POLITECNICO MILANO 1863

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Master Thesis in Computational Science and Engineering

About the convergence of the Graph Laplacian

Martino Milani

February 26, 2019

Abstract

Problem definition: find a convergence result for the Graph Laplacian on a graph approximating a sphere (WHICH GRAPH? Because for a full graph with $W(i,j) = e^{||x_i - x_j||^2}$ we already have it) to the Laplace-Beltrami operator on the sphere, given a deterministic sampling (HealPix).

Contents

1 Introductive Study

1.1 The Graph Laplacian (follwing Belkin and Niyogi)

In summary, the heat equation is indeed the key to approximating the Laplace operator. On can observe that given the heat equation

$$\frac{\partial}{\partial t}u(\mathbf{x},t) - \triangle u(\mathbf{x},t) = 0 \tag{1}$$

the corresponding solution, for an initial heat distribution $f(\mathbf{x})$ is given by $\mathbf{H}^t f(\mathbf{x})$ where $\mathbf{H}^t f(\mathbf{x})$ is the heat kernel convolution operator:

$$\mathbf{H}^t f(\mathbf{x}) = \int_{\mathcal{R}^k} f(\mathbf{y}) H^t(\mathbf{x},\mathbf{y}) d\mathbf{y}$$

$$H^t(\mathbf{x},\mathbf{y}) = (4\pi t)^{-\frac{k}{2}} e^{-\frac{||\mathbf{x}-\mathbf{y}||^2}{4t}}$$

Furthermore, it can be proved that

$$\lim_{t\to 0} \mathbf{H}^t f(\mathbf{x}) = f(\mathbf{x})$$

So, the Laplacian of the initial distribution f can be written in the following way:

$$-\triangle f(\mathbf{x}) = -\frac{\partial}{\partial t}\mathbf{H}^t f(\mathbf{x})|_{t=0} \tag{2}$$

Rewriting the rightmost term we obtain

$$-\frac{\partial}{\partial t}\mathbf{H}^t f(\mathbf{x})|_{t=0} = \lim_{t\to 0}\frac{1}{t}\left(\int_{\mathbb{R}^k} f(\mathbf{y}) H^t(\mathbf{x},\mathbf{y}) d\mathbf{y} - f(\mathbf{x})\right)$$

And thus the Laplacian can be written as follows:

$$-\triangle f(\mathbf{x}) = \lim_{t\to 0}\frac{1}{t}\left((4\pi t)^{-\frac{k}{2}} \int_{\mathbb{R}^k} f(\mathbf{y}) e^{-\frac{||\mathbf{x}-\mathbf{y}||^2}{4t}} d\mathbf{y} - f(\mathbf{x})(4\pi t)^{-\frac{k}{2}} \int_{\mathbb{R}^k} e^{-\frac{||\mathbf{x}-\mathbf{y}||^2}{4t}} d\mathbf{y}\right)$$

Writing the discrete version of the integrals involved using the point cloud $\mathcal{S}$ we have

$$\hat{\triangle}_{\mathcal{S}} f(\mathbf{x}) = \frac{1}{t}\frac{(4\pi t)^{-\frac{k}{2}}}{n}\left(f(\mathbf{x})\sum_i e^{-\frac{||\mathbf{x}_i-\mathbf{x}||^2}{4t}} - \sum_i f(\mathbf{x}_i) e^{-\frac{||\mathbf{x}_i-\mathbf{p}||^2}{4t}}\right) \tag{3}$$

We are now ready to define the Graph Laplacian:
Definition 1.1. [*Graph Laplacian*]
For a graph $G = (V,E)$ defined by the weight matrix W we define the Graph Laplacian L

$$L = D - W$$

Where D is a diagonal matrix such that $D(i,i) = \sum_j W(i,j)$

By defining a *full graph* on the point cloud $\mathcal{S}$ through the special weight matrix $W_n^t = e^{-\frac{||x_i - x_j||^2}{4t}}$ we can write the Laplacian on such particular graph as follows:

$$\mathbf{L}_n^t f(x) = f(x) \sum_j e^{-\frac{||x-x_j||^2}{4t}} - \sum_j f(x_j) e^{-\frac{||x-x_j||^2}{4t}}$$

