Convergence of Laplacian Eigenmaps

Mikhail Belkin*and Partha Niyogi†

May 15, 2008

Abstract

Geometrically based methods for various tasks of data analysis have attracted considerable attention over the last few years. In many of these algorithms, a central role is played by the eigenvectors of the graph Laplacian of a data-derived graph. In this paper, we show that if points are sampled uniformly at random from an unknown submanifold $\mathcal{M}$ of $\mathbb{R}^N$, then the eigenvectors of a suitably constructed graph Laplacian converge to the eigenfunctions of the Laplace Beltrami operator on $\mathcal{M}$. This basic result directly establishes the convergence of spectral manifold learning algorithms such as Laplacian Eigenmaps and Diffusion Maps. It also has implications for the understanding of geometric algorithms in data analysis, computational harmonic analysis, geometric random graphs, and graphics.

1 Introduction

The last several years have seen a flurry of activity in geometrically motivated approaches to data analysis and machine learning. The unifying premise behind these methods is the assumption that many types of high-dimensional natural data lie on or near a low-dimensional submanifold of $\mathbb{R}^N$. Collectively this class of learning algorithms is often referred to as *manifold learning*.

Some recent manifold algorithms include Isomap [31], Locally Linear Embedding (LLE) [28], Diffusion Maps [14], Hessian Eigenmaps [15] among others.

*The Ohio State University, Department of Computer Science and Engineering
†The University of Chicago, Departments of Computer Science and Statistics

In [3] we introduced an algorithmic framework based on the Laplace-Beltrami operator of a manifold to motivate using the graph Laplacian associated to point-cloud data for data representation and dimensionality reduction. We called the algorithm arising out of this point of view the Laplacian Eigenmaps.

Indeed, several recent manifold learning algorithms are closely related to the Laplacian. The eigenfunctions of the Laplacian are also eigenfunctions of the heat diffusion operators. The diffusion operators play an important role in a variety of algorithms for data analysis developed by Coifman and collaborators in a series of recent papers, see [14], and the special issue of Applied and Computational Harmonic Analysis [1]. These papers combine ideas from multiscale analysis and spectral geometry in many interesting ways to give rise to a suite of novel geometrically motivated algorithms for data processing. The Hessian Eigenmaps approach of Donoho and Grimes [15] uses the Frobenius norm of the Hessian matrix while the Laplacian is its trace. Finally, as observed in [3], the cost function that is minimized to obtain the embedding of LLE [28] can be viewed as an approximation to the squared Laplacian.

In the manifold learning setting, the underlying manifold is typically unknown. Therefore functional maps from the manifold need to be estimated using point cloud data. The common approximation strategy in these methods is to construct an adjacency graph associated to a point cloud. The underlying intuition has always been that since the graph is a proxy for the manifold, inference based on the structure of the graph corresponds to the desired inference based on the geometric structure of the manifold. Theoretical results to justify this intuition have been developed over the last few years. However, a proof of spectral convergence, necessary for guaranteeing convergence of algorithms, has been elusive.

1.1 Prior and Related Work

The problem of estimating geometric and topological invariants from point cloud data has recently attracted some attention. Some of the recent work includes estimating geometric invariants of the manifold, such as homology [32, 26], dimensionality [24, 20], geodesic distances [7], and comparing point clouds using Gromov-Hausdorff distance [22].

This paper relies on results obtained in [4, 2] for functional convergence of operators. However considerably more careful analysis is required to ensure spectral convergence, which guarantees convergence of the corresponding algorithms. To the best of our knowledge previous results are not sufficient to guarantee convergence for any spectral method in the manifold setting.

Empirical convergence of spectral clustering for a fixed kernel parameter t was shown in [25]. However the geometric case requires $t \to 0$. We also note results on approximating empirical eigenfunctions in [21] and work in [6].

The results in this paper as well as in [4, 2] are for the case of a uniform probability distribution on the manifold. Recently [16] provided deeper probabilistic analysis in that case. In a different context closely related ideas were considered in [30].

Lafon in [23] generalized pointwise convergence results from [2] to the important case of an arbitrary probability distribution on the manifold. We also note [8], where a similar result is shown for the case of a domain in $\mathbb{R}^n$. Those results were further generalized and presented with an empirical pointwise convergence theorem in [18]. A faster convergence rate was obtained in [29].

We observe that the arguments in this paper are likely to allow one to use their results to show convergence of eigenfunctions for a wide class of probability distributions on the manifold.

2 Main Result

The main result of this paper is to show convergence of the eigenvectors of the graph Laplacian associated to a point cloud dataset to eigenfunctions of the Laplace-Beltrami operator when the data is sampled from a uniform probability distribution on an embedded manifold. It is likely that this result can be extended to a larger class of probability distributions on the manifold.

In what follows we will assume that the manifold $\mathcal{M}$ is a compact infinitely differentiable Riemannian submanifold of $\mathbb{R}^N$ (without boundary). Recall now that the Laplace-Beltrami operator (for functions) $\Delta_\mathcal{M}$ on $\mathcal{M}$ is a differential operator $\Delta_\mathcal{M} : \mathcal{C}^2(\mathcal{M}) \to L^2(\mathcal{M})$ defined as

$$\Delta_\mathcal{M} f = -\operatorname{div}(\nabla f)$$

where ∇f is the gradient vector field and div denotes divergence.

$\Delta_\mathcal{M}$ is a positive semidefinite self-adjoint operator and has a discrete spectrum on a compact manifold. We will generally denote its ith smallest eigenvalue by λ_i (in increasing order). It is important to note the well-known fact that all eigenfunctions of the Laplace-Beltrami operator are infinitely differentiable functions. See [27] for an introduction to the subject.

We define the operator $\mathbf{L}_t : L^2(\mathcal{M}) \to L^2(\mathcal{M})$ as follows:

$$\mathbf{L}_t(f)(p) = \frac{1}{t(4\pi t)^{k/2}} \left(\int_{\mathcal{M}} e^{-\frac{\|p-y\|^2}{4t}} f(p)\, d\mu_y - \int_{\mathcal{M}} e^{-\frac{\|p-y\|^2}{4t}} f(y)\, d\mu_y \right)$$

where μ is the uniform measure on $\mathcal{M}$ obtained from the volume form. As shown in previous work, this operator serves as a functional approximation to the Laplace-Beltrami operator on $\mathcal{M}$.

If data points $x_1, \ldots, x_n \in \mathcal{M} \subset \mathbb{R}^N$ are obtained by sampling $\mathcal{M}$ according to μ, the corresponding empirical form of the operator is the following:

$$\hat{\mathbf{L}}_{t,n}(f)(p) = \frac{1}{t(4\pi t)^{k/2}} \left(\frac{1}{n} \sum e^{-\frac{\|p-x_i\|^2}{4t}} f(p) - \frac{1}{n} \sum_i e^{-\frac{\|p-x_i\|^2}{4t}} f(x_i) \right)$$

The operator $\hat{\mathbf{L}}_{t,n}$ is the *point cloud Laplacian* that forms the basis of the Laplacian Eigenmaps algorithm for manifold learning. It acts on functions $\mathcal{M} \to \mathbb{R}$ and may be viewed as an operator $\hat{\mathbf{L}}_{t,n} : C(\mathcal{M}) \to C(\mathcal{M})$ that is the sum of a multiplication operator and a finite rank operator. Consider the random (weighted) graph whose vertex set V is identified with the data points $x_1, \ldots, x_n$ and where the the weight matrix W is given by $W_{ij} = \frac{1}{n}\frac{1}{t(4\pi t)^{k/2}} e^{-\frac{\|x_i-x_j\|^2}{4t}}$. It is easy to see that for any $f : \mathcal{M} \to \mathbb{R}$, if one considers the restriction f_V of f to the graph, then $\hat{\mathbf{L}}_{t,n} f$ restricted to V is the same as the action of the graph Laplacian (matrix) $L = D - W$ on (the vector) f_V. Therefore the eigenvalues of the graph Laplacian coincide with those of $\hat{\mathbf{L}}_{t,n}$ (for small t) and the eigenfunctions of $\hat{\mathbf{L}}_{t,n}$ are naturally related to the eigenvectors of the graph Laplacian.

Our main theorem shows that that there is a way to choose a sequence t_n, such that the eigenfunctions of the empirical operators $\hat{\mathbf{L}}_{t_n,n}$ converge to the true eigenfunctions of the Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ in probability.

Theorem 2.1 *Let $\lambda^t_{n,i}$ be the ith eigenvalue of $\hat{\mathbf{L}}_{t,n}$ and $e^t_{n,i}$ be the corresponding eigenfunction (which, for each fixed i, will be shown to exist for t sufficiently small). Let λ_i and e_i be the corresponding eigenvalue and eigenfunction of $\frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_{\mathcal{M}}$ respectively. Then there exists a sequence $t_n \to 0$, such that*

$$\lim_{n\to\infty} \lambda^{t_n}_{n,i} = \lambda_i$$

$$\lim_{n\to\infty} \|e_{n,i}^{t_n}(x) - e_i(x)\| = 0$$

where the limits are taken in probability.

