Radial Basis Function Interpolation for Manifold Learning: Revisited

J. Wilson Peoples

Department of Mathematics
Penn State University

April 26, 2023

Joint work with John Harlim and Shixiao Jiang

Overview

1. Manifold Learning
2. Interpolation, Differentiation, and Projection
3. Estimating the Laplace-Beltrami Operator
4. Estimating the Bochner Laplacian

Basic Setup

- (M, g) is a compact Riemannian manifold of dimension d.

Basic Setup

- (M, g) is a compact Riemannian manifold of dimension d.
- $M \subseteq \mathbb{R}^n$, where $n \gg d$.

Basic Setup

- (M, g) is a compact Riemannian manifold of dimension d.
- $M \subseteq \mathbb{R}^n$, where $n \gg d$.
- A data set $X = \{x_1, \ldots, x_N\}$ of points sampled i.i.d. from M.

Basic Setup

- (M, g) is a compact Riemannian manifold of dimension d.
- $M \subseteq \mathbb{R}^n$, where $n \gg d$.
- A data set $X = \{x_1, \ldots, x_N\}$ of points sampled i.i.d. from M.

Goal of Manifold Learning

Use the data to construct a matrix which approximates an operator which encodes information about the manifold.

Basic Setup

- (M, g) is a compact Riemannian manifold of dimension d.
- $M \subseteq \mathbb{R}^n$, where $n \gg d$.
- A data set $X = \{x_1, \ldots, x_N\}$ of points sampled i.i.d. from M.

Goal of Manifold Learning

Use the data to construct a matrix which approximates an operator which encodes information about the manifold.

Example

$\Delta_M : C^\infty(M) \to C^\infty(M)$, the Laplace-Beltrami operator. In smooth local coordinates,

$$\Delta_M f = \frac{-1}{\sqrt{\det g}} \frac{\partial}{\partial \theta^i} \left(g^{ij} \sqrt{\det g} \frac{\partial f}{\partial \theta^j} \right).$$

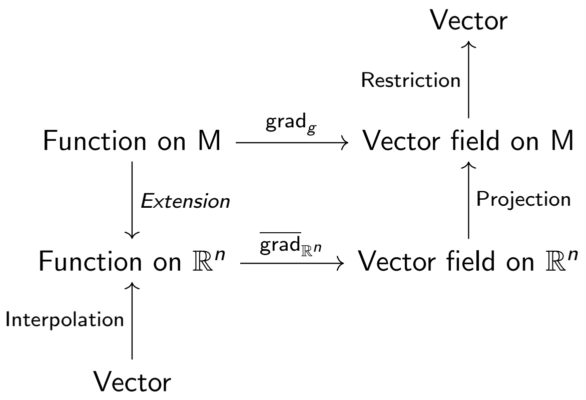
Interpolation, Differentiation, and Projection

$$\text{Function on M} \xrightarrow{\text{grad}_g} \text{Vector field on M}$$

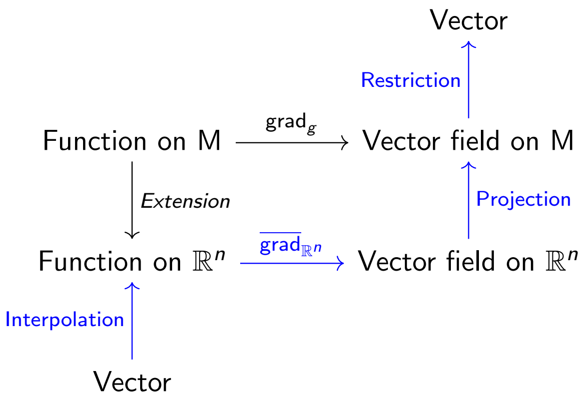
Interpolation, Differentiation, and Projection

$$\begin{array}{ccc} \text{Function on M} & \xrightarrow{\text{grad}_g} & \text{Vector field on M} \\ \Big\downarrow \textit{Extension} & & \Big\uparrow \text{Projection} \\ \text{Function on } \mathbb{R}^n & \xrightarrow{\overline{\text{grad}}_{\mathbb{R}^n}} & \text{Vector field on } \mathbb{R}^n \end{array}$$

