SPECTRAL CONVERGENCE OF GRAPH LAPLACIAN AND HEAT KERNEL RECONSTRUCTION IN $L^\infty$ FROM RANDOM SAMPLES

DAVID B DUNSON, HAU-TIENG WU, AND NAN WU

ABSTRACT. In the manifold setting, we provide a series of spectral convergence results quantifying how the eigenvectors and eigenvalues of the graph Laplacian converge to the eigenfunctions and eigenvalues of the Laplace-Beltrami operator in the $L^\infty$ sense. Based on these results, convergence of the proposed heat kernel approximation algorithm, as well as the convergence rate, to the exact heat kernel is guaranteed. To our knowledge, this is the first work exploring the spectral convergence in the $L^\infty$ sense and providing a numerical heat kernel reconstruction from the point cloud with theoretical guarantees.

1. INTRODUCTION

The graph Laplacian (GL) is a tool developed in the spectral graph theory community [9], and it has been widely applied in the machine learning society to design various algorithms. It has a natural interpretation as the infinitesimal generator of a random walk on a point cloud, or the associated graph.

Given data points sampled from a manifold, the GL is defined based on the affinity matrix constructed from determining how similar each pair of points are. It has been well known that under the manifold setup, the GL asymptotically converges to the Laplace-Beltrami operator. In the past decades, in addition to a rich methodological literature on the GL and its application, there is a lot of theoretical literature describing how the GL converges to the Laplace-Beltrami operator. We briefly review these results. The pointwise convergence of the GL to the Laplace-Beltrami operator has been discussed in several places, like [10, 28, 17], where the convergence rate is provided. The pointwise convergence result shows that when the GL is applied to the discretization of a test function over the data points, it converges to the value of the Laplacian of the test function at each point. In [10], the pointwise convergence is carried out when the sampling is non-uniform from a compact manifold with boundary, where the affinity is constructed by the $\alpha$-normalization with the Gaussian kernel (see (1) below); In [28], the convergence is carried out when the data points are uniformly i.i.d. sampled from a compact manifold without boundary, where the affinity matrix is constructed from the Gaussian kernel; In [17], the convergence is carried out when the i.i.d. sampling might be non-uniform from a compact manifold with boundary, where the affinity matrix is constructed from a compactly supported kernel. However, a pointwise convergence result might not guarantee a spectral convergence. To this end, there have been several results about the $L^2$ spectral convergence of GL on a compact manifold without boundary, with and without convergence rate. To our knowledge, the first result in this direction should be [2], where the authors prove the $L^2$ spectral convergence to the Laplace-Beltrami operator when the point cloud is i.i.d. uniformly sampled from a compact manifold without boundary, where the affinity matrix is constructed from...
the Gaussian kernel. In this work, the convergence rate is not provided. By using the variational method over the Dirichlet form, an \( L^2 \) spectral convergence to the Laplace-Beltrami operator is shown in [4] with a convergence rate, where the GL is constructed under a deterministic setting and the manifold is compact without boundary. This variational method is further extended in [31, 5] to study the same problem under a probabilistic setting, where the optimal transport technique is applied. In these papers, the authors prove the \( L^2 \) spectral convergence to a weighted Laplace-Beltrami operator that depends on the density function with a rate when the point cloud is sampled i.i.d. (non-)uniformly from a compact manifold without boundary and the affinity matrix is constructed from a compactly supported kernel. When the affinity matrix is constructed by renormalizing the Gaussian kernel, the authors in [35] prove an \( L^\infty \) spectral convergence with a rate to the Laplace-Beltrami operator when the point cloud is sampled (non-)uniformly i.i.d. on flat tori of any dimensions. Based on the results in [31, 5], the authors in [6] show an \( L^\infty \) spectral convergence with a rate to a weighted Laplace-Beltrami operator that depends on the density function when the point cloud is sampled i.i.d. (non-)uniformly from a compact manifold without boundary and the affinity matrix is constructed from a compactly supported kernel. We mention that the spectral convergence of the graph connection Laplacian is proved in [30], while the rate is not provided.

Recall that the Laplace-Beltrami operator is directly related to the heat kernel. Specifically, the heat kernel can be expressed as a series sum in terms of the eigenpairs of the Laplace-Beltrami operator. Therefore, if we can reconstruct the eigenvalues and eigenfunctions of the Laplace-Beltrami operator, we may recover the heat kernel. While there is a rich theoretical literature on the GL, however, to the best of our knowledge, our work is the first to use GL in approximating the heat kernel over an unknown manifold from a point cloud of data points. To achieve this purpose, we are also the first team to provide the spectral convergence of the eigenvectors of the GL to the eigenfunctions of the Laplace-Beltrami operator in the \( L^\infty \) sense with the convergence rate.

In Section 2, we provide background on the heat kernel on manifolds. In Section 3.1, we recall the GL. Section 3.2 contains our main theoretical results. We show the spectral convergence rate of the GL to the Laplace-Beltrami operator of the manifold in the \( L^\infty \) sense. We also provide the convergence rate from the heat kernel we recovered to the actual heat kernel. Section 4 consists of some fundamental facts to prove the main theorems. The proofs of the main theorems are organized in the Section 5 and Section 6. The proofs of some technical results are postponed to the Appendix. In the whole paper, we consider a compact and smooth \( d \)-dim manifold \( M \), and use \( \| \cdot \|_2 := \| \cdot \|_{L^2(M)} \) be the \( L^2 \) norm, \( \| \cdot \|_\infty := \| \cdot \|_{L^\infty(M)} \) be the \( L^\infty \) norm, and let \( \| \cdot \| \) be the operator norm in the relevant Banach space.

2. Background Knowledge about Heat Kernel

In this section, we recall the heat kernel of a manifold. Let \( M \) be a \( d \) dimensional smooth closed Riemannian manifold with the Riemannian metric \( g \). The heat kernel, \( H : M \times M \times [0, \infty) \rightarrow \mathbb{R} \), is the fundamental solution of the heat equation; namely, for any \( x, x' \in M \),

\[
\frac{\partial H(x, x', t)}{\partial t} = \Delta_x H(x, x', t),
\]

where \( \Delta_x \) is the Laplace-Beltrami operator of the manifold \( M \) acting on the first parameter of \( H \) and \( t > 0 \) is the diffusion time. The heat kernel satisfies the initial condition \( \lim_{t \rightarrow 0} H(x, x', t) = \delta_{x'}(x) \), where \( \delta_{x'} \) is a delta measure supported at \( x' \in M \), and the limit
is understood in the distributional sense. When the manifold is closed, the heat kernel on the manifold can be related to the geodesic distance by Varadhan’s formula \(32\).

\[
\lim_{t \to 0} -4t \log H(x, x', t) = d(x, x')^2,
\]

where \(d(x, x')\) is the geodesic distance between \(x\) and \(x'\). Take the special case when \(M = \mathbb{R}^d\) as an example. In this case, it is well known that the heat kernel is the Gaussian kernel:

\[
H(x, x', t) = (4\pi t)^{-d/2} e^{-\frac{|x-x'|^2}{4t}}.
\]

If we understand \(|x - x'|\) as the geodesic distance between \(x\) and \(x'\) in \(\mathbb{R}^d\), then based on the \(\mathbb{R}^d\) case and Varadhan’s formula, one may expect to use \((4\pi t)^{-d/2} e^{\frac{d|x-x'|^2}{4t}}\) to approximate the heat kernel on a manifold and use the methods like \([22, 29, 19, 21]\) to estimate the geodesic distance. Unfortunately, such an approximation can be quite inaccurate. When \(d|x-x'|\) and \(t\) are small, the heat kernel has the following expansion \([27]\).

\[
H(x, y, t) = (4\pi t)^{-d/2} e^{\frac{d|x-y|^2}{4t}} (u_0 + tu_1 + O(t^2)),
\]

where \(u_0(x, x') = \frac{1}{\sqrt{|\det(\exp_x^{-1} \circ \exp_y^{-1})|}}\) and \(\exp_x\) is the exponential map at \(x\). We mention that \(u_0(x, x') = 1 + O(d(x, x')^2)\), where the implied constant in \(O(d(x, x')^2)\) depends on the Ricci curvature of the manifold. Also, \(u_1\) is a continuous function of \(x\) and \(x'\) that only depends on the manifold. Such an expansion shows that using \((4\pi t)^{-d/2} e^{\frac{d|x-x'|^2}{4t}}\) to approximate the heat kernel leads to a large error when \(x\) and \(x'\) are close, even if we know the geodesic distance.

3. Graph Laplacian and Heat Kernel Approximation

The graph Laplacian (GL) is a fundamental tool in the spectral graph theory \([9]\). In this section, we start with defining the GL. We then show how to apply the kernel normalized GL to approximate the heat kernel of the manifold. Moreover, we provide a theoretical justification of the convergence rate of the approximation.

3.1. Graph Laplacian. Take a dataset \(\mathcal{X} := \{x_i\}_{i=1}^n \subset \mathbb{R}^D\). Consider a kernel function \(k_\varepsilon(x, x') = \exp\left(-\frac{|x-x'|^2}{2\varepsilon^2}\right)\), where \(\varepsilon > 0\) is the bandwidth. We mention that in practice we can consider a more general kernel, but to simplify the discussion, we focus on this Gaussian-like kernel.

Take \(\alpha \in [0, 1]\). We construct an \(n \times n\) affinity matrix \(W\) as

\[
W_{ij}^{(\alpha)} := \frac{k_\varepsilon(x_i, x_j)}{q_\varepsilon(x_i)^\alpha q_\varepsilon(x_j)^\alpha},
\]

where \(q_\varepsilon(x) := \sum_{i=1}^n k_\varepsilon(x, x_i)\) for \(x \in \mathbb{R}^D\). This affinity matrix contains affinities among pairs of points after \(\alpha\)-normalization \([19]\). Note that the term \(q_\varepsilon(x)\) is related to the kernel density estimation (KDE) evaluated at \(x\). The normalization of the kernel \(k_\varepsilon(x_i, x_j)\) by \(q_\varepsilon(x_i)q_\varepsilon(x_j)\) is aiming to remove the impact of the non-uniform sampling density. By defining an \(n \times n\) diagonal matrix \(D\) as

\[
D_n^{(\alpha)} = \sum_{j=1}^n W_{ij}^{(\alpha)},
\]

where
where \( i = 1, \ldots, n \), we define the row stochastic transition matrix as
\[
A^{(\alpha)} = D^{(\alpha)^{-1}}W^{(\alpha)}.
\]

Our main quantity of interest is the kernel normalized GL matrix, which is defined as
\[
L^{(\alpha)} := \frac{A^{(\alpha)} - I}{\epsilon^2}.
\]

In this work, we have interest in two special cases; \( \alpha = 0 \) and \( \alpha = 1 \). When \( \alpha = 1 \), we call \( L := L^{(1)} \) the kernel normalized GL, and \( W := W^{(1)} \), \( D := D^{(1)} \) and \( A := A^{(1)} \); when \( \alpha = 0 \), we call \( L^{un} := L^{(0)} \) the unnormalized GL, and \( W^{un} := W^{(0)} \), \( D^{un} := D^{(0)} \) and \( A^{un} := A^{(0)} \).

**Remark 1.** We shall comment on the nomination of \( L^{un} = L^{(0)} \) as the unnormalized GL. Sometimes, the name “unnormalized GL” refers to \( D^{(0)} - W^{(0)} \), while \( I - A^{(0)} \) is called the random walk GL. However, since \( A^{(1)} \) is also a transition matrix associated with a random walk, we decided to call \( L^{un} \) the unnormalized GL for the sake of comparing with \( L = L^{(1)} \) in which the kernel is normalized as in (1).

Note that \( L \) is a linear translation and scaling of \( A \), so on the linear algebra level, we only need to study \( A \) if we are interested in the spectral structure of \( L \). Since \( A \) is similar to \( \tilde{A} = D^{-1/2}WD^{-1/2} \), \( \tilde{A} \) is diagonalizable. The eigenvalues of \( A \) and \( \tilde{A} \) are the same, with the smallest eigenvalue bounded below by zero and all the values falling in \([0, 1]\) since the chosen Gaussian kernel is positive definite by the Bochner’s theorem. Hence, if \( \mu \) is an eigenvalue of \( -L \), then \( 0 \leq \mu \leq \frac{1}{\epsilon^2} \) and the smallest eigenvalue of \( -L \) is bounded from below by 0. The same discussion holds for \( L^{un} \).

### 3.2. Spectral convergence of the kernel normalized Graph Laplacian and the heat kernel reconstruction.

Now, we discuss how to estimate the heat kernel by the kernel normalized GL when the data are sampled from a closed Riemannian manifold \( M \) embedded in \( \mathbb{R}^D \).

**Assumption 1.** Let \( M \) be a \( d \)-dimensional smooth, closed and connected Riemannian manifold isometrically embedded in \( \mathbb{R}^D \) through \( t : M \rightarrow \mathbb{R}^D \). Suppose that \( p \) is a smooth probability density function (p.d.f.) on the manifold \( M \). We assume that \( p \) is bounded from below by \( p_0 > 0 \). Suppose the point cloud \( \mathcal{X} := \{x_1, \ldots, x_n\} \) is independently and identically (i.i.d.) sampled following the density function \( p \); that is, we have \( \mathcal{X} \subset M \). When \( p \) is a constant function, that is, \( p = \frac{1}{\#\mathcal{X}} \), we say that the sampling is uniform; otherwise it is non-uniform.

Let \( \Delta \) be the Laplace-Beltrami operator of \( M \). Let \( \sigma(-\Delta) := \{\lambda_i\}_{i=0}^\infty \) be the spectrum of \( -\Delta \). By standard elliptic theory, we have \( 0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \) and each eigenvalue has a finite multiplicity. Denote \( \phi_i \) to be the corresponding eigenfunction normalized in \( L^2(M) \); that is, for each \( i \in \mathbb{N} \), we have \( \Delta \phi_i = -\lambda_i \phi_i \). Denote \( \mu_{i,n,\epsilon} \) to be the \( i \)-th eigenvalue of \( -L \) with the associated eigenvector \( \tilde{\nu}_{i,n,\epsilon} \) normalized in the \( L^2 \) norm, where \( i = 1, \ldots, n \). We order \( \mu_{i,n,\epsilon} \) so that \( 0 = \mu_{0,n,\epsilon} \leq \mu_{1,n,\epsilon} \leq \cdots \leq \mu_{n-1,n,\epsilon} \leq \frac{1}{\epsilon^2} \).

Recall that the heat kernel has the following expression:
\[
H(x, x', t) = \sum_{i=0}^\infty e^{-\lambda_i t} \phi_i(x) \phi_i(x').
\]

Therefore, intuitively, if we are able to recover the eigenfunctions and eigenvalues of the Laplace-Beltrami operator through \( L \), we may be able to recover the heat kernel via the same formula, like \( \sum_{i=0}^\infty e^{-\mu_{i,n,\epsilon} t} \tilde{\nu}_{i,n,\epsilon} \tilde{\nu}_{i,n,\epsilon}^\top \). Below, we show that this intuition is correct, if we carefully carry out the estimation. We mention that the idea of using the kernel
normalized GL to gain benefit from the heat kernel has been applied in several algorithms, like the nonlinear dimension reduction algorithm diffusion map (DM) \cite{10}.

It has been well known that the matrix $L$ converges "pointwisely" to the Laplace-Beltrami operator of the manifold, which we summarize below.

**Theorem 1** (e.g. \cite{10,29}). Suppose $f \in C^4(M)$. if $\varepsilon = \varepsilon(n)$ so that \( \frac{\sqrt{\log n}}{\sqrt{n}} \to 0 \) and $\varepsilon \to 0$ as $n \to \infty$, then with probability greater than $1 - n^{-2}$, for all $x_i$, we have

\[
\sum_{j=1}^{n} L_{ij} f(x_j) = \Delta f(x_i) + O(\varepsilon^2) + O\left(\frac{\sqrt{\log n}}{n^{1/2} \varepsilon^{d/2+2}}\right).
\]

Since this fact has been proved in several places, we omit the proof. Note that Theorem 1 does not directly link the eigenvectors of $L$ and the eigenfunctions of the Laplace-Beltrami operator. Moreover, Theorem 1 is only a pointwise convergence result, which is not strong enough to guarantee the spectral convergence.

**Remark 2.** Denote the k-th moment of the kernel as $m_k := \int_{\mathbb{R}^d} \|u\|^k e^{-\|u\|^2/4\varepsilon^2} \, dx$, where $k = 0, 1, \ldots$. Note that we have $m_0 = (4\pi \varepsilon^2)^{d/2}$ and $m_2 = (4\pi \varepsilon^2)^{d/2} 2\varepsilon^2$, so a straightforward Taylor expansion in the normal coordinates shows that the coefficient is 1 in front of $\Delta f(x)$.

**Remark 3.** Note that the result shown in Theorem 1 is associated with the $1$-normalization; that is, $\alpha = 1$ in the $\alpha$-normalization proposed in \cite{10}. On the other hand, $L^{un}$ converges to a second order differential operator that is influenced by the p.d.f. if the p.d.f. is not uniform. In this case, the pointwise convergence rate is $O\left(\frac{\sqrt{\log n}}{\sqrt{n} \varepsilon^{d+2}}\right)$, which is faster than that shown in Theorem 1. The difference of the convergence rate comes from the KDE, $q$. Hence, if the dataset is uniformly sampled, one can simply consider $L^{un}$ to achieve a faster pointwise convergence rate.

Next we describe our main results about recovering the eigenvalues and eigenfunctions of the Laplace-Beltrami operator from the kernel normalized GL in the $L^\alpha$ sense. To state our main contribution, we need the following normalization. Since the eigenfunction $\phi_i$ of the Laplace-Beltrami operator is normalized in the $L^2(M)$ norm, in order to compare the $i$-th eigenvector of $L$ and $\phi_i$, we have to make sure that $\tilde{v}_{i,n,\varepsilon}$ is properly normalized.

**Definition 1.** Under Assumption 1, suppose $\tilde{v}$ is an eigenvector of $L$ which is normalized in the $L^2$ norm. Let $\mathcal{N}(i) = |B_{\varepsilon}^p(x_i) \cap \{t(x_1), \ldots, t(x_n)\}|$, which is the number of points in the $\varepsilon$ ball in the ambient space. Then, we define the $L^2$ norm of $\tilde{v}$ with respect to inverse estimated probability density $1/\hat{p}$ as:

\[
\|\tilde{v}\|_{L^2(1/\hat{p})} := \sqrt{\frac{S_{d-1}}{d} \sum_{i=1}^{n} \frac{n^2(i)}{\mathcal{N}(i)}}.
\]

Note that in this definition, the term $\frac{d\mathcal{N}(i)}{S_{d-1}|x_i|}$ is an estimation of the probability density $p$ at $x_i$ by the $0 - 1$ kernel over the Euclidean ball centered at $t(x_i)$ with radius $\varepsilon$. The reader may refer to Lemma 17 for details. A more sophisticated KDE can be applied here, but we are not concerned with it to simplify the discussion. Then, define

\[
v_{i,n,\varepsilon} = \frac{\tilde{v}_{i,n,\varepsilon}}{||\tilde{v}_{i,n,\varepsilon}\|_{L^2(1/\hat{p})}}.
\]

Intuitively, $v_{i,n,\varepsilon}$ can be regarded as a discretization of some function that is normalized in the $L^2(M)$ norm. A rigorous discussion can be found in Proposition 5 in Section 5. In the
following theorem, we show that $v_{i,n,\epsilon}$ is an approximation of $\phi_i$ over the $n$ data points with high probability.

The following theorem says that, on a closed manifold $M$, if we fix $K$ and we choose the bandwidth $\epsilon$ based on the number of data points $n$, then for $i < K$, with high probability, $\mu_{i,n,\epsilon}$ is an approximation of the $i$-th eigenvalue $\lambda_i$ of $-\Delta$. The proof of the theorem is in Section 5.

**Theorem 2. (Spectral convergence of the kernel normalized GL)** Under Assumption 2, suppose all the eigenvalues of $\Delta$ are simple. Fix $K \in \mathbb{N}$, and denote $\Gamma_K := \min_{1 \leq i \leq K} \text{dist}(\lambda_i, \sigma(-\Delta) \setminus \{\lambda_i\})$. Suppose

\[
\epsilon \leq \mathcal{K}_1 \min \left( \frac{\min(\Gamma_K, 1)}{\mathcal{K}_2 + \lambda_K^{d/2+3}}, \frac{1}{(\mathcal{K}_3 + \lambda_K^{(5d+7)/4})^2} \right),
\]

where $\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3 > 1$ are constants depending on $d$, $p_m$, the $C^2$ norm of $p$, and the volume, the injectivity radius, the curvature and the second fundamental form of the manifold. Then, when $n$ is sufficiently large so that $\epsilon = \epsilon(n) \geq \left( \frac{\log n}{n} \right)^{1/4}$, with probability greater than $1 - n^{-2}$, for all $0 \leq i < K$,

\[
|\mu_{i,n,\epsilon} - \lambda_i| \leq \Omega_1 \epsilon^{3/2}.
\]

Suppose (6) is satisfied. Then, when $n$ is sufficiently large so that $\epsilon = \epsilon(n) \geq \left( \frac{\log n}{n} \right)^{1/4}$, with probability greater than $1 - n^{-2}$, there are $a_i \in \{1, -1\}$ such that for all $0 \leq i < K$,

\[
\max_{x_j \in \mathcal{X}} |a_i v_{i,n,\epsilon}(j) - \phi_i(x_j)| \leq \Omega_2 \epsilon^{1/2}.
\]

$\Omega_1$ depends on $d$, the diameter of $M$, $p_m$, and the $C^2$ norm of $p$, and $\Omega_2$ depends on $d$, the diameter and the volume of $M$, $p_m$, and the $C^2$ norm of $p$.

Note that $\Gamma_K$ is the smallest spectral gap between the first $K$ eigenvalues. Hence, $\Gamma_K$ is a non-increasing function of $K$. When $d = 1$, $\Gamma_K = \lambda_2$ which is a constant only depending on the arclength of the curve. However, when $d \geq 2$, as $K$ increases, it is possible that $\Gamma_K = |\lambda_{i+1} - \lambda_i|$ for some $i$ close to $K$. No matter in which case, since we consider a fixed number of eigenvalues and eigenvectors, this eigengap dependence is fixed when $K$ is fixed.

Since $\lambda_K$ grows polynomially, the theorem emphasizes that if we want to recover more eigenvalues and eigenvectors, the $\epsilon$ needs to get polynomially smaller. In practice, we may only want to recover a fixed number of eigenvalues and eigenvectors, and we may not need to focus on this dependence. See Section 3.3 for a discussion in this direction.

We want to point it out that we assume that the eigenvalues of $\Delta$ are simple to simplify the discussion. In the case when the eigenvalues are not simple, the same proof still works by introducing the eigenprojection [8].

**Remark 4.** We shall mention that while the error of the eigenvalue and eigenvector estimation listed in (7) and (8) seems to simply depend on $\epsilon$, it is actually an interplay of $\epsilon$ and $n$. In the above theorem, we have two relationships between $\epsilon$ and $n$, one for eigenvalue and one for eigenvector. If we choose $\epsilon = \epsilon(n) = \left( \frac{\log n}{n} \right)^{1/4}$, asymptotically, the eigenvector converges at the rate $n^{-1/(8d+16)}$ up to a log factor, which is slow compared with the eigenvalue convergence rate at $n^{-3/(8d+26)}$ up to a log factor. Usually, when we apply GL to approximate the Laplace-Beltrami operator in the spectral sense, we use one
\[ n > \max \left( \frac{\lambda_2^d/2+5}{\min(\Gamma, 1)} \right)^{8d+26}, \right), (\lambda_3 + \lambda_n^{(5d+7)/4})^{8d+26} \right), \]
then for \( 0 \leq i \leq K, \)
\[ |\mu_{i,n,\varepsilon} - \lambda_i| \leq \Omega_1 \left( \frac{\log n}{n} \right)^{\frac{1}{8d+26}}. \]

Similarly, if
\[ n > \max \left( \frac{\lambda_2^d/2+5}{\min(\Gamma, 1)} \right)^{8d+16}, \right), (\lambda_3 + \lambda_n^{(5d+7)/4})^{8d+16} \right), \]
then for \( 0 \leq i \leq K, \)
\[ \max_{x_j \in \mathcal{E}} |d_i v_{i,n,\varepsilon}(j) - \phi_\varepsilon(x_j)| \leq \Omega_2 \left( \frac{\log n}{n} \right)^{\frac{1}{8d+16}}. \]

**Remark 5.** Another remark is about how to determine \( \varepsilon \) and \( K \) in practice. We shall emphasize that the asymptotic relationship between \( \varepsilon \) and \( n \), like \( \varepsilon = (\frac{\log n}{n})^{\frac{1}{8d+17}} \) or \( \varepsilon = (\frac{\log n}{n})^{\frac{1}{8d+17}} \), could be a guidance, but it is not practical to choose \( \varepsilon \) directly from this relationship. It is because the constant \( \Omega_2 \) is usually unknown for a generic manifold. To our knowledge, despite some practical rules of thumb, in general a good choice of \( \varepsilon \) depends on some try-and-errors. See [13] for a recent theoretically solid choice that aims to choose an \( \varepsilon \) that is robust to “big” high dimensional noise. However, when there is no noise or when noise is not big, how to choose an optimal \( \varepsilon \) for the spectral convergence purpose is still open. We face the similar challenge in determining the relation between \( n \) and \( K \). Indeed, while we could obtain an asymptotic relationship between \( K \) and \( n \) via applying Lemma 2 to (9) or (10), the relationship is not practical since the implied constant is in general difficult to control. To our knowledge, there does not exist a universal practical way to determine what \( K \) should be, and in practice it depends on try-and-errors. We will report our work in these practical challenges in our future work.

Now we introduce our heat kernel estimator. Fix \( K \in \mathbb{N} \) and \( t > 0 \). Consider the matrix
\[ H_{i,j}^{(K)}(t) = \sum_{i=0}^{K-1} e^{-\mu_{i,n,\varepsilon} t} v_{i,n,\varepsilon} v_{i,n,\varepsilon}^T \in \mathbb{R}^{n \times n}, \]
which will be our estimation of the heat kernel. Note that we sum \( i \) from 0 to \( K-1 \), which is based on the statement for \( i < K \) in the previous theorem.

To study how \( H_{i,j}^{(K)}(t) \) approximates the heat kernel \( H(\cdot, \cdot, t) \), below we list assumptions about the relationship among \( n, \varepsilon, K \) and \( t \).
Assumption 2. The parameters $n$, $\varepsilon$, $K$ and $t$ satisfy the following conditions.

1. Set $\Gamma_K := \min_{1 \leq i \leq K} \text{dist}(\lambda_i, \sigma(-\Delta) \setminus \{\lambda_i\})$. Suppose
   \[\varepsilon \leq \min \left(\mathcal{H}_1, \left(\frac{\min(\Gamma_K, 1)}{\mathcal{H}_2 + \lambda_K d/2 + 5}\right)^2, \frac{1}{\mathcal{H}_3 + \lambda_K^{(5d+7)/4}}, \frac{\mathcal{H}}{K^4 \lambda_K^{(d-1)/2}}\right),\]
   where $\mathcal{H}_1$ and $\mathcal{H}_2, \mathcal{H}_3 > 1$ are the constants in Theorem 2 and $\mathcal{H}$ depends on $d$ and the Ricci curvature.

2. $n$ is large enough so that $\left(\frac{\log n}{n}\right)^{1/3} \leq \varepsilon$.

3. $\frac{C \log K}{K^{2/d}} \leq t$, where $C$ is a constant depending on $d$ and the Ricci curvature and the diameter of $M$.