Note 1: We will assume that all eigenvalues are of multiplicity one. Otherwise corresponding eigenvectors are not unique and convergence of spectral projections may be obtained instead, using the same arguments. We also note that eigenfunctions are defined up to a change of sign (assuming they are norm one). To have a consistent way of choosing eigenfunctions, one takes an arbitrary function f, not perpendicular to any of the eigenfunctions and chooses eigenfunctions e, so that $\langle e, f\rangle > 0$.
Note 2: In the rest of our exposition, we will let $\mathrm{vol}(\mathcal{M}) = 1$.

Our main result has implications for a number of different subjects. Our own motivation was the analysis of algorithms for machine learning and the result directly proves the convergence of the Laplacian Eigenmaps algorithm and has consequences for a variety of algorithms that utilize ideas from spectral geometry. Theorem 2.1 may also be viewed as a result in random matrices where the spectrum (and corresponding spectral projections) of the random graph Laplacian converges to that of the underlying Laplace-Beltrami operator. This basic convergence result also has consequences for sensor networks (where sensors are dropped on a surface and their connectivity depends on distance), for graphics (where surfaces are modeled by point clouds), for computational harmonic analysis and scientific computing in a geometric setting (where eigenfunctions of the Laplacian need to be computed and can play an important role).

3 Structure of the proof

The proof of the main theorem consists of two main parts. One is spectral convergence of the functional approximation $\mathbf{L}_t$ to $\Delta_{\mathcal{M}}$ as $t \to 0$ and the other is spectral convergence of the empirical approximation $\hat{\mathbf{L}}_{t,n}$ to $\mathbf{L}_t$ as the number of data points n tends to infinity. These two types of convergence are then put together to obtain the main Theorem 1.

Part 1. The hard part of the proof is to show convergence of eigenvalues and eigenfunctions of the functional approximation $\mathbf{L}_t$ to those of $\Delta_{\mathcal{M}}$ as $t \to 0$. For that we will take a different functional approximation $\frac{\mathbf{1}-\mathbf{H}_t}{t}$ of $\Delta_{\mathcal{M}}$, where $\mathbf{H}_t$ is the heat operator. While $\frac{\mathbf{1}-\mathbf{H}_t}{t}$ does not converge *uniformly* to $\Delta_{\mathcal{M}}$ they share an eigenbasis and for each *fixed* i the ith eigenvalue of $\frac{\mathbf{1}-\mathbf{H}_t}{t}$ converges to the ith eigenvalue of $\Delta_{\mathcal{M}}$.

We will then consider the operator $\mathbf{R}_t = \frac{\mathbf{1}-\mathbf{H}_t}{t} - \mathbf{L}_t$. A careful analysis of this operator, which constitutes the bulk of this paper, shows that $\mathbf{R}_t$ is a small *relatively bounded* perturbation of $\frac{\mathbf{1}-\mathbf{H}_t}{t}$, in the sense that $\sup_{f\in L^2(\mathcal{M})} \| \frac{\langle \mathbf{R}_t f, f\rangle}{\langle \frac{\mathbf{1}-\mathbf{H}_t}{t} f, f\rangle} \| \to 0$ as $t \to 0$. This implies spectral convergence and leads to the following

Theorem 3.1 *Let $\lambda_i, \lambda_i^t, e_i, e_i^t$ be the ith smallest eigenvalues and the corresponding eigenfunctions*[1] *of $\Delta_{\mathcal{M}}$ and $\mathbf{L}_t$ respectively. Then*

$$\lim_{t\to 0} |\lambda_i - \lambda_i^t| = 0$$

$$\lim_{t\to 0} \|e_i - e_i^t\| = 0$$

Part 2. The second part is to show that the eigenfunctions of the empirical operator $\hat{\mathbf{L}}_{t,n}$ converge to the eigenfunctions of $\mathbf{L}_t$ as $n \to \infty$ in probability. That result follows readily from the previous work in [25] together with the analysis of the essential spectrum of $\mathbf{L}_t$. The following theorem is obtained:

Theorem 3.2 *For a fixed sufficiently small t, let $\lambda_{n,i}^t$ and λ_i^t be the ith eigenvalue of $\hat{\mathbf{L}}_{t,n}$ and $\mathbf{L}_t$ respectively. Let $e_{n,i}^t$ and e_i^t be the corresponding eigenfunctions. Then*

$$\lim_{n\to\infty} \lambda_{n,i}^t = \lambda_i^t$$

$$\lim_{n\to\infty} \|e_{n,i}^t(x) - e_i^t(x)\| = 0$$

with a.s. convergence, assuming that $\lambda_i^t \le \frac{1}{2t}$.

Observe that this implies convergence for any fixed i as soon as t is sufficiently small.

Symbolically these two theorems together with the final Theorem 2.1 can be represented by the following diagram:

$$\begin{tikzcd}
\mathbf{Eig}\ \hat{\mathbf{L}}_{t,n} \arrow[r, "n \to \infty"] \arrow[rr, bend right, "n \to \infty \quad t_n \to 0"'] & \mathbf{Eig}\ \mathbf{L}_t \arrow[r, "t \to 0"] & \mathbf{Eig}\ \Delta_{\mathcal{M}}
\end{tikzcd}$$

[1]For simplicity we assume that $\Delta_{\mathcal{M}}$ has no multiple eigenvectors. If such eigenvectors exist, spectral projections should be used instead of eigenvectors.

After demonstrating two types of convergence results in the top line of the diagram (notice that the left arrow is convergence almost surely but convergence of $\mathbf{L}_t$ to $\Delta_{\mathcal{M}}$ is deterministic), a simple argument shows that a sequence t_n can be chosen to guarantee convergence as in Theorem 2.1 and provides the bottom arrow.

Remark: It is worth noting that $\Delta_{\mathcal{M}}$ is viewed as an operator $C^2(\mathcal{M}) \to L^2(\mathcal{M})$, $\hat{\mathbf{L}}_{t,n}$ is an operator $C(\mathcal{M}) \to C(\mathcal{M})$ while $\mathbf{L}_t$ is an operator from $L^2(\mathcal{M}) \to L^2(\mathcal{M})$. Luckily the eigenfunctions of these operators are all in C^∞ and therefore contained in all the relevant function spaces. In proving Theorem 3.2, we consider both $\hat{\mathbf{L}}_{t,n}$ and $\mathbf{L}_t$ to be operators $C(\mathcal{M}) \to C(\mathcal{M})$ while in proving Theorem 3.1, we consider $\mathbf{L}_t$ to be $L^2 \to L^2$ and prove that it is a relatively bounded perturbation of another operator $\frac{\mathbf{1}-\mathbf{H}_t}{t} : L^2 \to L^2$.

Notation: Let us fix our notational convention for the variety of mathematical objects in this paper.

1. We will use bold letters to denote operators and capital letters to denote kernel functions for the associated integral operators. Thus, if $K(x,y)$ is a kernel, then $\mathbf{K}$ is the corresponding convolution operator $\mathbf{K}(f)(x) = \int_{\mathcal{M}} K(x,y)f(y)d\mu_y$.

2. In particular $G_t(x,y)$ will denote the Gaussian kernel

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

and $\mathbf{G}_t$ will denote the corresponding convolution over the manifold:

$$\mathbf{G}_t(f)(x) = (4\pi t)^{-\frac{k}{2}} \int_{\mathcal{M}} e^{-\frac{\|x-y\|^2}{4t}} f(y)\ d\mu_y$$

Similarly $H_t(x,y)$ will denote the *heat kernel* on $\mathcal{M}$ and $\mathbf{H}_t$ is the *heat operator*, which is convolution with the heat kernel.

3. For a function $f : \mathcal{M} \to \mathbb{R}$, $\|f\|_1, \|f\|, \|f\|_\infty, \|f\|_{H^s}$ will denote its norm in L^1, L^2, L^∞ and the Sobolev space H^s respectively. The last space, once s is fixed, we will also denote as H. For an operator $\mathbf{A}$, $\|\mathbf{A}\|$ denotes the L^2 operator norm:

$$\|\mathbf{A}\| = \sup_{\|f\|=1} \|\mathbf{A}f\|$$

4 Spectral Convergence of Functional Approximations.

4.1 Main Objects and Outline of the Proof

Let $\mathcal{M}$ be a smooth, compact k-dimensional submanifold of $\mathbb{R}^N$ with its Riemannian structure inherited from $\mathbb{R}^N$ and the corresponding induced measure μ.

We define the operator $\mathbf{L}_t : L^2(\mathcal{M}) \to L^2(\mathcal{M})$ as follows:

$$\mathbf{L}_t(f)(x) = \frac{1}{t(4\pi t)^{k/2}} \left(\int_{\mathcal{M}} e^{-\frac{\|x-y\|^2}{4t}} f(x)\ d\mu_y - \int_{\mathcal{M}} e^{-\frac{\|x-y\|^2}{4t}} f(y)\ d\mu_y \right)$$

As shown in previous work (see [4] and citations therein), this operator serves as a functional approximation (pointwise) to the Laplace-Beltrami operator on $\mathcal{M}$.

