RADIAL BASIS APPROXIMATION OF TENSOR FIELDS ON MANIFOLDS: FROM OPERATOR ESTIMATION TO MANIFOLD LEARNING

A PREPRINT

John Harlim
Department of Mathematics, Department of Meteorology and Atmospheric Science,
Institute for Computational and Data Sciences
The Pennsylvania State University, University Park, PA 16802, USA
jharlim@psu.edu

Shixiao Willing Jiang
Institute of Mathematical Sciences
ShanghaiTech University, Shanghai 201210, China
jiangshx@shanghaitech.edu.cn

John Wilson Peoples
Department of Mathematics
The Pennsylvania State University, University Park, PA 16802, USA
jwp5828@psu.edu

September 3, 2022

ABSTRACT

In this paper, we study the Radial Basis Function (RBF) approximation to differential operators on smooth tensor fields defined on closed Riemannian submanifolds of Euclidean space, identified by randomly sampled point cloud data. Theoretically, we establish the spectral convergence for the classical pointwise RBF discrete non-symmetric approximation of Laplacians. Numerically, we found that this formulation produces a very accurate estimation of leading spectra with large enough data, which leads to a computationally expensive task of solving an eigenvalue problem of not only large but also dense, non-symmetric matrix. However, when the size data is small and/or when the local tangent plane of the unknown manifold is poorly estimated, the accuracy deteriorates. Particularly, this formulation produces irrelevant eigenvalues, in the sense that they are not approximating any of the underlying Laplacian spectra. While these findings suggest that the RBF pointwise formulation may not be reliable to approximate Laplacians for manifold learning, it is still an effective method to approximate general differential operators on smooth manifolds for other applications, including solving PDEs and supervised learning. When the manifolds are unknown, i.e., are only identified by point cloud data, the error bound of the pointwise operator estimation depends on the accuracy of the approximate local tangent spaces. To improve this approximation accuracy, we develop a second-order local SVD technique for estimating local tangent spaces on the manifold that offsets the errors induced by the curvature in the classical first-order local SVD technique. For manifold learning, we introduce a symmetric RBF discrete approximation of the Laplacians induced by a weak formulation on appropriate Hilbert spaces. Such an approximation is motivated by the desirable properties of the resulting discrete estimators, namely: The guaranteed non-negative real-valued spectra and the orthogonality of the eigenvectors. Theoretically, we establish the convergence of the eigenpairs of both the Laplace-Beltrami operator and Bochner Laplacian in the limit of large data with Monte-Carlo convergence rates provided that smooth enough kernels are used. From the practical standpoint, this result provides a justification for representing functions on the manifolds (and eigenvector fields on the

tangent bundles) in terms of the approximate eigenbasis of the Laplace-Beltrami operator (and eigenvector fields of the Bochner Laplacian). Numerically, we provide supporting examples for the Laplace-Beltrami that acts on functions and various vector Laplacians, including the Bochner, Hodge, and Lichnerowicz Laplacians, with comparisons to graph-based approximations whenever appropriate.

***K*eywords** Radial Basis Functions (RBFs) · Laplace-Beltrami operators · vector Laplacians · manifold learning · operator estimation · local SVD

1 Introduction

Estimation of differential operators is an important computational task in applied mathematics and engineering science. While this estimation problem has been studied since the time of Euler in the mid-18th century, numerical differential equations emerges as an important sub-field of computational mathematics in the 1940s when modern computers were starting to be developed for solving differential equations. Among many available numerical methods, finite-difference, finite-volume, and finite-element methods are considered the most reliable algorithms to produce accurate solutions whenever one can control the distribution of points (or nodes) and meshes. Specification of the nodes, however, requires some knowledge of the domain and is subjected to the curse of dimension.

On the other hand, the Radial Basis Function (RBF) method [6] has been considered a promising alternative that can produce very accurate approximations [33, 53] even with randomly distributed nodes [30] on high-dimensional domains. Beyond the mesh-free approximation of differential operators, the deep connection of the RBF to kernel technique has also been documented in nonparametric statistical literature [10, 30] for machine learning applications. While kernel methods play a significant role in supervised machine learning [10, 15], in unsupervised learning, a kernel approach usually corresponds to the construction of a graph whose spectral properties [11, 50] can be used for clustering [35], dimensionality reduction [1], and manifold learning [12], among others. In the context of manifold learning, given a set of data that lie on a $d-$dimensional compact Riemannian sub-manifold of Euclidean domain $\mathbb{R}^n$, the objective is to represent observable with eigensolutions of an approximate Laplacian operator. For this purpose, spectral convergence type results, concerning the convergence of the graph Laplacian matrix induced by exponentially decaying kernels to the Laplacian operator on functions, are well-documented [2, 7, 47, 8, 18] for closed manifolds and [37] for manifolds with boundary. Beyond Laplacian on functions, graph-based approximation of connection Laplacian on vector fields [43] and a Galerkin-based approximation to Hodge Laplacian on 1-forms [4] have also been considered separately. Since these manifold learning methods fundamentally approximate Laplacian operators that act on smooth tensor fields defined on manifolds, it is natural to ask whether this learning problem can be solved using the RBF method. Another motivating point is that the available graph-based approaches can only approximate a limited type of differential operators whereas RBF can approximate arbitrary differential operators, including general k-Laplacians.

Indeed, RBF has been proposed to solve PDEs on 2D surfaces [26, 27, 38, 28]. In these papers, they showed that RBF solutions converge, especially when the point clouds are appropriately placed which requires some parameterization of the manifolds. When the surface parameterization is unknown, there are several approaches to characterize the manifolds, such as the closest point method [40], the orthogonal gradient [38], the moving least squares [32], and the local SVD method [17, 54, 48]. The first two methods require laying additional grid points on the ambient space, which can be numerically expensive when the ambient dimension is high. The moving least squares locally fit multivariate quadratic functions of the local coordinates approximated by PCA (or local SVD) on the metric tensor. While fast theoretical convergence rate when the data point is well sampled (see the detailed discussion in Section 2.2 of [32]), based on the presented numerical results in the paper, it is unclear whether the same convergence rate can be achieved when the data are randomly sampled. The local SVD method, which is also used in determining the local coordinates in the moving least-squares method, can readily approximate the local tangent space with the same accuracy as the Monte-Carlo error rate, $N^{-1/2}$, where N denotes the number of randomly sampled training data points. This information alone readily allows one to approximate differential operators on the manifolds. In this paper, we propose an improved scheme, which we call the second-order local SVD, that allows one to achieve a convergence rate of order $N^{-2/d}$ for randomly sampled data of d-dimensional manifold (which beats Monte-Carlo rate for $d \leq 3$), and subsequently improve the accuracy of the pointwise estimation of arbitrary differential operators.

From the perspective of approximation theory, the radial basis type kernel is universal in the sense that the induced Reproducing Kernel Hilbert Space (RKHS) is dense in the space of continuous and bounded function on a compact domain, under the standard uniform norm [45, 44]. While this property is very appealing, previous works on RBF suggest that the non-symmetric pointwise approximation to the Laplace-Beltrami operator can be numerically

problematic [26, 38, 28]. In particular, they numerically reported that when the number of training data is small the eigenvalues of the non-symmetric RBF Laplacian matrix that approximates the negative-definite Laplace-Beltrami operator are not only complex-valued, but they can also be on the positive half-plane. These papers also empirically reported that this issue can be overcome with more data points. The work in this paper is somewhat motivated by many open questions from these empirical results.

1.1 Contribution of this paper and a summary of our findings

One of the objectives of this paper is to assess the potential of the RBF method in solving the manifold learning problem, involving approximating the Laplacians acting on functions and vector fields of smooth manifolds, where the underlying manifold is identified by a set of randomly sample point cloud data. Specifically:

1. We study the non-symmetric Laplacian RBF matrix, which is a pointwise approximation to the Laplacian operator, that is used in [26, 38] to approximate the Laplace-Beltrami operator. Particularly:

a) We prove the spectral convergence of this non-symmetric estimator as the training data size grows to infinity, which justifies the empirical observations discussed in the last paragraph prior to this subsection. See Theorem 4.3 for the Laplace-Beltrami operator and Theorem 5.3 for the Bochner Laplacian on vector field. Based on the proof, we found that this approximation may break down on high-dimensional manifolds.

b) Through a numerical study in Section 6, we found that NRBF can produce a very accurate estimation of the leading spectral properties whenever the number of sample points used to approximate the RBF matrix is large enough and the local tangent space is sufficiently accurately approximated. In this paper, we propose an improved local SVD algorithm for approximating the local tangent spaces of the manifolds, where the improvement is essentially due to the additional steps designed to correct errors induced by the curvatures (see Section 3.2). We provide a theoretical error bound (see Theorem 3.2). We numerically find that this error (from the local tangent space approximation) dominates the errors induced by the non-symmetric RBF approximation of the leading eigensolutions (see Figures 6 and 9). On the upside, since this local SVD method is numerically not expensive even with very large data, applying NRBF with the accurately estimated local tangent space will give a very accurate estimation of the leading spectra with possibly expensive computational costs. Namely, solving eigenvalue problems of non-symmetric, dense discrete NRBF matrices, which can be very large, especially for the approximation of vector Laplacians. On the downside, since this pointwise estimation relies on the accuracy of the local tangent space approximation, such a high accuracy will not be attainable when the data is corrupted by noise.

c) Through numerical verification in Sections 6.3 and 6.4, we detect another issue with the non-symmetric formulation. That is, when the training data size is not large enough, the non-symmetric RBF Laplacian matrix tends to have eigenvalues that are somewhat irrelevant in the sense that they are not estimating the true eigenvalues. If we increase the training data, while the leading eigenvalues (that are closer to zero) are accurately estimated, the irrelevant estimates will not disappear. Their occurrence will be on higher modes. This issue can be problematic in manifold learning applications since the spectra of the underlying Laplacian to be estimated are unknown. As we have point out in previous point, large data size may not be numerically feasible for the non-symmetric formulation, especially in solving eigenvalue problems corresponding to Laplacians on vector fields.

2. To overcome the limitation of the non-symmetric formulation, we consider a symmetric discrete formulation induced by a weak approximation of the Laplacians on appropriate Hilbert spaces. Several advantages of this symmetric approximation are that the estimated eigenvalues are guaranteed to be non-negative real-valued, and the corresponding estimates for the eigenvectors (or eigenvector fields) are real-valued and orthogonal. The price we are paying to guarantee estimators with these nice properties is that the approximation is less accurate compared to the non-symmetric formulation provided that the latter works. Particularly, the error of the symmetric RBF is dominated by the Monte-Carlo approximation of the integral in the weak formulation. Our findings are based on:

a) A spectral convergence study with error bounds for the estimations of eigenvalues and eigenvectors (or eigenvector fields). See Theorems 4.1 and 4.2 for the approximation of eigenvalues and eigenfunctions of Laplace-Beltrami operator, respectively. See Theorems 5.1 and 5.2 for the approximation of eigenvalues and eigenvector fields of Bochner Laplacian, respectively. The error bounds decay with the Monte-Carlo error rate when we use relatively smooth kernels in the RBF approximation.

b) Numerical inspections on the estimation of Laplace-Beltrami operator. We show the empirical convergence as a function of training data size. Overall, we conclude that the symmetric RBF is competitive to the Diffusion Maps (DM) algorithm [12], a graph-based approach, in the estimation of leading eigenvalues

and eigenvectors. We also found that SRBF robustly produces more accurate estimates of non-leading eigenvalues.

c) Numerical inspections on the estimation of Bochner, Hodge, and Lichnerowicz Laplacians. We show the empirical convergence as a function of training data size. We numerically find that the symmetric RBF is more accurate compared to Spectral Exterior Calculus [4] in the estimation of the leading spectra of Hodge-Laplacian on the sphere S^2.

1.2 Organization of this paper

Section 2: We provide a detailed formulation for discrete approximation of differential operators on smooth manifolds, where we will focus on the RBF technique as an interpolant. Overall, the nature of the approximation is "exterior" in the sense that the tangential derivatives will be represented as a projection (or restriction) of appropriate ambient derivatives onto the local tangent space. To clarify this formulation, we overview the notion of the projection matrix that allows the exterior representation that will be realized through appropriate discrete tensors. We give a concrete discrete formulation for gradient, divergence, the Laplace-Beltrami operator, connection of a vector field, and the Bochner Laplacian. We discuss the symmetric and non-symmetric formulations of both Laplacians. We list the RBF approximation in Table 1, where we also include the estimation of the Hodge and Lichnerowicz Laplacian (which detailed derivations are reported in Appendix A).

Section 3: We present a novel algorithm to improve the characterization of the manifold from randomly sampled point cloud data, which subsequently allows us to improve the approximation of the projection matrix, which is explicitly not available when the manifold is unknown. The new local SVD method, which accounts for curvature errors, will improve the accuracy of RBF in the pointwise estimation of arbitrary differential operators.

Section 4: We deduce the spectral convergence of the estimation of the Laplace-Beltrami operator. For the symmetric RBF approximation, error bounds for both the eigenvalues and eigenfunctions estimations will be given in terms of the number of training data, the smoothness of the RBF kernel, and the dimension of the manifolds. For the non-symmetric RBF approximation, we deduce the consistency of the estimation of the eigenvalues in the limit of large training data. These results rely on a probabilistic type convergence result of the RBF interpolation error, which is reported in Appendix B.

Section 5: We deduce the spectral convergence of the estimation of the Bochner Laplacian. Here, we repeat the technique used in Section 4 and obtain the analogous results, except that different Hilbert spaces and more complicated tensorial representation are used since we are approximating an operator that acts on vector fields.

Section 6: We present numerical examples to inspect the non-symmetric and symmetric RBF spectral approximations. The first two examples focus on the approximations of Laplace-Beltrami on functions. The first example of the two-dimensional general torus is to verify the effectiveness of the approximation when the low-dimensional manifold is embedded in a high-dimensional ambient space (high co-dimension). The second example of the four-dimensional (and even five) flat torus is to verify the effectiveness of the approximation when the intrinsic dimension of the manifold is more than higher than in the first example. In the third example, we verify the accuracy of the spectral estimation of the Bochner, Hodge, and Lichnerowicz Laplacians on the sphere.

Section 7: We close this paper with a short summary and discussion of remaining and emerging open problems.

2 Basic formulation for estimating differential operators on manifolds

In this section, we formulate the detailed discrete approximations of the gradient, divergence, the Laplace-Beltrami operator, connection of a vector field, and finally, the Bochner Laplacian, on smooth closed manifolds. Both the Laplace-Beltrami operator and the Bochner Laplacian have two natural discrete estimators: symmetric and non-symmetric formulations. Since each formulation has its own practical and theoretical advantages, we describe in detail both approximations. Following similar derivations, we also report the discrete approximation to other differential operators that are relevant to manifold learning, e.g. the Hodge and Lichnerowicz Laplacian (see Appendix A for the detailed derivations).

In this paper, we consider estimating differential operator acting on functions (and vector fields) defined on a $d-$dimensional closed manifold M embedded in ambient space $\mathbb{R}^n$, where $d \leq n$. Each operator is estimated using an ambient space formulation followed by a projection onto the local tangent space of the manifold using a projection matrix $\mathbf{P}$. Before describing the discrete approximation of the operators, we introduce $\mathbf{P}$, discuss some of its basic properties, and quickly review the radial basis function (RBF) interpolation which is a convenient method to approximate functions and tensor fields from the point cloud training data $X = \{x_i\}_{i=1}^N$.

In the following, we periodically use the notation $\text{diag}(a_1,\dots,a_k)$ to denote a $k\times k$ diagonal matrix with the listed entries along the diagonal. We also define $f\mapsto R_N f=\mathbf{f}=(f(x_1),f(x_2),\dots,f(x_N))^\top$ for all $f:M\to\mathbb{R}$, which is a restriction operator to the function values on training data set $X=\{x_1,\dots,x_N\}$. In the remainder of this paper, we use boldface to denote discrete objects (vectors, matrices) and script font to denote operators on continuous objects.

Definition 2.1. *For any point $x\in M$, the local parameterization $\iota:O\subseteq\mathbb{R}^d\longrightarrow M\subseteq\mathbb{R}^n$, is defined through the following map, $\boldsymbol{\Theta}_{\iota^{-1}(x)}\longmapsto\mathbf{X}_x$. Here, O denotes a domain that contains the point $\iota^{-1}(x)$, which we denoted as $\boldsymbol{\Theta}_{\iota^{-1}(x)}$ in the canonical coordinates $\left(\frac{\partial}{\partial\theta^1}\Big|_{\iota^{-1}(x)},\dots,\frac{\partial}{\partial\theta^d}\Big|_{\iota^{-1}(x)}\right)$ and $\mathbf{X}_x$ is the embedded point represented in the ambient coordinates $\left(\frac{\partial}{\partial X^1}\Big|_x,\dots,\frac{\partial}{\partial X^n}\Big|_x\right)$. The pushforward $D\iota(x):T_{\iota^{-1}(x)}O\to T_xM$ is an $n\times d$ matrix given by $D\iota(x)=\left[\frac{\partial\mathbf{X}_x}{\partial\theta^1},\dots,\frac{\partial\mathbf{X}_x}{\partial\theta^d}\right]$, a matrix whose columns form a basis of T_xM in $\mathbb{R}^n$*

Definition 2.2. *The projection matrix $\mathbf{P}=P(x)\in\mathbb{R}^{n\times n}$ on $x\in M$ is defined with the matrix-valued function $P:M\to\text{Proj}(n,\mathbb{R})\subset\mathbb{R}^{n\times n}$,*

$$P=[P_{st}]_{s,t=1}^n:=\left[\frac{\partial X^s}{\partial\theta^i}g^{ij}\frac{\partial X^t}{\partial\theta^j}\right]_{s,t=1}^n,$$

where g^{ij} denotes the matrix entries of the inverse of the Riemannian metric tensor g_{ij}.

Since $[g^{ij}]_{d\times d}=(D\iota^\top D\iota)^{-1}$, we can equivalently define the projection matrix as

$$P:=D\iota(D\iota^\top D\iota)^{-1}D\iota^\top. \tag{1}$$

Before proving some basic properties of the projection matrix $\mathbf{P}$, we must fix a set of orthonormal vectors that span the tangent space T_xM for each $x\in M$. In particular, for any $x\in M$, let $\{\boldsymbol{\tau}_i\in\mathbb{R}^{n\times1}\}_{i=1}^d$ be the d orthonormal vectors that span T_xM and let $\{\mathbf{n}_i\in\mathbb{R}^{n\times1}\}_{i=1}^{n-d}$ be the $n-d$ orthonormal vectors that are orthogonal to T_xM. Here, we suppress the dependence of $\boldsymbol{\tau}_i$ and $\mathbf{n}_i$ on x to simplify the notation. Further, let $\mathbf{T}=[\boldsymbol{\tau}_1\dots\boldsymbol{\tau}_d]\in\mathbb{R}^{n\times d}$ and $\mathbf{N}=[\mathbf{n}_1\dots\mathbf{n}_{n-d}]\in\mathbb{R}^{n\times(n-d)}$. Since $\begin{pmatrix}\mathbf{T}&\mathbf{N}\end{pmatrix}\in\mathbb{R}^{n\times n}$ is an orthonormal matrix, one has the following relation,

$$\mathbf{I}=\begin{pmatrix}\mathbf{T}&\mathbf{N}\end{pmatrix}\begin{pmatrix}\mathbf{T}^\top\\\mathbf{N}^\top\end{pmatrix}=\mathbf{T}\mathbf{T}^\top+\mathbf{N}\mathbf{N}^\top=\sum_{i=1}^d\boldsymbol{\tau}_i\boldsymbol{\tau}_i^\top+\sum_{i=1}^{n-d}\mathbf{n}_i\mathbf{n}_i^\top.$$

We have the following proposition summarizing the basic properties of $\mathbf{P}$.

Proposition 2.1. *For any $x\in M$, let $\mathbf{P}=P(x)$ be the projection matrix defined in (1) and let $\mathbf{T}=[\boldsymbol{\tau}_1,\dots,\boldsymbol{\tau}_d]\in\mathbb{R}^{n\times d}$ be any d orthonormal vectors that span T_xM. Then*
(1) $\mathbf{P}$ is symmetric;
(2) $\mathbf{P}^2=\mathbf{P}$;
(3) $\mathbf{P}=P(x)=D\iota(x)(D\iota(x)^\top D\iota(x))^{-1}D\iota(x)^\top=\mathbf{T}\mathbf{T}^\top$.
(4) $\sum_{i=1}^n|\mathbf{p}_i|^2=d$, where $\mathbf{p}_i=(P_{1i}(x),\dots,P_{ni}(x))^\top$ is the ith column of $\mathbf{P}$.

Proof. Properties (1) and (2) are obvious from the definition of $\mathbf{P}$ in (1). Property (3) can be easily obtained by observing that both sides of the equation are orthogonal projections, and $\text{span}\left\{\frac{\partial\mathbf{X}_x}{\partial\theta^1},\dots,\frac{\partial\mathbf{X}_x}{\partial\theta^d}\right\}=\text{span}\{\boldsymbol{\tau}_1\dots\boldsymbol{\tau}_d\}$. To see (4), for each point $x\in M$, write $\mathbf{P}$ as

$$\mathbf{P}=\mathbf{T}\mathbf{T}^\top=\begin{bmatrix}P_{11}(x)&\cdots&P_{1n}(x)\\\vdots&\ddots&\vdots\\P_{n1}(x)&\cdots&P_{nn}(x)\end{bmatrix}=\begin{bmatrix}\mathbf{p}_1^\top\\\vdots\\\mathbf{p}_n^\top\end{bmatrix},$$

where $\mathbf{p}_i\in\mathbb{R}^{n\times1}$ for $i=1,\dots,n$ is the ith column of $\mathbf{P}$. It remains to observe the following chain of equalities:

$$\sum_{i=1}^n\left|\mathbf{p}_i\right|^2=\text{tr}\left(\mathbf{P}^\top\mathbf{P}\right)=\text{tr}\left(\mathbf{P}\right)=\text{tr}\left(\mathbf{T}\mathbf{T}^\top\right)=\text{tr}\left(\mathbf{T}^\top\mathbf{T}\right)=\text{tr}\left(\mathbf{I}_{d\times d}\right)=d.$$

□

Let $f:M\to\mathbb{R}$ an arbitrary some smooth function. Given function values $\mathbf{f}:=(f(x_1),\dots,f(x_N))^\top$ at $X=\{x_j\}_{j=1}^N$, the radial basis function (RBF) interpolant of f at x takes the form

$$I_{\phi_s}\mathbf{f}(x):=\sum_{k=1}^N c_k\phi_s\left(\|x-x_k\|\right), \tag{2}$$

where ϕ_s denotes the kernel function with shape parameter s. Here, one can interpret $I_{\phi_s} : \mathbb{R}^N \to C^\alpha(\mathbb{R}^n)$, where α denotes the smoothness of ϕ_s. In practice, common choices of kernel include the Gaussian $\phi_s(r) = \exp(-(sr)^2)$ [21], inverse quadratic function $\phi_s(r) = (1+(sr)^2)^{-1}$, or Matérn class kernel [28]. In our numerical examples, we have tested these kernels and they do not make too much difference when the shape parameters are tuned properly. However, we will point out that the Matérn kernel will be convenient for the theoretical analysis in Section 4 as it induces a reproducing kernel Hilbert space (RKHS) with Sobolev-like norm.

The expansion coefficients $\{c_k\}_{k=1}^N$ in (2) can be determined by a collocation method, which enforces the interpolation condition $I_{\phi_s}\mathbf{f}(x_j) = f(x_j)$ for all $j = 1,\dots,N$, or the following linear system with the interpolation matrix Φ:

$$\begin{bmatrix} \phi_s(\|x_1-x_1\|) & \phi_s(\|x_1-x_2\|) & \cdots & \phi_s(\|x_1-x_N\|) \\ \phi_s(\|x_2-x_1\|) & \phi_s(\|x_2-x_2\|) & \cdots & \phi_s(\|x_2-x_N\|) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_s(\|x_N-x_1\|) & \phi_s(\|x_N-x_2\|) & \cdots & \phi_s(\|x_N-x_N\|) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_N) \end{bmatrix}. \tag{3}$$

In general, better accuracy is obtained for flat kernels (small s) [see e.g., Chap. 16–17 of [20]], however, the corresponding system in (3) becomes increasingly ill-conditioned [25, 24, 21]. In this article, we will not focus on the shape parameter issues. Numerically, we will empirically choose the shape parameter and will use pseudo-inverse to solve the linear system in (3) when it is effectively singular.

With these backgrounds, we are now ready to discuss the RBF approximation to various differential operators.

2.1 Gradient of a function

Here, we consider estimating the gradient operator from the available training data set $X = \{x_1,\dots,x_N\}$. Before we discuss the numerical approximation, we briefly review the representation of the gradient operator. The gradient $\mathrm{grad}_g f$ of a smooth function $f \in C^\infty(M)$ can be written in intrinsic coordinates $\{\theta^1,\dots\theta^d\}$, and subsequently in extrinsic coordinates $\{X^1,\dots,X^n\}$ as,

$$\mathrm{grad}_g f = \frac{\partial f}{\partial \theta^j} g^{ij} \frac{\partial}{\partial \theta^i} = \frac{\partial X^s}{\partial \theta^j}\frac{\partial f}{\partial X^s} g^{ij} \frac{\partial X^t}{\partial \theta^i}\frac{\partial}{\partial X^t},$$

when there is an extension of f onto the neighborhood of M in $\mathbb{R}^n$. Writing the above as a column vector w.r.t. the basis for the ambient space $\left\{\frac{\partial}{\partial X^t}\right\}$, we obtain a column vector

$$\mathrm{grad}_g f = (\mathscr{G}_1 f, \mathscr{G}_2 f, \dots, \mathscr{G}_n f)^\top, \tag{4}$$

where

$$\mathscr{G}_t f = \left(\frac{\partial X^s}{\partial \theta^j}\frac{\partial f}{\partial X^s}\right) g^{ij} \frac{\partial X^t}{\partial \theta^i}. \tag{5}$$

To clarify the ambient representation in (5), we have used the fact that for any $f \in C^\infty(M)$, there exists a smooth extension $F \in C^\infty(\mathbb{R}^n)$ on an open neighborhood of M in $\mathbb{R}^n$ such that $F|_M = f$ (see *pp* 104 of [31]). Since such an extension is defined only on a neighborhood of M in $\mathbb{R}^n$, we abused the notation in (5) (and in the remainder of this paper) by using f for F. With this fact, we now define,

$$\overline{\mathrm{grad}}_{\mathbb{R}^n} f = \sum_t \frac{\partial f}{\partial X^t}\frac{\partial}{\partial X^t}.$$

Treating $\overline{\mathrm{grad}}_{\mathbb{R}^n} f$ at each point $x \in M$ as an $n \times 1$ column vector

$$(\overline{\mathrm{grad}}_{\mathbb{R}^n} f)(x) := \left(\frac{\partial f}{\partial X^1}(x), \frac{\partial f}{\partial X^2}(x), \dots, \frac{\partial f}{\partial X^n}(x)\right)^\top,$$

we see that multiplying by the projection matrix value function P yields a vector-vauled function,

$$P\overline{\mathrm{grad}}_{\mathbb{R}^n} f = \begin{bmatrix} \frac{\partial X^1}{\partial \theta^i} g^{ij} \frac{\partial X^s}{\partial \theta^j}\frac{\partial f}{\partial X^s} \\ \frac{\partial X^2}{\partial \theta^i} g^{ij} \frac{\partial X^s}{\partial \theta^j}\frac{\partial f}{\partial X^s} \\ \vdots \\ \frac{\partial X^n}{\partial \theta^i} g^{ij} \frac{\partial X^s}{\partial \theta^j}\frac{\partial f}{\partial X^s} \end{bmatrix} = \begin{bmatrix} \mathscr{G}_1 f \\ \mathscr{G}_2 f \\ \vdots \\ \mathscr{G}_n f \end{bmatrix},$$

where the above is understood pointwise at each $x \in M$. Using matrix notation, we have the following equality that holds at each point $x \in M$:

$$\operatorname{grad}_g f(x) = \mathbf{P}\overline{\operatorname{grad}}_{\mathbb{R}^n} f(x).$$

To practically use the representation above, one needs to take derivatives of the function f which is typically not available in many applications. In such a case, we consider the RBF interpolant $I_{\phi_s}\mathbf{f} \in C^\alpha(\mathbb{R}^n)$ in (2) which interpolates f on the available training data set $X = \{x_1, \dots, x_N\}$ by solving (3). First, we define

Definition 2.3. *Let $\mathscr{G}_t I_{\phi_s} : \mathbb{R}^N \to C^\alpha(\mathbb{R}^n)$ and $R_N : C(\mathbb{R}^n) \to \mathbb{R}^N$ be the restriction operator defined as $R_N f = \mathbf{f} = (f(x_1), \dots, f(x_N))^\top$ for any $f \in C(M)$ and $I_{\phi_s}\mathbf{f} \in C^\alpha(\mathbb{R}^n)$ is the RBF interpolant as defined in (2). We define a linear map $\mathbf{G}_t : \mathbb{R}^N \to \mathbb{R}^N$*

$$\mathbf{G}_t\mathbf{f} = R_N \mathscr{G}_t I_{\phi_s}\mathbf{f}. \tag{6}$$

as a discrete estimator of the differential operator $\mathscr{G}_t$, restricted on the training data X. On the right-hand-side, we understood $R_N \mathscr{G}_t I_{\phi_s}\mathbf{f} = \mathscr{G}_t I_{\phi_s}\mathbf{f}|_X \in \mathbb{R}^N$.

To conclude, we define a linear map $\mathbf{G} : \mathbb{R}^N \to \mathbb{R}^{nN}$ with,

$$\mathbf{G}\mathbf{f} = (\mathbf{G}_1\mathbf{f}, \mathbf{G}_2\mathbf{f}, \dots, \mathbf{G}_n\mathbf{f})^\top, \tag{7}$$

as a discrete estimator to the gradient restricted on the training data set X, in the sense that

$$\mathbf{G}\mathbf{f} = \mathbf{G}R_N f = R_N\left(\operatorname{grad}_g I_{\phi_s}\mathbf{f}\right) = \operatorname{grad}_g I_{\phi_s}\mathbf{f}|_X, \tag{8}$$

where the first and third equalities follow from the definition of R_N, and the second equality follows from (4), (6), (7).

2.2 Divergence of a vector field

The divergence of a vector field $u \in \mathfrak{X}(M)$ can be defined in terms of covariant differentiation:

$$\operatorname{div}_g u = \mathrm{C}_1^1(\nabla u),$$

where C_1^1 is the contraction of a $(1,1)$ tensor field, and ∇ is covariant differentiation, which we define presently. Given a vector field $u = u^i \frac{\partial}{\partial\theta^i}$, i.e. a $(1,0)$ tensor field, ∇u is defined to be a $(1,1)$ tensor field given by

$$\nabla u = \left(\frac{\partial u^i}{\partial\theta^k} + u^j \Gamma^i_{jk}\right)\frac{\partial}{\partial\theta^i} \otimes d\theta^k.$$

To determine a discrete estimator for the divergence, let us derive an ambient space formulation of the covariant differentiation.

In the ambient space, i.e. for a vector field $U = U^i \frac{\partial}{\partial X^i} \in \mathfrak{X}(\mathbb{R}^n)$, the Christoffel symbols vanish and we have that

$$\overline{\nabla}_{\mathbb{R}^n} U = \frac{\partial U^i}{\partial X^j}\frac{\partial}{\partial X^i} \otimes dX^j.$$

We will identify the above with a matrix-valued function with entries $\left[\overline{\nabla}_{\mathbb{R}^n} U\right]_{ij} = \frac{\partial U^i}{\partial X^j}$. Note that at any point $x \in M$, when the $n \times n$ matrix $\mathbf{P}$ acts on $\overline{\nabla}_{\mathbb{R}^n} U(x)$, it produces an $n \times n$ matrix with entries

$$\left[\mathbf{P}\overline{\nabla}_{\mathbb{R}^n} U(x)\right]_{ij} = \sum_t \left(\frac{\partial X^i_x}{\partial\theta^r} g^{rs}_x \frac{\partial X^t_x}{\partial\theta^s}\right)\left(\frac{\partial U^t(x)}{\partial X^j}\right).$$

We will show that when U is an extension of u, the above is precisely ∇u at each point $x \in M$. Let $u \in \mathfrak{X}(M)$ be a vector field on M, and let $U \in \mathfrak{X}(\mathbb{R}^n)$ be an extension of u to an open set $O \supseteq M$. Then U is related to u via the local parameterization as follows: $U(x) = D\iota(x)\, u(x)$, where $u(x) = (u^1(x), \dots, u^d(x))$ is the coordinate representation of the vector field u w.r.t. the basis $\left\{\frac{\partial}{\partial\theta^r}\right\}$. here one thing is whether we use the notation u as a vector field or the coordinate representation of a vector field. Using $[D\iota(x)]_{sr} = \frac{\partial X^s}{\partial\theta^r}$, we have the following equation relating the components of the vector fields:

$$U^s = u^r \frac{\partial X^s}{\partial\theta^r}. \tag{9}$$

Using this identity, we can derive

Proposition 2.2. *For any vector field $U = U^i \frac{\partial}{\partial X^i} \in \mathfrak{X}(\mathbb{R}^n)$ such that $U|_M = u \in \mathfrak{X}(M)$. Using the notation in Definition 2.1, we have*

$$\sum_r g^{ij} \frac{\partial X^r}{\partial \theta^j} \frac{\partial U^r}{\partial \theta^k} = \frac{\partial u^i}{\partial \theta^k} + u^p \Gamma^i_{pk}. \tag{10}$$

Proof. From a direct calculation, we have

$$\sum_r g^{ij} \frac{\partial X^r}{\partial \theta^j} \frac{\partial U^r}{\partial \theta^k} = \sum_r g^{ij} \frac{\partial X^r}{\partial \theta^j} \frac{\partial}{\partial \theta^k} \left(u^p \frac{\partial X^r}{\partial \theta^p} \right) = \sum_r g^{ij} \frac{\partial X^r}{\partial \theta^j} \left(\frac{\partial u^p}{\partial \theta^k} \frac{\partial X^r}{\partial \theta^p} + u^p \frac{\partial^2 X^r}{\partial \theta^k \partial \theta^p} \right),$$

where we have used (9) in the first equality above. Using the fact that $\sum_r \frac{\partial X^r}{\partial \theta^i} \frac{\partial X^r}{\partial \theta^k} = g_{ik}$,

$$\sum_r g^{ij} \frac{\partial X^r}{\partial \theta^j} \frac{\partial U^r}{\partial \theta^k} = g^{ij} g_{jp} \frac{\partial u^p}{\partial \theta^k} + \sum_r g^{ij} u^p \left(\frac{\partial X^r}{\partial \theta^j} \frac{\partial^2 X^r}{\partial \theta^k \partial \theta^p} \right) = \frac{\partial u^i}{\partial \theta^k} + \sum_r g^{ij} u^p \left(\frac{\partial X^r}{\partial \theta^j} \frac{\partial^2 X^r}{\partial \theta^k \partial \theta^p} \right). \tag{11}$$

It remains to observe that

$$\sum_r \left(\frac{\partial X^r}{\partial \theta^j} \frac{\partial^2 X^r}{\partial \theta^k \partial \theta^p} \right) = \left\langle \frac{\partial \mathbf{X}}{\partial \theta^j}, \frac{\partial^2 \mathbf{X}}{\partial \theta^k \partial \theta^p} \right\rangle = \left\langle \frac{\partial \mathbf{X}}{\partial \theta^j}, \sum_r \Gamma^r_{pk} \frac{\partial \mathbf{X}}{\partial \theta^r} + \mathbf{n} \right\rangle = \sum_r \Gamma^r_{pk} \left\langle \frac{\partial \mathbf{X}}{\partial \theta^j}, \frac{\partial \mathbf{X}}{\partial \theta^k} \right\rangle = \sum_r g_{jr} \Gamma^r_{pk},$$

where we have used the fact that $\frac{\partial^2 \mathbf{X}}{\partial \theta^k \partial \theta^p}$ is a linear combination of partials of $\mathbf{X}$ with Christoffel symbols as coefficients, plus a vector $\mathbf{n}$ orthogonal to the tangent space. Plugging this to (11) yields Equation (10). □

We now use equation (10) to rewrite ∇u in terms of ambient space coordinates:

$$\nabla u = \left(\frac{\partial u^i}{\partial \theta^k} + u^p \Gamma^i_{pk} \right) \frac{\partial}{\partial \theta^i} \otimes d\theta^k = \left(\sum_r g^{ij} \frac{\partial X^r}{\partial \theta^j} \frac{\partial U^r}{\partial X^l} \frac{\partial X^l}{\partial \theta^k} \right) \frac{\partial X^m}{\partial \theta^i} \frac{\partial}{\partial X^m} \otimes d\theta^k = \left(\frac{\partial X^m}{\partial \theta^i} \sum_r g^{ij} \frac{\partial X^r}{\partial \theta^j} \frac{\partial U^r}{\partial X^l} \right) \frac{\partial}{\partial X^m} \otimes dX^l.$$

Interpreting the above as the (m,l)-th entry of a matrix-valued function, the above line shows directly that we have the following relation:

$$\nabla u = P \overline{\nabla}_{\mathbb{R}^n} U$$

that is valid pointwisely at each $x \in M$, where again U is interpreted as a column vector written w.r.t. the basis $\left\{ \frac{\partial}{\partial X^i} \right\}$. In terms of earlier notation, we have that

$$P \overline{\nabla}_{\mathbb{R}^n} U = \begin{bmatrix} \mathscr{G}_1 U^1 & \mathscr{G}_1 U^2 & \dots & \mathscr{G}_1 U^n \\ \mathscr{G}_2 U^1 & \mathscr{G}_2 U^2 & \dots & \mathscr{G}_2 U^n \\ \vdots & \vdots & \ddots & \vdots \\ \mathscr{G}_n U^1 & \mathscr{G}_n U^2 & \dots & \mathscr{G}_n U^n \end{bmatrix}.$$

From the above, we obtain that

$$\operatorname{div}_g u = \mathrm{C}^1_1 (\nabla u) = \operatorname{tr}(P \overline{\nabla}_{\mathbb{R}^n} U) = \mathscr{G}_1 U^1 + \mathscr{G}_2 U^2 + \cdots + \mathscr{G}_n U^n.$$

To arrive at a discrete estimator, we use the notation in Definition 2.3 which approximates component of u with an RBF interpolant, follows by the restriction operator R_N. Specifically, denoting $\mathbf{U}^i = (U^i(x_1), \dots, U^i(x_N))^\top$, the discrete approximation to the divergence operator on the training data set X is given by the map $\mathbb{R}^{Nn} \to \mathbb{R}^N$ defined as,

$$\left(\mathbf{U}^1, \mathbf{U}^2, \dots, \mathbf{U}^n \right) \mapsto R_N (\operatorname{div}_g I_{\phi_s} \mathbf{U}) = \sum_{i=1}^n R_N \mathscr{G}_i I_{\phi_s} \mathbf{U}^i = \sum_{i=1}^n \mathbf{G}_i \mathbf{U}^i. \tag{12}$$

2.3 The Laplace-Beltrami Operator

The Laplace-Beltrami operator on a smooth function $f \in C^\infty(M)$ is given by $\Delta_M f = -\operatorname{div}_g \left(\operatorname{grad}_g f \right)$. Using the previous ambient space formulation, we can equivalently write

$$\Delta_M f = -(\mathscr{G}_1 \mathscr{G}_1 + \mathscr{G}_2 \mathscr{G}_2 + \cdots + \mathscr{G}_n \mathscr{G}_n) f.$$

This identity yields a natural way to estimate Δ_M by composing the discrete estimators for div_g in (12) and grad_g in (8). Particularly, a **non-symmetric estimator** of the Laplace-Beltrami operator is a map $\mathbb{R}^N \to \mathbb{R}^N$ given by

$$\mathbf{f} \mapsto -(\mathbf{G}_1 \mathbf{G}_1 + \cdots + \mathbf{G}_n \mathbf{G}_n) \mathbf{f}. \tag{13}$$

Remark 1. *The above discrete version of Laplace-Beltrami is not new. In particular, it has been well studied in the deterministic setting. See [19] for a detailed review of using the ambient space for estimation of Laplace-Beltrami, as well as [28]. In Section 4, we will study the spectral convergence of this nonsymmetric discretization in the setting where the data is sampled randomly from an unknown manifold. We will remark the advantages and disadvantages of such a non-symmetric formulation for manifold learning tasks.*

In weak form, for M without boundary, the Laplace-Beltrami operator can be written

$$\int_M h\Delta_M f \, d\text{Vol} = \int_M \langle \text{grad}_g h, \text{grad}_g f \rangle \, d\text{Vol}, \quad \forall f, h \in C^\infty(M),$$

where $\langle \cdot, \cdot \rangle$ denotes the Riemannian inner product of vector fields. Using the estimators from previous subsections, there are two natural ways to estimate the Laplace-Beltrami operator in this setting. Based on the weak formulation, we can estimate the gradient with the matrix $\mathbf{G} : \mathbb{R}^N \to \mathbb{R}^{Nn}$, then compose with the matrix adjoint of $\mathbf{G}$, where the domain and range of $\mathbf{G}$ are equipped with appropriate inner products approximating the corresponding inner products defined on the manifold. It turns out that the adjoint of $\mathbf{G}$ following this procedure is just the standard matrix transpose. In particular, we have that $\mathbf{G}^\top \mathbf{G} : \mathbb{R}^N \to \mathbb{R}^N$ given by

$$\mathbf{G}^\top \mathbf{G} \mathbf{f} := \left(\mathbf{G}_1^\top \mathbf{G}_1 + \mathbf{G}_2^\top \mathbf{G}_2 + \ldots + \mathbf{G}_n^\top \mathbf{G}_n \right) \mathbf{f}$$

is a **symmetric estimator** of the Laplace-Beltrami operator.

Remark 2. *The above symmetric formulation makes use of the discrete approximation of continuous inner products, and only holds when the data is sampled uniformly. For data sampled from a non-uniform density q, however, we can perform the standard technique of correcting for non-uniform data by dividing by the sampling density. For example, the symmetric approximation to the eigenvalue problem corresponds to solving $(\lambda, \mathbf{f})$ that satisfies*

$$\mathbf{f}^\top \mathbf{G}^\top \mathbf{G} \mathbf{f} := \mathbf{f}^\top \left(\mathbf{G}_1^\top \mathbf{Q}^{-1} \mathbf{G}_1 + \mathbf{G}_2^\top \mathbf{Q}^{-1} \mathbf{G}_2 + \ldots + \mathbf{G}_n^\top \mathbf{Q}^{-1} \mathbf{G}_n \right) \mathbf{f} = \lambda \mathbf{f}^\top \mathbf{Q}^{-1} \mathbf{f}, \tag{14}$$

where $\mathbf{Q} \in \mathbb{R}^{N \times N}$ is a diagonal matrix with diagonal entries $\{q(x_i)\}_{i=1,\ldots,N}$. This weighted Monte-Carlo provides an estimate for the $L^2(M)$ inner-product in the weak formulation. When q is unknown, one can approximate q using standard density estimation techniques, such as Kernel Density Estimation methods (see, for instance, [36, 5]).

2.4 Gradient of a vector field

As above, the projection matrix P in Definition 2.2 has been used for approximating operators acting on functions as matrix-vector multiplication. In the following, we first introduce a tangential projection tensor in order to derive identities for operators acting on tensor fields with extensions. Geometrically, the tangential projection tensor projects a vector field of $\mathfrak{X}(\mathbb{R}^n)$ onto a vector field of $\mathfrak{X}(M)$.