Based on the above assumption, we have the following theorem. The proof of the theorem is in Section 6.

Theorem 3. Suppose Assumptions 7 and 2 hold. Then, with probability greater than $1 - \frac{1}{n^{-2}}$
\[\sup_{x_i, x_j \in \mathcal{X}} |H_{(i,j)}^{(K)}(x_i, x_j) - H(x_i, x_j, t)| < \frac{\Omega_1}{K}(t+1)e^{-\Omega_2 t} + \Omega_3 \varepsilon,\]
where $\Omega_1$ depends on $d$, $p_m$, the $C^2$ norm of $p$, and the injectivity radius, the diameter, the volume and the curvature of $M$, $\Omega_3$ depends on $d$, and the Ricci curvature and the diameter of $M$, and $\Omega_5$ depends on the volume of $M$ and the $C^2$ norm of $p$.

The above assumption and theorem deserve a discussion. In the above theorem, the error between the heat kernel estimator and the actual heat kernel is characterized by $K$, $\varepsilon$ and $t$. The error is composed of two parts, the bound of the remainder term $\sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i(x) \phi_i(x')$ in the series expression of the heat kernel and the error between $H_{(i,j)}^{(K)}(x_i, x_j)$ and $\sum_{i=0}^{K-1} e^{-\lambda_i t} \phi_i(x_i) \phi_i(x_j)$. In fact, in the proof of the above theorem, we show that if $t$ is not too small, i.e. $\frac{C \log K}{K^{2/d}} \leq t$, then $e^{-\lambda_i t}$ is small and $|\sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i(x) \phi_i(x')|$ is bounded by $\frac{1}{K}(t+1)e^{-\Omega_1 t}$ up to a constant for all $x, x' \in M$. On the other hand, based on Theorem 2 (1) and (2) in Assumption 2 enable us to have a good control between the first $K$ eigenpairs of the kernel normalized GL and the corresponding eigenpairs of $-\Delta$. Together with the condition $\varepsilon \leq \frac{\mathcal{H}_1}{K^4 \lambda_K^{(d-1)/2}}$ in (1) in Assumption 2 we can conclude that the error between $H_{(i,j)}^{(K)}(x_i, x_j)$ and $\sum_{i=0}^{K-1} e^{-\lambda_i t} \phi_i(x_i) \phi_i(x_j)$ is also bounded by $\frac{1}{K}(t+1)e^{-\Omega_1 t}$ up to a constant plus $\Omega_3 \varepsilon$. Since we do not know the volume of $M$ and $\mu_{0, n, \varepsilon} = \lambda_0 = 0$, the term $\Omega_3 \varepsilon$ comes from the error generated when we renormalize $\tilde{v}_{0, n, \varepsilon}$ as in 5 to approximate $\phi_0$. Next, we discuss the dependence on the diffusion time $t$ in the error of the approximation. Note that decreases as $K$ increases; hence, the relationship (3) says that we can choose a small $t$, whenever $K$ is large enough. $(t+1)e^{-\Omega_1 t}$ is bounded up by a constant depending on $\Omega_4$. Hence, for a small diffusion time $t$, we need to choose a large $K$ in order to have a good approximation. Since $\varepsilon < \frac{1}{K}$ when $K$ is a large enough, the error in the approximation is dominated by $\frac{1}{K}$ for a small diffusion time $t$. For a large diffusion time $t$, since $(t+1)e^{-\Omega_1 t}$ goes to 0 as $t$ goes to infinity, the first term in the error bound is smaller when $t$ increases.

3.3. The spectral convergence under a special setup. In the previous subsection, we show the spectral convergence of the kernel normalized GL to the Laplace-Beltrami operator regardless of the distribution of the data points on the manifold in the $L^\infty$ sense so that the spectral geometry tool 3 can be applied for further data analysis. In this subsection, for the sake of comparing our results with other relevant result, we discuss the spectral
convergence under a special setup. Specifically, we focus on the case when the p.d.f. is uniform so that the normalization is not needed for the recovery of the Laplace-Beltrami operator, and the dependence on $\lambda_K$ is hidden in the implied constant but not shown in the relationship between $n$ and $\epsilon$.

Denote $\mu_{i,n,\epsilon}^m$ to be the $i$-th eigenvalue of $-L_{\epsilon}^m$ with the associated eigenvector $\hat{v}_{i,n,\epsilon}^m$ normalized in the $l^2$ norm, where $i = 1, \ldots, n$. We order $\mu_{i,n,\epsilon}^m$ so that $\mu_{0,n,\epsilon}^m \leq \mu_{1,n,\epsilon}^m \leq \cdots \leq \mu_{n-1,n,\epsilon}^m$. Let $\| \cdot \|_{\hat{l}^2(1/\hat{\beta})}$ be the normalized $l^2$ norm in Definition 1. Let

$$\hat{\lambda}_{i,n,\epsilon}^m = \frac{|\hat{v}_{i,n,\epsilon}^m|_{\hat{l}^2(1/\hat{\beta})}}{\| \hat{v}_{i,n,\epsilon}^m \|_{\hat{l}^2(1/\hat{\beta})}},$$

(12)

In the following theorem, we treat the eigengaps of $-\Delta$ and the eigenvalue $\lambda_K$ as constants and disregard the relation between $\epsilon$ and $K$. The proof of the theorem is almost the same the proof of Theorem 2 except simplification of several steps, so its proof is sketched and postponed to Appendix C.

**Theorem 4. (Spectral convergence of the unnormalized GL)** Suppose the sampling is uniform and all the eigenvalues of $\Delta$ are simple. Fix $K \in \mathbb{N}$. Let $\Gamma_K := \min_{1 \leq i \leq K} \text{dist}(\lambda_i, \sigma(-\Delta) \setminus \{ \lambda_1 \})$. For a sufficiently small $\epsilon > 0$, if $n$ is sufficiently large so that $\epsilon = \epsilon(n) \geq \left( \frac{\log n}{n} \right)^{1/12}$, then with probability greater than $1 - n^{-2}$, for all $0 \leq i < K$

$$|\mu_{i,n,\epsilon}^m - \lambda_i| \leq \Omega_{\epsilon}^{\text{un}} \epsilon^2,$$

where $\Omega_{\epsilon}^{\text{un}} > 0$ depends on $\Gamma_K$, $\lambda_K$, $d$ and the diameter, the volume, the injectivity radius, the curvature and the second fundamental form of the manifold. For a sufficiently small $\epsilon > 0$, if $n$ is sufficiently large so that $\epsilon = \epsilon(n) \geq \left( \frac{\log n}{n} \right)^{1/12}$, then with probability greater than $1 - n^{-2}$, there are $a_i \in \{ 1, -1 \}$ such that for all $0 \leq i < K$

$$\max_{x_j \in \mathcal{D}} |a_i \hat{v}_{i,n,\epsilon}^m(j) - \phi_i(x_j)| \leq \Omega_{\epsilon}^{\text{un}} \epsilon^2,$$

where $\Omega_{\epsilon}^{\text{un}} > 0$ depends on $\Gamma_K$, $\lambda_K$, $d$ and the diameter, the volume, the injectivity radius, the curvature and the second fundamental form of the manifold.

Note that remarks similar to Remarks 4 and 5 hold for Theorem 4. Also, we see that the relationship between $n$ and $\epsilon$ in Theorem 4 is different from that in Theorem 2. The difference mainly comes from the kernel normalization for the purpose of density estimation. By this theorem, when the point cloud is sampled uniformly, if we take $\epsilon = \left( \frac{\log n}{n} \right)^{1/12}$, we have

$$|\mu_{i,n,\epsilon}^m - \lambda_i| = O \left( \left( \frac{\log n}{n} \right)^{1/12} \right);$$

if we take $\epsilon = \left( \frac{\log n}{n} \right)^{1/8}$, we have for $a_i \in \{ 1, -1 \}$,

$$\max_{x_j \in \mathcal{D}} |a_i \hat{v}_{i,n,\epsilon}^m(j) - \phi_i(x_j)| = O \left( \left( \frac{\log n}{n} \right)^{1/8} \right).$$

We shall compare it with [5, 6]. In these papers, the authors also consider a $d$ dimensional compact manifold $M$ without boundary that is isometrically embedded in $\mathbb{R}^D$. In these papers, to construct the unnormalized GL, they use a compactly supported kernel of the form $\eta \left( \frac{\| x - y \|_{\mathcal{D}}}{\rho} \right)$, where $x, y \in M$ and $\eta : [0, \infty) \rightarrow [0, \infty)$ is a non-increasing Lipschitz continuous function with the compact support on $[0, 1]$. For $n$ points i.i.d. sampled from $M$ based
on a density function $\rho$, under this setup, in \cite{5}, the authors prove the $L^2$ spectral convergence of the unnormalized GL to the weighted Laplacian $\frac{1}{\rho} \nabla \cdot (\rho^2 \nabla)$ using the variational method for Dirichlet form and the techniques in optimal transport. Moreover, the authors show that the eigenvalue and eigenvector convergence is at a rate $O\left(\frac{\log n}{n^{d/4}}\right)$. Based on the result in \cite{5}, the authors in \cite{6} improve the $L^2$ convergence of the eigenvectors to the $L^\infty$ convergence. They show that the $L^\infty$ convergence rate of the eigenvectors matches the $L^2$ rate; that is, $O\left(\frac{\log n}{n^{d/4}}\right)$. Note that in the case when the sampling density is uniform, that is, $\rho$ is constant, then $\frac{1}{\rho} \nabla \cdot (\rho^2 \nabla)$ gives the Laplace-Beltrami operator. Hence, in this setup, our eigenvector convergence rate of unnormalized GL in Theorem 4 matches the $L^\infty$ result of \cite{6}. However, our eigenvalue convergence rate is lower. It is also worth mentioning that our results in Theorem 2 and Theorem 4 rely on the fact that the kernel function is Gaussian, while the kernel considered in \cite{6} is compactly supported and Lipschitz, which does not include the widely used Gaussian kernel. See Remark 7 for more details. We emphasize that the techniques used in the current paper and those used in \cite{6} are different, and it is potentially interesting to combine these techniques to obtain a more general result. We would expect that the optimal spectral convergence rate is better than what we have reported, in both kernel normalized and unnormalized cases.

4. SOME GENERIC FACTS

In this section, we summarize some fundamental facts we need in the whole proof. The first two lemmas are the facts about the eigenvalues and eigenfunctions of the Laplace-Beltrami operator.

**Lemma 1** (\cite{18,14}). For a compact Riemannian manifold $(M,g)$ and $l > 0$, we have the following bound for the $l$-th pair of eigenvalue $\lambda_l$ and normalized eigenfunction $\phi_l$ of the Laplace-Beltrami operator:

$$\|\phi_l\|_\infty \leq C_1 \lambda_l^{-\frac{1}{4}} \|\phi_l\|_2 = C_1 \lambda_l^{-\frac{d-1}{4}},$$

where $C_1$ is a constant depending on the injectivity radius and sectional curvature of the manifold $M$.

**Lemma 2** (\cite{16}). For a $d$ dimensional compact and connected Riemannian manifold $(M,g)$, the eigenvalues of the Laplace-Beltrami operator, $0 = \lambda_0 < \lambda_1 \leq \ldots$, satisfy

$$C_2^{1+d/\sqrt{d}} \frac{\text{diam}(M)^{-2} t^{2/d}}{d} \leq \lambda_l \leq \frac{(d-1)^2}{4} \kappa + \tilde{C}_2 V(M)^{-3/2} t^{2/d}$$

for all $l \geq 1$, where the $\text{Ric}_g \geq -(d-1) \kappa g$ for $\kappa \geq 0$, and $C_2 > 0$ and $\tilde{C}_2 > 0$ are constants depending on $d$ only.

The following lemma describes the behavior of the heat kernel. The proof of can be found in \cite{27}, \cite{12} and \cite{15}.

**Lemma 3.** Suppose $M$ is a $d$ dimensional closed smooth manifold. For $x, y \in M$, $d(x,y)$ is the geodesic distance between $x$ and $y$.

(a) Fix $\gamma > 0$ small enough. For $d(x,y) < \gamma$, we have when $t \to 0$:

$$H(x,y,t) = (4\pi t)^{-d/2} e^{-\frac{d(x,y)^2}{4t}} (u_0 + t u_1 + O(t^2)).$$
Let $A$ and $B$ be compact self-adjoint operators from $L^2(M)$ to $L^2(M)$. Let $(\cdot, \cdot)$ be the inner product of $L^2(M)$. Assume the eigenvalues of $A$, $\lambda_i(A)$, $l = 0, 1, \ldots$, are simple and positive and the eigenvalues of $B$, $\lambda_i(B)$, $l = 0, 1, \ldots$, are simple and bounded from below so that $1 = \lambda_0(A) > \lambda_1(A) > \lambda_2(A) > \cdots \geq 0$ and $\lambda_0(B) > \lambda_1(B) > \lambda_2(B) > \cdots$. Also denote $\{u_i\}$ to be the orthonormal eigenfunctions of $A$ and $\{w_i\}$ to be orthonormal eigenfunctions of $B$. For $i = 1, 2, \ldots$, denote

$$\gamma_i(B) := \min(\lambda_i(B) - \lambda_{i-1}(B), \lambda_{i+1}(B) - \lambda_i(B)).$$

Let $E := A - B$. Then, for $\epsilon > 0$ we have the following statements:

(a) If $\left| \frac{\langle f, f \rangle}{(A)f, f} \right| \leq \epsilon$ for all $f \in L^2$, then for all $i > 0$, we have

$$\left| \frac{1 - \lambda_i(B)}{1 - \lambda_i(A)} - 1 \right| \leq \epsilon.$$

(b) If $\|Bu_i - \lambda_i(B)u_i\| \leq \epsilon$, where $i \geq 1$, then we can find $a = 1$ or $-1$ so that

$$\|aw_i - u_i\| \leq \frac{2\epsilon}{\gamma_i(B)}$$

Moreover,

$$\langle u_i, w_i \rangle \geq 1 - \frac{\epsilon}{\gamma_i(B)}.$$
(c) For \( i, j \geq 0 \), when \((u_i, w_j) \neq 0\), the associated eigenvalues satisfy
\[
|\lambda_i(A) - \lambda_j(B)| \leq \frac{\|Ew_j\|_2}{\|(u_i, w_j)\|}.
\]

**Remark 6.** Note that (b) can be understood as a variation of the Davis-Kahan theorem. Later we will apply the lemma to operators of the form \( I - C \) when \( C \) is compact and self-adjoint. Clearly, \( I - C \) is not compact but shares the same eigenfunctions with \( C \), and the eigenvalues of \( I - C \) are simply those of \( 1 - \lambda \) when \( \lambda \) is an eigenvalue of \( C \). Specifically, later we will set \( A \) to be the integral operator with the heat kernel, and \( B \) to be the integral operator with a diffusion kernel. We mention that (c) is also used in [5] to prove the spectral convergence rate in the \( L^2 \) sense.

**Proof.** The first statement can be found in [2, Proposition 4.4], so we omit it.

For the second statement, we express \( u_i = b_i w_i + \sum_{j=0, j \neq i}^\infty b_j w_j \). By a direct expansion and the assumption, we have
\[
\|Bu_i - \lambda_i(B)u_i\|_2 = \|B\left(b_i w_i + \sum_{j=0, j \neq i}^\infty b_j w_j\right) - \lambda_i(B)\left(b_i w_i + \sum_{j=0, j \neq i}^\infty b_j w_j\right)\|_2
\]
\[
= \|B\left(\sum_{j=0, j \neq i}^\infty b_j w_j\right) - \lambda_i(B)\left(\sum_{j=0, j \neq i}^\infty b_j w_j\right)\|_2
\]
\[
= \|\sum_{j=0, j \neq i}^\infty (\lambda_j(B) - \lambda_i(B))b_j w_j\|_2
\]
\[
= \sqrt{\sum_{j=0, j \neq i}^\infty |\lambda_j(B) - \lambda_i(B)|^2 |b_j|^2} \leq \varepsilon.
\]

Therefore, we have
\[
\|u_i - (u_i, w_i)w_i\|_2 = \left\| \sum_{j=0, j \neq i}^\infty b_j w_j \right\|_2 = \sqrt{\sum_{j=0, j \neq i}^\infty |b_j|^2}
\]
\[
\leq \frac{1}{\gamma(B)} \sqrt{\sum_{j=0, j \neq i}^\infty |\lambda_j(B) - \lambda_i(B)|^2 |b_j|^2} \leq \frac{\varepsilon}{\gamma(B)}.
\]

By Lemma [4], we have
\[
\|aw_i - u_i\|_2 \leq 2\|u_i - (u_i, w_i)w_i\|_2 \leq \frac{2\varepsilon}{\gamma(B)},
\]
which leads to the conclusion of \(|(u_i, w_i)| \geq 1 - \frac{\varepsilon}{\gamma(B)}\) from a direct expansion via the polarization identity.

For the third statement, the self-adjoint assumption leads to
\[
\lambda_i(A)(u_i, w_j) = (Au_i, w_j) = (u_i, Aw_j) = (u_i, Bw_j) + (u_i, (A - B)w_j)
\]
\[
= \lambda_j(B)(u_i, w_j) + (u_i, Ew_j).
\]

Hence, when \((u_i, w_j) \neq 0\), by Cauchy-Schwartz, we have
\[
|\lambda_i(A) - \lambda_j(B)| = \frac{|(u_i, Ew_j)|}{|(u_i, w_j)|} \leq \frac{\|Ew_j\|_2}{|(u_i, w_j)|}.
\]

Recall the definition of collective compact convergence.
Definition 2. Let \((E, \| \cdot \|_E)\) be an arbitrary Banach space, \(B \subset E\) be the unit ball centered at \(0\), and \(\{T_n\}_{n=1}^\infty\) be a set of bounded linear operators on \(E\). The set \(\{T_n\}_{n=1}^\infty\) is called collectively compact if the set \(\bigcup_n T_n(B)\) is relatively compact in \((E, \| \cdot \|_E)\). The sequence \(\{T_n\}_{n=1}^\infty\) is said to converge collectively compactly to an operator \(T\), if it converges pointwise and there exists some \(N \in \mathbb{N}\) such that the sequence of operators \(\{T_n - T\}_{n=N}^\infty\) is collectively compact.

Also recall the definition of a resolvent.

Definition 3. Let \(T\) be a compact linear operator on an arbitrary Banach space \((E, \| \cdot \|_E)\) and denote \(\sigma(T) \subset \mathbb{C}\) to be the associated spectrum. Then for \(z \in \mathbb{C} \setminus \sigma(T)\), the resolvent operator is defined as \(R_z(T) := (z - T)^{-1}\). For an eigenvalue \(\lambda \in \sigma(T)\), let \(\Gamma_r(\lambda) \subset \mathbb{C}\) be a circle centered at \(\lambda\) with the radius \(r > 0\).

The following generic theorem is the key toward the spectral convergence. Specifically, it quantifies the convergence of the corresponding eigenfunctions for a sequence of operators \(\{T_n\}_{n=1}^\infty\) that collectively compactly converges to an operator \(T\).

Theorem 5. Let \((E, \| \cdot \|_E)\) be an arbitrary Banach space. Let \(\{T_n\}_{n=1}^\infty\) and \(T\) be compact linear operators on \(E\) such that \(\{T_n\}_{n=1}^\infty\) converges to \(T\) collectively compactly. For an eigenvalue \(\lambda \in \sigma(T)\), denote the corresponding spectral projection by \(\text{Pr}_\lambda\). Let \(D \subset \mathbb{C}\) be an open neighborhood of \(\lambda\) such that \(\sigma(T) \cap D = \{\lambda\}\). There exists some \(N \in \mathbb{N}\) such that for all \(n > N\), \(\sigma(T_n) \cap D = \{\lambda_n\}\). Let \(\text{Pr}_{\lambda_n}\) be the corresponding spectral projection associated with \(\lambda_n\) of \(T_n\). Let \(r < |\lambda|\) and \(r < \text{dist}(\{\lambda\}, \sigma(T) \setminus \{\lambda\})\). Then for every \(x \in \text{Pr}_\lambda(E)\), we have

\[
\|x - \text{Pr}_{\lambda_n} x\|_E \leq \max_{z \in \Gamma_r(\lambda)} \frac{2r\|R_z(T)||}{\min_{z \in \Gamma_r(\lambda)} |z|} \left(\|(T_n - T)x\|_E + \|R_z(T)_x\|_E\|T_n - T\||_E\right).
\]

This theorem is a restatement of Equation (5) in Theorem 3 in [1]. Specifically, we let the set \(F = \Gamma_r(\lambda)\), \(M = \max_{z \in \Gamma_r(\lambda)} \|R_z(T)\|\), and \(c_0 = \min_{z \in \Gamma_r(\lambda)} |z|\) in [1] Lemma on page 460], set \(\varepsilon = r\) in [1] Theorem 3, and note that \(\|R_z(T)_x\|_E \leq M\|x\|_E\).

5. Proof of Theorem

5.1. Some quantities needed in the proof and their properties. Following Assumption [1] for the point cloud \(\mathcal{X} := \{x_i\}_{i=1}^n\) i.i.d. sampled from the random vector with the density function \(p\) supported on the manifold \(M\), we denote the empirical measure associated with the measure \(d\mathcal{P} := pdV_M\), where \(dV_M\) is the Riemannian volume measure, as

\[
P_n := \frac{1}{n} \sum_{i=1}^n \delta_{x_i},
\]

where \(\delta_{x_i}\) is the delta measure supported on \(x_i\). We start by introducing the following quantities.

Definition 4. For any continuous function \(f(x,y)\) on \(M \times M\), we define two operators, \(P\) and \(P_n\), with respect to the density measure and empirical measure as

\[Pf(x) := \int_M f(x,y)d\mathcal{P}(y),\]

\[P_nf(x) := \int_M f(x,y)d\mathcal{P}_n(y) = \frac{1}{n} \sum_{i=1}^n f(x,x_i).\]
For the kernel
\[ k_\varepsilon(x,y) = \exp \left( - \frac{\|t(x) - t(y)\|_D^2}{4\varepsilon^2} \right), \]
we have the following quantities:
\[ d_{n,\varepsilon}(x) := P_n k_\varepsilon(x), \quad d_\varepsilon(x) := P k_\varepsilon(x), \]
\[ Q_{n,\varepsilon}(x,y) := \frac{k_\varepsilon(x,y)}{d_{n,\varepsilon}(x)d_{n,\varepsilon}(y)}, \quad Q_\varepsilon(x,y) := \frac{k_\varepsilon(x,y)}{d_\varepsilon(x)d_\varepsilon(y)}. \]

Note that we have \( n d_{n,\varepsilon}(x) = q_\varepsilon(x) \) defined in [1]. We also introduce the following three operators.

**Definition 5.** For any \( f(x) \in C(M) \), we define two operators from \( C(M) \) to \( C(M) \):
\[ T_{n,\varepsilon} f(x) = \frac{P_n Q_{n,\varepsilon} f(x)}{P_n Q_{n,\varepsilon}(x)}, \quad T_\varepsilon f(x) = \frac{P Q_\varepsilon f(x)}{P Q_\varepsilon(x)}. \]
Moreover, we define an intermediate operator connecting \( T_{n,\varepsilon} \) and \( T_\varepsilon \) that is also from \( C(M) \) to \( C(M) \):
\[ \hat{T}_{n,\varepsilon} f(x) = \frac{P_n Q_\varepsilon f(x)}{P Q_\varepsilon(x)}. \]

We first state some known facts about the operator \( T_\varepsilon \). The proof can be found in [10, Lemma 8]. In [10], the Laplace-Beltrami operator differs from ours by a negative sign, and the bandwidth is \( \sqrt{\varepsilon} \) rather than \( \varepsilon \); otherwise the proof is the same, so we omit the proof.

**Lemma 6.** We have the following results.

(a) \( d_\varepsilon(x) = m_0 p(x) e^d + O(\varepsilon^{d+2}) \), where \( m_0 := \int_{\mathbb{R}^d} e^{-\|x\|^2} \, du \) and the implied constant in \( O(\varepsilon^{d+2}) \) depends on \( p_m \) and the \( C^2 \) norm of \( p \). Hence,
\[ Q_\varepsilon(x,y) = \frac{k_\varepsilon(x,y)}{m_0^2 p(x)p(y) e^{2d} (1 + O(\varepsilon^2))}, \]
where the implied constant in \( O(\varepsilon^2) \) depends on \( p_m \) and the \( C^2 \) norm of \( p \).

(b) \( T_\varepsilon \) is a self-adjoint operator on \( L^2(M) \).

(c) For any \( f \in L^2(M) \), \( T_\varepsilon f(x) \) is a smooth function on \( M \). In particular, the eigenfunctions of \( T_\varepsilon \) are smooth. Moreover,
\[ T_\varepsilon f(x) = \frac{\int_M k_\varepsilon(x,y)f(y) \, dV_M}{\int_M k_\varepsilon(x,y) \, dV_M} + O(\varepsilon^2), \]
where the implied constant in \( O(\varepsilon^2) \) depends on \( p_m \) and the \( C^2 \) norm of \( p \) and \( \|f\|_2 \).

(d) For any function \( f \in C^4(M) \),
\[ \frac{T_\varepsilon f(x) - f(x)}{\varepsilon^2} = \Delta f(x) + O(\varepsilon^2), \]
where the implied constant in \( O(\varepsilon^2) \) depends on the \( L^\infty \) norm of \( \Delta^2 f \), \( p_m \), the \( C^2 \) norm of \( p \) and the Ricci curvature and the second fundamental form of the manifold.

Define \( H_\varepsilon \) to be the integral operator associated with the heat kernel on the manifold \( M \) with the diffusion time \( \varepsilon^2 \); that is, for a suitable function \( f \), we have
\[ H_\varepsilon f(x) := \int_M H(x,y,\varepsilon^2)f(y) \, dV_M(y). \]
Let $$R_\varepsilon := \frac{I - H_\varepsilon}{\varepsilon^2} - \frac{I - T_\varepsilon}{\varepsilon^2}.$$ We discuss the properties of $$R_\varepsilon$$ in the next two lemmas.

