Lecture 20

9 Manifold learning

We are now going to add geometry on top of the spectral graph theory developed over the previous lectures. To do so we introduce the notion of a smooth manifold, which provides a mathematical framework for us to work with data sampled from complex geometric shapes. The Laplacian will once again play a fundamental role, and the coupling of graph theory and geometry will allow us to interpret the graph Laplacian as a discrete approximation of the continous Laplacian operator over a manifold. This interpretation will imply that the eigenvectors and eigenvalues of the graph Laplacians that we compute characterize the underlying manifold that our data is sampled from.

9.1 The Laplacian and the heat equation

Adapted from [11, Chapter 2.1.3]

Manifold learning is intimately related to heat diffusion. We review the heat equation on $\mathbb{R}^d$ now.

Let $\Omega \subset \mathbb{R}^d$ be an open, bounded domain. The Laplace operator is defined as:

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}.$$

Recall that the gradient of a function $f : \mathbb{R}^d \to \mathbb{R}$ 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 $\nabla f : \mathbb{R}^d \to \mathbb{R}^d$ defines a vector field over $\mathbb{R}^d$. Let $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$ be a general vector field,

$$\mathbf{F}(x) = (F_1(x), \ldots, F_d(x)).$$

The divergence of $\mathbf{F}$ is:

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

It follows that the Laplace operator can be defined as:

$$\Delta f(x) = \operatorname{div} \nabla f(x).$$

Define the related Laplacian operator $\mathcal{L}$ as

$$\mathcal{L} = -\Delta.$$

The closure of Ω is the closed domain

$$\overline{\Omega} = \Omega \cup \partial\Omega,$$

where $\partial\Omega$ is the boundary of Ω and is simply defined as the set difference $\overline{\Omega} \setminus \Omega$. The Laplacian eigenvalue problem on $\overline{\Omega}$ is

$$\mathcal{L} f(x) = \lambda f(x), \quad x \in \Omega,$$

with one of the following boundary conditions:

$$\begin{aligned} f(x) &= 0, \quad x \in \partial\Omega, \quad \text{(Dirichlet)} \\ \frac{\partial f}{\partial \mathbf{n}}(x) &= 0, \quad x \in \partial\Omega, \quad \text{(Neumann)} \end{aligned}$$

where $\mathbf{n}$ is the normal to the boundary $\partial\Omega$. We will consider Neumann boundary conditions, unless otherwise stated.

It is known that $\mathcal{L}$ is a self-adjoint operator. Since Ω is bounded, $\mathcal{L}$ has a countable set of eigenvalues, and each eigenvalue has finite multiplicity. Additionally, all eigenvalues of $\mathcal{L}$ are non-negative; order them with multiplicity as:

$$0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k \leq \cdots \to \infty.$$

Let the corresponding eigenfunctions be denoted by $\phi_k : \mathbb{R}^d \to \mathbb{R}$, so that

$$\mathcal{L}\phi_k(x) = \lambda_k \phi_k(x).$$

With Neumann boundary conditions, $\lambda_1 = 0$ and $\phi_1(x) \equiv C$ is a constant function assuming Ω is connected. Notice the similarity with the graph Laplacian.

The collection $\{\phi_k\}_{k=1}^{\infty}$ forms an ONB for $\mathbf{L}^2(\Omega)$, so that:

$$\forall\, f \in \mathbf{L}^2(\Omega), \quad f = \sum_{k=1}^{\infty} \langle f, \phi_k \rangle \phi_k.$$

Since $\langle \phi_1, \phi_k \rangle = 0$ for $k \geq 2$ and ϕ_1 is constant, we see that

$$\forall\, k \geq 2, \quad \int_\Omega \phi_k(x)\, dx = 0.$$

Thus, again paralleling the graph treatment, we see that ϕ_k are oscillating functions, which can be interpreted as Fourier modes for the domain Ω. The larger λ_k, the more ϕ_k oscillates over Ω.

Let $u(x,t)$ be a function defined on $\overline{\Omega} \times [0,\infty)$, with $x \in \overline{\Omega}$ and time $t \in [0,\infty)$. Let

$$u_t = \frac{\partial u}{\partial t}.$$

Consider the heat equation with Neumann boundary condition:

$$\begin{aligned} u_t &= \Delta u, && \text{in } \Omega \times (0,\infty), \\ \frac{\partial u}{\partial \mathbf{n}} &= 0, && \text{in } \partial\Omega \times (0,\infty), \\ u(x,0) &= f(x), && \text{in } \Omega. \end{aligned}$$

Given the initial solution f, the solution u can be computed through the fundamental solution, otherwise known as the heat kernel. The heat kernel $K(x,y;t)$ satisfies

$$\begin{aligned} K_t &= \Delta_x K, \\ \lim_{t \to 0} K(x,y;t) &= \delta(x-y). \end{aligned}$$

With the heat kernel K, the solution u can be computed as:

$$u(x,t) = \int_\Omega K(x,y;t) f(y)\, dy.$$

The heat kernel can be further related to Δ through the exponential formula, which states:

$$K(x,y;t) = \sum_{k=1}^{\infty} e^{-\lambda_k t} \phi_k(x) \phi_k(y), \quad t \in (0,\infty), \;\; x,y \in \overline{\Omega}.$$

It follows that:

$$u(x,t) = \sum_{k=1}^{\infty} e^{-t\lambda_k} \langle f, \phi_k \rangle \phi_k(x).$$

We thus define the integral operator:

$$e^{-t\Delta}f(x) = \int_{\Omega} K(x,y;t)f(y)\,dy,$$

and will at times call $e^{-t\Delta}$ the heat kernel.

Exercise 31 (optional). Let $A : \mathbf{L}^2(\overline{\Omega}) \to \mathbf{L}^2(\overline{\Omega})$ be a positive semi-definite operator acting on square integrable functions $f : \overline{\Omega} \to \mathbb{R}$. Define a new operator $e^A : \mathbf{L}^2(\overline{\Omega}) \to \mathbf{L}^2(\overline{\Omega})$ as:

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

Prove that if λ is an eigenvalue of A with eigenfunction $\varphi \in \mathbf{L}^2(\overline{\Omega})$, then e^λ is an eigenvalue of e^A with the same eigenfunction φ.

Exercise 32 (optional). Using the definition in the previous exercise, prove that $u(x,t) = e^{t\Delta}f(x)$ is the solution to the heat equation with initial condition $u(x,0) = f(x)$. In other words, prove

$$\frac{\partial}{\partial t}\left(e^{t\Delta}f\right)(x) = \Delta\left(e^{t\Delta}f\right)(x)$$

Exercise 33 (optional). Using Mercer's Theorem and the fact that the heat kernel is continuous, conclude from the previous two exercises that the heat kernel can be written as:

$$K(x,y;t) = \sum_{k=1}^{\infty} e^{-\lambda_k t}\phi_k(x)\phi_k(y), \quad t \in (0,\infty), \;\; x,y \in \overline{\Omega}.$$

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.