Definition 2.4. *The tangential projection tensor $\mathscr{P} : \mathfrak{X}(\mathbb{R}^n) \to \mathfrak{X}(M)$ is defined as*

$$\mathscr{P} = \delta_{sr} \frac{\partial X^s}{\partial \theta^i} g^{ij} \frac{\partial X^t}{\partial \theta^j} \mathrm{d}X^r \otimes \frac{\partial}{\partial X^t}, \tag{15}$$

where δ_{sr} is the Kronecker delta function. In particular, for a vector field $Y = Y^k \frac{\partial}{\partial X^k} \in \mathfrak{X}(\mathbb{R}^n)$, one has

$$\begin{aligned} \mathscr{P}(Y|_M) &= \delta_{sr} Y^k \frac{\partial X^s}{\partial \theta^i} g^{ij} \frac{\partial X^t}{\partial \theta^j} \mathrm{d}X^r \left(\frac{\partial}{\partial X^k} \right) \otimes \frac{\partial}{\partial X^t} = \delta_{sr} Y^r \frac{\partial X^s}{\partial \theta^i} g^{ij} \frac{\partial X^t}{\partial \theta^j} \frac{\partial}{\partial X^t} \\ &= \delta_{sr} Y^r \frac{\partial X^s}{\partial \theta^i} g^{ij} \frac{\partial}{\partial \theta^j} \in \mathfrak{X}(M). \end{aligned}$$

For convenience, we simplify our notation as $\mathscr{P} Y := \mathscr{P}(Y|_M)$ in the rest of the paper since we only concern about the points restricted on manifold M. Obviously, for any vector field $v \in \mathfrak{X}(M)$, we have $\mathscr{P} v = v \in \mathfrak{X}(M)$.

The gradient of a vector field $u \in \mathfrak{X}(M)$ is defined by $\text{grad}_g u = \sharp \nabla u$, where $\sharp$ is the standard musical isomorphism notation to raise the index, in this case from a $(1,1)$ tensor to a $(2,0)$ tensor. In local intrinsic coordinates, one can calculate

$$\text{grad}_g u = g^{kj} \left(\frac{\partial u^i}{\partial \theta^k} + u^p \Gamma^i_{pk} \right) \frac{\partial}{\partial \theta^i} \otimes \frac{\partial}{\partial \theta^j}.$$

Using (10), we can rewrite the above as

$$\begin{aligned}
\mathrm{grad}_g u &= g^{km}\left(\delta_{rs} g^{ij}\frac{\partial X^r}{\partial\theta^j}\frac{\partial U^s}{\partial\theta^k}\right)\frac{\partial}{\partial\theta^i}\otimes\frac{\partial}{\partial\theta^m},\\
&= g^{km}\delta_{rs}g^{ij}\frac{\partial X^r}{\partial\theta^j}\left(\frac{\partial U^s}{\partial X^a}\frac{\partial X^a}{\partial\theta^k}\right)\left(\frac{\partial X^b}{\partial\theta^i}\frac{\partial}{\partial X^b}\right)\otimes\left(\frac{\partial}{\partial X^c}\frac{\partial X^c}{\partial\theta^m}\right),\\
&= \delta^{ea}\left(\delta_{ep}\frac{\partial X^c}{\partial\theta^k}g^{km}\frac{\partial X^p}{\partial\theta^m}\right)\left(\delta_{rs}\frac{\partial X^b}{\partial\theta^i}g^{ij}\frac{\partial X^r}{\partial\theta^j}\right)\frac{\partial U^s}{\partial X^a}\frac{\partial}{\partial X^b}\otimes\frac{\partial}{\partial X^c},\\
&= \mathscr{P}_1\mathscr{P}_2\left(\delta^{ea}\frac{\partial U^r}{\partial X^a}\frac{\partial}{\partial X^r}\otimes\frac{\partial}{\partial X^e}\right)\equiv\mathscr{P}_1\mathscr{P}_2\overline{\mathrm{grad}}_{\mathbb{R}^n}U, \qquad (16)
\end{aligned}$$

where $U = U^s\frac{\partial}{\partial X^s} = u^p\frac{\partial X^s}{\partial\theta^p}\frac{\partial}{\partial X^s}$, and $\mathscr{P}_1$=$\delta_{ts}\frac{\partial X^t}{\partial\theta^k}g^{km}\frac{\partial X^b}{\partial\theta^m}\mathrm{d}X^s\otimes\frac{\partial}{\partial X^b}$ acting on the first tensor field, and $\mathscr{P}_2 = \delta_{rq}\frac{\partial X^c}{\partial\theta^i}g^{ij}\frac{\partial X^r}{\partial\theta^j}\mathrm{d}X^q\otimes\frac{\partial}{\partial X^c}$ acting on the second vector field. Evaluating at each $x\in M$ and using the notation in Definition (1), Eq. (16) can be written in a matrix form as,

$$\mathrm{grad}_g u(x) = \mathbf{P}\left(\overline{\mathrm{grad}}_{\mathbb{R}^n}U(x)\right)\mathbf{P} = \begin{bmatrix} \mathscr{G}_1U^1(x) & \cdots & \mathscr{G}_1U^n(x)\\ \vdots & \ddots & \vdots\\ \mathscr{G}_nU^1(x) & \cdots & \mathscr{G}_nU^n(x)\end{bmatrix}_{n\times n}\begin{bmatrix} P_{11}(x) & \cdots & P_{1n}(x)\\ \vdots & \ddots & \vdots\\ P_{n1}(x) & \cdots & P_{nn}(x)\end{bmatrix}_{n\times n}.$$

Interpreting $U(x) = (U^1(x),\ldots,U^n(x))^\top\in\mathbb{R}^{n\times 1}$ as a vector, one sees immediately that

$$\mathrm{grad}_g u(x) = (\mathscr{H}_1U(x),\ldots,\mathscr{H}_nU(x)), \qquad (17)$$

where each component can be rewritten as,

$$\mathscr{H}_iU(x) := \mathbf{P}\mathrm{diag}(\mathscr{G}_i,\ldots,\mathscr{G}_i)\,U(x),$$

owing to the symmetry of $\mathbf{P}$.

To write the discrete approximation on the training data set $X=\{x_1,\ldots,x_n\}$, we define $\mathbf{U}^i = (U^i(x_1),\ldots,U^i(x_N))^\top\in\mathbb{R}^{N\times 1}$ and $\mathbf{U} = ((\mathbf{U}^1)^\top,\ldots,(\mathbf{U}^n)^\top)^\top\in\mathbb{R}^{nN\times 1}$. Consider now the map $\mathbf{H}_i:\mathbb{R}^{nN}\to\mathbb{R}^{nN}$ defined by

$$\mathbf{H}_i\mathbf{U} := R_N\mathscr{H}_i I_{\phi_s}\mathbf{U}, \qquad (18)$$

where the interpolation is defined on each $\mathbf{U}^i\in\mathbb{R}^N$ such that $I_{\phi_s}\mathbf{U}^i\in C^\alpha(\mathbb{R}^n)$ and the restriction R_N is applied on each row. Relating to $\mathbf{G}_i$ in Definition 2.3, one can write

$$\mathbf{H}_i\mathbf{U} = \mathbf{P}^\otimes\begin{bmatrix}\mathbf{G}_i & & \\ & \ddots & \\ & & \mathbf{G}_i\end{bmatrix}_{Nn\times Nn}\begin{bmatrix}\mathbf{U}^1\\ \vdots\\ \mathbf{U}^n\end{bmatrix}_{Nn\times 1},$$

where tensor projection matrix $\mathbf{P}^\otimes$ is given by

$$\mathbf{P}^\otimes = \sum_{k=1}^N\begin{bmatrix} P_{11}(x_k) & \cdots & P_{1n}(x_k)\\ \vdots & \ddots & \vdots\\ P_{n1}(x_k) & \cdots & P_{nn}(x_k)\end{bmatrix}_{n\times n}\otimes[\delta_{kk}]_{N\times N} = \begin{bmatrix}\mathrm{diag}(\mathbf{p}_{11}) & \cdots & \mathrm{diag}(\mathbf{p}_{1n})\\ \vdots & \ddots & \vdots\\ \mathrm{diag}(\mathbf{p}_{n1}) & \cdots & \mathrm{diag}(\mathbf{p}_{nn})\end{bmatrix}_{Nn\times Nn},$$

where $\otimes$ is the Kronecker product between two matrices, δ_{kk} has value 1 for the entry in kth row and kth column and has 0 values elsewhere, and $\mathbf{p}_{ij} = \left(P_{ij}(x_1),\ldots,P_{ij}(x_N)\right)\in\mathbb{R}^{N\times 1}$. Finally, consider the linear map $\mathbf{H}:\mathbb{R}^{nN}\to\mathbb{R}^{nN\times n}$ given by

$$\mathbf{H}\mathbf{U} := [\mathbf{H}_1\mathbf{U},\ldots,\mathbf{H}_n\mathbf{U}] = R_N\mathrm{grad}_g I_{\phi_s}\mathbf{U}, \qquad (19)$$

as an estimator of the gradient of any vector field U restricted on the training data set X, where on the right-hand-side, the restriction is done to each function entry-wise which results in an $N\times 1$ column vector. In the last equality, we have used the representation in (17) for the gradient of the interpolated vector field $I_{\phi_s}\mathbf{U}$ whose components are functions in $C^\alpha(\mathbb{R}^n)$ and the identity in (18).

2.5 Divergence of a (2,0) tensor field

Let v be a $(2,0)$ tensor field of $v = v^{jk}\frac{\partial}{\partial\theta^j}\otimes\frac{\partial}{\partial\theta^k}$ and V be the corresponding extension in ambient space. The divergence of v is defined as

$$\mathrm{div}_1^1(v) = C_1^1(\nabla v), \tag{20}$$

where C_1^1 denotes the contraction operator. Following a similar layout as before, we obtain an ambient space formulation of the divergence of a $(2,0)$ tensor field (see Appendix A for the detailed derivations). Interpreting $V(x)$ as a matrix with columns $\bar{V}_i(x)$, the divergence of a $(2,0)$ tensor field evaluated at any $x\in M$ can be written as

$$\mathrm{div}_1^1(v(x)) = \mathbf{P}\mathrm{tr}_1^1\left(\mathbf{P}\bar{\nabla}_{\mathbb{R}^n}(V(x))\right) = \sum_i \mathbf{P}\mathrm{diag}(\mathscr{G}_i,\ldots,\mathscr{G}_i)\bar{V}_i(x) = \sum_i \mathscr{H}_i\bar{V}_i(x).$$

Using the same procedure as before, employing div_1^1 on the RBF interpolant (2,0) tensor field $I_{\phi_s}\bar{\mathbf{V}}_i$, where $\bar{\mathbf{V}}_i = \bar{V}_i|_X\in\mathbb{R}^{nN\times 1}$ denotes the restriction of $\bar{V}_i$ on the training data, we arrive at an estimator of the divergence div_1^1 of a $(2,0)$ tensor field. Namely, replacing each $\mathscr{H}_i$ with the discrete version $\mathbf{H}_i$ as defined in (18), we obtain a map $\mathbb{R}^{nN\times n}\to\mathbb{R}^{nN}$ given by

$$[\bar{\mathbf{V}}_1,\ldots,\bar{\mathbf{V}}_n]\mapsto \mathbf{H}_1\bar{\mathbf{V}}_1+\cdots+\mathbf{H}_n\bar{\mathbf{V}}_n.$$

2.6 Bochner Laplacian

The Bochner Laplacian $\Delta_B:\mathfrak{X}(M)\to\mathfrak{X}(M)$ is defined by

$$\Delta_B u = -\mathrm{div}_1^1\left(\mathrm{grad}_g u\right),$$

where div_1^1 is in fact the formal adjoint of grad_g acting on vector fields. In particular, for $u,v\in\mathfrak{X}(M)$, we define the following inner-product:

$$\langle u,v\rangle_{L^2(\mathfrak{X}(M))} := \int_M \langle u,v\rangle_x d\mathrm{Vol}(x).$$

Similarly, for any two smooth $(2,0)$ tensor fields $a = a_{ij}\frac{\partial}{\partial\theta^i}\otimes\frac{\partial}{\partial\theta^j}$, $b = b_{ij}\frac{\partial}{\partial\theta^i}\otimes\frac{\partial}{\partial\theta^j}$, we define

$$\langle a,b\rangle_{L^2(T^{(2,0)}TM)} := \int_M a_{ij}(x)b_{kl}(x)\left\langle\frac{\partial}{\partial\theta^i},\frac{\partial}{\partial\theta^k}\right\rangle_x\left\langle\frac{\partial}{\partial\theta^j},\frac{\partial}{\partial\theta^l}\right\rangle_x d\mathrm{Vol}(x).$$

With respect to these inner-products, we have that

$$\langle\mathrm{grad}_g u,a\rangle_{L^2(T^{(2,0)}TM)} = -\langle u,\mathrm{div}_1^1 a\rangle_{L^2(\mathfrak{X}(M))},$$

for any $u\in\mathfrak{X}$, $a\in T^{(2,0)}TM$. A natural way to estimate Δ_B is to compose the discrete estimators for div_1^1 and grad_g. In particular, a **non-symmetric estimator** of the Bochner Laplacian is a map $\mathbb{R}^{nN}\to\mathbb{R}^{nN}$ given by

$$\mathbf{U}\mapsto -(\mathbf{H}_1\mathbf{H}_1+\cdots+\mathbf{H}_n\mathbf{H}_n)\mathbf{U},$$

where $\mathbf{U} = U|_X\in\mathbb{R}^{nN\times 1}$ and $\mathbf{H}_i$ are as defined in (18).

The second formulation relies on the fact that div_1^1 is indeed the formal adjoint of grad_g acting on a vector field such that,

$$\int_M\langle\Delta_B u,v\rangle_x d\mathrm{Vol}(x) = \int_M\langle\mathrm{grad}_g u,\mathrm{grad}_g v\rangle_x d\mathrm{Vol}(x),\qquad \forall u,v\in\mathfrak{X}(M),$$

where the inner-product on the left is the Riemannian inner product of vector fields, and the inner-product on the right is the Riemannian inner product of $(2,0)$ tensor fields. Similar to the symmetric discrete estimator of the Laplace-Beltrami operator, we take advantages of the ambient space formulations in previous two subsections to approximate the inner product in (21) with appropriate normalized inner products in Euclidean space.

First, we notice that the transpose of the map $\mathbf{H}$

$$\mathbf{U}\mapsto\left[\mathbf{H}_1\mathbf{U},\mathbf{H}_2\mathbf{U},\ldots,\mathbf{H}_n\mathbf{U}\right]$$

is given by the standard transpose. Due to the possibility, however, that $\mathbf{H}^\top$ may produce a vector corresponding to a vector field with components normal to the manifold which are nonzero, there is a need to compose such an estimator with $\mathbf{P}$. With this consideration, we have $\mathbf{P}^\otimes\mathbf{H}^\top\mathbf{H}\mathbf{P}^\otimes:\mathbb{R}^{nN}\to\mathbb{R}^{nN}$ given by

$$\mathbf{P}^\otimes\mathbf{H}^\top\mathbf{H}\mathbf{P}^\otimes\mathbf{U} = \mathbf{P}^\otimes\mathbf{H}_1^\top\mathbf{H}_1\mathbf{P}^\otimes\mathbf{U}+\mathbf{P}^\otimes\mathbf{H}_2^\top\mathbf{H}_2\mathbf{P}^\otimes\mathbf{U}+\cdots+\mathbf{P}^\otimes\mathbf{H}_n^\top\mathbf{H}_n\mathbf{P}^\otimes\mathbf{U}$$

as a **symmetric estimator** of the Bochner Laplacian on vector fields.

Table 1: RBF formulation for functions and vector fields from Riemannian geometry.

Object	Continuous operator	Discrete matrix
gradient	$\text{grad}_g : C^\infty(M) \to \mathfrak{X}(M)$	$\text{grad}_g : \mathbb{R}^N \to \mathbb{R}^{N\times n}$
functions	$f \mapsto \left[\mathscr{G}_1 f, \dots, \mathscr{G}_n f\right]$	$\mathbf{f} \mapsto (\mathbf{G}_1\mathbf{f}, \dots, \mathbf{G}_n\mathbf{f})$
divergence	$\text{div}_g : \mathfrak{X}(M) \to C^\infty(M)$	$\text{div}_g : \mathbb{R}^{N\times n} \to \mathbb{R}^{N\times 1}$
vector fields	$U \mapsto \mathscr{G}_1 U^1 + \cdots + \mathscr{G}_n U^n$	$\mathbf{U} \mapsto \mathbf{G}_1\mathbf{U}^1 + \cdots + \mathbf{G}_n\mathbf{U}^n$
Laplace-Beltrami	$\Delta_g : C^\infty(M) \to C^\infty(M)$	$\Delta_g : \mathbb{R}^{N\times 1} \to \mathbb{R}^{N\times 1}$
non-symmetric	$f \mapsto -(\mathscr{G}_1\mathscr{G}_1 + \cdots + \mathscr{G}_n\mathscr{G}_n) f$	$\mathbf{f} \mapsto -(\mathbf{G}_1\mathbf{G}_1 + \cdots + \mathbf{G}_n\mathbf{G}_n)\mathbf{f}$
symmetric	$f \mapsto \left(\mathscr{G}_1^\top\mathscr{G}_1 + \cdots + \mathscr{G}_n^\top\mathscr{G}_n\right) f$	$\mathbf{f} \mapsto \left(\mathbf{G}_1^\top\mathbf{G}_1 + \cdots + \mathbf{G}_n^\top\mathbf{G}_n\right)\mathbf{f}$
gradient	$\text{grad}_g : \mathfrak{X}(M) \to \mathfrak{X}(M)\times\mathfrak{X}(M)$	$\text{grad}_g : \mathbb{R}^{Nn\times 1} \to \mathbb{R}^{Nn\times n}$
vector fields	$U \mapsto [\mathscr{H}_1 U, \dots, \mathscr{H}_n U]$	$\mathbf{U} \mapsto [\mathbf{H}_1\mathbf{U}, \dots, \mathbf{H}_n\mathbf{U}]$
divergence	$\text{div}_1^1 : \mathfrak{X}(M)\times\mathfrak{X}(M) \to \mathfrak{X}(M)$	$\text{div}_1^1 : \mathbb{R}^{Nn\times n} \to \mathbb{R}^{Nn\times 1}$
(2,0) tensor fields	$V \mapsto \mathscr{H}_1\bar{V}_1 + \cdots + \mathscr{H}_n\bar{V}_n$	$[\bar{\mathbf{V}}_1, \dots, \bar{\mathbf{V}}_n] \mapsto \mathbf{H}_1\bar{\mathbf{V}}_1 + \cdots + \mathbf{H}_n\bar{\mathbf{V}}_n$
	$\bar{V}_i$ is the ith row of V	$\bar{\mathbf{V}}_i = [\mathbf{V}_{i1}, \dots, \mathbf{V}_{in}] \in \mathbb{R}^{Nn\times 1}$
Bochner Laplacian	$\Delta_B : \mathfrak{X}(M) \to \mathfrak{X}(M)$	$\Delta_B : \mathbb{R}^{Nn\times 1} \to \mathbb{R}^{Nn\times 1}$
non-symmetric	$U \mapsto -(\mathscr{H}_1\mathscr{H}_1 + \cdots + \mathscr{H}_n\mathscr{H}_n)U$	$\mathbf{U} \mapsto -(\mathbf{H}_1\mathbf{H}_1 + \cdots + \mathbf{H}_n\mathbf{H}_n)\mathbf{U}$
symmetric	$U \mapsto (\mathscr{H}_1^\top\mathscr{H}_1 + \cdots + \mathscr{H}_n^\top\mathscr{H}_n)U$	$\mathbf{U} \mapsto (\mathbf{P}^\otimes\mathbf{H}_1^\top\mathbf{H}_1\mathbf{P}^\otimes + \cdots + \mathbf{P}^\otimes\mathbf{H}_n^\top\mathbf{H}_n\mathbf{P}^\otimes)\mathbf{U}$
Hodge Laplacian	$\Delta_H : \mathfrak{X}(M) \to \mathfrak{X}(M)$	$\Delta_H : \mathbb{R}^{Nn\times 1} \to \mathbb{R}^{Nn\times 1}$
vector fields	$U \mapsto -\begin{bmatrix}\mathscr{H}_1\\ \vdots\\ \mathscr{H}_n\end{bmatrix} \cdot \text{Ant} \begin{bmatrix}\mathscr{H}_1 U\\ \vdots\\ \mathscr{H}_n U\end{bmatrix}$	$\mathbf{U} \mapsto -\begin{bmatrix}\mathbf{H}_1\\ \vdots\\ \mathbf{H}_n\end{bmatrix} \cdot \text{Ant} \begin{bmatrix}\mathbf{H}_1\mathbf{U}\\ \vdots\\ \mathbf{H}_n\mathbf{U}\end{bmatrix}$
Ant is the anti-symmetric part	$-\begin{bmatrix}\mathscr{G}_1\\ \vdots\\ \mathscr{G}_n\end{bmatrix} \left(\sum_{k=1}^n \mathscr{G}_k U^k\right)$	$-\begin{bmatrix}\mathbf{G}_1\\ \vdots\\ \mathbf{G}_n\end{bmatrix} \left(\sum_{k=1}^n \mathbf{G}_k\mathbf{U}^k\right)$
Lichnerowicz Lap.	$\Delta_L : \mathfrak{X}(M) \to \mathfrak{X}(M)$	$\Delta_L : \mathbb{R}^{Nn\times 1} \to \mathbb{R}^{Nn\times 1}$
vector fields	$U \mapsto -\begin{bmatrix}\mathscr{H}_1\\ \vdots\\ \mathscr{H}_n\end{bmatrix} \cdot \text{Sym} \begin{bmatrix}\mathscr{H}_1 U\\ \vdots\\ \mathscr{H}_n U\end{bmatrix}$	$\mathbf{U} \mapsto -\begin{bmatrix}\mathbf{H}_1\\ \vdots\\ \mathbf{H}_n\end{bmatrix} \cdot \text{Sym} \begin{bmatrix}\mathbf{H}_1\mathbf{U}\\ \vdots\\ \mathbf{H}_n\mathbf{U}\end{bmatrix}$
covariant derivative	$\nabla : \mathfrak{X}(M)\times\mathfrak{X}(M) \to \mathfrak{X}(M)$	$\nabla : \mathbb{R}^{Nn\times 1}\times\mathbb{R}^{Nn\times 1} \to \mathbb{R}^{Nn\times 1}$
	$(U, Y) \mapsto P\bar{\nabla}_U Y$	$(\mathbf{U}, \mathbf{Y}) \mapsto \mathbf{P}^\otimes\bar{\nabla}_{\mathbf{U}}\mathbf{Y}$

Remark 3. *Again, we note that this symmetric formulation makes use of approximating continuous inner products, and hence obviously holds only for uniform data. For data sampled from a non-uniform density q, we perform the same trick mentioned in Remark 2.*

We conclude this section with a list of RBF discrete formulation in Table 1. We also add a discrete approximation for Hodge and Lichnerowicz Laplacians, and covariant derivative (whose detailed derivation can be found in Appendix A).

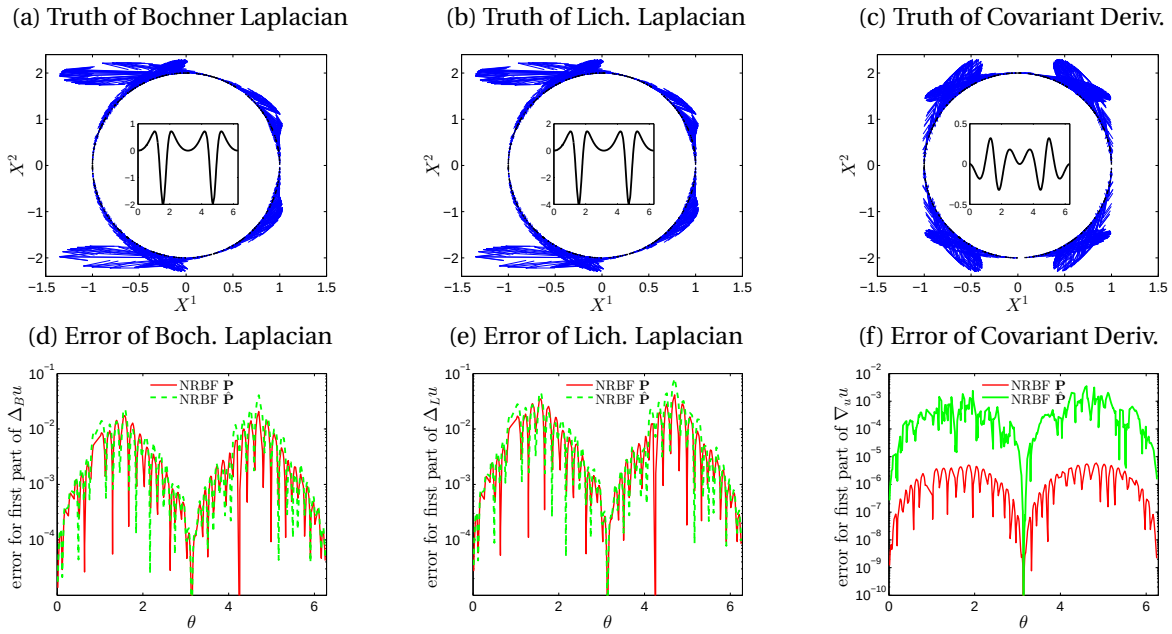

Figure 1: **1D ellipse in $\mathbb{R}^2$.** The upper panels display the truth of (a) Bochner Laplacian, (b) Lichnerowicz Laplacian, and (c) covariant derivative of a vector field. The insets of upper panels display the first components of these operator approximations. The bottom panels display the errors of NRBF approximations using analytic $\mathbf{P}$ (red curve) and approximated $\hat{\mathbf{P}}$ (green curve) for (d) Bochner Laplacian, (e) Lichnerowicz Laplacian, and (f) covariant derivative. The Gaussian kernel with shape parameter $s = 1.5$ was used. The $N = 400$ data points are randomly distributed.

2.7 Numerical verification for operator approximation

We now show the non-symmetric RBF (NRBF) approximations for vector Laplacians and covariant derivative. The manifold is a one-dimensional full ellipse,

$$x = (x^1, x^2) = (\cos\theta, a\sin\theta), \tag{21}$$

defined with the Riemannian metric $g = \sin^2\theta + a^2\cos^2\theta$ for $0 \le \theta \le 2\pi$,where $a = 2 > 1$. The $N = 400$ data points are randomly distributed on the ellipse. The Gaussian kernel with the shape parameter $s = 1.5$ was used.

We first approximate the vector Laplacians. We take a vector field $u = u^1 \frac{\partial}{\partial\theta}$ with $u^1(x) \equiv u^1(x(\theta)) = \sin\theta$. The Bochner Laplacian acting on u can be calculated as $\Delta_B u = g^{-1} u^1_{,11} \frac{\partial}{\partial\theta}$, where $u^1_{,11} = \frac{\partial^2 u^1}{\partial\theta^2} + \frac{\partial u^1}{\partial\theta}\Gamma^1_{11} + u^1 \frac{\partial\Gamma^1_{11}}{\partial\theta}$ with $\Gamma^1_{11} = \frac{1}{2} g^{-1} \frac{\partial g}{\partial\theta}$. Numerically, Fig. 1(a) shows the Bochner Laplacian $\Delta_B u = \bar{\Delta}_B U$, which is a 2×1 vector lying in the tangent space of each given point. The inset of Fig. 1(a) displays the first vector component of the Bochner Laplacian as a function of the intrinsic coordinate θ. Figure 1(d) displays the error for the first vector component of Bochner Laplacian as a function of θ. Here, we show the errors of using the analytic $\mathbf{P}$ and an approximated $\hat{\mathbf{P}}$. Here (and in the remainder of this paper), we used the notation $\hat{\mathbf{P}}$ to denote the approximated $\mathbf{P}$ obtained from the second-order method discussed in Section 3. One can clearly see that the errors for NRBF using both analytic $\mathbf{P}$ and approximated $\hat{\mathbf{P}}$ are small about 0.01. Since Hodge Laplacian is identical to Bochner Laplacian on a 1D manifold, the results for Hodge Laplacian are almost the same as those for Bochner [not shown here]. The Lichnerowicz Laplacian is the double of Bochner Laplacian, $\Delta_L u = 2 g^{-1} u^1_{,11} \frac{\partial}{\partial\theta}$, as shown in Fig. 1(b). The error for the first vector component of Lichnerowicz Laplacian is also nearly doubled as shown in Fig. 1(e).

We next approximate the covariant derivative. The covariant derivative can be calculated as $\nabla_u u = u^1 (\frac{\partial u^1}{\partial\theta} + u^1\Gamma^1_{11}) \frac{\partial}{\partial\theta}$ as shown in Fig. 1(c). Figure 1(f) displays the error of NRBF approximation for the first vector component of covariant derivative $\nabla_u u$. One can see from Fig. 1(f) that the error for analytic $\mathbf{P}$ (red) is very small about 10^{-6} and the error for approximated $\hat{\mathbf{P}}$ (green) is about 10^{-3}.

3 Estimation of the projection matrix

When the manifold M is unknown and identified only by a point cloud data $X = \{x_1, \dots, x_N\}$, where $x_i \in M$, we do not immediately have access to the matrix-valued function P. In this section, we first give a quick overview of the existing first-order local SVD method for estimating $\mathbf{P} = P(x)$ on each $x \in M$. Subsequently, we present a novel second-order method (which is a local SVD method that corrects the estimation error induced by the curvature) under the assumption that the dataset X lies on a C^3 d-dimensional Riemannian manifold M embedded in $\mathbb{R}^n$.

Let $x, y \in X$ such that $\left|y - x\right| = O(s)$. Define γ to be a geodesic, connecting x and y. The curve is parametrized by the arc-length,

$$s = \int_0^s \left|\gamma'(t)\right| dt,$$

where $\gamma(0) = x, \gamma(s) = y$. Taking derivative with respect to s, we obtain constant velocity, $1 = \left|\gamma'(t)\right|$ for all $0 \le t \le s$. Let $\mathbf{s} = (s_1, \dots, s_d)$ be the geodesic normal coordinate of y defined by an exponential map $\exp_x : T_xM \to M$. Then $\mathbf{s}$ satisfies

$$s\gamma'(0) = \mathbf{s} = \exp_x^{-1}(y),$$

where

$$s^2 = s^2\left|\gamma'(0)\right|^2 = |\mathbf{s}|^2 = \sum_{i=1}^d s_i^2.$$

For any point $x, y \in M$, let ι be the local parametrization of manifold such that $\iota(\mathbf{s}) = y$ and $\iota(\mathbf{0}) = x$. Consider the Taylor expansion of $\iota(\mathbf{s})$ centered at $\mathbf{0}$,

$$\iota(\mathbf{s}) = \iota(\mathbf{0}) + \sum_{i=1}^d s_i \frac{\partial \iota(\mathbf{0})}{\partial s_i} + \frac{1}{2}\sum_{i,j=1}^d s_i s_j \frac{\partial^2 \iota(\mathbf{0})}{\partial s_i \partial s_j} + O(s^3). \tag{22}$$

Since the Riemannian metric tensor at the based point $\mathbf{0}$ is an identity matrix, $\boldsymbol{\tau}_i = \frac{\partial \iota(\mathbf{0})}{\partial s_i} \Big/ \left|\frac{\partial \iota(\mathbf{0})}{\partial s_i}\right| = \frac{\partial \iota(\mathbf{0})}{\partial s_i}$ are d orthonormal tangent vectors that span T_xM.

3.1 First-order local SVD method

The classical local SVD method [17, 54, 48] uses the difference vector $y - x = \iota(\mathbf{s}) - \iota(\mathbf{0})$ to estimate $\mathbf{T} = (\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_d)$ (up to an orthogonal rotation) and subsequently use it to approximate $\mathbf{P} = \mathbf{T}\mathbf{T}^\top$. Numerically, the first-order local SVD proceeds as follows:

1. For each $x \in X$, let $\{y_1, \dots, y_K\} \subset X$ be the K-neareast neighbor (one can also use a radius neighbor) of x. Construct the distance matrix $\mathbf{D} := [\mathbf{D}_1, \dots, \mathbf{D}_K] \in \mathbb{R}^{n\times K}$, where $K > d$ and $\mathbf{D}_i := y_i - x$.
2. Take a singular value decomposition of $\mathbf{D} = \mathbf{U}\Sigma\mathbf{V}^\top$. Then the leading $d-$ columns of $\mathbf{U}$ consists of $\tilde{\mathbf{T}}$ which approximates a span of column vectors of $\mathbf{T}$, which forms a basis of T_xM.
3. Approximate $\mathbf{P}$ with $\tilde{\mathbf{P}} = \tilde{\mathbf{T}}\tilde{\mathbf{T}}^\top$.

Based on the Taylor's expansion in (22), such an approximation can only provide an estimate with accuracy $\|\tilde{\mathbf{P}} - \mathbf{P}\|_F = O(s)$, which is an order-one scheme. For uniformly sampled data, we state the following definition and probabilistic type convergence result (Theorem 2 of [48]) for this local SVD method, which will be useful in our convergence study.

Definition 3.1. *For each point $x \in M$, where M is a d-dimensional smooth manifold embedded in $\mathbb{R}^n$, where $d+1 \le n$. We define $N_\epsilon(x) = M \cap B_x(\sqrt{\epsilon})$, where $B_x(\sqrt{\epsilon})$ denotes the Euclidean ball (in $\mathbb{R}^n$) centered at x and radius $\sqrt{\epsilon}$. If M has a positive injectivity radius at $\sqrt{\epsilon} > 0$ at $x \in M$, then there is a diffeomorphism between $N_\epsilon(x)$ and T_xM. In such a case, there exists a local one-to-one map $T_xM \ni \mathbf{s} \mapsto y = \exp_x(\mathbf{s}) := (\mathbf{s}, f_1(\mathbf{s}), \dots f_{n-d}(\mathbf{s})) \in M \subset \mathbb{R}^n$, for all $y \in M$, with smooth functions $f_\ell : T_xM \to \mathbb{R}$. We also denote the maximum principal curvature at x as K_{max}.*

Theorem 3.1. *Suppose that $\{x_i \in M\}_{i=1}^K$ are point cloud data points such that their orthogonal projections, $\mathbf{s}_i \sim U[-\sqrt{\epsilon}, \sqrt{\epsilon}]^d \subset T_xM$ at x are i.i.d. Let $\mathbf{W} \in \mathbb{R}^{(n-d)\times(n-d)}$ be a matrix with components given as,*

$$W_{ij} = \mathbb{E}_{\mathbf{s}\sim U[-\sqrt{\epsilon},\sqrt{\epsilon}]^d}[f_{q,i}(\mathbf{s}) f_{q,j}(\mathbf{s})],$$

where $f_{q,\ell}$ denotes a quadratic form of f_ℓ, which is the second-order Taylor expansion about the base point $\mathbf{0}$, representing the curvature of M. If $\mathbf{W}$ is a dense matrix, for any $\tau \in (0,1)$, in high probability,

$$\|\tilde{\mathbf{P}} - \mathbf{P}\|_F \le \sqrt{2}\tau,$$

if $\epsilon = O(n^{-1}d^{-2}|K_{max}|^{-2})$ and $K = O(\tau^{-2}d^2 \log n)$. Here $\|\cdot\|_F$ denotes the standard Frobernius matrix norm.

Remark 4. *We should point out that the result above implies that the local SVD has a Monte-Carlo error rate, $\tau = O(K^{-1/2})$ and the choice of local neighbor of radius $\sqrt{\epsilon} = O(s)$ should be inversely proportional to the maximum principal curvature and the dimension of the manifold and ambient space. So, to expect an error $\tau = O(\epsilon^{1/2})$, by balancing $n^{-1}d^{-2}|K_{max}|^{-2} \sim K^{-1}d^2 \log n$, this result suggests that one should choose $K = O(d^4 n \log n |K_{max}|^2)$.*

3.2 Second-order local SVD method

In this section, we devise an improved scheme to achieve the tangent space approximation with accuracy up to order of $O(s^2)$, by accounting for the Hessian components in (22). The algorithm proceeds as follows:

1. Perform the first-order local SVD algorithm and attain $\tilde{\mathbf{T}} = [\tilde{\boldsymbol{t}}_1, \dots, \tilde{\boldsymbol{t}}_d] \in \mathbb{R}^{n\times d}$.
2. For each neighbor $\{y_i\}_{i=1,\dots,k}$ of x, compute $\tilde{\mathbf{s}}^{(i)} = (\tilde{s}_1^{(i)}, \dots, \tilde{s}_d^{(i)})$, where
$$\tilde{s}_j^{(i)} = \mathbf{D}_i^\top \tilde{\boldsymbol{t}}_j, \quad i = 1, \dots, k, j = 1, \dots, d,$$
where $\mathbf{D}_i := y_i - x$ is the ith column of $\mathbf{D} \in \mathbb{R}^{n\times d}$.
3. Approximate the Hessian $\mathbf{Y}_p = \frac{\partial^2 \iota(\mathbf{0})}{\partial s_i \partial s_j} \in \mathbb{R}^n$ using the following ordinary least squared regression, with $p = 1, \dots, D = d(d+1)/2$ denoting the upper triangular components ($p \rightarrowtail (i,j)$ such that $i \le j$) of symmetric Hessian matrix. Notice that for each $y_\ell \in \{y_1, \dots, y_K\}$ neighbor of x, the equation (22) can be written as
$$\sum_{i,j=1}^{d} s_i^{(\ell)} s_j^{(\ell)} \frac{\partial^2 \iota(\mathbf{0})}{\partial s_i \partial s_j} = 2\left(\iota(\mathbf{s}) - \iota(\mathbf{0})\right) - 2\sum_{i=1}^{d} s_i^{(\ell)} \boldsymbol{\tau}_i + O(s^3), \qquad \ell = 1, \dots K. \tag{23}$$
In compact form, we can rewrite (23) as a linear system,
$$\mathbf{A}\mathbf{Y} = 2\mathbf{D}^\top - 2\mathbf{R} + O(s^3), \tag{24}$$
where $\mathbf{R}^\top = (\mathbf{r}_1, \dots, \mathbf{r}_K) \in \mathbb{R}^{n\times K}$ denotes the order-s residual term.
$$\boldsymbol{r}_j = \sum_{i=1}^{d} s_i^{(j)} \boldsymbol{\tau}_i.$$
Here, $\mathbf{Y} \in \mathbb{R}^{D\times n}$ is a matrix whose pth row is $\mathbf{Y}_p$ and
$$\mathbf{A} = \begin{pmatrix} (s_1^{(1)})^2 & (s_2^{(1)})^2 & \dots & (s_d^{(1)})^2 & 2(s_1^{(1)}s_2^{(1)}) & \dots & 2(s_{d-1}^{(1)}s_d^{(1)}) \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ (s_1^{(k)})^2 & (s_2^{(k)})^2 & \dots & (s_d^{(k)})^2 & 2(s_1^{(k)}s_2^{(k)}) & \dots & 2(s_{d-1}^{(k)}s_d^{(k)}) \end{pmatrix} \in \mathbb{R}^{K\times D}. \tag{25}$$
With the choice of K in Remark 4, $K > D$, we approximate $\mathbf{Y}$ by solving an over-determined linear problem
$$\tilde{\mathbf{A}}\mathbf{Y} = 2\mathbf{D}^\top,$$
where $\tilde{\mathbf{A}}$ is defined as in (25) except that $s_i^{(j)}$ in the matrix entries is replaced by $\tilde{s}_i^{(j)}$. The regression solution is given by $\tilde{\mathbf{Y}} = (\tilde{\mathbf{A}}^\top\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^\top\mathbf{D}^\top$. Here $\tilde{\mathbf{Y}}_{ij} = \tilde{Y}_i^{(j)}$, $i = 1, \dots, D$, $j = 1, \dots, n$. Here, each row of $\tilde{\mathbf{Y}}$ is denoted as $\tilde{\mathbf{Y}}_p = (\tilde{Y}_p^{(1)}, \dots, \tilde{Y}_p^{(n)}) \in \mathbb{R}^{1\times n}$, which is an estimator of $\mathbf{Y}_p$.
4. Apply SVD to
$$2\tilde{\mathbf{R}}^\top := 2\mathbf{D} - (\tilde{\mathbf{A}}\tilde{\mathbf{Y}})^\top \in \mathbb{R}^{n\times K}. \tag{26}$$
Let the leading d left singular vectors be denoted as $\hat{\boldsymbol{\Phi}} = [\hat{\boldsymbol{\phi}}_1, \dots, \hat{\boldsymbol{\phi}}_d] \in \mathbb{R}^{n\times d}$, which is an estimator of $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_d]$, where $\boldsymbol{\phi}_j$ are the leading d left singular vectors of $\mathbf{R}$ as defined in (24). We define $\hat{\mathbf{P}} = \hat{\boldsymbol{\Phi}}\hat{\boldsymbol{\Phi}}^\top$ as the estimator for $\mathbf{P} = \boldsymbol{\Phi}\boldsymbol{\Phi}^\top$, where the last equality is valid due to Proposition 2.1(3).

Figure 2 shows the manifold learning results on a torus with randomly distributed data. One can see that error of the first-order local SVD method is $O(N^{-1/2})$ whereas second-order method is $O(N^{-1})$.

Theoretically, we can deduce the following error bound.

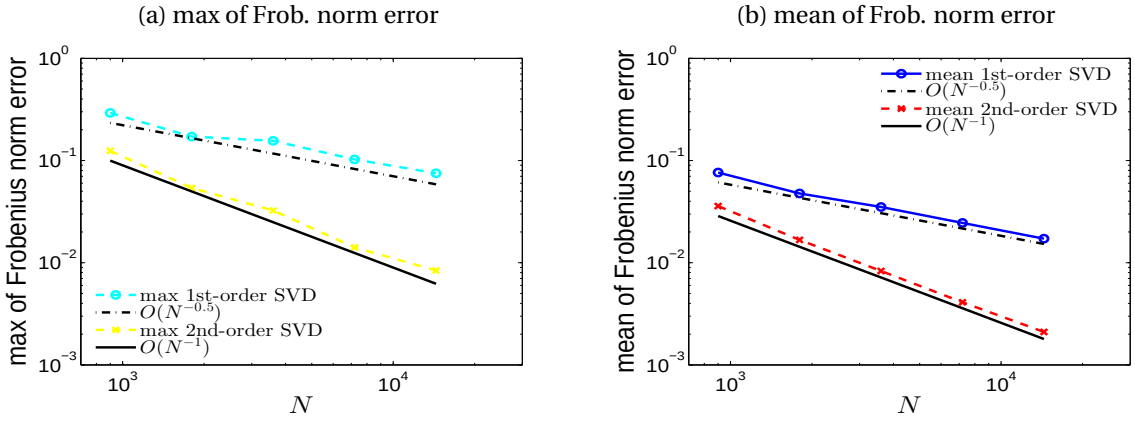

Figure 2: **2D torus in $\mathbb{R}^3$.** Comparison of convergence rates between 1st-order SVD and our 2nd-order SVD for approximating the tangential projection matrix $\mathbf{P}$. Panels (a) and (b) show the maximum and the mean of Frobenius norm errors, respectively. The error of $||\tilde{\mathbf{P}}-\mathbf{P}||_F$ is $O(N^{-1/2})$ for the 1st-order SVD whereas the error of $||\hat{\mathbf{P}}-\mathbf{P}||_F$ is $O(N^{-1})$ for our 2nd-order SVD. The $N = 3600$ data points are randomly distributed.