**Lemma 7.** Let $$f \in L^2(M)$$. There exists a constant $$C_3$$ depending on $$p_m$$, the $$C^2$$ norm of $$p$$ and the curvature of the manifold $$M$$, such that when $$\varepsilon > 0$$ is sufficiently small, $$\|R_\varepsilon f\|_2 \leq C_3 \|f\|_2.$$

**Proof.** By Lemma 6 (c),
\[
T_\varepsilon f(x) = \frac{\int_M k_\varepsilon(x, y) f(y) dV_M}{\int_M k_\varepsilon(x, y) dV_M} + O(\varepsilon^2)
\]
\[
= \frac{\int_M (4\pi \varepsilon^2)^{-d/2} k_\varepsilon(x, y) f(y) dV_M}{\int_M (4\pi \varepsilon^2)^{-d/2} k_\varepsilon(x, y) dV_M} + O(\varepsilon^2).
\]
Note that $$\int_M (4\pi \varepsilon^2)^{-d/2} k_\varepsilon(x, y) dV_M = 1 + O(\varepsilon^2)$$ by Lemma 6 (a). Hence,
\[
T_\varepsilon f(x) = \int_M (4\pi \varepsilon^2)^{-d/2} k_\varepsilon(x, y) f(y) dV_M + O(\varepsilon^2),
\]
where the implied constant in $$O(\varepsilon^2)$$ depends on $$p_m$$, the $$C^2$$ norm of $$p$$ and $$\|f\|_2$$. Thus,
\[
\|R_\varepsilon f\|_2 \leq \left\| \frac{(I - H_\varepsilon)f(x)}{\varepsilon^2} - \frac{f(x) - \int_M (4\pi \varepsilon^2)^{-d/2} k_\varepsilon(x, y) f(y) dV_M}{\varepsilon^2} \right\|_2 + \tilde{C}_1 \|f\|_2
\]
\[
= \left\| \frac{H_\varepsilon f(x) - \int_M (4\pi \varepsilon^2)^{-d/2} k_\varepsilon(x, y) f(y) dV_M}{\varepsilon^2} \right\|_2 + \tilde{C}_1 \|f\|_2,
\]
where $$\tilde{C}_1$$ depends on $$p_m$$ and the $$C^2$$ norm of $$p$$. To control $$\|H_\varepsilon f(x) - \int_M (4\pi \varepsilon^2)^{-d/2} k_\varepsilon(x, y) f(y) dV_M\|_2$$, note that
\[
\|H_\varepsilon f(x) - \int_M (4\pi \varepsilon^2)^{-d/2} k_\varepsilon(x, y) f(y) dV_M\|_2 \leq \left\| H_\varepsilon f(x) - \int_M (4\pi \varepsilon^2)^{-d/2} e^{-\frac{d(x, y)^2}{4\varepsilon^2}} f(y) dV_M \right\|_2 + \left\| \int_M (4\pi \varepsilon^2)^{-d/2} e^{-\frac{d(x, y)^2}{4\varepsilon^2}} f(y) dV_M - \int_M (4\pi \varepsilon^2)^{-d/2} k_\varepsilon(x, y) f(y) dV_M \right\|.
\]
The former term on the right hand side can be bounded by Lemma 5 (a), while the later term can be bounded by using the expansion of $$d^2(x, y)$$ in terms of $$\|t(x) - t(y)\|^2_{\text{H}^0}$$. In conclusion, we have
\[
\|R_\varepsilon f\|_2 \leq \tilde{C}_2 \|f\|_2,
\]
where $$\tilde{C}_2$$ depends on the curvature of the manifold. We finish the proof by setting $$C_3 = \tilde{C}_1 + \tilde{C}_2$$.

**Remark 7.** Note that the proof of the above lemma relies on the fact that $$k_\varepsilon(x, y)$$ is Gaussian. In particular, we use the expansion of $$H(x, y, \varepsilon^2)$$ and $$(4\pi \varepsilon^2)^{-d/2} k_\varepsilon(x, y)$$ in terms of $$(4\pi \varepsilon^2)^{-d/2} e^{-\frac{d(x, y)^2}{4\varepsilon^2}}$$ in [19], which is related to the parametrix method [22].

**Lemma 8.** Fix $$K \in \mathbb{N}$$. Suppose $$\phi_k \in C^\infty(M)$$ is the $$K$$-th eigenfunction of $$\Delta$$ and $$\|\phi_k\|_2 = 1$$. Then, there is a constant $$C_4$$ depending on $$p_m$$, the $$C^2$$ norm of $$p$$, the Ricci curvature and the second fundamental form of the manifold, so that when $$\varepsilon > 0$$ is sufficiently small,
\[
\|R_\varepsilon \phi_k\|_2 \leq C_4 \varepsilon^2 \left( 1 + \lambda_k^{d/2+2} \right).
\]
We mention that this is the main step controlling the $L^2$ convergence rate due to the fundamental Sobolev embedding theorem. Also note how the bound depends on the associated eigenvalue. By the lower bound of the eigenvalues shown in Lemma 2, the bound exponentially increases as $K$ increases.

**Proof.** By Lemma 6 (d), $R_x \phi (x) \leq \tilde{C}_1 \| \phi \|_{C^1(M)}$, where $\tilde{C}_1$ is a constant depending on $p_m$, the $C^2$ norm of $p$ and the Ricci curvature and the second fundamental form of the manifold. Moreover, by the Sobolev Embedding theorem [25, Theorem 9.2], there is a constant $\tilde{C}_2 > 0$ depending on the Ricci curvature of the manifold so that

$$\| \phi \|_{C^4(M)} \leq \tilde{C}_2 \| \phi \|_{H^2} \leq \tilde{C}_2 (1 + \| \Delta^{d/2+5} \phi \|_2) \leq \tilde{C}_2 (1 + \lambda_K^{d/2+5}).$$

By taking $C_4 = \tilde{C}_1 \tilde{C}_2$, the conclusion follows. 

5.2. **Spectral convergence of** $\frac{L - H\gamma}{e^2}$ **to** $-\Delta$. In this subsection we show the spectral convergence of $\frac{L - H\gamma}{e^2}$ to $-\Delta$. We prove the main results in this subsection by extending the argument in [10, 2] with several new arguments and results. The proof is long and contains results of independent interest, so we divide it into several parts.

Recall that the Laplace-Beltrami operator $\Delta$ is not a compact operator. To show the convergence of the eigenvalues and the corresponding eigenvectors, we compare $\frac{L - H\gamma}{e^2}$ and $\frac{L - H\gamma}{e^2}$. For the $i$-th eigenvalue of $-\Delta$, $\lambda_i$, we know that the $i$-th eigenvalue of $\frac{L - H\gamma}{e^2}$ is $\frac{1 - e^{-\lambda_i e^2}}{e^2}$, which converges to $\lambda_i$ when $e$ goes to 0. We will show the convergence of the $i$-th eigenvalue of $\frac{L - H\gamma}{e^2}$ to the $i$-th eigenvalue of $\frac{L - H\gamma}{e^2}$ when $i$ is fixed. Moreover, since $\frac{L - H\gamma}{e^2}$ and $\Delta$ share the same eigenfunctions, the convergence of eigenfunctions will follow from the convergence of eigenvalues.

**Proposition 1.** Assume that the eigenvalues of $\Delta$ are simple. Suppose $(\lambda_{i,e}, \phi_{i,e})$ is the $i$-th eigenpair of $\frac{L - H\gamma}{e^2}$ and $(\lambda_i, \phi_i)$ is the $i$-th eigenpair of $-\Delta$. Assume both $\phi_{i,e}$ and $\phi_i$ are normalized in the $L^2$ norm. For $K \in \mathbb{N}$, denote

$$\Gamma_K := \min_{1 \leq i \leq K} \text{dist}(\lambda_i, \sigma(-\Delta) \setminus \{\lambda_i\}).$$

Suppose (6) is satisfied and $e > 0$ is sufficiently small. Then, there are $a_i \in \{-1, 1\}$ such that for all $1 \leq i < K$,

$$|\lambda_{i,e} - \lambda_i| \leq e^{3/2}, \quad \|a_i \phi_{i,e} - \phi_i\|_2 \leq e^{3/2}, \quad \|a_i \phi_{i,e} - \phi_i\|_\infty \leq e.$$

The results shown in (22) have been highly simplified. A more delicate error bound depending on $\lambda_K$ and other terms can be found in (25) and (51). Note that $\Gamma_K = \min\{\gamma_i(\Delta) | i = 1, \ldots, K\}$, where $\gamma_i(\Delta)$ is defined in (14). The requirement that $e \leq \Gamma_K$ is somehow intuitive – to recover the first few eigenvalues with small gaps, we need a small bandwidth.

We now describe the idea of proving Proposition 1. We first use Lemma 5 (a) to obtain a convergence of the eigenvalues at a possibly slow convergence rate. With the convergence of eigenvalues, we can apply Lemma 5 (b) to show that the angle between the $i$-th eigenfunction of $\frac{L - H\gamma}{e^2}$ and the $i$-th eigenfunction of $-\Delta$ in $L^2(M)$ is well bounded from above; for example, this angle is at most $\frac{\pi}{2}$. Then we use Lemma 8 and Lemma 5 (c) to achieve the final convergence rate of eigenvalues. With the eigenvalue convergence, we apply Lemma 5 (b) again to show the convergence rate of eigenfunctions in $L^2(M)$ through the convergence rate of the eigenvalues and Lemma 5 (b). At last, we turn the convergence in $L^2(M)$ to the convergence in $L^\infty(M)$ with a slower convergence rate by the Sobolev embedding.
In this step, we control the operator norm. Note that Lemma 7 does not imply the convergence, and Lemma 8 only holds for a subset of the whole spectrum.

**Proof.** The proof is divided into few steps.

**Step 1, control \( R_k \):** In this step, we control \( \| R_k \|_2 \). For any \( f \in L^2(M) \) with \( \| f \|_2 = 1 \), we have \( f = \sum_{i=0}^{\infty} a_i \phi_i \), where \( \sum_{i=0}^{\infty} a_i^2 = 1 \). Take \( J = J(\varepsilon) \in \mathbb{N} \), which we will fix later. Let \( f = f_1 + f_2 \), where \( f_1 = \sum_{i=0}^{J} a_i \phi_i \) and \( f_2 = \sum_{i=J+1}^{\infty} a_i \phi_i \). Then,

\[
\| (R_k f, f) \|_2 \leq \| (R_k f_1, f_1) \|_2 + 2 \| (R_k f_1, f_2) \|_2 + \| (R_k f_2, f_2) \|_2
\]

\[
\leq \| f_1 \|_2 \| f_1 \|_2 + 2 \| f_1 \|_2 \| f_2 \|_2 + \| f_2 \|_2 \| f_2 \|_2
\]

\[
\leq 3 \| R_k f_1 \|_2 + \| R_k f_2 \|_2 \| f_2 \|_2.
\]

By Lemma 8,

\[
\| R_k f_1 \|_2 = \left\| R_k \left( \sum_{i=1}^{J} a_i \phi_i \right) \right\|_2 \leq \sum_{i=1}^{J} \| R_k \phi_i \|_2
\]

\[
\leq \sum_{i=1}^{J} C_4 \varepsilon^2 (1 + \lambda_i^{d/2+5}) \leq C_4 J \varepsilon^2 (1 + \lambda_j^{d/2+5}).
\]

Moreover,

\[
\left| \left( \frac{I - H_k}{\varepsilon^2} f , f \right) \right| = \left( \sum_{i=1}^{\infty} \frac{1 - e^{-\varepsilon^2 \lambda_i}}{\varepsilon^2} a_i \phi_i \right) \left( \sum_{i=1}^{\infty} a_i \phi_i \right)
\]

\[
= \sum_{i=1}^{\infty} a_i^2 \frac{1 - e^{-\varepsilon^2 \lambda_i}}{\varepsilon^2} > \frac{1}{2} \min \left( \frac{1}{\varepsilon}, \frac{1}{R} \right) = \frac{\lambda_1}{2},
\]

where we use the fact that when \( 0 < \varepsilon^2 < 0.1 \), we have [2] Theorem 4.3

\[
\frac{1 - e^{-\varepsilon^2}}{\varepsilon^2} \geq \frac{1}{2} \min \left( \frac{1}{\varepsilon}, \frac{1}{R} \right).
\]

Therefore,

\[
\| (I - H_k f, f) \|_2 \leq \frac{6 C_4 J}{\lambda_1} (1 + \lambda_j^{d/2+5}) \varepsilon^2.
\]

By Lemma 2, when \( \varepsilon > 0 \) is sufficiently small, we can take \( J \approx e^{-\frac{3d}{2(d+10)}} \) so that \( e^{-\frac{3d}{2(d+10)}} \leq \lambda_1^{d/2+5} \leq 2e^{-\frac{d}{4}}. \) Thus,

\[
\| (I - H_k f, f) \|_2 \leq \frac{18 C_4 e^{2(d+10)}}{\lambda_1} e^{\frac{3d}{2(d+10)}}.
\]

For \( f_2 \), which is in general a \( L^2 \) function, we apply Lemma 7

\[
\| R_k f_2 \|_2 \| f_2 \|_2 \leq C_5 \| f_2 \|_2^2.
\]
Since \( \{ \phi_i \} \) forms an orthonormal basis, we have \( \left( \frac{I - H_\varepsilon}{\varepsilon^2} f_2, f_1 \right) = \left( \frac{I - H_\varepsilon}{\varepsilon^2} f_1, f_2 \right) = 0. \) For \( f_2 \), we control \( \left( \frac{I - H_\varepsilon}{\varepsilon^2}, f, f \right) \) in a way different from (23).

\[
\left| \left( \frac{I - H_\varepsilon}{\varepsilon^2} f, f \right) \right| = \left( \frac{I - H_\varepsilon}{\varepsilon^2} f_1, f_1 \right) + \left( \frac{I - H_\varepsilon}{\varepsilon^2} f_2, f_2 \right) \geq \left( \frac{I - H_\varepsilon}{\varepsilon^2} f_2, f_2 \right) \\
= \sum_{i \neq f+1} d_i^2 \frac{1 - e^{-2\lambda_i}}{\varepsilon^2} \geq \frac{1}{2} \min \left( \lambda_{J+1}, \frac{1}{\varepsilon} \right) \| f_2 \|_2^2 \geq \frac{1}{2} \min \left( \lambda_J, \frac{1}{\varepsilon} \right) \| f_2 \|_2^2,
\]

where we use (24) in the second last step. Hence, by the choice of \( J \), we have

\[
\frac{\| R_\varepsilon f_2 \|_2}{\| \left( \frac{I - H_\varepsilon}{\varepsilon^2}, f, f \right) \|} \leq 2C_3 \max \left( \varepsilon, \frac{1}{\lambda_J} \right) \leq 2C_3 \varepsilon^{\frac{3}{3m}}.
\]

By putting the above together, since \( \varepsilon^{\frac{3}{3m}} > \varepsilon^{\frac{d^2+4d+10}{2d(d+1)}} \), by the lower bound of \( \lambda_1 \) stated in Lemma 2 (or the Faber-Krahn lower bound of the first non-trivial eigenvalue), there is a constant \( \tilde{C}_1 > 0 \) satisfying

\[
\frac{18C_4}{\lambda_1} + 2C_3 \leq \tilde{C}_1
\]

such that

(25) \[
\frac{\| R_\varepsilon f_2 \|_2}{\| \left( \frac{I - H_\varepsilon}{\varepsilon^2}, f, f \right) \|} \leq \tilde{C}_1 \varepsilon^{\frac{3}{3m}}.
\]

**Step 2, bound the angle between two eigenfunctions:** Suppose \( \lambda_{i,e} \) is the \( i \)-th eigenvalue of \( \frac{I - H_{e\varepsilon}}{\varepsilon^2} \). Note that the \( i \)-th eigenvalue of \( \frac{1 - H_\varepsilon}{\varepsilon^2} \) is \( \frac{1 - e^{-2\lambda_i}}{\varepsilon^2} \). By Lemma 5 (a), (25) lead to

\[
\left| \lambda_{i,e} - \frac{1 - e^{-2\lambda_i}}{\varepsilon^2} \right| \leq \tilde{C}_1 \varepsilon^{\frac{3}{3m}} \left( 1 - e^{-2\lambda_i} \right)
\]

for all \( i \). Hence, for any \( i, e \in \mathbb{N} \), we have

(26) \[
\left| (1 - \lambda_{i,e} e^2) - e^{-2\lambda_i} \right| \leq \tilde{C}_1 \varepsilon^{\frac{3}{3m}} (1 - e^{-2\lambda_i}) \leq \tilde{C}_1 \varepsilon^{\frac{3}{3m} + 2} \lambda_i,
\]

where we use the trivial bound \(|1 - e^{-t}| \leq x \) for \( x > 0 \). Hence, trivially

(27) \[
\| (1 - \lambda_{i,e} e^2) \phi_i - e^{-2\lambda_i} \phi_i \|_2 \leq \tilde{C}_1 \varepsilon^{\frac{3}{3m} + 2} \lambda_i.
\]

On the other hand, by Lemma 8 when \( i < K \), we have

(28) \[
\| T_e \phi_i - e^{-2\lambda_i} \phi_i \|_2 = \| (I - H_e) \phi_i - (I - T_e) \phi_i \|_2 \leq C_4 e^4 (1 + \lambda_K^{d/2 + 5}).
\]

Thus, if we combine (27) and (28), when \( i < K \), we have

(29) \[
\left\| \frac{I - T_e}{\varepsilon^2} \phi_i - \lambda_i \phi_i \right\|_2 \leq C_4 e^2 (1 + \lambda_K^{d/2 + 5}) + \tilde{C}_1 e^{\frac{3}{3m} + 2} \lambda_K.
\]

Next, we first find the lower bounds for the gaps between \( 1 - \lambda_{i-1,e} e^2 \), \( 1 - \lambda_{i,e} e^2 \) and \( 1 - \lambda_{i+1,e} e^2 \). First, note that (26) implies that

\[
1 - \lambda_{i+1,e} e^2 \geq e^{-2\lambda_i e} + \tilde{C}_1 e^{\frac{3}{3m} + 2} \lambda_{i+1},
\]

\[
1 - \lambda_{i,e} e^2 \geq e^{-2\lambda_i} - \tilde{C}_1 e^{\frac{3}{3m} + 2} (\lambda_i + \lambda_{i+1}).
\]

Thus,

(30) \[
(1 - \lambda_{i,e} e^2) - (1 - \lambda_{i+1,e} e^2) \geq (e^{-2\lambda_i} - e^{-2\lambda_{i+1}}) - \tilde{C}_1 e^{\frac{3}{3m} + 2} (\lambda_i + \lambda_{i+1}).
\]
To bound \((1 - \lambda_{i,e} \varepsilon^2) - (1 - \lambda_{i+1,e} \varepsilon^2)\), we control \(e^{-\varepsilon^2 \lambda_i} - e^{-\varepsilon^2 \lambda_{i+1}}\) and \(\tilde{C_1} \varepsilon^{\frac{3}{\pi^2 \hbar}} (\lambda_i + \lambda_{i+1})\) separately. When

\[
\lambda_K \varepsilon^2 \leq 1,
\]

which holds under the assumption of \(\varepsilon\), we have \(\varepsilon^2 (\lambda_{i+1} - \lambda_{i}) \leq \varepsilon^2 \lambda_{i+1} \leq \varepsilon^2 \lambda_K \leq 1\) since \(i < K\). Combined with the fact that \(e^{-\varepsilon^2 \lambda_i} \geq e^{-1}\), we have

\[
e^{-\varepsilon^2 \lambda_i} - e^{-\varepsilon^2 \lambda_{i+1}} = e^{-\varepsilon^2 \lambda_i} [1 - e^{-\varepsilon^2 (\lambda_{i+1} - \lambda_{i})}] \geq \frac{1}{e} \left( \frac{1}{2} (\lambda_{i+1} - \lambda_{i}) \right) \varepsilon^2 \geq \frac{1}{6} (\lambda_{i+1} - \lambda_{i}) \varepsilon^2.
\]

Next we control \(\tilde{C_1} \varepsilon^{\frac{3}{\pi^2 \hbar}} (\lambda_i + \lambda_{i+1})\), and we want to have the following bound:

\[
-\tilde{C_1} \varepsilon^{\frac{3}{\pi^2 \hbar}} (\lambda_{i+1} + \lambda_i) \geq -\frac{1}{12} (\lambda_{i+1} - \lambda_{i}) \varepsilon^2,
\]

which is equivalent to requiring

\[
\varepsilon^{\frac{3}{\pi^2 \hbar}} \leq \frac{1}{12} \frac{\lambda_{i+1} - \lambda_{i}}{\lambda_{i+1} + \lambda_{i}} \leq \frac{\Gamma_K}{24 \tilde{C_1} \lambda_K}.
\]

By plugging (32) and (33) into (30), we have

\[
(1 - \lambda_{i,e} \varepsilon^2) - (1 - \lambda_{i+1,e} \varepsilon^2) \geq \frac{1}{12} (\lambda_{i+1} - \lambda_{i}) \varepsilon^2.
\]

A similar argument leads to

\[
\gamma \left( \frac{I - T_e}{\varepsilon^2} \right) \geq \frac{1}{12} \min \left\{ (\lambda_i - \lambda_{i-1}), (\lambda_{i+1} - \lambda_{i}) \right\} \geq \frac{1}{12} \Gamma_K,
\]

by the assumption of \(\Gamma_K\). Next, with (29) and (35), we apply Lemma 5 (b) to get

\[
|\langle \phi_i, \phi_i,e \rangle| \geq 1 - 12 \frac{C_4 \varepsilon^2 (1 + \lambda_K^{d/2+5}) + \tilde{C_1} \varepsilon^{\frac{3}{\pi^2 \hbar}} \lambda_K}{\min \{ (\lambda_i - \lambda_{i-1}), (\lambda_{i+1} - \lambda_{i}) \}} \geq 1 - 12 \frac{C_4 \varepsilon^2 (1 + \lambda_K^{d/2+5}) + \tilde{C_1} \varepsilon^{\frac{3}{\pi^2 \hbar}} \lambda_K}{\Gamma_K}.
\]

If

\[
12 \frac{C_4 \varepsilon^2 (1 + \lambda_K^{d/2+5}) + \tilde{C_1} \varepsilon^{\frac{3}{\pi^2 \hbar}} \lambda_K}{\Gamma_K} < \frac{1}{2},
\]

which holds when

\[
\varepsilon^2 < \frac{\Gamma_K}{48 C_4 (1 + \lambda_K^{d/2+5})}
\]

and

\[
\varepsilon^{\frac{3}{\pi^2 \hbar}} < \frac{\Gamma_K}{48 \tilde{C_1} \lambda_K},
\]

then we have

\[
|\langle \phi_i, \phi_i,e \rangle| \geq \frac{1}{2}.
\]

**Step 3, find the final bounds:** With the lower bound of the angle among eigenfunctions shown in (38), we boost the bound of the eigenvalues and hence the eigenfunctions. By Lemma 8 (38), and Lemma 5(c), we have

\[
\left| \lambda_{i,e} - \frac{1 - e^{-\varepsilon^2 \lambda_i}}{\varepsilon^2} \right| \leq \frac{\| R_e (\phi_i) \|}{|\langle \phi_i, \phi_i,e \rangle|} \leq 2 C_4 \varepsilon^2 (1 + \lambda_K^{d/2+5})
\]
By summing up the above two inequalities, we have
\[ |\lambda_{i,e} - \lambda_i| \leq \left( 2C_4 + 2C_4 \lambda_N^{d/2+5} + \lambda_N^2 \right) \epsilon^2. \]
Taking \( C_2 = \max\{2C_4, 1\} \), we have
\[ |\lambda_{i,e} - \lambda_i| \leq C_2 \left( 1 + \lambda_N^{d/2+5} + \lambda_N^2 \right) \epsilon^2. \]
Next, by (43), we have
\[ \left\| \lambda_{i,e} \phi_i - \left( \frac{1 - e^{-e^2 \lambda_i}}{e^2} \right) \phi_i \right\|_2 \leq 2C_4 \epsilon^2 (1 + \lambda_N^{d/2+5}). \]
By Lemma 8
\[ \left\| \frac{I - T_e}{e^2} \phi_i - \frac{I - H_e}{e^2} \phi_i \right\|_2 = \left\| \frac{I - T_e}{e^2} \phi_i - \left( \frac{1 - e^{-e^2 \lambda_i}}{e^2} \right) \phi_i \right\|_2 \leq C_4 \epsilon^2 (1 + \lambda_N^{d/2+5}). \]
So, by combining (42) and (43), we have for all \( i \leq K \):
\[ \left\| \frac{I - T_e}{e^2} \phi_i - \lambda_{i,e} \phi_i \right\|_2 \leq 3C_4 \epsilon^2 \left( 1 + \lambda_N^{d/2+5} \right). \]
Recall that in (35), the eigengap for \( \lambda_{i,e} \) is controlled by \( \gamma \left( \frac{I - T_e}{e^2} \right) \geq \frac{1}{12} \Gamma_K \). As a result, by Lemma 5(b), there are \( a_i \in \{-1, 1\} \) such that
\[ \|a_i \phi_i - \phi_i\|_2 \leq \frac{72C_4 \epsilon^2 (1 + \lambda_N^{d/2+5})}{\Gamma_K}. \]
Let us temporarily define \( C_{1,K} = C_2 \left( 1 + \lambda_N^{d/2+5} + \lambda_N^2 \right) \) and \( C_{2,K} = \frac{72C_4 (1 + \lambda_N^{d/2+5})}{\Gamma_K} \). We have
\[ |\lambda_{i,e} - \lambda_i| \leq C_{1,K} \epsilon^2, \]
\[ \|a_i \phi_i - \phi_i\|_2 \leq C_{2,K} \epsilon^2. \]

With the above \( L^2 \) bound of the eigenfunction estimate, we use the Sobolev Embedding theorem [25, Theorem 9.2] to control the convergence rate in \( L^\infty \) norm. There is a constant \( C_3 \) depending on the Ricci curvature of the manifold, such that
\[ \|a_i \phi_i - \phi_i\|_\infty \leq C_3 \left( \|a_i \phi_i - \phi_i\|_2 + \|\Delta^d (a_i \phi_i - \phi_i)\|_2 \right) \]
\[ = C_3 \left[ \|a_i \phi_i - \phi_i\|_2 + \|a_i \left( \Delta^d - \left( \frac{T_e - I}{e^2} \right)^d \right) \phi_i + \left( \frac{T_e - I}{e^2} \right)^d a_i \phi_i - \Delta^d \phi \|_2 \right] \]
\[ \leq C_3 \left[ \|a_i \phi_i - \phi_i\|_2 + C_3 \left( \frac{T_e - I}{e^2} \right)^d a_i \phi_i - \Delta^d \phi \|_2 \right] \]
\[ + C_3 \left[ \left( \Delta^d - \left( \frac{T_e - I}{e^2} \right)^d \right) \phi_i \right\|^2_2. \]
Now we bound the right hand side term by term. The second term of (46) is controlled by
\[ \left\| \frac{T_e - I}{e^2} \right\|_d a_i \phi_i - \Delta^d \phi \right\|_2 = \|\lambda_{i,e}^d a_i \phi_i - \lambda_{i,e}^d \phi \|_2 \]
\[ \leq \|\lambda_{i,e}^d a_i \phi_i - \lambda_{i,e}^d a_i \phi_i - \lambda_{i,e}^d a_i \phi_i - \lambda_{i,e}^d \phi \|_2 \]
\[ \leq |\lambda_{i,e}^d - \lambda_{i,e}^d| + \lambda_{i,e}^d \|a_i \phi_i - \phi\|_2 \| \cdot | \leq 2^d d \lambda_{i,e}^{d-1} |\lambda_{i,e}^d - \lambda_{i,e}^d| + \lambda_{i,e}^d \|a_i \phi_i - \phi\|_2, \]
where in the last step, we use the bound \( \lambda_\epsilon \leq \lambda + C_1 e^2 \leq 2 \lambda \), which holds when

\[
\epsilon^2 \leq \frac{\lambda}{C_{1,K}}.
\]

For the third term of (45), by Lemma 6(d) and a straightforward expansion, there is a constant \( \tilde{C}_4 \) depending on \( d, p_m, \) the \( C^2 \) norm of \( p, \) the Ricci curvature, the second fundamental form of the manifold such that

\[
\| \left( \Delta^d - \frac{T_e - I}{\epsilon^2} \right)^d \phi_{\epsilon,x} \|_2 \leq \tilde{C}_4 \left( \| \Delta \phi_{\epsilon,x} \|_2 - \frac{T_e - I}{\epsilon^2} \phi_{\epsilon,x} \|_2 \right) \left\| \left( \frac{T_e - I}{\epsilon^2} \right)^{d+1} \phi_{\epsilon,x} \right\|_\infty.
\]

Based on Lemma 15 in Appendix, there is a constant \( C_\epsilon \) depending on \( d, \) the volume, the curvature and the diameter of \( M \) such that

\[
\left\| \left( \Delta^d - \left( \frac{T_e - I}{\epsilon^2} \right)^d \phi_{\epsilon,x} \right) \phi_{\epsilon,x} \right\|_2 \leq \tilde{C}_4 \left( \| \Delta \phi_{\epsilon,x} \|_2 - \frac{T_e - I}{\epsilon^2} \phi_{\epsilon,x} \|_2 \right) \left\| \left( \frac{T_e - I}{\epsilon^2} \right)^{d+1} \phi_{\epsilon,x} \right\|_\infty \leq \tilde{C}_4 2^{d+1} \lambda^d + (1 + C_{1,K} + (C_6 + 2 \lambda) C_{2,K} + 4 \lambda_4^2) \| \phi_{\epsilon,x} \|_\infty \epsilon^2.
\]

In the last step, we use again \( \lambda_\epsilon \leq \lambda + C_{1,K} \epsilon^2 \leq 2 \lambda \), when \( \epsilon^2 \leq \frac{\lambda}{C_{1,K}} \).