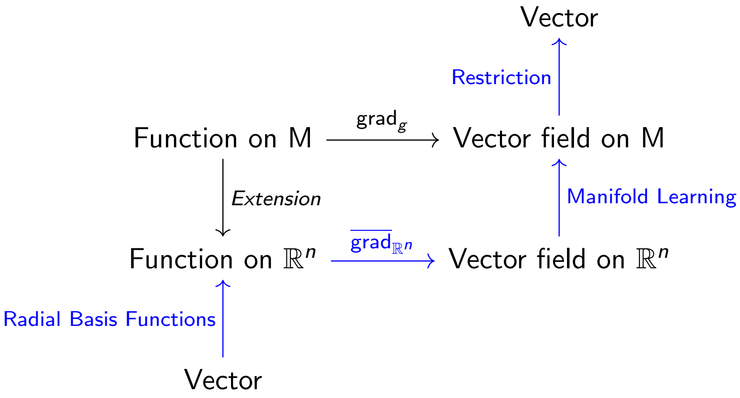
Interpolation, Differentiation, and Projection



Interpolation, Differentiation, and Projection



Interpolation, Differentiation, and Projection



New Additions

- Key fact:

$$\mathrm{grad}_g f = \mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n} F$$

New Additions

- Key fact:

$$\mathrm{grad}_g f = \mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n} F$$

- Much work has been done within in this framework (see, for instance, [Fuselier and Wright, 2012]).

New Additions

- Key fact:

$$\operatorname{grad}_g f = \mathbf{P}\overline{\operatorname{grad}}_{\mathbb{R}^n} F$$

- Much work has been done within in this framework (see, for instance, [Fuselier and Wright, 2012]).
- We adapt this framework to the manifold learning setup.

New Additions

- Key fact:

$$\text{grad}_g f = \mathbf{P}\overline{\text{grad}}_{\mathbb{R}^n} F$$

- Much work has been done within in this framework (see, for instance, [Fuselier and Wright, 2012]).
- We adapt this framework to the manifold learning setup.
- We alter existing methods to obtain symmetric estimators in this framework.

New Additions

- Key fact:
$$\mathrm{grad}_g f = \mathbf{P}\overline{\mathrm{grad}}_{\mathbb{R}^n} F$$
- Much work has been done within in this framework (see, for instance, [Fuselier and Wright, 2012]).
- We adapt this framework to the manifold learning setup.
- We alter existing methods to obtain symmetric estimators in this framework.
- We extend such approximations to more general differential operators between tensor bundles.

Learning P

- At each $x \in M$, consider $T_x M \subseteq T_x \mathbb{R}^n$.

Learning P

- At each $x \in M$, consider $T_xM \subseteq T_x\mathbb{R}^n$.
- There is an $n \times n$ matrix orthogonal projection matrix $\mathbf{P} := T_x\mathbb{R}^n \to T_xM \subseteq T_x\mathbb{R}^n$.

Learning P

- At each $x \in M$, consider $T_x M \subseteq T_x \mathbb{R}^n$.
- There is an $n \times n$ matrix orthogonal projection matrix $\mathbf{P} := T_x \mathbb{R}^n \to T_x M \subseteq T_x \mathbb{R}^n$.
- The entries of $\mathbf{P}$ can be written in terms of the Riemanian matrix g and the embedding $(\theta^1, \ldots, \theta^d) \to (X^1, \ldots, X^n)$:

$$[\mathbf{P}]_{ij} = \frac{\partial X^i}{\partial \theta^r} g^{rs} \frac{\partial X^j}{\partial \theta^s}.$$

Learning P

- At each $x \in M$, consider $T_x M \subseteq T_x \mathbb{R}^n$.
- There is an $n \times n$ matrix orthogonal projection matrix $\mathbf{P} := T_x \mathbb{R}^n \to T_x M \subseteq T_x \mathbb{R}^n$.
- The entries of $\mathbf{P}$ can be written in terms of the Riemanian matrix g and the embedding $(\theta^1, \ldots, \theta^d) \to (X^1, \ldots, X^n)$:

$$[\mathbf{P}]_{ij} = \frac{\partial X^i}{\partial \theta^r} g^{rs} \frac{\partial X^j}{\partial \theta^s}.$$

- Methods exist for approximating $\mathbf{P}$ [Zhang and Zha, 2004, Tyagi, et al, 2013].