Putting eq. 3 and definition 1.1 together we arrive at the following equation, that

$$\hat{\triangle}_{\mathcal{S}} f(\mathbf{x}) = \frac{(4\pi t)^{-\frac{k}{2}}}{n} \mathbf{L}_n^t(f)(\mathbf{x})$$

1.2 Towards a Theoretical Foundation for Laplacian-Based Manifold Methods (Belkin & Niyogi, 2005)

In this paper it is proved that the *random* graph Laplacian defined with the special weight matrix $W_n^t = e^{-\frac{||x_i - x_j||^2}{4t}}$ of a random point cloud converges to the Laplace-Beltrami operator on a general(!!) manifold $\mathcal{M}$ *in probability.*

Given a set of points $\mathcal{S}_n = \{x_1, ..., x_n\} \subset \mathbb{R}^k$, the weight matrix is set to be

$$W_n^t(i,j) = e^{-\frac{||x_i - x_j||^2}{4t}}$$

This weight matrix is set to be this way in analogy with the Heat Kernel on $\mathbb{R}^k$.

The main two results presented in such paper are the following:
Theorem 1.1. [*Convergence of the Random Graph Laplacian*]
Let data points $\{x_1, ..., x_n\}$ in $\mathbb{R}^N$ be sampled form a uniform distribution on a manifold $\mathcal{M} \subset \mathbb{R}^N$.
Put $t_n = n^{-\frac{1}{k+2+\alpha}}$, where $\alpha > 0$ and let $f \in C^\infty(\mathcal{M})$.
Then there is a consant C such that *in probability*

$$\lim_{n\to\infty} C \frac{(4\pi t_n)^{-\frac{k+2}{2}}}{n} L_n^{t_n} f(x) = \triangle_{\mathcal{M}} f(x)$$

Theorem 1.2. [*Uniform Convergence of the Random Graph Laplacian*]
Let data points $\{x_1, ..., x_n\}$ in $\mathbb{R}^N$ be sampled form a uniform distribution on a compact manifold $\mathcal{M} \subset \mathbb{R}^N$. Take the space $\mathcal{F} = \{f \in C^\infty, \triangle f \text{is Lipschitz}\}$. Then there exists a sequence of numbers $t_n \to 0$ and a constant C such that in probability

$$\lim_{n\to\infty} \sup_{x\in\mathcal{M}, f\in\mathcal{F}} \left| C \frac{(4\pi t_n)^{-\frac{k+2}{2}}}{n} L_n^{t_n} f(x) - \triangle_{\mathcal{M}} f(x) \right| = 0$$

Differences with what we want to prove

1. Our sampling is deterministic: HealPix. Plus, the graph is not full. Deferrard & Perraudin use the same weight scheme as Belking & Nyiogi, connecting the 8 (or 7^4) neighboring pixels in the HEALpix hierarchy. However, in Belking & Nyiogi the graph is FULL. However, as the sampling increases, maybe we can show convergence of some quantity to the integral of equation (9), and then with a plug and play the whole paper works!
2. Our graph is not fully connected

Questions

1. How come that the Laplacian defined on our graph actually seems to act as the Laplacian of Belkin & Nyiogi? **Answer:**

1.3 Computing Fourier Transforms and Convolutions on the 2-Sphere (Driscoll and Healy, 1994)

If we find a basis of minimal subspaces invariant (a vector space of functions on the sphere is invariant if all of the operators $\Lambda(g), g \in SO(3)$ take each function in the space back into the space) under all the rotations of $SO(3)$, then we simplify a lot the analysis of rotation-invariant operators.