The purpose of this paper is to extend the previous results to the eigenvalues and eigenfunctions. This turns out to need some careful estimates.

Let $\mathbf{H}_t$ be the heat operator for the Riemannian manifold $\mathcal{M}$. We define

$$\mathbf{R}_t = \frac{\mathbf{1} - \mathbf{H}_t}{t} - \mathbf{L}_t$$

The idea is to give an estimate on the remainder term $\mathbf{R}_t$, that will imply that $\mathbf{R}_t$ is dominated by $\frac{\mathbf{1}-\mathbf{H}_t}{t}$. This, in turn, will imply convergence of the spectrum and eigenfunctions.

We will need two estimates for the size of the perturbation $\mathbf{R}_t$, which are given in the following two propositions:

Proposition 4.1 *Let $f \in L^2$. There exists $C \in \mathbb{R}$, such that for all sufficiently small values of t*

$$\|\mathbf{R}_t f\| \leq C\|f\|$$

Proposition 4.2 *Let f belong to the Sobolev space $H^{\frac{k}{2}+1}$. There exists $C \in \mathbb{R}$, such that for all sufficiently small values of t*

$$\|\mathbf{R}_t f\| \leq C\sqrt{t}\|f\|_{H^{\frac{k}{2}+1}}$$

These propositions, will allow us to derive the main technical result of the paper (whose proof together with the proofs of propositions will be presented in the next few sections):

Theorem 4.3 *Let $t \in (0, 0.1)$. Then there exists a constant $C > 0$, independent of t, such that the following inequality holds:*

$$\sup_{f \in L^2} \frac{|\langle \mathbf{R}_t f, f\rangle|}{\langle \frac{\mathbf{1}-\mathbf{H}_t}{t} f, f\rangle} \leq C t^{\frac{2}{k+6}}$$

In particular,

$$\lim_{t \to 0} \sup_{f \in L^2} \frac{\langle \mathbf{R}_t f, f\rangle}{\langle \frac{\mathbf{1}-\mathbf{H}_t}{t} f, f\rangle} = 0$$

and hence $\mathbf{R}_t$ is dominated by $\frac{\mathbf{1}-\mathbf{H}_t}{t}$.

This result, establishing a *relative bound* on the size of the perturbation implies spectral convergence and hence establishes Theorem 2.1 as outlined below.

Observation. The *ith* smallest eigenvalue of $\frac{\mathbf{1}-\mathbf{H}_t}{t}$ (denoted by $\lambda_i(\frac{\mathbf{1}-\mathbf{H}_t}{t})$) is equal to $(1 - e^{-\lambda_i t})/t$. Thus $\lim_{t\to 0} \lambda_i(\frac{\mathbf{1}-\mathbf{H}_t}{t}) = \lambda_i$. The corresponding eigenvector $e_i(\frac{\mathbf{1}-\mathbf{H}_t}{t})$ is the same as e_i. Thus we see that it is sufficient to show that for a fixed i and small t, the ith eigenfunction and eigenvalue of $\mathbf{L}_t$ are close to those of $\frac{\mathbf{1}-\mathbf{H}_t}{t}$.

We now provide the argument for convergence of eigenvalues. For convergence of eigenvectors additional assumptions about the eigengap are needed. However spectral projections can be shown to converge without those assumptions ([19]).

Proposition 4.4 *Let $\mathbf{A}, \mathbf{B}$ be positive, self-adjoint operators with discrete spectrum that may be arranged in increasing order. Let $\mathbf{D} = \mathbf{A} - \mathbf{B}$ and $\lambda_1(\mathbf{A}) \leq \lambda_2(\mathbf{A}) \leq \ldots$ and $\lambda_1(\mathbf{B}) \leq \lambda_2(\mathbf{B}) \leq \ldots$ denote the eigenvalues of A and B respectively. Assume that for all $f \in L^2$*

$$\left| \frac{\langle \mathbf{D} f, f\rangle}{\langle \mathbf{A} f, f\rangle} \right| \leq \epsilon$$

Then for all k, we have $1 - \epsilon \leq \lambda_k(\mathbf{B})/\lambda_k(\mathbf{A}) \leq 1 + \epsilon$.

Proof:

For any $f \in L^2$, we have

$$|\langle \mathbf{A} f, f\rangle| \leq |\langle \mathbf{B} f, f\rangle| + |\langle \mathbf{D} f, f\rangle| \leq |\langle \mathbf{B} f, f\rangle| + \epsilon |\langle \mathbf{A} f, f\rangle|$$

By the same token,

$$|\langle \mathbf{A}f, f\rangle| \geq |\langle \mathbf{B}f, f\rangle| - |\langle \mathbf{D}f, f\rangle| \geq |\langle \mathbf{B}f, f\rangle| - \epsilon |\langle \mathbf{A}f, f\rangle|$$

Putting these together, we have

$$(1-\epsilon)|\langle \mathbf{A}f, f\rangle| \leq |\langle \mathbf{B}f, f\rangle| \leq (1+\epsilon)|\langle \mathbf{A}f, f\rangle|$$

Let H be an arbitrary k-dimensional subspace of L^2 and $H^\perp$ its orthogonal complement. Then

$$(1-\epsilon) \max_H \min_{f \in H^\perp} |\langle \mathbf{A}f, f\rangle| \leq \max_H \min_{f \in H^\perp} |\langle \mathbf{B}f, f\rangle| \leq (1+\epsilon) \max_H \min_{f \in H^\perp} |\langle \mathbf{A}f, f\rangle|$$

By the Courant-Fischer theorem, the result follows. □

4.2 Proof of Theorem 4.3.

Technical note: We will consider all functions to be orthogonal to the constant function in L^2. This can be done without a loss of generality and is needed at several points in the proofs. We will also normalize all eigenfuctions to norm 1 in L^2.

Before proceeding with the main result we need to formulate the following

Lemma 4.5 *Let $f = \sum_{\lambda_i \leq \alpha} a_i e_i$ be a bandlimited function in terms of eigenfunctions of the Laplace-Beltrami operator. Then for some constant $C > 0$, we have*

$$\|f\|_{H^{\frac{k}{2}+1}} \leq C\alpha^{\frac{k+2}{4}} \|f\| \tag{1}$$

In particular, if e is an eigenvector of $\Delta_\mathcal{M}$ with eigenvalue λ, then

$$\|e\|_{H^{\frac{k}{2}+1}} \leq C\lambda^{\frac{k+2}{4}} \tag{2}$$

Proof:$\Delta_\mathcal{M}^{-1}$ is a bounded operator $H^p \to H^{p+2}$. Recalling that $L^2 = H^0$ we obtain

$$\|f\|_{H^{\frac{k}{2}+1}} \leq C\|(\Delta_\mathcal{M})^{\frac{k+2}{4}} f\|_{H^0} = C\|\sum_{\lambda_i \leq \alpha} \lambda_i^{\frac{k+2}{4}} a_i e_i\| \leq C\alpha^{\frac{k+2}{4}} \|f\|$$

□

Now we can proceed with the proof of the central theorem 4.3.

Proof:[Theorem 4.3]

Let $e_i(x)$ be the ith eigenfunction of $\Delta_{\mathcal{M}}$ and let λ_i be the corresponding eigenvalue. Recall that e_i's form an orthonormal basis of $L^2(\mathcal{M})$. Thus any function $f \in L^2(\mathcal{M})$ can be written uniquely as

$$f(x) = \sum_{i=0}^{\infty} a_i e_i(x)$$

where a_i's are such that $\sum a_i^2 < \infty$.

Recall also that

$$\mathbf{H}_t f = \exp(-t\Delta_{\mathcal{M}}) f \tag{3}$$

Thus

$$\mathbf{H}_t e_i = \exp(-t\lambda_i) e_i \tag{4}$$

$$\frac{\mathbf{1} - \mathbf{H}_t}{t} e_i = \frac{1 - e^{-\lambda_i t}}{t} e_i \tag{5}$$

Now let us fix t and consider the function $\phi(x) = \frac{1-e^{-xt}}{t}$ for positive x. It is easy to check that ϕ is a concave and increasing function of x.

Put $x_0 = 1/\sqrt{t}$. We have:

$$\phi(0) = 0 \qquad \phi(x_0) = \frac{1 - e^{-\sqrt{t}}}{t} \qquad \frac{\phi(x_0)}{x_0} = \frac{1 - e^{-\sqrt{t}}}{\sqrt{t}}$$

Splitting the positive real line in two intervals $[0, x_0]$, $[x_0, \infty)$ and using concavity and monotonicity we observe that

$$\phi(x) \geq \min\left(\frac{1 - e^{-\sqrt{t}}}{\sqrt{t}} x, \frac{1 - e^{-\sqrt{t}}}{t} \right)$$

Note that $\lim_{t \to 0} \frac{1-e^{-\sqrt{t}}}{\sqrt{t}} = 1$.

