map of a region on the surface. For a concrete example, consider a map giving a planar representation of a country, a region on the earth, a curved surface.

Example 1. The unit circle $S^1 \subset \mathbb{R}^2$ defined as

$$S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$$

is a smooth 1-manifold. Intuitively, for each point $p \in S^2$, there is a local neighborhood around p which is diffeomorphic to a subset of $\mathbb{R}$. We can write down precisely two charts φ_N and φ_S for S^1 to make this precise. Let $N = (0, 1)$ be the "north pole" and $S = (0, -1)$ the "south pole," and set $U_N = S^1 \setminus \{N\}$ and $U_S = S^1 \setminus \{S\}$. Define the maps $\varphi_N^{-1} : U_N \to \mathbb{R}$ and $\varphi_S^{-1} : U_S \to \mathbb{R}$ as

$$\varphi_N^{-1}(x, y) = \frac{1}{1 - y} x$$
$$\varphi_S^{-1}(x, y) = \frac{1}{1 + y} x$$

The above maps are called the stereographic projections from the north/south pole, respectively. With these definitions, the maps $\varphi_N : \mathbb{R} \to U_N$ and

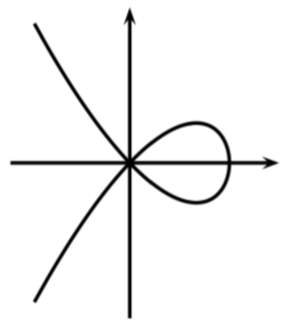

Figure 26: Example of a non-manifold.

$\varphi_S : \mathbb{R} \to U_S$ are given by:

$$\varphi_N(x) = \frac{1}{x^2+1}(2x, x^2-1)$$
$$\varphi_S(x) = \frac{1}{x^2+1}(2x, 1-x^2)$$

One can verify that these two maps satisfy the conditions of a smooth manifold. The chart φ_N^{-1} assigns local coordinates to points in the "lower half" of S^1 (all of U_N), while φ_S^{-1} assigns local coordinates to points in the "upper half" of S^1 (all of U_S). The chart φ_N is centered at S, while the chart φ_S is centered at N, since $\varphi_N(S) = 0$ and $\varphi_S(N) = 0$.

Example 2. For some clarity, let's now look at a non-example. Consider the curve in $\mathbb{R}^2$ given by the zero locus (i.e., the zeros) of the equation:

$$y^2 = x^2 - x^3,$$

which is:

$$\mathcal{X} = \{(x, y) \in \mathbb{R}^2 : y^2 = x^2 - x^3\}.$$

The data set $\mathcal{X}$ is a curve, which is plotted in Figure 26. The curve $\mathcal{X}$ can be parameterized as:

$$\forall\, t \in \mathbb{R}, \quad x(t) = 1 - t^2$$
$$y(t) = t(1 - t^2)$$

We claim that $\mathcal{X}$ is not a manifold. The problem is that the curve has a self intersection at the origin. Intuitively, the local neighborhood around

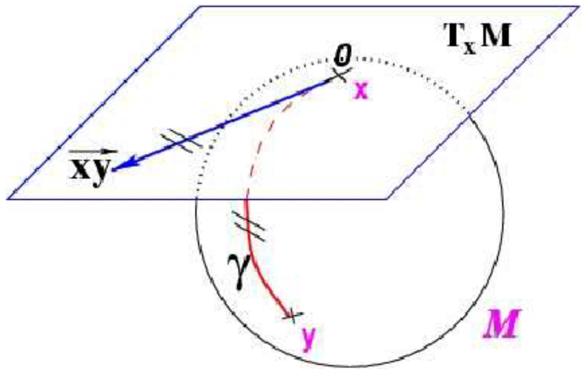

Figure 27: S^2 manifold with tangent space and geodesic.

