Lecture 22

9.3 Riemannian manifolds

Adapted from [11, Chapters 2.2.3, 2.2.4]

A *Riemannian manifold* $(\mathcal{M}, g)$ is a smooth manifold $\mathcal{M}$ equipped with an inner product $\langle \cdot, \cdot \rangle_p$ on the tangent space $T_p\mathcal{M}$ at each point $p \in \mathcal{M}$ that varies smoothly from point to point in the sense that if X, Y are vector fields on $\mathcal{M}$, then $p \mapsto \langle X(p), Y(p) \rangle_p$ is a smooth function. As we shall see below, the inner product will be defined in terms of $g(p)$, which is called the *Riemannian metric tensor*.

A Riemannian metric (tensor) makes it possible to define various geometric notions on a Riemannian manifold, such as angles, lengths of curves, areas (or volumes), curvature, gradients of functions and divergence of vector fields. In other words, it allows us to do calculus on manifolds!

We now make this more precise using what we have developed in the previous section. Recall that the Jacobian matrix $J(\varphi)(a)$ is the matrix representation of the derivative $D\varphi(a)$ of the parameterization φ of a point $p \in \mathcal{M}$ (where $\varphi(a) = p$) in the standard bases of $\mathbb{R}^d$ and $\mathbb{R}^N$. The Riemannian metric $g(p)$ is defined as:

$$g(p) = J(\varphi)(a)^T J(\varphi)(a), \quad \varphi(a) = p \in \mathcal{M}.$$

The metric $g(p)$ is a $d \times d$ positive definite matrix, whose entries vary smoothly in p. Let $X, Y \in T_p\mathcal{M}$ be represented by

$$\begin{aligned} X &= J(\varphi)(a)x, \quad x \in \mathbb{R}^d, \\ Y &= J(\varphi)(a)y, \quad y \in \mathbb{R}^d. \end{aligned}$$

The vectors $x, y \in \mathbb{R}^d$ are the local coordinates of the tangent vectors X, Y, which we introduced at the end of the last lecture. We are going to use the local coordinate representations to define the inner product $\langle X, Y \rangle_p$, and many subsequent quantities. To that end, we define:

$$\langle X, Y \rangle_p = x^T g(p) y.$$

The inner product defines the norm of $X \in T_p\mathcal{M}$ by:

$$\|X\|_p^2 = x^T g(p) x.$$

Using the inner product $\langle \cdot, \cdot \rangle_p$ we can define distances on connected Riemannian manifolds. Let $\gamma : [0,1] \to \mathcal{M}$ be a parameterized curve in $\mathcal{M}$ going from p to q:

$$\forall\, t \in [0,1], \quad \gamma(t) = (\gamma_1(t), \ldots, \gamma_N(t)) \in \mathcal{M}, \quad \gamma(0) = p, \;\; \gamma(1) = q.$$

Suppose that γ is differentiable,

$$\gamma'(t) = (\gamma_1'(t), \ldots, \gamma_N'(t)).$$

Then $\gamma'(t) \in T_r\mathcal{M}$, where $\gamma(t) = r$. The length of γ, written as $\operatorname{len}(\gamma)$, is defined as:

$$\operatorname{len}(\gamma) = \int_0^1 \|\gamma'(t)\|_{\gamma(t)}\, dt.$$

Notice that this generalizes the notion of arc length from calculus. Indeed, if $h : [0,1] \to \mathbb{R}^d$ is a curve in $\mathbb{R}^d$, then the length of h is given by:

$$\operatorname{len}(h) = \int_0^1 \|h'(t)\|\, dt,$$

where $\|h'(t)\|$ is the standard Euclidean norm of $h'(t)$. Now let's return to the manifold setting. Let $\Gamma(p,q)$ denote the set of all differentiable curves connecting p and q. We then define the *(geodesic) distance* on $\mathcal{M}$, denoted $d_{\mathcal{M}}(p,q)$, as:

$$d_{\mathcal{M}}(p,q) = \inf_{\gamma \in \Gamma(p,q)} \operatorname{len}(\gamma).$$

It is the length of the shortest curve in $\mathcal{M}$ connecting p and q; see again Figure 27.