Therefore, if we plug (45), (47) and (50) into (46), we have

\[
\| a_i \phi_{\epsilon,x} - \phi_i \|_\infty \leq \tilde{C}_3 (C_{2,K} (1 + \lambda_4^d) + 2^d \lambda^d - 1 C_{1,K}) \epsilon^2 + \tilde{C}_5 2^{d+1} \lambda^d + (1 + C_{1,K} + (C_6 + 2 \lambda) C_{2,K} + 4 \lambda_4^2) \| \phi_{\epsilon,x} \|_\infty \epsilon^2.
\]

where \( \tilde{C}_3 := \tilde{C}_3 \max \{ 2^d d, 2^{d+1} \tilde{C}_4 \}. \) If we require \( \tilde{C}_3 (C_{2,K} (1 + \lambda_4^d) + 2^d \lambda^d - 1 C_{1,K}) \epsilon^2 \leq 1 \) and \( \tilde{C}_3 2^{d+1} \lambda^d + (1 + C_{1,K} + (C_6 + 2 \lambda) C_{2,K} + 4 \lambda_4^2) \| \phi_{\epsilon,x} \|_\infty \epsilon^2 \leq \frac{1}{2}, \) then we have

\[
\| \phi_{\epsilon,x} \|_\infty - \| \phi_i \|_\infty \leq \| a_i \phi_{\epsilon,x} - \phi_i \|_\infty \leq \frac{1}{2} \| \phi_{\epsilon,x} \|_\infty + 1.
\]

Hence, by Lemma 1, we have

\[
\| \phi_{\epsilon,x} \|_\infty \leq 2 \| \phi_i \|_\infty + 2 \leq 2 (C_1 \lambda_{d+4}^d + 1).
\]

Thus, (51) becomes

\[
\| a_i \phi_{\epsilon,x} - \phi_i \|_\infty \leq \tilde{C}_5 (C_{2,K} (1 + \lambda_4^d) + 2^d \lambda^d - 1 C_{1,K}) \epsilon^2 + 2 \tilde{C}_5 2^{d+1} \lambda^d + (1 + C_{1,K} + (C_6 + 2 \lambda) C_{2,K} + 4 \lambda_4^2) (C_1 \lambda_{d+4}^d + 1) \epsilon^2.
\]
Finally, we list the relationship

\[ \mathcal{C}_5 C_{2,K} (1 + \lambda_K^d) \epsilon \leq \frac{1}{4}, \]

\[ \mathcal{C}_5 \lambda_K^{d-1} C_{1,K} \epsilon \leq \frac{1}{4}, \]

\[ 2\mathcal{C}_5 \lambda_K^{d+1} (1 + C_{1,K} + 4\lambda_K^2) (C_1 \lambda_K^{d+1} + 1) \epsilon \leq \frac{1}{4}, \]

\[ 2\mathcal{C}_5 \lambda_K^{d+1} (C_6^d + 2\lambda_K) (C_1 \lambda_K^{d+1} + 1) \epsilon \leq \frac{1}{4}, \]

then \( \| a_i \phi_{i,e} - \phi_i \|_\infty \leq \epsilon \).

**Step 4, manage the relationship among parameters:** Finally, we list the relationship in Table 5.2 between \( \epsilon \) and \( \lambda_K \) to obtain the convergence of eigenvalues and eigenvectors. Note that (1") is stronger than (31), and (2")-(5") come from (34), (36), (37), and (48). If (6")-(10") are satisfied, then we have

| No. | Relationship |
|-----|-------------|
| (1") | \( \lambda_K \epsilon^2 \leq \frac{1}{4} \) |
| (2") | \( \epsilon \leq \frac{1}{4 C_1 \lambda_K} \) |
| (3") | \( \epsilon^2 \leq \frac{\epsilon}{4 C_1 (1 + \lambda_K^{d+1})} \) |
| (4") | \( \epsilon \leq \frac{\epsilon}{2 \mathcal{C}_5 (1 + \lambda_K^{d+1})} \) |
| (5") | \( \epsilon \leq \frac{\epsilon}{4 \mathcal{C}_1 \lambda_K} \) |
| (6") | \( \tilde{C}_2 (1 + \lambda_K^{d/2+5} + \lambda_K^2) \epsilon^{1/2} \leq 1 \) |
| (7") | \( \frac{72 C_1 (1 + \lambda_K^{d/2+5})}{\mathcal{C}_1} \epsilon^{1/2} \leq 1 \) |
| (8") | \( \tilde{C}_5 (1 + \lambda_K^d) \frac{72 C_1 (1 + \lambda_K^{d/2+5})}{\mathcal{C}_1} \epsilon \leq \frac{\epsilon}{4} \) |
| (9") | \( \tilde{C}_5 \tilde{C}_2 \lambda_K^{d-1} (1 + \lambda_K^{d/2+5} + \lambda_K^2) \epsilon \leq \frac{\epsilon}{4} \) |
| (10") | \( 2\tilde{C}_5 \lambda_K^{d+1} (1 + \tilde{C}_2 (1 + \lambda_K^{d/2+5} + \lambda_K^2) + 4\lambda_K^2) (C_1 \lambda_K^{d+1} + 1) \epsilon \leq \frac{\epsilon}{4} \) |
| (11") | \( \epsilon \leq \frac{\min(f_{K,1})}{C_1 \lambda_K^{d-1/4-3\epsilon+1}} \) |
| (12") | \( C_1 \lambda_K^{d-1/4} \epsilon^{1/4} \leq 1 \) |
| (13") | \( \frac{\epsilon^2}{3 \epsilon_{\text{max}}} \lambda_K^{d-1/2} \epsilon \leq \frac{\epsilon}{3} \) |
| (14") | \( 144 \tilde{C}_5 \lambda_K^{d+1} (C_6^d + 2\lambda_K) \frac{(1 + \lambda_K^{d/2+5})}{\mathcal{C}_1 \lambda_K^{d+1}} (C_1 \lambda_K^{d+1} + 1) \epsilon \leq \frac{\epsilon}{4} \) |

(6")-(10") are satisfied, then we have

\[ | \lambda_{i,e} - \lambda_i | \leq \epsilon \frac{\epsilon}{2}, \quad \| a_i \phi_{i,e} - \phi_i \|_2 \leq \epsilon \frac{\epsilon}{2}, \quad \| a_i \phi_{i,e} - \phi_i \|_\infty \leq \epsilon. \]
Moreover, (11"), (12") and (13") will be used in the proof of propositions later. Note that (2"), (4"), (5"), (6"), (7") and (11") are satisfied whenever we have

\[(59) \quad \epsilon^{1/2} \leq \frac{\tilde{C}_8 \min(\Gamma_K, 1)}{d + \lambda_K^{d/2+5}}.\]

Observe that if \(\lambda_K < 1\), then \(\lambda_K^{d/2+5} + \lambda_K^{1/2} + \lambda_K^{(d+10)/6} + \lambda_K^{(d-1)/4} < 4\). If \(\lambda_K \geq 1\), then \(\lambda_K^{d/2+5} + \lambda_K^{1/2} + \lambda_K^{(d+10)/6} + \lambda_K^{(d-1)/4} \leq 4\lambda_K^{d/2+5}\). Hence, (59) is true whenever

\[(60) \quad \epsilon^{1/2} \leq \frac{\tilde{C}_6 \min(\Gamma_K, 1)}{d + \lambda_K^{d/2+5}}.\]

where \(\tilde{C}_7 > 1\). Based on the requirement (59), (1"), (8"), (9") and (13") are satisfied when we further require \(\epsilon^2 \leq \frac{\tilde{C}_8}{\lambda_K}\), \(\epsilon^{1/2} \leq \frac{\tilde{C}_6}{d + \lambda_K}\), \(\epsilon \leq \frac{\tilde{C}_7}{\lambda_K}\) and \(\epsilon^{1/2} \leq \frac{\tilde{C}_8}{d + \lambda_K}\) respectively. Hence it is sufficient to require

\[(61) \quad \epsilon^{1/2} \leq \frac{\tilde{C}_8}{1 + \lambda_K^{d/2}}.\]

Based on (60) and (61), (10") and (14") hold when

\[(62) \quad \epsilon^{1/2} \leq \frac{\tilde{C}_9}{\lambda_K^{d/2+2}(1 + C_1\lambda_K^{d/4})}.\]

If we combine (60) and (62), we set the relation between \(\epsilon\) and \(\lambda_K\) to be

\[(63) \quad \epsilon \leq \mathcal{K}_1 \min(\lambda_K^{2/3}, 1) \left(\frac{\min(\Gamma_K, 1)}{d + \lambda_K^{d/2+5}}\right)^2 \frac{1}{\lambda_K^{(5d+7)/4}}\]

where \(\mathcal{K}_2, \mathcal{K}_3 > 1\). Note that \(\mathcal{K}_1, \mathcal{K}_2, \) and \(\mathcal{K}_3\) depend on all the constants in Table 5.2 hence they depend on \(d, p_m,\) the \(C^2\) norm of \(p\) and the volume, the injectivity radius, the curvature and the second fundamental form of the manifold.

5.3. Relation between the operator \(T_{n,\epsilon}\) and the operator \(T_\epsilon\). We now show that the sequence of operators \(T_{n,\epsilon}\) collectively compactly converge to \(T_\epsilon\). We need to study the following spaces of functions.

**Definition 6.** Take \(u \in C(M)\). Define

\[\mathcal{K}_\epsilon = \{k_\epsilon(x, \cdot) | x \in M\}\]

\[u \mathcal{D}_\epsilon = \{u(\cdot)Q_\epsilon(x, \cdot) | x \in M\}\]

\[\mathcal{Q}_\epsilon = \{Q_\epsilon(x, \cdot)Q_\epsilon(y, \cdot) | x, y \in M\}\]

In general, we cannot show \(\sup_{f \in C(M)} |P_n f - P f|\) goes to 0. However, the following proposition shows that it is true for functions in the above spaces. The similar statements can be found in several places. However, for the sake of self-containedness, we provide a proof in the Appendix.

**Proposition 2.** *(Gilvenko-Cantelli Class)* Take \(u \in C(M)\) with \(\|u\|_\infty \leq 1\). Fix \(\epsilon > 0\) small enough. Then there is a constant \(C_{gc} > 0\) depending on \(d,\) the diameter of \(M, p_m,\) and the
such that

\[
\sup_{f \in \mathcal{X}} |P_n f - Pf| \leq \frac{C_{gc}}{\sqrt{n}} \left( \sqrt{-\log \varepsilon + \log n} \right),
\]

\[
\sup_{f \in \mathcal{U}} |P_n f - Pf| \leq \frac{C_{gc}}{\sqrt{m}e^{2d}} \left( \sqrt{-\log \varepsilon + \log n} \right),
\]

\[
\sup_{f \in \mathcal{D}} |P_n f - Pf| \leq \frac{C_{gc}}{\sqrt{m}e^d} \left( \sqrt{-\log \varepsilon + \log n} \right).
\]

In the next few Lemmas, we provide bounds so that we can apply Theorem 5. Recall that the operators $T_n, T_{n,e}$ and $\hat{T}_n, \hat{T}_{n,e}$ in Definition 5 are operators from $(C(M), \| \cdot \|_{\infty})$ to $(C(M), \| \cdot \|_{\infty})$. We provide the bounds of them in the next lemma.

**Lemma 9.** Suppose $f \in C(M)$ and $\| f \|_{\infty} \leq 1$. For $\varepsilon > 0$ small enough, there are constants $C_5 > 0$, $C_6 > 0$ and $C_7 > 0$ depending on $p_m$ and the $C^0$ norm of $p$ so that:

(a) For any $x \in M$, $C_5 e^d \leq d_{f}(x) \leq C_6 e^d$.

(b) For any $x \in M$, $\| PQ_n f(x) \| \leq \frac{C_6}{C_5} e^d$.

(c) For any $x \in M$, $PQ_n(x) \geq \frac{C_5}{C_6} e^d$.

(d) $\| T_n \| \leq \frac{C_3}{C_7}$.

Fix $\varepsilon > 0$ small enough. Suppose $n$ is large enough so that $\frac{1}{\sqrt{4n}} (\sqrt{-\log \varepsilon + \log n}) < \mathcal{C}_1$, where $\mathcal{C}_1$ depends on $d$, the diameter of $M$, $p_m$, and the $C^0$ norm of $p$. Then, with probability greater than $1 - n^{-2}$, the following bounds hold.

(a') For any $x \in M$, $C_5 e^d \leq d_{n,e}(x) \leq C_6 e^d$.

(b') For any $x \in M$, $|PQ_n f(x)| \leq \frac{C_6}{C_5} e^d$ and $|P_n Q_n f(x)| \leq \frac{C_6}{C_5} e^d$.

(c') For any $x \in M$, $P_n Q_n(x) \geq \frac{C_5}{C_6} e^d$.

(d') $\| T_{n,e} \| \leq \frac{C_3}{C_7}$ and $\| \hat{T}_{n,e} \| \leq \frac{C_3}{C_7}$.

Moreover, there is a constant $C_7 > 0$ depending on $d$, the diameter of $M$, $p_m$, and the $C^0$ norm of $p$, such that with probability greater than $1 - n^{-2}$,

(e') For any $x, y \in M$, $|Q_n f(x) - Q_n f(y)| \leq \frac{C_7}{\sqrt{4n}} (\sqrt{-\log \varepsilon + \log n})$.

**Proof.** The proof of (a) follows from Lemma 8(a), i.e. there are constants $0 < \bar{C}_1 \leq \bar{C}_2$ such that

$\bar{C}_1 e^d \leq d_f(x) \leq \bar{C}_2 e^d$,

where $\bar{C}_1$ and $\bar{C}_2$ depend on $p_m$ and the $C^0$ norm of $p$. To prove (a'), by Proposition 2, we have

$|d_{n,e}(x) - d_f(x)| \leq \frac{C_{gc}}{\sqrt{n}} \left( \sqrt{-\log \varepsilon + \log n} \right)$.

Hence, as long as $\frac{C_{gc}}{\sqrt{n}} (\frac{1}{e} + \log n) < \frac{\bar{C}_2}{2} e^d$, which is equivalent to

\[
\frac{1}{\sqrt{4n}} \left( \sqrt{-\log \varepsilon + \log n} \right) < \frac{\bar{C}_1}{2C_{gc}} = \mathcal{C}_1,
\]

we have

\[
\sup_{f \in \mathcal{X}} |P_n f - Pf| \leq \frac{C_{gc}}{\sqrt{n}} \left( \sqrt{-\log \varepsilon + \log n} \right),
\]

\[
\sup_{f \in \mathcal{U}} |P_n f - Pf| \leq \frac{C_{gc}}{\sqrt{m}e^{2d}} \left( \sqrt{-\log \varepsilon + \log n} \right),
\]

\[
\sup_{f \in \mathcal{D}} |P_n f - Pf| \leq \frac{C_{gc}}{\sqrt{m}e^d} \left( \sqrt{-\log \varepsilon + \log n} \right).
\]
we have
\[ \frac{\hat{C}_1}{2} \epsilon^d \leq d_{n,\epsilon}(x) \leq 2\hat{C}_2 \epsilon^d. \]

Note that \( \sqrt{\frac{\log n}{n}} \) goes to 0 as \( n \) goes to infinity. Hence, the term \( \frac{1}{\sqrt{n \epsilon^d}} (\sqrt{- \log \epsilon} + \sqrt{\log n}) \) will be smaller than the constant \( c_1 \), when \( n \) is large enough. The conclusion follows by setting \( C_5 := \frac{3}{4} \) and \( C_6 := 2\hat{C}_2 \).

For (b), by a direct expansion and the bound of \( d_{\epsilon} \) in (a), we have
\[ |PQf(x)| = \left| \int_{M} k_{\epsilon}(x, y) f(y) dP(y) \right| \leq \frac{\sup_{x \in M} d_{\epsilon}(x)}{(\inf_{x \in M} d_{\epsilon}(x))^2} \leq \frac{C_6}{C_5^2 \epsilon^d}. \]

For (b'), with probability greater than \( 1 - n^{-2} \), the following bounds hold:
\[ |P_n Q_{n,\epsilon} f(x)| = \frac{1}{n} \sum_{i=1}^{n} k_{\epsilon}(x, x_i) f(x_i) \leq \frac{\sup_{x \in M} d_{n,\epsilon}(x)}{(\inf_{x \in M} d_{n,\epsilon}(x))^2} \leq \frac{C_6}{C_5^2 \epsilon^d}, \]
\[ |P_n Q_{n,\epsilon} f(x)| = \frac{1}{n} \sum_{i=1}^{n} k_{\epsilon}(x, x_i) f(x_i) \leq \frac{\sup_{x \in M} d_{n,\epsilon}(x)}{(\inf_{x \in M} d_{n,\epsilon}(x))^2} \leq \frac{C_6}{C_5^2 \epsilon^d}, \]
where we apply the bound for \( d_{n,\epsilon} \) in (a').

For (c), again, by (a),
\[ Q_{\epsilon}(x, y) \geq \frac{k_{\epsilon}(x, y)}{(\sup_{x \in M} d_{\epsilon}(x))^2} \geq \frac{k_{\epsilon}(x, y)}{C_5^2 \epsilon^d}, \]
and hence \( P Q_{\epsilon}(x) \geq \frac{d_{\epsilon}(x)}{C_5^2 \epsilon^d} \geq \frac{C_5}{C_5^2 \epsilon^d} \). (c') follows from (a') and a direct expansion:
\[ Q_{n,\epsilon}(x, y) \geq \frac{k_{\epsilon}(x, y)}{(\sup_{x \in M} d_{n,\epsilon}(x))^2} \geq \frac{k_{\epsilon}(x, y)}{C_5^2 \epsilon^d}. \]
Hence \( P_n Q_{n,\epsilon}(x) \geq \frac{d_{n,\epsilon}(x)}{C_5^2 \epsilon^d} \geq \frac{C_5}{C_5^2 \epsilon^d} \). Similarly, (d) follows from (b) and (c) and (d') follows from (b') and (c').

For (e'), we have
\[ |Q_{n,\epsilon}(x, y) - Q_{\epsilon}(x, y)| \leq |k_{\epsilon}(x, y)| \frac{1}{d_{n,\epsilon}(x) d_{n,\epsilon}(y)} - \frac{1}{d_{\epsilon}(x) d_{\epsilon}(y)} \]
\[ \leq \frac{d_{\epsilon}(x) d_{\epsilon}(y) - d_{n,\epsilon}(x) d_{n,\epsilon}(y)}{d_{n,\epsilon}(x) d_{n,\epsilon}(y) d_{\epsilon}(x) d_{\epsilon}(y)} \]
\[ \leq \frac{|d_{\epsilon}(x) - d_{n,\epsilon}(x)|}{d_{\epsilon}(x) d_{n,\epsilon}(y) d_{\epsilon}(y)} + \frac{|d_{\epsilon}(y) - d_{n,\epsilon}(y)|}{d_{n,\epsilon}(x) d_{n,\epsilon}(y) d_{\epsilon}(x)} \]
\[ \leq \frac{C_5}{\sqrt{n \epsilon^p}} \left( \sqrt{- \log \epsilon} + \sqrt{\log n} \right) \]
for a constant \( C_5 > 0 \), where the second bound holds since \( |k_{\epsilon}(x, y)| \leq 1 \) and we use (a) and Proposition 12 in the last step. \( \square \)

Next, we need to bound the term \( \| T_{n,\epsilon} u - T_{\epsilon} u \| \), which is needed when we apply Theorem 5.

**Lemma 10.** Fix \( u \in C(M) \) with \( \| u \|_{\infty} \leq 1 \). Fix any \( \epsilon \) small enough. Suppose \( n \) is large enough so that \( \frac{1}{\sqrt{n \epsilon^p}} (\sqrt{- \log \epsilon} + \sqrt{\log n}) < c_1 \), where \( c_1 \) is defined in (67). There is a
constant \( C_8 > 0 \) depending on \( d \), the diameter of \( M \), \( p_m \) and the \( C^0 \) norm of \( p \) such that with probability greater than \( 1 - n^{-2} \), we have
\[
|T_{n,e}u(x) - T_{e}u(x)| \leq \frac{C_8}{\sqrt{n}e^{2d}} \left( \sqrt{-\log \epsilon} + \sqrt{\log n} \right)
\]
for any \( x \in M \).

**Proof.** By a direct expansion and triangular inequality,
\[
|T_{n,e}u(x) - T_{e}u(x)| = \left| \frac{P_nQ_n e u(x)}{PQ_e(x)} - \frac{PQ_e u(x)}{PQ_e(x)} \right| + \left| \frac{P_nQ_n e u(x) - P_nQ_n e u(x)}{PQ_e(x)} \right|
\]
which is further bounded by
\[
\left( \frac{|P_nQ_n e u(x)|}{P_nQ_n e(x)PQ_e(x)} + \frac{1}{PQ_e(x)} \right) |P_nQ_n e u(x) - P_nQ_n e u(x)|
\]
By Lemma 9
\[
\frac{|P_nQ_n e u(x)|}{P_nQ_n e(x)PQ_e(x)} + \frac{1}{PQ_e(x)} \leq \left( \frac{C_6}{C_5} + \frac{C_6}{C_5} \right) e^d,
\]
where \( C_5 \) and \( C_6 \) are the constants in Lemma 9. By \((e')\) in Lemma 9
\[
|P_nQ_n e u(x) - P_nQ_n e u(x)| \leq \frac{C_7}{\sqrt{n}e^{2d}} \left( \sqrt{-\log \epsilon} + \sqrt{\log n} \right).
\]
By Proposition 2
\[
|P_nQ e u(x) - PQ e u(x)| \leq \frac{C_6}{\sqrt{n}e^{2d}} \left( \sqrt{-\log \epsilon} + \sqrt{\log n} \right).
\]
Hence, by putting the above together, \( |T_{n,e}u(x) - T_{e}u(x)| \leq \frac{C_8}{\sqrt{n}e^{2d} + 1} \), for some constant \( C_8 > 0 \) as claimed.

Next, we need to bound the term \( \| (T_{e} - T_{n,e})T_{n,e} \| \) so that we can apply Theorem 5. Note that for any fixed \( \epsilon \), \( \| T_{e} - T_{n,e} \| \) does not go to 0 in general as \( n \) goes to infinity, neither does \( \| T_{e} - \hat{T}_{n,e} \| \). Otherwise, the spectral convergence is straightforward and there is no need to introduce the concept of collective compact convergence. Indeed, although we can establish the inequality shown in Lemma 10, the inequality only holds for a single function. To achieve \( \| T_{e} - T_{n,e} \| \to 0 \), we need to control all \( u \in C^0 \) satisfying \( \| u \|_\infty \leq 1 \). However, this cannot be achieved directly with the uniform bound. This bound is the critical step in controlling the convergence rate when we have a finite sampling set. In other words, we cannot show that \( \|(T_{e} - T_{n,e})T_{n,e}\| \) goes to 0 by showing that \( \|(T_{e} - T_{n,e})\| \) goes to 0. To bypass this difficulty, we need the following intermediate lemma.
Proof. For any $f \in C(M)$ such that $\|f\|_\infty \leq 1$,

$$
\sup_{x \in M} |\hat{T}_{n,e} f(x) - T_{n,e} f(x)| \leq \sup_{x \in M} \left| \frac{P_n Q_{n,e} f(x)}{P_{Q,e}(x)} - \frac{P_n Q_{e} f(x)}{P_{Q,e}(x)} \right|
$$

$$
\leq \sup_{x \in M} \left( \frac{|P_n Q_{n,e} f(x)|}{P_{Q,e}(x)} |P_n Q_{n,e} f(x) - P_{Q,e} f(x)| \right) + \sup_{x \in M} \left( \frac{1}{P_{Q,e}(x)} |P_n Q_{n,e} f(x) - P_{n Q_e} f(x)| \right).
$$

We control the last step term by term. First,

$$
|P_n Q_{n,e} f(x) - P_{Q,e} f(x)| \leq |P_n Q_{n,e} f(x) - P_n Q_{e} f(x)| + |P_n Q_{e} f(x) - P_{Q,e} f(x)|
$$

$$
\leq \left( \frac{C_7}{n^{e^d}} + \frac{C_{2e}}{n^{2e^d}} \right) \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right)
$$

by Proposition 2 and (e') in Lemma 9. Moreover,

$$
|P_n Q_{n,e} f(x) - P_{n Q_e} f(x)| \leq \frac{C_7}{n^{e^d}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right)
$$

by (e') in Lemma 9. In summary,

$$
\sup_{x \in M} |\hat{T}_{n,e} f(x) - T_{n,e} f(x)| \leq \frac{C_9}{n^{e^d}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right)
$$

for some constant $C_9 > 0$, which is our first claim.

For the second claim, similarly, for any $f \in C(M)$ such that $\|f\|_\infty \leq 1$,

$$
\sup_{x \in M} \left| (T_{e} - \hat{T}_{n,e}) T_{n,e} f(x) \right|
$$

$$
= \sup_{x \in M} \left| \frac{1}{P_{Q,e}(x)} \left( \int_M Q_{e}(x,y) \int_M Q_{e}(y,z) dP_n(z) dP(y) \right) \right|
$$

$$
- \int_M Q_{e}(x,y) \int_M Q_{e}(y,z) dP_n(z) dP(y)
$$

is bounded by

$$
\frac{1}{\inf_{x \in M} P_{Q,e}(x)} \sup_{x \in M} \left| \int_M Q_{e}(x,y) \int_M Q_{e}(y,z) dP_n(z) dP(y) \right|
$$

$$
- \int_M Q_{e}(x,y) \int_M Q_{e}(y,z) dP_n(z) dP(y)
$$

$$
\leq \frac{1}{\inf_{x \in M} P_{Q,e}(x)} \sup_{x \in M} \left| \int_M Q_{e}(x,y) \int_M Q_{e}(y,z) dP_n(z) dP(y) \right|
$$

$$
- \int_M Q_{e}(x,y) \int_M Q_{e}(y,z) dP_n(z) dP(y)
$$

$$
= \frac{1}{\inf_{x \in M} P_{Q,e}(x)} \sup_{x \in M} \left| \int_M Q_{e}(x,y) Q_{e}(y,z) dP(y) \right|
$$

$$
- \int_M Q_{e}(x,y) Q_{e}(y,z) dP(y) \right| f(z) dP_n(z) \right|,
$$

Moreover, the diameter of $M$, $\rho_n$ and the $C^0$ norm of $\rho$ such that with probability greater than $1 - n^{-2}$,

$$
\|T_{n,e} - T_{n,e}\| \leq \frac{C_9}{n^{e^d}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right), \quad \|(T_{e} - \hat{T}_{n,e}) T_{n,e}\| \leq \frac{C_{10}}{n^{e^d}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right).
$$

Lemma 11. Fix $\varepsilon > 0$ small enough. Suppose $n$ is sufficiently large so that $\frac{1}{\sqrt{n \varepsilon}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right) < \xi_1$, where $\xi_1$ is defined in (67). There are constants $C_9$ and $C_{10}$ depending on $d$, the diameter of $M$, $\rho_n$ and the $C^0$ norm of $\rho$ such that with probability greater than $1 - n^{-2}$,
where we use Fubini’s Theorem in the last equality. As a result, by Proposition 2 and Lemma 9,

\[
\sup_{x \in M} |(T_e - T_{n,e})T_{n,e} f(x)| \leq \frac{C_4 e^{2d}}{C_3^2 \sqrt{n e^{3d}}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right) = \frac{C_{10}}{\sqrt{n e^{2d}}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right),
\]

by setting \( C_{10} := \frac{C_4 C_6}{C_3^2} > 0 \), and hence the proof follows.