Theorem 3.2. *Let the assumptions in Theorem 3.1 be valid, particularly $s \sim \epsilon^{1/2}$. Suppose that the matrix $\mathbf{D} \in \mathbb{R}^{n\times K}$ is defined as in Step 1 of the algorithm with a fixed K chosen as in Remark 4 which implies that $K > D = \frac{1}{2}d(d+1)$ to ensure $\tau = \epsilon^{1/2} > 0$. Assume that $\kappa_2(\mathbf{A})\|\mathbf{A}\|_2 = K\omega(\epsilon^{3/2})$ as $\epsilon \to 0$. and the eigenvalues $\{\lambda_i\}_{i=1,\dots,n}$ of $\mathbf{R}^\top\mathbf{R}$, where $\mathbf{R}^\top \in \mathbb{R}^{n\times K}$ as defined in* (24)*, are simple with spectral gap $g_i := \min_{j\neq i}|\lambda_i - \lambda_j| > c\epsilon$ for some $c > 0$ and all $i = 1,\dots,n$. Let $\hat{\mathbf{P}} = \hat{\mathbf{\Phi}}\hat{\mathbf{\Phi}}^\top$ be the second-order estimator of $\mathbf{P}$, where columns of $\hat{\mathbf{\Phi}}$ are the leading d left singular vectors of $\tilde{\mathbf{R}}$ as defined in* (26)*. Then, with high probability,*

$$\left\|\hat{\mathbf{P}}-\mathbf{P}\right\|_F = O(\epsilon),$$

as $\epsilon \to 0$.

Before proving this result, we state a known result from perturbation theory. For convenience, we only rewrite a simplified version of Theorem 5.3.1 in [49] (i.e., here there is no perturbation on $\mathbf{D}$) for approximating $\mathbf{Y}^{(i)}$ that satisfies (each column of (24) in Step 3 of the algorithm),

$$\mathbf{A}\mathbf{Y}^{(i)} = 2\mathbf{D}^{(i)} - 2\mathbf{r}^{(i)} + O(\epsilon^{3/2}), \quad \forall i = 1,\dots,n, \tag{27}$$

by a least-squares solution to the perturbed problem,

$$\tilde{\mathbf{Y}}^{(i)} := \arg\min_{\mathbf{X}\in\mathbb{R}^D} \|\tilde{\mathbf{A}}\mathbf{X} - 2\mathbf{D}^{(i)}\|^2.$$

Here, the matrix $\tilde{\mathbf{A}} \in \mathbb{R}^{K\times D}$ is defined exactly as $\mathbf{A}$ in (25) with every $s_j^{(i)}$ in the matrix entries replaced by $\tilde{s}_j^{(i)}$, $\mathbf{D}^{(i)} \in \mathbb{R}^K$ denotes a column vector that is the ith row of $\mathbf{D}$, and $\mathbf{r}^{(i)} \in \mathbb{R}^K$ denotes the ith column vector of $\mathbf{R}$.

Lemma 3.1. *Suppose that $\mathbf{A}$ has full rank and $\delta := \frac{\|\tilde{\mathbf{A}}-\mathbf{A}\|_2}{\|\mathbf{A}\|_2} < \kappa_2(\mathbf{A})$, where $\kappa_2(\mathbf{A})$ denotes the condition number based on spectral matrix norm, $\|\cdot\|_2$. Then*

$$\frac{\|\tilde{\mathbf{Y}}^{(i)} - \mathbf{Y}^{(i)}\|}{\|\mathbf{Y}^{(i)}\|} \leq 2\kappa_{LS}(\mathbf{A})\delta.$$

where $\kappa_{LS} = \kappa_2(\mathbf{A}) + \kappa_2^2(\mathbf{A})\frac{2\|\mathbf{r}^{(i)}\|}{\|\mathbf{A}\|_2\|\mathbf{Y}^{(i)}\|}$.

This bound states that the least-squares condition number is roughly $\kappa_2(A)$ if the residual $\|\mathbf{r}^{(i)}\| = O(\epsilon^{1/2})$ is small such that the relative error is of $O(\kappa_2(A)^2)$. We are now ready to prove Theorem 3.2.

Proof. First, Eq. (22) implies that,

$$s_j^{(i)} = \mathbf{D}_i^\top \boldsymbol{\tau}_j + O(\epsilon),$$

where we set $s = O(\epsilon^{1/2})$. Since $\tilde{\mathbf{T}} = \mathbf{Q}\mathbf{T} + O(\epsilon^{1/2})$ for some orthogonal matrix $\mathbf{Q}$, we can write

$$s_j^{(i)} = \mathbf{D}_i^\top \boldsymbol{t}_j + O(\epsilon),$$

where $\boldsymbol{t}_j = \mathbf{Q}\boldsymbol{\tau}_j$, using the fact that components of $\mathbf{D}_i$ are also $O(\epsilon^{1/2})$. Based on Step 2 of the Algorithm above and Theorem 3.1, we have in high probability, for any $i = 1, \ldots, k$,

$$|\tilde{s}_j^{(i)} - s_j^{(i)}| = |\mathbf{D}_i^\top(\tilde{\boldsymbol{t}}_j - \boldsymbol{t}_j)| + O(\epsilon) \leq \|\mathbf{D}_i\|\|\tilde{\boldsymbol{t}}_j - \boldsymbol{t}_j\| + O(\epsilon) \leq c\epsilon^{1/2}\|\tilde{\mathbf{P}} - \mathbf{P}\|_F + O(\epsilon) \leq 2c\epsilon.$$

for some $c > 0$, as $\epsilon \to 0$, using the fact that $\|\mathbf{D}_i\| = O(\epsilon)$. This implies that,

$$|\tilde{s}_i^{(k)}\tilde{s}_j^{(k)} - s_i^{(k)}s_j^{(k)}| \leq |\tilde{s}_i^{(k)}||\tilde{s}_j^{(k)} - s_j^{(k)}| + |s_j^{(k)}||\tilde{s}_i^{(k)} - s_i^{(k)}| = O(\epsilon^{3/2}),$$

where we used again $|s_j| = O(\epsilon^{1/2})$. This means, $\|\mathbf{A} - \tilde{\mathbf{A}}\|_2 \leq \sqrt{KD}\|\mathbf{A} - \tilde{\mathbf{A}}\|_{max} = O(K\epsilon^{3/2})$.

We note that the components of each row of matrix $\mathbf{A}$ forms a homogeneous polynomial of degree-2, so they are linearly independent. Since the data points are uniformly sampled from $[-\sqrt{\epsilon}, \sqrt{\epsilon}]^d$, for large enough samples, $K > D$, these samples will not lie on a subspace of dimension strictly less than d. Thus, rank (A) = D almost surely.

Furthermore, there exists a constant $C > 0$ such that,

$$\frac{\|\mathbf{A} - \tilde{\mathbf{A}}\|_2}{\|\mathbf{A}\|_2} \leq \frac{CK\epsilon^{3/2}}{\|\mathbf{A}\|_2} < \kappa_2(\mathbf{A}),$$

where we used the assumption $\|\mathbf{A}\|_2\kappa_2(\mathbf{A}) = K\omega(\epsilon^{3/2})$, as $\epsilon \to 0$ for the last inequality. By Lemma 3.1, it is clear that,

$$\frac{\|\tilde{\mathbf{Y}}^{(i)} - \mathbf{Y}^{(i)}\|}{\|\mathbf{Y}^{(i)}\|} \leq \frac{2CK\epsilon^{3/2}\kappa_{LS}}{\|\mathbf{A}\|_2}.$$

Substracting each column of (26) from (27), we have

$$\begin{aligned} 2\|\tilde{\mathbf{r}}^{(i)} - \mathbf{r}^{(i)}\| &= \|\mathbf{A}\mathbf{Y}^{(i)} - \tilde{\mathbf{A}}\tilde{\mathbf{Y}}^{(i)}\| + O(\epsilon^{3/2}) \\ &\leq \|\mathbf{A}\|_2\|\mathbf{Y}^{(i)} - \tilde{\mathbf{Y}}^{(i)}\| + \|\tilde{\mathbf{Y}}^{(i)}\|\|\mathbf{A} - \tilde{\mathbf{A}}\|_2 + O(\epsilon^{3/2}) \\ &\leq 2CK\epsilon^{3/2}\kappa_{LS}\|\mathbf{Y}^{(i)}\| + CK\epsilon^{3/2}\|\mathbf{A}\|_2\|\mathbf{Y}^{(i)}\| + O(\epsilon^{3/2}) = O(\epsilon^{3/2}). \end{aligned}$$

By definition, each row of $\mathbf{R}$ is $\boldsymbol{r}_i \in \text{Span}\{\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_d\}$ and it is not difficult to deduce,

$$\|\mathbf{R}^\top\mathbf{R} - \tilde{\mathbf{R}}^\top\tilde{\mathbf{R}}\|_2 = O(\epsilon^2).$$

Let $\{\boldsymbol{\phi}_i\}_{i=1,\ldots,n}$ and $\{\tilde{\boldsymbol{\phi}}_i\}_{i=1,\ldots,n}$ be the unit eigenvectors of $\mathbf{R}^\top\mathbf{R}$ and $\tilde{\mathbf{R}}^\top\tilde{\mathbf{R}}$, respectively. By Theorem 5.4 of [16] and the assumption on the spectral gap of $\mathbf{R}^\top\mathbf{R}$, $g_i > c\epsilon$ for some $c > 0$, the acute angle $\boldsymbol{\phi}_i$ and $\tilde{\boldsymbol{\phi}}_i$ satisfies,

$$\sin(2\theta) \leq 2\frac{\|\mathbf{R}^\top\mathbf{R} - \tilde{\mathbf{R}}^\top\tilde{\mathbf{R}}\|_2}{g_i} = O(\epsilon).$$

Then,

$$\begin{aligned} \|\boldsymbol{\phi}_i - \tilde{\boldsymbol{\phi}}_i\|^2 &= \|\boldsymbol{\phi}_i\|^2 + \|\tilde{\boldsymbol{\phi}}_i\|^2 - 2\boldsymbol{\phi}_i^\top\tilde{\boldsymbol{\phi}}_i. \\ &= 2(1 - \cos(\theta)) = 2\left(1 - \sqrt{1 - \sin^2(\theta)}\right) = 2\left(1 - (1 - \frac{1}{2}\epsilon^2) + O(\epsilon^4)\right) = O(\epsilon^2). \end{aligned} \tag{28}$$

By Proposition 2.1(3), it is clear that $\mathbf{P} := \boldsymbol{\Phi}\boldsymbol{\Phi}^\top$. Based on the step 4 of the Algorithm, $\hat{\mathbf{P}} := \hat{\boldsymbol{\Phi}}\hat{\boldsymbol{\Phi}}^\top$ and from the error bound in (28), the proof is complete.

□

Remark 5. *The result above is consistent with the intuition that the scheme is of order $s^2 = O(\epsilon)$. The numerical result in the torus suggests of rate $\epsilon = O(N^{-1})$. For non-simple eigenvalues of $\mathbf{R}^\top\mathbf{R}$, one could follow a procedure similar to that used in [37] to show the the convergence of eigenvectors. For general d-dimensional manifolds, the error rate is expected to be $\epsilon \sim N^{-2/d}$, due to the fact that $s \propto N^{-1/d}$.*

4 Spectral Convergence for Laplace-Beltrami

In the present section, we prove spectral convergence results for estimators $\mathbf{G}^\top\mathbf{G}$ (symmetric) and $\mathbf{L}_N := -\sum_i \mathbf{G}_i\mathbf{G}_i$ (non-symmetric) to the Laplace-Beltrami operator Δ_M. We will see that the symmetry of the operator $\mathbf{G}^\top\mathbf{G}$ allows convergence of the eigenvectors of $\mathbf{G}^\top\mathbf{G}$ to be proved as well, while the convergence of eigenvectors of the non-symmetric estimator, though observed numerically, remains an open problem. The results and proof technique in

this section can be mimicked to show analogous results for the Bochner Laplacian acting on vector fields. For the pointwise analysis in each subsection below, we keep a general perspective that

$$\|I_{\phi_s} f - f\|_{L^2(M)} \leq \epsilon_f$$

with probability higher than $1-\delta_f$, in order to track how the approximation of differential operators depends on the interpolation error. For a detailed analysis of interpolation error in this setting, see [30]. Here is a slight adaptation of a result from [30], the proof of which can be found in Appendix B.

Lemma 4.1. *Let ϕ_s be a kernel whose RKHS norm is norm equivalent to a Sobolev space of order $\alpha > n/2$. Then with probability $1-\frac{1}{N}$, we have*

$$\|I_{\phi_s} f - f\|_{L^2(M)} = O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right),$$

for all $f \in C^\infty(M)$.

The result as stated in the appendix is slightly more general, and holds for all functions of sufficient Sobolev smoothness. However, our results require only C^∞ functions. The above result will be cited when deriving final rates and probabilities for the convergence of eigenvalues and eigenvectors. For the discrete approximation, we need to define an inner product over restricted functions such that it is consistent with the inner product of $L^2(M)$ as the number of data points N approaches infinity. In particular, we have the following definition.

Definition 4.1. *Given two vectors $\mathbf{f}, \mathbf{h} \in \mathbb{R}^N$, we define*

$$\langle \mathbf{f}, \mathbf{h} \rangle_{L^2(\mu_N)} := \frac{1}{N} \mathbf{f}^\top \mathbf{h}.$$

Similarly, we denote by $\|\cdot\|_{L^2(\mu_N)}$ the norm induced by the above inner-product.

We remark that when $\mathbf{f}$ and $\mathbf{h}$ are restrictions of functions f and h, respectively, then the above can be evaluated as $\langle \mathbf{f}, \mathbf{h} \rangle_{L^2(\mu_N)} = \frac{1}{N}\sum_{i=1}^N f(x_i)h(x_i)$.

The main results of this section are as follows. First, we have the following eigenvalue convergence result.

Theorem 4.1. *(convergence of eigenvalues: symmetric formulation) Let λ_i denote the i-th eigenvalue of Δ_M, enumerated $\lambda_1 \leq \lambda_2 \leq \ldots$, and let $1 \leq i < N$. Assume that $\mathbf{G}$ is as defined in* (6) *with interpolation operator that satisfies the hypothesis in Lemma 4.1. Then there exists an eigenvalue $\hat{\lambda}_i$ of $\mathbf{G}^\top \mathbf{G}$ such that*

$$\left|\lambda_i - \hat{\lambda}_i\right| \leq O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right),$$

with probability greater than $1-\frac{12}{N}$.

The first error term can be seen as coming from discretizing a continuous operator, while the second error term comes from the fact that continuous estimators in our setting differ from the true Laplace-Beltrami by precomposing with interpolation. The same rate holds true for convergence of eigenvectors, though in this case, the constants involved depend heavily on the multiplicity of the eigenvalue and several mild assumptions that we will state now.

Given a dataset of points $X = \{x_1, \ldots x_N\} \subseteq M$ and a function $f : M \to \mathbb{R}$, recall that the interpolation $I_{\phi_s} f$ only depends on $f(x_1), \ldots, f(x_N)$. Hence, I_{ϕ_s} can be viewed, after restricting to the manifold, as a map either from $\{f : M \to \mathbb{R}\} \to C^{\alpha - \frac{(n-d)}{2}}(M)$, or as a map

$$I_{\phi_s} : \mathbb{R}^N \to C^{\alpha - \frac{(n-d)}{2}}(M).$$

For details regarding the loss of regularity which occurs when restricting, see the beginning of Section 2.3 in [27]. In this section, our focus is continuous estimators, so we presently regard I_{ϕ_s} as a map

$$I_{\phi_s} : \{f : M \to \mathbb{R}\} \to C^{\alpha - \frac{(n-d)}{2}}(M).$$

In the remainder of this section, we need the following definition.

Definition 4.2. *Denote the L^2 error between $I_{\phi_s} f$ and f by ϵ_f. i.e.,*

$$\epsilon_f := \|I_{\phi_s} f - f\|_{L^2(M)}$$

For concentration bound, we will also define a parameter $0 \leq \delta_f \leq 1$ to probabilistically characterize an upper bound for ϵ_f. For example, Lemma 4.1 states that $\epsilon_f = O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right)$, with probability higher than $1-\delta_f$, where $\delta_f = 1/N$.

For the next spectral result, we consider the following assumptions, which are not necessary but simplify the error bounds. First, we define the formal adjoint of $\text{grad}_g I_{\phi_s}$ to be the unique linear operator $(\text{grad}_g I_{\phi_s})^* : \mathfrak{X}(M) \to C^\infty(M)$ satisfying

$$\langle (\text{grad}_g I_{\phi_s}) f, u \rangle_{L^2(\mathfrak{X}(M))} = \langle f, (\text{grad}_g I_{\phi_s})^* u \rangle_{L^2(M)}$$

for any $f \in C^\infty(M), u \in \mathfrak{X}(M)$. We are now ready to state the assumptions.

Assumption 4.1. *For a fixed $f \in C^\infty(M)$, we assume that for sufficiently large datasize $|X|$, the interpolation of $(\text{grad}_g I_{\phi_s})^*(\text{grad}_g I_{\phi_s}) f$ approximates $(\text{grad}_g I_{\phi_s})^*(\text{grad}_g I_{\phi_s}) f$ just as accurate as the interpolation of $\Delta_M f$ approximates $\Delta_M f$. i.e.,*

$$\begin{aligned} \epsilon_{(\text{grad}_g I_{\phi_s})^*(\text{grad}_g I_{\phi_s}) f} &:= \| I_{\phi_s}(\text{grad}_g I_{\phi_s})^*(\text{grad}_g I_{\phi_s}) f - (\text{grad}_g I_{\phi_s})^*(\text{grad}_g I_{\phi_s}) f \|_{L^2(M)} \\ &\leq C_1 \| I_{\phi_s} \Delta_M f - \Delta_M \|_{L^2(M)} := C_1 \epsilon_{\Delta_M f}, \end{aligned}$$

for some constant C_1 depending on f. As we will show in next subsection, the operator $I_{\phi_s}(\text{grad}_g I_{\phi_s})^*(\text{grad}_g I_{\phi_s})$ weakly converges to Δ_M. Hence, this assumption basically implies that the approximation error induced by the approximate operator $I_{\phi_s}(\text{grad}_g I_{\phi_s})^*(\text{grad}_g I_{\phi_s})$ is negligible compared to the interpolation error of the second-order differential operator, Δ_M.

Assumption 4.2. *There exists a constant $C_2 > 0$ depending only on $f \in C^\infty(M)$ and the fixed kernel ϕ_s such that,*

$$\left\| \Delta_M I_{\phi_s} \Delta_M f \right\|_{L^2(M)} + \left\| \Delta_M I_{\phi_s} (\text{grad}_g I_{\phi_s})^* (\text{grad}_g I_{\phi_s}) f \right\|_{L^2(M)} \leq C_2 \left\| \Delta_M^2 f \right\|_{L^2(M)}. \tag{29}$$

As we have alluded to in the previous discussion, we will show that $I_{\phi_s}(\text{grad}_g I_{\phi_s})^*(\text{grad}_g I_{\phi_s})$ weakly converges to Δ_M. This assumption implies that for sufficiently large N, the Laplacian of these functions are close to each other, i.e.,

$$\| \Delta_M I_{\phi_s} (\text{grad}_g I_{\phi_s})^* (\text{grad}_g I_{\phi_s}) f \|_{L^2(M)} \approx \| \Delta_M^2 f \|_{L^2(M)}.$$

Moreover, the inequality in (29) *also assumes that the Laplace-Beltrami of the interpolation error is small, i.e.,*

$$\| \Delta_M I_{\phi_s} \Delta_M f \|_{L^2(M)} \approx \| \Delta_M^2 f \|_{L^2(M)}.$$

With these assumptions, our result is the following.

Theorem 4.2. *Let $\epsilon_{\lambda_\ell} := |\lambda_\ell - \hat{\lambda}_\ell|$ denote the error in approximating the ℓ-th distinct eigenvalue, λ_ℓ, as defined in Theorem 4.1. Let Assumptions 4.1 and 4.2 be valid. For any ℓ, there is a constant c_ℓ such that whenever $\epsilon_{\lambda_{\ell-1}}, \epsilon_{\lambda_{\ell+1}} < c_\ell$, then with probability higher than $1 - \left(\frac{2m^2+4m+24}{N}\right)$, we have the following situation: for any normalized eigenvector $\mathbf{u}$ of $\mathbf{G}^\top \mathbf{G}$ with eigenvalue $\hat{\lambda}_\ell$, there is a normalized eigenfunction f of Δ_M with eigenvalue λ_ℓ such that,*

$$\| R_N f - \mathbf{u} \|_{L^2(\mu_N)} = O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right),$$

where m is the geometric multiplicity of eigenvalue λ_ℓ.

For the non-symmetric estimator, we additionally need an interpolation of vector-field result. Interpolation of vector fields in our settting is done componentwise in the ambient space coordinates. Applying Lemma 4.1 on each component, we obtain the following lemma. A proof can be found in Appendix B.

Lemma 4.2. *For any $u = u^i \frac{\partial}{\partial \theta^i} \in \mathfrak{X}(M)$, we have that with probability $1 - \frac{n}{N}$,*

$$\| u - I_{\phi_s} u \|_{L^2(\mathfrak{X}(M))} = O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right).$$

Before stating the theorem of spectral convergence in the non-symmetric setting, we must first introduce a sparseness coefficient k_i. Let $f_1, f_2, \ldots f_N$ be the first N smooth, orthonormal eigenfunctions of Δ_M. Notice that for $1 \leq i, j \leq N$,

$$\langle R_N f_i, R_N f_j \rangle_{L^2(\mu_N)} = \delta_{ij} + O\left(\frac{1}{\sqrt{N}}\right)$$

with high probability, since the eigenfunctions are orthonormal in $L^2(M)$. Perform Gram-Schmidt orthogonalization process on $\{R_N f_1, \ldots, R_N f_N\}$, and denote the resulting vectors by $\{\tilde{\mathbf{f}}_1, \ldots, \tilde{\mathbf{f}}_N\}$. We have the following useful definition.

Definition 4.3. *We define the i-th sparseness coefficient for the Laplace-Beltrami k_i to be the number of nonzero entries in the i-th column of $\mathbf{L}_N$ when written w.r.t. the basis $\{\tilde{\mathbf{f}}_i\}$. In Section 4.2.3, we will denote the matrix coefficients of $\mathbf{L}_N$ in this basis by $a_{ij} := \tilde{\mathbf{f}}_i^\top \mathbf{L}_N \tilde{\mathbf{f}}_j$.*

Consider the Gershgorin discs corresponding to the matrix $[a_{ij}]$, which is similar to $\mathbf{L}_N$, so they both have the same eigenvalues. Particularly, consider the N discs $\mathscr{C}_1,\dots,\mathscr{C}_N$, where $\mathscr{C}_j$ is centered at a_{jj} with radius $\sum_{k\neq j}|a_{kj}|$. It will be shown that for a fixed i, the center of $\mathscr{C}_i$ converges to λ_i as $N\to\infty$, and the radius converges to 0. Therefore, if the spectrum of Δ_M grows sufficiently fast, then one expects that for a fixed i, discs $\mathscr{C}_j$ with $j\gg i$ are disjoint from $\mathscr{C}_i$ since their centers are far away. Similarly, for j which is not much smaller or larger than i, then for sufficiently large N, since the radii of $\mathscr{C}_j$ and $\mathscr{C}_i$ approach 0 while the centers approach distinct values λ_j and λ_i, it follows that $\mathscr{C}_i$ and $\mathscr{C}_j$ are disjoint as well. This discussion motivates the following assumption.

Assumption 4.3. *Consider a fixed eigenvalue of Δ_M with multiplicity m. I.e., there is ℓ such that $\lambda_{\ell+1}=\cdots=\lambda_{\ell+m}$. We assume there is N sufficiently large so that $\cup_{j=1}^m \mathscr{C}_{\ell+j}$, the union of m Gershgorin discs of the matrix $[a_{ij}]$, is disjoint from the other $N-m$ discs.*

The eigenvalue convergence result for the non-symmetric formulation is formally stated in the following theorem.

Theorem 4.3. *Fix any eigenvalue $\lambda_{\ell+1}=\cdots=\lambda_{\ell+m}$ of multiplicity m. Additionally, let the Assumption 4.3 be valid. Then with high probability, there exists a sequence $\{\mu_N\}_{N=1}^\infty$, where μ_N is an eigenvalue of $\mathbf{L}_N$, such that*

$$|\lambda_{\ell+1}-\mu_N|\to 0 \text{ as } N\to\infty.$$

Before proving the main results for both symmetric and non-symmetric approximations, we prove the necessary pointwise and weak convergence results. It is important to note that the results of this section use analytic **P**. This allows us to conclude that, depending on the smoothness of the kernel, any error slower than Monte-Carlo convergence observed numerically is introduced through the approximation of **P**, as discussed in the previous section.

4.1 Symmetric Formulation

In this subsection, we study the consistency of the symmetric matrix,

$$\mathbf{G}^\top\mathbf{G}=\sum_{i=1}^n \mathbf{G}_i^\top\mathbf{G}_i,$$

as an approximation of the Laplace-Beltrami operator. We begin by focusing on the continuous (unrestricted) operator acting on a fixed, smooth function f, $\sum_i(\mathscr{G}_i I_{\phi_s})^*(\mathscr{G}_i I_{\phi_s})f$ and prove its convergence to $\Delta_M f$ with high probability. Such results depend on the accuracy with which I_{ϕ_s} can approximate f and its derivatives. We then quantify the error obtained when restricting to the dataset and constructing a matrix. Finally, we prove convergence of eigenvalues and eigenvectors. Since the estimator in this case is symmetric, convergence of eigenvalues requires only a weak convergence result. To prove convergence of eigenvectors, however, we need convergence of our estimator in L^2 sense.

4.1.1 Pointwise and weak convergence results: interpolation error

Using Cauchy-Schwarz, paried with the fact that the formal adjoint of grad_g is div_g, we immediately have the following Lemma.

Lemma 4.3. *Let $f\in C^\infty(M)$, and let $u\in\mathfrak{X}(M)$. Then with probability higher than $1-\delta_f$,*

$$\left|\langle \mathrm{grad}_g f-\mathrm{grad}_g I_{\phi_s}f,u\rangle_{L^2(\mathfrak{X}(M))}\right|\leq \epsilon_f\|\mathrm{div}_g(u)\|_{L^2(M)}.$$

A little more work also yields the following.

Lemma 4.4. *Let $f\in C^\infty(M)$. Then with probability higher than $1-\delta_f$,*

$$\left|\|\mathrm{grad}_g f\|^2_{L^2(\mathfrak{X}(M))}-\|\mathrm{grad}_g I_{\phi_s}f\|^2_{L^2(\mathfrak{X}(M))}\right|\leq \epsilon_f\left(\|\Delta_M f\|_{L^2(M)}+\|\Delta_M I_{\phi_s}f\|_{L^2(M)}\right).$$

Proof. Without ambiguity, we use the notation $\langle\cdot,\cdot\rangle$ for both inner products with respect to $L^2(M)$ and $L^2(\mathfrak{X}(M))$ in the derivations below to simplify the notation. We begin by adding and subtracting the mixed term $\langle\mathrm{grad}_g f,\mathrm{grad}_g I_{\phi_s}f\rangle$:

$$\begin{aligned}\left|\|\mathrm{grad}_g f\|^2-\langle\mathrm{grad}_g f,\mathrm{grad}_g I_{\phi_s}f\rangle+\langle\mathrm{grad}_g f,\mathrm{grad}_g I_{\phi_s}f\rangle-\|\mathrm{grad}_g I_{\phi_s}f\|^2\right| &\leq \\ \left|\|\mathrm{grad}_g f\|^2-\langle\mathrm{grad}_g f,\mathrm{grad}_g I_{\phi_s}f\rangle\right| &+ \left|\langle\mathrm{grad}_g f,\mathrm{grad}_g I_{\phi_s}f\rangle-\|\mathrm{grad}_g I_{\phi_s}f\|^2\right|.\end{aligned}$$

By the previous lemma, the first term is $\epsilon_f\|\Delta_M f\|_{L^2(M)}$, while the second term is $\epsilon_f\|\Delta_M I_{\phi_s}f\|_{L^2(M)}$. □

While results above allow us to prove a spectral convergence result, we will need the following result for studying the convergence of eigenvectors.

Corollary 4.1. *Let $f \in C^\infty(M)$. Then with probability higher than $1-\delta_f$,*

$$\|\mathrm{grad}_g f - \mathrm{grad}_g I_{\phi_s} f\|^2_{L^2(\mathfrak{X}(M))} \leq \epsilon_f \left(\|\Delta_M f\|_{L^2(M)} + \|\Delta_M I_{\phi_s} f\|_{L^2(M)}\right).$$

Proof. Expanding yields

$$\|\mathrm{grad}_g f - \mathrm{grad}_g I_{\phi_s} f\|^2_{L^2(\mathfrak{X}(M))} = \langle \mathrm{grad}_g f, \mathrm{grad}_g f\rangle - \langle \mathrm{grad}_g f, \mathrm{grad}_g I_{\phi_s} f\rangle - \langle \mathrm{grad}_g I_{\phi_s} f, \mathrm{grad}_g f\rangle + \langle \mathrm{grad}_g I_{\phi_s} f, \mathrm{grad}_g I_{\phi_s} f\rangle.$$

Grouping the first two and last two terms, the desired result is immediate. □

We now prove that $\Delta_M f$ can be weakly approximated, $\|\Delta_M f\|_{L^2(M)}$ can be approximated, and finally that Δ_M can be approximated in L^2 norm.

Lemma 4.5. *Let $f, h \in C^\infty(M)$. Then with probability higher than $1-\delta_f-\delta_h$,*

$$\left|\langle \Delta_M f, h\rangle_{L^2(M)} - \langle (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) f, h\rangle_{L^2(M)}\right| \leq \epsilon_h \|\Delta_M f\|_{L^2(M)} + \epsilon_f \|\Delta_M I_{\phi_s} h\|_{L^2(M)}.$$

Here $0<\delta_h<1$ and $\epsilon_h>0$ are defined exactly as in (4.2) *with f replaced by h.*

Proof. We again add and subtract a mixed term.

$$\begin{aligned}\langle \Delta_M f, h\rangle - \left\langle (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) f, h\right\rangle &= \langle \Delta_M f, h\rangle - \langle \mathrm{grad}_g f, \mathrm{grad}_g I_{\phi_s} h\rangle \\ &\quad + \langle \mathrm{grad}_g f, \mathrm{grad}_g I_{\phi_s} h\rangle - \left\langle (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) f, h\right\rangle\end{aligned}$$

The first two terms are bounded by $\epsilon_h \|\Delta_M f\|_{L^2(M)}$ while the second two terms are bounded by $\epsilon_f \|\Delta_M I_{\phi_s} h\|_{L^2(M)}$ □

Using similar arguments as above, we now have a convergence in L^2 norm result.

Lemma 4.6. *Let $f \in C^\infty(M)$ such that the Assumptions 4.1 and 4.2 are valid. Then for sufficiently large dataset X, with probability higher than $1-\delta_f-\delta_{\Delta_M f}$,*

$$\|\Delta_M f - (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) f\|^2_{L^2(M)} \leq \epsilon_f C_2 \|\Delta^2_M f\|_{L^2(M)} + 2C_1 \epsilon_{\Delta_M f} \|\Delta_M f\|_{L^2(M)}.$$

Proof. Expanding $\|\Delta_M f - (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) f\|^2_{L^2(M)}$ yields

$$\begin{aligned}\|\Delta_M f - (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) f\|^2_{L^2(M)} &= \langle \Delta_M f, \Delta_M f\rangle - \langle \Delta_M f, (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) f\rangle \\ &- \langle (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) f, \Delta_M f\rangle \\ &+ \langle (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) f, (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) f\rangle.\end{aligned}$$

We use the previous Lemma twice, once on the first two terms (which gives an error of $\epsilon_{\Delta_M f}\|\Delta_M f\|_{L^2(M)} + \epsilon_f \|\Delta_M I_{\phi_s} \Delta_M f\|_{L^2(M)}$) and once on the last two terms (which gives an error of $\epsilon_{(\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) f}\|\Delta_M f\|_{L^2(M)} + \epsilon_f \|\Delta_M I_{\phi_s} (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) f\|_{L^2(M)}$). Using the inequalities in the Assumptions 4.1 and 4.2, the proof is complete. □

4.1.2 Pointwise and weak convergence results: empirical error

We now quantify the error obtained when discretizing our estimators on the dataset X. The results of this section are primarily based on the law of large numbers. The following Lemma is a direct consequence of a standard concentration result.

Lemma 4.7. *Let $f, h \in C^\infty(M)$. Then*

$$\mathbb{P}_X\left(\left|\langle \mathbf{Gf}, \mathbf{Gh}\rangle_{L^2(\mu_N)} - \langle \mathrm{grad}_g I_{\phi_s} f, \mathrm{grad}_g I_{\phi_s} h\rangle_{L^2(\mathfrak{X}(M))}\right| \geq \epsilon\right) \leq 2\exp\left(\frac{-2\epsilon^2 N}{D^2}\right),$$

where D is a constant depending on the kernel ϕ_s, as well as f and h.

Proof. Using the fact that $\text{grad}_g I_{\phi_s} f = (\mathscr{G}_1 I_{\phi_s} f, \mathscr{G}_2 I_{\phi_s} f, \ldots, \mathscr{G}_n I_{\phi_s} f)^\top$ as defined in (4), it is clear that,

$$\langle \text{grad}_g I_{\phi_s} f, \text{grad}_g I_{\phi_s} f \rangle_{L^2(\mathfrak{X}(M))} = \int_M \sum_{i=1}^n \left(\mathscr{G}_i I_{\phi_s} f(x)\right)\left(\mathscr{G}_i I_{\phi_s} h(x)\right) d\text{Vol}(x),$$

and we can see immediately that the result follows from a concentration inequality on the random variable $\sum_{i=1}^n \left(\mathscr{G}_i I_{\phi_s} f(x)\right)\left(\mathscr{G}_i I_{\phi_s} h(x)\right)$. The range of I_{ϕ_s} is in $C^{\alpha-\frac{(n-d)}{2}}(M)$, that is, $I_{\phi_s} f, I_{\phi_s} h$ are $C^{\alpha-\frac{(n-d)}{2}}(M)$, and since $\mathscr{G}_i$ are simply differential operators, we see that the random variable is bounded away from the mean by a constant D, depending on the kernel ϕ_s. The desired result follows immediately by Hoeffding's inequality. □

Since $\mathbf{G}$ is simply the restricted version of $\text{grad}_g I_{\phi_s}$, the same reasoning as above gives the following result, which will be needed to prove convergence of eigenvectors.

Lemma 4.8. *Let $f \in C^\infty(M)$. Then*

$$\mathbb{P}_X\left(\left|\left\|\mathbf{G}^\top \mathbf{G}\mathbf{f} - R_N \Delta_M f\right\|_{L^2(\mu_N)}^2 - \left\|(\text{grad}_g I_{\phi_s})^* (\text{grad}_g I_{\phi_s}) f - \Delta_M f\right\|_{L^2(M)}^2\right| \geq \epsilon\right) \leq 2\exp\left(\frac{-2\epsilon^2 N}{C^2}\right)$$

where C is a constant depending on ϕ and f.

4.1.3 Proof of Spectral Convergence

Here, we prove Theorem 4.1, the convergence of eigenvalues result for the Laplace-Beltrami operator.

Proof. Enumerate the eigenvalues of $\mathbf{G}^\top\mathbf{G}$ and label them $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \cdots \leq \hat{\lambda}_N$. Let $\mathscr{S}_i' \subseteq C^\infty(M)$ denote an i-dimensional subspace of smooth functions on which the quantity $\max_{f\in\mathscr{S}_i} \frac{\langle \mathbf{G}^\top\mathbf{G} R_N f, R_N f\rangle_{L^2(\mu_N)}}{\|R_N f\|_{L^2(\mu_N)}}$ achieves its minimum. Let $\tilde{f} \in \mathscr{S}_i'$ be the function on which the maximum $\max_{f\in\mathscr{S}_i'} \langle \Delta_M f, f\rangle_{L^2(M)}$ occurs. WLOG, assume that $\|\tilde{f}\|_{L^2(M)}^2 = 1$. Assume that N is sufficiently large so that by Hoeffding's inequality $\left|\|R_N\tilde{f}\|_{L^2(\mu_N)}^2 - 1\right| \leq \frac{\text{Const}}{\sqrt{N}} \leq 1/2$, with probability $1-\frac{2}{N}$, so that $\|R_N\tilde{f}\|_{L^2(\mu_N)}^2$ is bounded away from zero. Hence, we can Taylor expand $\frac{\langle \mathbf{G}^\top\mathbf{G} R_N\tilde{f}, R_N\tilde{f}\rangle_{L^2(\mu_N)}}{\|R_N\tilde{f}\|_{L^2(\mu_N)}^2}$ to obtain

$$\frac{\langle \mathbf{G}^\top\mathbf{G} R_N\tilde{f}, R_N\tilde{f}\rangle_{L^2(\mu_N)}}{\|R_N\tilde{f}\|_{L^2(\mu_N)}^2} = \langle \mathbf{G}^\top\mathbf{G} R_N\tilde{f}, R_N\tilde{f}\rangle_{L^2(\mu_N)} - \frac{\text{Const}\langle \mathbf{G}^\top\mathbf{G} R_N\tilde{f}, R_N\tilde{f}\rangle_{L^2(\mu_N)}}{\sqrt{N}}.$$

By Lemma 4.7, with probability higher than $1-\frac{2}{N}$, we have that

$$\left|\langle \mathbf{G}^\top\mathbf{G} R_N\tilde{f}, R_N\tilde{f}\rangle_{L^2(\mu_N)} - \langle \text{grad}_g I_{\phi_s}\tilde{f}, \text{grad}_g I_{\phi_s}\tilde{f}\rangle_{L^2(\mathfrak{X}(M))}\right| = O\left(\frac{1}{\sqrt{N}}\right),$$

where we have chosen $\epsilon = \sqrt{\frac{\log(N)}{N}}$ and ignored the log factor.

Combining the two bounds above with Lemma 4.5 and Lemma 4.1, we obtain that

$$\langle \Delta_M\tilde{f}, \tilde{f}\rangle_{L^2(M)} \leq \frac{\langle \mathbf{G}^\top\mathbf{G} R_N\tilde{f}, R_N\tilde{f}\rangle_{L^2(\mu_N)}}{\|R_N\tilde{f}\|_{L^2(\mu_N)}^2} + O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right),$$

with probability higher than $1-\frac{6}{N}$. Since $\tilde{f}$ is the function on which $\langle \Delta_M f, f\rangle_{L^2(M)}$ achieves its maximum over all $f \in \mathscr{S}_i'$, and since certainly

$$\frac{\langle \mathbf{G}^\top\mathbf{G} R_N\tilde{f}, R_N\tilde{f}\rangle_{L^2(\mu_N)}}{\|R_N\tilde{f}\|_{L^2(\mu_N)}^2} \leq \max_{f\in\mathscr{S}_i'} \frac{\langle \mathbf{G}^\top\mathbf{G} R_N f, R_N f\rangle_{L^2(\mu_N)}}{\|R_N f\|_{L^2(\mu_N)}^2},$$

we have the following:

$$\max_{f\in\mathscr{S}_i'} \langle \Delta_M f, f\rangle_{L^2(M)} \leq \max_{f\in\mathscr{S}_i'} \frac{\langle \mathbf{G}^\top\mathbf{G} R_N f, R_N f\rangle_{L^2(\mu_N)}}{\|R_N f\|_{L^2(\mu_N)}^2} + O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right).$$

But we assumed that $\mathscr{S}_i'$ is the exact subspace on which $\max_{f\in\mathscr{S}_i} \frac{\langle \mathbf{G}^\top \mathbf{G} R_N f, R_N f\rangle_{L^2(\mu_N)}}{\|R_N f\|^2_{L^2(\mu_N)}}$ achieves its minimum. Hence,

$$\max_{f\in\mathscr{S}_i'} \langle \Delta_M f, f\rangle_{L^2(M)} \leq \hat{\lambda}_i + O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right).$$

But the left-hand-side certainly bounds from above by the minimum of $\max_{f\in\mathscr{S}_i} \langle \Delta_M f, f\rangle_{L^2(M)}$ over all i-dimensional smooth subspaces $\mathscr{S}_i$. Hence,

$$\lambda_i \leq \hat{\lambda}_i + O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right).$$

The same argument yields that $\hat{\lambda}_i \leq \lambda_i + O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right)$, with probability higher than $1-\frac{6}{N}$. This completes the proof. □

4.1.4 Proof of Eigenvector Convergence

We now prove the convergence of eigenvectors. The outline of this proof follows the arguments in the convergence analysis found in [37] and [8]. It is important to note that since the matrix $\mathbf{G}^\top\mathbf{G}$ is symmetric, there exists an orthonormal basis of eigenvectors of $\mathbf{G}^\top\mathbf{G}$, which is key in the following proof of Theorem 4.2.