Therefore for positive and sufficiently small t (say, $0 < t < 0.1$)

$$\phi(x) \geq \min\left(\frac{1}{2} x, \frac{1}{2\sqrt{t}} \right)$$

Thus

$$\left\langle \frac{\mathbf{1} - \mathbf{H}_t}{t} e_i, e_i \right\rangle = \frac{1 - e^{-\lambda_i t}}{t} \geq \frac{1}{2} \min\left(\lambda_i, \frac{1}{\sqrt{t}} \right) \tag{6}$$

Now take $f \in L^2$, $f(x) = \sum_{i=1}^{\infty} a_i e_i(x)$. Without loss of generality we can assume that $\|f\|_2 = 1$. For any $\alpha > 0$, we can split f as a sum of f_1 and f_2 as follows:

$$f_1 = \sum_{\lambda_i \le \alpha} a_i e_i, \qquad f_2 = \sum_{\lambda_i > \alpha} a_i e_i$$

It is clear that $f = f_1 + f_2$ and, since f_1 and f_2 are orthogonal, $\|f\|^2 = \|f_1\|^2 + \|f_2\|^2$. We will now deal separately with f_1 and with f_2.

Notice that $\mathbf{R}_t$ is self-adjoint and thus

$$\langle \mathbf{R}_t f, f\rangle = \langle \mathbf{R}_t f_1, f_1\rangle + 2\langle \mathbf{R}_t f_1, f_2\rangle + \langle \mathbf{R}_t f_2, f_2\rangle$$

Using the Cauchy-Schwartz and triangle inequalities, we have

$$|\langle \mathbf{R}_t f, f\rangle| \le 3\|\mathbf{R}_t f_1\| + \|\mathbf{R}_t f_2\|\|f_2\| \tag{7}$$

We now give a bound for $\frac{\|\mathbf{R}_t f_1\|}{\langle \frac{\mathbf{1}-\mathbf{H}_t}{t} f, f\rangle}$. We see that by Proposition 4.2,

$$\|\mathbf{R}_t f_1\| < C\sqrt{t}\|f_1\|_{H^{\frac{k}{2}+1}}$$

Using the fact that f_1 is band-limited by α and applying Lemma 4.5, we get

$$\|\mathbf{R}_t f_1\| < C\sqrt{t}\left\|\sum_{\lambda_i \le \alpha} a_i e_i\right\|_{H^{\frac{k}{2}+1}} < C_1\sqrt{t}\alpha^{\frac{k+2}{4}}$$

On the other hand, from Inequality 6

$$\left\langle \frac{\mathbf{1}-\mathbf{H}_t}{t} f, f\right\rangle = \left\langle \sum_i \frac{1-e^{-t\lambda_i}}{t} a_i e_i, \sum_i a_i e_i \right\rangle$$

$$= \sum_i a_i^2 \frac{1-e^{-t\lambda_i}}{t} > \frac{1}{2}\sum_i a_i^2 \min(\lambda_i, \frac{1}{\sqrt{t}})$$

Therefore, for $0 < t < 0.1$, we obtain

$$\left\langle \frac{\mathbf{1}-\mathbf{H}_t}{t} f, f\right\rangle > \frac{1}{2}\lambda_1 \sum_i a_i^2 = \frac{\lambda_1}{2}$$

Thus,

$$\frac{\|\mathbf{R}_t f_1\|}{\langle \frac{\mathbf{1}-\mathbf{H}_t}{t} f, f\rangle} < \frac{2C_1}{\lambda_1}\sqrt{t}\alpha^{\frac{k+2}{4}} \tag{8}$$

We will now bound $\frac{\|\mathbf{R}_t f_2\|\|f_2\|}{\langle \frac{\mathbf{1}-\mathbf{H}_t}{t} f,f\rangle}$. By applying Proposition 4.1, we have

$$\|\mathbf{R}_t f_2\|\|f_2\| \leq C_3 \|f_2\|^2$$

On the other hand,

$$\left\langle \frac{\mathbf{1}-\mathbf{H}_t}{t} f, f \right\rangle \geq \left\langle \frac{\mathbf{1}-\mathbf{H}_t}{t} f_2, f_2 \right\rangle \geq \sum_{\lambda_i > \alpha} a_i^2 \frac{1}{2} \min(\alpha, \frac{1}{\sqrt{t}}) \geq \frac{1}{2} \min(\alpha, \frac{1}{\sqrt{t}}) \|f_2\|^2$$

Thus,

$$\frac{\|\mathbf{R}_t f_2\|\|f_2\|}{\langle \frac{\mathbf{1}-\mathbf{H}_t}{t} f, f\rangle} \leq \frac{C_3}{\min(\alpha, \frac{1}{\sqrt{t}})} \leq C_3 \max(\frac{1}{\alpha}, \sqrt{t}) \tag{9}$$

Putting this together with inequalities (7,8), we get

$$\frac{\langle \mathbf{R}_t f, f\rangle}{\langle \frac{\mathbf{1}-\mathbf{H}_t}{t} f, f\rangle} \leq C_4 \left(\sqrt{t}\alpha^{\frac{k+2}{4}} + \max(\frac{1}{\alpha}, \sqrt{t}) \right) \tag{10}$$

Letting $\alpha = t^{-\frac{2}{k+6}}$, we recover the theorem.

□

4.3 Proof of Estimate for L^2 norm of $\mathbf{R}^t$ (Proposition 4.1).

Let $\mathcal{M}$ be a smooth, compact, k-dimensional Riemannian submanifold of $\mathbb{R}^N$. Following [26], we characterize the complexity of the embedding of $\mathcal{M}$ by a quantity τ defined as the largest number having the property: The open normal bundle about $\mathcal{M}$ of radius r is imbedded in $\mathbb{R}^N$ for every $r < \tau$. $\frac{1}{\tau}$ bounds the norm of the second fundamental form and therefore provides a bound on curvature and nearness of self-intersection of $\mathcal{M}$. τ is also called the reach or the rolling ball condition in computational geometry. Since $\mathcal{M}$ is compact and smooth, we are guaranteed that $\tau > 0$.

Let H_t be the heat kernel on the manifold $\mathcal{M}$. Let the *ambient* Gaussian kernel be

$$G_t(p,q) = \frac{1}{(4\pi t)^{k/2}} e^{-\frac{\|p-q\|^2}{4t}}$$

and let E_t be the *geodesic* Gaussian kernel given by

$$E_t(p,q) = \frac{1}{(4\pi t)^{k/2}} e^{-\frac{d^2(p,q)}{4t}}$$

where $d(p,q)$ is the *geodesic* distance between $p,q \in \mathcal{M}$. Each kernel is naturally associated with an integral operator on the manifold denoted by $\mathbf{H}_t$, $\mathbf{G}_t$ and $\mathbf{E}_t$ respectively. We will start with the main result and then prove the necessary technical ingredients.

PROOF:[Proposition 4.1]

Recall that

$$\mathbf{L}_t f(x) = \frac{1}{t}\left(f(x)\int_{\mathcal{M}} G_t(x,y)d\mu_y - \int_{\mathcal{M}} f(y)G_t(x,y)d\mu_y \right)$$

Thus

$$\mathbf{R}_t f = \frac{1}{t}\left(1 - \int_{\mathcal{M}} G_t(x,y)d\mu_y\right) f - \frac{1}{t}(\mathbf{H}_t - \mathbf{G}_t)f$$

Applying Proposition 4.6, we see that the norm of the operator $\mathbf{H}_t - \mathbf{G}_t$ is bounded by Ct.

On the other hand, it is easily verified that the norm of the multiplication operator $\mathbf{M}_g f = fg$ is bounded by $\sup|g|$, $\|\mathbf{M}_g\| \leq \sup|g|$. Hence applying Proposition 4.11, we see that the norm of the multiplication operator in Eq. (4.3) is bounded by $C't$.

Putting these two observation together we arrive at Proposition 4.1. □

Proposition 4.6 *For some constant C depending only on the manifold and independent of t*

$$\|\mathbf{H}_t - \mathbf{G}_t\| \leq Ct$$

PROOF: We observe that

$$\|\mathbf{H}_t - \mathbf{G}_t\| \leq \|\mathbf{H}_t - \mathbf{E}_t\| + \|\mathbf{E}_t - \mathbf{G}_t\|$$

by the triangle inequality. The first term may be bounded by Lemma 4.13 while the second may be bounded by Lemma 4.7. The result follows.