**Lemma 12.** Fix \( \varepsilon > 0 \) small enough. Suppose \( n \) is sufficiently large so that \( \frac{1}{\sqrt{n e^d}} (\sqrt{-\log \varepsilon} + \sqrt{\log n}) < \mathcal{C}_1 \), where \( \mathcal{C}_1 \) is defined in (67). There is a constant \( C_{11} > 0 \) depending on \( d \), the diameter of \( M \), \( p_m \) and the \( C^0 \) norm of \( p \), such that with probability greater than \( 1 - n^{-2} \), we have

\[
\|(T_e - T_{n,e})T_{n,e}\| \leq \frac{C_{11}}{\sqrt{n e^{2d}}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right).
\]

**Proof.** By a direct expansion and Lemma 11 we have

\[
\|(T_e - T_{n,e})T_{n,e}\| \leq \|T_e T_{n,e} - T_e T_{n,e}\| + \|T_e T_{n,e} - T_{n,e} T_{n,e}\|
\]

\[
+ \|T_{n,e} T_{n,e} - T_{n,e} T_{n,e}\| + \|T_{n,e} T_{n,e} - T_{n,e} T_{n,e}\|
\]

\[
\leq (\|T_e\| + \|T_{n,e}\| + \|T_{n,e}\| + \|T_{n,e}\|) \|T_{n,e} - T_{n,e}\| + \|T_e - T_{n,e}\| \|T_{n,e}\|
\]

\[
\leq \frac{3C_6^3}{C_3^2} \frac{C_9}{\sqrt{n e^{2d}}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right) + \frac{C_{10}}{\sqrt{n e^{2d}}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right).
\]

The conclusion follows by setting \( C_{11} := \frac{3C_6^3}{C_3^2} C_9 + C_{10} \).

In the next Lemma, we show that \( T_e \) is a compact operator.

**Lemma 13.** \( T_e \) is a compact operator. In particular, for any function \( f \in C(M) \) such that \( \|f\|_\infty \leq 1 \), when \( \|t(x) - t(x')\|_{\mathbb{R}^D} \) is small enough, for \( x, x' \in M \), we have

\[
|T_e f(x) - T_e f(x')| \leq C_{12} e^{-d} \|t(x) - t(x')\|_{\mathbb{R}^D},
\]

where \( C_{12} > 0 \) only depends on the curvature of \( M \), \( p_m \) and the \( C^1 \) norm of \( p \).

**Proof.** First, by (d) in Lemma 9 we have \( \|T_e\| \leq \frac{C_1}{C_3} \). Next, since \( Q_{\varepsilon}(\cdot, y) \) is a smooth function on \( M \) and \( M \) is compact, for \( x, x' \in M \) with \( \|t(x) - t(x')\|_{\mathbb{R}^D} \) small enough, we have

\[
|Q_{\varepsilon}(x, y) - Q_{\varepsilon}(x', y)| \leq \tilde{C}_1 e^{-2d} d(x, x') \leq 2\tilde{C}_1 e^{-2d} \|t(x) - t(x')\|_{\mathbb{R}^D},
\]

where \( \tilde{C}_1 > 0 \) depends on the curvature of \( M \), \( p_m \) and the \( C^1 \) norm of \( p \). Here \( d(x, x') \) is the geodesic distance between \( x \) and \( x' \), and we use the fact that \( d(x, x') \) can be bounded by \( 2\|t(x) - t(x')\|_{\mathbb{R}^D} \) whenever \( \|t(x) - t(x')\|_{\mathbb{R}^D} \) is small enough. Thus,

\[
\left| \int_M Q_{\varepsilon}(x, y) dP(y) - \int_M Q_{\varepsilon}(x', y) dP(y) \right| \leq \int_M |Q_{\varepsilon}(x, y) - Q_{\varepsilon}(x', y)| dP(y)
\]

\[
\leq 2\tilde{C}_1 e^{-2d} \|t(x) - t(x')\|_{\mathbb{R}^D},
\]

where we use the fact that \( \int_M dP(y) = 1 \). Similarly, for any function \( f \in C(M) \) such that \( \|f\|_\infty \leq 1 \),

\[
\left| \int_M Q_{\varepsilon}(x, y) f(y) dP(y) - \int_M Q_{\varepsilon}(x, y) f(y) dP(y) \right| \leq 2\tilde{C}_1 e^{-2d} \|t(x) - t(x')\|_{\mathbb{R}^D}.
\]
As a result,
\[ |T_x f(x) - T_x f(x')| = \left| \int_M Q_\varepsilon(x,y) f(y) dP(y) - \int_M Q_\varepsilon(x',y) f(y) dP(y) \right| \]
is bounded by
\[
\frac{\left| \int_M Q_\varepsilon(x,y) f(y) dP(y) - \int_M Q_\varepsilon(x',y) f(y) dP(y) \right|}{\int_M Q_\varepsilon(x,y) dP(y)} + \frac{\left| \int_M Q_\varepsilon(x',y) f(y) dP(y) \right|}{\int_M Q_\varepsilon(x,y) dP(y)} \varepsilon \sum_{i=1}^\varepsilon \left( \int_M Q_\varepsilon(x,y) dP(y) \right) \sup_{\varepsilon' \in M} \left| \int_M Q_\varepsilon(x,y) dP(y) - \int_M Q_\varepsilon(x',y) dP(y) \right| \]\
\[
\leq C_1 \varepsilon^{-d} \left| \int (x) - \int (x') \right|_{\mathbb{R}^d},
\]
for a constant $C_1 > 0$. Hence, for a sequence of functions $\{f_\varepsilon\}_{\varepsilon=1}^\infty \subset C(M)$ such that $\|f_\varepsilon\|_{\infty} \leq 1$, $\{T_\varepsilon f_\varepsilon\}_{\varepsilon=1}^\infty$ is equi-continuous. Then, $T_\varepsilon$ is compact by the Arzelà-Ascoli theorem. \hfill \square

In the next Lemma, we show that $T_{n,\varepsilon}$ almost surely converges compactly to $T_\varepsilon$.

**Lemma 14.** For $\varepsilon > 0$ small enough, as $n \to \infty$, $T_{n,\varepsilon}$ almost surely converges collectively compactly to $T_\varepsilon$ on the Banach space $(C(M), \|\cdot\|_\infty)$.

**Proof.** Denote $B_1 \subset C(M)$ to be the unit ball in $(C(M), \|\cdot\|_\infty)$ centered at 0. By Lemma 10, $T_{n,\varepsilon}$ almost surely converges pointwise to $T_\varepsilon$. Hence, we only need to show that $\cup_{n=1}^\infty (T_{n,\varepsilon} - T_\varepsilon)B_1$ is almost surely relatively compact. Since $T_\varepsilon$ is a compact operator, it suffices to show that $\cup_{n=1}^\infty T_{n,\varepsilon}B_1 \subset C(M)$ is almost surely relatively compact.

For any sequence $\{g_\varepsilon\}_{\varepsilon=1}^\infty \subset \cup_{n=1}^\infty T_{n,\varepsilon}B_1$, without loss of generality, assume we can choose a subsequence of the form $\{T_{n,\varepsilon}f_1, T_{n,\varepsilon}f_2, \cdots\}$. Clearly, there is a $N := N(\varepsilon)$ such that if $i > N(\varepsilon)$, then $\frac{1}{\sqrt{x^d}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log i} \right) < \gamma_1$, where $\gamma_1$ is defined in (67). Hence, by (d’) in Lemma 9 with probability greater than $1 - \frac{1}{x^d}$, $\|T_{n,\varepsilon}f_i\|_\infty$ is bounded by a constant. Hence, by the Borel-Cantelli lemma, $\{T_{N+1,\varepsilon}f_N, T_{N+1,\varepsilon}f_{N+1}, \cdots\}$ is almost surely bounded.

Consider any element $T_{j,\varepsilon}f_j$, where $j \geq N$, in the above subsequence. By (69), we know that if $\|\int (x) - \int (x')\|_{\mathbb{R}^d}$ is small enough,
\[
|Q_\varepsilon(x,y) - Q_\varepsilon(x',y)| \leq 2\tilde{C}_1 \varepsilon^{-2d} \|\int (x) - \int (x')\|_{\mathbb{R}^d}.
\]
Moreover, by (68) in the proof of Lemma 9 with probability greater than $1 - j^{-2}$, for any $x,y \in M$,
\[
|Q_{j,\varepsilon}(x,y) - Q_\varepsilon(x,y)| \leq \frac{C_7}{\sqrt{\varepsilon J^d}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log j} \right).
\]
Hence, with probability greater than $1 - j^{-2}$,
\[
|Q_{j,\varepsilon}(x,y) - Q_{j,\varepsilon}(x',y)| \leq 2\tilde{C}_1 \varepsilon^{-2d} \|\int (x) - \int (x')\|_{\mathbb{R}^d} + \frac{2C_7}{\sqrt{\varepsilon J^d}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log j} \right).
\]
It follows from the Arzelà-Ascoli theorem that

Therefore,

\begin{align*}
\text{Lemma 15.} \quad & \text{Proof can also be found in [34].} \\
& \text{standard approach, but we provide details for the sake of self-containedness. A similar}
\end{align*}

is bounded by

\begin{align*}
|T_{j,\epsilon}f_j(x) - T_{j,\epsilon}f_j(x')| &= \left| \int_M Q_{j,\epsilon}(x, y) f_j(y) dP_j(y) - \int_M Q_{j,\epsilon}(x', y) f_j(y) dP_j(y) \right|
\end{align*}

for \( \tilde{C} > 0 \), where we use (b') and (c') in Lemma \[9\]. As a result, for any \( \delta > 0 \), there exists \( N'(\delta, \epsilon) \in \mathbb{N} \) so that when \( j > N'(\delta, \epsilon) \),

\begin{align*}
&\frac{1}{\sqrt{\epsilon \log \delta}} \left( \sqrt{-\log \epsilon} + \sqrt{\log j} \right)
\end{align*}

As a result, we can further refine the subsequence of \( \{T_{N,\epsilon, f_N, T_{N+1,\epsilon, f_{N+1}}, \cdots}\} \) into a subsequence that is almost surely bounded and equicontinuous by the Borel-Cantelli Lemma. It follows from the Arzelà-Ascoli theorem that \( \cup_{n=1}^\infty T_{n,\epsilon} B_1 \subset C(M) \) is almost surely relatively compact.

Next, we introduce the following lemma to bound the resolvent. The proof follows the standard approach, but we provide details for the sake of self-containment. A similar proof can also be found in [43].

\begin{align*}
\text{Lemma 15.} \quad & \text{Let } \lambda \text{ be an isolated eigenvalue of an operator } T_\epsilon. \text{ If } r < \frac{1}{2} \text{dist}(\{\lambda\}, \sigma(T_\epsilon) \setminus \{\lambda\}), \text{ then}
\end{align*}

\begin{align*}
\max_{z \in \Gamma_\epsilon(\lambda)} |R_z(T_\epsilon)| \leq \frac{1}{r}
\end{align*}

\begin{align*}
\text{Proof.} \quad & \text{Fix any } z \in \Gamma_\epsilon(\lambda). \text{ Set } |\lambda_z| = |R_z(T_\epsilon)|, \text{ where } \lambda_z \text{ is the eigenvalue of } R_z(T_\epsilon) \text{ with the largest magnitude. If } v \text{ is the corresponding eigenvector, we have } (zI - T_\epsilon)^{-1}v = \lambda_z v.
\end{align*}

\begin{align*}
&\text{Therefore, } \langle zI - T_\epsilon \rangle v = \frac{1}{z} v. \text{ We conclude that } T_\epsilon v = (z - \frac{1}{z}) v; \text{ that is, } z - \frac{1}{z} \text{ is an eigenvalue of } T_\epsilon. \text{ Since } r < \frac{1}{2} \text{dist}(\{\lambda\}, \sigma(T_\epsilon) \setminus \{\lambda\}), \text{ we have dist}(z, \sigma(T_\epsilon)) = r \leq |z - (z - \frac{1}{z})|.
\end{align*}

\begin{align*}
\text{We conclude that } |\lambda_z| \leq \frac{1}{r}.
\end{align*}

Suppose \( \lambda_{i,\epsilon} \) is the \( i \)-th smallest eigenvalue of \( \frac{I - T_\epsilon}{\epsilon^2} \), then \( \tilde{\lambda}_{i,\epsilon} = 1 - \lambda_{i,\epsilon} \epsilon^2 \) is the \( i \)-th largest eigenvalue of \( T_\epsilon \). In other words, we have \( \lambda_0, \epsilon \leq \lambda_{2,\epsilon} \leq \cdots \leq \lambda_0 \leq \tilde{\lambda}_{0,\epsilon} \) for \( T_\epsilon \). Note that \( \frac{I - T_\epsilon}{\epsilon^2} \) and \( T_\epsilon \) share the same eigenfunctions; hence, we denote \( \phi_{i,\epsilon} \) to be the eigenfunction corresponding to \( \lambda_{i,\epsilon} \) (hence to \( \tilde{\lambda}_{i,\epsilon} \)). Similarly, let \( \lambda_{i,\epsilon} \) be the \( i \)-th smallest eigenvalue of \( \frac{I - T_\epsilon}{\epsilon^2} \), and \( \tilde{\lambda}_{i,\epsilon} = 1 - \lambda_{i,\epsilon} \epsilon^2 \) be the \( i \)-th largest eigenvalue of \( T_\epsilon \). Denote \( \phi_{i,\epsilon} \) to be the eigenfunction corresponding to \( \lambda_{i,\epsilon} \) (hence to \( \tilde{\lambda}_{i,\epsilon} \)). We assume that both \( \phi_{i,\epsilon} \) and \( \phi_{i,\epsilon} \) are normalized in the \( L^2 \) norm. The following proposition describes the spectral convergence of the operator \( T_{n,\epsilon} \) to the operator \( T_\epsilon \). Since we assume that the eigenvalues of \( \Delta \) are
simple, due to the convergence of the eigenvalues, we will have that for any $K \in \mathbb{N}$, there exists $\varepsilon > 0$ sufficiently small so that the first $K$ eigenvalues of $T_\varepsilon$ are simple. Therefore, even though we study the spectral convergence from $T_{n,\varepsilon}$ to $T_\varepsilon$ in the next proposition, to be consistent, we still only assume that the eigenvalues of $\Delta$ are simple rather than making the assumption on $T_\varepsilon$.

**Proposition 3.** Fix $K \in \mathbb{N}$. Assume that the eigenvalues of $\Delta$ are simple. Suppose $\varepsilon > 0$ is sufficiently small so that (6) is satisfied. If $n$ is large enough so that $\frac{\sqrt{n}e^2}{\sqrt{-\log \varepsilon + \sqrt{\log n}}} < \varepsilon_1$, where $\varepsilon_1 > 0$ is defined in (67), then there are $a_i \in \{1, -1\}$ such that with probability greater than $1 - n^{-2}$, for all $1 \leq i < K$, we have

$$|\bar{\lambda}_{i,\varepsilon} - \lambda_{i,n,\varepsilon}| \leq \frac{C_{13}}{\sqrt{n}e^{2d/3}} \left( \sqrt{-\log \varepsilon + \sqrt{\log n}} \right),$$

$$\|a_i \phi_{i,n,\varepsilon} - \phi_{i,\varepsilon}\| \leq \frac{C_{13}}{\sqrt{n}e^{2d/3}} \left( \sqrt{-\log \varepsilon + \sqrt{\log n}} \right).$$

where $C_{13} > 0$ is a constant depending on $d$, the diameter of $M$, $\rho_m$ and the $C^0$ norm of $p$.

**Proof.** Suppose $\lambda_i$, $i = 1, \ldots, K$, are the eigenvalues of $-\Delta$ so that $0 < \lambda_1 < \lambda_2 < \ldots$. Recall the definition (21):

$$\Gamma_K = \min_{1 \leq i < K} \text{dist}(\lambda_i, \sigma(-\Delta) \setminus \{\lambda_i\}).$$

Consider an $i < K$. Based on the choice of $\varepsilon$, we have shown in (35) in the proof of Proposition 1 that

$$\gamma \left( \frac{I - T_\varepsilon}{e^2} \right) \geq \frac{1}{12} \Gamma_K.$$

Hence,

$$\text{dist}(\{\bar{\lambda}_{i,\varepsilon}\}, \sigma(T_\varepsilon) \setminus \{\bar{\lambda}_{i,\varepsilon}\}) > \frac{\Gamma_K}{12} e^2.$$

Choose $r = \frac{\Gamma_K e^2}{24}$. By Lemma 15 we then have

$$\max_{z \in \Gamma_i(\bar{\lambda}_{i,\varepsilon})} \|R_z(T_\varepsilon)\| \leq 1 = \frac{1}{r} = \frac{\Gamma_K e^2}{24}.$$

Moreover, if $\varepsilon$ is small enough, $|z| > \frac{1}{2}$ for $z \in \Gamma_i(\bar{\lambda}_{i,\varepsilon})$.

Let $\text{Pr}_{\phi_{i,n,\varepsilon}}$ be the projection onto $\text{span}\{\phi_{i,n,\varepsilon}\}$. By Theorem 5 we have

$$\|\phi_{i,\varepsilon} - \text{Pr}_{\phi_{i,n,\varepsilon}} \phi_{i,\varepsilon}\|_{\infty} \leq \max_{z \in \Gamma_i} \frac{2r \|R_z(T_\varepsilon)\|}{\bar{\lambda}_{z,\varepsilon} - r} \left( \|T_{n,\varepsilon} - T_\varepsilon\|_{\infty} + \|R_z(T_\varepsilon)\phi_{i,\varepsilon}\|_{\infty} \right),$$

where $z$ we use the fact that since $T_\varepsilon$ and $T_{n,\varepsilon}$ are compact self-adjoint operators, the eigenvalues are all real, and the fact that $\min_{z \in \Gamma_i(\bar{\lambda}_{i,\varepsilon})} = \bar{\lambda}_{i,\varepsilon} - r$. By Proposition 1 we have $\lambda_{i,\varepsilon} \leq \bar{\lambda}_{i,\varepsilon} e^2$, and hence

$$\bar{\lambda}_{i,\varepsilon} - r = 1 - \lambda_{i,\varepsilon} e^2 - \frac{\Gamma_K e^2}{24} \geq 1 - (\lambda_i + \varepsilon^2) e^2 - \frac{\Gamma_K e^2}{24}$$

$$(73) \geq 1 - (2\lambda_K + \varepsilon^2) e^2 \geq \frac{1}{2},$$
where in the last step, we use (11') in Table 5.2. By Lemma 13, \( R_\epsilon(T_e) \leq 1, \| R_\epsilon(T_e) \phi_{i,e} \| \leq \| R_\epsilon(T_e) \| \| \phi_{i,e} \| \) and (72), we have

\[
\| \phi_{i,e} - \text{Pr}_{\phi_{i,e}} \phi_{i,e} \| \leq 4 \left( \left\| (T_{n,e} - T_e) \frac{\phi_{i,e}}{\| \phi_{i,e} \|} \right\| + \frac{24}{\Gamma K \epsilon^2} \| (T_e - T_{n,e}) T_{n,e} \| \right) \| \phi_{i,e} \|. 
\]

Next, by Lemma 10 and Lemma 12 suppose \( n \) is large enough so that \( \frac{1}{\sqrt{n \epsilon}} \left( \sqrt{- \log \epsilon} + \sqrt{\log n} \right) \epsilon < \epsilon_1 \), with probability greater than \( 1 - n^{-2} \), we further have

(74)

\[
\| \phi_{i,e} - \text{Pr}_{\phi_{i,e}} \phi_{i,e} \| \leq \left( \frac{4C_8}{\sqrt{n \epsilon^2} \delta} + \frac{96C_{11}}{\Gamma K} \frac{1}{\sqrt{n \epsilon^2} \delta + 2} \right) \left( \sqrt{- \log \epsilon} + \sqrt{\log n} \right) \| \phi_{i,e} \|. 
\]

By (74) and Lemma 4 there are \( a_i \in \{1, -1\} \) such that

(75)

\[
\| a_i \phi_{i,n,e} - \phi_{i,e} \| \leq 2 \| \phi_{i,n,e} - \text{Pr}_{\phi_{i,e}} \phi_{i,e} \| 
\]

Next, we discuss the eigenvalues. We have

\[
| \lambda_{i,e} - \lambda_{i,n,e} \phi_{i,e} \| = | \lambda_{i,e} \phi_{i,e} - \lambda_{i,n,e} \phi_{i,e} \| 
\]

\[
\leq \| \lambda_{i,e} \phi_{i,e} - \lambda_{i,n,e} a_i \phi_{i,n,e} \| + \| \lambda_{i,n,e} a_i \phi_{i,n,e} - \lambda_{i,n,e} \phi_{i,e} \| 
\]

\[
\leq \| T_e \phi_{i,e} - T_{n,e} a_i \phi_{i,n,e} \| + | \lambda_{i,n,e} | a_i \phi_{i,n,e} - \phi_{i,e} \|, 
\]

where we control the difference by taking \( \phi_{i,n,e} \) into account. The right hand side is further bounded by

\[
\| T_e \phi_{i,e} - T_{n,e} \phi_{i,e} \| + \| T_{n,e} \phi_{i,e} - T_{n,e} a_i \phi_{i,n,e} \| + | \lambda_{i,n,e} | a_i \phi_{i,n,e} - \phi_{i,e} \| 
\]

\[
\leq \left( \| T_e - T_{n,e} \phi_{i,e} \| \| \phi_{i,e} \| + 2 \| T_{n,e} \phi_{i,e} - a_i \phi_{i,n,e} \| \right) | \lambda_{i,n,e} | a_i \phi_{i,n,e} - \phi_{i,e} \|
\]

\[
\leq \left[ \frac{C_8}{\sqrt{n \epsilon^2} \delta} + \frac{2C_3}{C^3} \left( \frac{8C_8}{\sqrt{n \epsilon^2} \delta} + \frac{192C_{11}}{\Gamma K} \frac{1}{\sqrt{n \epsilon^2} \delta + 2} \right) \right] \left( \sqrt{- \log \epsilon} + \sqrt{\log n} \right) | \lambda_{i,n,e} | a_i \phi_{i,n,e} - \phi_{i,e} \|, 
\]

where we use Lemma 2 and Lemma 10 in the last step. Hence, we can cancel \( \| \phi_{i,e} \| \) on both sides, and get

\[
| \lambda_{i,e} - \lambda_{i,n,e} | \leq \left[ \frac{C_8}{\sqrt{n \epsilon^2} \delta} + \frac{2C_3^2}{C^2} \left( \frac{8C_8}{\sqrt{n \epsilon^2} \delta} + \frac{192C_{11}}{\Gamma K} \frac{1}{\sqrt{n \epsilon^2} \delta + 2} \right) \right] \left( \sqrt{- \log \epsilon} + \sqrt{\log n} \right). 
\]

Next, we simplify (75). By Lemma 13 \( \| \phi_{i} \| \leq C_1 \lambda_{i,K}^{\frac{d+1}{2}} \). Hence, by Proposition 1 we have \( \| \phi_{i,e} \| \leq C_1 \lambda_{i,K}^{\frac{d+1}{2}} + \epsilon \leq 2C_1 \lambda_{i,K}^{\frac{d+1}{2}} \) when \( \epsilon \) is sufficiently small. Therefore, if

(76)

\[
\epsilon \leq \frac{\min(\Gamma,K,1)}{2C_1 \lambda_{i,K}^{\frac{d+1}{2}} + 1}, 
\]

then \( \| \phi_{i,e} \|, \| \phi_{i,e} \|, \frac{1}{\Gamma K} \) are all bounded above by \( \frac{1}{\Gamma K} \). Note that (76) is (11") in Table 5.2 and it is sufficient to require \( \epsilon \leq \lambda_1 \min \left( \left( \frac{\min(\Gamma,K)}{\lambda_{i,K}^{d+1}} \right)^2, \frac{\Gamma K}{(\lambda_{i,K}^{d+1} + 1)^2} \right) \).
In conclusion, we have
\[ \| \phi_i - \phi_{i,n} \|_{\infty} \leq \frac{C_{13}}{\sqrt{n}e^{2d+3}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right) \]
and
\[ |\lambda_{i,n} - \lambda_{i,e}| \leq \frac{\| \lambda_{i,n} - \lambda_{i,e} \|}{\sqrt{\log n}} \leq \frac{C_{13}}{\sqrt{n}e^{2d+3}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right), \]
where \( C_{13} > 0 \) is a constant depending on the constants from \( C_5 \) to \( C_{11} \), and hence it depends on \( d \), the diameter of \( M \), \( p_{\text{min}} \) and the \( C^0 \) norm of \( p \).

**Remark 8.** We explain the reason why the eigenvalue's convergence rate is slower than that of the eigenfunction that is suggested in the previous proposition. Note that we can write
\[ |\lambda_{i,n} - \lambda_{i,e}| \leq \frac{\| T_{n,e} - T_{n,n} \| \phi_i}{\sqrt{n}e^{2d+3}} \leq \frac{C_{13}}{\sqrt{n}e^{2d+3}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right), \]
where \( C_{13} > 0 \) is a constant depending on the constants from \( C_5 \) to \( C_{11} \), and hence it depends on \( d \), the diameter of \( M \), \( p_{\text{min}} \) and the \( C^0 \) norm of \( p \).

We have a good control of the first term on the right hand side so that it matches the convergence rate of the eigenfunctions. However, we do not have a good control of the operator norm of \( T_{n,e} - T_{n,n} \) in \( L^{\infty}(M) \) since we use a trivial bound. Such trivial bound leads to an extra \( \frac{1}{\varepsilon^4} \), and hence the convergence rate of eigenvalues is slower.

In the previous proposition, we assume \( \varepsilon \) is fixed and small enough and \( n \) is large enough, then we characterize the spectral convergence rate of \( \frac{I - T_{n,e}}{\varepsilon^2} \) to \( \frac{I - T_{n,n}}{\varepsilon^2} \) in terms of \( n \) and \( \varepsilon \). Next, we impose a relation between \( n \) and \( \varepsilon \) to further simplify the above proposition.

**Corollary 1.** Fix \( K \in \mathbb{N} \). Assume that the eigenvalues of \( \Delta \) are simple. Suppose (6) holds and \( n \) is sufficiently large so that \( \varepsilon = \varepsilon(n) \geq \left( \log n \right)^{1/13} \). Then with probability greater than \( 1 - n^{-2} \), for all \( 1 \leq i < K \), we have
\[ |\lambda_{i,n} - \lambda_{i,e}| \leq 2C_{13} \varepsilon^2. \]

Suppose (6) holds and \( n \) is sufficiently large so that \( \varepsilon = \varepsilon(n) \geq \left( \log n \right)^{1/13} \). Then there are \( a_i \in \{1, -1\} \) such that with probability greater than \( 1 - n^{-2} \), for all \( 1 \leq i < K \), we have
\[ \| a_i \phi_{i,n,e} - \phi_{i,e} \|_{\infty} \leq 2C_{13} \varepsilon. \]

The constant \( C_{13} > 0 \) is defined in Proposition 7.