9.4 The Laplace-Beltrami operator

Adapted from [11, Chapter 2.3.2]

Now let $f : \mathcal{M} \to \mathbb{R}$ be a function on the Riemannian manifold $\mathcal{M}$. We are going to define the analogue of the Laplacian applied to f, which is called the Laplace-Beltrami operator. To do so, we first define the gradient and divergence on Riemannian manifolds. As before, we do not do things in full generality and we simplify some concepts, which hopefully gives some additional clarity.

We begin by recalling the situation in $\mathbb{R}^d$. Let $f : \mathbb{R}^d \to \mathbb{R}$ and let $v \in \mathbb{R}^d$ be a vector. The direction derivative of f in the direction v is defined as:

$$\partial_v f(x) = \lim_{h \to 0} \frac{f(x+hv) - f(x)}{h}.$$

Recall as well that the gradient of f is defined as:

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_d}(x) \right).$$

The gradient vector $\nabla f(x)$ is the unique vector such that

$$\partial_v f(x) = \nabla f(x) \cdot v,$$

where $\nabla f(x) \cdot v$ is the standard inner (dot) product in $\mathbb{R}^d$.

Let us now generalize these ideas to Riemannian manifolds $\mathcal{M}$. Let $f : \mathcal{M} \to \mathbb{R}$, $p \in \mathcal{M}$, and $X \in T_p\mathcal{M}$. To define the directional derivative of f in the direction X, let $\gamma : [-1, 1] \to \mathcal{M}$ be a differentiable curve with $\gamma(0) = p$ and $\gamma'(0) = X$. Then the directional derivative is defined as:

$$\partial_X f(p) = \frac{d}{dt} f \circ \gamma(t) \bigg|_{t=0}$$

The definition is independent of γ, so long as $\gamma(0) = p$ and $\gamma'(0) = X$.

A vector field on $\mathcal{M}$ is a vector valued function $X(p) \in T_p\mathcal{M}$ mapping $\mathcal{M}$ into its tanget space. The gradient ∇f is defined as the unique vector field over $\mathcal{M}$ such that

$$\forall\, X \in T_p\mathcal{M}, \quad \langle \nabla f(p), X \rangle_p = \partial_X f(p).$$

One can write $\nabla f(p)$ explicitly in terms of local coordinates and the metric tensor g. We do not do so here since we will not use it.

Now we generalize the notion of divergence. For $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$, with $\mathbf{F}(x) = (F_1(x), \ldots, F_d(x))$, recall that the divergence is defined as:

$$\operatorname{div} \mathbf{F}(x) = \nabla \cdot \mathbf{F}(x) = \frac{\partial F_1}{\partial x_1}(x) + \cdots + \frac{\partial F_d}{\partial x_d}(x).$$

Recall that the Jacobian matrix of $\mathbf{F}$ is:

$$J(\mathbf{F})(x) = \frac{\partial F_i}{\partial x_j}(x), \quad i, j = 1, \ldots, d.$$

Notice that:

$$\operatorname{div} \mathbf{F}(x) = \operatorname{Tr}\Big[J(\mathbf{F})(x)\Big],$$

i.e., the trace of the Jacobian matrix. Notice as well that the i^{th} row of the Jacobian matrix is $\nabla F_i(x)$. Considering $J(\mathbf{F})(x)$ as a linear operator mapping $\mathbb{R}^d$ to $\mathbb{R}^d$, we have for $v \in \mathbb{R}^d$:

$$J(\mathbf{F})(x)v = (\nabla F_1(x) \cdot v, \ldots, \nabla F_d(x) \cdot v) = (\partial_v F_1(x), \ldots, \partial_v F_d(x)) = \partial_v \mathbf{F}(x).$$

Thus $J(\mathbf{F})(x)$ is the linear map that takes v to $\partial_v \mathbf{F}(x)$, and the divergence is the trace of this linear map. This is how we will define the divergence on a Riemannian manifold $\mathcal{M}$.