Things to keep in mind from section 2, "Preliminaries"

1. any rotation $g \in SO(3)$ can be written in the well-known Euler angle decomposition: $g = u(\phi)a(\theta)u(\psi)$ determined uniquely for almost all g. Remember that any point on the sphere
2. $\omega(\theta, \phi) = (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)$. In fact, the 2-sphere is a quotient of the rotation group $SO(3)$ and inherits its natural coordinate system from that of the group.
3. $\Lambda(g)f(\omega) = f(g^{-1}\omega)$
4. invariant volume measure on $SO(3)$ is $dg = \sin\theta \; d\theta \; d\phi \; d\psi$, invariant volume measure on the sphere is $d\omega = \sin\theta \; d\theta \; d\phi$
5. The invariant subspace of degree l harmonic polynomials restricted to the sphere is called the space of *spherical harmonics of degree l*. Spherical harmonics of different degree are orthogonal to one another
6. In coordinates,
$$Y_l^m(\theta, \phi) = (-1)^m \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos\theta) e^{im\phi}$$
where P_l^m are Legendre functions.

7. Of all the possible basis for $L^2(S^2)$ the spherical harmonics uniquely exploit the symmetries of the sphere. Under a rotation$g$, each spherical harmonic of degree l is transformed into a linear combination of only those spherical harmonics of same degree l. Thus the effect of a rotation on a function expressed in the basis of the spherical harmonics is a multiplication by a semi-infinite block-diagonal matrix with the $(2l+1) \times (2l+1)$ blocks for each $l \geq 0$ given by
$$D^{(l)}(g) = \left(D^{(l)}_{m,n}\right)(g) = \left(D^{(l)}_{m,n}\right)(u(\phi)a(\theta)u(\psi)) = e^{-im\psi}d^{(l)}_{m,n}(\cos\theta)e^{-in\phi}$$
The effect of all of this is to block-diagonalize rotationally invariant operators; namely, convolution operators obtained as weighted averages of the rotation operators by functions or kernels. For example the Laplace-Beltrami operator, that acts diagonally on the spherical harmonic basis.

8.
Definition 1.2. [*Left Convolution*]
$k \star f(\omega) = \left(\int_{g\in SO(3)} dg\ k(g\eta)\Lambda(g)\right) f(\omega) = \int_{g\in SO(3)} k(g\eta)f(g^{-1}\omega)dg$ where η is the north pole.

Since the convolution is a linear combination of rotation operators $\Lambda(g)$, it follows that also the convolution must be block diagonalized. Indeed,
$$(f \hat{\star} h)(l,m) = 2\pi\sqrt{\frac{4\pi}{2l+1}}\hat{f}(l,m)\hat{h}(l,0)$$

9. We desire the ability to sample a band-limited function on the sphere so that the integrals defining the Fourier coefficients can be efficiently evaluated as weighted sums of the samples. In this paper the authors present a sampling result for band-limited functions that can exactly recover the original function obtaining its transform as weighted sum of the sampled values of that function, **given an equiangular sampling of the sphere**. This is stated in the following theorem:
Theorem 1.3. [*Shannon on the sphere*]
Let $f(\theta,\phi)$ be a band-limited function on S^2 such that $\hat{f}(l,m) = 0$ for $l \geq b$. Then
$$\hat{f}(m,l) = \frac{\sqrt{2\pi}}{2b}\sum_{j=0}^{2b-1}\sum_{k=0}^{2b-1} a_j^{(b)} f(\theta_j,\phi_k)\bar{Y}_l^m(\theta_j,\phi_k) \tag{4}$$
for $l \leq b, |m| \leq l$. Here $\theta_j = \frac{\pi j}{2b}$, $\phi_k = \frac{\pi k}{b}$, and the coefficients $a_k^{(b)}$ are defined in equation (5) of that paper.