**Proof.** When \( \left( \frac{\log n}{n} \right)^{1/13} \leq \varepsilon \), we have
\[ \frac{\sqrt{\log n}}{\sqrt{n}} \frac{1}{e^{2d+3}} \leq \varepsilon^4 \quad \text{and} \quad -\log \varepsilon \leq \frac{\log n}{4d + 13}. \]
When \( \left( \frac{\log n}{n} \right)^{1/13} \leq \varepsilon \), we have
\[ \frac{\sqrt{\log n}}{\sqrt{n}} \frac{1}{e^{2d+3}} \leq \varepsilon \quad \text{and} \quad -\log \varepsilon \leq \frac{\log n}{4d + 8}. \]
Hence, we can check that the condition \( \frac{1}{\sqrt{\log n}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right) < \varepsilon \) is satisfied in both cases when \( n \) is sufficiently large, where \( \varepsilon \) is defined in Proposition 3. The conclusion follows by substituting the above inequality into the bounds into Proposition 3. □

5.4. Relation between \( I - A \) and the operator \( I - T_{n,e} \). The following proposition relates the eigenfunctions of \( I - T_{n,e} \) and the eigenvectors of the kernel normalized GL matrix \( I - A \) by using the Nyström extension.

**Proposition 4.** There is a bijection between the eigenpairs of \( I - A \) and the eigenpairs of \( I - T_{n,e} \) in the following sense:

1. If \((\lambda, f)\) is an eigenpair of \( I - T_{n,e} \), where \( f \in L^\infty(M) \), then \((\lambda, v)\) is an eigenpair of \( I - A \), where \( v \in \mathbb{R}^n \) and \( v(i) = f(x_i) \) for any \( i \).
2. If \((\lambda, v)\) is an eigenpair of \( I - A \), where \( v \in \mathbb{R}^n \), let

\[
f(x) := \frac{1}{\lambda} \frac{1}{n} \sum_{i=1}^{n} \frac{k_e(x_i)}{\sqrt[2]{q_e(x_i)}} v(i),
\]

then \((\lambda, f)\) is an eigenpair of \( I - T_{n,e} \). Moreover, \( f \in C^\infty(M) \) and \( v(i) = f(x_i) \) for any \( i \).

**Proof.** It is sufficient to prove the theorem for \( T_{n,e} \) and \( A \). For any \( g \in C(M) \), denote \( \tilde{g} = (g(x_1), \cdots, g(x_n))^\top \). Fix any \( x_k \in \{x_i\}_{i=1}^n \). By the definition of \( A \), we have

\[
Ag(x_k) = \sum_{i=1}^{n} A_{ki} g(x_i) = \sum_{i=1}^{n} \frac{W_{ki} g(x_i)}{D_{ki}} = \sum_{i=1}^{n} \frac{W_{ki} g(x_i)}{\sum_{i=1}^{n} W_{ki}}.
\]

Therefore,

\[
= \frac{1}{n} \sum_{i=1}^{n} \frac{k_e(x_i)}{\sqrt[2]{q_e(x_i)}} g(x_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{k_e(x_i)}{\sqrt[2]{q_e(x_i)}} g(x_i).
\]

Hence, if \( \lambda \) is an eigenvalue of \( T_{n,e} \) and \( f(x) \) is the corresponding eigenfunction, then for any \( x_k \), we have \( T_{n,e} f(x_k) = \lambda f(x_k) \). Therefore, the vector \( v \in \mathbb{R}^n \) with \( v(k) = f(x_k) \) satisfies \( Av = \lambda v \).

For the second statement, observe that

\[
f(x_j) = \frac{1}{\lambda} \frac{1}{n} \sum_{i=1}^{n} \frac{k_e(x_i)}{\sqrt[2]{q_e(x_i)}} v(i) = \frac{1}{\lambda} \frac{1}{n} \sum_{i=1}^{n} \frac{k_e(x_i)}{\sqrt[2]{q_e(x_i)}} \frac{1}{\sqrt[2]{q_e(x_j)}} v(j).\]

Next,

\[
T_{n,e} f(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{k_e(x_i)}{\sqrt[2]{q_e(x_i)}} \frac{1}{\sqrt[2]{q_e(x_j)}} f(x_j) = \frac{1}{n} \sum_{i=1}^{n} \frac{k_e(x_i)}{\sqrt[2]{q_e(x_i)}} \frac{1}{\sqrt[2]{q_e(x_j)}} v(i) = \frac{1}{n} \sum_{i=1}^{n} \frac{k_e(x_i)}{\sqrt[2]{q_e(x_i)}} = \lambda f(x).
\]

At last, the smoothness of \( f \) implies that it is a finite summation and quotient of nonzero smooth functions. Finally, note that \( W \) is positively definite by the Bochner theorem. We thus finish the proof. □
5.5. Renormalization of the eigenvectors of $\frac{L-\Delta}{\varepsilon^2}$. Denote $\mu_{i,n,\varepsilon}$ to be the $i$-th eigenvalue of $\frac{L-\Delta}{\varepsilon^2}$ with the associated eigenvector $\tilde{v}_{i,n,\varepsilon}$ normalized in the $\ell^2$ norm. By Proposition 4,

\begin{equation}
\tilde{\phi}_{i,n,\varepsilon}(x) := \frac{1}{1-\mu_{i,n,\varepsilon}\varepsilon^2} \frac{\sum_{j=1}^{n} k_e(x,j)}{\phi(x,j)} \tilde{v}_{i,n,\varepsilon}(j)
\end{equation}

is the $i$-th eigenfunction of $\frac{L-\Delta}{\varepsilon^2}$ with $\tilde{\phi}_{i,n,\varepsilon}(x_j) = \tilde{v}_{i,n,\varepsilon}(j)$. Note that by Corollary 1 and Proposition 3, it is intuitive to expect that $\tilde{\phi}_{i,n,\varepsilon}$ is close to a discretization of an eigenfunction of $T_\varepsilon$. In the previous propositions, we compare the eigenfunctions of operators $\frac{L-\Delta}{\varepsilon^2}$, $\frac{L-\Delta}{\varepsilon^2}$, and $\Delta$ normalized in $L^2(M)$, we would expect to “normalize” $\tilde{v}_{i,n,\varepsilon}$ in the $L^2(M)$ sense. In this subsection, we will make this intuition rigorous. We have the following lemma by using the convergence of $\frac{\tilde{\phi}_{i,n,\varepsilon}(x_j)}{\|\tilde{\phi}_{i,n,\varepsilon}\|_2}$ to $\phi(x_j)$.

**Lemma 16.** Fix $K \in \mathbb{N}$. Suppose (6) holds and $n$ is sufficiently large so that $\varepsilon = \varepsilon(n) \geq (\frac{\log n}{n})^{\frac{1}{m+n}}$. For all $1 \leq i < K$, with probability greater than $1 - n^{-2}$, we have

\begin{equation}
\max_{x_j} \left| \frac{\tilde{\phi}_{i,n,\varepsilon}(x_j)}{\|\tilde{\phi}_{i,n,\varepsilon}\|_2} \right| \varepsilon^{\frac{1}{2}} \leq 2
\end{equation}

and

\begin{equation}
\left| \frac{1}{n} \sum_{j=1}^{n} \tilde{\phi}_{i,n,\varepsilon}^2(x_j) \right| - \frac{1}{n} \sum_{j=1}^{n} \left| \frac{\tilde{\phi}_{i,n,\varepsilon}(x_j)}{\phi(x_j)} \right|^2 \leq \frac{4(C_{13} + 1)}{p_m} \varepsilon^{\frac{1}{2}},
\end{equation}

where $C_{13} > 0$ is defined in Proposition 3 and depends on $d$, the diameter of $M$, $p_m$ and the $C_0$ norm of $p$.

**Proof.** By Proposition 1 and Corollary 1, there are $a_i \in \{1, -1\}$ such that with probability greater than $1 - n^{-2}$,

\begin{equation}
\max_{x_j} \left| a_i \frac{\tilde{\phi}_{i,n,\varepsilon}(x_j)}{\|\tilde{\phi}_{i,n,\varepsilon}\|_2} - \phi(x_j) \right| \leq 2C_{13}\varepsilon + \varepsilon \leq (2C_{13} + 1)\varepsilon,
\end{equation}

where the last bound holds when $\varepsilon$ is sufficiently small. By the triangular inequality, we have

\begin{equation}
\max_{x_j} \left| \frac{\tilde{\phi}_{i,n,\varepsilon}(x_j)}{\|\tilde{\phi}_{i,n,\varepsilon}\|_2} \right| \leq \|\phi\|_\infty + (2C_{13} + 1)\varepsilon.
\end{equation}

Therefore,

\begin{equation}
\max_{x_j} \left| \frac{\tilde{\phi}_{i,n,\varepsilon}(x_j)}{\|\tilde{\phi}_{i,n,\varepsilon}\|_2} \right| \varepsilon^{\frac{1}{2}} \leq \|\phi\|_\infty \varepsilon^{\frac{1}{2}} + (2C_{13} + 1)\varepsilon^{\frac{5}{2}}.
\end{equation}

By Lemma 1, $\|\phi\|_\infty \varepsilon^{\frac{1}{2}} \leq C_1 \lambda_{\varepsilon}^{\frac{d-1}{d}} \varepsilon^{\frac{1}{2}}$. Note that $C_1 \lambda_{\varepsilon}^{\frac{d-1}{d}} \varepsilon^{\frac{1}{2}} \leq 1$ which is (12”) in Table 5.2. Hence,

\begin{equation}
\|\phi\|_\infty \varepsilon^{\frac{1}{2}} \leq 1
\end{equation}

when (6) is satisfied. Therefore, we conclude that

\begin{equation}
\max_{x_j} \left| \frac{\tilde{\phi}_{i,n,\varepsilon}(x_j)}{\|\tilde{\phi}_{i,n,\varepsilon}\|_2} \right| \varepsilon^{\frac{1}{2}} \leq 2,
\end{equation}

when (6) is satisfied. Therefore, we conclude that
where the last bound holds when $\epsilon$ is sufficiently small. By the triangular inequality again, we also have
\[
\max_{x_j} \left| \frac{1}{n} \sum_{j=1}^n \frac{\phi_{i,n,e}^2(x_j)}{p(x_j)} - \frac{1}{n} \sum_{j=1}^n \frac{\phi_{i,n,e}^2(x_j)}{p(x_j)} \right| = O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2}} \right),
\]
and hence the claim follows from
\[
36 \quad DAVID B DUNSON, HAU-TIENG WU, AND NAN WU
\]
\[
\max_{x_j} \left| \frac{1}{n} \sum_{j=1}^n \frac{\phi_{i,n,e}^2(x_j)}{p(x_j)} - \frac{1}{n} \sum_{j=1}^n \frac{\phi_{i,n,e}^2(x_j)}{p(x_j)} \right| \leq O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2}} \right),
\]
Therefore, we conclude that
\[
\max_{x_j} \left| \frac{1}{n} \sum_{j=1}^n \frac{\phi_{i,n,e}^2(x_j)}{p(x_j)} - \frac{1}{n} \sum_{j=1}^n \frac{\phi_{i,n,e}^2(x_j)}{p(x_j)} \right| \leq O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2}} \right),
\]
and hence the claim follows from
\[
\left| \frac{1}{n} \sum_{j=1}^n \frac{\phi_{i,n,e}^2(x_j)}{p(x_j)} - \frac{1}{n} \sum_{j=1}^n \frac{\phi_{i,n,e}^2(x_j)}{p(x_j)} \right| = O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2}} \right),
\]
where we use the trivial bound $\frac{1}{n} \sum_{j=1}^n \frac{1}{p(x_j)} \leq 1/p_m$. By using $\|\phi_i\|_\infty \frac{1}{n} \leq 1$ again, the conclusion follows when $\epsilon$ is small enough.

Let $\mathbb{N}(j) := \{b_{\epsilon}^p(t(x_j)) \cap \{t(x_1), \ldots, t(x_n)\}\}$. By usual kernel density estimation (Lemma B.5 and Lemma E.1 in [36]), we have the following lemma.

**Lemma 17.** When $n$ is large enough, we have with probability greater than $1 - n^{-2}$ that for all $j = 1, \ldots, n$
\[
\hat{p}(x_j) := \frac{d\mathbb{N}(j)}{|S^{d-1}| n \epsilon^d} = p(x_j) + O(\epsilon^2) + O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2}} \right),
\]
where $|S^{d-1}|$ is the volume of $S^{d-1}$ and the implied constant in $O(\epsilon^2)$ depends on the $C^2$ norm of $p$.

Since $\phi_i$ is normalized in $L^2(M)$, in next lemma, we show that we can construct an approximation of the constant 1, and hence an approximation of $\|\phi_{i,n,e}\|_2$ by using $\phi_i$.

**Lemma 18.** Fix $K \in \mathbb{N}$. Suppose (6) holds and $n$ is sufficiently large so that $\epsilon = \epsilon(n) \geq (\log n)^{-\beta}$. For all $0 \leq i < K$, we have with probability greater than $1 - n^{-2}$ that
\[
\left| \frac{1}{n} \sum_{j=1}^n \frac{\phi_i^2(x_j)}{p(x_j)} - 1 \right| < \epsilon.
\]

**Proof.** Suppose $p$ is the p.d.f. of the random variable $X$ with the range $M$. Define a random variable $F := \frac{\phi_i^2}{p(x_j)}$. Then, $\mathbb{E}[F] = \|\phi_i\|_2^2 = 1$ and $F_j := \frac{\phi_i^2(x_j)}{p(x_j)}$ can be regarded as i.i.d. samples from $F$. By Lemma [1] we have
\[
b := \frac{\|\phi_i^2\|_\infty}{p} \leq C_1 \lambda_K^{(d-1)/2},
\]
where $C_1$ is defined in Lemma [1] and
\[
\mathbb{E}[F^2] = \int_M \phi_i^4 dx = \frac{C_1^2}{\lambda_K^{(d-1)/2}} \int_M \phi_i^2 dx = \frac{C_1^2}{\lambda_K^{(d-1)/2}} \|\phi_i\|_2^2 \leq b.
\]
Hence,
\[ \sigma^2 := \text{Var}(F) \leq b - 1. \]

We apply Bernstein’s inequality to provide a large deviation bound. Recall that for \( \beta > 0 \), the Bernstein’s inequality is
\[
\Pr \left\{ \left| \frac{1}{n} \sum_{j=1}^{n} F_j - \mathbb{E}[F] \right| > \beta \right\} \leq e^{-\frac{n\beta^2}{2\sigma^2 + \frac{\beta^2}{2}}}.
\]

For \( \beta = \varepsilon < 1 \),
\[
\frac{n\varepsilon^2}{2\sigma^2 + \frac{\beta^2}{2}} \geq \frac{n\varepsilon^2}{3b}.
\]

Note that \( \frac{n\varepsilon^2}{2\sigma^2 + \frac{\beta^2}{2}} \geq n\varepsilon^3 \) when \( 3b\varepsilon \leq 1 \), which is satisfied when
\[
\frac{C_1^2}{p_m} \sqrt{d-1/2} \varepsilon \leq \frac{1}{3},
\]
which is (13") in Table 5.2. Note that (13") in Table 5.2 is satisfied when 6 is satisfied. Therefore, we have
\[
\Pr \left\{ \left| \frac{1}{n} \sum_{j=1}^{n} F_j - \mathbb{E}[F] \right| > \varepsilon \right\} \leq e^{-n\varepsilon^3}.
\]

Note that \( e^{-n\varepsilon^3} \leq \frac{1}{n^2} \) when \( \varepsilon \geq \frac{2\log n}{n} \), which is satisfied by the assumption \( (\frac{\log n}{n})^{\frac{1}{d+\pi}} \leq \varepsilon \).

If we combine the above three lemmas, we have the following proposition.

**Proposition 5.** Fix \( K \in \mathbb{N} \). Suppose 6 holds and \( n \) is sufficiently large so that \( \varepsilon = \varepsilon(n) \geq (\frac{\log n}{n})^{\frac{1}{d+\pi}} \). Then, with probability greater than \( 1 - n^{-2} \), for all \( 1 \leq i \leq K \), we have
\[
\max_j \left| \frac{\tilde{v}_{i,n,e}(j)}{\|\tilde{v}_{i,n,e}\|_{\tilde{\phi}_i(\varepsilon)}} - \frac{\tilde{\phi}_{i,n,e}(x_j)}{\|\tilde{\phi}_{i,n,e}\|_2} \right| \leq \mathcal{A}_d \varepsilon^{1/2},
\]
where the constant \( \mathcal{A}_d > 0 \) is a constant depending on \( d \), the diameter of \( M \), \( p_m \), and the \( C^2 \) norm of \( p \).

**Proof.** By Lemma 17 with probability greater than \( 1 - n^{-2} \),
\[
\frac{|S^{d-1}|\varepsilon d}{d} \sum_{j=1}^{n} \frac{\tilde{v}_{i,n,e}(j)}{N(j)} = \frac{1}{n} \sum_{j=1}^{n} \frac{\tilde{v}_{i,n,e}(j)}{\|\tilde{v}_{i,n,e}\|_{\tilde{\phi}_i(\varepsilon)}} \left[ 1 + O(\varepsilon^2) + O\left( \frac{\sqrt{\log(n)}}{n^{1/2}d^{1/2}} \right) \right],
\]
where the constant in \( O(\varepsilon^2) \) depends on the \( C^2 \) norm of \( p \). Hence,
\[
\frac{|S^{d-1}|\varepsilon d}{d} \sum_{j=1}^{n} \frac{\tilde{v}_{i,n,e}(j)}{N(j)} \left[ 1 + O(\varepsilon^2) + O\left( \frac{\sqrt{\log(n)}}{n^{1/2}d^{1/2}} \right) \right] = \frac{1}{n} \sum_{j=1}^{n} \frac{\tilde{v}_{i,n,e}(j)}{p(x_j)}.
\]

By the assumption \( (\frac{\log n}{n})^{\frac{1}{d+\pi}} = \varepsilon \), \( \sqrt{\frac{\log(n)}{n^{1/2}d^{1/2}}} \leq \varepsilon^2 \). Therefore, we have
\[
\frac{|S^{d-1}|\varepsilon d}{d} \sum_{j=1}^{n} \frac{\tilde{v}_{i,n,e}(j)}{N(j)} \left( 1 + O(\varepsilon^2) \right) = \frac{1}{n} \sum_{j=1}^{n} \frac{\tilde{v}_{i,n,e}(j)}{p(x_j)} = \frac{1}{n} \sum_{j=1}^{n} \frac{\tilde{\phi}_{i,n,e}(x_j)}{p(x_j)},
\]
where the constant in $O(\varepsilon^2)$ depends on $d$, the diameter of $M$, $p_m$, and the $C^2$ norm of $p$. By Lemma 18, with probability greater than $1 - n^{-2}$,

$$\frac{1}{n} \sum_{j=1}^{n} \left| \frac{\hat{\phi}_{i,n,e}(x_j)}{p(x_j)} - \|\hat{\phi}_{i,n,e}\|_2^2 \right| < \varepsilon \|\hat{\phi}_{i,n,e}\|_2^2.$$  

(82)

Hence, by (82) and Lemma 16, when $\varepsilon$ is sufficiently small, we have

$$\frac{1}{n} \sum_{j=1}^{n} \hat{\phi}_{i,n,e}^2(x_j) - \|\hat{\phi}_{i,n,e}\|_2^2 \leq \frac{8(C_{13} + 1)}{p_m} \|\hat{\phi}_{i,n,e}\|_2^2 \varepsilon^2,$$

where $C_{13}$ depends on $d$, the diameter of $M$, $p_m$ and the $C^0$ norm of $p$. Hence,

$$\left| \frac{S_{d-1}^d \varepsilon d}{d} \sum_{j=1}^{n} \kappa_{i,n,e}(j) \frac{\varepsilon^2}{\mathbb{N}(j)} \mathbb{N}(j) - \|\hat{\phi}_{i,n,e}\|_2^2 \right| \leq \frac{8(C_{13} + 1)}{p_m} \|\hat{\phi}_{i,n,e}\|_2^2 \varepsilon^2,$$

where the constant in $O(\varepsilon^2)$ depends on $d$, the diameter of $M$, $p_m$, and the $C^2$ norm of $p$. A straightforward expansion shows that,

$$\left| \frac{S_{d-1}^d \varepsilon d}{d} \sum_{j=1}^{n} \kappa_{i,n,e}(j) \frac{\varepsilon^2}{\mathbb{N}(j)} \mathbb{N}(j) - \|\hat{\phi}_{i,n,e}\|_2^2 \right| \leq \tilde{C}_1 \varepsilon^2 \|\hat{\phi}_{i,n,e}\|_2^2,$$

for a constant $\tilde{C}_1$ depending on $d$, the diameter of $M$, $p_m$, and the $C^2$ norm of $p$. Simply expand the absolute value and take the square root, we have

$$\frac{1 - \tilde{C}_1 \varepsilon^2}{\tilde{C}_1} \|\hat{\phi}_{i,n,e}\|_2 \leq \|\hat{v}_{i,n,e}\|_{L^2(1/\beta)} \leq \frac{1 + \tilde{C}_1 \varepsilon^2}{\tilde{C}_1} \|\hat{\phi}_{i,n,e}\|_2.$$

(83)

At last, by the fact that $\hat{\phi}_{i,n,e}(x_j) = \hat{v}_{i,n,e}(j)$, when $\varepsilon$ is sufficiently small, we have

$$\frac{\hat{v}_{i,n,e}(j)}{\|\hat{v}_{i,n,e}\|_{L^2(1/\beta)}} - \frac{\hat{\phi}_{i,n,e}(x_j)}{\|\hat{\phi}_{i,n,e}\|_2} \leq \frac{\tilde{C}_1}{1 + \tilde{C}_1 \varepsilon^2} \frac{\varepsilon^2}{\|\hat{\phi}_{i,n,e}\|_2} \|\hat{\phi}_{i,n,e}(x_j)\| \leq \frac{2\tilde{C}_1}{1 + \tilde{C}_1 \varepsilon^2} e^{1/2} < 2\tilde{C}_1 e^{1/2} = \mathcal{K}_4 e^{1/2},$$

where we apply Lemma 16 again in the second step, and $\mathcal{K}_4 = 2\tilde{C}_1 > 0$ is a constant depending on $d$, the diameter of $M$, $p_m$, and the $C^2$ norm of $p$. □

5.6. **Finish the Proof of Theorem 2** With the above preparation, we are finally ready to prove Theorem 2.

**Proof.** Combining Propositions 1-4 and Corollary 1 when $\left(\frac{\log n}{n}\right)^{1/13} \leq \varepsilon$, we have for each $1 \leq i < K$, \n
$$|\mu_{i,n,e} - \lambda_i| \leq \varepsilon^{3/2} + 2C_{13} \varepsilon^{3/2}.$$ \n
Further with Proposition 5 when $\left(\frac{\log n}{n}\right)^{1/13} \leq \varepsilon$, we know that there are $a_i \in \{1, -1\}$ such that for all $1 \leq i < K$, with probability greater than $1 - n^{-2}$,

$$\max_{x_j} |a_i \hat{v}_{i,n,e}(j) - \hat{\phi}(x_j)| \leq 2C_{13} \varepsilon + \varepsilon + \mathcal{K}_4 e^{1/2}.$$
In summary, we have
\[
|\mu_{i,n} - \lambda_i| \leq \Omega_1 \epsilon^{3/2},
\]
\[
|1 - e^{2\mu_{i,n} - e^{-t^2\lambda_i}}| \leq \Omega_1 \epsilon,
\]
\[
\max_{x_j} |a_{i\nu_{i,n}(j)} - \Phi(x_j)| \leq \Omega_2 \epsilon^{1/2},
\]
where \(\Omega_1 = 2C_{13} + 1\) and \(\Omega_2 = 2C_{13} + \mathcal{K}_4 + 1\). Hence \(\Omega_1\) and \(\Omega_2\) depend on \(d\), the diameter of \(M\), \(p_m\), and the \(C^2\) norm of \(p\).

Note that Lemma 1 does not include \(\lambda_0 = 0\) case, so we need to further prove the case when \(i = 0\). Recall that \(\mu_{0,n} = \lambda_0\) is the smallest eigenvalue of \(\text{L} - \text{A} \frac{\text{d}}{\text{d}}\). The eigenvalues of \(A\) are between \([0, 1]\) and 1 is an eigenvalue of \(A\) corresponding to the constant eigenvector. Therefore \(\mu_{0,n} = 0\). And \(\mu_{0,n} - \lambda_0 \leq \Omega_1 \epsilon^{3/2}\) holds trivially. If \(\nu_{0,n}\) is the eigenvector corresponding to \(\mu_{0,n}\) normalized in \(\mathbb{R}^2\), then \(\nu_{0,n} = [\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}]\) or \([-\frac{1}{\sqrt{n}}, \cdots, -\frac{1}{\sqrt{n}}]\).

We know that \(\Phi_0 = \frac{1}{\sqrt{\text{Vol}(\mathcal{M})}}\), hence by the same argument as Lemma 18 we have with probability greater than \(1 - n^{-2}\)
\[
\left|\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\text{Vol}(\mathcal{M})} - 1\right| < \epsilon.
\]
By Lemma 17 with probability greater than \(1 - n^{-2}\),
\[
\frac{\|\Phi_0\|_d^2}{\text{Vol}(\mathcal{M})} \geq \frac{1}{n^2} \sum_{j=1}^{n} \frac{1}{\text{Vol}(\mathcal{M})} = \frac{1}{n^2} \sum_{j=1}^{n} \frac{1}{\text{Vol}(\mathcal{M})} \left[1 + O(\epsilon) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2}}\right)\right].
\]
When \(\frac{\log(n)}{n^{1/2} \epsilon^{d/2}} \leq \epsilon\), with (84), the above equation implies that
\[
\frac{\|\Phi_0\|_d^2}{\text{Vol}(\mathcal{M})} \geq \frac{1}{n^2} \sum_{j=1}^{n} \frac{1}{\text{Vol}(\mathcal{M})} = \frac{1}{n} \left[1 + O(\epsilon)\right].
\]
Therefore, \(\nu_{0,n}(j) = \frac{1}{\sqrt{\text{Vol}(\mathcal{M})}}(1 + O(\epsilon))\), where the implied constant depends on the volume of \(M\), and the \(C^2\) norm of \(p\). In this case,
\[
\max_{x_j} |a_{i\nu_{i,n}(j)} - \Phi_0(x_j)| \leq \Omega_2 \epsilon,
\]
where \(\Omega_2\) depends on the volume of \(M\), and the \(C^2\) norm of \(p\). With the \(i > 0\) cases, the conclusion follows.