□

We will now state the main technical result of this subsection:

Lemma 4.7 *For some constant C independent of t*

$$\|\mathbf{E}_t - \mathbf{G}_t\| \leq Ct$$

PROOF:
Consider a point $p \in \mathcal{M}$. We see that

$$(\mathbf{G}_t - \mathbf{E}_t)(f)(p) = \frac{1}{(4\pi t)^{k/2}} \int_{\mathcal{M}} (e^{-\frac{\|p-q\|^2}{4t}} - e^{-\frac{d^2(p,q)}{4t}}) f(q) d\mu$$

Consider the geodesic ball $B_\epsilon(p) = \{q \in \mathcal{M} | d(q,p) < \epsilon\}$, which is the set of all points in $\mathcal{M}$ whose geodesic distance from p is less than ϵ. Choose $\epsilon < \min(\tau/2, 1)$. We break the integral above in two parts, $(\mathbf{G}_t - \mathbf{E}_t)f = A + B$,

$$A = \frac{1}{(4\pi t)^{k/2}} \int_{B_\epsilon(p)} (e^{-\frac{\|p-q\|^2}{4t}} - e^{-\frac{d^2(p,q)}{4t}}) f(q) d\mu$$

$$B = \frac{1}{(4\pi t)^{k/2}} \int_{\mathcal{M} \setminus B_\epsilon(p)} (e^{\frac{\|p-q\|^2}{4t}} - e^{-\frac{d^2(p,q)}{4t}}) f(q) d\mu$$

From Lemma 4.10, we see that $|A| < C_1 t \mathbf{E}_{2t}(|f|)(p)$ for some $C_1 > 0$.. similarly, from Lemma 4.8, we have that $|B| < C_2 t \|f\|$. Define $h_1(p) = C_1 t \mathbf{E}_{2t}(|f|)(p)$ and $h_2(p) = C_2 t \|f\|$ (a constant function). Then, we see that

$$|(\mathbf{E}_t - \mathbf{G}_t)(f)(p)| \leq h_1(p) + h_2(p)$$

from which it follows that

$$\|(\mathbf{E}_t - \mathbf{G}_t) f\| \leq \|h_1\| + \|h_2\|$$

¿From Corollary 4.16 we see that $\|h_1\| = C_1 t \|\mathbf{E}_{2t} f\| < C_3 t \|f\|$. Since the manifold is compact $\|h_2\| < C_2 t \|f\|$. Putting these inequalities together, we see that there is a constant $C > 0$ such that

$$\|(\mathbf{E}_t - \mathbf{G}_t) f\| \leq Ct \|f\|$$

The proposition is proved. $\square$

We will now give necessary bounds for B and A.

Lemma 4.8 *For some constant C and t sufficiently small*

$$|B| = \frac{1}{(4\pi t)^{k/2}} \left| \int_{\mathcal{M} \setminus B_\epsilon(p)} \left(e^{-\frac{\|p-q\|^2}{4t}} - e^{-\frac{d^2(p,q)}{4t}} \right) f(q) d\mu \right| \leq Ct \|f\|$$

PROOF: We observe that $\|p-q\| \leq d(p,q)$ and hence by the triangle inequality

$$|B| \leq \frac{2}{(4\pi t)^{k/2}} \int_{\mathcal{M}\backslash B_\epsilon(p)} e^{-\frac{\|p-q\|^2}{4t}} |f(q)| d\mu(q)$$

Now let $X = \inf_{q \in \mathcal{M}\backslash B_\epsilon(p)} \|q-p\|$. Clearly,

$$|B| \leq 2\frac{e^{-\frac{X^2}{4t}}}{(4\pi t)^{k/2}} \int_{\mathcal{M}\backslash B_\epsilon(p)} |f(q)| d\mu(q) \leq 2\frac{e^{-\frac{X^2}{4t}}}{(4\pi t)^{k/2}} \int_{\mathcal{M}} |f(q)| d\mu(q)$$

By Schwartz inequality, we have $\int_{\mathcal{M}} |f| d\mu \leq \|\mathbf{1}\| \|f\| = \mathrm{vol}(\mathcal{M}) \|f\|$. Therefore,

$$|B| \leq 2\frac{e^{-\frac{X^2}{4t}}}{(4\pi t)^{k/2}} \mathrm{vol}(\mathcal{M}) \|f\|$$

It now only remains to check what X is.

We use the fact (proved in [26]) that for any two points $p, q \in \mathcal{M}$, such that $\|p-q\| < \tau/2$, we have that

$$\|p-q\| \leq d(p,q) \leq \tau - \tau\sqrt{1 - 2\frac{\|p-q\|}{\tau}}$$

Therefore, we see that

$$\|p-q\| < \alpha \implies d(p,q) \leq \epsilon$$

for $\alpha = (\tau/2)\left(1 - (1 - \frac{\epsilon}{\tau})^2\right)$. For all points $q \in \mathcal{M}\backslash B_\epsilon(p)$, we have $d(p,q) > \epsilon$ and hence $\|p-q\| > \alpha$. Therefore

$$0 < \alpha \leq \inf_{q \in \mathcal{M}\backslash B_\epsilon(p)} \|p-q\| = X$$

leading to $e^{-\frac{X^2}{4t}} \leq e^{-\frac{\alpha^2}{4t}}$. Since exponential function increases faster than a polynomial, the lemma is proved. □

Now we turn to bounding $A = \frac{1}{(4\pi t)^{k/2}} \int_{B_\epsilon(p)} \left(e^{-\frac{\|p-q\|^2}{4t}} - e^{-\frac{d^2(p,q)}{4t}}\right) f(q) d\mu$. Since $p, q \in B_\epsilon(p)$ are nearby, we can resort to the following lemma (appears in [5]).

Lemma 4.9 *For any two points $p, q \in \mathcal{M}$ such that $d(p,q) < 1$, the relationship between the Euclidean distance and geodesic distance is given by*

$$\|p-q\|^2 = d^2(p,q) - h(p,q)$$

where $h(p,q) = O(d^4(p,q))$. In other words, there exists an $a > 0$ such that $|h(p,q)| \leq ad^4(p,q)$ for all such p, q. The constant a depends upon the embedding of the manifold and bounds on the third derivatives of the embedding coordinates.

Using this lemma, we can now prove

Lemma 4.10 *For any point $p \in \mathcal{M}$ and point $q \in B_\epsilon(p)$, we have*

$$|E_t(p,q) - G_t(p,q)| \leq CtE_{2t}(p,q)$$

Proof:Using the exponential map $\exp : B_\epsilon \to \mathcal{M}$ from an ϵ-ball in T_p (centered at the origin) to $\mathcal{M}$, we can write in exponential coordinates to let $q = \exp(x)$. Then, we see that since $d(p,q) = \|x\|$, we have by Lemma 4.9 that $\|p-q\|^2 \geq \|x\|^2 - a\|x\|^4$. Hence, we have

$$E_t(p,q) = \frac{1}{(4\pi t)^{k/2}} e^{\frac{-\|x\|^2}{4t}}$$

and

$$G_t(p,q) = \frac{1}{(4\pi t)^{k/2}} e^{-\frac{\|p-q\|^2}{4t}} \leq \frac{1}{(4\pi t)^{k/2}} e^{-\frac{\|x\|^2 - a\|x\|^4}{4t}}$$

Therefore,

$$|E_t - G_t| \leq \frac{1}{(4\pi t)^{k/2}} \left(e^{-\frac{\|x\|^2 - a\|x\|^4}{4t}} - e^{-\frac{\|x\|^2}{4t}} \right)$$

To prove the lemma, it is sufficient to show

$$\frac{1}{(4\pi t)^{k/2}} \left(e^{-\frac{\|x\|^2 - a\|x\|^4}{4t}} - e^{-\frac{\|x\|^2}{4t}} \right) \leq Ct \frac{1}{(4\pi t)^{k/2}} e^{-\frac{\|x\|^2}{8t}}$$

for some constant C. Canceling common terms, we reduce the above inequality to

$$e^{-\frac{\|x\|^2}{8t}} \left(e^{\frac{a\|x\|^4}{4t}} - 1 \right) \leq Ct$$

Putting $z = \|x\|^2$, we see that this is equivalent to showing

$$f(z) = e^{-\frac{z}{8t}} \left(e^{\frac{az^2}{4t}} - 1 \right) \leq Ct$$

Examining $f(z)$, we see that the derivative $f'(z)$ is given by

$$f'(z) = \frac{1}{8t}\left((4az-1)e^{\frac{2az^2-z}{8t}} + e^{-\frac{z}{8t}}\right)$$

Notice that $f'(z) < 0$ if

$$1 \leq (1-4az)e^{\frac{az^2}{4t}}$$

Since $e^{\frac{az^2}{4t}} \geq 1 + \frac{az^2}{4t}$, we see that $f'(z) < 0$ as long as $1 + \frac{az^2}{4t} \geq \frac{1}{1-4az}$ or $\frac{z}{4t} \geq \frac{4}{1-4az}$. Since we are working in the ball $B_\epsilon(p)$, we have $z < \epsilon^2$. By choosing ϵ so that $4a\epsilon^2 < \frac{1}{2}$, we see that $f'(z) < 0$ for $z \geq 32t$. Therefore, for $z \geq 0$, $f(z)$ attains its maximum when $z \leq 32t$.

For $z \leq 32t$, and for t sufficiently small, we have

$$|f(z)| \leq e^{\frac{az^2}{4t}} - 1 \leq e^{256at} - 1 \leq 512at$$

The lemma is proved. □

Proposition 4.11

$$\left|1 - \int_{\mathcal{M}} G_t d\mu\right| \leq Ct$$

for some constant $C > 0$ that depends only on the manifold $\mathcal{M}$ and is independent of t.