10. To be efficient to calculate the Fourier coefficients $\hat{f}(m,l)$ they separate the Legendre part and the exponential part of the spherical harming to leverage a Fast-Fourier transform rewriting eq. (4) in the following way
$$\hat{f}(m,l) = q_l^m \sum_{k=0}^{b-1} a_k^{(b/2)} P_l^m(\cos\ k\theta)\sum_{j=0}^{b-1} e^{-imj\phi} f(k\theta, j\phi) \tag{5}$$
where $\theta = \pi/b$, $\phi = 2\pi/b$. In this way the inner sum can be calculated for each fixed k and for all m by means fo the fast Fourier Transform. Once obtained the inner sum, the outer sum can be calculated with a Legendre transform.

2 Daniel Spielman's notes, Jupyter Notebook "Experience1" and meeting of Tuesday 26 February

One interesting thing:
Theorem 2.1. *Lemma 5.2.1 of Daniel Spielman's Lecture Notes of the course of Spectral Graph Theory (2015)* The Laplacian of the 2-Nearest-Neighbours ring graph R_n on n equispaced vertices has eigenvectors

$$\mathbf{x}_k(u) = \cos(2k\pi u/n)$$
$$\mathbf{y}_k(u) = \sin(2k\pi u/n)$$

for $0 \leq k \leq n/2$, ignoring $\mathbf{y}_0 = \mathbf{0}$ and for even n ignoring $\mathbf{y}_{n/2}$ for the same reason. Eigenvectors $\mathbf{x}_k, \mathbf{y}_k$ have eigenvalues $2 - 2\cos(2\pi k/n)$

This lemma make us realize the fact that for a Graph Laplacian it is possible to have eigenvectors equal to the spherical harmonics sampled in the nodes, but has completely different eigenvalues! Furthermore, if I artificially build a matrix $L' = U\Lambda' U^T$ such that $L = U\Lambda U^T$ but with $\Lambda' = \text{diag}\{-1, -4, -4, -9, -9, ...\}$ the eigenvalues of the operator

$$\triangle\Phi(\rho,\theta) = \frac{1}{\rho}\partial_\rho\left(\frac{1}{\rho}\partial_\rho\Phi\right) + \frac{1}{\rho^2}\partial_{\theta\theta}\Phi$$

Laplace-Beltrami on the circle So, this tells us that the problem proposed by Michaël

$$W^\star = \arg\min_W ||U - \hat{U}||, \quad L = D - W, \quad LU = U\Lambda\hat{U} = \text{"Spherical harmonics sampled in the point set P"} \tag{6}$$

could be badly posed, in the sens that once found the weights $W^\star$ that give us the $U^\star$ realizing the minimum above, we have no clue of which eigenvalues to set in order to retrieve a true Graph Laplacian (positive diagonal, negative entries, symmetric, row/column sum equal to 0).
Nathanaël proposes a different approach. After having observed that given a non-uniform sampling (that's the setting we are aiming at) the vectors of the spherical harmonics sampled in the point set P won't be orthogonal, and thus a spectral decomposition of the Laplacian will never realize Michaël's problem 6 he proposes the following problem:

$$\arg\min_W ||L\hat{U} - \hat{U}\hat{\Lambda}||_{\mathcal{F}}, \quad L = D - W, \quad W > 0, \quad D = \text{diag}\{\sum_i (W)_{i,j}\} \tag{7}$$

where $\hat{\Lambda}, \hat{U}$ are respectively the eigenvalues and eigenvectors of $\triangle_{\mathcal{M}}$ the operator of Laplace-Beltrami sampled in the point set P. In this way we get the graph laplacian that has the closest spectral decomposition to the "true" one. The problem 7 can be rewritten in the following form

$$[w^\star\ \lambda^\star]^T = \arg\min_W ||A[w\ \lambda]^T||_{\mathcal{F}} \tag{8}$$

However, Michaël is not convinced by the fact that the form of the problem seems redundant: $\lambda = \lambda(w)$ is a function of w. So, the optimization is done on both λ, w that can't respect the relationship $\lambda = \lambda(w)$. At the end only one of the two arguments of the minimization problem $[w^\star \ \lambda^\star]^T$ will be used to obtain the second according to the relation $\lambda = \lambda(w)$.

Waiting for Nathanël's notes to make this passage clearer

References