$(0,0)$ is not like $\mathbb{R}$. To make this more precise, if $\mathcal{X}$ where a manifold then by definition there would exist an open subset $\Omega \subset \mathbb{R}$ and $U \subset \mathcal{X}$, with $(0,0) \in U$, and a local parameterization $\varphi : \Omega \to U$ such that φ is a diffeomorphism. However, $U \setminus \{(0,0)\}$ must necessarily consist of four separate connected components, while $\Omega \setminus \{\varphi^{-1}(0,0)\}$ consists of two connected components. So φ is not even a homeomorphism, let alone a diffeomorphism.

We now define the notion of a tangent space of manifold. Let $p \in \mathcal{M}$, the tangent space, denote $T_p\mathcal{M}$, will vary with p. It is a generalization of the notion of a tangent to a one dimensional curve $y = f(x)$. Figure 27 gives an illustration for $\mathcal{M} = S^2$, the 2-sphere.

Definition 2. Let $\mathcal{M} \subset \mathbb{R}^N$ be a d-dimensional manifold, $p \in \mathcal{M}$, $U \subset \mathcal{M}$ with $p \in U$, $\Omega \subset \mathbb{R}^d$, and $\varphi : \Omega \to U$ a parameterization of $\mathcal{M}$ at p with $\varphi(a) = p$. The *tangent space of $\mathcal{M}$ at p*, denoted $T_p\mathcal{M}$, is defined as the image of $D\varphi(a) : \mathbb{R}^d \to \mathbb{R}^N$:

$$T_p\mathcal{M} = \text{image}\, D\varphi(a).$$

A vector $X \in T_p\mathcal{M}$ is called a tangent vector.

We can identify $D\varphi(a)$ with its $N \times d$ Jacobian matrix:

$$D\varphi(a) \sim J(\varphi)(a)_{ij} = \frac{\partial \varphi_i}{\partial x_j}(a),$$

where $\varphi = (\varphi_1, \ldots, \varphi_N)$. Define

$$\frac{\partial \varphi}{\partial x_j}(a) = \left(\frac{\partial \varphi_1}{\partial x_j}(a), \ldots, \frac{\partial \varphi_N}{\partial x_j}(a) \right)^T \in \mathbb{R}^N.$$

Figure 28: The two dimensional torus, $\mathbb{T}^2$

It follows that the set

$$\left\{ \frac{\partial \varphi}{\partial x_1}(a), \ldots, \frac{\partial \varphi}{\partial x_d}(a) \right\}$$

is a basis for $T_p\mathcal{M}$. Thus any $X \in T_p\mathcal{M}$ can be represented uniquely as:

$$X = \sum_{j=1}^{d} \alpha_j \frac{\partial \varphi}{\partial x_j}(a) = J(\varphi)(a)\alpha, \quad \alpha = (\alpha_1, \ldots, \alpha_d)^T.$$

If $\alpha \neq 0$, it represents a direction in $\mathbb{R}^d$ and X is a direction in the tangent space $T_p\mathcal{M}$.

Exercise 34 (optional). Define the d-dimensional torus, denote $\mathbb{T}^d$, as:

$$\mathbb{T}^d = \underbrace{S^1 \times \cdots \times S^1}_{d \text{ times}} \subset \mathbb{R}^{2d}.$$

See Figure 28 for a picture of $\mathbb{T}^2$. Prove that $\mathbb{T}^d$ is a d-dimensional smooth manifold.

References

[1] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation and Machine Learning. The MIT Press, 2002.

[2] Afonso S. Bandeira. Ten lectures and forty-two open problems in the mathematics of data science. MIT course *Topics in Mathematics of Data Science*, 2015.

[3] Jon Shlens. A tutorial on principal component analysis. arXiv:1404.1100, 2014.

[4] Karl Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, Series 6*, 2(11):559–572, 1901.

[5] V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72(114):507–536, 1967.

[6] J. Baik, G. Ben-Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

[7] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

[8] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer-Verlag New York, 2nd edition, 2009.

[9] Athanasios Tsanas and Angeliki Xifara. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, 49:560–567, 2012.

[10] Daniel A. Spielman. Spectral graph theory. *Yale Course Notes*, Fall, 2009.

[11] Jianzhong Wang. *Geometric Structure of High-Dimensional Data and Dimensionality Reduction*. Springer, 2012.

[12] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[13] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21:5–30, 2006.