Proof:We begin by noting that

$$\int_{\mathcal{M}} G_t d\mu = \int_{B_\epsilon(p)} G_t\, d\mu + \int_{\mathcal{M}\setminus B_\epsilon(p)} G_t d\,\mu$$

By the same arguments used in Lemma 4.8, the quantity $\int_{\mathcal{M}\setminus B_\epsilon(p)} G_t$ can be controlled as t goes to 0. We therefore concentrate on the term $\int_{B_\epsilon(p)} G_t$. Switching to exponential coordinates by writing $q = \exp(x)$ as before and using the fact that $e^z = 1 + O(ze^z)$, we can write $\int_{B_\epsilon(p)} G_t\, d\mu = D + F$, where

$$D = \frac{1}{(4\pi t)^{k/2}} \int_{B_\epsilon} e^{-\frac{\|x\|^2}{4t}} \sqrt{\det(g)} dx$$

$$F = O\left(\frac{1}{(4\pi t)^{k/2}} \int_{B_\epsilon} \frac{\|x\|^4}{4t} e^{-\frac{\|x\|^2 - a\|x\|^4}{4t}} \sqrt{\det(g)} dx\right)$$

Here g is the metric tensor in exponetial coordinates and we will use the fact that the quantity $\sqrt{\det(g)}$ can be written as

$$\sqrt{\det(g)} = 1 - \frac{1}{6} x^T R x + O(\|x\|^3), \tag{11}$$

where R is the Ricci curvature tensor.

Consider first the term F. Notice first that for $\|x\| \leq \frac{1}{\sqrt{2a}}$, we have that $e^{-\frac{\|x\|^2 - a\|x\|^4}{4t}} \leq e^{-\frac{\|x\|^2}{8t}}$. Therefore for some C'

$$F < \frac{C'}{(4\pi t)^{k/2}} \int_{B_\epsilon} e^{-\frac{\|x\|^2}{8t}} \frac{\|x\|^4}{4t} \sqrt{\det(g)} dx$$

Using the Eq. 11, we see that F is upper bounded by $\frac{C'}{4t}(u_4 + u_6 + u_7)$ where u_4, u_6, u_7 are the fourth, sixth, and seventh moments of the Gaussian distribution with variance $4t$. Hence $u_4 + u_6 + u_7 = O(t^2)$ and $F = O(t)$.

The first term D can be written as

$$D = \frac{1}{(4\pi t)^{k/2}} \int_{B_\epsilon} e^{-\frac{\|x\|^2}{4t}} \left(1 - \frac{1}{6} x^T R x + O(\|x\|^3)\right) dx$$

Using similar reasoning, it is easy to check that this quantity is also $1 + O(t)$ and the lemma is established. $\square$

We now turn to bounding $\|\mathbf{H}_t - \mathbf{E}_t\|$. We start with the following

Lemma 4.12 *Let $\mathcal{M}$ be a smooth compact Riemannian manifold. As above, let H_t be the corresponding heat kernel. Fix $\epsilon > 0$. For all $p, q \in \mathcal{M}$ such that $d(p, q) \geq \epsilon$, we have that for small t*

$$H_t(p, q) \leq C t^{3/2}$$

where $C > 0$ depends on ϵ and the invariants of the manifold.

Proof:The proof follows directly from estimates on the heat kernel obtained in the literature. See, for example, Theorem 1.1 of [17]. $\square$

Lemma 4.13

$$\|\mathbf{H}_t - \mathbf{E}_t\| \leq Ct$$

where C is a constant depending only upon the manifold.

PROOF:We begin by evaluating

$$(\mathbf{H}_t - \mathbf{E}_t)(f)(p) = \int_{\mathcal{M}} (H_t(p,q) - E_t(p,q))f(q)d\mu(q)$$

As before, we break this integral into a local and a global part. for the local part, we consider a geodesic ball around p given by $B_\epsilon(p) = \{q \in \mathcal{M} | d(p,q) < \epsilon\}$. We therefore have

$$(\mathbf{H}_t - \mathbf{E}_t)\,(f)(p) = A + B$$

$$A = \int_{B_\epsilon(p)} (H_t(p,q) - E_t(p,q))f(q)\,d\mu(q)$$

$$B = \int_{\mathcal{M}\setminus B_\epsilon(p)} (H_t(p,q) - E_t(p,q))f(q)\,d\mu(q)$$

We bound A and B separately. For the first term, we see that

$$A \leq \int_{B_\epsilon(p)} |H_t(p,q) - E_t(p,q)|\,|f(q)|\,d\mu(q)$$

Using Lemma 4.14, we see that on $B_\epsilon(p)$, there exists a constant C' such that

$$|H_t(p,q) - E_t(p,q)| \leq C't(H_{2t}(p,q) + H_t(p,q) + 1)$$

Therefore, we have

$$|A| \leq C't \int_{B_\epsilon(p)} (H_{2t}(p,q) + H_t(p,q) + 1)\,|f(q)|d\mu(q) \leq C''t\,(\mathbf{H}_{2t}|f|(p) + \mathbf{H}_t|f|(p) + \|f\|)$$

For the set $\mathcal{M} \setminus B_\epsilon(p)$, we have

$$|B| \leq \int_{\mathcal{M}\setminus B_\epsilon(p)} H_t|f(q)|d\mu(q) + \int_{\mathcal{M}\setminus B_\epsilon(p)} E_t|f(q)|d\mu(q)$$

Using using the arguments from Lemma 4.8 and Lemma 4.12, we know that $|E_t(p,q)|$ and $|H_t(p,q)|$ can be bounded by $O(t)$ when $d(p,q) > \epsilon$. Therefore,

$$|B| \leq C_1 t \int_{\mathcal{M}} |f(q)|d\mu(q) = C_1 t\|f\|_1 \leq C_2 t\|f\|$$

Putting these together, we have that

$$|\left(\mathbf{H}_t - \mathbf{E}_t\right)(f)(p)| \leq Ct\left((\mathbf{H}_{2t} + \mathbf{H}_t)(|f|)(p) + \|f\|_2\right)$$

Hence, by a standard argument,

$$\|(\mathbf{H}_t - \mathbf{E}_t)f\| \leq Ct(\|\mathbf{H}_{2t}(|f|)\| + \|\mathbf{H}_t(|f|)\| + \|f\|)$$

Using Lemma 4.15, we obtain

$$\|(\mathbf{H}_t - \mathbf{E}_t)f\| \leq 3Ct\|f\|$$

and the proposition is proved. □

Lemma 4.14 *For sufficiently small t and for all p, q sufficiently close, we have*

$$|H_t(p, q) - E_t(p, q)| \leq Ct(H_{2t}(p, q) + H_t(p, q) + 1)$$

Proof:¿From the asymptotic expansion of the heat kernel, it is known (see Rosenberg, 1997) that for p, q sufficiently close, there exist continuous functions $u_0(p, q)$ and $u_1(p, q)$ such that

$$|H_t(p, q) - E_t(p, q)(u_0(p, q) + tu_1)(p, q)| < C't$$

from which it follows that

$$|H_t(p, q) - E_t(p, q)| \leq E_t(p, q)|u_0(p, q) - 1| + tE_t(p, q)|u_1(p, q)| + C't$$

Using compactness of $\mathcal{M}$, we let $M = \sup_{p,q \in \mathcal{M}} |u_1(p, q)| < \infty$. Therefore, we have

$$|H_t - E_t| \leq E_t|u_0 - 1| + tE_tM + C't$$

Now using the fact that $u_0(p, q) = \det^{-\frac{1}{2}}(g_{ij}(q))$ (again, g is the metric tensor at point q; see Rosenberg, 1997), and using the asymptotic expansion in Eq. 11 of $\det(g)$, we have for a compact $\mathcal{M}$ that $u_0 = 1 + O(\|x\|^2)$. Therefore,

$$E_t|u_0 - 1| \leq C'' \frac{1}{(4\pi t)^{k/2}} e^{-\frac{\|x\|^2}{4t}} \|x\|^2$$

Letting $z = \frac{\|x\|}{\sqrt{t}}$, we have

$$\frac{1}{(4\pi t)^{k/2}} e^{-\frac{\|x\|^2}{4t}} \|x\|^2 = t \frac{1}{(4\pi t)^{k/2}} e^{-\frac{z^2}{4}} z^2 \leq t C_1 \frac{1}{(4\pi t)^{k/2}} e^{-\frac{z^2}{8}} = C_2 t E_{2t}$$

where the penultimate inequality makes use of the fact that $e^{-\frac{z^2}{4}} z^2 = e^{-\frac{z^2}{8}} e^{-\frac{z^2}{8}} z^2$ and that $e^{-\frac{z^2}{8}} z^2$ is a bounded function of z.

Therefore

$$|H_t - E_t| \leq Ct(E_{2t} + E_t + 1) \tag{12}$$

for some constant $C > 0$.