6. Proof of Theorem \[\]

Note that
\[
\left|\text{H}^K_{E_d}(ij) - \text{H}(x_i, x_j, t)\right|
\leq \sum_{l=0}^{K-1} \left(\int e^{-\mu_{i,n} t} \nu_{i,n}(i) \nu_{i,n}(j) - e^{-\lambda_l t} \Phi_0 \Phi_0(x_j)\right) + \sum_{l=K}^{\infty} e^{-\lambda_l t} \Phi_0 \Phi_0(x_j).
\]
Observe that by the Cauchy-Schwarz inequality, we have
\[
\sum_{l=K}^{\infty} e^{-\lambda_l t} \Phi_0 \Phi_0(x_j) \leq \sup_{x_j \in M} \left|\sum_{l=K}^{\infty} e^{-\lambda_l t} \Phi_0 \Phi_0(x_j)\right|.
\]
where $\tilde{\epsilon}_1$ depends on $d$, the Ricci curvature and the diameter of $M$. By Lemma 2

$$\sum_{x = M} e^{-\lambda t} \phi(x)^2 \leq \tilde{\epsilon}_1 t^{-\frac{d}{2}} \int_{xM} s^2 e^{-s} ds,$$

where $C_2$ depends on $d$, the Ricci curvature and the diameter of $M$. Note that the Gamma function with $z > 0$,

$$\int_z^\infty s^{\alpha} e^{-s} ds \leq \epsilon^{\alpha + 1} e^{-z},$$

when $z = \alpha + 1$. Hence, if $C_2 K^{\frac{d}{2}} t \geq \frac{d}{2} + 1$, i.e. $\frac{1 + \frac{d}{2}}{c_2} K^{\frac{d}{2}} t \leq t$, then

$$\sum_{l = K}^{\infty} e^{-\lambda l^d} \phi_l(x) \phi_l(y) \leq \sup_{x \in M} e^{-\lambda l^d} \phi_l(x)^2 \leq \tilde{\epsilon}_1 t^{-\frac{d}{2}} \int_{C_2 K^{\frac{d}{2}} t}^{\infty} s^2 e^{-s} ds$$

$$\leq \tilde{\epsilon}_1 t^{-\frac{d}{2}} (C_2 K^{\frac{d}{2}} t)^{\frac{d}{2} + 1} e^{-C_2 K^{\frac{d}{2}} t} = \tilde{\epsilon}_1 C_2^{\frac{d}{2} + 1} K^{\frac{d}{2} + 1} e^{-C_2 K^{\frac{d}{2}} t} \leq 2 \tilde{\epsilon}_1 C_2^{d} K^{\frac{d}{2} + 1} e^{-C_2 K^{\frac{d}{2}} t},$$

Next, we control the term with $l < K$. By the triangle inequality, we have

$$\left| \sum_{l = 0}^{K-1} (e^{-\mu_{n,e} t} v_{l,n,e}(i) v_{l,n,e}(j) - e^{-\lambda l^d} \phi_l(x_i) \phi_l(x_j)) \right|$$

$$\leq \sum_{l = 0}^{K-1} \left( e^{-\mu_{n,e} t} - e^{-\lambda l^d} \right) \left| \phi_l(x_i) \phi_l(x_j) \right| + \left| v_{l,n,e}(i) v_{l,n,e}(j) - \phi_l(x_i) \phi_l(x_j) \right| e^{-\mu_{n,e} t}.$$
Moreover, since $1 \leq l$,
\[|\phi_i(x_i)\phi_j(x_j)| \leq C_1^2 \lambda_i^{(d-1)/2} \leq C_1^2 \lambda_K^{(d-1)/2}.\]

Therefore,
\[
\sum_{l=0}^{K-1} e^{-\mu_{l,n,t}l} - e^{-\lambda_l^2} ||\phi_i(x_i)\phi_j(x_j)|| \leq \Omega_1 C_2^2 \lambda_K^{(d-1)/2} \frac{3}{3} e^2 t e^{-C_2 l}.\]

Next, we bound the term $|v_{l,n,e}(i)v_{l,n,e}(j) - \phi_i(x_i)\phi_j(x_j)|e^{-\mu_{l,n,t}l}$. Note that for $l \geq 1$
\[
|v_{l,n,e}(i)v_{l,n,e}(j) - \phi_i(x_i)\phi_j(x_j)|e^{-\mu_{l,n,t}l} \leq |v_{l,n,e}(i)v_{l,n,e}(j) - \phi_i(x_i)\phi_j(x_j)|e^{-\lambda_l^2 l} \leq |v_{l,n,e}(i)v_{l,n,e}(j) - \phi_i(x_i)\phi_j(x_j)|e^{-C_2 l}.
\]

Also, for $1 \leq l < K$,
\[
|v_{l,n,e}(i)v_{l,n,e}(j) - \phi_i(x_i)\phi_j(x_j)| \\
\leq \max_{x_i} |a_n v_{l,n,e}(i) - \phi_i(x_i)| (\max_{x_i} |\phi_i(x_i)| + \max_{x_i} |v_{l,n,e}(i)|) \\
\leq (2 \max_{x_i} |\phi_i(x_i)| + \Omega_2 e^{1/2}) \max_{x_i} |a_n v_{l,n,e}(i) - \phi_i(x_i)| \\
\leq (2C_1 \lambda_K^{d/4} + \Omega_2 e^{1/2}) \Omega_2 e^{1/2} \leq 2C_1 \lambda_K^{d/4} \Omega_2 e^{1/2} + \Omega_2 e,
\]

where the last two bounds come from Theorem 2.

When $l = 0$, similarly, by [85], we have
\[
|v_{0,n,e}(i)v_{0,n,e}(j) - \phi_0(x_i)\phi_0(x_j)||e^{-\mu_{0,n,t}l}| \\
= |v_{0,n,e}(i)v_{0,n,e}(j) - \phi_0(x_i)\phi_0(x_j)| \leq \left(\frac{2}{\sqrt{\text{Vol}(M)}} + \tilde{c}_2 \epsilon\right) \tilde{c}_2 \epsilon \leq \tilde{c}_3 \epsilon,
\]

for $\tilde{c}_3 > 0$ depending on the volume of $M$, and the $C^2$ norm of $p$. Hence,
\[
\sum_{l=0}^{K-1} \left(|v_{l,n,e}(i)v_{l,n,e}(j) - \phi_i(x_i)\phi_j(x_j)||e^{-\mu_{l,n,t}l}|\right) \leq K(2C_1 \lambda_K^{d/4} \Omega_2 e^{1/2} + \Omega_2 e) e^{-C_2 l} + \tilde{c}_3 \epsilon.
\]

If we sum the above terms together, we have
\[
|v_{K,(i)}^{(j)} - H(x_i,x_j,t)| \\
\leq \tilde{c}_1 \frac{C_2^2}{K} e^{-\frac{C_2 l}{K}} + K \left(\frac{2\Omega_1 C_2}{eC_2} K \lambda_K^{(d-1)/2} e^2 t e^{-\frac{C_2 l}{K}} + 2C_1 \lambda_K^{d+1} \Omega_2 e^{1/2} e^{-C_2 l} + \Omega_2 e e^{-C_2 l}\right) + \tilde{c}_3 \epsilon \\
\leq \left[\tilde{c}_1 \frac{C_2^{d/2}}{K} + \frac{2\Omega_1 C_2}{eC_2} K \lambda_K^{d-1/2} e^{1/2} + 2C_1 \lambda_K^{d+1} \Omega_2 e^{1/2} e^{-C_2 l} + \Omega_2 e e^{-C_2 l}\right] \\
\times (t + 1) e^{-\frac{C_2 l}{K}} + \tilde{c}_3 \epsilon.
\]

By Lemma 2 $\frac{\lambda_K}{C_2} \geq 1$. Hence, if $\epsilon < \frac{C_2^{(d-1)/2}}{K^4 \lambda_K^{(d-1)/2}} = \frac{\epsilon}{K^4 \lambda_K^{(d-1)/2}}$, where $\lambda_K$ depends on $d$, the Ricci curvature and the diameter of $M, K^2 \left(\frac{\lambda_K}{C_2}\right)^{d-1/2} e^{1/2}, K \left(\frac{\lambda_K}{C_2}\right)^{d+1} e^{1/2}$, and $\epsilon$ are all
bounded by $\frac{1}{K}$. In conclusion, we have

$$||H_{ij}^{(K)} - H(x_i, x_j, t)|| \leq \left( \tilde{c}_1 C_2^d + \frac{2\Omega_1 C_1^2}{e} C_2^{d-3} + 2C_1 C_2^{d-1} \Omega_2 + \Omega_2^2 \right) \frac{1}{K} (t + 1) e^{-\frac{\Omega_2}{2} t} + \tilde{c}_3 \epsilon$$

= $\frac{\Omega_3}{K} (t + 1) e^{-\Omega_4 t} + \Omega_5 \epsilon$,

where $\Omega_3 := \tilde{c}_1 C_2^d + \frac{2\Omega_1 C_1^2}{e} C_2^{d-3} + 2C_1 C_2^{d-1} \Omega_2 + \Omega_2^2$. Observe that $\Omega_3$ depends on $d$, $p_m$, the $C^2$ norm of $p$, the injectivity radius, the diameter, the volume and the curvature of $M$. $\Omega_4 = \frac{C_2}{2}$ depends on $d$, the Ricci curvature and the diameter of $M$. $\Omega_5 = \tilde{c}_3$ depends on the volume of $M$, and the $C^2$ norm of $p$.

7. DISCUSSION

The heat kernel approximation we propose in this paper relies on the series representation of the heat kernel in terms of the eigenpairs of the Laplace-Beltrami operator. Since the kernel normalized GL associated with a point cloud sampled from a manifold can recover the Laplace-Beltrami operator regardless of the distribution of the data points on the manifold, we propose to approximate the heat kernel based on the eigenpairs of the kernel normalized GL. To this end, we study the spectral convergence of the eigenvalues and eigenvectors of the kernel normalized GL to the eigenvalues and eigenfunctions of the Laplace-Beltrami operator, and provide the convergence rate. With this result, the proposed approximation to the heat kernel is quantified with a convergence rate.

As an alternative to the GL, one can potentially leverage on other algorithms that produce estimates of the Laplace-Beltrami operator in order to obtain different approximations to the heat kernel. As in practice the data can have arbitrary non-uniform distributions over unknown manifolds, it is important for the Laplace-Beltrami operator to be efficiently and consistently estimated regardless of the distribution of the data. It has been shown that the well-known unsupervised learning algorithm, Locally Linear Embeddings (LLEs), provides another choice in this regard. Indeed, it has been shown in [36] that if the regularization is properly chosen, the LLE asymptotically approximates the Laplace-Beltrami operator even if the sampling scheme is non-uniform. In other words, the KDE step commonly applied in machine learning to handle the non-uniform sampling is implicitly carried out in the LLE.

The methods applied in this work serve for the convergence rate of the eigenvalues and the $L^\infty$ convergence rate of the eigenvectors of the kernel normalized GL to the eigenvalues and eigenfunctions of the Laplace-Beltrami operator, even if the sampling on the manifold is non-uniform. In order to relate the discrete eigenvectors to the continuous eigenfunctions, an interpolation is necessary (e.g. we introduce the operator $T_{n,\epsilon}$ in the definition 5). Since we do not know any relationship between the position of the sampling and the geometry of the manifold, the analysis of the interpolation in Lemma 10 is based on the generic case. This step seems to be the main bottleneck of improving the convergence rate in our analysis. While we provide a convergence rate, it does not fully match the numerical findings in practice. In [31,5], the authors claim a sharper eigenvalue convergence rate and an eigenvector convergence rate in the $L^2$ sense when the 1-normalization step (kernel density estimation) is not applied. Besides an independent application of a similar eigenvalue control idea in Lemma 5 to improve the estimate, the techniques applied in [5] are completely different from ours. Since the asymptotical operator considered in [5] is
not the Laplace-Beltrami operator when the sampling is non-uniform, and obtaining the optimal convergence rate is out of the scope of this paper, we refer readers with interest to [5] for details. We will explore whether their approaches can be developed to obtain a sharper $L^\infty$ convergence rate in our future work.

There are many other interesting directions motivated by this work. The most important direction is its practical application in statistics. For example, we can develop an algorithm to handle the Gaussian process regression problem on an unknown embedded manifold in a Euclidean space by using the heat kernel of the manifold as the covariance kernel in the Gaussian process. It is not straightforward to define a valid covariance kernel for the Gaussian process regression when the predictors are on a nonlinear manifold [7, 20], which might have non-trivial topology. Although the heat kernel provides a natural and canonical choice theoretically, it is analytically intractable to directly evaluate, and current Bayesian methods rely on computationally demanding Monte Carlo approximations [24]. With an efficient and accurate approximation of the heat kernel relying on the diffusion property of the GL, we may have a diffusion-based Gaussian process modeling framework for the regression purpose. This development has immediate applications to various fields, including biomedical signals and ecological datasets. Next, it is interesting to develop broader classes of covariance kernels on unknown manifolds beyond the heat kernel; for example, including anisotropy and non-stationarity inspired by the background knowledge, so that the regression can be carried out in a more adaptive way. Additional important directions include approximating the heat kernel under when there exist unbounded measurement errors and more intricate structure in the data, such as disconnected manifolds. Indeed, while we provide a preliminary analysis of the error-in-variable result, the analysis might be further improved with more suitable analysis tools. Another natural question to ask is if we could achieve a similar result for the connection Laplacian associated with a principal bundle via the graph connection Laplacian [29, 30], so that more topological information could be obtained.

**ACKNOWLEDGEMENT**

Hau-Tieng Wu thanks the valuable discussion with Professor Jeff Calder and Professor Nicolas Garcia Trillos. The authors would like to thank the anonymous reviewers for their constructive comments that helped improve this paper.

**REFERENCES**

[1] Kendall E Atkinson. The numerical solution of the eigenvalue problem for compact integral operators. *Transactions of the American Mathematical Society*, 129(3):458–465, 1967.

[2] Mikhail Belkin and Partha Niyogi. Convergence of Laplacian eigenmaps. In *Advances in Neural Information Processing Systems*, pages 129–136, 2007.

[3] Pierre Bérard, Gérard Besson, and Sylvain Gallot. Embedding Riemannian manifolds by their heat kernel. *Geometric & Functional Analysis GAFA*, 4(4):373–398, 1994.

[4] Dmitri Burago, Sergei Ivanov, and Yaroslav Kurylev. A graph discretization of the Laplace-Beltrami operator. *Journal of Spectral Theory*, 4(4):675–714, 2015.

[5] Jeff Calder and Nicolas Garcia Trillos. Improved spectral convergence rates for graph Laplacians on epsilon-graphs and k-nn graphs. *arXiv preprint arXiv:1910.13476*, 2019.

[6] Jeff Calder, Nicolas Garcia Trillos, and Marta Lewicka. Lipschitz regularity of graph Laplacians on random data clouds. *arXiv preprint arXiv:2007.06679*, 2020.

[7] Ismaël Castillo, Gérard Kerkyacharian, and Dominique Picard. Thomas Bayes’ walk on manifolds. *Probability Theory and Related Fields*, 158(3-4):665–710, 2014.

[8] Françoise Chatelin. *Spectral approximation of linear operators*. SIAM, 2011.

[9] Fan Chung. *Spectral Graph Theory*. American Mathematical Society, 1996.
Lemma 19. Suppose \((\lambda_k, \phi_k)\) is the \(k\)-th eigenpair of \(-\Delta\) with \(\|\phi_k\|_2 = 1\) and \((\lambda_{k,\epsilon}, \phi_{k,\epsilon})\) the \(k\)-th eigenpair of \(\frac{I - \epsilon D}{\epsilon^2}\) with \(\|\phi_{k,\epsilon}\|_2 = 1\), where we assume \(\epsilon > 0\) is small enough. Take \(K \geq k\) so that \(\lambda_k \leq \lambda_K\) and assume that \(\lambda_K\) and \(\epsilon\) satisfy (6") in Table 5.2. Then, when
For the term \( \epsilon > 0 \) is sufficiently small,

\[
\left\| \frac{T_\epsilon - I}{\epsilon^2} \phi_{k,\epsilon} - \Delta \phi_{k,\epsilon} \right\|_2 \leq 2 \left( 1 + C_1(\lambda_K) + (C'_0 + 2\lambda_K)C_2(\lambda_K) + 4\lambda_K^2 \right) \epsilon^2,
\]

where \( C_1(\lambda_K) > 0 \) and \( C_2(\lambda_K) > 0 \) are constants depending on \( \lambda_K \) but independent of \( \epsilon \) and \( C'_0 > 0 \) depends on \( d \) and the volume, the curvature and the diameter of \( M \).

**Proof.** First of all, we have the \( L^2 \) convergence \( |\lambda_k - \lambda_{k,\epsilon}| \leq C_1(\lambda_K)\epsilon^2 \) and \( \|\phi_k - a\phi_{k,\epsilon}\|_2 \leq C_2(\lambda_K)\epsilon^2 \), where \( a \in \{1, -1\} \), according to (44) and (45). Recall that \( H_\epsilon \) is the integral operator associated with the heat kernel on the manifold \( M \) with the diffusion time \( \epsilon^2 \), in other words, we have

\[
H_\epsilon f(x) := \int_M H(x, y, \epsilon^2) f(y) dV_M(y).
\]

Thus, when \( \epsilon > 0 \) is sufficiently small, a trivial bound leads to

\[
\left\| \frac{T_\epsilon - I}{\epsilon^2} \phi_{k,\epsilon} - \Delta \phi_{k,\epsilon} \right\|_2 \leq -\lambda_k \phi_{k,\epsilon} + \frac{\Delta H_\epsilon \phi_{k,\epsilon}}{1 - \epsilon^2 \lambda_k} + \frac{\Delta H_\epsilon \phi_{k,\epsilon}}{1 - \epsilon^2 \lambda_k} - \frac{\Delta T_\epsilon \phi_{k,\epsilon}}{1 - \epsilon^2 \lambda_k} \leq -\lambda_{k,\epsilon} (1 - \epsilon^2 \lambda_{k,\epsilon}) \phi_{k,\epsilon} - \Delta H_\epsilon \phi_{k,\epsilon} \leq 2 \left\| \frac{T_\epsilon - I}{\epsilon^2} \phi_{k,\epsilon} - \Delta H_\epsilon \phi_{k,\epsilon} \right\|_2 + 2 \left\| \frac{T_\epsilon - I}{\epsilon^2} \phi_{k,\epsilon} - \Delta H_\epsilon \phi_{k,\epsilon} \right\|_2.
\]

We first bound the term \( \|\Delta T_\epsilon \phi_{k,\epsilon} - \Delta H_\epsilon \phi_{k,\epsilon}\|_2 \) in (95). Recall that \( \{\phi_i\}_{i=0}^\infty \) form an orthonormal basis of \( L^2(M) \). We thus have

\[
\phi_{k,\epsilon} = \sum_{i=0}^\infty b_i \phi_i =: b_k \phi_k + \mathbf{f},
\]

where since \( \|\phi_k - \phi_{k,\epsilon}\|_2 \leq C_2(\lambda_K)\epsilon^2 \), we have \( |b_k| \leq 1 \) and \( \|\mathbf{f}\|_2 \leq C_2(\lambda_K)\epsilon^2 \) by the triangle inequality. Therefore,

\[
\|\Delta T_\epsilon \phi_k - \Delta H_\epsilon \phi_k\|_2 \leq \|\Delta T_\epsilon \mathbf{f}\|_2 + \|\Delta H_\epsilon \mathbf{f}\|_2.
\]

By Lemma 8 and integration by parts,

\[
\|\Delta T_\epsilon \phi_k - \Delta H_\epsilon \phi_k\|_2 = \lambda_k \|T_\epsilon \phi_k - H_\epsilon \phi_k\|_2 \leq C_4 \epsilon^4 \lambda_k^d (1 + \lambda_k^{d/2+5}).
\]

For the term \( \|\Delta T_\epsilon \mathbf{f} - \Delta H_\epsilon \mathbf{f}\|_2 \), we need a further exploration of \( T_\epsilon \) and hence a comparison with \( H_\epsilon \). Recall that \( k_\epsilon(x, y) = \exp \left( -\frac{||x-y||^2}{4\epsilon^2} \right) \), and based on the definition of \( T_\epsilon \), we have

\[
T_\epsilon f(x) = \int_M \frac{k_\epsilon(x, y) f(y) p(y) dV_M(y)}{\int_M k_\epsilon(x, y) p(y) dV_M(y)} = \frac{\int_M k_\epsilon(x, y) f(y) dV_M(y)}{\int_M k_\epsilon(x, y) p(y) dV_M(y)}.
\]

where

\[
\mathcal{K}_\epsilon(x, y) := \frac{\int_M k_\epsilon(x, y) p(y) dV_M(y)}{\int_M k_\epsilon(x, y) p(y) dV_M(y)}.
\]
We now compare $K_{e}(x, y)$ with $H(x, y, t)$ when $t = e^{2}$. By Lemma 3, we have the following facts about $H(x, y, t)$ and its space and time derivatives. There is $γ > 0$ and $B_{γ}(x)$ is the geodesic ball in $M$ around $x$ with the radius $γ$, so that when $ε > 0$ is small enough, we have

\begin{align}
H(x, y, 2e^{2}) & \geq C_{1}′e^{-d}e^{-\frac{d(x, y)^{2}}{2e^{2}}} & \text{if } y \in B_{γ}(x), \\
|\nabla H(x, y, ε^{2})| & \leq C_{H}e^{-d-1}e^{-\frac{d(x, y)^{2}}{2e^{2}}} & \text{if } y \in B_{γ}(x), \\
|\Delta_{H}H(x, y, ε^{2})| & = |\partial_{i}H(x, y, ε^{2})| \leq C_{H}e^{-d-2}e^{-\frac{d(x, y)^{2}}{2e^{2}}} & \text{if } y \in B_{γ}(x), \\
|\Delta_{H}H(x, y, ε^{2})| & = |\partial_{i}H(x, y, ε^{2})| \leq C_{H}ε^{2} & \text{if } y \in B_{γ}(x),
\end{align}

where $C_{1}′ > 0$ depends on $d$ and the Ricci curvature of $M$ and $C_{H} > 0$ depends on $d$, the volume, the Ricci curvature and the diameter of $M$. In other words,

\begin{align}
|\nabla H(x, y, ε^{2})| & \leq \frac{C_{H}}{C_{1}′}H(x, y, 2ε^{2}) & \text{if } y \in B_{γ}(x), \\
|\Delta_{H}H(x, y, ε^{2})| & \leq \frac{C_{H}}{C_{1}′}H(x, y, 2ε^{2}) & \text{if } y \in B_{γ}(x), \\
|\Delta_{H}(x, y, ε^{2})d(x, y)^{2}| & \leq \frac{8C_{H}}{C_{1}′}H(x, y, 4ε^{2}) & \text{if } y \in B_{γ}(x).
\end{align}

By a straightforward expansion of $K_{e}(x, y)$ as in Lemma 7 and Lemma 8 in [10] and by Lemma 3, we have the following facts. There is $γ > 0$ so that when $ε > 0$ is small enough, we have

\begin{align}
K_{e}(x, y) = H(x, y, ε^{2})(w_{1}(x, y) + w_{2}(x, y, ε^{2})ε^{2}) & \text{ if } y \in B_{γ}(x), \\
|\Delta_{e}K_{e}(x, y)| & \leq C_{2}′ε^{2} & \text{if } y \in B_{γ}(x),
\end{align}

where $C_{2}′ > 0$ depends on $d$ and the curvature of the manifold and $w$ is a smooth function of $x, y$ satisfying

\begin{align}
|w_{1}(x, y) - 1| & \leq C_{3}′d(x, y)^{2} & \text{for all } x, y \in B_{γ}(x), \\
|w_{2}(x, y, ε^{2})| & \leq C_{4}′ & \text{for all } x, y \in B_{γ}(x), \\
\|w_{2}(x, y, ε^{2})\|_{C^{2}(B_{γ}(x))} & \leq C_{5}′ & \text{for all } x, y \in B_{γ}(x),
\end{align}

where $C_{3}′, C_{4}′, C_{5}′ > 0$ depend on $d$ and the curvature of the manifold and $C_{5}′ > 0$ depends on $d$ and the second order covariant derivatives of the curvature tensor. In order to bound $\|\Delta_{t}f - ∆H_{t}f\|_{2}$, we first bound $|\Delta_{t}f(x) - ∆H_{t}f(x)|$. By a straightforward bound,

\[
\left|\int_{M}\Delta_{e}K_{e}(x, y)f(y)dV_{M}(y) - \int_{M}\Delta_{H}(x, y, ε^{2})f(y)dV_{M}(y)\right|
\leq \left|\int_{B_{γ}(x)}[\Delta_{e}K_{e}(x, y) - \Delta_{H}(x, y, ε^{2})]f(y)dV_{M}(y)\right| + \int_{B_{γ}(x)}|\Delta_{e}K_{e}(x, y)|f(y)dV_{M}(y)
\]

\[
+ \left|\int_{B_{γ}(x)}\Delta_{H}(x, y, ε^{2})f(y)dV_{M}(y)\right|,
\]

where

\[
\int_{B_{γ}(x)}[\Delta_{e}K_{e}(x, y)]f(y)dV_{M}(y) + \int_{B_{γ}(x)}\Delta_{H}(x, y, ε^{2})f(y)dV_{M}(y)
\leq (C_{2}′ + C_{H})\Vol(M)ε^{2}\|f\|_{2} \leq \Vol(M)\|f\|_{2}
\]
when \( \epsilon \) is sufficiently small due to (102) and (107). It remains to bound the main term over \( B_r(x) \). By (106), we have

\[
\left| \int_{B_r(x)} \left[ \Delta_x \mathbb{K}_r(x, y) - \Delta_x H(x, y, \epsilon^2) \right] f(y) dV_M(y) \right|
\]

\[
= \left| \int_{B_r(x)} \Delta_x H(x, y, \epsilon^2) \left[ w_1(x, y) - 1 \right] f(y) dV_M(y) \right| + \left| \int_{B_r(x)} \Delta_x \left[ H(x, y, \epsilon^2) w_2(x, y, \epsilon^2) \epsilon^2 \right] f(y) dV_M(y) \right|.
\]

For the first term on the right hand side,

\[
\left| \int_{B_r(x)} \Delta_x H(x, y, \epsilon^2) \left[ w_1(x, y) - 1 \right] f(y) dV_M(y) \right|
\]

\[
\leq \int_{B_r(x)} |\Delta_x H(x, y, \epsilon^2)| |w_1(x, y) - 1||f(y)| dV_M(y) \leq \int_{B_r(x)} C_3 |\Delta_x H(x, y, \epsilon^2)| d(x, y)^2 |f(y)| dV_M(y)
\]

\[
\leq \int_{B_r(x)} \frac{8C_H C_4^2}{C_1} H(x, y, 4\epsilon^2) |f(y)| dV_M(y)
\]

\[
\leq \frac{8C_H C_4^2}{C_1} H(x, y, 4\epsilon^2) |f(y)| dV_M(y).
\]

Note that we use (108) and (105) in the third and fourth step. Again by a direct expansion,

\[
\left| \int_{B_r(x)} \Delta_x \left[ H(x, y, \epsilon^2) w_2(x, y, \epsilon^2) \epsilon^2 \right] f(y) dV_M(y) \right|
\]

is bounded by

\[
\epsilon^2 \int_{B_r(x)} |\Delta_x H(x, y, \epsilon^2)| w_2(x, y, \epsilon^2) |f(y)| dV_M(y)
\]

\[
+ 2\epsilon^2 \int_{B_r(x)} |\nabla_x H(x, y, \epsilon^2)||\nabla_x w_2(x, y, \epsilon^2)||f(y)| dV_M(y)
\]

\[
+ \epsilon^2 \int_{B_r(x)} H(x, y, \epsilon^2) |\Delta_x w_2(x, y, \epsilon^2)||f(y)| dV_M(y)
\]

\[
\leq \frac{C_H C_4^2}{C_1} \int_{B_r(x)} H(x, y, 2\epsilon^2) |f(y)| dV_M(y) + 2\epsilon \frac{C_H C_4^2}{C_1} \int_{B_r(x)} H(x, y, 2\epsilon^2) |f(y)| dV_M(y)
\]

\[
+ \epsilon^2 C_5 \int_{B_r(x)} H(x, y, \epsilon^2) |f(y)| dV_M(y)
\]

\[
\leq \left( \frac{C_H C_4^2}{C_1} + 1 \right) \int_{B_r(x)} H(x, y, 2\epsilon^2) |f(y)| dV_M(y) + \int_{B_r(x)} H(x, y, \epsilon^2) |f(y)| dV_M(y).
\]

Note that we use (103), (102), (109) and (110) in the first bound, and the second bound holds when \( \epsilon \) is sufficiently small. Denote

\[
u(x, t) = \int_M H(x, y, t) |f(y)| dV_M(y).\]

Obviously we have

\[
\int_{B_r(x)} H(x, y, \epsilon^2) |f(y)| dV_M(y) \leq u(x, 2\epsilon^2) \quad \text{and} \quad \int_{B_r(x)} H(x, y, \epsilon^2) |f(y)| dV_M(y) \leq u(x, \epsilon^2).
\]

Hence, we have the bound

\[
\left| \int_M \Delta_x \mathbb{K}_r(x, y) f(y) dV_M(y) - \int_M \Delta_x H(x, y, \epsilon^2) f(y) dV_M(y) \right|
\]

\[
\leq \frac{8C_H C_4^2}{C_1} u(x, 4\epsilon^2) + \left( \frac{C_H C_4^2}{C_1} + 1 \right) u(x, 2\epsilon^2) + u(x, \epsilon^2) + \sqrt{\text{Vol}(M)} \|f\|_2.
\]
Finally, since \( u(x,t) \) is the solution to the heat equation with initial condition \(|f|\), we have 
\[ \|u(x,t)\|_2 \leq \|f\|_2 \text{ for any } t \geq 0. \]
Hence, we have shown that
\[ \|\Delta T_e f - \Delta \Phi_e \|_2 \leq \left( \frac{8CH C_i'}{C_i'} + C_i' + 2 + \text{Vol}(M) \right) \|f\|_2. \]  
(111)

Substitute (97) and (111) into (96), we have
\[ \|\Delta T_e \Phi_{k,e} - \Delta \Phi_e \Phi_{k,e}\|_2 \leq C_4 e^4 \lambda_k \left( 1 + \lambda_k^{d/2 + 2} \right) \left( \frac{8CH C_i'}{C_i'} + C_i' + 2 + \text{Vol}(M) \right) \|f\|_2. \]  
(112)

where we use relation (6") in table 5.2 in the last step.