Proof. Fix any $\ell \in \mathbb{N}$. For convenience, we let ϵ_{λ_ℓ} denote the error in approximating the ℓ-th eigenvalue, from the previous section. Similarly, we let δ_{λ_ℓ} denote the quantity such that eigenvalue approximation occurs with probability higher than $1-\delta_{\lambda_\ell}$. Let m be the geometric multiplicity of the eigenvalue λ_ℓ, i.e., there is an i such that $\lambda_{i+1} = \lambda_{i+2} = \cdots = \lambda_\ell = \cdots = \lambda_{i+m}$. Let

$$c_\ell = \frac{1}{2}\min\left\{|\lambda_\ell - \lambda_i|, |\lambda_\ell - \lambda_{i+m+1}|\right\}.$$

By Theorem 4.1, if $\epsilon_{\lambda_i}, \epsilon_{\lambda_{i+m+1}} < c_\ell$, then with probability $1-\delta_{\lambda_i} - \delta_{\lambda_{i+m+1}}$,

$$|\hat{\lambda}_i - \lambda_i| < c_\ell, \quad |\hat{\lambda}_{i+m+1} - \lambda_{i+m+1}| < c_\ell.$$

Let $\hat{\mathbf{u}}_1, \ldots \hat{\mathbf{u}}_N$ be an orthonormal basis of $L^2(\mu_N)$ consisting of eigenvectors of $\mathbf{G}^\top\mathbf{G}$, where $\hat{\mathbf{u}}_j$ has eigenvalue $\hat{\lambda}_j$. Let S be the m dimensional subspace of $L^2(\mu_N)$ corresponding to the span of $\{\hat{\mathbf{u}}_j\}_{j=i+1}^{i+m}$, and let P_S (resp. $P_S^\perp$) denote the projection onto S (resp. orthogonal complement of S). Let f be a norm 1 eigenfunction of Δ_M corresponding to eigenvalue λ_ℓ. Notice that

$$P_S^\perp R_N \Delta_M f = \lambda_\ell P_S^\perp R_N f = \lambda_\ell \sum_{j\neq i+1,\ldots,i+m} \langle R_N f, \hat{\mathbf{u}}_j\rangle_{L^2(\mu_N)} \hat{\mathbf{u}}_j.$$

Similarly,

$$P_S^\perp \mathbf{G}^\top \mathbf{G} R_N f = \sum_{j\neq i+1,\ldots,i+m} \hat{\lambda}_j \langle R_N f, \hat{\mathbf{u}}_j\rangle_{L^2(\mu_N)} \hat{\mathbf{u}}_j.$$

Hence,

$$\begin{aligned}
\left\| P_S^\perp R_N \Delta_M f \quad - \quad P_S^\perp \mathbf{G}^\top \mathbf{G} R_N f \right\|_{L^2(\mu_N)} &= \left\| \sum_{j\neq i+1,\ldots,i+m} (\lambda_\ell - \hat{\lambda}_j) \langle R_N f, \hat{\mathbf{u}}_j\rangle_{L^2(\mu_N)} \hat{\mathbf{u}}_j \right\|_{L^2(\mu_N)} \\
&\geq \min\left\{|\lambda_\ell - \hat{\lambda}_i|, |\lambda_\ell - \hat{\lambda}_{i+m+1}|\right\} \left\| \sum_{j\neq i+1,\ldots,i+m} \langle R_N f, \hat{\mathbf{u}}_j\rangle_{L^2(\mu_N)} \hat{\mathbf{u}}_j \right\|_{L^2(\mu_N)} \\
&\geq \min\left\{|\lambda_\ell - \hat{\lambda}_i|, |\lambda_\ell - \hat{\lambda}_{i+m+1}|\right\} \left\| P_S^\perp R_N f \right\|_{L^2(\mu_N)}.
\end{aligned}$$

But $P_S^\perp$ is an orthogonal projection, so

$$\begin{aligned}
\min\left\{|\lambda_\ell - \hat{\lambda}_i|, |\lambda_\ell - \hat{\lambda}_{i+m+1}|\right\} \|P_S^\perp R_N f\|_{L^2(\mu_N)} &\leq \left\| P_S^\perp R_N \Delta_M f - P_S^\perp \mathbf{G}^\top \mathbf{G} R_N f \right\|_{L^2(\mu_N)} \\
&\leq \|R_N \Delta_M f - \mathbf{G}^\top \mathbf{G} R_N f\|_{L^2(\mu_N)}.
\end{aligned}$$

Without loss of generality, assume that $\min\left\{|\lambda_\ell - \hat{\lambda}_i|, |\lambda_\ell - \hat{\lambda}_{i+m+1}|\right\} = |\lambda_\ell - \hat{\lambda}_i|$. Notice that

$$|\lambda_\ell - \hat{\lambda}_i| \geq \left||\lambda_\ell - \lambda_i| - |\lambda_i - \hat{\lambda}_i|\right| > c_\ell,$$

by the hypothesis. Hence,

$$\|P_S^\perp R_N f\|^2_{L^2(\mu_N)} \leq \frac{1}{c_\ell^2}\|R_N\Delta_M f - \mathbf{G}^\top \mathbf{G} R_N f\|^2_{L^2(\mu_N)}.$$

By Lemma 4.8 paired with Lemma 4.6, this upper bound is smaller than $O\left(\frac{1}{\sqrt{N}}\right) + \epsilon_f C_f \|\Delta_M^2 f\|_{L^2(M)} + 2\epsilon_{\Delta_M f}\|\Delta_M f\|_{L^2(M)}$ with probability higher than $1 - \frac{2}{N} - \delta_f - \delta_{\Delta_M f}$. Notice that $P_S^\perp R_N f = R_N f - P_S R_N f$. Hence, if $\{f_1, f_2, \ldots, f_m\}$ are an orthonormal basis for the eigenspace corresponding to λ_ℓ, applying the above reasoning m times, we see that with a total probability of $1 - \frac{2}{N} - \delta_{f_1} - \delta_{\Delta_M f_1} - \cdots - \frac{2}{N} - \delta_{f_m} - \delta_{\Delta_M f_m}$,

$$\|R_N f_j - P_S R_N f_j\|^2_{L^2(\mu_N)} = \frac{1}{c_\ell^2}\left(O\left(\frac{1}{\sqrt{N}}\right) + \epsilon_{f_j} C_{f_j}\|\Delta_M^2 f_j\|_{L^2(M)} + 2\epsilon_{\Delta_M f_j}\|\Delta_M f_j\|_{L^2(M)}\right), \qquad \text{for } j = 1, 2, \ldots m.$$

Let C_ℓ denote an upper bound on the essential supremum of the eigenvectors $\{f_1, f_2, \ldots, f_m\}$. For any i, j,

$$\left| f_i(x) f_j(x) - \int_M f_j(y) f_i(y) d\mu(y) \right| \leq C_\ell^2 (1 + \mathrm{Vol}(M)).$$

Hence, using Hoeffding's inequality with $\alpha = 2\sqrt{2} C_\ell \sqrt{\frac{\log(N)}{N}}$, with probability $1 - \frac{2}{N}$,

$$\left| \frac{1}{N}\sum_{l=1}^N f_i(x_l) f_j(x_l) - \int_{\mathscr{M}} f_i(y) f_j(y) d\mu(y) \right| < \alpha.$$

Since $\{f_1, \ldots, f_m\}$ are orthonormal in $L^2(M)$, by Hoeffding's inequality used m^2 times, we see that probability higher than $1 - \frac{2m^2}{N}$,

$$\langle R_N f_i, R_N f_j \rangle_{L^2(\mu_N)} = \delta_{ij} + O\left(\frac{1}{\sqrt{N}}\right).$$

Hence, with a total probability higher than $1 - \delta_{\lambda_i} - \delta_{\lambda_{i+m+1}} - \frac{2m^2}{N} - \frac{2}{N} - \delta_{f_1} - \delta_{\Delta_M f_1} - \cdots - \frac{2}{N} - \delta_{f_m} - \delta_{\Delta_M f_m}$ we have that

$$\begin{aligned}\langle P_S R_N f_i, P_S R_N f_j \rangle_{L^2(\mu_N)} &= \langle R_N f_i, R_N f_j \rangle_{L^2(\mu_N)} - \langle R_N f_i - P_S R_N f_i, R_N f_j - P_S R_N f_j \rangle_{L^2(\mu_N)} \\ &= \delta_{ij} + O\left(\frac{1}{\sqrt{N}}\right) + \sqrt{\mathrm{Error}(i)}\sqrt{\mathrm{Error}(j)},\end{aligned}$$

where $\mathrm{Error}(j) = \frac{1}{c_\ell^2}\left(O\left(\frac{1}{\sqrt{N}}\right) + \epsilon_{f_j} C_{f_j}\|\Delta_M^2 f_j\|_{L^2(M)} + 2\epsilon_{\Delta_M f_j}\|\Delta_M f_j\|_{L^2(M)}\right)$. Letting $\mathbf{v}_1 = \frac{P_S R_N f_1}{\|P_S R_N f_1\|_{L^2(\mu_N)}}$, we see that

$$\|P_S R_N f_1 - \mathbf{v}_1\|^2_{L^2(\mu_N)} = O\left(1/\sqrt{N}\right) + O\left(\mathrm{Error}(1)\right).$$

Similarly, letting $\tilde{\mathbf{v}}_2 = P_S R_N f_2 - \frac{\langle P_S R_N f_1, P_S R_N f_2\rangle_{L^2(\mu_N)}}{\|P_S R_N f_1\|^2_{L^2(\mu_N)}} P_S R_N f_1$, and $\mathbf{v}_2 = \frac{\tilde{\mathbf{v}}_2}{\|\tilde{\mathbf{v}}_2\|_{L^2(\mu_N)}}$ it is easy to see that

$$\|P_S R_N f_2 - \tilde{\mathbf{v}}_2\|^2_{L^2(\mu_N)} = O\left(1/\sqrt{N}\right) + O\left(\mathrm{Error}(2)\right),$$

and hence,

$$\|P_S R_N f_2 - \mathbf{v}_2\|^2_{L^2(\mu_N)} = O\left(1/\sqrt{N}\right) + O\left(\mathrm{Error}(2)\right).$$

Continuing in this way, we see that the Gram-Schmidt procedure on $\{P_S R_N f_j\}_{j=1}^m$, yields an orthonormal set of m vectors $\{\mathbf{v}_j\}_{j=1}^m$ spanning S such that

$$\|P_S R_N f_j - \mathbf{v}_j\|^2_{L^2(\mu_N)} = O\left(1/\sqrt{N}\right) + O\left(\mathrm{Error}(j)\right), \qquad j = 1, 2, \ldots, m.$$

and therefore

$$\begin{aligned}\|R_N f_j - \mathbf{v}_j\|_{L^2(\mu_N)} &\leq \|R_N f_j - P_S R_N f_j\|_{L^2(\mu_N)} + \|P_S R_N f_j - \mathbf{v}_j\|_{L^2(\mu_N)} \\ &= 2\sqrt{O\left(1/\sqrt{N}\right) + O\left(\mathrm{Error}(j)\right)}, \qquad j = 1, 2, \ldots, m,\end{aligned}$$

Therefore, for any eigenvector $\mathbf{u} = \sum_{j=1}^m b_j \mathbf{v}_j$ with $L^2(\mu_N)$ norm 1, notice that $f = \sum_{j=1}^m b_j f_j$ is a $L^2(M)$ norm 1 eigenfunction of Δ_M with eigenvalue λ_ℓ. Indeed,

$$\Delta_M f = \sum_{j=1}^m b_j \Delta_M f_j = \lambda_\ell f,$$

and

$$\|f\|^2_{L^2(M)} = \langle f, f\rangle_{L^2(M)} = \sum_{i=1}^{m}\sum_{j=1}^{m} b_i b_j \langle f_j, f_i\rangle_{L^2(M)} = \sum_{j=1}^{m} b_j^2 = 1,$$

where the last equality follows from the fact that $\|\mathbf{u}\|^2_{L^2(\mu_N)} = \|\sum_{i=1}^m b_i \mathbf{v}_i\|^2_{L^2(\mu_N)} = \sum_{i=1}^m b_i^2 = 1$. Moreover, the function f also satisfies

$$\|R_N f - \mathbf{u}\|^2_{L^2(\mu_N)} \leq \sum_{j=1}^{m} |b_j|^2 \|R_N f_j - \mathbf{v}_j\|^2_{L^2(\mu_N)} = O\left(1/\sqrt{N}\right) + \sum_{j=1}^{m} O\big(\text{Error}(j)\big).$$

Using Lemma 4.1 and collecting the probabilities, the above holds with probability higher than $1 - \left(\frac{2m^2+4m+24}{N}\right)$. Moreover, each Error$(j)$ is on the order of $O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right)$. This completes the proof. □

4.2 Non-symmetric Formulation

In this subsection, we present the spectral convergence study for an approximation of the Laplace-Beltrami operator approximated by the non-symmetric matrix

$$\mathbf{L}_N := -\sum_{i=1}^{n} \mathbf{G}_i \mathbf{G}_i.$$

We begin by focusing on the continuous (unrestricted) operator acting on a fixed, smooth function f, $-\text{div}_g\left(I_{\phi_s}\text{grad}_g I_{\phi_s} f\right)$ and prove its convergence to $\Delta_M f$ with high probability. Analogous to the previous subsection, these results depend on the accuracy with which I_{ϕ_s} can approximate f and its derivatives. We also quantify the error induced by restriction to the dataset. Finally, we prove convergence of eigenvalues in this setting. The estimator in this case is non-symmetric, and the techniques involved in the proof are considerably different from those in the previous section.

4.2.1 Pointwise and weak convergence results: interpolation error

We will use the same notation as before, that is, for a smooth function f, we denote by ϵ_f the L^2 error between f and $I_{\phi_s} f$, which holds with probability higher than $1-\delta_f$. Similarly, for a vector field u, we denote by ϵ_u the $L^2(\mathfrak{X}(M))$ error between $I_{\phi_s} u$ and u, which holds with probability higher than $1-\delta_u$. Recall that the interpolation for a vector field is performed component-wise in the ambient space coordinates.

Proposition 4.1. *Let $f, h \in C^\infty(M)$. Then with probability greater than $1-\delta_f-\delta_{\text{grad}_g I_\phi f}$,*

$$\left|\left\langle \Delta_M f + \text{div}_g\left(I_{\phi_s}\text{grad}_g I_{\phi_s} f\right), h\right\rangle_{L^2(M)}\right| \leq \epsilon_f \|h\|_{L^2(M)} + \epsilon_{\text{grad}_g I_{\phi_s} f}\|\text{grad}_g h\|_{L^2(\mathfrak{X}(M))}.$$

Proof. We note that,

$$\begin{aligned}\left\langle \Delta_M f + \text{div}_g\left(I_{\phi_s}\text{grad}_g I_{\phi_s} f\right), h\right\rangle_{L^2(M)} &= \langle -\text{grad}_g f + (I_{\phi_s}\text{grad}_g I_{\phi_s} f), \text{grad}_g h\rangle_{L^2(\mathfrak{X}(M))} \\ &= \left\langle -\text{grad}_g f + \text{grad}_g I_{\phi_s} f, \text{grad}_g h\right\rangle_{L^2(\mathfrak{X}(M))} \\ &\quad + \left\langle -\text{grad}_g I_{\phi_s} f + I_{\phi_s}\text{grad}_g I_{\phi_s} f, \text{grad}_g h\right\rangle_{L^2(\mathfrak{X}(M))}.\end{aligned}$$

Notice that the first term is $\langle -\text{grad}_g f + \text{grad}_g I_{\phi_s} f, \text{grad}_g h\rangle_{L^2(\mathfrak{X}(M))} = \langle -f + I_{\phi_s} f, \Delta_M h\rangle_{L^2(M)}$. By the interpolation result, this is bounded by $\epsilon_f \|\Delta_M h\|_{L^2(M)}$ with probability $1-\delta_f$. Similarly, the second term is simply the interpolation error on the vector field $\text{grad}_g I_{\phi_s} f$ times $\|\text{grad}_g h\|_{L^2(\mathfrak{X}(M))}$. This completes the proof. □

Proposition 4.2. *Let $f \in C^\infty(M)$. Then with probability higher than $1-2\delta_f - 2\delta_{\text{grad}_g I_{\phi_s} f}$,*

$$\left\|\Delta_M f + \text{div}_g\left(I_{\phi_s}\text{grad}_g I_{\phi_s} f\right)\right\|^2_{L^2(M)} = \epsilon_f\left(\|\psi_1\|_{L^2(M)} + \|\psi_2\|_{L^2(M)}\right) + \epsilon_{\text{grad}_g I_{\phi_s} f}\left(\|\text{grad}_g \psi_1\|_{L^2(\mathfrak{X}(M))} + \|\text{grad}_g \psi_2\|_{L^2(\mathfrak{X}(M))}\right)$$

where $\psi_1 = \Delta_M f$ and $\psi_2 = \text{div}_g\left(I_{\phi_s}\text{grad}_g I_{\phi_s} f\right)$.

Proof. Notice that

$$\begin{aligned}\left\|\Delta_M f + \operatorname{div}_g\left(I_{\phi_s}\operatorname{grad}_g I_{\phi_s} f\right)\right\|^2_{L^2(M)} &= \left\langle \Delta_M f + \operatorname{div}_g\left(I_{\phi_s}\operatorname{grad}_g I_{\phi_s} f\right), \Delta_M f\right\rangle_{L^2(M)} \\ &\quad + \left\langle \Delta_M f + \operatorname{div}_g\left(I_{\phi_s}\operatorname{grad}_g I_{\phi_s} f\right), \operatorname{div}_g\left(I_{\phi_s}\operatorname{grad}_g I_{\phi_s} f\right)\right\rangle_{L^2(M)} \\ &\leq \left|\left\langle \Delta_M f + \operatorname{div}_g\left(I_{\phi_s}\operatorname{grad}_g I_{\phi_s} f\right), \Delta_M f\right\rangle_{L^2(M)}\right| \\ &\quad + \left|\left\langle \Delta_M f + \operatorname{div}_g\left(I_{\phi_s}\operatorname{grad}_g I_{\phi_s} f\right), \operatorname{div}_g\left(I_{\phi_s}\operatorname{grad}_g I_{\phi_s} f\right)\right\rangle_{L^2(M)}\right|.\end{aligned}$$

Using Proposition 4.1 twice, once with $h = \Delta_M f$, and once with $h = \operatorname{div}_g\left(I_{\phi_s}\operatorname{grad}_g I_{\phi_s} f\right)$, we obtain the desired result. □

4.2.2 Pointwise and weak convergence results: empirical error

Using the same methods as Section 4.1.2, on the function $-\operatorname{div}_g\left(I_{\phi_s}\operatorname{grad}_g I_{\phi_s} f\right)$, we obtain the following restriction result, which is an immediate consequence of Hoeffding's inequality.

Lemma 4.9. *Let $f \in C^\infty(M)$. Then we have that*

$$\mathbb{P}_X\left(\left|\|\Delta_M f + \operatorname{div}_g\left(I_{\phi_s}\operatorname{grad}_g I_{\phi_s} f\right)\|^2_{L^2(M)} - \|R_N\Delta_M f - \mathbf{L}_N R_N f\|^2_{L^2(\mu_N)} \geq \epsilon\right|\right) \leq 2\exp\left(\frac{-2\epsilon^2 N}{D^2}\right)$$

where the constant D depends on f and the kernel ϕ_s.

Given N smooth linearly independent eigenfunctions $f_1, \ldots, f_N$ of the Laplace-Beltrami operator with eigenvalues $\lambda_1, \lambda_2, \ldots \lambda_N$, it is clear that the restriction $R_N f_1, \ldots, R_N f_N$ are linearly independent, and hence, form a basis of $\mathbb{R}^N$. In this case, we can define a matrix on the basis $\{R_N f_1, \ldots, R_N f_N\}$ by

$$\Delta_M^{(N)} R_N f_i = \lambda_i R_N f_i.$$

Letting $V = [R_N f_1, \ldots R_N f_N]$, and $\Lambda = \operatorname{diag}(\lambda_1, \ldots \lambda_N)$ we see that

$$(\Delta_M^{(N)})_{ij} = (V\Lambda V^{-1})_{ij}.$$

We have the following theorem regarding the distance between the matrices $\mathbf{L}_N$ and $\Delta_M^{(N)}$.

Proposition 4.3. *Let $f \in C^\infty(M)$. Denote $\mathbf{f} := R_N f$. With probability greater than $1 - \frac{4}{N} - 2\delta_f - 2\delta_{\operatorname{grad}_g I_{\phi_s} f}$,*

$$\begin{aligned}\|\Delta_M^{(N)}\mathbf{f} - \mathbf{L}_N\mathbf{f}\|^2_{L^2(\mu_N)} &= O\left(1/\sqrt{N}\right) + \epsilon_f\left(\|\psi_1\|_{L^2(M)} + \|\psi_2\|_{L^2(M)}\right) \\ &\quad + \epsilon_{\operatorname{grad}_g I_{\phi_s} f}\left(\|\operatorname{grad}_g \psi_1\|_{L^2(\mathfrak{X}(M))} + \|\operatorname{grad}_g \psi_2\|_{L^2(\mathfrak{X}(M))}\right)\end{aligned}$$

where $\psi_1 = \Delta_M f$ and $\psi_2 = \operatorname{div}_g\left(I_{\phi_s}\operatorname{grad}_g I_{\phi_s} f\right)$.

Proof. For any function $f = \sum_{i=1}^N a_i f_i$, it is clear that $R_N\Delta_M f = \Delta_M^{(N)} R_N f$. Hence for any h we have that $\langle \Delta_M^{(N)} R_N f, R_N h\rangle_{L^2(\mu_N)} = \langle R_N\Delta_M f, R_N h\rangle_{L^2(\mu_N)}$. Using the previous lemma, with probability greater than $1 - \frac{2}{N}$,

$$\left|\|\Delta_M^{(N)}\mathbf{f} - \mathbf{L}_N\mathbf{f}\|^2_{L^2(\mu_N)} - \|\Delta_M f + \operatorname{div}_g\left(I_{\phi_s}\operatorname{grad}_g I_{\phi_s} f\right)\|^2_{L^2(M)}\right| = O\left(1/\sqrt{N}\right).$$

Combining with Proposition 4.2 , the desired result follows. □

4.2.3 Proof of Spectral Convergence

Let $f_1, f_2, \ldots f_N$ be the first N smooth, orthonormal eigenfunctions of Δ_M. Notice that for $1 \leq i, j \leq N$,

$$\langle R_N f_i, R_N f_j\rangle_{L^2(\mu_N)} = \delta_{ij} + O\left(\frac{1}{\sqrt{N}}\right)$$

with high probability, since the eigenfunctions are orthonormal in $L^2(M)$. Let $\Delta_M^{(N)}$ denote the matrix satisfying

$$\Delta_M^{(N)} R_N f_i = \lambda_i R_N f_i \qquad 1 \leq i \leq N.$$

Perform Gram-Schmidt orthogonalization process on $\{R_N f_1, \dots, R_N f_N\}$, and denote the resulting vectors by $\{\tilde{\mathbf{f}}_1, \dots, \tilde{\mathbf{f}}_N\}$. Finally, define a matrix $\widehat{\Delta}_M^{(N)}$ on this basis by

$$\widehat{\Delta}_M^{(N)} \tilde{\mathbf{f}}_j := \lambda_i \tilde{\mathbf{f}}_j, \qquad j = 1, \dots, N.$$

Our analysis of the eigenvalues of $\mathbf{L}_N$ and Δ_M goes as follows. On one hand,

$$\|\widehat{\Delta}_M^{(N)} \tilde{\mathbf{f}}_i - \mathbf{L}_N \tilde{\mathbf{f}}_i\|_{L^2(\mu_N)} \leq \|\widehat{\Delta}_M^{(N)} \tilde{\mathbf{f}}_i - \Delta_M^{(N)} \tilde{\mathbf{f}}_i\|_{L^2(\mu_N)} + \|\Delta_M^{(N)} \tilde{\mathbf{f}}_i - \mathbf{L}_N \tilde{\mathbf{f}}_i\|_{L^2(\mu_N)}. \tag{30}$$

For fixed i, the first term on the right converges to 0 after $N \to \infty$. This can be seen by writing the Gram-Schmidt orthogonalization process explicitly, and using the earlier observation that $\{R_N f_i\}$ are *almost* orthonormal. The second term also converges to 0 as $N \to \infty$. This can be seen via a slight adaptation of the results from the previous section. The left-hand-side, however, can also be evaluated as

$$\|\widehat{\Delta}_M^{(N)} \tilde{\mathbf{f}}_i - \mathbf{L}_N \tilde{\mathbf{f}}_i\|_{L^2(\mu_N)}^2 = \|\lambda_i \tilde{\mathbf{f}}_i - \sum_{j=1}^{N} a_{ji} \tilde{\mathbf{f}}_j\|_{L^2(\mu_N)}^2 = \|(\lambda_i - a_{ii}) \tilde{\mathbf{f}}_i - \sum_{j \neq i} a_{ji} \tilde{\mathbf{f}}_j\|_{L^2(\mu_N)}^2 = |\lambda_i - a_{ii}|^2 + \sum_{j \neq i} |a_{ji}|^2,$$

where we recall that $[a_{ij}]$ are the coefficients of $\mathbf{L}_N$ in the basis $\{\tilde{\mathbf{f}}_i\}_{i=1}^N$, and the above holds since $\{\tilde{\mathbf{f}}_j\}_{i=1}^N$ is an orthonormal basis.

We can then use Gershgorin's circle Theorem to guarantee the existence of an eigenvalue of $\mathbf{L}_N$ in a Gershgorin circle with center converging to λ_i, and radius converging to 0. This leads precisely to a spectral convergence result. We formalize these observations in the following proof of Theorem 4.3.

Proof. Fix any $i \in \{\ell + 1, \dots, \ell + m\}$. First, with probability $1 - \frac{2i}{N}$ we have that

$$\langle R_N f_i, R_N f_j \rangle_{L^2(\mu_N)} = \delta_{ij} + O\left(\frac{1}{\sqrt{N}}\right)$$

for $1 \leq j \leq i$, by Hoeffding's inequality. Note that $\tilde{\mathbf{f}}_i$ is in the span of $\{R_N f_1, \dots, R_N f_i\}$. Hence, there is a continuous function $\tilde{f}$ (defined with the same linear combination as the discrete vectors), with $\tilde{f} \in \text{Span}\{f_1, \dots, f_i\}$ such that $R_N \tilde{f} = \tilde{\mathbf{f}}_i$. Since $\tilde{f} \in \text{Span}\{f_1, \dots, f_i\}$, its supremum can be bounded depending on $\{f_1, \dots, f_i\}$, and hence independent of N. It follows immediately then from Proposition 4.3 combined with Lemma 4.2 that the term

$$\|\Delta_M^{(N)} \tilde{\mathbf{f}}_i - \mathbf{L}_N \tilde{\mathbf{f}}_i\|_{L^2(\mu_N)}^2 = O\left(1/\sqrt{N}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right)$$

with a total probability of $1 - \frac{6}{N} - \frac{2n}{N} - \frac{2i}{N}$. For the first term of (30), notice that

$$\begin{aligned} \|\widehat{\Delta}_M^{(N)} \tilde{\mathbf{f}}_i - \Delta_M^{(N)} \tilde{\mathbf{f}}_i\|_{L^2(\mu_N)} &\leq \|\widehat{\Delta}_M^{(N)} \tilde{\mathbf{f}}_i - \Delta_M^{(N)} R_N f_i\|_{L^2(\mu_N)} + \|\Delta_M^{(N)} R_N f_i - \Delta_M^{(N)} \tilde{\mathbf{f}}_i\|_{L^2(\mu_N)} \\ &= \|\lambda_i \tilde{\mathbf{f}}_i - \lambda_i \mathbf{f}_i\|_{L^2(\mu_N)} + \|\Delta_M^{(N)} R_N f_i - \Delta_M^{(N)} \tilde{\mathbf{f}}_i\|_{L^2(\mu_N)} = O(1/\sqrt{N}) + \|\Delta_M^{(N)} R_N f_i - \Delta_M^{(N)} \tilde{\mathbf{f}}_i\|_{L^2(\mu_N)}, \end{aligned}$$

The Monte-Carlo rate in the above equation is derived below. Since $\tilde{\mathbf{f}}_i - R_N f_i$ is in the span of $\{R_N f_1, \dots R_N f_i\}$, it follows that

$$\|\Delta_M^{(N)} R_N f_i - \Delta_M^{(N)} \tilde{\mathbf{f}}_i\|_{L^2(\mu_N)} \leq \lambda_i \|R_N f_i - \tilde{\mathbf{f}}_i\|_{L^2(\mu_N)}$$

But

$$\begin{aligned} R_N f_i - \tilde{\mathbf{f}}_i &= R_N f_i - \frac{R_N f_i - \sum_{j=1}^{i-1} \langle R_N f_i, \tilde{\mathbf{f}}_j \rangle_{L^2(\mu_N)} \tilde{\mathbf{f}}_j}{\left\| R_N f_i - \sum_{j=1}^{i-1} \langle R_N f_i, \tilde{\mathbf{f}}_j \rangle_{L^2(\mu_N)} \tilde{\mathbf{f}}_j \right\|_{L^2(\mu_N)}} \\ &= \frac{R_N f_i \left(\left\| R_N f_i - \sum_{j=1}^{i-1} \langle R_N f_i, \tilde{\mathbf{f}}_j \rangle_{L^2(\mu_N)} \tilde{\mathbf{f}}_j \right\|_{L^2(\mu_N)} \right) - R_N f_i + \sum_{j=1}^{i-1} \langle R_N f_i, \tilde{\mathbf{f}}_j \rangle_{L^2(\mu_N)} \tilde{\mathbf{f}}_j}{\left\| R_N f_i - \sum_{j=1}^{i-1} \langle R_N f_i, \tilde{\mathbf{f}}_j \rangle_{L^2(\mu_N)} \tilde{\mathbf{f}}_j \right\|_{L^2(\mu_N)}} \end{aligned}$$

Using that $\left\| R_N f_i - \sum_{j=1}^{i-1} \langle R_N f_i, \tilde{\mathbf{f}}_j \rangle_{L^2(\mu_N)} \tilde{\mathbf{f}}_j \right\|_{L^2(\mu_N)} = 1 + O\left(\frac{1}{N^{1/4}}\right)$, and $\| \sum_{j=1}^{i-1} \langle R_N f_i, \tilde{\mathbf{f}}_j \rangle_{L^2(\mu_N)} \tilde{\mathbf{f}}_j \|_{L^2(\mu_N)} = O\left(\frac{1}{N^{1/4}}\right)$, we see that

$$\|R_N f_i - \tilde{\mathbf{f}}_i\|_{L^2(\mu_N)}^2 = O\left(\frac{1}{N^{1/2}}\right).$$

The above derivations involving norms can be seen directly from the Gram-Schmidt process and the fact that $\langle R_N f_k, R_N f_j\rangle_{L^2(\mu_N)} = \delta_{kj} + O\left(1/N^{1/2}\right)$. This shows that

$$\|\widehat{\Delta}_M^{(N)}\tilde{\mathbf{f}}_i - \mathbf{L}_N\tilde{\mathbf{f}}_i\|^2_{L^2(\mu_N)} = O\left(1/\sqrt{N}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right)$$

with probability greater than $1 - \frac{6}{N} - \frac{2n}{N} - \frac{2i}{N}$. To see this gives the desired bound, since

$$\|\widehat{\Delta}_M^{(N)}\tilde{\mathbf{f}}_i - \mathbf{L}_N\tilde{\mathbf{f}}_i\|_{L^2(\mu_N)} = \left(|\lambda_i - a_{ii}|^2 + \sum_{j\neq i}|a_{ji}|^2\right)^{1/2},$$

we have that the center of the Gershgorin circle a_{ii} satisfies

$$|\lambda_i - a_{ii}| = \sqrt{O\left(1/\sqrt{N}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right)}.$$

Since the i-th column has k_i nonzero entries, we see that

$$\sum_{j\neq i}|a_{ji}| \leq (k_i)^{1/2}\cdot\left(\sum_{j\neq i}|a_{ji}|^2\right)^{1/2} = k_i^{1/2}\sqrt{O\left(1/\sqrt{N}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right)}.$$

Performing the above procedure for $i = \ell+1,\ldots,\ell+m$, and using the Assumption 4.3, the union of these m discs is disjoint from the others. Hence, by Gershgorin circle theorem, there exists an eigenvalue of $\mathbf{L}_N$ laying in the union of these ℓ Gershgorin circles. This is the desired eigenvalue μ_N. Note that this event holds with probability higher than $1 - \sum_{i=\ell+1}^{\ell+m}\left(6/N - 2n/N - \frac{2i}{N}\right)$. Sending $N\to\infty$, we see that $\mu_N \to \lambda_{\ell+1}$, since the center of each circle approaches $\lambda_{\ell+1}$ and the radii converge to 0. This completes the proof. □

Remark 6. *We note that Assumption 4.3, used above, is only valid when λ_j, the value which the center of the j-th disc approaches, grows with N sufficiently fast to account for the potentially large radii of circles $\mathscr{C}_j$ with $j \gg i$, for fixed i. A consequence of Weyl's formula is that the spectrum of Δ_M grows like $O(i^{\frac{2}{d}})$. See, for instance, [13]. It follows that while this assumption is reasonable for manifolds of lower dimension, it may break down when the dimension is larger than 4, at which point the N-th smallest eigenvalue of Δ_M grows as slow as $O(\sqrt{N})$.*

Remark 7. *Though the convergence of eigenvectors in this setting is numerically observed, the theoretical result ensuring this with high probability remains open. The key issue that prevents the techniques of Section 4.1.4 from being used in this setting is the asymmetry of the estimator $\mathbf{L}_N$. In particular, since $\mathbf{L}_N$ is a non-symmetric matrix, there does not exist an orthonormal basis of $\mathbb{R}^N$ consisting of eigenvectors of $\mathbf{L}_N$. This is an important aspect of the symmetric matrix $\mathbf{G}^\top\mathbf{G}$, which allowed the proof in Section 4.1.4 to be carried out.*

5 Spectral Convergence for the Bochner Laplacian

In this section, we prove spectral convergence of discrete estimators to the Bochner Laplacian on vector fields. The Bochner Laplacian on vector fields is defined in such a way that makes the theoretical discussion in this setting almost identical to that of the Laplace-Beltrami operator. In particular, we can again derive convergence results in terms of the interpolation error on vector fields. Since interpolation of vector fields is especially useful for this chapter, we restate the interpolation of vector fields result, which is simply an application of Lemma 4.1 n times.

Lemma 5.1. *Let ϕ_s be a kernel whose RKHS norm is norm equivalent to a Sobolev space of order $\alpha > n/2$. Then with probability $1 - \frac{n}{N}$, we have*

$$\|I_{\phi_s}u - u\|_{L^2(\mathfrak{X}(M))} = O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right),$$

for all $u \in \mathfrak{X}(M)$.

The main results of this section are as follows. First, we have the following eigenvalue convergence result.

Theorem 5.1. *(convergence of eigenvalues: symmetric formulation) Let λ_i denote the i-th eigenvalue of Δ_B, enumerated $\lambda_1 \leq \lambda_2 \leq \ldots$. For $1 \leq i < N$, there exists an eigenvalue $\hat{\lambda}_i$ of $\mathbf{P}^{\otimes}\mathbf{H}^\top\mathbf{H}\mathbf{P}^{\otimes}$ such that*

$$\left|\lambda_i - \hat{\lambda}_i\right| \leq O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right),$$

with probability greater than $1 - \left(\frac{4n+4}{N}\right)$.

We need the following definition, analogous to Definition 4.2, but in the setting of vector fields.

Definition 5.1. *Denote the L^2 error between the vector fields $I_{\phi_s} v$ and v by ϵ_v. i.e.,*

$$\epsilon_v := \| I_{\phi_s} v - v \|_{L^2(\mathfrak{X}(M)}$$

We will also define a parameter $0 \leq \delta_v \leq 1$ to probabilistically characterize an upper bound for ϵ_v. For example, Lemma 5.1 states that $\epsilon_v = O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right)$ with probability higher than $1-\delta_v$, where $\delta_v = n/N$.

Similarly to the previous section, we have the following two assumptions which are not needed, but vastly simplify error bounds of the necessary lemmas.

Assumption 5.1. *For a fixed vector field v, we assume that for sufficiently large datasize $|X|$, the interpolation of $(\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) v$ approximates $(\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) v$ just as accurate as the interpolation of $\Delta_B v$ approximates $\Delta_B v$. i.e.,*

$$\begin{aligned} \epsilon_{(\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) v} &:= \| I_{\phi_s} (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) v - (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) v \|_{L^2(\mathfrak{X}(M))} \\ &\leq C_3 \| I_{\phi_s} \Delta_B v - \Delta_B v \|_{L^2(\mathfrak{X}(M))} := C_3 \epsilon_{\Delta_B v}, \end{aligned}$$

for some constant $C_3 > 0$ depending on v. The intuitive explanation for this assumption is exactly as in Assumption 4.1.

Assumption 5.2. *There exists a constant $C_4 > 0$ depending only on $v \in \mathfrak{X}(M)$ and the fixed kernel ϕ_s such that,*

$$\left\| \Delta_B I_{\phi_s} \Delta_B v \right\|_{L^2(\mathfrak{X}(M))} + \left\| \Delta_B I_{\phi_s} (\mathrm{grad}_g I_{\phi_s})^* (\mathrm{grad}_g I_{\phi_s}) v \right\|_{L^2(\mathfrak{X}(M))} \leq C_4 \left\| \Delta_B^2 v \right\|_{L^2(\mathfrak{X}(M)} .$$

Again, the detailed description of this assumption is the exact same as in Assumption 4.2.

The same rate holds true for convergence of eigenvectors, just as in the Laplace-Beltrami case. In fact, the proof of convergence of eigenvectors follows the same argument in Section 4.1.4. The main result is formally stated in the following theorem.

Theorem 5.2. *Let $\epsilon_{\lambda_\ell} := |\lambda_\ell - \hat{\lambda}_\ell|$ denote the error in approximating the ℓ-th distinct eigenvalue, following the notation in Theorem 5.1. Let the Assumptions 5.1, and 5.2 be valid. For any ℓ, assume that there is a constant c_ℓ such that if $\epsilon_{\lambda_{\ell-1}}, \epsilon_{\lambda_{\ell+1}} < c_\ell$, then with probability higher than $1 - \left(\frac{2m^2+4m+12n+12}{N}\right)$, we have the following situation: for any normalized eigenvector $\mathbf{U}$ of $\mathbf{P}^{\otimes} \mathbf{H}^\top \mathbf{H} \mathbf{P}^{\otimes}$ with eigenvalue $\hat{\lambda}_\ell$, there is a normalized eigenvector field v of Δ_B with eigenvalue λ_ℓ such that*

$$\| R_N v - \mathbf{U} \|_{L^2(\mu_N)} = O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right),$$

where m is the geometric multiplicity of eigenvalue λ_ℓ.

Remark 8. *Since the exact same proof as the eigenvector convergence for the Laplace Beltrami operator is valid, we do not rewrite it in full detail. Instead, we simply note that while Theorem 4.2 is proved using Theorem 4.1, as well as Lemmas 4.1, 4.6, and 4.8, similarly, we have that Theorem 5.2 can be proved with the exact same argument, using instead Theorem 5.1, along with Lemmas 5.1, 5.5, and 5.7. Statements and proofs of the lemmas mentioned above can be found below, in the present section.*

For the non-symmetric estimator, we additionally need an interpolation of $(2,0)$ tensor-field result. A proof of the following lemma can be found in Appendix B.

Lemma 5.2. *For any $a = \sum_{ij} a_{ij} \frac{\partial}{\partial \theta^i} \otimes \frac{\partial}{\partial \theta^j} \in L^2(T^{(2,0)}(TM))$, we have that with probability $1 - \frac{n^2}{N}$,*

$$\| a - I_{\phi_s} a \|_{L^2(T^{(2,0)}(TM))} = \left(\int_M \langle a - I_{\phi_s} a, a - I_{\phi_s} a \rangle_x dV(x) \right)^{1/2} = O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right).$$

In the same way that the orthonormal basis $\{\tilde{\mathbf{f}}_1, \ldots, \tilde{\mathbf{f}}_N\}$ is obtained for $\mathbb{R}^N$, we can similarly obtain an orthonormal basis $\{\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_{nN}\}$ for $\mathbb{R}^{nN}$ by orthonormalizing the restricted eigenvector fields of the Bochner Laplacian. This brings us to the following definition of sparseness coefficient for the Bochner Laplacian.

Definition 5.2. *We define the sparseness coefficient for the Bochner Laplacian k_i to be the number of nonzero entries in the i-th column of $\mathbf{L}_N^B$ when written w.r.t. the basis $\{\tilde{\mathbf{v}}_i\}$. We will denote the matrix coefficients of $\mathbf{L}_N^B$ written in this basis by $a_{ij} = \tilde{\mathbf{v}}_i^\top \mathbf{L}_N^B \tilde{\mathbf{v}}_j$.*

Remark 9. *Since we do not make use of the sparseness coefficient for the Laplace Beltrami operator at all in this section, we use the same letter k_i to denote the sparseness coefficient both for convenience, and to emphasize the similarity of theoretical analysis in this setting between the Bochner Laplacian and the Laplace Beltrami operator.*

We additionally have an assumption analagous to Assumption 4.3.

Assumption 5.3. *Fix an eigenvalue $\lambda_{\ell+1} = \cdots = \lambda_{\ell+m}$ of the Bochner Laplacian with multiplicity m. We assume again that for sufficiently large N, the union of the Gershgorin discs $\cup_{j=1}^{m} \mathscr{C}_{\ell+j}$ corresponding to columns $\ell+1,\ldots,\ell+m$ of the matrix $[a_{ij}]$, is disjoint from the remaining $N-m$ discs.*

Using this definition, the eigenvalue convergence result for the non-symmetric formulation of the Bochner Laplacian is formally stated in the following theorem.

Theorem 5.3. *Fix any eigenvalue $\lambda_{\ell+1} = \cdots = \lambda_{\ell+m}$ of the Bochner Laplacian of multiplicity m. Additionally, let Assumption 5.3 be valid. Then, with high probability, there exists a sequence $\{\mu_N\}_{N=1}^{\infty}$, where μ_N is an eigenvalue of $\mathbf{L}_N^B$, such that*

$$|\lambda_{\ell+1} - \mu_N| \to 0 \text{ as } N \to \infty.$$

5.1 Interpolation of (2,0)-tensor fields and approximation of inner-products

In the Bochner Laplacian discussion, we need to interpolate $(2,0)$ tensor fields, and approximate the corresponding continuous inner-product with a discrete inner-product. We outline the strategy for doing so presently. Given a $(2,0)$ tensor field $a = a_{ij}\frac{\partial}{\partial\theta^i}\otimes\frac{\partial}{\partial\theta^j}$, we can extend a to $A = A_{ij}\frac{\partial}{\partial X^i}\otimes\frac{\partial}{\partial X^j}$, defined on a neighborhood of M in $\mathbb{R}^n$, and write it as an $n\times n$ matrix in the basis $\frac{\partial}{\partial X^i}\otimes\frac{\partial}{\partial X^j}$. We define $I_{\phi_s}A$ to be the $(2,0)$ tensor field with components $I_{\phi_s}A_{ij}$ in the above ambient space basis. Recall that if $a,b\in T^{(2,0)}TM$, then by definition the Riemannian inner-product at x is given by

$$\langle a,b\rangle_x = \sum_{i,j,k,l} a_{ij}b_{kl}\left\langle\frac{\partial}{\partial\theta^i},\frac{\partial}{\partial\theta^k}\right\rangle_x\left\langle\frac{\partial}{\partial\theta^j},\frac{\partial}{\partial\theta^l}\right\rangle_x.$$

Performing a change of basis to the ambient space coordinates, a computation shows that

$$\langle a,b\rangle_x = \operatorname{tr}(A(x)^\top B(x)),$$

where A,B are the extensions of a,b and thought of as $n\times n$ matrices written w.r.t. the ambient space coordinates. Hence, defining the restriction of a $(2,0)$ tensor field $A = A_{ij}\frac{\partial}{\partial X^i}\otimes\frac{\partial}{\partial X^i}$ to be the tensor $R_NA\in\mathbb{R}^{n\times n\times N}$ with components,

$$(R_NA)_{ijk} = A_{ij}(x_k),$$

it follows that the inner-product on $\mathbb{R}^{n\times n\times N}$ given by

$$\langle R_NA,R_NB\rangle_{L^2(\mu_{N,n\times n})} := \frac{1}{N}\sum_{k=1}^{N}\operatorname{tr}\left(A(x_k)^\top B(x_k)\right)$$

approximates the continuous inner-product on $(2,0)$ tensor fields over M. In a previous notion of discrete $(2,0)$ tensor fields $[\mathbf{U}_1,\ldots,\mathbf{U}_n]\in\mathbb{R}^{nN\times n}$, each column $\mathbf{U}_i$ was thought of as a restricted vector field, equipped with the inner-product

$$\langle[\mathbf{U}_1,\ldots,\mathbf{U}_n],[\mathbf{V}_1,\ldots,\mathbf{V}_n]\rangle_{L^2(\mu_{N,N\times n})} := \frac{1}{N}\sum_{i=1}^{n}\mathbf{U}_i\cdot\mathbf{V}_i.$$

These two notions are equivalent in the following sense. Define $\Phi:\mathbb{R}^{nN\times n}\to\mathbb{R}^{n\times n\times N}$ by

$$\Phi\mathbf{E}_{(i-1)N+k,j} = \mathbf{E}_{i,j,k}\qquad i,j = 1,2,\ldots,n \text{ and } k = 1,\ldots,N,$$

where $\mathbf{E}_{(i-1)N+k,j}$ denotes the $nN\times n$ matrix with a one in entry $((i-1)N+k,j)$ and zeros elsewhere. Similarly for $\mathbf{E}_{i,j,k}$. One can easily check that Φ is an isometric isomorphism between the two inner product spaces defined above. In what follows, we will use this identification to consider $\mathbf{H}:\mathbb{R}^{nN}\to\mathbb{R}^{nN\times n}$ as a map with range in $\mathbb{R}^{n\times n\times N}$. This will be useful for computing the adjoint. Denote by $L^2(\mu_{N,n})$ the inner-product on $\mathbb{R}^{nN}$ which approximates $L^2(\mathfrak{X}(M))$:

$$\langle\mathbf{U},\mathbf{V}\rangle_{L^2(\mu_{N,n})} = \frac{1}{N}\sum_{j=1}^{n}\mathbf{U}^j\cdot\mathbf{V}^j,$$

where $\mathbf{U}^j \in \mathbb{R}^N$, and $\mathbf{U} = ((\mathbf{U}^1)^\top, \dots, (\mathbf{U}^n)^\top)^\top$. Using the identification above, it is simple to check that the transpose of $\Phi\mathbf{H} : \mathbb{R}^{nN} \to \mathbb{R}^{n\times n\times N}$ with respect to inner-products defined above is given by,

$$[\tilde{\mathbf{U}}^1, \dots, \tilde{\mathbf{U}}^n] \mapsto [\mathbf{U}^1, \dots, \mathbf{U}^n] \mapsto \sum_i \mathbf{H}_i^\top \mathbf{U}^i,$$

where $[\tilde{\mathbf{U}}^1, \dots, \tilde{\mathbf{U}}^n] \in \mathbb{R}^{n\times n\times N}$. Since Φ is an isometric isomorphism, $\mathbf{H}^\top\Phi^\top\Phi\mathbf{H} = \mathbf{H}^\top\mathbf{H}$. This shows that $\mathbf{H}^\top\mathbf{H} : \mathbb{R}^{nN} \to \mathbb{R}^{nN}$ is indeed given by the formula

$$\mathbf{H}^\top\mathbf{H} = \sum_i \mathbf{H}_i^\top \mathbf{H}_i,$$

which will be used extensively in the following calculations.