Finally, to prove the lemma, we notice that

$$|H_t - E_t| \leq E_t |u_0 - 1| + t E_t M + Bt \leq M'(E_t(d^2(p,q) + t) + t) \leq 2\epsilon M' E_t + \epsilon$$

for sufficiently small $t > 0$. Therefore, we have

$$E_t(1 - 2\epsilon M') \leq H_t + \epsilon$$

ultimately proving that there exists a constant $P > 0$ such that

$$E_t \leq P(H_t + 1)$$

Combining this with Eq. 12, the lemma is proved. □

Lemma 4.15 *Let $f \in L^2$. Then for any $t \geq 0$*

$$\|\mathbf{H}_t f\| \leq \|f\|$$

Proof:Write $f = \sum a_i e_i$. $\|f\|^2 = \sum a_i^2$. We have $\mathbf{H}_t f = \sum e^{-t\lambda_i} a_i e_i$ and

$$\|\mathbf{H}_t f\|^2 = \sum (e^{-t\lambda_i} a_i)^2 \leq \sum a_i^2$$

as $e^{-t\lambda_i} \leq 1$. □

Notice that in combination with Lemma 4.13, we have the following corollary:

Corollary 4.16 *Let $f \in L^2$. Then there exists a constant C such that for t sufficiently small*

$$\|\mathbf{E}_t f\| \leq C\|f\|$$

4.4 Proof of estimate for $H^{\frac{k}{2}+1}$ norm of $\mathbf{R}^t$ (Proposition 4.2).

To simplify the notation in this section we will denote the Sobolev space $H^{\frac{k}{2}+1}$ by H. We will first need the following standard fact (for reference see, e.g., **??**, Chapter 4):

Lemma 4.17 *Let $f \in H$. Then f is Lipschitz with the Lipschitz constant bounded by $C\|f\|_H$ for some universal constant C.*

Observe that for a smoothly embedded compact $\mathcal{M}$ Lipschitz functions on $\mathcal{M}$ are also Lipschitz in terms of the ambient distance. The ratio of the Lipschitz constants depends on the embedding of $\mathcal{M}$.

Proposition 4.18 *Let $f \in H$. Then there exists a constant C, such that for t sufficiently small*

$$\|\mathbf{R}_t f\|_2 \leq C\sqrt{t}\|f\|_H$$

PROOF:

We begin by using the fact that the constant function is an eigenfunction of $\mathbf{H}_t$ so that $1 = \int_{\mathcal{M}} H_t(x,y)d\mu(y)$. Therefore, we can write

$$\mathbf{R}_t f(p) = \frac{1}{t}\int_{\mathcal{M}} (H_t(p,q) - G_t(p,q))(f(p) - f(q))d\mu(q)$$

We bound this integral by writing it as a sum of two parts, choosing an appropriate $\epsilon > 0$ and considering the set $B_\epsilon(p) = \{q | d(p,q) < \epsilon\}$ and its complement in the usual way. Thus we have two integrals given by

$$A = \int_{B_\epsilon(p)} (H_t(p,q) - G_t(p,q))(f(p) - f(q))d\mu(q)$$

and

$$B = \int_{\mathcal{M}\setminus B_\epsilon(p)} (H_t(p,q) - G_t(p,q))(f(p) - f(q))d\mu(q)$$

respectively. Let us begin by bounding A. Using the exponential map $\exp : T_p \to \mathcal{M}$, we write $q = \exp(x)$ and get

$$A = \int_{B_\epsilon} (H_t - G_t)(f(0) - f(x))\sqrt{\det(g)}dx$$

Now we use the fact that for small $\|x\|$ and small t, we have

$$H_t = \frac{1}{(4\pi t)^{k/2}} e^{-\frac{\|x\|^2}{4t}} (u_0 + tu_1 + t^2 u_2) + O(t^2)$$

Additionally, using the fact that $u_0 = \det^{-\frac{1}{2}}(g)$, and the asymptotic expansion of $\det(g) = 1 - \frac{1}{6} x^T R x + O(\|x\|^3)$, we see that for small $\|x\|$ and sufficiently small t, we have

$$H_t = E_t + O(E_t(\|x\|^2 + t) + t^2)$$

By Lemma 4.10, we have

$$G_t = E_t + O(tE_{2t})$$

Therefore,

$$|H_t - G_t| \leq C(E_t(\|x\|^2 + t) + t^2 + tE_{2t})$$

Since the Lipschitz constant of f is controlled by H norm, we have

$$|f(0) - f(x)| \leq C\|f\|_H \|x\|$$

and

$$|A| \leq C\|f\|_H \frac{1}{(4\pi t)^{k/2}} \int_{B_\epsilon} \left((E_t(\|x\|^2 + t) + t^2) + tE_{2t}\right) \|x\| \sqrt{det(g)} dx$$

It is easy to check that this gives us

$$|A| \leq Ct^{3/2} \|f\|_H$$

For bounding B, we simply make use of the fact that on $\mathcal{M} \setminus B_\epsilon$, both H_t and G_t are $O(t^{3/2})$ (Lemma 4.12 and the argument of lemma 4.8) and $|f(p) - f(q)| = O(\|f\|_H \sup_{p,q \in \mathcal{M}} d(p,q))$ to see that

$$|B| \leq Ct^{3/2} \|f\|_H$$

Putting these together, we see that

$$|\mathbf{R}_t f(p)| = |\frac{1}{t}(A + B)| \leq Ct^{1/2} \|f\|_H$$

The proposition follows immediately.

□

5 Spectral Convergence of Empirical Approximation

The discussion in section relies on technical results obtained in previous work.

We will now need the following

Proposition 5.1 *For t sufficiently small*

$$\mathrm{SpecEss}\,(\mathbf{L}_t) \subset \left(\frac{1}{2}t^{-1}, \infty\right)$$

where SpecEss *denotes the essential spectrum of the operator.*

PROOF:As noted before $\mathbf{L}_t f$ is a difference of a multiplication operator and a compact operator

$$L^t f(x) = g(x)f(x) - \mathbf{K}f \tag{13}$$

where

$$g(x) = (4\pi t)^{-\frac{k+2}{2}} \int_{\mathcal{M}} e^{-\frac{\|x-y\|^2}{4t}}\, d\mu_y$$

and $\mathbf{K}f$ is a convolution with a Gaussian. As noted in [25], it is a fact in basic perturbation theory (e.g., [12]) that

$$\mathrm{SpecEss}\,(\mathbf{L}_t) = \mathrm{rg}\, g$$

where $\mathrm{rg}\, g$ is the range of the function $g : \mathcal{M} \to \mathbb{R}$. To estimate $\mathrm{rg}\, g$ observe first that

$$\lim_{t\to\infty} (4\pi t)^{-\frac{k}{2}} \int_{\mathcal{M}} e^{-\frac{\|p-y\|^2}{4t}}\, d\mu_y = 1$$

We thus see that for t sufficiently small

$$(4\pi t)^{-\frac{k}{2}} \int_{\mathcal{M}} e^{-\frac{\|p-y\|^2}{4t}}\, d\mu_y > \frac{1}{2}$$

and hence $g(t) > \frac{1}{2}t^{-1}$. □

It is a well-known fact that all eigenfunctions of $\Delta_{\mathcal{M}}$ are infinitely differentiable.

Lemma 5.2 *Let e^t be an eigenfunction of $\mathbf{L}_t$, $\mathbf{L}_t e^t = \lambda^t e^t$, $\lambda^t < \frac{1}{2}t^{-1}$. Then $e^t \in \mathcal{C}^\infty$.*

PROOF:Write $\mathbf{L}_t e(x) = g(x)e(x) - \mathbf{K}e(x)$ as in Eq. 13. We have

$$e(x) = \frac{\mathbf{K}e(x)}{\lambda^t - g(x)}$$

Since $\mathbf{K}$ is a convolution operator $\mathbf{K}e \in \mathcal{C}^\infty$. Since $\lambda^t \notin \text{rg } g$, we see that the ratio is in $\mathcal{C}^\infty$ as well. □

We see that Theorem 3.2 follows easily:

PROOF:[Theorem 3.2] By the Proposition 5.1 we see that the part of the spectrum of L^t between 0 and $\frac{1}{2}t^{-1}$ is discrete. It is a standard fact of functional analysis (follows from a general form of the Spectral Theorem, e.g. [12]) that such points are eigenvalues and there are corresponding eigenspaces of finite dimension. For simplicity we will assume that all multiplicities are one.

Consider now $\lambda_i^t \in [0, \frac{1}{2}t^{-1}]$ and the corresponding eigenfunction e_i^t.

The Theorem 4 then follows from Theorem 23 and Proposition 25 in [25], which show convergence of spectral properties for the empirical operators. □

6 Main Theorem

We are finally in position to prove the main Theorem 2.1.