To that end once again let $p \in \mathcal{M}$ be a point on the manifold, $X \in T_p\mathcal{M}$ a tangent vector at p, and $Y(p) \in T_p\mathcal{M}$ a vector field over $\mathcal{M}$. Note that $Y(p) = (Y_1(p), \ldots, Y_N(p))$, and define the linear map $M(p, Y) : T_p\mathcal{M} \to \mathbb{R}^N$ as:

$$M(p, Y)X = \partial_X Y(p) = (\partial_X Y_1(p), \ldots, \partial_X Y_N(p)).$$

The divergence of Y is then defined as:

$$\operatorname{div} Y(p) = \operatorname{Tr}\big[M(p, Y)\big]$$

With the generalizations of the gradient and divergence to Riemannian manifolds in hand, we define the *Laplace-Beltrami* operator Δ as:

$$\Delta f(p) = \operatorname{div} \nabla f(p),$$

which mirrors exactly the definition of the Laplacian on Euclidean space.

9.5 Integration on Riemannian manifolds

Adapted from [11, Chapter 2.3.2]

The final ingredient we will need is how to define integration on Riemannian manifolds $\mathcal{M}$. We make the greatly simplifying assumption that there exists a single local parameterization $\varphi : \mathcal{M} \to \mathbb{R}^d$, which in general is not true and is not necessary to define integration, but greatly simplifies the presentation while still giving the idea of how to define integration over general Riemannian manifolds.

Let $g(p)$ be the corresponding Riemannian metric. Define $|g(p)| = \det g(p) > 0$. The volume of $\mathcal{M}$ is defined as:

$$\mathrm{Vol}(\mathcal{M}) = \int_{\varphi(\mathcal{M})} \sqrt{|g(\varphi^{-1}(x))|}\, dx.$$

The definition of volume is independent of the parameterization φ. The integral of $f : \mathcal{M} \to \mathbb{R}$ is then defined as:

$$\int_{\mathcal{M}} f(p)\, dV(p) = \int_{\varphi(\mathcal{M})} f(\varphi^{-1}(x)) \sqrt{|g(\varphi^{-1}(x))|}\, dx.$$

The space $\mathbf{L}^2(\mathcal{M})$ is defined as those functions $f : \mathcal{M} \to \mathbb{R}$ for which

$$\|f\|^2 = \int_{\mathcal{M}} |f(p)|^2\, dV(p) < \infty.$$

It is a Hilbert space equipped with the inner product:

$$\langle f, h \rangle = \int_{\mathcal{M}} f(p)h(p)\, dV(p).$$

We remark that the following standard formulas hold for manifolds as well:

$$\begin{aligned} \int_{\mathcal{M}} f(p)\Delta h(p)\, dV(p) &= \int_{\mathcal{M}} h(p)\Delta f(p)\, dV(p) \\ \int_{\mathcal{M}} \|\nabla f(p)\|_p^2\, dV(p) &= -\int_{\mathcal{M}} f(p)\Delta f(p)\, dV(p) \end{aligned} \tag{61}$$

As writing the variable p is often cumbersome, we will often simply write

$$\int_{\mathcal{M}} f = \int_{\mathcal{M}} f(p)\, dV(p)$$

for the integral of f over $\mathcal{M}$. The formulas above, for example, would then be written as $\int_{\mathcal{M}} f\Delta h = \int_{\mathcal{M}} h\Delta f$ and $\int_{\mathcal{M}} \|\nabla f\|^2 = -\int_{\mathcal{M}} f\Delta f$.

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.