5.2 Symmetric Formulation

In this subsection, we discuss theoretical results concerning the Bochner Laplacian approximated by the symmetric matrix

$$\mathbf{P}^\otimes\mathbf{H}^\top\mathbf{H}\mathbf{P}^\otimes = \sum_{i=1}^n \mathbf{P}^\otimes\mathbf{H}_i^\top\mathbf{H}_i\mathbf{P}^\otimes.$$

This subsection is organized analogously to Section 4.1 that studies the spectral convergence of the symmetric approximation to the Laplace-Beltrami operator. We first investigate the continuous counterpart of the above discrete estimator, and prove its convergence in terms of interpolation error to the Bochner Laplacian in the weak sense, as well as in $L^2(\mathfrak{X}(M))$ sense. We then apply law of large numbers results to quantity the error obtained by discretizing to the dataset. Finally, due to the similarity of the setup compared to the Laplace-Beltrami operator, we can use the same arguments to derive eigenvalue and eigenvector convergence results.

5.2.1 Pointwise and weak convergence results: interpolation error

Using Cauchy-Schwarz, along with the fact that the formal adjoint of grad_g acting on vector fields is div_1^1 as defined in (20), we immediately have the following result.

Lemma 5.3. *Let $u \in \mathfrak{X}(M)$, and let $a \in T^{(2,0)}TM$. Then w.p. higher than $1-\delta_u$,*

$$\left|\langle \mathrm{grad}_g u - \mathrm{grad}_g I_{\phi_s} u, a\rangle_{L^2(T^{(2,0)}TM)}\right| \le \epsilon_u \|\mathrm{div}_1^1(a)\|_{L^2(\mathfrak{X}(M))}.$$

We also have the following norm result.

Corollary 5.1. *Let $u \in \mathfrak{X}(M)$. Then with probability higher than $1-\delta_u$,*

$$\|\mathrm{grad}_g u - \mathrm{grad}_g I_{\phi_s} u\|^2_{L^2(T^{(2,0)}TM)} \le \epsilon_u \left(\|\Delta_B u\|_{L^2(\mathfrak{X}(M))} + \|\Delta_B I_{\phi_s} u\|_{L^2(\mathfrak{X}(M))}\right).$$

Proof. Note that,

$$\|\mathrm{grad}_g u - \mathrm{grad}_g I_{\phi_s} u\|^2 = \langle \mathrm{grad}_g u, \mathrm{grad}_g u\rangle - \langle \mathrm{grad}_g u, \mathrm{grad}_g I_{\phi_s} u\rangle - \langle \mathrm{grad}_g I_{\phi_s} u, \mathrm{grad}_g u\rangle + \langle \mathrm{grad}_g I_{\phi_s} u, \mathrm{grad}_g I_{\phi_s} u\rangle.$$

Grouping the first two and last two terms, the desired result is immediate. □

Using the same reasoning as before, we can deduce a weak convergence result.

Lemma 5.4. *Let $v, w \in \mathfrak{X}(M)$. Then with probability higher than $1-\delta_v-\delta_w$,*

$$\left|\langle \Delta_B v, w\rangle_{L^2(\mathfrak{X}(M))} - \langle \mathrm{grad}_g I_{\phi_s} v, \mathrm{grad}_g I_{\phi_s} w\rangle_{L^2(T^{(2,0)}TM)}\right| \le \epsilon_w \|\Delta_B v\|_{L^2(\mathfrak{X}(M))} + \epsilon_v \|\Delta_B I_{\phi_s} w\|_{L^2(\mathfrak{X}(M))}.$$

Proof. We again add and subtract a mixed term.

$$\langle \Delta_B v, w\rangle - \langle \mathrm{grad}_g v, \mathrm{grad}_g I_{\phi_s} w\rangle + \langle \mathrm{grad}_g v, \mathrm{grad}_g I_{\phi_s} w\rangle - \left\langle \mathrm{grad}_g I_{\phi_s} v, \mathrm{grad}_g I_{\phi_s} w\right\rangle$$

The first two terms are bounded by $\epsilon_w\|\Delta_B v\|_{L^2(\mathfrak{X}(M))}$ while the second two terms are bounded by $\epsilon_v\|\Delta_B I_{\phi_s} w\|_{L^2(\mathfrak{X}(M))}$ □

Similarly, we can derive an L^2 convergence result. First, similar to the previous section, let us denote the formal adjoint of $\mathrm{grad}_g I_{\phi_s} : \mathfrak{X}(M) \to T^{(2,0)}TM$ by $(\mathrm{grad}_g I_{\phi_s})^* : T^{(2,0)}TM \to \mathfrak{X}(M)$. The exact same proof as before yields the following Lemma.

Lemma 5.5. *Let $v \in \mathfrak{X}(M)$, and let Assumptions 5.1 and 5.2 be valid. Then for sufficiently large dataset X, with probability higher than $1-\delta_v-\delta_{\Delta_B v}$,*

$$\|\Delta_B v - (\mathrm{grad}_g I_{\phi_s})^*(\mathrm{grad}_g I_{\phi_s})v\|^2 \leq \epsilon_v C_v \|\Delta_B^2 v\| \quad + \quad 2\epsilon_{\Delta_B v}\|\Delta_B v\|.$$

Proof. Expanding $\|\Delta_B v - (\mathrm{grad}_g I_{\phi_s})^*(\mathrm{grad}_g I_{\phi_s})v\|^2$ yields

$$\begin{aligned}\|\Delta_B v - (\mathrm{grad}_g I_{\phi_s})^*(\mathrm{grad}_g I_{\phi_s})v\|^2 &= \langle \Delta_B v, \Delta_B v\rangle - \langle \Delta_B v, (\mathrm{grad}_g I_{\phi_s})^*(\mathrm{grad}_g I_{\phi_s})v\rangle \\ &- \langle (\mathrm{grad}_g I_{\phi_s})^*(\mathrm{grad}_g I_{\phi_s})v, \Delta_B v\rangle \\ &+ \langle (\mathrm{grad}_g I_{\phi_s})^*(\mathrm{grad}_g I_{\phi_s})v, (\mathrm{grad}_g I_{\phi_s})^*(\mathrm{grad}_g I_{\phi_s})v\rangle.\end{aligned}$$

We use the previous Lemma twice, once on the first two terms (which gives an error of $\epsilon_{\Delta_B v}\|\Delta_B v\| + \epsilon_v\|\Delta_B I_{\phi_s}\Delta_B v\|$) and once on the last two terms (which gives an error of $\epsilon_{(\mathrm{grad}_g I_{\phi_s})^*(\mathrm{grad}_g I_{\phi_s})v}\|\Delta_B v\| + \epsilon_v\|\Delta_B I_{\phi_s}(\mathrm{grad}_g I_{\phi_s})^*(\mathrm{grad}_g I_{\phi_s})v\|$). Using Assumptions 5.1 and 5.2, we obtain the desired result. □

5.2.2 Pointwise and weak convergence results: empirical error

The following is a simple concentration result.

Lemma 5.6. *Let $u, v \in \mathfrak{X}(M)$. Then*

$$\mathbb{P}\left(\left|\langle \mathbf{H}\mathbf{P}^{\otimes} R_N u, \mathbf{H}\mathbf{P}^{\otimes} R_N v\rangle_{L^2(\mu_{N,n\times n})} - \langle \mathrm{grad}_g I_{\phi_s} u, \mathrm{grad}_g I_{\phi_s} v\rangle_{L^2(T^{(2,0)}TM)}\right| \geq \epsilon\right) \leq 2\exp\left(\frac{-2\epsilon^2 N}{D^2}\right),$$

where D is a constant depending on the kernel ϕ_s, as well as u and v.

Proof. We note that

$$\begin{aligned}\langle \mathrm{grad}_g I_{\phi_s} u, \mathrm{grad}_g I_{\phi_s} v\rangle &= \int_M \mathrm{tr}\left([\mathscr{H}_1 U, \dots \mathscr{H}_n U]^\top [\mathscr{H}_1 V, \dots \mathscr{H}_n V]\right) d\mathrm{Vol}(x) \\ &= \int_M \sum_{i=1}^n \langle \mathscr{H}_i U, \mathscr{H}_i V\rangle_x \, d\mathrm{Vol}(x) = \int_M \sum_{i=1}^n \langle \mathscr{H}_i \mathbf{P}U, \mathscr{H}_i \mathbf{P}V\rangle_x \, d\mathrm{Vol}(x),\end{aligned}$$

where U and V are extensions of u, v onto an $\mathbb{R}^n$ neighborhood of M. The last equality comes from the fact that since U, V extend u, v, then at each point $x \in M$, we have $\mathbf{P}U = U, \mathbf{P}V = v$. We can see immediately that the result follows by using a concentration inequality on the random variable $\sum_{i=1}^n \langle \mathscr{H}_i \mathbf{P}U, \mathscr{H}_i \mathbf{P}V\rangle_x$. Plugging in to Hoeffding's inequality and using smoothness assumptions give the desired result. □

Again, $\mathbf{H}$ is simply the restricted version of $\mathrm{grad}_g I_{\phi_s}$, as defined in (18). Hence, the same reasoning as above yields the norm convergence result, which plays a part in proving the convergence of eigenvectors.

Lemma 5.7. *Let $v \in \mathfrak{X}(M)$. Then*

$$\mathbb{P}\left(\left|\left\|\mathbf{P}^{\otimes}\mathbf{H}^\top\mathbf{H}\mathbf{P}^{\otimes} v - R_N \Delta_B v\right\|^2_{L^2(\mu_N,n)} - \|(\mathrm{grad}_g I_{\phi_s})^*(\mathrm{grad}_g I_{\phi_s})v - \Delta_B v\|^2_{L^2(\mathfrak{X}(M))}\right| \geq \epsilon\right) \leq 2\exp\left(\frac{-2\epsilon^2 N}{C^2}\right)$$

where C is a constant depending on ϕ_s and v.

5.2.3 Proof of Spectral Convergence

Here, we prove Theorem 5.1, the convergence of eigenvalue result for the Bochner Laplacian operator.

Proof. Enumerate the eigenvalues of $\mathbf{P}^{\otimes}\mathbf{H}^\top\mathbf{H}\mathbf{P}^{\otimes}$ and label them $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \cdots \leq \hat{\lambda}_N$. Let $\mathscr{S}_i' \subseteq \mathfrak{X}(M)$ denote an i-dimensional subspace of smooth functions on which the quantity $\max_{v\in\mathscr{S}_i} \frac{\langle \mathbf{P}^{\otimes}\mathbf{H}^\top\mathbf{H}\mathbf{P}^{\otimes} R_N v, R_N v\rangle_{L^2(\mu_{N,n})}}{\|R_N v\|_{L^2(\mu_{N,n})}}$ achieves its minimum. Let $\tilde{v} \in \mathscr{S}_i'$ be the smooth vector field on which the maximum $\max_{v\in\mathscr{S}_i'}\langle \Delta_B v, v\rangle$ occurs. WLOG, assume that $\|\tilde{v}\|^2_{L^2(\mathfrak{X}(M))} = 1$. Assume that N is sufficiently large so that by Hoeffding's inequality $\left|\|R_N\tilde{v}\|^2_{L^2(\mu_{N,n})} - 1\right| \leq \frac{\mathrm{Const}}{\sqrt{N}} \leq$

1/2, with probability $1-\frac{2}{N}$, so that $\|R_N\tilde{v}\|^2_{L^2(\mu_{N,n})}$ is bounded away from zero. Hence, we can Taylor expand the denominator term of $\frac{\langle \mathbf{P}^{\otimes}\mathbf{H}^{\top}\mathbf{H}\mathbf{P}^{\otimes}R_N\tilde{v},R_N\tilde{v}\rangle_{L^2(\mu_{N,n})}}{\|R_N\tilde{v}\|^2_{L^2(\mu_{N,n})}}$ to obtain

$$\frac{\langle \mathbf{P}^{\otimes}\mathbf{H}^{\top}\mathbf{H}\mathbf{P}^{\otimes}R_N\tilde{v},R_N\tilde{v}\rangle_{L^2(\mu_{N,n})}}{\|R_N\tilde{v}\|^2_{L^2(\mu_{N,n})}} = \langle \mathbf{P}^{\otimes}\mathbf{H}^{\top}\mathbf{H}\mathbf{P}^{\otimes}R_N\tilde{v},R_N\tilde{v}\rangle_{L^2(\mu_{N,n})} - \frac{\text{Const}\langle \mathbf{P}^{\otimes}\mathbf{H}^{\top}\mathbf{H}\mathbf{P}^{\otimes}R_N\tilde{v},R_N\tilde{v}\rangle_{L^2(\mu_{N,n})}}{\sqrt{N}}.$$

Note that with probability higher than $1-\frac{2}{N}$, we have that

$$\left|\langle \mathbf{P}^{\otimes}\mathbf{H}^{\top}\mathbf{H}\mathbf{P}^{\otimes}R_N\tilde{v},R_N\tilde{v}\rangle_{L^2(\mu_{N,n})} - \langle \text{grad}_g I_{\phi_s}\tilde{v},\text{grad}_g I_{\phi_s}\tilde{v}\rangle_{L^2(T^{(1,1)}(TM))}\right| = O\left(\frac{1}{\sqrt{N}}\right),$$

by Lemma 5.6, choosing $\epsilon=\sqrt{\frac{\log(N)}{N}}$ and ignoring log factors. Combining with Lemma 5.4 and plugging in the result from Lemma 5.1, we obtain that

$$\langle \Delta_B\tilde{v},\tilde{v}\rangle_{L^2(\mathfrak{X}(M))} \leq \frac{\langle \mathbf{P}^{\otimes}\mathbf{H}^{\top}\mathbf{H}\mathbf{P}^{\otimes}R_N\tilde{v},R_N\tilde{v}\rangle_{L^2(\mu_{N,n})}}{\|R_N\tilde{v}\|^2_{L^2(\mu_{N,n})}} + O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right),$$

with total probability higher than $1-\frac{2n+2}{N}$. Since $\tilde{v}=\arg\max_{v\in\mathscr{S}_i'}\langle\Delta_B v,v\rangle_{L^2(\mathfrak{X}(M))}$, and since,

$$\frac{\langle \mathbf{P}^{\otimes}\mathbf{H}^{\top}\mathbf{H}\mathbf{P}^{\otimes}R_N\tilde{v},R_N\tilde{v}\rangle_{L^2(\mu_{N,n})}}{\|R_N\tilde{v}\|_{L^2(\mu_{N,n})}} \leq \max_{v\in\mathscr{S}_i'}\frac{\langle \mathbf{P}^{\otimes}\mathbf{H}^{\top}\mathbf{H}\mathbf{P}^{\otimes}R_N v,R_N v\rangle_{L^2(\mu_{N,n})}}{\|R_N v\|_{L^2(\mu_{N,n})}},$$

we have the following:

$$\max_{v\in\mathscr{S}_i'}\langle\Delta_B v,v\rangle_{L^2(\mathfrak{X}(M))} \leq \max_{v\in\mathscr{S}_i'}\frac{\langle \mathbf{P}^{\otimes}\mathbf{H}^{\top}\mathbf{H}\mathbf{P}^{\otimes}R_N v,R_N v\rangle_{L^2(\mu_{N,n})}}{\|R_N v\|_{L^2(\mu_{N,n})}} + O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{-\frac{(2\alpha-d)}{2d}}\right).$$

Since $\mathscr{S}_i'$ is the exact subspace on which $\max_{v\in\mathscr{S}_i}\frac{\langle \mathbf{P}^{\otimes}\mathbf{H}^{\top}\mathbf{H}\mathbf{P}^{\otimes}R_N v,R_N v\rangle_{L^2(\mu_{N,n})}}{\|R_N v\|_{L^2(\mu_{N,n})}}$ achieves its minimum, then

$$\max_{v\in\mathscr{S}_i'}\langle\Delta_B v,v\rangle_{L^2(\mathfrak{X}(M))} \leq \hat{\lambda}_i + O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right).$$

But the left-hand-side certainly bounds above the minimum of $\max_{v\in\mathscr{S}_i}\langle\Delta_B v,v\rangle_{L^2(\mathfrak{X}(M))}$ over all i-dimensional smooth subspaces $\mathscr{S}_i$. Hence,

$$\lambda_i \leq \hat{\lambda}_i + O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right).$$

The same argument yields that $\hat{\lambda}_i \leq \lambda_i + O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right)$, with probability higher than $1-\frac{2+2n}{N}$. This completes the proof. □

5.3 Non-symmetric Formulation

In this subsection, we investigate the non-symmetric estimator to the Bochner-Laplacian, given by

$$\mathbf{L}_N^B := -\sum_{i=1}^n \mathbf{H}_i\mathbf{H}_i.$$

Similar to the symmetric formulation, this subsection is organized exactly as the non-symmetric subsection for the Laplace-Beltrami operator. We begin by proving results between the continuous version of $\mathbf{L}_N^B$ and Δ_B. Then, after quantifying the error involved in discretization, we prove spectral convergence. The proof of the final result is the same as the proof for spectral convergence for the non-symmetric estimator of the Laplace-Beltrami operator.

5.3.1 Pointwise and weak convergence results: interpolation error

Here, we state the analog of Propositions 4.1 and 4.2 for approximation of Bochner Laplacian. The proofs follow immediately with the same reasoning as Propositions 4.1 and 4.2, respectively.

Lemma 5.8. *Let $v, w \in \mathfrak{X}(M)$. Then with probability greater than $1-\delta_v-\delta_{\text{grad}_g I_{\phi_s} v}$,*

$$\left|\langle \Delta_B v + \text{div}_1^1\left(I_{\phi_s}\text{grad}_g I_{\phi_s} v\right), w\rangle\right| \leq \epsilon_v \|w\|_{L^2(\mathfrak{X}(M))} + \epsilon_{\text{grad}_g I_{\phi_s} v}\|\text{grad}_g w\|_{L^2(T^{(2,0)}TM)}.$$

Lemma 5.9. *Let $v \in \mathfrak{X}(M)$. Then with probability higher than $1-2\delta_v-2\delta_{\text{grad}_g I_{\phi_s} v}$,*

$$\begin{aligned}\|\Delta_B v + \text{div}_1^1\left(I_{\phi_s}\text{grad}_g I_{\phi_s} v\right)\|^2 \quad &\leq \quad \epsilon_v\left(\|w_1\|_{L^2(\mathfrak{X}(M))} + \|w_2\|_{L^2(\mathfrak{X}(M))}\right) \\ &\quad + \epsilon_{\text{grad}_g I_{\phi_s} v}\left(\|\text{grad}_g w_1\|_{L^2(T^{(2,0)}TM)} + \|\text{grad}_g w_2\|_{L^2(T^{(2,0)}TM)}\right),\end{aligned}$$

where $w_1 = \Delta_B v$ and $w_2 = \text{div}_1^1\left(I_{\phi_s}\text{grad}_g I_{\phi_s} v\right)$.

5.3.2 Pointwise and weak convergence results: empirical error

The following is an immediate consequence of Hoeffding's inequality.

Lemma 5.10. *Let $v \in \mathfrak{X}(M)$. Then we have that*

$$\mathbb{P}\left(\left|\|\Delta_B v + \text{div}_1^1\left(I_{\phi_s}\text{grad}_g I_{\phi_s} v\right)\|^2_{L^2(\mathfrak{X}(M))} - \|R_N\Delta_B v - \mathbf{L}_N^B R_N v\|^2_{L^2(\mu_{N,n})} \geq \epsilon\right|\right) \leq 2\exp\left(\frac{-2\epsilon^2 N}{C^2}\right)$$

where the constant $C>0$ depends on v and the kernel ϕ_s.

Given nN smooth linearly independent eigenvector fields $v_1,\ldots,v_{nN}$ of the Bochner Laplacian with eigenvalues $\lambda_1,\lambda_2,\ldots\lambda_{nN}$, the restriction $R_N v_1,\ldots,R_N v_{nN}$ is linearly independent, and hence forms a basis of $\mathbb{R}^{nN}$. In this case, we can define a matrix on the basis $\{R_N v_1,\ldots,R_N v_{nN}\}$ by

$$\Delta_B^{(N)} R_N v_i = \lambda_i R_N v_i.$$

Letting $V = [R_N v_1,\ldots R_N v_{nN}]$, and $\Lambda = \text{diag}(\lambda_1,\ldots\lambda_{nN})$ we see that

$$(\Delta_B^{(N)})_{ij} = (V\Lambda V^{-1})_{ij},$$

where Λ is a diagonal matrix whose diagonal entries are $\lambda_1,\ldots,\lambda_{nN}$. We have the following lemma regarding the distance between the matrices $\mathbf{L}_N^B$ and $\Delta_B^{(N)}$.

Lemma 5.11. *Let $\mathbf{U} \in \mathbb{R}^{nN}$. With probability greater than $1-\frac{4}{N}-2\delta_u-2\delta_{\text{grad}_g I_{\phi_s} u}$,*

$$\begin{aligned}\|\Delta_B^{(N)}\mathbf{U} - \mathbf{L}_N^B\mathbf{U}\|^2_{L^2(\mu_{N,n})} = O\left(1/\sqrt{N}\right) \quad &+ \quad \epsilon_v\left(\|w_1\|_{L^2(\mathfrak{X}(M))} + \|w_2\|_{L^2(\mathfrak{X}(M))}\right) \\ &+ \quad \epsilon_{\text{grad}_g I_{\phi_s} u}\left(\|\text{grad}_g w_1\|_{L^2(T^{(1,1)}(TM))} + \|\text{grad}_g w_2\|_{L^2(T^{(2,0)}TM)}\right),\end{aligned}$$

where $\mathbf{U} = R_N u$, $w_1 = \Delta_B u$ and $w_2 = \text{div}_1^1\left(I_{\phi_s}\text{grad}_g I_{\phi_s} u\right)$.

Proof. For any vector field u that can be written $u = \sum_{i=1}^{nN} a_i v_i$, it is clear that $R_N\Delta_B u = \Delta_B^{(N)} R_N u$. And hence for any w we have that $\langle \Delta_B^{(N)} R_N u, R_N w\rangle_{L^2(\mu_{N,n})} = \langle R_N\Delta_B u, R_N w\rangle_{L^2(\mu_{N,n})}$. Using Lemma 5.10, with probability greater than $1-\frac{2}{N}$,

$$\left|\|\Delta_B^{(N)} R_N u - \mathbf{L}_N^B R_N u\|^2_{L^2(\mu_{N,n})} - \|\Delta_B u + \text{div}_1^1\left(I_{\phi_s}\text{grad}_g I_{\phi_s} u\right)\|^2_{L^2(\mathfrak{X}(M))}\right| = O\left(1/\sqrt{N}\right).$$

Combining with Lemma 5.9 , the desired result follows. □

5.3.3 Outline of the proof of Spectral Convergence

Let $v_1, v_2, \ldots v_{nN}$ be the first nN smooth, orthonormal eigenvector fields of Δ_B. Notice that for $1 \leq i, j \leq N$,

$$\langle R_N v_i, R_N v_j\rangle_{L^2(\mu_{N,n})} = \delta_{ij} + O\left(\frac{1}{\sqrt{N}}\right)$$

with high probability, since the eigenvector fields are orthonormal in $L^2(\mathfrak{X}(M))$. Let $\Delta_B^{(N)}$ denote the matrix satisfying

$$\Delta_B^{(N)} R_N v_i = \lambda_i R_N v_i \qquad 1 \leq i \leq nN.$$

Perform Gram-Schmidt orthogonalization process on $\{R_N v_1, \ldots, R_N v_{nN}\}$, and denote the resulting vectors by $\{\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_{nN}\}$. This is precisely the basis involved in the Definition 5.2. Finally, define a matrix $\widehat{\Delta}_B^{(N)}$ by

$$\widehat{\Delta}_B^{(N)} \tilde{\mathbf{v}}_i = \lambda_i \tilde{\mathbf{v}}_i.$$

Our analysis of the eigenvalues of $\mathbf{L}_N^B$ and Δ_B goes as follows. On one hand,

$$\|\widehat{\Delta}_B^{(N)} \tilde{\mathbf{v}}_i - \mathbf{L}_N^B \tilde{\mathbf{v}}_i\|_{L^2(\mu_{N,n})} \leq \|\widehat{\Delta}_B^{(N)} \tilde{\mathbf{v}}_i - \Delta_B^{(N)} \tilde{\mathbf{v}}_i\|_{L^2(\mu_{N,n})} + \|\Delta_B^{(N)} \tilde{\mathbf{v}}_i - \mathbf{L}_N^B \tilde{\mathbf{v}}_i\|_{L^2(\mu_{N,n})}.$$

For fixed $i \in \{\ell+1, \ldots, \ell+m\}$, the first term on the right converges to 0 after $N \to \infty$. This can be seen by writing the Gram-Schmidt orthogonalization process explicitly, and using the earlier observation that $\{R_N v_i\}$ are *almost* orthonormal. The second term also converges to 0 as $N \to \infty$. This can be seen via a slight adaptation of the results from the previous section. The left-hand-side, however, can also be evaluated as

$$\|\widehat{\Delta}_B^{(N)} \tilde{\mathbf{v}}_i - \mathbf{L}_N^B \tilde{\mathbf{v}}_i\|_{L^2(\mu_{N,n})} = \left(|\lambda_i - a_{ii}|^2 + \sum_{j \neq i} |a_{ji}|^2 \right)^{1/2},$$

where a_{ij} are the entries of the matrix $\mathbf{L}_N^B$ in the basis $\{\tilde{\mathbf{v}}_1, \ldots \tilde{\mathbf{v}}_{nN}\}$. We can then use Gershgorin's circle Theorem on the matrix $[a_{ij}]$, which is similar to $\mathbf{L}_N^B$ and hence shares the same eigenvalues, to guarantee the existence of an eigenvalue of $\mathbf{L}_N^B$ in a Gershgorin circle with center converging to λ_i, and radius converging to 0. This leads precisely to a spectral convergence result. Again, the argument in the remaining of the proof is identical to that for the Laplace-Beltrami operator, where the main changes are on inner product spaces of vector fields and (2,0)-tensor fields with key Lemma 5.11 replacing Proposition 4.3 and Lemma 5.2 replacing Lemma 4.2. Tracking probabilities, we see that the spectral consistency holds with probability higher than $1 - \frac{2i}{N} - \frac{4}{N} - \frac{2n}{N} - \frac{2n^2}{N}$. Repeating the same argument above for $i = \ell+1, \ldots, \ell+m$, we see that the convergence, again, holds with probability approaching 1 as $N \to \infty$. Again since the radii of the Gershgorin circles approach 0, and the centers approach $\lambda_{\ell+1}$, we see that we can find a sequence of eigenvalues approaching $\lambda_{\ell+1}$.

6 Numerical study of eigenvalue problems

In this section, we first discuss two examples of eigenvalue problems of functions defined on simple manifolds: one being a 2D generalized torus embedded in $\mathbb{R}^{21}$ and the other 4D flat torus embedded in $\mathbb{R}^{16}$. In these two examples, we will compare the results between the Non-symmetric and Symmetric RBFs, which we refer to as NRBF and SRBF respectively, using analytic $\mathbf{P}$ and the approximated $\hat{\mathbf{P}}$, as well as diffusion maps (DM). When the manifold is unknown, as is often the case in practical applications, one does not have access to analytic $\mathbf{P}$. Hence, it is most reasonable to compare DM and RBF with $\hat{\mathbf{P}}$. Next, we discuss an example of eigenvalue problems of vector fields defined on a sphere. Numerically, we compare the results between the RBF method and the analytic truth. Since the size of our vector Laplacian approximation is $Nn \times Nn$, the current RBF methods are only numerically feasible for datasets with small ambient dimension n.

6.1 numerical setups

In the following, we introduce our numerical setups for DM, NRBF and SRBF methods of finding approximate solutions to eigenvalue problems associated to Laplace-Beltrami or vector Laplacians on manifolds.

Parameter Specification for RBF: For the implementation of RBF methods, there are two groups of kernels to be used. One group includes infinitely smooth RBFs, such as Gaussian (GA), multi-quadric (MQ), inverse multiquadric (IMQ), inverse quadratic (IQ), and Bessel (BE) [24, 34, 23]. The other group includes piecewise smooth RBFs, such as Polyharmonic spline (PHS), Wendland (WE), and Matérn (MA) [34, 26, 51, 52]. In this work, we only apply GA and IQ kernels and test their numerical performances. To compute the interpolant matrix Φ in (3), all points are connected and we did not use KNN truncations. The shape parameter s is manually tuned but fixed for different N when we examine the convergence of eigenmodes.

Despite not needing a structured mesh, many RBF techniques impose strict requirements for uniformity of the underlying data points. For uniformly distributed grid points, it often occurs that the operator approximation error decreases rapidly with the number of data N until the calculation breaks down due to the increasing ill-conditioning of the interpolant matrix Φ defined in (3) [46, 41, 34]. In this numerical section, we consider data points randomly distributed on manifolds, which means that two neighboring points can be very close to each other. In this case, the interpolant matrix Φ involved in most of the global RBF techniques tends to be ill-condition or even singular for sufficiently large N. To resolve such an ill-conditioning issue, we apply the pseudo inversion instead of the direct

inversion in approximating the interpolant matrix Φ. In our implementation, we will take the tolerance parameter of pseudo-inverse around $10^{-9} \sim 10^{-4}$.

If the parametrization or the level set representation of the manifold is known, we can apply the analytic tangential projection matrix $\mathbf{P}$ for constructing the RBF Laplacian matrix. If the parametrization is unknown, that is, only the point cloud is given, we can first learn $\hat{\mathbf{P}}$ using the 2nd-order SVD method and then construct the RBF Laplacian. Notice that we can also construct the Laplacian matrix using $\tilde{\mathbf{P}}$ estimated from the 1st-order SVD (not shown in this work). We found that the results of eigenvalues and eigenvectors using $\tilde{\mathbf{P}}$ are not as good as those using $\hat{\mathbf{P}}$ from our 2nd-order SVD. We also notice that the estimation of $\mathbf{P}$ and the construction of the Laplacian matrix can be performed separately using two sets of points. For example, one can use $10,000$ points to approximate $\mathbf{P}$ but use only 2500 points to construct the Laplacian matrix. This allows one to leverage large amounts of data in the estimation of $\mathbf{P}$, which may not be computationally feasible with graph Laplacian-based approximators such as the diffusion maps algorithm.

For SRBF, the estimated sampling density is needed if the distribution of the dataset is unknown. (Note that for NRBF, the sampling density is not needed for constructing Laplacian.) In our numerical experiments, we apply the MATLAB built-in kernel density estimations (KDEs) function mvksdensity.m for approximating the sampling density. We also apply Silverman's rule of thumb [42] for auto-tuning the bandwidth parameter in KDEs.

Eigenvalue problem solver for RBF: For NRBF, we apply the non-symmetric estimator in (13) for solving the eigenvalue problem. The NRBF eigenvectors might be complex-valued and are only linearly independent (i.e., they are not necessarily orthonormal). For SRBF, we apply the symmetric estimator in (14) for solving the generalized eigenvalue problem. When the sampling density q is unknown, we employ the symmetric formulation with the estimated sampling density $\tilde{q}(x)$ obtained from KDE.

Since we used pseudo inversion to resolve the ill-conditioning issue of Φ, the resulting RBF Laplacian matrix will be of low rank, L, and will have many zero eigenvalues, depending on the choice of tolerance in the pseudo-inverse algorithm. Two issues naturally arise in this situation. First, it becomes difficult to compute the eigenspace corresponding to the zero eigenvalue(s), especially for the eigenvector-field problem. At this moment, we have not developed appropriate schemes to detect the existence of the nullspace and estimate the harmonic functions in this nullspace if it exists. Second, finding even the leading nonzero eigenvalues (that are close to zero) can be numerically expensive. Based on the rank of the RBF Φ, for symmetric approximation, one can use the ordered real-valued eigenvalues to attain the nontrivial L eigenvalues in descending (ascending) order when L is small (large). For the nonsymmetric approximation, one can also employ a similar idea by sorting the magnitude of the eigenvalues (since the eigenvalues may be complex-valued). This naive method, however, can be very expensive when the number of data points N is large and when the rank of the matrix L is neither $O(10)$ nor close to N.

Comparison of eigenvectors for repeating eigenvalues: When an eigenvalue is non-simple, one needs to be careful in quantifying the errors of eigenvectors for these repeated eigenvalues since even the set of true orthonormal eigenvectors is only unique up to a rotation matrix. To quantify the errors of eigenvectors, we apply the following Ordinary Least Square (OLS) method. Let $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_m)$ be the true eigenfunctions located at X corresponding to one repeated eigenvalue λ_i with multiplicity m, and let $\tilde{\mathbf{U}} = (\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_m)$ be their DM or RBF approximations. Assume that the linear regression model is written as $\mathbf{F} = \tilde{\mathbf{U}}\beta + \varepsilon$, where ε is an $N \times m$ matrix representing the errors and β is a $m \times m$ matrix representing the regression coefficients. The coefficients matrix β can be approximated using OLS by $\hat{\beta} = (\tilde{\mathbf{U}}^\top \tilde{\mathbf{U}})^{-1}(\tilde{\mathbf{U}}^\top \mathbf{F})$. The rotated DM or RBF eigenvectors can be written as a linear combination, $\hat{\mathbf{V}} = \tilde{\mathbf{U}}\hat{\beta}$, where these new $\hat{\mathbf{V}} = (\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_m)$ are in $\mathrm{Span}\{\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_m\}$. Finally, we can quantify the pointwise errors of eigenvectors between $\mathbf{f}_j$ and $\hat{\mathbf{v}}_j$ for each $j = 1, \ldots, m$. For eigenvector fields, we can follow a similar outline to quantify the errors of vector fields. Incidentally, we mention that there are many ways to measure eigenvector error since the approximation of the rotational coefficient matrix β is not unique. Here, we only provide a practical metric that we will use in our numerical examples below.

6.2 2D general torus

In this section, we investigate the eigenmodes of the Laplace-Beltrami operator on a general torus. The parameterization of the general torus is given by

$$x = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^{n-2} \\ x^{n-1} \\ x^n \end{bmatrix} = \begin{bmatrix} (a+\cos\theta)\cos\phi \\ (a+\cos\theta)\sin\phi \\ \vdots \\ \frac{2}{n-1}(a+\cos\theta)\cos\frac{n-1}{2}\phi \\ \frac{2}{n-1}(a+\cos\theta)\sin\frac{n-1}{2}\phi \\ \sqrt{\sum_{i=1}^{(n-1)/2}\frac{1}{i^2}}\sin\theta \end{bmatrix}, \tag{31}$$

where the two intrinsic coordinates $0 \le \theta, \phi \le 2\pi$ and the radius $a = 2 > 1$. The Riemannian metric is

$$g = \begin{bmatrix} b_{\frac{n-1}{2}} & 0 \\ 0 & \frac{n-1}{2}(a+\cos\theta)^2 \end{bmatrix}, \tag{32}$$

where $b_{\frac{n-1}{2}} := \sum_{i=1}^{(n-1)/2}\frac{1}{i^2}$. We solve the following eigenvalue problem for Laplace-Beltrami operator:

$$\Delta_g \psi_k = \frac{-1}{(a+\cos\theta)}\left[\frac{\partial}{\partial\theta}\left((a+\cos\theta)\frac{1}{b_{\frac{n-1}{2}}}\frac{\partial\psi_k}{\partial\theta}\right) + \frac{\partial}{\partial\phi}\left(\frac{2}{n-1}\frac{1}{(a+\cos\theta)}\frac{\partial\psi_k}{\partial\phi}\right)\right] = \lambda_k\psi_k, \tag{33}$$

where λ_k and ψ_k are the eigenvalues and eigenfunctions, respectively. After separation of variables (that is, we set $\psi_k = \Phi_k(\phi)\Theta_k(\theta)$ and substitute ψ_k back into (33) to deduce the equations for Φ_k and Θ_k), we obtain:

$$\begin{aligned} \frac{\mathrm{d}^2\Phi_k}{\mathrm{d}\phi^2} + m_k^2\Phi_k &= 0, \\ \frac{\mathrm{d}}{\mathrm{d}\theta}\left((a+\cos\theta)\frac{\mathrm{d}\Theta_k}{\mathrm{d}\theta}\right) - \frac{b_{\frac{n-1}{2}}m_k^2}{\frac{n-1}{2}(a+\cos\theta)}\Theta_k &= -b_{\frac{n-1}{2}}(a+\cos\theta)\lambda_k\Theta_k. \end{aligned}$$

The eigenvalues to the first equation are $m_k^2 = k^2$ with $k = 0,1,2,\ldots$ and the associated eigenvectors are $\Phi_k = \{1, \sin k\phi, \cos k\phi\}$. The second eigenvalue problem can be written in the Sturm–Liouville form and then numerically solved on a fine uniform grid with N_θ points $\{\theta_j = \frac{2\pi j}{N_\theta}\}_{j=0}^{N_\theta-1}$ [39]. The eigenvalues λ_k associated with the eigenfunctions ψ_k obtained above are referred to as the true semi-analytic solutions to the eigenvalue problem (33).

In our numerical implementation, data points are randomly sampled from the general torus with uniform distribution in intrinsic coordinates. To apply the DM, we use $K = O(\sqrt{N})$ nearest neighbors to construct a sparse Laplacian matrix for computational efficiency and auto-tune the kernel bandwidth ϵ. In detail, $k = 31, 51, 71, 101$ for $1024, 2500, 5000, 10000$, respectively. To apply NRBF, we use IQ kernel with $s = 0.5$. To apply SRBF, we use IQ kernel with $s = 0.1$. Figure 3 shows the comparison of eigenfunctions for modes $k = 2, 9, 13$ among the semi-analytic truth, NRBF with $\mathbf{P}$ and $\hat{\mathbf{P}}$, SRBF with $\mathbf{P}$ and $\hat{\mathbf{P}}$, and DM. One can see from the first row of Fig. 3 that when $k = 2$ is very small, all the methods can provide excellent approximations of eigenfunctions. For larger k, such as 9 and 13, NRBF methods with $\mathbf{P}$ or $\hat{\mathbf{P}}$ provide more accurate approximations compared to SRBF with $\mathbf{P}$ and $\hat{\mathbf{P}}$ and DM. In fact, the eigenvectors obtained from NRBF with $\mathbf{P}$ are accurate and very smooth as seen from the second column of Fig. 3.

Figure 4 further quantifies the errors of eigenvalues and eigenfunctions for all the methods. One can see that NRBF with $\mathbf{P}$ performs much better than all other methods on this 2D manifold example. When the manifold is unknown, one can diagnose the manifold learning capabilities of the symmetric and nonsymmetric RBF using $\hat{\mathbf{P}}$ compared to DM. One can see that NRBF with $\hat{\mathbf{P}}$ (red curve) performs better than the other two methods. One can also see that DM (blue curve) performs slightly better than SRBF with $\hat{\mathbf{P}}$ (cyan curve) in estimating the leading eigenvalues but somewhat worst in estimating eigenvalues corresponding to the higher modes (Fig. 4(b)). In terms of eigenvectors estimation, they are comparable (blue and cyan curves Fig. 4(c)). Additionally, SRBF with $\mathbf{P}$ (yellow curve) and SRBF with $\hat{\mathbf{P}}$ (cyan curve) produce comparable accuracies in terms of the eigenvector estimation. This result, where no advantage is observed using the analytic $\mathbf{P}$ over the approximated $\hat{\mathbf{P}}$, is consistent with the theory which suggests that the error bound is dominated by the Monte-Carlo error rate provided smooth enough kernels are used.

In the previous two figures, we showed the SRBF estimates corresponding to a specific choice of $s = 0.1$. Now let us check the robustness of the method with respect to other choices of the shape parameter. In Fig. 5, we show the errors in the estimation of eigenvalues and eigenvectors. Specifically, we report the average errors of low modes

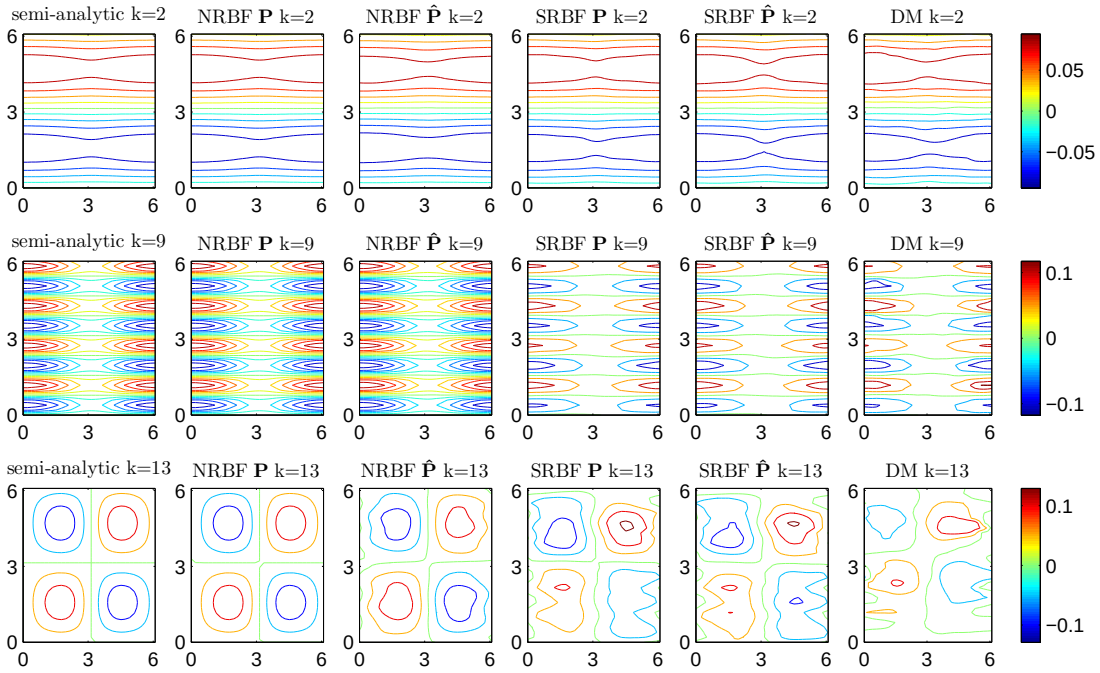

Figure 3: **2D general torus in $\mathbb{R}^{21}$.** Comparison of eigenfunctions of Laplace -Beltrami for $k = 2, 9, 13$ among NRBF using $\mathbf{P}$ and $\hat{\mathbf{P}}$, SRBF using $\mathbf{P}$ and $\hat{\mathbf{P}}$, and DM. For NRBF, IQ kernel with $s = 0.5$ is used, and for SRBF, IQ kernel with $s = 0.1$ is used. The horizontal and vertical axes correspond to θ and ϕ, respectively. The randomly distributed $N = 2500$ data points on manifold are used for computing the eigenvalue problem. Then, the eigenvectors are generalized onto the 32×32 well-sampled grid points $\{\theta_i, \phi_j\} = \{\frac{2\pi i}{32}, \frac{2\pi j}{32}\}_{i,j=0}^{31}$ for plotting.