PROOF:[Theorem 2.1] ¿From Theorems 3.1 and Theorem 3.2 we obtain the following convergence results:

$$\textbf{Eig } \hat{\mathbf{L}}_{t,n} \xrightarrow{n \to \infty} \textbf{Eig } \mathbf{L}_t \xrightarrow{t \to 0} \textbf{Eig } \Delta_{\mathcal{M}}$$

where the first convergence is almost surely for $\lambda_i \leq \frac{1}{2}t^{-1}$. To see how to choose a sequence t_n, we express t and n in terms of a common integer parameter j. Let $t = \frac{1}{j}$. For every j, i.e., every $t_j = \frac{1}{j}$, pick n_j such that

$$\forall i \text{ such that } \lambda_i^{t_j} < \frac{1}{2t_j}, \mathbb{P}\left\{\|e_{n_j,i}^{t_j} - e_i^{t_j}\| \geq \frac{1}{j}\right\} \leq \frac{1}{j}$$

Such an n_j always exists by the convergence implied in Theorem 3.2. We arrange it so that n_j is an increasing sequence. Then for any $i \in \mathbb{N}$, we see that

$$\mathbb{P}\left\{\|e_{n_j,i}^{t_j} - e_i\| > \epsilon\right\} \leq \mathbb{P}\left\{\|e_{n_j,i}^{t_j} - e_i^{t_j}\| + \|e_i^{t_j} - e_i\| > \epsilon\right\}$$

By the convergence implied in Theorem 3.1, there exists a J such that for all $j > J$, we have $\|e_i^{t_j} - e_i\| \leq \frac{\epsilon}{2}$. Therefore, for all $j > J$, we have

$$\mathbb{P}\left\{\|e_{n_j,i}^{t_j} - e_i\| > \epsilon\right\} \leq \mathbb{P}\left\{\|e_{n_j,i}^{t_j} - e_i^{t_j}\| > \frac{\epsilon}{2}\right\}$$

On the other hand, for all $j > \max(2\lambda_i^{t_j}, \frac{2}{\epsilon})$, we have

$$\mathbb{P}\left\{\|e_{n_j,i}^{t_j} - e_i^{t_j}\| > \frac{\epsilon}{2}\right\} \leq \mathbb{P}\left\{\|e_{n_j,i}^{t_j} - e_i^{t_j}\| > \frac{1}{j}\right\} \leq \frac{1}{j}$$

Thus it follows that for any $\epsilon > 0$ and any $i \in \mathbb{N}$,

$$\lim_{j\to\infty} \mathbb{P}\left\{\|e_{n_j,i}^{t_j} - e_i\| > \epsilon\right\} = 0$$

Inverting the relationship between t_j and n_j allows us to choose a sequence t_n such that

$$\lim_{n\to\infty} \mathbb{P}\left\{\|e_{n,i}^{t_n} - e_i\| > \epsilon\right\} = 0$$

$\square$

A Note on Rates: Rates of convergence may be easily derived from the exposition here with some additional considerations. In [25], the rate of convergence of $\hat{\mathbf{L}}_n^t$ to $\mathbf{L}^t$ as a function of n was obtained for each fixed t. From Theorem 4.3, we explicitly obtain the rate of convergence of $\mathbf{L}^t$ to $\frac{\mathbf{1}-\mathbf{H}_t}{t}$ as a function of t. Putting these together, appropriate rates may be obtained. A preliminary analysis suggests that any t_n satisfying $\lim_{n\to\infty} t_n = 0$ and $\lim_{n\to\infty} nt_n^{k+2} = \infty$ is sufficient to guarantee convergence. We leave a more complete analysis for the future.

7 Conclusions

The graph Laplacian associated to the data, the corresponding point cloud Laplace operator and the manifold Laplace-Beltrami operator play a central role in our understanding of a class of algorithms for data analysis. In this paper we have shown that the eigenfunctions of the point cloud graph Laplacian converge to eigenfunctions of the manifold Laplacian in probability. This provides a first consistency result for the Laplacian Eigenmaps algorithms.

This basic result has implications for a number of additional algorithms in data analysis, clustering, and machine learning that use ideas from spectral geometry. It also suggests how to perform computational harmonic analysis on an unknown manifold, how to reconstruct the heat kernel (for large t), and how to solve partial differential equations of of a suitable type from random point samples. There are connections to geometric random graphs, random matrices, and the applications of these to algorithm analysis and design in graphics, scientific computing, and sensor networks.

A variety of questions arise for the future. These include a more complete analysis of rates of convergence, an understanding of the robustness of our estimates to noise in the data, and generalizations of this kind of result to other operators. In a more geometric direction, it makes one wonder if, more generally, one might be able to recover the Laplace-Beltrami operator on differential forms from random samples and thus construct a probabilistic convergence theory to parallel the deterministic convergence theory as seen in the work of Dodziuk and Patodi [9, 10].

Acknowledgements. We thank Vladimir Koltchinskii, Peter Constantin, Ulrike von Luxburg, and Todd Dupont for helpful conversations.

References

[1] Applied and Computational Harmonic Analysis, Special Issue: Diffusion Maps and Wavelets, Volume 21, Issue 1, Pages 1-144 (July 2006).

[2] M. Belkin, *Problems of Learning on Manifolds*, The University of Chicago, Ph.D. Dissertation, 2003.

[3] M. Belkin, P. Niyogi, *Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering*, NIPS 2001.

[4] M. Belkin, P. Niyogi, *Towards a Theoretical Foundation for Laplacian-Based Manifold Methods*, COLT 2005.

[5] M. Belkin, P. Niyogi, *Towards a Theoretical Foundation for Laplacian-Based Manifold Methods*, to appear, Journal of Computer and System Sciences, 2007.

[6] Y. Bengio, P. Vincent, J.-F. Paiement, O. Delalleau, M. Ouimet, and N. Le Roux. *Spectral clustering and kernel PCA are learning eigenfunctions.* Technical Report TR 1239, University of Montreal, 2003.

[7] M. Bernstein, V. de Silva, J.C. Langford, J.B. Tenenbaum, *Graph approximations to geodesics on embedded manifolds*, Technical Report, 2000.

[8] O. Bousquet, O. Chapelle, M. Hein, *Measure Based Regularization*, NIPS 2003.

[9] J. Dodziuk, *Finite Difference Approach to the Hodge Theory of Harmonic Forms*, American Journal of Mathematics, vol. 98, 1976.

[10] J. Dodziuk and V. K. Patodi, *Riemannian structures and Triangulations of Manifolds,* Journal of the Indian Mathematical Society, vol. 40, 1976.

[11] T. Kato, *Perturbation Theory for Linear Operators*, Springer, 1966.

[12] F. Chatelin. (1983). *Spectral Approximation ok Linear Operators.* Computer Science and Applied Mathematics.

[13] F. R. K. Chung. (1997). *Spectral Graph Theory.* Regional Conference Series in Mathematics, number 92.

[14] R.R.Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner and S. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure definition of data*, submitted to the Proceedings of the National Academy of Sciences (2004).

[15] D. L. Donoho, C. E. Grimes, *Hessian Eigenmaps: new locally linear embedding techniques for high-dimensional data*, Proceedings of the National Academy of Arts and Sciences vol. 100 pp. 5591-5596.

[16] E. Gine and V. Koltchinskii. Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results. Proceedings of the 4th International Conference on High Dimensional Probability., pp. 238-259. IMS Lecture Notes 51, Beachwood, OH

[17] A. Grigoryan, *Gaussian upper bounds on the heat kernel for arbitrary manifolds*, Journal of Differential Geometry, vol. 45, 1997.

[18] M. Hein, J.-Y. Audibert, U. von Luxburg, *From Graphs to Manifolds – Weak and Strong Pointwise Consistency of Graph Laplacians*, COLT 2005.

[19] T. Kato, *Perturbation Theory For Linear Operators*, Springer-Verlag, 1966.

[20] B. Kegl. *Intrinsic dimension estimation using packing numbers.* NIPS 2002.

[21] V. Koltchinskii, E. Gine. *Random matrix approximation of spectra of integral operators*. Bernoulli, 6(1):113 167, 2000.

[22] F. Memoli, G. Sapiro, *Comparing Point Clouds*, IMA Technical Report, 2004.

[23] S. Lafon, *Diffusion Maps and Geodesic Harmonics*, Ph. D. Thesis, Yale University, 2004.

[24] E. Levina and P.J. Bickel (2005). *Maximum Likelihood Estimation of Intrinsic Dimension.*, NIPS 04.

[25] U. von Luxburg, M. Belkin, O. Bousquet, *Consistency of Spectral Clustering*, Max Planck Institute for Biological Cybernetics Technical Report TR 134, 2004.

[26] P. Niyogi, S. Smale, S. Weinberger, *Finding the Homology of Submanifolds with High Confidence from Random Samples*, Univ. of Chicago Technical Report TR-2004-08, 2004.

[27] S. Rosenberg, *The Laplacian on a Riemannian Manifold*, Cambridge University Press, 1997.

[28] Sam T. Roweis, Lawrence K. Saul. (2000). *Nonlinear Dimensionality Reduction by Locally Linear Embedding*, Science, vol 290.

[29] A. Singer, *From graph to manifold Laplacian: the convergence rate*, Applied and Computational Harmonic Analysis, 21 (1), 135-144 (2006).

[30] O.G. Smolyanov, H.v. Weizscker, and O. Wittich. *Brownian motion on a manifold as limit of stepwise conditioned standard Brownian motions.* In Stochastic processes, Physics and Geometry: New Interplays. II: A Volume in Honour of S. Albeverio, volume 29 of Can. Math. Soc. Conf. Proc., pages 589-602. Am. Math. Soc., 2000.

[31] J.B.Tenenbaum, V. de Silva, J. C. Langford. (2000). *A Global Geometric Framework for Nonlinear Dimensionality Reduction*, Science, Vol 290.

[32] A. Zomorodian, G. Carlsson, *Computing persistent homology*, 20th ACM Symposium on Computational Geometry, 2004.