Next, we bound \( -\lambda_k e \| \Phi_{k,e} - \Delta \Phi_e \Phi_{k,e} \|_2 \) in (95). Recall that \( \Phi_{k,e} = \sum_{t=0}^{\infty} b_t \phi_t = b_k \phi_k + f \). Hence,
\[ \| - \lambda_k e (1 - e^2 \lambda_k) \Phi_{k,e} - \Delta \Phi_e \|_2 \leq \| - \lambda_k e b_k \phi_k - \Delta \Phi_e \|_2 + e^2 \lambda_k^2 + C_4 \lambda_k^2 \|f\|_2. \]  
By using the similar method as above, we can show that
\[ \| - \lambda_k e b_k \phi_k - \Delta \Phi_e \|_2 \leq 2 e^2 \lambda_k^2 + C_4 \lambda_k^2 \|f\|_2. \]  
Hence,
\[ \| - \lambda_k e (1 - e^2 \lambda_k) \Phi_{k,e} - \Delta \Phi_e \|_2 \leq 2 e^2 \lambda_k^2 + C_4 \lambda_k^2 \|f\|_2. \]  
(113)

Finally, if we substitute (112) and (113) into (95), we have
\[ \left\| \frac{T_e - I}{e^2} \Phi_{k,e} - \Delta \Phi_{k,e} \right\|_2 \leq 2 \left( 1 + C_4 \lambda_k^2 \lambda_k^2 + C_4 \lambda_k^2 \lambda_k^2 + C_4 \lambda_k^2 \lambda_k^2 \right) \|f\|_2. \]  
The statement of the lemma follows.
\[ \square \]

Appendix B. Proof of Proposition 2

Let \( \mathcal{F} \subset C(M) \). The covering number \( \mathcal{N} (\mathcal{F}, \gamma, \| \cdot \|_N) \) is the smallest number of balls with radius \( \gamma \) in the norm \( \| \cdot \|_N \) that cover \( \mathcal{F} \). In this proof, we consider three spaces, \( \mathcal{X}_e \), \( u \mathcal{D}_e \) and \( \mathcal{L}_e \mathcal{D}_e \). The norms that we consider are \( \| \cdot \|_N = \| \cdot \|_\infty \) and \( \| \cdot \|_N = \| \cdot \|_L^2(p_n) \).

Observe that the cover number satisfies the basic property: if \( \gamma_i \leq \gamma_2 \), then
\[ \mathcal{N} (\mathcal{F}, \gamma_1, \| \cdot \|_N) \geq \mathcal{N} (\mathcal{F}, \gamma_2, \| \cdot \|_N). \]

Recall that for the point cloud \( \mathcal{X} := \{ x_i \}_{i=1}^n \) i.i.d. sampled from the random vector with the density function \( p \) supported on the manifold \( M \), for \( f \in C(M) \), we have
\[ \|f\|_{L^2(p_n)} = \left( \frac{1}{n} \sum_{i=1}^{n} f(x_i)^2 \right)^{1/2}. \]
Clearly we have for \( f \in C(M) \), \( \|f\|_{L^2(p_n)} \leq \|f\|_\infty \). Hence,
\[ \mathcal{N} (\mathcal{F}, \gamma, \| \cdot \|_{L^2(p_n)}) \leq \mathcal{N} (\mathcal{F}, \gamma, \| \cdot \|_\infty). \]  
(114)
In the next Lemma, we bound the covering number \( \mathcal{N}(\mathcal{K}_E, \gamma, \| \cdot \|_\infty) \). Then, we will use \( \mathcal{N}(\mathcal{K}_E, \gamma, \| \cdot \|_\infty) \) to bound the covering corresponding number for the spaces \( \mathcal{K}_E, u\mathcal{D}_E \) and \( \mathcal{D}_E \).

**Lemma 20.** For \( \gamma > 0 \) and \( \varepsilon > 0 \) sufficiently small, we have

\[
\mathcal{N}(\mathcal{K}_E, \gamma, \| \cdot \|_\infty) \leq \left( \frac{4 \text{diam}(M)}{\varepsilon^2} \right)^{\left( \frac{d(3d+11)}{4} \right)} .
\]

**Proof.** Recall that \( M \) is isometrically embedded in \( \mathbb{R}^D \) through \( t \). By the Nash embedding theorem [23], there is an isometric embedding \( E : M \to \mathbb{R}^{s'} \), where \( s' \leq s := \frac{d(3d+11)}{2} \leq D \).

We claim that, for any \( \varepsilon_1 > 0 \), the covering number \( \mathcal{N}(E(M), \varepsilon_1, \| \cdot \|_{\mathbb{R}^s}) \) is bounded by \( \left( \frac{4 \text{diam}(M)}{\varepsilon_1} \right)^{s} \). Choose any point \( p \in E(M) \). Clearly, \( E(M) \) is contained in the \( s \) dimensional cube in \( \mathbb{R}^{s'} \) centered at \( p \) with the side length \( 2 \text{diam}(M) \). The cube can be covered by \( \left( \frac{4 \text{diam}(M)}{\varepsilon_1} \right)^{s} \) Euclidean balls of radius \( \frac{\text{diam}(M)}{2} \), denoted as \( \{ B_{\varepsilon_1/2}(z_i) \} \), where \( z_i \in \mathbb{R}^s \). Note that \( \{ z_i \} \) may not be on \( E(M) \).

Next, we apply the following argument. If \( B_{\varepsilon_1/2}(z_i) \) intersects \( E(M) \), we could take \( z'_i \in B_{\varepsilon_1/2}(z_i) \cap E(M) \), where \( B_{\varepsilon_1/2}(z_i) \subset B_{\varepsilon_1}(z'_i) \). Otherwise, we let \( z'_i = z_i \). Obviously, we have

\[
E(M) \subseteq \bigcup_{\varepsilon_1/2(z'_i) \cap E(M) \neq \emptyset} B_{\varepsilon_1/2}(z'_i) \subseteq \bigcup_{\varepsilon_1/2(z'_i) \cap E(M) \neq \emptyset} B_{\varepsilon_1}(z'_i).
\]

Then, we claim that the number of \( z'_i \) in \( E(M) \) such that \( B_{\varepsilon_1/2}(z'_i) \cap E(M) \neq \emptyset \) is bounded by the total number of \( \{ z_i \} \), which is \( \left( \frac{4 \text{diam}(M)}{\varepsilon_1} \right)^{s} \). Indeed, for any \( x, y \in M \) so that \( \| E(x) - E(y) \|_{\mathbb{R}^s} \) is small enough, it is well known that

\[
\| E(x) - E(y) \|_{\mathbb{R}^s} \leq d(x, y) \leq 2 \| E(x) - E(y) \|_{\mathbb{R}^s}.
\]

See, for example, [36, Lemma B.3]. Hence, if \( \varepsilon_1 \) is small enough and \( \{ B_{\varepsilon_1}(z'_i) \} \) with \( z'_i \in E(M) \) is a cover of \( E(M) \), then the above inequality implies that the balls \( \{ B_{\varepsilon_1}(w_i) \} \) with \( w_i = t(E^{-1}(z'_i)) \in \mathbb{R}^D \) cover \( t(M) \). Hence, when \( \varepsilon_1 \) is small enough, the covering number \( \mathcal{N}(t(M), \varepsilon_1, \| \cdot \|_{\mathbb{R}^D}) \) is bounded by \( \left( \frac{4 \text{diam}(M)}{\varepsilon_1} \right)^{s} \).

For \( \gamma > 0 \), fix \( \varepsilon \) small enough so that \( \varepsilon_1 = 2\varepsilon \gamma \) is sufficiently small; in other words, \( \gamma = \frac{\varepsilon_1}{2\varepsilon} \). For \( x, x_i \in M \), if \( \| E(x) - E(x_i) \|_{\mathbb{R}^s} \leq \varepsilon_1 \), by the mean value theorem,

\[
\max_y | k_{\varepsilon}(x, y) - k_{\varepsilon}(x_i, y) | \leq \frac{\varepsilon_1}{2\varepsilon^2}.
\]

Hence, if the Euclidean balls centered at \( \{ w_i \} \subset E(M) \) of radius \( \varepsilon_1 = 2\varepsilon \gamma \) cover \( E(M) \), then the \( L^m \) balls centered at \( k_{\varepsilon}(t^{-1}(w_i), \cdot) \) of radius \( \gamma \) can cover the space \( K_{\varepsilon} \). We substitute \( \varepsilon_1 = 2\varepsilon \gamma \) into the bound of \( \mathcal{N}(t(M), \varepsilon_1, \| \cdot \|_{\mathbb{R}^D}) \), then we prove the lemma. \( \Box \)

**Lemma 21.** Take \( u \in C(M) \) so that \( \| u \|_{\infty} \leq 1 \). There is a constant \( \tilde{C}_1 > 0 \) depending on \( p_m \) and the \( C^0 \) norm of \( p \) such that if \( \varepsilon \) is small enough, for all \( f \in \tilde{C}_1 \varepsilon^{2d} u \mathcal{D}_E \) and \( f \in \tilde{C}_1 \varepsilon^{4d} \mathcal{D}_E \), we have \( \| f \|_{\infty} \leq 1. \) Moreover,

\[
\mathcal{N}(\tilde{C}_1 \varepsilon^{2d} u \mathcal{D}_E, \gamma, \| \cdot \|_\infty) \leq \left( \frac{\tilde{C}_2}{\varepsilon^{d+2} \gamma} \right)^{\frac{d(3d+11)}{2}},
\]

\[
\mathcal{N}(\tilde{C}_1 \varepsilon^{4d} \mathcal{D}_E, \gamma, \| \cdot \|_\infty) \leq \left( \frac{\tilde{C}_2}{\varepsilon^{d+2} \gamma} \right)^{\frac{d(3d+11)}{2}}.
\]
where $C_2 > 0$ is a constant depending on $d$, the diameter of $M$, $p_m$ and the $C^0$ norm of $p$.

**Proof.** First, it follows from (a) in Lemma 6 that if $\varepsilon$ is small enough, for any $x, y \in M$, $|Q_\varepsilon(x, y)| \leq \frac{C}{\varepsilon^2}$, where $C$ depends on $p_m$ and the $C^0$ norm of $p$. Take $\tilde{C}_1 = \min \{\frac{1}{\varepsilon^2}, \varepsilon^2\}$, then for all $f \in \tilde{C}_1 \varepsilon^{2d} u \mathcal{D}_\varepsilon$ and $f \in \tilde{C}_1 \varepsilon^{d} \mathcal{D}_\varepsilon \mathcal{D}_\varepsilon$, we have $\|f\|_\infty \leq 1$.

By Lemma 2 for any $x \in M$, $C_5 \varepsilon^d \leq d_\varepsilon(x) \leq C_6 \varepsilon^d$, where $C_5$ and $C_6$ depend on $p_m$ and the $C^0$ norm of $p$. Moreover,

$$
|d_\varepsilon(x) - d_\varepsilon(y)| = \left| \int_M (k_\varepsilon(x, z') - k_\varepsilon(y, z') dP(z') \right| \leq \|k_\varepsilon(x, \cdot) - k_\varepsilon(y, \cdot)\|_\infty.
$$

Hence,

$$
\|\tilde{C}_1 \varepsilon^{2d} u(z) Q_\varepsilon(x, z) - \tilde{C}_1 \varepsilon^{2d} u(z) Q_\varepsilon(y, z)\| = \frac{2\tilde{C}_1}{C_2^3 \varepsilon^{2d}} \|k_\varepsilon(x, \cdot) - k_\varepsilon(y, \cdot)\|_\infty.
$$

Since $\mathcal{N}(K_\varepsilon, \gamma, \| \cdot \|_\infty) \leq \left( \frac{4\text{diam}(M)}{e^2 \gamma} \right)^{\frac{d(3d+11)}{2}}$, we have $\mathcal{N}(u \mathcal{D}_\varepsilon, \gamma, \| \cdot \|_\infty) \leq \left( \frac{8\text{diam}(M)}{C_5^2 \varepsilon^{d+2} \gamma} \right)^{\frac{d(3d+11)}{2}}$.

Note that $\frac{8\text{diam}(M)}{C_5} \varepsilon^d$ depends on $d$, the diameter of $M$, $p_m$ and the $C^0$ norm of $p$. Similarly, we can prove the bound for $\mathcal{N}(\tilde{C}_1 \varepsilon^{d} \mathcal{D}_\varepsilon \mathcal{D}_\varepsilon, \gamma, \| \cdot \|_\infty)$.

Now we are ready to prove Proposition 2.

**Proof.** (Proof of Proposition 2)

Let $\mathcal{F} \subset C(M)$ and $\|f\|_\infty \leq 1$ for all $f \in \mathcal{F}$. Based on the entropy bound in Proposition 19 in [33], there is a constant $c_1 > 0$ such that for all $n$, with probability greater than $1 - \delta$,

$$
\sup_{f \in \mathcal{F}} |P_n f - Pf| \leq \frac{c_1}{n^{1/2}} \int_0^{\infty} \sqrt{\log \mathcal{N}(\mathcal{F}, \gamma, \| \cdot \|_{L^2(P_n)})} d\gamma + \sqrt{\frac{\log 2}{2n}} \\frac{\log \frac{2}{\delta}}{\delta}.
$$

Since $\mathcal{N}(\mathcal{F}, \gamma, \| \cdot \|_{L^2(P_n)}) \leq \mathcal{N}(\mathcal{F}, \gamma, \| \cdot \|_\infty)$,

$$
\sup_{f \in \mathcal{F}} |P_n f - Pf| \leq \frac{c_1}{n^{1/2}} \int_0^{\infty} \sqrt{\log \mathcal{N}(\mathcal{F}, \gamma, \| \cdot \|_\infty)} d\gamma + \sqrt{\frac{\log 2}{2n}} \\frac{\log \frac{2}{\delta}}{\delta}.
$$

where we use the fact that the diameter of the set $\mathcal{F}$ is bounded by $2$ in the last step. Note that by Lemma 21 for any $f \in \mathcal{F}_\varepsilon \cup \tilde{C}_1 \varepsilon^{2d} u \mathcal{D}_\varepsilon \cup \tilde{C}_1 \varepsilon^{d} \mathcal{D}_\varepsilon \mathcal{D}_\varepsilon$, we have $\|f\|_\infty \leq 1$.

We first discuss the case $\mathcal{F} = K_\varepsilon$. By Lemma 20 $\log \mathcal{N}(K_\varepsilon, \gamma, \| \cdot \|_\infty) = \frac{d(3d+11)}{2} \log \left( \frac{4\text{diam}(M)}{e^2 \gamma} \right)$.

Let $\gamma' = \frac{\varepsilon^2}{4\text{diam}(M)}$. If $\gamma \leq \gamma'$, then

$$
\log \mathcal{N}(K_\varepsilon, \gamma, \| \cdot \|_\infty) \leq \frac{d(3d+11)}{2} \log \left( \frac{4\text{diam}(M)}{e^2 \gamma} \right) = \frac{d(3d+11)}{2} \log \left( \frac{1}{\gamma \gamma'} \right) \leq d(3d+11) \log \left( \frac{1}{\gamma} \right) \leq d(3d+11) \left( \log \left( \frac{1}{\gamma} \right) \right)^2.
$$
If $\gamma > \gamma'$,
\[
\log \mathcal{N}(\mathcal{H}_E, \gamma, \cdot ||_\infty) \leq \log \mathcal{N}(\mathcal{H}_E, \gamma', \cdot ||_\infty) = 2d(3d + 11) \log \left( \frac{2\sqrt{\text{diam}(M)}}{\varepsilon} \right).
\]
Hence,
\[
\int_0^2 \sqrt{\log \mathcal{N}(K_E, \gamma, \cdot ||_\infty)} d\gamma
\leq \int_0^\gamma \sqrt{d(3d + 11)} \log \left( \frac{1}{\gamma'} \right) d\gamma + (2 - \gamma') \sqrt{2d(3d + 11)} \sqrt{\log \left( \frac{2\sqrt{\text{diam}(M)}}{\varepsilon} \right)}
\]
\leq \sqrt{d(3d + 11)} \gamma' \log \left( \frac{1}{\gamma'} \right) + \sqrt{2d(3d + 11)} \sqrt{\log \left( \frac{2\sqrt{\text{diam}(M)}}{\varepsilon} \right)}
\leq \sqrt{d(3d + 11)} + 2\sqrt{2d(3d + 11)} \sqrt{\log \left( \frac{2\sqrt{\text{diam}(M)}}{\varepsilon} \right)} + 2\sqrt{2d(3d + 11)} \sqrt{\log \left( \frac{1}{\varepsilon} \right)}
\leq c_2 \sqrt{-\log \varepsilon},
\]
where we use the trivial bound $\gamma' \log(\frac{1}{\gamma'}) < 1$ since $\gamma'$ is of order $\varepsilon^2$ and the fact that $\varepsilon$ is small enough in the last step, and $c_2 > 0$ is a constant depending on $d$ and the diameter of $M$. Therefore, if we choose $\delta = \frac{1}{n^2}$,
\[
(121) \quad \sup_{f \in \mathcal{H}_E} |P_n f - Pf| \leq \frac{c_2 \sqrt{-\log \varepsilon}}{\sqrt{n}} + \sqrt{\log \frac{2n^2}{2n}} \leq \frac{c_3}{\sqrt{n}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right),
\]
where $c_3 > 0$ is a constant depending on $d$ and the diameter of $M$. Similarly, by using Lemma 21, we can show that
\[
(122) \quad \sup_{f \in \mathcal{H}_E} |P_n f - Pf| \leq \frac{c_4}{\sqrt{n}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right),
\]
\[
(123) \quad \sup_{f \in \mathcal{H}_E} |P_n f - Pf| \leq \frac{c_5}{\sqrt{n}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right),
\]
where $c_4 > 0$ and $c_5 > 0$ are constants depending on $d$, the diameter of $M$, $p_m$, and the $C^0$ norm of $p$. Hence, we have
\[
(124) \quad \sup_{f \in \mathcal{H}_E} |P_n f - Pf| \leq \frac{c_4}{\sqrt{n}e^{2d}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right),
\]
\[
(125) \quad \sup_{f \in \mathcal{H}_E} |P_n f - Pf| \leq \frac{c_5}{\sqrt{n}e^{2d}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right).
\]
The conclusion follows by taking $C_{\mathcal{E}} := \max\{c_3, c_4, c_5\}$.

**APPENDIX C. PROOF OF THEOREM 3**

In addition to notations used in the proof of Theorem 2 that are summarized in Section 5.1, we need the following new notations.

**Definition 7.** For any function $f(x) \in C(M)$, we define two operators from $C(M)$ to $C(M)$ with the $L^\infty$ norm:
\[
T_{n,t}^m f(x) = \frac{P_n f(x)}{d_{n,t}(x)}, \quad T_{\varepsilon}^m f(x) = \frac{P_{\varepsilon} f(x)}{d_{\varepsilon}(x)}.
\]
Based on the same argument as that for Lemma 6, we have the following lemma, which proof is omitted.

**Lemma 22.**

1. For any \( f \in L^2(M) \), \( T^\varepsilon f(x) = \int_M k^\varepsilon(x,y) f(y) dV_M \) is a smooth function on \( M \). In particular, the eigenfunctions of \( T^\varepsilon \) are smooth.

2. For any function \( f \in C^4(M) \),
   \[
   \frac{T^\varepsilon f(x) - f(x)}{\varepsilon^2} = \Delta f(x) + O(\varepsilon^2),
   \]
   where the implied constant in \( O(\varepsilon^2) \) depends on the \( C^4 \) norms of \( f \) and the volume and the Ricci curvature of the manifold.

Suppose \( (\lambda_{i,n,k}^m, \phi_{i,n,k}^m) \) is the \( i \)-th eigenpair of \( \frac{1-T^\varepsilon}{\varepsilon^2} \), \( (\lambda_{i,k}^m, \phi_{i,k}^m) \) is the \( i \)-th eigenpair of \( \frac{1-T_{\Delta}^\varepsilon}{\varepsilon^2} \), and \( (\lambda_i, \phi_i) \) is the \( i \)-th eigenpair of \( -\Delta \). Assume \( \phi_{i,n,k}^m \), \( \phi_{i,k}^m \), and \( \phi_i \) are normalized in the \( L^2 \) norm. Hence, we have the following proposition about the spectral convergence rate from \( \frac{1-T_{\Delta}^\varepsilon}{\varepsilon^2} \) to \( -\Delta \). The proof follows by combining Lemma 22 and the same method as in Lemmas 7, Lemma 8, and Proposition 1.

**Proposition 6.** Assume that the eigenvalues of \( \Delta \) are simple. Let \( \Gamma_K := \min_{1 \leq i \leq K} \text{dist}(\lambda_i, \sigma(-\Delta) \setminus \{\lambda_i\}) \). Suppose \( \varepsilon \) is small enough. Then, there are \( a_i \in \{-1, 1\} \) such that for all \( 1 \leq i < K \),
   \[
   |\lambda_{i,k}^m - \lambda_i| \leq C_1 \varepsilon^2, \quad \|a_i \phi_{i,k}^m - \phi_i\|_\infty \leq C_1 \varepsilon^2,
   \]
   where \( C_1 \) is a constant depending on \( \Gamma_K, \lambda_K, d \), and the injectivity radius, the sectional curvature and the second fundamental form of the manifold.

By using the similar argument as Proposition 3, we can show the convergence rate of \( \lambda_{i,n,k}^m \) to \( \lambda_{i,k}^m \) and \( \phi_{i,n,k}^m \) to \( \phi_{i,k}^m \).

**Proposition 7.** Assume that the eigenvalues of \( \Delta \) are simple. Let \( \Gamma_K := \min_{1 \leq i \leq K} \text{dist}(\lambda_i, \sigma(-\Delta) \setminus \{\lambda_i\}) \). Fix \( \varepsilon \) small enough. If \( n \) is large enough so that \( \frac{1}{\sqrt{n\varepsilon^2}} (\sqrt{-\log \varepsilon} + \sqrt{\log n}) < \gamma_1 \), where \( \gamma_1 \) depends on \( d \) and the volume and the diameter of \( M \), then there are \( a_i \in \{1, -1\} \) such that with probability greater than \( 1 - n^{-2} \), for all \( 1 \leq i < K \), we have
   \[
   |\lambda_{i,n,k}^m - \lambda_{i,k}^m| \leq \frac{C_2}{\sqrt{n\varepsilon^{d+1}}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right),
   \]
   \[
   \|a_i \phi_{i,n,k}^m - \phi_{i,k}^m\|_\infty \leq \frac{C_2}{\sqrt{n\varepsilon^{d+2}}} \left( \sqrt{-\log \varepsilon} + \sqrt{\log n} \right),
   \]
   where \( C_2 > 0 \) is a constant depending on \( \Gamma_K, \lambda_K, \sigma, \rho \).

Next, if we impose the relation between \( \varepsilon \) and \( n \) we have the following corollary.

**Corollary 2.** Assume that the eigenvalues of \( \Delta \) are simple. Fix \( \varepsilon \) small enough. Let \( \Gamma_K := \min_{1 \leq i \leq K} \text{dist}(\lambda_i, \sigma(-\Delta) \setminus \{\lambda_i\}) \). For a sufficiently small \( \varepsilon > 0 \), if \( n \) is sufficiently large so that \( \varepsilon = \varepsilon(n) \geq \left( \frac{\log n}{n} \right)^{-1} \), then with probability greater than \( 1 - n^{-2} \), for all \( 1 \leq i < K \), we have
   \[
   |\lambda_{i,n,k}^m - \lambda_{i,k}^m| \leq 2C_2 \varepsilon^2.
   \]
   For a sufficiently small \( \varepsilon > 0 \), if \( n \) is sufficiently large so that \( \varepsilon = \varepsilon(n) \geq \left( \frac{\log n}{n} \right)^{-1} \), then there are \( a_i \in \{1, -1\} \) such that with probability greater than \( 1 - n^{-2} \), for all \( 1 \leq i < K \),
   \[
   \|a_i \phi_{i,n,k}^m - \phi_{i,k}^m\|_\infty \leq 2C_2 \varepsilon^2.
   \]
   \( C_2 > 0 \) is a constant depending on \( \Gamma_K, \lambda_K, \sigma, \rho \).
Finally, we need to control the error in renormalizing the eigenvector $\tilde{v}_{i,n,\varepsilon}$ by $\| \cdot \|_{l^2(1/\beta)}$ norm in the following proposition. The proof is the same as the proof of Proposition 4 and Proposition 5.

**Proposition 8.** Let $\Gamma_K := \min_{1 \leq i \leq K} \text{dist}(\lambda_i, \sigma(-\Delta) \setminus \{\lambda_i\})$. For a sufficiently small $\varepsilon > 0$, if $n$ is sufficiently large so that $\varepsilon = \varepsilon(n) \geq \left( \frac{\log n}{n} \right)^{\frac{1}{2d+8}}$, then with probability greater than $1 - \frac{1}{n^2}$, for all $1 \leq i < K$, we have

$$\max_j \left| \frac{\tilde{v}_{i,n,\varepsilon}(f)}{\|\tilde{v}_{i,n,\varepsilon}\|_{l^2(1/\beta)}} - \phi_{i,n,\varepsilon}(x_j) \right| \leq C_3 \varepsilon^2,$$

where the constant $C_3 > 0$ is a constant depending on $\Gamma_K$, $\lambda_K$, $d$ and the volume and the diameter of $M$.

Finally, in Theorem 4, the proof of the cases $1 \leq i < K$ follows by combining Proposition 6, Corollary 2 and Proposition 8. For the case $i = 0$, the proof is the same as the case $i = 0$ in Theorem 2.

**David B Dunson, Department of Statistical Science, Duke University**  
*Email address: dunson@duke.edu*

**Hau-Tieng Wu, Departments of Mathematics and Department of Statistical Science, Duke University**  
*Email address: hauwu@math.duke.edu*

**Nan Wu, Department of Mathematics, Duke University**  
*Email address: nan.wu@duke.edu*