(between modes 2-5) for DM (blue), SRBF with $\mathbf{P}$ (yellow), and SRBF with $\hat{\mathbf{P}}$ (cyan). For these low modes, notice that the eigenvalue estimates from SRBF with $\hat{\mathbf{P}}$ can be made as accurate as the diffusion maps with a slightly larger shape parameter, $s = 0.3$. For this shape parameter value, the SRBF with $\hat{\mathbf{P}}$ produces an even more accurate estimation of the leading eigenvalues. However, the accuracy of the SRBF eigenvectors decreases slightly under this parameter value. We also report the average errors of high modes (between modes 21-30) for DM (red), SRBF with $\mathbf{P}$ (green), and SRBF with $\hat{\mathbf{P}}$ (black). For these high modes, both SRBFs are uniformly more accurate than DM in the estimation of eigenvalues, but the accuracies of the estimation of eigenvectors are comparable. These empirical results suggest the SRBF has a slight advantage over DM, especially in the estimation of non-leading spectrum.

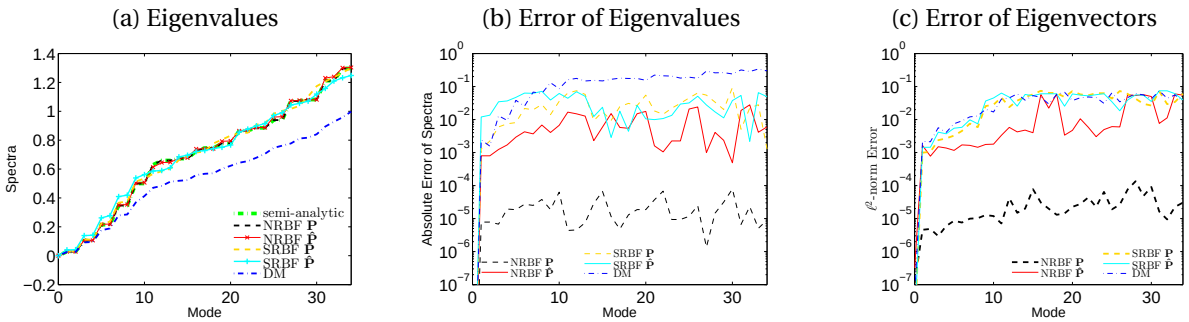

Figure 4: **2D general torus in $\mathbb{R}^{21}$.** Comparison of (a) eigenvalues, (b) error of eigenvalues, and (c) error of eigenvectors, among NRBF, SRBF, and DM. For NRBF, IQ kernel with $s = 0.5$ is used, and for SRBF, IQ kernel with $s = 0.1$ is used. For SRBF, the KDE estimated $\tilde{q}$ is used. The randomly distributed $N = 2500$ data points on manifold are used for solving the eigenvalue problem.

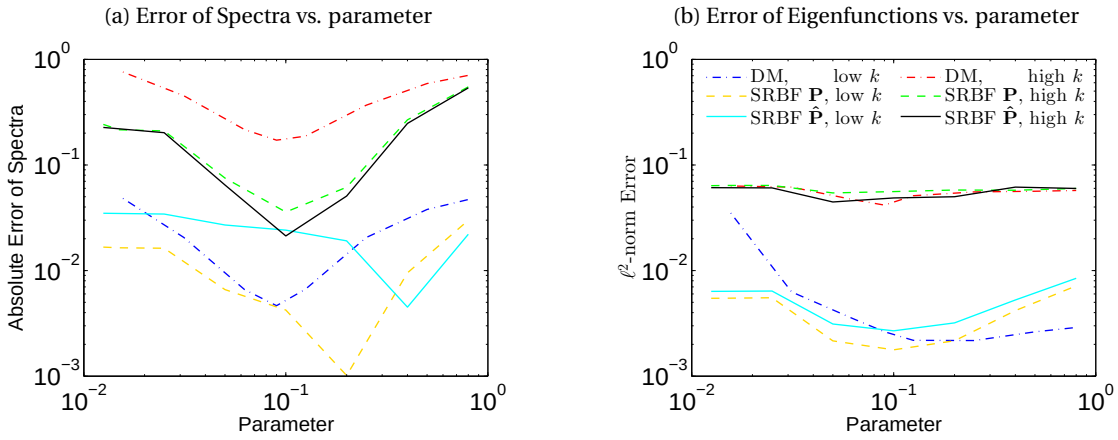

Figure 5: **2D general torus in $\mathbb{R}^{21}$.** Sensitivity of (a) eigenvalues and (b) eigenfunctions with respect to the parameter (bandwidth for DM and shape parameter for SRBF using true $\mathbf{P}$ and SRBF using estimated $\hat{\mathbf{P}}$). For each parameter, plotted are the averages of errors of eigenvalues or eigenvectors of the 2nd-5th modes for low k modes (blue, yellow and cyan curves) and of the 21st-30th modes for high k modes (red, green, and black curves). For SRBF, IQ kernel is used and shape parameter s ranges from 10^{-2} to 10^{0}. For DM, $K = 50$ nearest neighbors are used and bandwidth width ϵ ranges from 10^{-2} to 10^{0}. In this experiment, we fixed the $N = 2500$ data points which are randomly distributed on the general torus with uniform distribution in intrinsic coordinate.

In Fig. 6, we examine the convergence of eigenvalues and eigenvectors for NRBF with $\hat{\mathbf{P}}$ in the case of unknown manifold. For NRBF with $\hat{\mathbf{P}}$, since the eigenvalues are complex value, Fig. 6(a) displays all the eigenvalues on a complex plane. Here are four observations from our numerical results:

1. When N increases, more eigenvalues with large magnitudes flow to the "tail" as a cluster packet.
2. The magnitude of imaginary parts decays as N increases.
3. For the leading modes with small magnitudes, NRBF eigenvalues converge fast to the real axis and converge to the true spectra at the same time.
4. It appears that all of the eigenvalues lie in the right half plane with positive real parts for this 2D manifold as long as N is large enough ($N > 1000$). Notice that this result is consistent with the previous result reported in [28]. In that paper, the authors considered the negative definite Laplace-Beltrami operator and numerically observed that all eigenvalues are in the left half plane with negative real parts for many complicated 2D manifolds for large enough data. As far as our knowledge, Theorem 4.3 is the first theoretical justification for this empirical observation.

In Fig. 6(b)-(d), we would like to verify that the convergence rate of the NRBF is dominated by the error rate in the estimation of $\mathbf{P}$. In all numerical experiments in these panels, we solve eigenvalue problems of discrete approximation with a fixed 2500 data points as in previous examples. Here, we verify the error rate in terms of the number of points used to construct $\hat{\mathbf{P}}$, which we denote as N_p. For the 2D manifolds, we found that the convergence rate (panel (b)) for the leading 12 modes decay with the rate N_p^{-1}, which is consistent with the theoretical rate deduced in Theorem 3.2 and the discussion in Remark 5. This rate is faster than the Monte-Carlo rate even for randomly distributed data. In panels (c)-(d), we report the detailed errors in the eigenvalue and eigenvector estimation for each mode. This result suggests that if N is large enough (as we point out in bullet point 4 above), one can attain accurate estimation by improving the accuracy of the estimation of $\hat{\mathbf{P}}$ by increasing the sample size, N_p.

In Fig. 7, we examine the convergence of eigenvalues and eigenvectors for DM and SRBF with $\hat{\mathbf{P}}$ in the case of the unknown manifold. Previously in Figs. 3-5, we showed result with $N = 2500$ fixed, now we examine the convergence rate as N increases. For DM and SRBF with $\hat{\mathbf{P}}$, the Laplacian matrix is always symmetric positive definite, so that their eigenvalues and eigenvectors must be real-valued and their eigenvalues must be positive. Figures 7(c)-(e) display the errors of eigenvalues and eigenvectors for DM and SRBF for $N = 1024, 2500, 5000, 10000$. For robustness, we report estimates from 16 experiments, where each estimate corresponds to independent randomly drawn data (see thin lines). The thicker lines in each panel correspond to the average of these experiments. One can observe from Figs. 7(a) and (b) that the errors of the leading four modes for both methods are comparable and decay on the order of $N^{-1/2}$. This rate is consistent with Theorems 4.1 and 4.2 with smooth kernels for the SRBF. On the

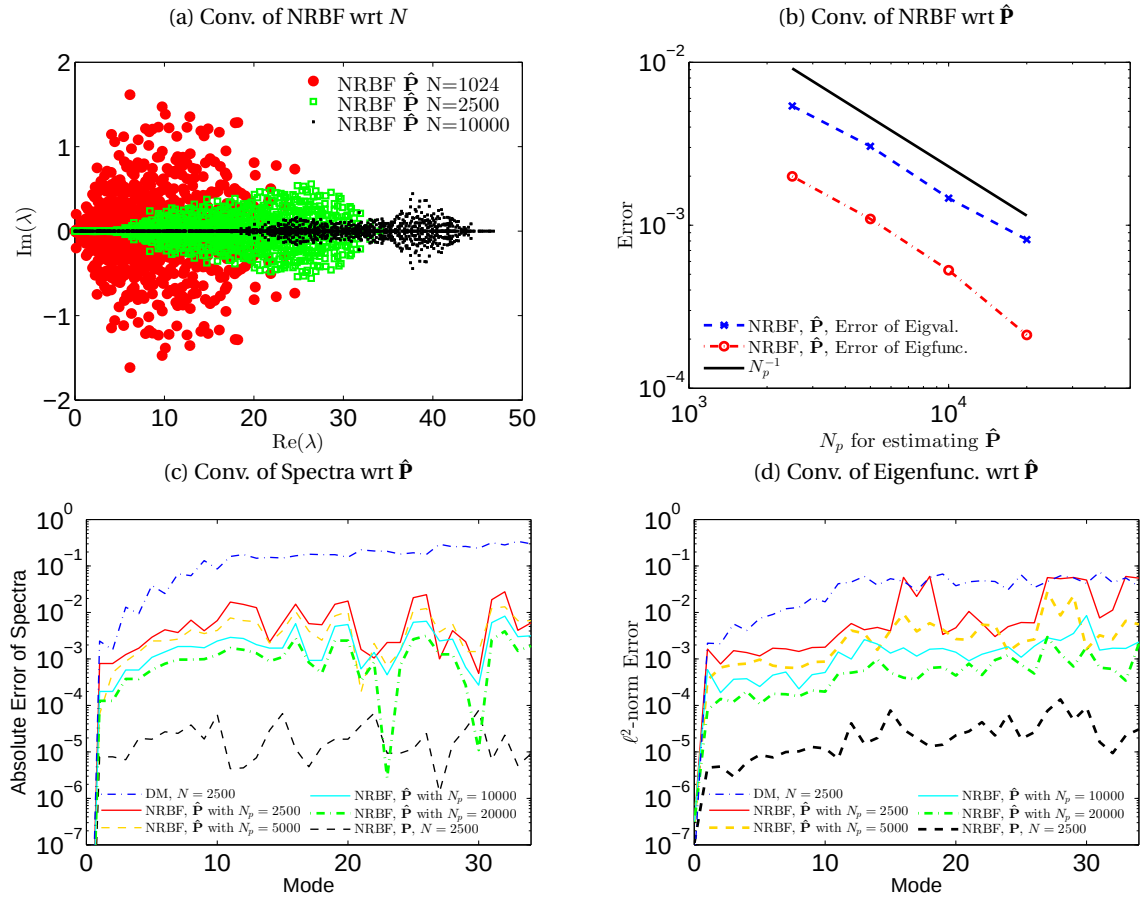

Figure 6: **2D general torus in $\mathbb{R}^{21}$.** Convergence examination for NRBF method. Here N denotes the number of points for solving eigenvalue problem and N_p denotes the number of points for estimating $\hat{\mathbf{P}}$. (a) Convergence of eigenvalues with respect to N for NRBF using $\hat{\mathbf{P}}$. In panel (a), N data points are used for both approximating $\hat{\mathbf{P}}$ and evaluating NRBF matrices. In panels (c) and (d), shown is the convergence of NRBF eigenvalues and NRBF eigenfunctions with respect to $\hat{\mathbf{P}}$ using different N_p number of points. The same 2500 data points are used for solving NRBF eigenvalue problem but different N_p data points are used for approximating $\hat{\mathbf{P}}$ at the 2500 data points. (b) For each N_p, plotted are the averages of errors of eigenvalues or eigenfunctions for the leading 12 modes (2nd-13rd modes). The convergence rate of eigenvalues and eigenfunctions are both N_p^{-1}. IQ kernel with $s = 0.5$ was fixed for all cases. The data points are randomly distributed on the general torus with uniform distribution in intrinsic coordinate.

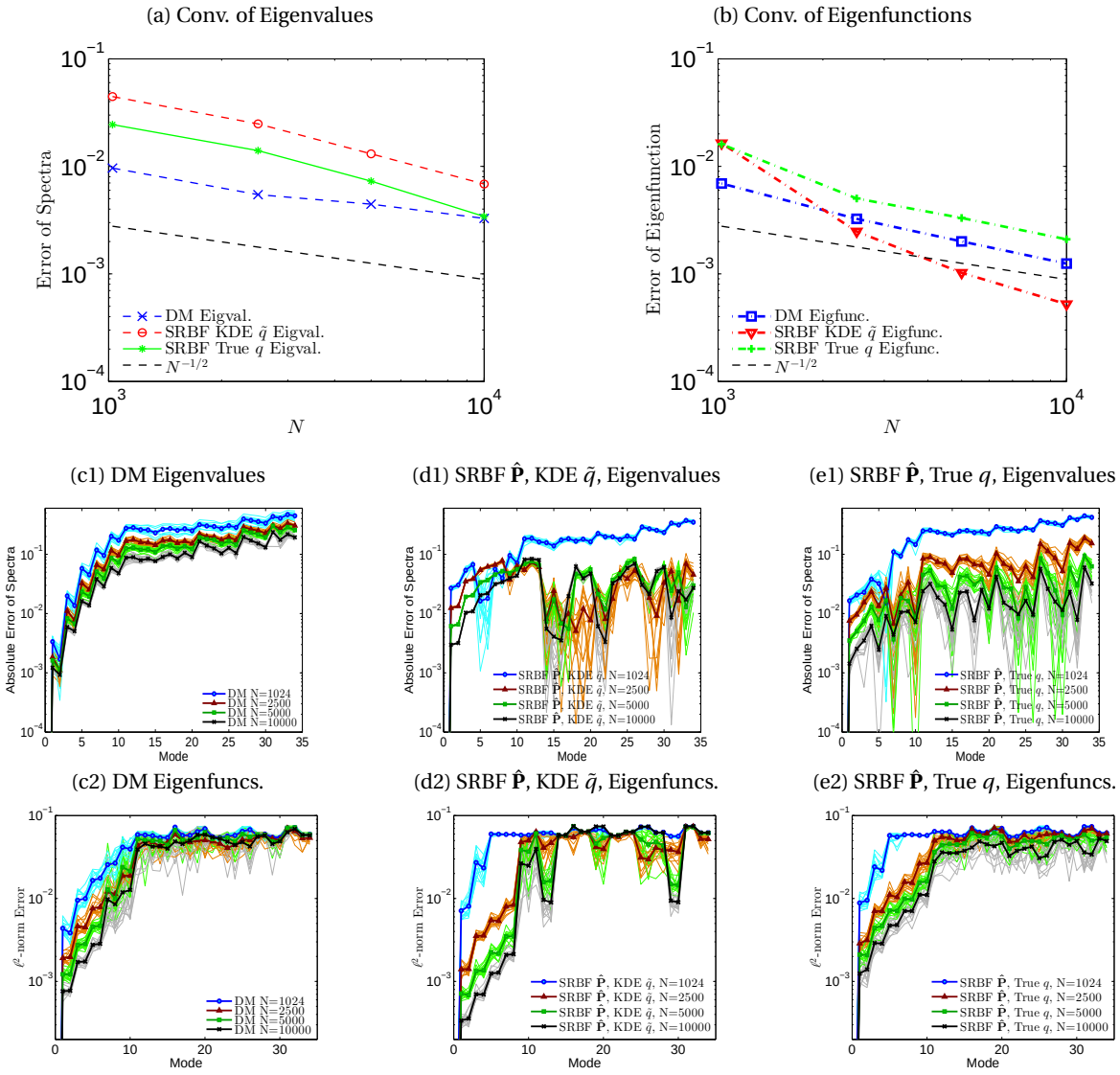

Figure 7: **2D general torus in $\mathbb{R}^{21}$.** Convergence of (a) eigenvalues and (b) eigenfunctions for DM and SRBF using $\hat{\mathbf{P}}$. For each N, plotted are the averages of errors of eigenvalues or eigenvectors for the leading four modes (2nd-5th modes). For SRBF, IQ kernel with $s = 0.1$ was fixed for different N. Comparison of errors of eigenvalues for (c1) DM, (d1) SRBF using estimated sampling density $\tilde{q}$, (e1) SRBF using the true sampling density q. Plotted in (c2)-(e2) are the corresponding comparison of errors of eigenvectors. For each N, 16 independent trials are run and depicted by light color. For each trial, the randomly distributed data points on manifold are used for computation. For each N, the average of all 16 trials are depicted by dark color.

other hand, this rate is faster than the theoretical convergence rate predicted in [8], $N^{-\frac{1}{d+4}}$. One can also see that SRBF using KDE $\tilde{q}$ (red dash-dotted curve) provides a slightly faster convergence rate for the error of eigenfunction compare to those of SRBF with analytic q, which is counterintuitive. In the next example, we will show the opposite result, which is more intuitive. Lastly, more detailed errors of the eigenvalues and eigenvectors estimates for each leading mode are reported in Figs. 7(c)-(e), from which one can also see the convergence for each leading mode and the slight advantage of SRBF over DM in the estimation of non-leading eigenvalues (consistent to the result with a fixed $N = 2500$ shown in Figs. 4-5.

6.3 4D flat torus

We consider a $d-$dimensional manifold embeded in $\mathbb{R}^{2md}$ with the following parameterization,

$$x = \frac{1}{\sqrt{1+\cdots+m^2}} \begin{pmatrix} \cos(t_1), & \sin(t_1), & \cdots & \cos(mt_1), & \sin(mt_1), \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ \cos(t_d), & \sin(t_d), & \cdots & \cos(mt_d), & \sin(mt_d) \end{pmatrix},$$

with $0 \le t_1 \le 2\pi, \cdots, 0 \le t_d \le 2\pi$. The Riemannian metric is given by a $d \times d$ identity matrix $\mathbf{I}_d$. The Laplace-Beltrami operator can be computed as $\Delta_g u = -\sum_{i=1}^{d} \frac{\partial^2 u}{\partial t_i^2}$. For each dimension t_i, the eigenvalues of operator $-\frac{\partial^2}{\partial t_i^2}$ are $\{0,1,1,4,4,\ldots,k^2,k^2,\ldots\}$. The exact spectrum and multiplicities of the Laplace-Beltrami operator on the general flat torus depends on the intrinsic dimension d of the manifold. In this section, we study the eigen modes of Laplace-Beltrami operator for a flat torus with dimension $d = 4$ and ambient space $\mathbb{R}^{2md} = \mathbb{R}^{16}$. In this case, the spectra of the flat torus can be calculated as

$$\begin{array}{lccccc} \text{spectra} & 0 & 1 & 2 & 3 & 4 \\ \text{mode } k & 1 & 2 \sim 9 & 10 \sim 33 & 34 \sim 65 & 66 \sim 89 \end{array}, \tag{34}$$

where the eigenvalues $0,1,2,3,4$ have multiplicities of $1,8,24,32,24$, respectively.

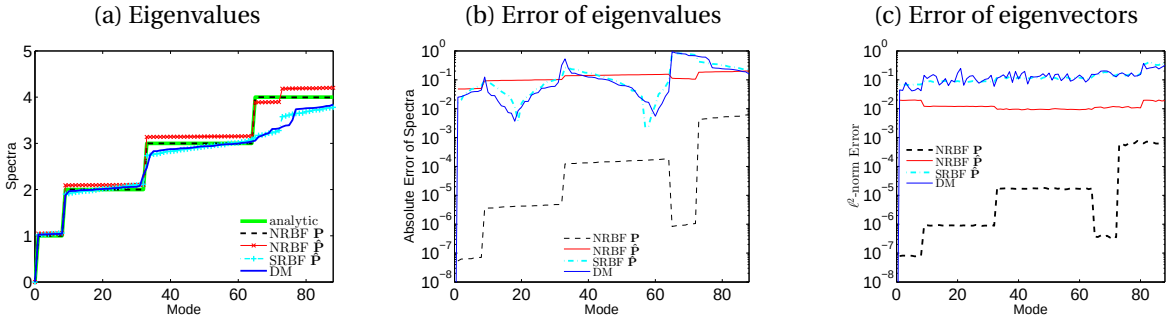

Figure 8: **4D flat torus in $\mathbb{R}^{16}$.** Comparison of (a) eigenvalues, (b) error of eigenvalues, and (c) error of eigenvectors, among NRBF, SRBF, and DM. For NRBF, GA kernel with $s = 0.5$ is used, and for SRBF, IQ kernel with $s = 0.3$ is used. The $N = 30,000$ data points are randomly distributed on the flat torus.

Numerically, data points are randomly distributed on the flat torus with uniform distribution. To apply DM, we set the K-nearest bandwidth parameter to $K \ge C\sqrt{N}$ to construct a sparse Laplacian matrix and auto-tune the kernel bandwidth ϵ accordingly. In detail, $k = 50, 70, 140$ for $5000, 10000, 30000$, respectively. To apply NRBF, we use GA kernel with $s = 0.5$. To apply SRBF, we use IQ kernel with $s = 0.3$. Figure 8 shows the results of eigenvalues and eigenfunctions for NRBF with $\mathbf{P}$ and $\hat{\mathbf{P}}$, SRBF with $\hat{\mathbf{P}}$, and DM. One can see from Figs. 8(b) and (c) that when $N = 30000$ is large enough, NRBF with $\mathbf{P}$ (black dashed curve) performs much better than the other methods. One can also see that the other three methods, NRBF with $\hat{\mathbf{P}}$ (red curve), SRBF with $\hat{\mathbf{P}}$ (cyan curve), and DM (blue curve), are comparable when the manifold is assumed to be unknown.

For the NRBF method using $\hat{\mathbf{P}}$, all the four observations in Example 6.2 persist. However, we should point out that for the last observation (for all numerical eigenvalues in the right half plane with positive real parts) to hold, the number of points N has to be around 20000 as can be seen from Fig. 9(a). When $N = 10000$ is not large enough, there are many irrelevant eigenvalues with negative real parts (blue crosses). Moreover, the leading eigenvalues (yellow dots) are not close to the truth (red circles) [the inset of Fig. 9(a)]. When N increases to 20000 or 30000, the irrelevant eigenvalues do not completely disappear (they appear near eigenvalues with larger real components) even

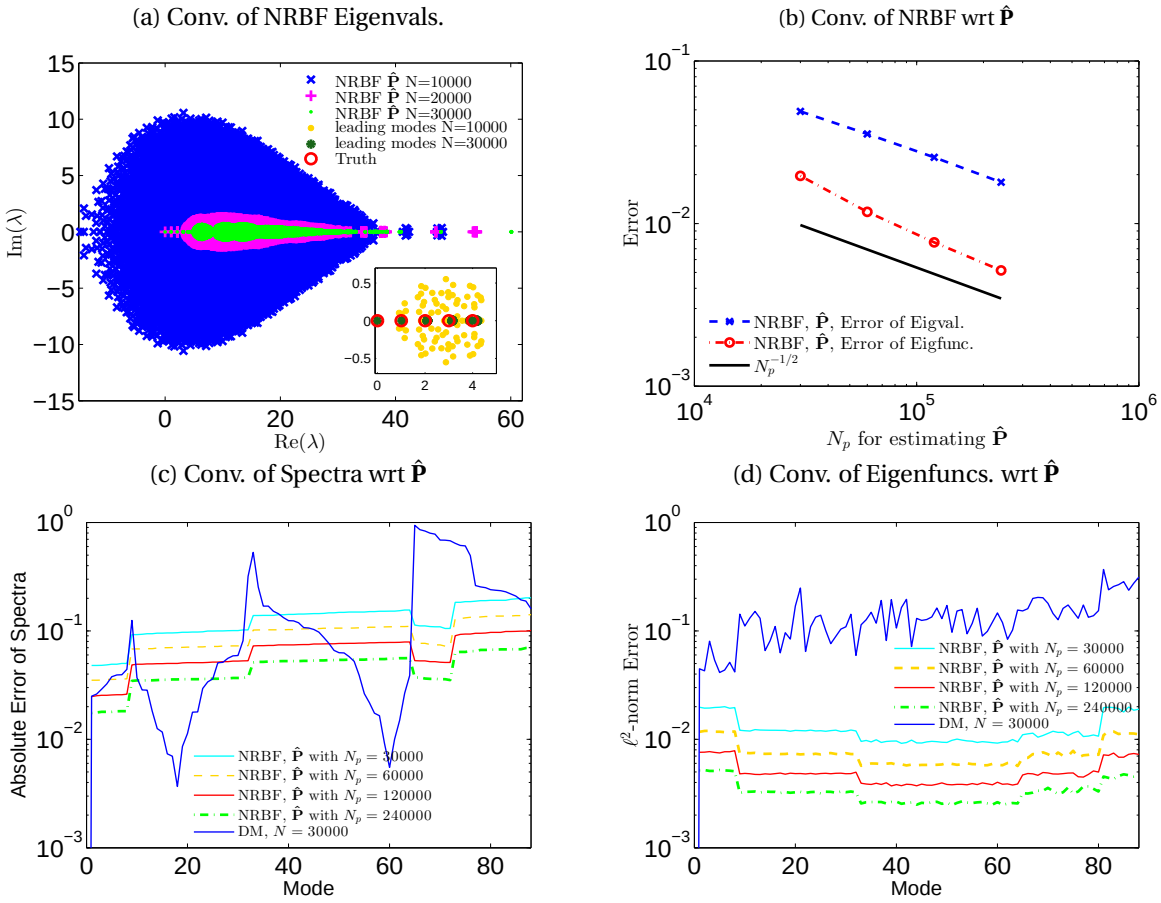

Figure 9: **4D flat torus in $\mathbb{R}^{16}$.** (a) Convergence of eigenvalues with respect to N for NRBF using $\hat{\mathbf{P}}$. For panel (a), N data points are used for both approximating $\hat{\mathbf{P}}$ and evaluating NRBF matrices. The leading modes in legend are referred to as the 89 numerical NRBF eigenvalues closest to the truth listed in equation (34). Convergence of (c) NRBF eigenvalues and (d) NRBF eigenfunctions with respect to $\hat{\mathbf{P}}$ estimated from different N_p number of points. For panels (c) and (d), the same $N = 30,000$ data points are used for evaluating NRBF matrices but different N_p data points are used for approximating $\hat{\mathbf{P}}$ at those 30,000 data points. In panel (b), plotted is the convergence rate of $O(N_p^{-1/2})$ for the leading 8 modes (2nd-9th modes). GA kernel with $s = 0.5$ was fixed for all N. The data points are randomly distributed on the flat torus. The results of DM with $N = 30,000$ are also plotted for comparison.

though NRBF eigenvalues (magenta and green dots) lie in the right half plane. Here, the leading NRBF eigenvalues (dark green dots) approximate the truth (red circles) more accurately [the inset of Fig. 9(a)].

For NRBF, we also repeat the experiment with $N = 10000$ data points using the analytic $\mathbf{P}$. We found that the profile of blue crosses in Fig. 9(a) still persists even if the analytic $\mathbf{P}$ was used to replace the approximated $\hat{\mathbf{P}}$ [not shown]. The only slight difference was that the errors of the leading modes became smaller if $\hat{\mathbf{P}}$ was replaced with $\mathbf{P}$ [red and black curves in Fig. 8(b)]. This numerical result suggests that these irrelevant eigenvalues (blue crosses in Fig. 9(a)) are due to not enough data points rather than the inaccuracy of $\hat{\mathbf{P}}$.

In contrast, when ($N = 30000$) data points are used, we already observed in Fig. 8(b) and (c) that NRBF with $\mathbf{P}$ (black dashed curve) performs much better than NRBF with $\hat{\mathbf{P}}$ (red solid curve). One can expect the errors of the NRBF with $\hat{\mathbf{P}}$ (red curve in Fig. 8(b)) to decay to the errors of the NRBF with $\mathbf{P}$ (black curve in Fig. 8(b)) as the data points used to learn the tangential projection matrices $\hat{\mathbf{P}}$ is increased (beyond 30,000 points) while the same fixed $N = 30,000$ data points are used to construct the Laplacian matrix for the eigenvalue problem. This scenario is practically useful since the approximation of $\hat{\mathbf{P}}$ is computationally cheap even with a large number of data, whereas solving the eigenvalue problem of a dense, non-symmetric RBF matrix is very expensive when N is large. In Figs. 9(b) and (c) for NRBF, we demonstrate the improvement with this estimation scenario, especially with respect to the

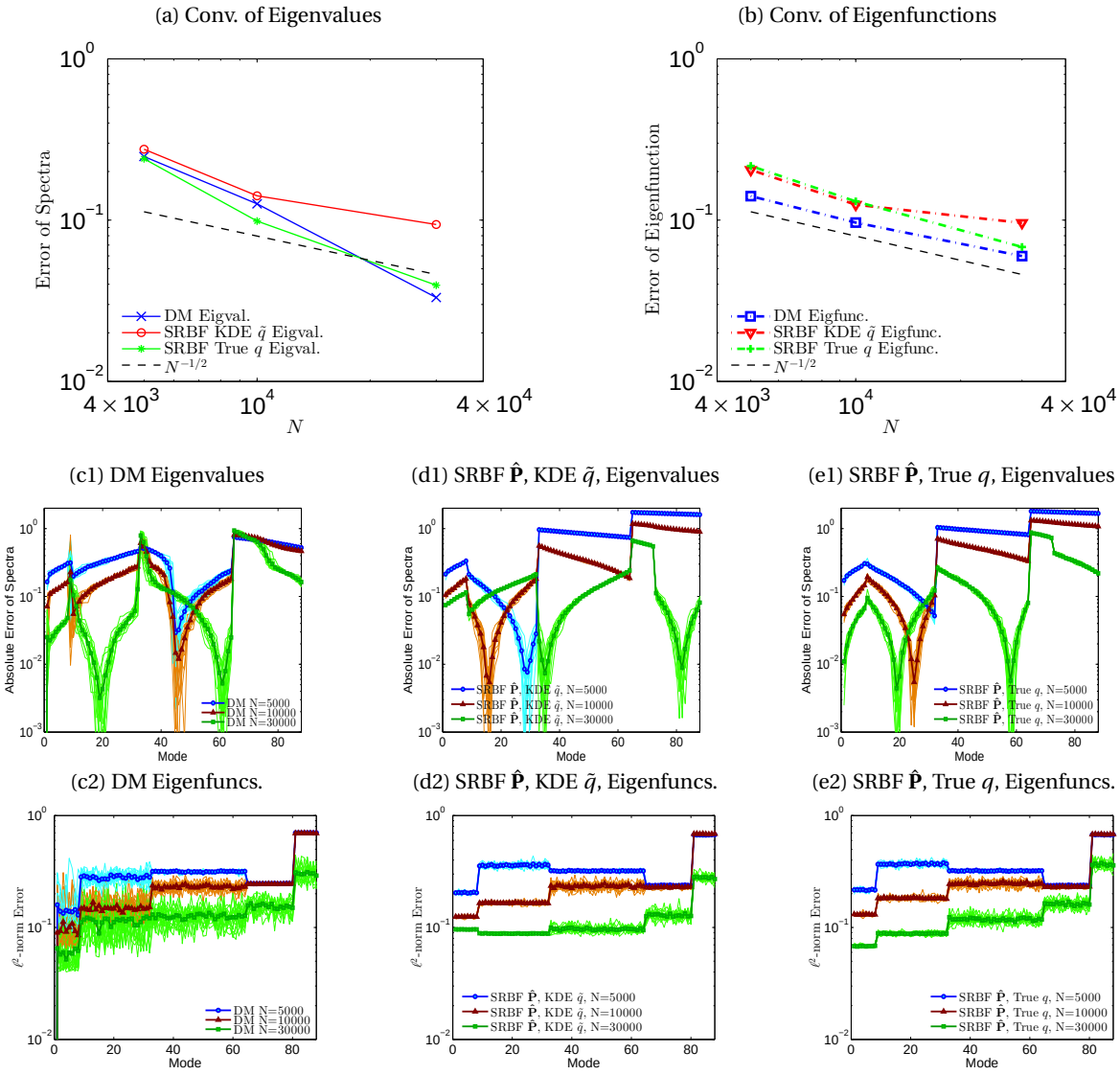

Figure 10: **4D flat torus in $\mathbb{R}^{16}$.** Convergence of (a) eigenvalues and (b) eigenvectors for DM and SRBF using $\hat{\mathbf{P}}$. For each N, plotted are the averages of errors of eigenvalues or eigenvectors for the leading eight modes (2nd-9th modes). For SRBF, IQ kernel with $s = 0.3$ was fixed for different N. Plotted in the middle row are the errors of eigenvalues for (c1) DM, (d1) SRBF using estimated sampling density $\tilde{q}$, and (e1) SRBF using the true sampling density q. Plotted in (c2)-(e2) are the corresponding errors of eigenvectors. 16 independent trials are run and depicted by light color for each trial. The average of all 16 trials are depicted by dark color. The data points for computation are randomly distributed on manifold.

number of points N_p used to construct $\hat{\mathbf{P}}$. We found that the rate is $N_p^{-1/2}$ which is consistent with our expectation as noted in Remark 5 for 4-dimensional examples.

We also inspect the convergence of NRBF as a function of the dimension of the manifold. In particular, we compare the numerical result with a 3D flat torus example in $\mathbb{R}^{12}$ as well as a 5D flat torus example in $\mathbb{R}^{20}$. For the 3D flat torus, we found that irrelevant eigenvalues appear in the left half plane when $N = 3,000$, and those eigenvalues disappear when N increases to $10,000$ [not shown]. For the 5D flat torus, we found that irrelevant eigenvalues always exist in the left half plane even when N increases to $30,000$ [not shown]. Based on these experiments and the theoretical consistency in Theorem 4.3, we believe that more data points are needed for accurate estimations of the leading order spectra of NRBF in higher dimensional cases.

We now compare results between SRBF and DM. Figure 10 displays the errors of eigenmodes for SRBF and DM for $N = 5000$,10000, and 30000. One can see from Figs. 10(a) and (b) that the errors of eigenvalues and eigenvectors of the leading eight modes decrease on order of $N^{-1/2}$ for DM and SRBF with $\hat{\mathbf{P}}$ and true q. The convergence can also be examined for each leading mode for SRBF using $\hat{\mathbf{P}}$ and true q in Figs. 10(e1) and (e2). However, the convergence is not clear for SRBF using $\hat{\mathbf{P}}$ and KDE $\tilde{q}$ as shown in Figs. 10(a)(b)(d). This implies that $\hat{\mathbf{P}}$ is accurate enough, whereas the sampling density $\tilde{q}$ is not. We suspect that the use of KDE for the density estimation in 4D may not be optimal. This leaves room for future investigations with more accurate density estimation methods.

6.4 2D sphere

In this section, we study the eigenvector field problem for the Hodge, Bochner and Lichnerowicz Laplacians, on a 2D unit sphere. The parameterization of the unit sphere is given by

$$\boldsymbol{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix}, \quad \text{for} \quad \begin{matrix} \theta \in [0,\pi] \\ \phi \in [0,2\pi) \end{matrix}, \tag{35}$$

with Riemannian metric given as,

$$g = \begin{bmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{bmatrix}. \tag{36}$$

The analytical solution of the eigenvalue problem for the Laplace-Beltrami operator on the unit sphere consists of the set of Laplace's spherical harmonics with corresponding eigenvalue $\lambda = l(l+1)$ for $l \in \mathbb{N}^+$. In particular, the leading spherical harmonics can be written in Cartesian coordinate as

$$\begin{array}{ll} \psi = x_i, & \text{for } \lambda = 2, \\ \psi = 3x_i x_k - \delta_{ik} r^2, & \text{for } \lambda = 6, \\ \psi = 15 x_i x_k x_n - 3\delta_{ik} r^2 x_n - 3\delta_{kn} r^2 x_i - 3\delta_{ni} r^2 x_k, & \text{for } \lambda = 12, \end{array} \tag{37}$$

for indices $i,k,n = 1,2,3$, where $(x_1,x_2,x_3) = (x,y,z)$ and $r^2 = x^2+y^2+z^2$.

We start with the analysis of spectra and corresponding eigenvector fields of Hodge Laplacian on a unit sphere. The Hodge Laplacian defined on a k-form is

$$\Delta_H = \mathrm{d}_{k-1}\mathrm{d}_{k-1}{}^* + \mathrm{d}_k^*\mathrm{d}_k,$$

which is self-adjoint and positive definite. Here, d is the exterior derivative and d^* is the adjoint of d given by

$$\mathrm{d}_{k-1}^* = (-1)^{d(k-1)+1} * \mathrm{d}_{d-k} * : \Omega^k(M) \to \Omega^{k-1}(M),$$

where d denotes the dimension of manifold M and $* : \Omega^k(M) \to \Omega^{d-k}(M)$ is the standard Hodge star operator. Obviously, one has $\mathrm{dd} = 0$ and $\mathrm{d}^*\mathrm{d}^* = 0$ and the following diagram:

$$\Omega^0(M) \underset{\mathrm{d}_0^*}{\overset{\mathrm{d}_0}{\rightleftarrows}} \Omega^1(M) \underset{\mathrm{d}_1^*}{\overset{\mathrm{d}_1}{\rightleftarrows}} \Omega^2(M) \underset{\mathrm{d}_2^*}{\overset{\mathrm{d}_2}{\rightleftarrows}} \cdots,$$

with $\Omega^0(M) = C^\infty(M)$. We note that Hodge Laplacian Δ_H reduces to Laplace-Beltrami Δ_M when acting on functions. To obtain the solution of eigenvalue problem for Hodge of 1-form, we can apply the results from [22] or we can provide one derivation based on the following proposition:

Proposition 6.1. *Let M be a $d-$dimensional manifold and Δ_H be the Hodge Laplacian on k-form. Then,*

(1) $\Omega^k(M) = \mathrm{im}\,\mathrm{d}_{k-1} \oplus \mathrm{im}\,\mathrm{d}_k^* \oplus \ker\Delta_H$.

(2) $\lambda(\Delta_H)\setminus\{0\} = \lambda\left(\mathrm{d}_{k-1}\mathrm{d}_{k-1}^*\right)\setminus\{0\} \cup \lambda\left(\mathrm{d}_k^*\mathrm{d}_k\right)\setminus\{0\}$.

(3) $\lambda\left(\mathrm{d}_k\mathrm{d}_k^*\right)\setminus\{0\} = \lambda\left(\mathrm{d}_k^*\mathrm{d}_k\right)\setminus\{0\}$ *for* $k = 0, 1, \ldots, d$.

(4) $\lambda\left(\mathrm{d}_k\mathrm{d}_k^*\right) = \lambda\left(\mathrm{d}_{d-k-1}^*\mathrm{d}_{d-k-1}\right)$ *for* $k = 0, 1, \ldots, d-1$.

Proof. (1) is the Hodge Decomposition Theorem. The proof of (2) and (3) can be referred to as *pp* 138 of [29]. For (4), we first show that $\lambda\left(\mathrm{d}_k\mathrm{d}_k^*\right) \subseteq \lambda\left(\mathrm{d}_{d-k-1}^*\mathrm{d}_{d-k-1}\right)$. Assume that $\mathrm{d}_k\mathrm{d}_k^*\alpha = \lambda\alpha$ for $\alpha \in \Omega^{k+1}(M)$. Taking Hodge star $*$ at both sides, one arrives at the result, $(-1)^{dk+1} * \mathrm{d}_k * \mathrm{d}_{d-k-1} * \alpha = \mathrm{d}_{d-k-1}^*\mathrm{d}_{d-k-1}(*\alpha) = \lambda * \alpha$. Vice versa one can prove $\lambda\left(\mathrm{d}_k\mathrm{d}_k^*\right) \supseteq \lambda\left(\mathrm{d}_{d-k-1}^*\mathrm{d}_{d-k-1}\right)$. □

Corollary 6.1. *Let M be a 2D surface embedded in $\mathbb{R}^3$ and $\Delta_H = \sharp\left(\mathrm{dd}^* + \mathrm{d}^*\mathrm{d}\right)\flat$ be the Hodge Laplacian on vector fields $\mathfrak{X}(M)$. Then, the non-trivial eigenvalues of Hodge Laplacian, $\lambda(\Delta_H)$, are identical with the non-trivial eigenvalues of Laplace-Beltrami operator, $\lambda(\Delta_M)\setminus\{0\}$, where the number of eigenvalues of the Hodge Laplacian doubles those of the Laplace-Beltrami operator. Specifically, the corresponding eigenvector fields to nonzero eigenvalues are $\mathbf{P}\nabla_{\mathbb{R}^3}\psi$ and $\boldsymbol{n}\times\nabla_{\mathbb{R}^3}\psi$, where ψ is the eigenfunction of Laplace-Beltrami Δ_M and $\boldsymbol{n}$ is the normal vector to surface M.*

Proof. First, we have $\lambda\left(\mathrm{d}_0\mathrm{d}_0^*\right)\setminus\{0\} = \lambda\left(\mathrm{d}_0^*\mathrm{d}_0\right)\setminus\{0\} = \lambda(\Delta_M)\setminus\{0\}$ using Proposition 6.1 (3). Second, we have $\lambda\left(\mathrm{d}_1^*\mathrm{d}_1\right)\setminus\{0\} = \lambda\left(\mathrm{d}_0\mathrm{d}_0^*\right)\setminus\{0\} = \lambda(\Delta_M)\setminus\{0\}$ using Proposition 6.1 (3) and (4). Using Proposition 6.1 (2), we obtain the first part in Corollary 6.1. That is, the eigenvalues of Hodge Laplacian for 1-forms or vector fields are exactly of eigenvalues of Laplace-Beltrami, and the number of each eigenvalue of Hodge Laplacian doubles the multiplicity of the corresponding eigenspace of the Laplace-Beltrami operator.

To simplify the notation, we use d to denote d_k for arbitrary k, which is implicitly identified by whichever $k-$form it acts on. We now examine the eigenforms. Assume that ψ is an eigenfunction for the Laplace-Beltrami operator Δ_M associated with the eigenvalue λ, that is, $\Delta_M\psi = \mathrm{d}_0^*\mathrm{d}_0\psi = \lambda\psi$. Then, one can show that $\mathrm{d}\psi \in \Omega^1(M)$ is an eigenform of Δ_H,

$$\Delta_H\mathrm{d}\psi = \left(\mathrm{dd}^* + \mathrm{d}^*\mathrm{d}\right)\mathrm{d}\psi = \mathrm{dd}^*\mathrm{d}\psi = \mathrm{d}\lambda\psi = \lambda\mathrm{d}\psi. \tag{38}$$

One can also show that $*\mathrm{d}\psi \in \Omega^1(M)$ is an eigenform,

$$\begin{aligned}\Delta_H * \mathrm{d}\psi &= \left(\mathrm{dd}^* + \mathrm{d}^*\mathrm{d}\right) * \mathrm{d}\psi = \mathrm{d}^*\mathrm{d} * \mathrm{d}\psi \\ &= -*\mathrm{d}*\mathrm{d}*\mathrm{d}\psi = *\mathrm{dd}^*\mathrm{d}\psi = *\mathrm{d}\lambda\psi = \lambda * \mathrm{d}\psi,\end{aligned} \tag{39}$$

where we have used $\mathrm{d}_1^* = (-1)^{2(1)+1} * \mathrm{d}_0 *$ and $\mathrm{d}_0^* = (-1)^1 * \mathrm{d}_1 *$. Thus, based on Hodge Decomposition Theorem 6.1 (1), harmonic forms and above eigenforms $\mathrm{d}\psi$ and $*\mathrm{d}\psi$ form a complete space of $\Omega^1(M)$.

Last, we compute the corresponding eigenvector fields. The corresponding eigenvector field for $\mathrm{d}\psi$ is

$$\sharp\mathrm{d}\psi = g^{ij}\psi_i\frac{\partial}{\partial\theta^j} = \mathbf{P}\nabla_{\mathbb{R}^3}\psi. \tag{40}$$

Assume that $\{\theta^1, \theta^2\}$ is the local normal coordinate so that the Riemannian metric is locally identity $g_{ij} = \delta_{ij}$ at point $\boldsymbol{x}$. The corresponding eigenvector field for $*\mathrm{d}\psi$ is

$$\begin{aligned}\sharp * \mathrm{d}\psi &= \sharp * \psi_i\mathrm{d}\theta^i = \sharp\delta_{ij}^{12}\psi^i\mathrm{d}\theta^j = \sharp\left(\psi^1\mathrm{d}\theta^2 - \psi^2\mathrm{d}\theta^1\right) \\ &= -\frac{\partial\psi}{\partial\theta^2}\frac{\partial}{\partial\theta^1} + \frac{\partial\psi}{\partial\theta^1}\frac{\partial}{\partial\theta^2} = -(\nabla_{\mathbb{R}^3}\psi\cdot\frac{\partial\boldsymbol{x}}{\partial\theta^2})\frac{\partial}{\partial\theta^1} + (\nabla_{\mathbb{R}^3}\psi\cdot\frac{\partial\boldsymbol{x}}{\partial\theta^1})\frac{\partial}{\partial\theta^2} \\ &= \nabla_{\mathbb{R}^3}\psi\times(\frac{\partial\boldsymbol{x}}{\partial\theta^2}\times\frac{\partial\boldsymbol{x}}{\partial\theta^1}) = \boldsymbol{n}\times\nabla_{\mathbb{R}^3}\psi,\end{aligned} \tag{41}$$

where the equality in last line holds true for 2D surface in $\mathbb{R}^3$. □

Based on (40) and (41), one can immediately obtain the leading eigenvalues and corresponding eigenvector fields for Hodge Laplacian. When $\lambda^H = 2$, the 6 corresponding Hodge eigen vector fields are

$$U_{1,2,3} = \boldsymbol{n}\times\nabla_{\mathbb{R}^3}\psi = \left[\begin{array}{c|c|c} y & -z & 0 \\ -x & 0 & z \\ 0 & x & -y \end{array}\right], U_{4,5,6} = \mathbf{P}\nabla_{\mathbb{R}^3}\psi = \left[\begin{array}{c|c|c} xz & xy & -y^2-z^2 \\ yz & -x^2-z^2 & xy \\ -x^2-y^2 & zy & xz \end{array}\right],$$

where $U_{1,2,3}$ are computed from the curl formula (41) and $U_{4,5,6}$ are computed from the projection formula (40) by taking $\psi = x_i$ $(i = 1,2,3)$ in (37). When $\lambda^H = 6$, the 10 Hodge eigenvector fields are

$$U_{7\sim11} = \boldsymbol{n} \times \nabla_{\mathbb{R}^3}\psi = \left[\begin{array}{c|c|c|c|c} -xz & y^2-z^2 & xy & 0 & -yz \\ yz & -xy & z^2-x^2 & xz & 0 \\ x^2-y^2 & xz & -yz & -xy & xy \end{array}\right],$$

$$U_{12\sim16} = \mathbf{P}\nabla_{\mathbb{R}^3}\psi = \left[\begin{array}{c|c|c|c|c} y-2x^2y & z-2x^2z & -2xyz & x-x^3 & -xy^2 \\ x-2xy^2 & -2xyz & z-2y^2z & -x^2y & y-y^3 \\ -2xyz & x-2xz^2 & y-2yz^2 & -x^2z & -y^2z \end{array}\right],$$

where $U_{7\sim11}$ are computed from the curl (41) and $U_{12\sim16}$ are computed from the projection (40) by taking $\psi = 3x_ix_k - \delta_{ik}r^2$ in (37).

For the Bochner Laplacian, we notice that it is different from the Hodge Laplacian by a Ricci tensor and Ricci curvature is constant on the sphere. Therefore, the Bochner and Hodge Laplacians share the same eigenvector fields but have different eigenvalues. We examined that $U_{1\sim6}$ are eigenvector fields for $\lambda^B = 1$ and $U_{7\sim16}$ are eigenvector fields for $\lambda^B = 5$. In general, the spectrum are $\lambda^B = l(l+1) - 1$ for the Bochner Laplacian and $\lambda^H = l(l+1)$ for the Hodge Laplacian for $l \in \mathbb{N}^+$.

For the Lichnerowicz Laplacian, we can verify that $U_{1,2,3}$ are in the null space $\ker\Delta_L$, which is often referred to as the Killing field. We can further verify that $U_{4,5,6}$ correspond to $\lambda^L = 2$, $U_{7\sim11}$ correspond to $\lambda^L = 4$, and $U_{12\sim16}$ correspond to $\lambda^L = 10$. Moreover, we verify that $U_{17\sim23}$ corresponds to $\lambda^L = 10$, where $U_{17\sim23}$ are computed from the curl equation (41) by taking $\psi = 15x_ix_kx_n - 3\delta_{ik}r^2x_n - 3\delta_{kn}r^2x_i - 3\delta_{ni}r^2x_k$. We refer to the eigenvalues λ associated with the eigenvector fields U obtained above as the analytic solutions to the eigenvalue problem of vector Laplacians.

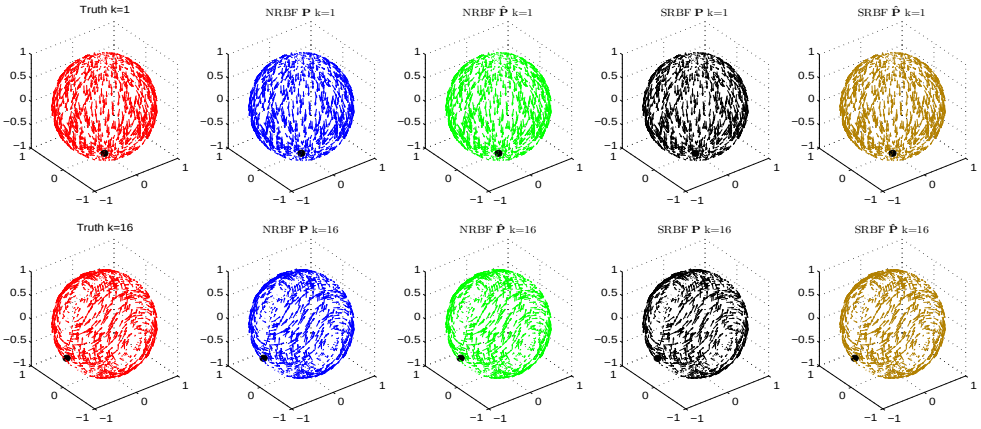

Figure 11: **2D Sphere in $\mathbb{R}^3$.** Comparison of eigen-vector fields of Bochner Laplacian for $k = 1, 16$ among NRBF using $\mathbf{P}$ and $\hat{\mathbf{P}}$, and SRBF using $\mathbf{P}$ and $\hat{\mathbf{P}}$. Black dots correspond the poles where the vector fields vanish. For NRBF, GA kernel with $s = 1.0$ is used, and for SRBF, IQ kernel with $s = 0.5$ is used. The $N = 1024$ data points are randomly distributed on manifold.

In the following, we compute eigenvalues and associated eigenvector fields of the Hodge, Bochner, and Lichnerowicz Laplacians on a unit sphere. We use uniformly random sample data points for one trial comparison in most of the following figures, except for Fig. 14 in which we use well-sampled data points for verifying the convergence of the SRBF method. Figure 11 displays the comparison of eigenvector fields of Bochner Laplacian for $k = 1$ and 16. When $\mathbf{P}$ is used, we have examined that these vector fields at each point are orthogonal to the normal directions $\boldsymbol{n} = \boldsymbol{x} = (x, y, z)$ of the sphere. When $\hat{\mathbf{P}}$ is used (i.e. in an unknown manifold scenario), the vector fields lie in the approximated tangent space which is orthogonal to the normal $\boldsymbol{x}$ within numerical accuracy. Based on Hairy ball theorem, a vector field vanishes at one or more points on the sphere. Here, we plot the poles where the vector field vanishes with black dots.

Figure 12 displays the comparison of the leading eigenvalues and eigenvector fields for the Hodge, Bochner, and Lichnerowicz Laplacians. For NRBF, one can see from Fig. 12(a)-(c) that both eigenvalues and eigenvector fields can be approximated very accurately using either the analytic $\mathbf{P}$ or estimated $\hat{\mathbf{P}}$. This small error result can be expected using analytic $\mathbf{P}$ for the known manifold. It is a little unexpected that the estimation produces such a small error

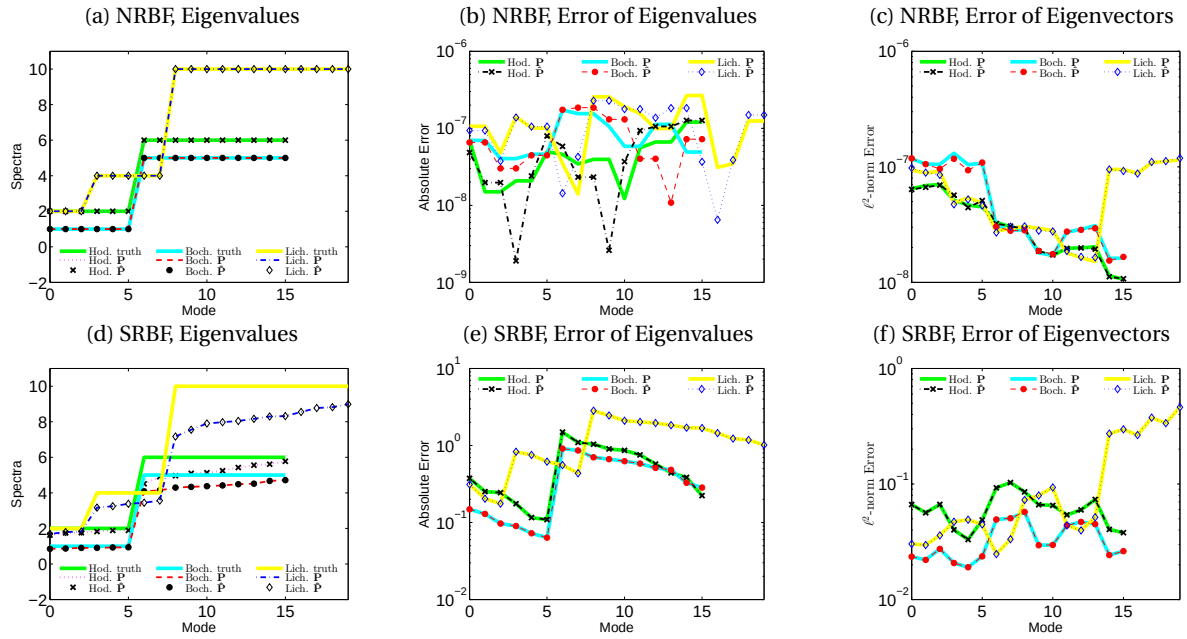

Figure 12: (Color online) 2D Sphere in $\mathbb{R}^3$. Comparison of (a)(d) eigenvalues, (b)(e) error of eigenvalues, and (c)(f) error of eigenvector-fields for Hodge, Bochner, Lichnerowicz Laplacians. (a)(b)(c) correspond to NRBF and (d)(e)(f) correspond to SRBF results. For NRBF, GA kernel with $s = 1.0$ is used, and for SRBF, IQ kernel with $s = 0.5$ is used. The data points are random with $N = 1024$.

result using estimated $\hat{\mathbf{P}}$ when the manifold is unknown. After further inspection, we found that our 2nd order SVD provides a super-convergence for $\hat{\mathbf{P}}$ on this particular 2D sphere example. For SRBF, one can see from Fig. 12(d)-(f) that eigenvalues and eigenvector fields can be approximated within certain errors for both $\mathbf{P}$ and $\hat{\mathbf{P}}$. This means that the Monte-Carlo error dominates the error for $\hat{\mathbf{P}}$ in this example. Notice that our results are more accurate compared to the estimates from the spectral exterior calculus (SEC) proposed in [4], which is a modified version of DM for computing eigenforms. Based on re-running their code, which is available from [3], using the same number of uniformly distributed random points, $N = 1024$, we found that the first six their leading spectra are $1.4564, 1.7293, 1.7489, 1.8207, 1.8773, 1.9145$, which are farther from the truth, 2, compared to the SRBF estimates reported in Table 2.

To inspect more of these spectra, we compare the leading eigenvalues of the truth, NRBF, and SRBF with $\hat{\mathbf{P}}$ up to $k = 80$ in Table 2. For NRBF, one can see from Table 2 that the leading NRBF eigenvalues are often in excellent agreement with the truth. Then, irrelevant eigenvalues may appear but are followed by the true eigenvalues again. For instance, the first 30 Bochner eigenvalues are accurately approximated, the $31^{\text{st}} \sim 40^{\text{th}}$ eigenvalues are irrelevant, and $41^{\text{st}} \sim 58^{\text{th}}$ eigenvalues become very accurate again. To further understand the irrelevant spectra phenomenon, we also solve the problem with different kernels and/or tune different shape parameters. Interestingly, we found that the true eigenvalues can always be accurately approximated even under different kernels or shape parameters. Unfortunately, the irrelevant eigenvalues also appear with different ordering for different kernels or shape parameters [not shown here]. In these experiments, the data points are always fixed. Based on the theoretical result, we expect that these irrelevant estimates can be avoided for the leading modes with larger training datasets, but cannot be completely eliminated for the non-leading modes. In contrast, SRBFs do not exhibit such an issue despite the fact that the eigenvalue estimates are not as accurate as NRBFs. Several advantages of SRBF are that their eigenvalues are real-valued, well-ordered, and do not produce irrelevant estimates.

Figure 13 shows the convergence of eigenvalues of NRBF methods for vector Laplacians. The data in this figure are randomly distributed. One can see that the four observations in Example 6.2 are still valid for the behavior of NRBF eigenvalues. Figure 14 shows the convergence of eigenvalues and eigenvectors of the SRBF method for vector Laplacians. For an efficient computational labor, we show results with well-sampled data points $\{\theta_i, \phi_j\}$, defined as equally spaced points in both direction of the intrinsic coordinates $[0, \pi] \times [0, 2\pi)$ such that the total number of points are $N = 32^2, 45^2, 71^2, 100^2$. We use the approximated $\hat{\mathbf{P}}$ and the true sampling density q to construct the Laplacians. One can see from Figs. 14(a) and (b) that the errors of eigenvalues and eigenvectors for all vector

Table 2: 2D sphere in $\mathbb{R}^3$. Comparison of eigenvalues of Bochner, Hodge, and Lichnerowicz Laplacians from NRBF and SRBF using approximated $\hat{\mathbf{P}}$. The $N = 1024$ data points are randomly distributed on the sphere. The eigenvalues of NRBF are shown with their negative absolute values when they are complex. For NRBF, GA kernel with $s = 1.0$ is used, and for SRBF, IQ kernel with $s = 0.5$ is used.

k	Boch True	Boch NRBF	Boch SRBF	Hodg True	Hodg NRBF	Hodg SRBF	Lich True	Lich NRBF	Lich SRBF	k	Boch True	Boch NRBF	Boch SRBF	Hodg True	Hodg NRBF	Hodg SRBF	Lich True	Lich NRBF	Lich SRBF
1	1	1.0	0.85	2	2.0	1.63	0	–	–	41	19	19.0	15.7	20	20.0	15.3	28	28.0	19.0
2	1	1.0	0.87	2	2.0	1.75	0	–	–	42	19	19.0	15.8	20	20.0	15.5	28	28.0	19.8
3	1	1.0	0.90	2	2.0	1.76	0	–	–	43	19	19.0	16.1	20	20.0	15.8	28	28.0	20.4
4	1	1.0	0.91	2	2.0	1.82	2	2.0	1.69	44	19	19.0	16.1	20	20.0	15.8	28	28.0	20.7
5	1	1.0	0.93	2	2.0	1.88	2	2.0	1.80	45	19	19.0	16.4	20	20.0	16.2	28	28.0	21.4
6	1	1.0	0.94	2	2.0	1.89	2	2.0	1.82	46	19	19.0	16.6	20	20.0	16.4	28	28.0	21.6
7	5	5.0	4.09	6	6.0	4.51	4	4.0	3.17	47	19	19.0	17.0	20	20.0	16.5	28	28.0	22.1
8	5	5.0	4.14	6	6.0	4.91	4	4.0	3.25	48	19	19.0	17.1	20	20.0	16.7	28	28.0	22.7
9	5	5.0	4.30	6	6.0	4.96	4	4.0	3.38	49	29	19.0	18.9	30	30.0	16.8	28	28.0	23.3
10	5	5.0	4.34	6	6.0	5.10	4	4.0	3.44	50	29	19.0	19.2	30	30.0	17.3	28	28.0	23.9
11	5	5.0	4.38	6	6.0	5.14	4	4.0	3.56	51	29	19.0	19.7	30	30.0	17.4	38	33.1	24.2
12	5	5.0	4.42	6	6.0	5.24	10	10.0	7.17	52	29	19.0	20.1	30	30.0	17.6	38	33.4	24.9
13	5	5.0	4.49	6	6.0	5.43	10	10.0	7.55	53	29	19.0	20.9	30	30.0	17.7	38	33.9	25.6
14	5	5.0	4.52	6	6.0	5.56	10	10.0	7.90	54	29	19.0	21.1	30	30.0	17.8	38	34.5	26.2
15	5	5.0	4.67	6	6.0	5.62	10	10.0	7.97	55	29	19.0	21.4	30	30.0	17.9	38	35.0	26.4
16	5	5.0	4.72	6	6.0	5.78	10	10.0	8.05	56	29	19.0	21.5	30	30.0	18.2	38	35.2	26.8
17	11	11.0	8.27	12	12.0	7.92	10	10.0	8.17	57	29	19.0	22.1	30	30.0	18.4	38	36.5	27.0
18	11	11.0	8.46	12	12.0	8.38	10	10.0	8.30	58	29	19.0	22.3	30	30.0	18.6	38	36.9	27.3
19	11	11.0	8.65	12	12.0	8.64	10	10.0	8.32	59	29	19.1	22.6	30	30.0	18.7	38	36.9	27.8
20	11	11.0	8.80	12	12.0	9.26	10	10.0	8.55	60	29	19.1	22.9	30	30.0	18.8	40	37.7	28.0
21	11	11.0	8.90	12	12.0	9.44	10	10.0	8.77	61	29	19.5	23.6	30	30.0	19.0	40	37.9	28.3
22	11	11.0	9.03	12	12.0	9.61	10	10.0	8.82	62	29	20.1	23.9	30	30.0	19.1	40	37.9	28.7
23	11	11.0	9.13	12	12.0	9.74	10	10.0	8.99	63	29	20.1	24.2	30	30.0	19.2	40	38.0	29.3
24	11	11.0	9.27	12	12.0	9.99	18	18.0	12.2	64	29	20.6	24.4	30	30.0	19.5	40	38.0	29.4
25	11	11.0	9.47	12	12.0	10.2	18	18.0	12.7	65	29	20.9	24.8	30	30.0	19.6	40	38.0	30.2
26	11	11.0	9.56	12	12.0	10.5	18	18.0	13.2	66	29	21.1	25.2	30	30.0	19.8	40	38.0	30.7
27	11	11.0	9.87	12	12.0	10.6	18	18.0	13.4	67	29	21.6	25.7	30	30.0	20.0	40	38.0	31.3
28	11	11.0	9.98	12	12.0	10.9	18	18.0	13.6	68	29	21.6	25.7	30	30.0	20.1	40	38.0	31.5
29	11	11.0	10.3	12	12.0	11.4	18	18.0	13.8	69	29	21.8	26.1	30	30.0	20.5	40	38.0	32.4
30	11	11.0	10.5	12	12.0	11.6	18	18.0	14.1	70	29	22.0	26.2	30	30.0	20.7	40	38.0	32.6
31	19	16.7	13.3	20	20.0	11.7	18	18.0	14.7	71	41	22.4	26.9	42	42.0	20.7	40	38.0	33.7
32	19	16.8	13.5	20	20.0	12.6	18	18.0	15.6	72	41	22.8	27.0	42	42.0	21.0	40	38.5	33.7
33	19	17.1	13.7	20	20.0	12.8	22	22.0	15.9	73	41	23.3	27.4	42	42.0	21.1	54	40.0	34.8
34	19	17.4	13.8	20	20.0	13.1	22	22.0	16.3	74	41	23.7	27.6	42	42.0	21.2	54	40.0	34.9
35	19	17.7	14.1	20	20.0	13.8	22	22.0	16.6	75	41	24.3	28.0	42	42.0	21.6	54	40.0	35.7
36	19	18.0	14.4	20	20.0	13.9	22	22.0	17.2	76	41	25.4	28.2	42	42.0	21.7	54	40.0	35.9
37	19	18.5	14.6	20	20.0	14.2	22	22.0	17.3	77	41	25.8	28.6	42	42.0	21.9	54	40.0	36.4
38	19	18.6	14.8	20	20.0	14.6	22	22.0	17.7	78	41	29.0	28.8	42	42.0	22.0	54	40.0	36.8
39	19	18.6	15.2	20	20.0	14.7	22	22.0	18.3	79	41	29.0	29.2	42	42.0	22.3	54	40.0	37.2
40	19	18.9	15.3	20	20.0	15.1	28	28.0	18.7	80	41	29.0	29.5	42	42.0	22.5	54	40.0	38.7

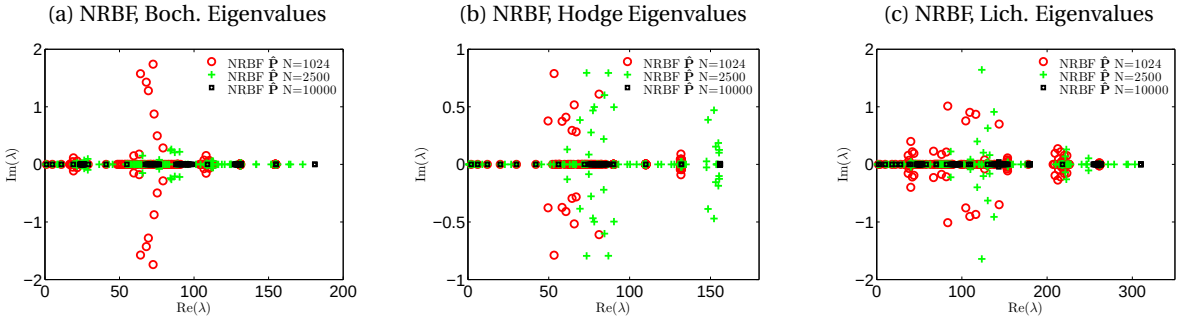

Figure 13: **2D Sphere in $\mathbb{R}^3$.** Convergence of eigenvalues for NRBF using $\hat{\mathbf{P}}$ for (a) Bochner, (b) Hodge, and (c) Lichnerowicz Laplacians. GA kernel with $s = 1.0$ was fixed for different N. The data are randomly distributed.

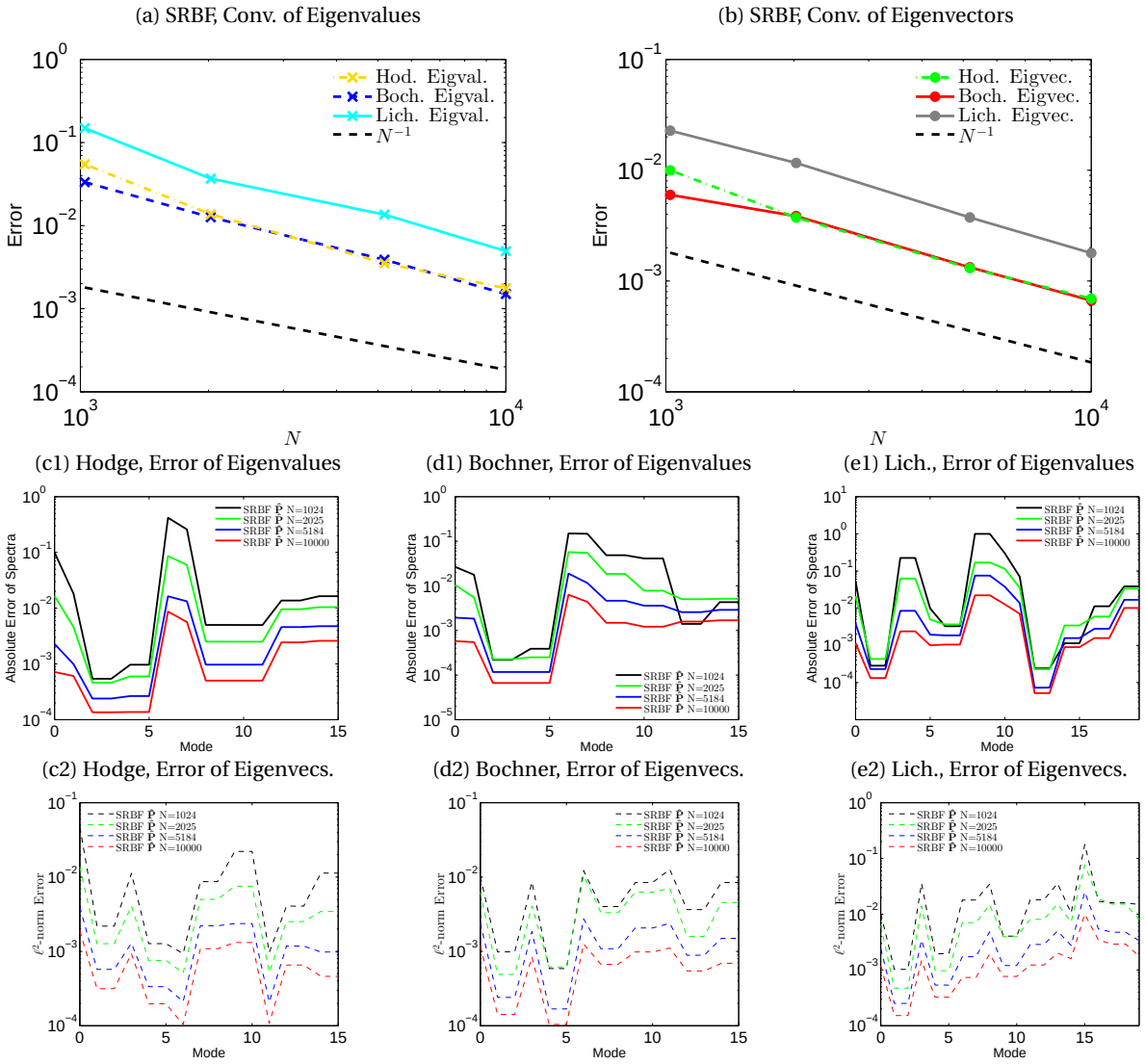

Figure 14: **2D Sphere in $\mathbb{R}^3$.** The data are uniformly well-sampled on manifold and the true sampling density q was used. Convergence of (a) eigenvalues and (b) eigenvectors for the average of the leading 16 modes for Bochner and Hodge Laplacians and for the average of the leading 20 modes for the Lichnerowicz Laplacian for SRBF using $\hat{\mathbf{P}}$ and true q. Plotted in the second row are the errors of eigenvalues for (c1) Hodge, (d1) Bochner, and (e1) Lichnerowicz Laplacians for each leading mode for SRBF. Plotted in (c2)-(e2) are the corresponding errors of eigenvectors. IQ kernel with $s = 0.5$ was fixed for SRBF for different N.

Laplacians decay on order of N^{-1}. More details can be found in Figs. 14(c)-(e) for the leading eigenmodes. Notice that for the well-sampled data, the errors here are relatively small compared to those in the case of random data shown in Figs. 12(d)-(f). Also, notice that the error rates here decay on order of N^{-1} for well-sampled data, which is faster than Monte Carlo error rates of $N^{-1/2}$ for random data.

7 Summary and Discussion

In this paper, we studied the Radial Basis Function (RBF) approximation to differential operators on smooth, closed Riemannian submanifolds of Euclidean space, identified by randomly sampled point cloud data. From this study, we conclude that:

- The classical pointwise RBF formulation is an effective and accurate method for approximating general differential operators on closed manifolds. For unknown manifolds, i.e., identified only by point clouds, the approximation error is dominated by the accuracy of the estimation of local tangent space. With sufficient regularity on the functions and manifolds, we improve the local SVD technique to achieve a second-order accuracy, by appropriately accounting for the curvatures of the manifolds.
- The pointwise non-symmetric RBF formulation can produce very accurate estimations of the leading eigensolutions of Laplacians when the size of data used to construct the RBF matrix N is large enough and the projection operator $\mathbf{P}$ is accurately estimated. We provided a theoretical result that supports this observation (see Theorem 4.3 for the Laplace-Beltrami operator and Theorem 5.3 for the Bochner Laplacian) without sharp error bounds. We empirically found that the accuracy is significantly better than those produced by the graph-based method. This conclusion is supported by the results in Figures 6(b) and 9(b)), where for sufficiently large N, one can improve the estimates by increasing $N_p \gg N$, the number of points used to estimate $\mathbf{P}$. While the second-order local SVD method is numerically efficient even for large N_p, the ultimate computational cost depends on the size of N as we are solving the eigenvalue problem of dense, non-symmetric matrices of size $N \times N$ for Laplace-Beltrami or $nN \times nN$ for vector Laplacians. When both N and N_p are not large enough, this approximation produces eigensolutions that are not consistent with the spectral properties of the Laplacians in the sense that the eigenvalue estimates can be complex-valued and do not correspond to any of the underlying real-valued spectra. Since the spectral accuracy of NRBF relies on a very accurate estimation of the projection operator $\mathbf{P}$, we believe that this method may not be robust when the data is corrupted by noises, and thus, may not be reliable for manifold learning.
- The proposed symmetric RBF approximation for Laplacians overcomes the issues in the pointwise formulation. The only caveat with the symmetric formulations is that this approximation may not be as accurate as the pointwise approximation whenever the latter works. Based on our analysis, the error is dominated by the Monte-Carlo error rate of discretizing integrals in the weak formulation of appropriate Hilbert spaces, provided sufficiently smooth kernels are used. See Theorems 4.1 and 4.2 for the approximation of eigenvalues and eigenfunctions of Laplace-Beltrami operator, respectively. See Theorems 5.1 and 5.2 for the approximation of eigenvalues and eigenvector fields of Bochner Laplacian, respectively. Empirically, we found that SRBF consistently produces more accurate estimations of the non-leading eigenvalues than graph-based methods for Laplacian on functions. On the leading eigenvalues and all eigenvectors, they are comparable. These encouraging results suggest that the symmetric RBF discrete Laplacian approximation has a huge potential for manifold learning, especially when no graph-based approximations are available.

While the second bullet point above clearly identifies the advantages and weaknesses of NRBF, the last bullet point raises an important question. Namely, what are the advantages (and disadvantages) of the symmetric RBF approximation over the graph-based approximation for manifold learning?
Advantages:

- The symmetric RBF formulation is rather general. While we have demonstrated the symmetric discrete approximation for Laplacians on functions and vector fields, in principle, one can approximate k-Laplacians. In contrast, graph-based formulation requires specific choices of kernels to estimate Laplacians on general tensor fields. As far as our knowledge, to estimate the Laplace-Beltrami operator that acts on functions, one can use Gaussian kernels with normalization induced by the diffusion maps asymptotic expansion [12]. To estimate Rough Laplacian (which is equivalent to Bochner Laplacian), one can use the same asymptotic expansion with the heat kernel for vector fields, which designed required careful approximation to a parallel transport operator [43]. The point we would like to emphasize is that one may have to design appropriate kernels for other Laplacian operators. An alternative to a somewhat graph-based approach that has guaranteed spectral properties is the Spectral Exterior Calculus [4]. This approach represents differential forms

(and vector fields) with a Galerkin expansion of eigensolutions of the graph-based approximation to the Laplace-Beltrami operator. Based on our empirical test (using their code that is available at [3]), we found that the RBF symmetric formulation gives more accurate results than SEC.

- Based on our theoretical analysis, the symmetric RBF formulation spectral convergence rate is as fast as the Monte-Carlo rate, $O(N^{-1/2})$, where N denotes the size of training data, provided that one uses smooth kernels. This rate is faster than the theoretical rate reported in [8] of order $O(N^{-\frac{1}{d+4}})$, although we numerically verify that the rates are comparable in our test examples (see Figures 7 and 10). Based on the numerical comparison with DM in the spectral estimation of the Laplace-Beltrami operator, we also found that the SRBF is more accurate than DM in the estimation of non-leading spectra and comparable to DM in the estimations of the leading spectra and all eigenvectors.

Disadvantages:

- Parameter tuning: There are two key parameters to be tuned whenever RBF is used. First, the tolerance parameter in the pseudo-inverse, which determines the rank of the discrete approximation. Smaller tolerance parameter value increases the condition number, while too large of tolerance parameter value reduces the number of the non-trivial eigensolutions that can be estimated. Second, more accurate approximations seem to occur with smaller value of shape parameters, which yields a denser matrix such that the corresponding eigenvalue problem becomes expensive, especially for large data. This is the opposite of the graph-based approach which tends to be more accurate with smaller bandwidth parameters (inversely proportional to shape parameter) which induce sparser matrices.
- For the discrete approximation corresponding to an unweighted L^2-space, one needs to de-bias the sampling distribution induced by the random data with appropriate Monte-Carlo sampling weights (or density function). When the sampling density of the data is unknown, one needs to approximate it as in the graph-based approximation. Hence, the accuracy of the estimation depends on the accuracy of the density estimation, which we did not analyze in this paper as we only implement the classical kernel density estimation technique. In the case when the operator to be estimated is defined weakly over an L^2-space that is weighted with respect to the sampling distribution, then one does not need to estimate the sampling density function whereas the graph Laplacian-based approach still needs it. Particularly, graph Laplacian approximation that based on the diffusion maps asymptotic expansion imposes a "right normalization" over the square root of the approximate density.

We conclude this paper with some open issues that arise from this study:

- For the non-symmetric approximation of the Laplacians, our convergence study only provides consistency of the approximation in the limit of large data without a sharp error bound. We believe different proving techniques are needed to achieve the error bound reflected in the empirical numerical study. As mentioned in Remark 7, convergence of eigenvector estimation remains an open problem.
- While we believe the symmetric formulation should be robust to noise error within Monte-Carlo scaling, this claim remains to be theoretically justified. Particularly, further investigation to extend Theorem 3.1 for noisy data is needed.

Acknowledgment

The research of J.H. was partially supported by the NSF grants DMS-1854299, DMS-2207328, and the ONR grant N00014-22-1-2193. S. J. was supported by the NSFC Grant No. 12101408 and the HPC Platform of ShanghaiTech University. S. J. also thanks Chengjian Yao for useful discussion.

A More Operators on Manifolds

In this appendix, we show explicitly how other operators, which are not the primary focus of this paper, can be approximated in this setting. In particular, we derive formulas for approximations of the Lichnerowicz and Hodge Laplacians, as well as the covariant derivative. Before deriving such results, we explicitly compute the matrix form of divergence of a $(2,0)$ tensor field, which will be needed in the following subsections. Throughout the following calculations, we will use the following shorthand notation for simplification:

$$v^{jk}_{,s} := v^{mk}\Gamma^{j}_{sm} + \frac{\partial v^{jk}}{\partial \theta^{s}} + v^{jm}\Gamma^{k}_{sm},$$

where $v = v^{jk}\frac{\partial}{\partial\theta^j}\otimes\frac{\partial}{\partial\theta^k}$ is a $(2,0)$ tensor field. While we do not prove convergence in the probabilistic sense, the authors suspect that techniques similar to those used in Section 4 can be used to obtain such results, at least for convergence in the pointwise sense. In Section 6, convergence is demonstrated numerically for the approximations derived in this appendix.

A.1 Derivation of divergence of a (2,0) tensor field

Let v be a $(2,0)$ tensor field of $v = v^{jk}\frac{\partial}{\partial\theta^j}\otimes\frac{\partial}{\partial\theta^k}$. The divergence of v is defined as

$$\mathrm{div}_1^r(v) = C_1^r(\nabla v)$$

for $r=1$ or $r=2$, where C_1^r denotes the contraction. More explicitly, the divergence $\mathrm{div}_1^r(v)$ can be calculated as,

$$\mathrm{div}_1^r(v) = C_1^r\left(v^{jk}_{,s}\frac{\partial}{\partial\theta^j}\otimes\frac{\partial}{\partial\theta^k}\otimes\mathrm{d}\theta^s\right) = \begin{cases} v^{jk}_{,j}\frac{\partial}{\partial\theta^k}, & r=1 \\ v^{jk}_{,k}\frac{\partial}{\partial\theta^j} & r=2 \end{cases}. \tag{42}$$

where $v^{jk}_{,s} = v^{mk}\Gamma^j_{sm} + \frac{\partial v^{jk}}{\partial\theta^s} + v^{jm}\Gamma^k_{sm}$. For any $p\in M$, we can extend v to V, defined on a neighborhood of p with ambient space representation

$$V = \frac{\partial X^s}{\partial\theta^j}v^{jk}\frac{\partial X^t}{\partial\theta^k}\frac{\partial}{\partial X^s}\otimes\frac{\partial}{\partial X^t} \equiv V^{st}\frac{\partial}{\partial X^s}\otimes\frac{\partial}{\partial X^t},$$

so that $V|_p = v|_p$ agrees at $p\in M$. We show the following formula for the divergence for a $(2,0)$ tensor field:

$$\mathrm{div}_1^r(v) = \mathscr{P}C_1^r\left(\mathscr{P}_3\bar{\nabla}_{\mathbb{R}^n}V\right), \tag{43}$$

where the right hand side is defined as

$$\begin{aligned}\mathscr{P}C_1^r\left(\mathscr{P}_3\bar{\nabla}_{\mathbb{R}^n}V\right) &\equiv \mathscr{P}C_1^r\left(\left[\delta_{cq}\frac{\partial X^c}{\partial\theta^a}g^{ab}\frac{\partial X^d}{\partial\theta^b}\mathrm{d}X^q\otimes\frac{\partial}{\partial X^d}\right]\left[\frac{\partial V^{st}}{\partial X^m}\frac{\partial}{\partial X^s}\otimes\frac{\partial}{\partial X^t}\otimes\mathrm{d}X^m\right]\right)\\ &= \mathscr{P}C_1^r\left(\delta_{cq}\frac{\partial X^c}{\partial\theta^a}g^{ab}\frac{\partial X^m}{\partial\theta^b}\frac{\partial V^{st}}{\partial X^m}\frac{\partial}{\partial X^s}\otimes\frac{\partial}{\partial X^t}\otimes\mathrm{d}X^q\right).\end{aligned}$$

Here, $\mathscr{P}_3$ acts on the last 1-form component $\mathrm{d}X^m$. For $r=1$, RHS of (43) can be calculated as,

$$\begin{aligned}\mathscr{P}C_1^r\left(\mathscr{P}_3\bar{\nabla}_{\mathbb{R}^n}V\right) &= \mathscr{P}\left(\delta_{cs}\frac{\partial X^c}{\partial\theta^a}g^{ab}\frac{\partial X^m}{\partial\theta^b}\frac{\partial V^{st}}{\partial X^m}\frac{\partial}{\partial X^t}\right) = \mathscr{P}\left(\sum_{s=1}^n\frac{\partial X^s}{\partial\theta^a}g^{ab}\frac{\partial}{\partial\theta^b}\left(\frac{\partial X^s}{\partial\theta^j}v^{jk}\frac{\partial X^t}{\partial\theta^k}\right)\frac{\partial}{\partial X^t}\right)\\ &= \mathscr{P}\left(\sum_{s=1}^n\frac{\partial X^s}{\partial\theta^a}g^{ab}\left(\frac{\partial^2X^s}{\partial\theta^b\partial\theta^j}v^{jk}\frac{\partial X^t}{\partial\theta^k}+\frac{\partial X^s}{\partial\theta^j}\frac{\partial v^{jk}}{\partial\theta^b}\frac{\partial X^t}{\partial\theta^k}+\frac{\partial X^s}{\partial\theta^j}v^{jk}\frac{\partial^2X^t}{\partial\theta^b\partial\theta^k}\right)\frac{\partial}{\partial X^t}\right),\end{aligned}$$

where the second line follows from Leibniz rule. Unraveling definitions, one obtains

$$\begin{aligned}\mathscr{P}C_1^r\left(\mathscr{P}_3\bar{\nabla}_{\mathbb{R}^n}V\right) &= \mathscr{P}\left(\left(g^{ab}v^{jk}\frac{\partial X^t}{\partial\theta^k}g_{am}\Gamma^m_{bj}+g_{aj}g^{ab}\frac{\partial v^{jk}}{\partial\theta^b}\frac{\partial X^t}{\partial\theta^k}+g_{aj}g^{ab}v^{jk}\frac{\partial^2X^t}{\partial\theta^b\partial\theta^k}\right)\frac{\partial}{\partial X^t}\right)\\ &= \left(v^{jk}\frac{\partial X^t}{\partial\theta^k}\Gamma^m_{mj}+\frac{\partial v^{jk}}{\partial\theta^j}\frac{\partial X^t}{\partial\theta^k}\right)\frac{\partial}{\partial X^t}+\mathscr{P}\left(v^{jk}\frac{\partial^2X^t}{\partial\theta^j\partial\theta^k}\frac{\partial}{\partial X^t}\right),\end{aligned}$$

where we have used that $\mathscr{P}v = v$ for $v\in\mathfrak{X}(M)$, which holds for the first and second terms above. Using the definition of projection tensor and the ambient space formulation of the connection for the third term above, one obtains

$$\begin{aligned}\mathscr{P}C_1^r\left(\mathscr{P}_3\bar{\nabla}_{\mathbb{R}^n}V\right) &= \left(v^{jk}\frac{\partial X^t}{\partial\theta^k}\Gamma^m_{mj}+\frac{\partial v^{jk}}{\partial\theta^j}\frac{\partial X^t}{\partial\theta^k}\right)\frac{\partial}{\partial X^t}+\left(\delta_{il}\frac{\partial X^i}{\partial\theta^r}g^{rp}\frac{\partial X^h}{\partial\theta^p}\mathrm{d}X^l\otimes\frac{\partial}{\partial X^h}\right)\left(v^{jk}\frac{\partial^2X^t}{\partial\theta^j\partial\theta^k}\frac{\partial}{\partial X^t}\right)\\ &= \left(v^{jk}\frac{\partial X^t}{\partial\theta^k}\Gamma^m_{mj}+\frac{\partial v^{jk}}{\partial\theta^j}\frac{\partial X^t}{\partial\theta^k}\right)\frac{\partial}{\partial X^t}+\sum_{t=1}^n g^{rp}\frac{\partial X^h}{\partial\theta^p}v^{jk}\frac{\partial^2X^t}{\partial\theta^j\partial\theta^k}\frac{\partial X^t}{\partial\theta^r}\frac{\partial}{\partial X^h}\\ &= \left(v^{jk}\frac{\partial X^t}{\partial\theta^k}\Gamma^m_{mj}+\frac{\partial v^{jk}}{\partial\theta^j}\frac{\partial X^t}{\partial\theta^k}\right)\frac{\partial}{\partial X^t}+g^{rp}v^{jk}g_{rm}\Gamma^m_{jk}\frac{\partial X^h}{\partial\theta^p}\frac{\partial}{\partial X^h}\\ &= \left(v^{jk}\Gamma^m_{mj}+\frac{\partial v^{jk}}{\partial\theta^j}+v^{jm}\Gamma^k_{jm}\right)\frac{\partial X^t}{\partial\theta^k}\frac{\partial}{\partial X^t} = v^{jk}_{,j}\frac{\partial X^t}{\partial\theta^k}\frac{\partial}{\partial X^t} = v^{jk}_{,j}\frac{\partial}{\partial\theta^k} = \mathrm{div}_1^1(v).\end{aligned}$$

Following the same steps, we can obtain $\mathrm{div}_1^r(v) = \mathscr{P}C_1^r\left(\mathscr{P}_3\bar{\nabla}_{\mathbb{R}^n}V\right)$ for $r=2$. In matrix form, the divergence of a $(2,0)$ tensor field can be written as

$$\mathrm{div}_1^r(v) = \mathbf{P}\mathrm{tr}_1^r\left(\mathbf{P}\bar{\nabla}_{\mathbb{R}^n}(V)\right).$$

The above formula is needed when deriving approximations for the Lichnerowicz and Hodge Laplacian.

A.2 RBF approximation for Lichnerowicz Laplacian

The Lichnerowicz Laplacian can be computed as

$$\begin{aligned} -\Delta_L u &\equiv 2\mathrm{div}_1^1(Su) = 2C_1^1(\nabla Su) = C_1^1\left[\left(g^{ij}u^k_{,is} + g^{ki}u^j_{,is}\right)\frac{\partial}{\partial\theta^j}\otimes\frac{\partial}{\partial\theta^k}\otimes\mathrm{d}\theta^s\right] \\ &= \left(g^{ij}u^k_{,ij} + g^{ki}u^j_{,ij}\right)\frac{\partial}{\partial\theta^k} = \mathrm{div}_1^1(\mathrm{grad}_g u) + \mathrm{div}_1^2(\mathrm{grad}_g u), \end{aligned} \tag{44}$$

where S denotes the symmetric tensor, defined as

$$Su = \frac{1}{2}\left(g^{ki}u^j_{,i} + g^{ij}u^k_{,i}\right)\frac{\partial}{\partial\theta^j}\otimes\frac{\partial}{\partial\theta^k}.$$

In matrix form, (44) can be written as,

$$-\Delta_L u = \mathrm{div}_1^1(\mathrm{grad}_g u) + \mathrm{div}_1^2(\mathrm{grad}_g u) = \mathbf{P}\mathrm{tr}_1^1\left(\mathbf{P}\nabla_{\mathbb{R}^n}\left(\mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}U\mathbf{P}\right)\right) + \mathbf{P}\mathrm{tr}_1^2\left(\mathbf{P}\nabla_{\mathbb{R}^n}\left(\mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}U\mathbf{P}\right)\right), \tag{45}$$

at each $x\in M$. In weak form, one also has the following formula for the Lichnerowicz Laplacian:

$$\Delta_L u = -\mathrm{div}_1^1\left(\mathrm{grad}_g u + (\mathrm{grad}_g u)^\top\right)$$

for a vector field $u\in\mathfrak{X}(M)$. Hence, after extension to Euclidean space, the Lichnerowicz Laplacian can be written in matrix form in the following way:

$$\begin{aligned} \bar{\Delta}_L U &= -\mathbf{P}\mathrm{tr}_1^1\left(\mathbf{P}\nabla_{\mathbb{R}^n}(\mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}U\mathbf{P} + (\mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}U\mathbf{P})^\top)\right) \\ &= -\mathbf{P}\mathrm{tr}_1^1\left(\mathbf{P}\nabla_{\mathbb{R}^n}(\mathrm{Sym}(\mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}U\mathbf{P}))\right), \end{aligned} \tag{46}$$

where $\mathrm{Sym}(\mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}U\mathbf{P}) = \mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}U\mathbf{P} + (\mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}U\mathbf{P})^\top$ is a $(2,0)$ tensor field.

One approach for approximating the Lichnerowicz Laplacian requires a fixed vector field U. This approach is useful for studying pointwise convergence numerically, but does not allow spectral convergence to be studied. A second approach does not require a fixed vector field, and thus allows spectral results to be studied. We will begin with the approach for a fixed vector field U. Following the notation and methodology used for the Bochner Laplacian, one can approximate Lichnerowicz Laplacian as,

$$\mathbf{U}\mapsto -\begin{bmatrix}\mathbf{H}_1\\ \vdots\\ \mathbf{H}_n\end{bmatrix}\cdot\mathrm{Sym}\begin{bmatrix}\mathbf{H}_1\mathbf{U}\\ \vdots\\ \mathbf{H}_n\mathbf{U}\end{bmatrix} := -\begin{bmatrix}\mathbf{H}_1\\ \vdots\\ \mathbf{H}_n\end{bmatrix}\cdot\mathrm{Sym}\begin{bmatrix}\bar{\mathbf{V}}_1\\ \vdots\\ \bar{\mathbf{V}}_n\end{bmatrix},$$

where $\bar{\mathbf{V}}_i := \mathbf{H}_i\mathbf{U} = [\mathbf{V}_{i1},\ldots,\mathbf{V}_{in}]^\top$ is the ith row of the tensor $\mathbf{V} = \left[\mathbf{V}_{ij}\right]_{i,j=1}^n$. The symmetric operator can be written in detail as

$$\mathrm{Sym}\begin{bmatrix}\bar{\mathbf{V}}_1\\ \vdots\\ \bar{\mathbf{V}}_n\end{bmatrix} = \mathrm{Sym}\begin{bmatrix}\mathbf{V}_{11} & \mathbf{V}_{12} & \cdots & \mathbf{V}_{1n}\\ \mathbf{V}_{21} & \mathbf{V}_{22} & \cdots & \vdots\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{V}_{n1} & \cdots & \cdots & \mathbf{V}_{nn}\end{bmatrix} = \begin{bmatrix}2\mathbf{V}_{11} & \mathbf{V}_{12}+\mathbf{V}_{21} & \cdots & \mathbf{V}_{1n}+\mathbf{V}_{n1}\\ \mathbf{V}_{21}+\mathbf{V}_{12} & 2\mathbf{V}_{22} & \cdots & \vdots\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{V}_{n1}+\mathbf{V}_{1n} & \cdots & \cdots & 2\mathbf{V}_{nn}\end{bmatrix}.$$

Then Lichnerowicz Laplacian can be written as

$$\mathbf{U}\mapsto -\mathbf{H}_1\left(\bar{\mathbf{V}}_1 + \mathbf{V}|_1\right) + \cdots + \mathbf{H}_n\left(\bar{\mathbf{V}}_n + \mathbf{V}|_n\right), \tag{47}$$

where $\mathbf{V}|_i = [\mathbf{V}_{1i},\ldots,\mathbf{V}_{ni}]^\top$ is the ith column of the tensor $\mathbf{V} = \left[\mathbf{V}_{ij}\right]_{i,j=1}^n$.

We now derive the second method of approximating the Lichnerowicz Laplacian which does not require a fixed vector field. Since the $(2,0)$ tensor $V_{ij} = [\text{grad}_g u]_{ij}$, we see that the ith column of $\left[V_{ij}\right]_{i,j=1}^{n}$ can be calculated as,

$$
\begin{aligned}
V|_i &= \begin{bmatrix} V_{1i} \\ V_{2i} \\ \vdots \\ V_{ni} \end{bmatrix}_{n\times 1} = \begin{bmatrix} \mathscr{G}_1 U^1 & \mathscr{G}_1 U^2 & \cdots & \mathscr{G}_1 U^n \\ \mathscr{G}_2 U^1 & \mathscr{G}_2 U^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathscr{G}_n U^1 & \cdots & \cdots & \mathscr{G}_n U^n \end{bmatrix}_{n\times n} \begin{bmatrix} P_{1i} \\ P_{2i} \\ \vdots \\ P_{ni} \end{bmatrix}_{n\times 1} \\
&= \begin{bmatrix} P_{1i}\mathscr{G}_1 U^1 + \cdots + P_{ni}\mathscr{G}_1 U^n \\ P_{1i}\mathscr{G}_2 U^1 + \cdots + P_{ni}\mathscr{G}_2 U^n \\ \vdots \\ P_{1i}\mathscr{G}_n U^1 + \cdots + P_{ni}\mathscr{G}_n U^n \end{bmatrix}_{n\times 1} = \begin{bmatrix} P_{1i}\mathscr{G}_1 & P_{2i}\mathscr{G}_1 & \cdots & P_{ni}\mathscr{G}_1 \\ P_{1i}\mathscr{G}_2 & P_{2i}\mathscr{G}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ P_{1i}\mathscr{G}_n & \cdots & \cdots & P_{ni}\mathscr{G}_n \end{bmatrix}_{n\times n} \begin{bmatrix} U^1 \\ U^2 \\ \vdots \\ U^n \end{bmatrix}_{n\times 1}.
\end{aligned} \tag{48}
$$

After discretization using RBF approximation, above (48) can be approximated by

$$
\mathbf{V}|_i = \begin{bmatrix} \text{diag}\left(\mathbf{p}_{1i}\right)\mathbf{G}_1 & \text{diag}\left(\mathbf{p}_{2i}\right)\mathbf{G}_1 & \cdots & \text{diag}\left(\mathbf{p}_{ni}\right)\mathbf{G}_1 \\ \text{diag}\left(\mathbf{p}_{1i}\right)\mathbf{G}_2 & \text{diag}\left(\mathbf{p}_{2i}\right)\mathbf{G}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \text{diag}\left(\mathbf{p}_{1i}\right)\mathbf{G}_n & \cdots & \cdots & \text{diag}\left(\mathbf{p}_{ni}\right)\mathbf{G}_n \end{bmatrix}_{Nn\times Nn} \begin{bmatrix} \mathbf{U}^1 \\ \mathbf{U}^2 \\ \vdots \\ \mathbf{U}^n \end{bmatrix}_{Nn\times 1} := \mathbf{S}_i \mathbf{U},
$$

where $\mathbf{p}_{ij} = \left(P_{ij}(\boldsymbol{x}_1), \ldots, P_{ij}(\boldsymbol{x}_N)\right) \in \mathbb{R}^{N\times 1}$. Then, Lichnerowicz Laplacian in (47) can be derived as,

$$
\begin{aligned}
\mathbf{U} &\mapsto -\mathbf{H}_1\left(\bar{\mathbf{V}}_1 + \mathbf{V}|_1\right) + \cdots + \mathbf{H}_n\left(\bar{\mathbf{V}}_n + \mathbf{V}|_n\right) \\
&= -\left[\mathbf{H}_1\left(\mathbf{H}_1 + \mathbf{S}_1\right) + \cdots + \mathbf{H}_n\left(\mathbf{H}_n + \mathbf{S}_n\right)\right]\mathbf{U}.
\end{aligned}
$$

Thus, the symmetric operator can be expressed as

$$
\text{Sym}\begin{bmatrix} \mathbf{H}_1\mathbf{U} \\ \vdots \\ \mathbf{H}_n\mathbf{U} \end{bmatrix} = \begin{bmatrix} \left(\mathbf{H}_1 + \mathbf{S}_1\right)\mathbf{U} \\ \vdots \\ \left(\mathbf{H}_n + \mathbf{S}_n\right)\mathbf{U} \end{bmatrix}, \tag{49}
$$

and Lichnerowicz Laplacian matrix given by $-\mathbf{H}_1\left(\mathbf{H}_1 + \mathbf{S}_1\right) - \cdots - \mathbf{H}_n\left(\mathbf{H}_n + \mathbf{S}_n\right)$ can be used to study the spectral properties of the Lichnerowicz Laplacian.

A.3 RBF approximation for Hodge Laplacian

The Hodge Laplacian is initially defined on k forms as

$$
\Delta_H = \left(\mathrm{d}\mathrm{d}^* + \mathrm{d}^*\mathrm{d}\right),
$$

where such Hodge Laplacian is negative definite. Here, d is the exterior derivative and d^* is the adjoint of d given by

$$
\mathrm{d}^* = (-1)^{dr+d+1} * \circ \mathrm{d} \circ * : \Omega^r(M) \to \Omega^{r-1}(M),
$$

where $*$ is Hodge star operator. See [9] for details. The Hodge Laplacian for a vector field $u = u^k \frac{\partial}{\partial \theta^k}$ is defined as

$$
\Delta_H u = \sharp\left(\mathrm{d}\mathrm{d}^* + \mathrm{d}^*\mathrm{d}\right)\flat u,
$$

where $\sharp$ and $\flat$ are the standard musical isomorphisms. First, we can compute

$$
\sharp\mathrm{d}\mathrm{d}^*\flat u = \sharp\mathrm{d}(-g^{jk}u_{j,k}) = -g^{jk}u^i_{,ij}\frac{\partial}{\partial\theta^k} = -\text{grad}_g\left(\text{div}_g(u)\right), \tag{50}
$$

where $\flat u = u_i \mathrm{d}\theta^i = g_{ij}u^j \mathrm{d}\theta^i$. Next, we can compute

$$
\begin{aligned}
\sharp\mathrm{d}^*\mathrm{d}\flat u &= \sharp\mathrm{d}^*\mathrm{d}\left(u_i \mathrm{d}\theta^i\right) = \sharp\mathrm{d}^*\left(\frac{1}{2!}\left(u_{j,i} - u_{i,j}\right)\mathrm{d}\theta^i \wedge \mathrm{d}\theta^j\right) = \sharp\left(-g^{ij}\left(u_{k,ij} - u_{i,kj}\right)\mathrm{d}\theta^k\right) \\
&= -g^{ij}u^k_{,ij}\frac{\partial}{\partial\theta^k} + g^{kj}u^i_{,ji}\frac{\partial}{\partial\theta^k} = -\text{div}_1^1(\text{grad}_g u) + \text{div}_1^2(\text{grad}_g u),
\end{aligned} \tag{51}
$$

where in last equality we have used definitions of gradient and divergence. Lastly, from (50) and (51), we can compute the Hodge Laplacian as

$$\begin{aligned}\Delta_H u &= \sharp\left(\mathrm{d}^*\mathrm{d}+\mathrm{d}\mathrm{d}^*\right)\flat u = -\left(g^{ij}u^k_{,ij} - g^{kj}u^i_{,ji} + g^{jk}u^i_{,ij}\right)\frac{\partial}{\partial\theta^k} \\ &= \Delta_B u + g^{jk}\left(u^i_{,ji} - u^i_{,ij}\right)\frac{\partial}{\partial\theta^k} = \Delta_B u + g^{jk}u^m R_{mj}\frac{\partial}{\partial\theta^k} \equiv \Delta_B u + \mathrm{Ri}(u), \end{aligned} \tag{52}$$

where the Ricci identity was used in the second line. One can see from (52) that the Hodge Laplacian Δ_H is different from the Bochner Laplacian Δ_B by a Ricci tensor $\mathrm{Ri}(u) \equiv g^{jk}u^m R_{mj}\frac{\partial}{\partial\theta^k}$ (see *pp* 478 of [9] for k-form for example).

Now we present the Hodge Laplacian written in matrix form. First, using previous formulas for the divergence and gradient, we have

$$\sharp\mathrm{d}\mathrm{d}^*\flat u = -\mathrm{grad}_g\left(\mathrm{div}_g(u)\right) = -\mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}\left(\mathrm{tr}^1_1\left[\mathbf{P}\bar{\nabla}_{\mathbb{R}^n}\mathbf{U}\right]\right).$$

Secondly, we have

$$\begin{aligned}\sharp\mathrm{d}^*\mathrm{d}\flat u &= -g^{ij}u^k_{,ij}\frac{\partial}{\partial\theta^k} + g^{kj}u^i_{,ji}\frac{\partial}{\partial\theta^k} = -\mathrm{div}^1_1(\mathrm{grad}_g u) + \mathrm{div}^2_1(\mathrm{grad}_g u) \\ &= -\mathbf{P}\mathrm{tr}^1_1\left(\mathbf{P}\bar{\nabla}_{\mathbb{R}^n}(\mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}U\mathbf{P})\right) + \mathbf{P}\mathrm{tr}^2_1\left(\mathbf{P}\bar{\nabla}_{\mathbb{R}^n}(\mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}U\mathbf{P})\right)\end{aligned}$$

Then, we can calculate Hodge Laplacian as

$$\begin{aligned}\Delta_H u &= \sharp\left(\mathrm{d}^*\mathrm{d}+\mathrm{d}\mathrm{d}^*\right)\flat u = -\left(g^{ij}u^k_{,ij} - (g^{kj}u^i_{,ji} - g^{jk}u^i_{,ij})\right)\frac{\partial}{\partial\theta^k} \\ &= -\mathrm{div}^1_1(\mathrm{grad}_g u) + \left[\mathrm{div}^2_1\left(\mathrm{grad}_g u\right) - \mathrm{grad}_g\left(\mathrm{div}(u)\right)\right] \\ &= -\mathbf{P}\mathrm{tr}^1_1\left(\mathbf{P}\bar{\nabla}_{\mathbb{R}^n}(\mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}U\mathbf{P})\right) + \left[\mathbf{P}\mathrm{tr}^2_1\left(\mathbf{P}\bar{\nabla}_{\mathbb{R}^n}(\mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n}U\mathbf{P})\right) - \mathbf{P}\mathrm{grad}_{\mathbb{R}^n}\left(\mathrm{tr}^1_1\left[\mathbf{P}\bar{\nabla}_{\mathbb{R}^n}U\right]\right)\right],\end{aligned} \tag{53}$$

where the last line holds at each $x \in M$. The first term of (53) indeed is the Bochner Laplacian $\Delta_B u$. Hence, it only remains to compute the second and third terms. At each $x \in M$, we have

$$\begin{aligned}\mathbf{P}\mathrm{tr}^2_1\left(\mathbf{P}\nabla_{\mathbb{R}^n}\mathbf{V}\right) &= \begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{n1} & \cdots & P_{nn}\end{bmatrix}_{n\times n}\begin{bmatrix} \sum_{k=1}^n \mathscr{G}_k V_{1k} \\ \vdots \\ \sum_{k=1}^n \mathscr{G}_k V_{nk}\end{bmatrix}_{n\times 1} \\ &= \begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{n1} & \cdots & P_{nn}\end{bmatrix}_{n\times n}\begin{bmatrix} \mathscr{G}_1 V_{11} + \mathscr{G}_2 V_{12} + \cdots + \mathscr{G}_n V_{1n} \\ \vdots \\ \mathscr{G}_1 V_{n1} + \mathscr{G}_2 V_{n2} + \cdots + \mathscr{G}_n V_{nn}\end{bmatrix}_{n\times 1},\end{aligned} \tag{54}$$

and the third term as

$$\mathbf{P}\mathrm{grad}_{\mathbb{R}^n}\left(\mathrm{tr}^1_1\left[\mathbf{P}\bar{\nabla}_{\mathbb{R}^n}U\right]\right) = \mathbf{P}\mathrm{grad}_{\mathbb{R}^n}\left(\sum_{k=1}^n \mathscr{G}_k U_k\right) = \begin{bmatrix}\mathscr{G}_1 \\ \vdots \\ \mathscr{G}_n\end{bmatrix}\left(\sum_{k=1}^n \mathscr{G}_k U_k\right). \tag{55}$$

Substituting (54),(55), as well as the formula for the Bochner Laplacian, into (53), we obtain the formula for Hodge Laplacian.

Another way to write the Hodge Laplacian is given by

$$-\Delta_H u = \mathrm{div}^1_1(\mathrm{grad}_g u) - \mathrm{div}^2_1(\mathrm{grad}_g u) + \mathrm{grad}_g\left(\mathrm{div}_g(u)\right) = \mathrm{div}^1_1(\mathrm{grad}_g u - (\mathrm{grad}_g u)^\top) + \mathrm{grad}_g\left(\mathrm{div}_g(u)\right). \tag{56}$$

The first term, involving the anti-symmetric part of $\mathrm{grad}_g u$, after pre-composing with the interpolating operator can be approximated by

$$\mathrm{Ant}\begin{bmatrix}\mathbf{H}_1\mathbf{U} \\ \vdots \\ \mathbf{H}_n\mathbf{U}\end{bmatrix} = \begin{bmatrix}(\mathbf{H}_1 - \mathbf{S}_1)\mathbf{U} \\ \vdots \\ (\mathbf{H}_n - \mathbf{S}_n)\mathbf{U}\end{bmatrix}.$$

Thus, we only need to consider the RBF approximation for the last term $\mathrm{grad}_g\left(\mathrm{div}_g(u)\right)$ in (56). Using the formula for gradient of a function and divergence of a vector field in (55) directly, we can obtain that

$$\begin{aligned}\mathrm{grad}_g\left(\mathrm{div}_g(u)\right) &= \begin{bmatrix}\mathscr{G}_1\\ \vdots\\ \mathscr{G}_n\end{bmatrix}\sum_{i=1}^{n}\mathscr{G}_i U^i = \begin{bmatrix}\mathscr{G}_1\sum_{i=1}^{n}\mathscr{G}_i U^i\\ \vdots\\ \mathscr{G}_n\sum_{i=1}^{n}\mathscr{G}_i U^i\end{bmatrix}\\ &= \begin{bmatrix}\mathscr{G}_1 & 0 & \cdots & 0\\ 0 & \mathscr{G}_2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \mathscr{G}_n\end{bmatrix}_{n\times n}\begin{bmatrix}\mathscr{G}_1 & \mathscr{G}_2 & \cdots & \mathscr{G}_n\\ \mathscr{G}_1 & \mathscr{G}_2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \vdots\\ \mathscr{G}_1 & \cdots & \cdots & \mathscr{G}_n\end{bmatrix}_{n\times n}\begin{bmatrix}U^1\\ U^2\\ \vdots\\ U^n\end{bmatrix}_{n\times 1}.\end{aligned}$$

After discretization, we obtain that this term can be approximated by

$$\mathbf{U}\mapsto\begin{bmatrix}\mathbf{G}_1 & 0 & \cdots & 0\\ 0 & \mathbf{G}_2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \mathbf{G}_n\end{bmatrix}_{Nn\times Nn}\begin{bmatrix}\mathbf{G}_1 & \mathbf{G}_2 & \cdots & \mathbf{G}_n\\ \mathbf{G}_1 & \mathbf{G}_2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \vdots\\ \mathbf{G}_1 & \cdots & \cdots & \mathbf{G}_n\end{bmatrix}_{Nn\times Nn}\begin{bmatrix}\mathbf{U}^1\\ \mathbf{U}^2\\ \vdots\\ \mathbf{U}^n\end{bmatrix}_{Nn\times 1}:=\mathbf{TU}.$$

It follows that the Hodge Laplacian matrix can be approximated by

$$\bar{\Delta}_H = -\left(\mathbf{H}_1\left(\mathbf{H}_1-\mathbf{S}_1\right)+\cdots+\mathbf{H}_n\left(\mathbf{H}_n-\mathbf{S}_n\right)\right)-\mathbf{T}. \tag{57}$$

A.4 RBF approximation for covariant derivative

In the following, we examine $\nabla_u y = P\bar{\nabla}_U Y$ via direct calculation, where the RHS will be defined below. Let $u = u^i\frac{\partial}{\partial\theta^i}\in\mathfrak{X}(M)$ and $y = y^i\frac{\partial}{\partial\theta^i}\in\mathfrak{X}(M)$ be two vector fields, where $u^i, y^i\in C^\infty(M)$ are smooth functions. Then, the covariant derivative is defined as

$$\nabla_u y = u^k y^i_{,k}\frac{\partial}{\partial\theta^i} = u^k\left(\frac{\partial y^i}{\partial\theta^k}+y^j\Gamma^i_{jk}\right)\frac{\partial}{\partial\theta^i},$$

where the covariant derivative operator $\nabla:\mathfrak{X}(M)\times\mathfrak{X}(M)\to\mathfrak{X}(M)$ takes vector fields u and y in $\mathfrak{X}(M)$ to a vector field $\nabla_u y$ in $\mathfrak{X}(M)$. We can rewrite covariant derivative as

$$\begin{aligned}\nabla_u y &= u^k\left(\delta_{rs}g^{ij}\frac{\partial X^r}{\partial\theta^j}\frac{\partial Y^s}{\partial\theta^k}\right)\frac{\partial}{\partial\theta^i} = u^k\delta_{rs}g^{ij}\frac{\partial X^r}{\partial\theta^j}\left(\frac{\partial Y^s}{\partial X^m}\frac{\partial X^m}{\partial\theta^k}\right)\left(\frac{\partial X^p}{\partial\theta^i}\frac{\partial}{\partial X^p}\right)\\ &= \delta_{rs}\frac{\partial X^p}{\partial\theta^i}g^{ij}\frac{\partial X^r}{\partial\theta^j}\frac{\partial Y^s}{\partial X^m}U^m\frac{\partial}{\partial X^p}\\ &= \left(\delta_{rs}\frac{\partial X^p}{\partial\theta^i}g^{ij}\frac{\partial X^r}{\partial\theta^j}\mathrm{d}X^s\otimes\frac{\partial}{\partial X^p}\right)\left(U^m\frac{\partial Y^k}{\partial X^m}\frac{\partial}{\partial X^k}\right):=\mathscr{P}\bar{\nabla}_U Y,\end{aligned} \tag{58}$$

where the first line follows from the chain rule, the second line follows from $U^m = u^k\frac{\partial X^m}{\partial\theta^k}$, and the last line defines $\bar{\nabla}_U Y\equiv U^m\frac{\partial Y^k}{\partial X^m}\frac{\partial}{\partial X^k}$ for the covariant derivative in Euclidean space. In matrix-vector form, (58) is straightforward to write as

$$\nabla_u y = \mathbf{P}\bar{\nabla}_U Y = \begin{bmatrix}P_{11} & \cdots & P_{1n}\\ \vdots & \ddots & \vdots\\ P_{n1} & \cdots & P_{nn}\end{bmatrix}\begin{bmatrix}\frac{\partial Y^1}{\partial X^1} & \cdots & \frac{\partial Y^1}{\partial X^n}\\ \vdots & \ddots & \vdots\\ \frac{\partial Y^n}{\partial X^1} & \cdots & \frac{\partial Y^n}{\partial X^n}\end{bmatrix}\begin{bmatrix}U^1\\ \vdots\\ U^n\end{bmatrix}. \tag{59}$$

The RBF approximation for covariant derivative is slightly different from above since it is nonlinear involving two vector fields U and Y as in (59). We first compute the covariant derivative $\bar{\nabla}_U Y$ in Euclidean space. For each Y^r, its RBF interpolant we can use the RBF interpolant to evaluate the rth row of $\bar{\nabla}_U Y$ in (59) at all node locations. This yields

$$\bar{\nabla}_{\mathbf{U}}\mathbf{Y}^r := \sum_{k=1}^{n}\frac{\partial I_\phi Y^r}{\partial X^k}U^k|_X.$$

Concatenating all the $\bar{\nabla}_{\mathbf{U}}\mathbf{Y}^r$ to form an augmented vector $\bar{\nabla}_{\mathbf{U}}\mathbf{Y} = (\bar{\nabla}_{\mathbf{U}}\mathbf{Y}^1,\ldots,\bar{\nabla}_{\mathbf{U}}\mathbf{Y}^n)\in\mathbb{R}^{Nn\times 1}$, we can obtain the RBF formula for covariant derivative as

$$\mathbf{P}\bar{\nabla}_U I_{\phi_s}Y = \mathbf{P}^{\otimes}_{Nn\times Nn}\bar{\nabla}_{\mathbf{U}}\mathbf{Y}. \tag{60}$$

at each point in X.

B Interpolation Error

In this appendix, we develop results regarding interpolation error in the probabilistic setting. Namely, we formally state and prove interpolation results for functions, vector fields, and (1, 1) tensor fields. The setting for this section is self-contained, and can be summarized as follows. Let $X = \{x_1, \dots, x_N\}$ be finitely many uniformly sampled data points of a closed, smooth Riemannian manifold M of dimension d, embedded in a higher dimensional euclidean space $M \subseteq \mathbb{R}^n$. We assume additionally that the injectivity radius $\iota(M)$ is bounded away from zero from below by a constant $r > 0$.

B.1 Probabilistic mesh size result

In this setting, we have the following result from [14]:

Lemma B.1. *(Proposition 14 in [14]) Let $B_\delta(x)$ denote a geodesic ball of radius δ around a point $x \in M$. For $\delta < \iota(M)/2$, we have*

$$Vol(B_\delta(x)) \geq C(d)\delta^d$$

where $C(d)$ is a constant depending only on the dimension d of the manifold.

Consider the quantity

$$h_{X,M} = \sup_{x \in M} \min_{x_i \in X} \|x - x_i\|_g$$

often referred to as mesh-size. We will show that this quantity converges to 0, in high probability, after $N \to \infty$. Here, $\|\cdot\|_g$ denotes the geodesic distance.

Lemma B.2. *We have the following result regarding the mesh size $h_{X,M}$:*

$$\mathbb{P}_{X \sim \mathscr{U}}\left(h_{X,M} > \delta\right) \leq \exp(-CN\delta^d).$$

Here, $C = C(d)/Vol(M)$, and $\mathscr{U}$ denotes the uniform distribution on M.

Proof. Suppose $h_{X,M} > \delta$, so there is $x \in M$ such that $\min_{x_i \in X} \|x - x_i\|_g \geq \delta$. So $B_\delta(x) \cap X = \emptyset$. In other words, each $x_i \in X$ is in $M \setminus B_\delta(x)$. This has measure $1 - \frac{\text{Vol}(B_\delta(x)}{\text{Vol}(M)} \leq 1 - C\delta^d$. Hence, $\min_{x_i \in X} \|x - x_i\|_g \geq \delta$ occurs with probability less than $(1 - C\delta^d)^N \leq \exp(-CN\delta^d)$. This completes the proof. □

B.2 Interpolation of functions, vector fields, and (2,0) tensors

Given a radial basis function $\phi_s : M \times M \to \mathbb{R}$, consider the interpolation map

$$I_{\phi_s} f = \sum_{i=1}^{N} c_i \phi_s(\cdot, x_i), \tag{61}$$

where c_i's are chosen so that $I_{\phi_s} f|_X = f|_X$. In the following discussion, we assume that the kernel ϕ_s is a Mercel kernel that is at least C^2. Our probabilistic interpolation result relies heavily on Theorem 10 from [27]. We adapt this theorem to our notation and state it here, for convenience.

Theorem B.1. *(adapted from Theorem 10 in [27]) Let M be a d-dimensional submanifold of $\mathbb{R}^n$, and let ϕ_s be a kernel with RKHS norm equivalent to a Sobolev space of order $\alpha > n/2$ Then there exists a constant h_M such that whenever a finite note set X satisfies $h_{X,M} < h_M$, then for all $f \in H^{\alpha - \frac{(n-d)}{2}}(M)$ we have*

$$\|f - I_{\phi_s} f\|_{L^2(M)} \leq C h_{X,M}^{\alpha - \frac{(n-d)}{2}} \|f\|_{H^{\alpha - \frac{(n-d)}{2}}(M)}.$$

The following lemma uses the above result, paired with the fact that with high probability, the mesh size $h_{X,M}$ is small as $N \to \infty$.

Lemma B.3. *Let ϕ_s be a kernel whose RKHS norm equivalent to Sobolev space of order $\alpha > n/2$. Then there is sufficiently large $N = |X|$ such that with probability higher than $1 - \frac{1}{N}$, for all $f \in H^{\alpha - \frac{(n-d)}{2}}(M)$, we have*

$$\|I_{\phi_s} f - f\|_{L^2(M)} = O\left(N^{\frac{-2\alpha + (n-d)}{2d}}\right).$$

Proof. By Lemma B.2, choosing $\delta = \left(\frac{\log N}{CN}\right)^{1/d}$, we have that $h_{X,M} = O\left(N^{-1/d}\right)$ with probability $1 - \frac{1}{N}$. For N sufficiently large, it follows that $h_{X,M} \leq h_M$, where h_M is as in Theorem 10 in [27]. Hence, the hypotheses of Theorem 10 in [27] are satisfied (with $q = 2$ and $\mu = 0$, in their notation) and we have that (in our notation)

$$\|I_{\phi_s} f - f\|_{L^2(M)} = O(h_{X,M}^{\alpha-(n-d)/2}).$$

Using the scaling of the mesh size with N, we obtain the desired result. □

For the interpolation of vector fields, recall that given a vector field $u = u^i \frac{\partial}{\partial \theta^i}$ on M, we can extend it to a smooth vector field $U = U^i \frac{\partial}{\partial X^i}$ defined on an open $\mathbb{R}^n$ neighborhood of M. Moreover, if U and V are extensions of u and v respectively, then we have that

$$\langle u, v \rangle_x = \begin{bmatrix} U^1(x) \\ U^2(x) \\ \vdots \\ U^n(x) \end{bmatrix} \cdot \begin{bmatrix} V^1(x) \\ V^2(x) \\ \vdots \\ V^n(x) \end{bmatrix}$$

at each $x \in M$, where $\langle \cdot, \cdot \rangle_x$ denotes the Riemannian innerproduct, and $\cdot$ denotes the standard euclidean innerproduct. Recall that the interpolation of a vector field is defined componentwise in the ambient space coordinates:

$$I_{\phi_s} u = I_{\phi_s} U^i \frac{\partial}{\partial X^i},$$

where again U^i denotes the ambient space components of the extension of u to a neighborhood of M. We are now ready to prove the following lemma.

Lemma B.4. *For any $u = u^i \frac{\partial}{\partial \theta^i} \in \mathfrak{X}(M)$, we have that with probability $1 - \frac{n}{N}$,*

$$\|u - I_{\phi_s} u\|_{L^2(\mathfrak{X}(M))} = O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right).$$

Proof. Notice that

$$\|u - I_{\phi_s} u\|^2_{L^2(\mathfrak{X}(M))} = \int_M \langle u - I_{\phi_s} u, u - I_{\phi_s} u \rangle_x d\mathrm{Vol}(x) = \sum_{i=1}^n \int_M (U^i(x) - I_{\phi_s} U^i(x))^2 d\mathrm{Vol}(x).$$

Using Lemma B.3 n times, we see that with probability higher than $1 - \frac{n}{N}$,

$$\left(\int_M (U^i(x) - I_{\phi_s} U^i(x))^2 d\mathrm{Vol}(x)\right)^{1/2} = O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right). \qquad i = 1, 2, \ldots, n.$$

Hence,

$$\|u - I_{\phi_s} u\|_{L^2(\mathfrak{X}(M))} = O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right).$$

which was to be demonstrated. □

Similarly, if $a = a_{ij} \frac{\partial}{\partial \theta^i} \otimes \frac{\partial}{\partial \theta^j}$ is a $(2,0)$ tensor field, we can extend a to $A = A_{ij} \frac{\partial}{\partial X^i} \otimes \frac{\partial}{\partial X^j}$, defined on an $\mathbb{R}^n$ neighborhood of M. Moreover, if a and b are $(2,0)$ tensor fields on M with extensions A and B respectively, we have that

$$\langle a, b \rangle_x = \mathrm{tr}(A(x)^\top B(x))$$

for each $x \in M$, where $A(x), B(x)$ are thought of $n \times n$ matrices with components $A_{ij}(x), B_{ij}(x)$ respectively. We can now prove the following lemma.

Lemma B.5. *For any $a = \sum_{ij} a_{ij} \frac{\partial}{\partial \theta^i} \otimes \frac{\partial}{\partial \theta^j} \in T^{(2,0)}TM)$, we have that with probability $1 - \frac{n^2}{N}$,*

$$\|a - I_{\phi_s} a\|_{L^2(T^{(2,0)}TM)} = O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right).$$

Proof. Again, notice that

$$\langle a - I_{\phi_s} a, a - I_{\phi_s} a \rangle_x = \sum_{j=1}^n \sum_{i=1}^n (A_{ij}(x) - I_{\phi_s} A_{ij}(x))^2,$$

for every $x \in M$. Integrating the above over M and using Lemma B.3 n^2 times, we see that

$$\int_M \langle a - I_{\phi_s} a, a - I_{\phi_s} a \rangle_x d\mathrm{Vol} = O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right),$$

with probability higher than $1 - \frac{n^2}{N}$. Taking the square root yields the final result. This completes the proof. □

References

[1] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[2] Mikhail Belkin and Partha Niyogi. Convergence of laplacian eigenmaps. *Advances in Neural Information Processing Systems*, 19:129, 2007.

[3] T. Berry. MATLAB Code for Diffusion Forecast and Spectral Exterior Calculus, 2018.

[4] Tyrus Berry and Dimitrios Giannakis. Spectral exterior calculus. *Communications on Pure and Applied Mathematics*, 73(4):689–770, 2020.

[5] Tyrus Berry and Timothy Sauer. Density estimation on manifolds with boundary. *Computational Statistics & Data Analysis*, 107:1–17, 2017.

[6] Martin D Buhmann. *Radial basis functions: theory and implementations*, volume 12. Cambridge university press, 2003.

[7] S. Burago and Y. Kurylev. A graph discretization of the laplace beltrami operator. *Journal of Spectral Theory*, 4:675–714, 2014.

[8] J. Calder and N. García Trillos. Improved spectral convergence rates for graph laplacians on ϵ-graphs and k-nn graphs. *arXiv preprint*, 2019.

[9] Bennett Chow, Sun-Chin Chu, David Glickenstein, Christine Guenther, Jim Isenberg, Tom Ivey, Dan Knopf, Peng Lu, Feng Luo, and Lei Ni. *The Ricci flow: techniques and applications*. American Mathematical Society Providence, 2007.

[10] Andreas Christmann and Ingo Steinwart. *Support vector machines*. Springer, 2008.

[11] Fan R.K. Chung. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.

[12] R.R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.

[13] Bruno Colbois, Ahmad El Soufi, and Alessandro Savo. Eigenvalues of the Laplacian on a compact manifold with density. *Communications in Analysis and Geometry*, 23(3):639–670, 2015.

[14] Christopher B Croke. Some isoperimetric inequalities and eigenvalue estimates. In *Annales scientifiques de l'École normale supérieure*, volume 13, pages 419–435, 1980.

[15] Felipe Cucker and Steve Smale. On the mathematical foundations of learning. *Bulletin of the American Mathematical Society*, 39(1):1–49, 2002.

[16] James W Demmel. *Applied numerical linear algebra*. SIAM, 1997.

[17] David L Donoho and Carrie Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences*, 100(10):5591–5596, 2003.

[18] D.B. Dunson, Hau-Tieng Wu, and N. Wu. Spectral convergence of graph laplacian and heat kernel reconstruction in l^∞ from random samples. *Applied and Computational Harmonic Analysis*, 55:282–336, 2021.

[19] Gerhard Dziuk and Charles M Elliott. Finite element methods for surface pdes. *Acta Numerica*, 22:289–396, 2013.

[20] Gregory E Fasshauer. *Meshfree approximation methods with MATLAB*, volume 6. World Scientific, 2007.

[21] Gregory E Fasshauer and Michael J McCourt. Stable evaluation of gaussian radial basis function interpolants. *SIAM Journal on Scientific Computing*, 34(2):A737–A762, 2012.

[22] G.B. Folland. Harmonic analysis of the de rham complex on the sphere. *Journal Für Die Reine Und Angewandte Mathematik*, 1989(398), 1989.

[23] B. Fornberg and E. Lehto. Stabilization of rbf-generated finite difference methods for convective pdes. *Journal of Computational Physics*, 230(6):2270–2285, 2011.

[24] Bengt Fornberg, Elisabeth Larsson, and Natasha Flyer. Stable computations with gaussian radial basis functions. *SIAM Journal on Scientific Computing*, 33(2):869–892, 2011.

[25] Bengt Fornberg and Grady Wright. Stable computation of multiquadric interpolants for all values of the shape parameter. *Computers & Mathematics with Applications*, 48(5-6):853–867, 2004.

[26] Edward Fuselier and Grady Wright. Stability and error estimates for vector field interpolation and decomposition on the sphere with rbfs. *SIAM J. Numer. Anal.*, 2009.

[27] Edward Fuselier and Grady B Wright. Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates. *SIAM Journal on Numerical Analysis*, 50(3):1753–1776, 2012.

[28] Edward J Fuselier and Grady B Wright. A high-order kernel method for diffusion and reaction-diffusion equations on surfaces. *Journal of Scientific Computing*, 56(3):535–565, 2013.

[29] Jürgen Jost and Jeurgen Jost. *Riemannian geometry and geometric analysis*, volume 42005. Springer, 2008.

[30] Motonobu Kanagawa, Philipp Hennig, Dino Sejdinovic, and Bharath K Sriperumbudur. Gaussian processes and kernel methods: A review on connections and equivalences. *arXiv preprint arXiv:1807.02582*, 2018.

[31] John M Lee. *Introduction to Riemannian manifolds*. Springer, 2018.

[32] Jian Liang and Hongkai Zhao. Solving partial differential equations on point clouds. *SIAM Journal on Scientific Computing*, 35(3):A1461–A1486, 2013.

[33] Francis J Narcowich and Joseph D Ward. Norms of inverses and condition numbers for matrices associated with scattered data. *Journal of Approximation Theory*, 64(1):69–94, 1991.

[34] Natasha, Flyer, Bengt, and Fornberg. Solving pdes with radial basis functions. *Acta Numerica*, 2015.

[35] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002.

[36] Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

[37] J Wilson Peoples and John Harlim. Spectral convergence of symmetrized graph laplacian on manifolds with boundary. *arXiv preprint arXiv:2110.06988*, 2021.

[38] Cécile Piret. The orthogonal gradients method: A radial basis functions method for solving partial differential equations on arbitrary surfaces. *Journal of Computational Physics*, 231(14):4662–4675, 2012.

[39] John D Pryce. *Numerical Solution of Sturm-Liouville Problems*. Oxford University Press, 1993.

[40] Steven J Ruuth and Barry Merriman. A simple embedding method for solving partial differential equations on surfaces. *Journal of Computational Physics*, 227(3):1943–1961, 2008.

[41] R. Schaback. Error estimates and condition numbers for radial basis function interpolation. *Advances in Computational Mathematics*, 3(3):251–264, 1995.

[42] Bernard W Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.

[43] Amit Singer and Hau-Tieng Wu. Spectral convergence of the connection Laplacian from random samples. *Information and Inference: A Journal of the IMA*, 6(1):58–123, 2017.

[44] Bharath K Sriperumbudur, Kenji Fukumizu, and Gert RG Lanckriet. Universality, characteristic kernels and RKHS embedding of measures. *Journal of Machine Learning Research*, 12(7), 2011.

[45] Ingo Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of machine learning research*, 2(Nov):67–93, 2001.

[46] A. E Tarwater. Parameter study of hardy's multiquadric method for scattered data interpolation. *Technical report UCRL-54670, Lawrence Livermore National Laboratory*, 1985.

[47] Nicolás García Trillos, Moritz Gerlach, Matthias Hein, and Dejan Slepčev. Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace–Beltrami operator. *Foundations of Computational Mathematics*, 20(4):827–887, 2020.

[48] Hemant Tyagi, Elıf Vural, and Pascal Frossard. Tangent space estimation for smooth embeddings of riemannian manifolds. *Information and Inference: A Journal of the IMA*, 2(1):69–114, 2013.

[49] Charles F Van Loan and G Golub. Matrix computations (johns hopkins studies in mathematical sciences). *Matrix Computations*, 1996.

[50] Ulrike Von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, pages 555–586, 2008.

[51] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4(1):389–396, 1995.

[52] H. Wendland. *Scattered Data Approximation*. Cambridge University Press, 2005.

[53] Zong-min Wu and Robert Schaback. Local error estimates for radial basis function interpolation of scattered data. *IMA journal of Numerical Analysis*, 13(1):13–27, 1993.

[54] Zhenyue Zhang and Hongyuan Zha. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM journal on scientific computing*, 26(1):313–338, 2004.