Learning P

- Note that if $\mathbf{T} = (\tau_1, \ldots, \tau_d)$ denotes a matrix with orthonormal columns that span $T_x M$, then $\mathbf{P} := \mathbf{T}\mathbf{T}^\top$.

Learning $\mathbf{P}$

- Note that if $\mathbf{T} = (\tau_1, \ldots, \tau_d)$ denotes a matrix with orthonormal columns that span $T_x M$, then $\mathbf{P} := \mathbf{T}\mathbf{T}^\top$.
- Methods for learning $\mathbf{P}$ relate the distance $y - x$, for points y close to x, to the directions τ_i.

Learning $\mathbf{P}$

- Note that if $\mathbf{T} = (\tau_1, \ldots, \tau_d)$ denotes a matrix with orthonormal columns that span $T_x M$, then $\mathbf{P} := \mathbf{T}\mathbf{T}^\top$.
- Methods for learning $\mathbf{P}$ relate the distance $y - x$, for points y close to x, to the directions τ_i.
- We propose a novel, related method which achieves a faster convergence rate by correcting for curvature.

Second order approximation

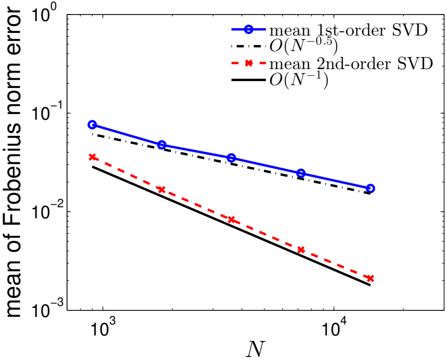


Figure: Mean of Frobenius error $\|\mathbf{P} - \hat{\mathbf{P}}\|_F$ as a function of N, on the $2D$ torus in $\mathbb{R}^3$.

Interpolation using RBFs

- Given function values $\mathbf{f} := (f(x_1), \ldots, f(x_N))^\top$ at $X = \{x_j\}_{j=1}^N$, the radial basis function (RBF) interpolant of f at x takes the form

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

Interpolation using RBFs

- Given function values $\mathbf{f} := (f(x_1), \ldots, f(x_N))^\top$ at $X = \{x_j\}_{j=1}^N$, the radial basis function (RBF) interpolant of f at x takes the form

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

- One common choice of kernel is the Gaussian $\phi_s(r) = \exp(-(sr)^2)$.

Interpolation using RBFs

- Given function values $\mathbf{f} := (f(x_1), \ldots, f(x_N))^\top$ at $X = \{x_j\}_{j=1}^N$, the radial basis function (RBF) interpolant of f at x takes the form

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

- One common choice of kernel is the Gaussian $\phi_s(r) = \exp(-(sr)^2)$.
- Theoretical advantage: A certain class of RBFs have Reproducing Kernel Hilbert Space norm which is equivalent to Sobolev space norms [Fuselier and Wright, 2012].

Laplace-Beltrami Estimators

- Let $\mathbf{G} : \mathbb{R}^N \to \mathbb{R}^{nN}$ denote the estimate of grad_g from before.

Laplace-Beltrami Estimators

- Let $\mathbf{G} : \mathbb{R}^N \to \mathbb{R}^{nN}$ denote the estimate of grad_g from before.
- In analogy with the Laplace-Beltrami operator, it's natural to consider

$$\mathbf{G}^\top \mathbf{G} : \mathbb{R}^N \to \mathbb{R}^N$$

as an estimator of Δ_M.

Laplace-Beltrami Estimators

- Let $\mathbf{G} : \mathbb{R}^N \to \mathbb{R}^{nN}$ denote the estimate of grad_g from before.
- In analogy with the Laplace-Beltrami operator, it's natural to consider

$$\mathbf{G}^\top \mathbf{G} : \mathbb{R}^N \to \mathbb{R}^N$$

  as an estimator of Δ_M.
- Alternatively, one can estimate Δ_M with

$$\mathbf{L}_N := -(\mathbf{G}_1\mathbf{G}_1 + \cdots + \mathbf{G}_n\mathbf{G}_n) : \mathbb{R}^N \to \mathbb{R}^N.$$

Convergence of Eigenvalues

Theorem

Let λ_i denote the i-th eigenvalue of Δ_M, enumerated $\lambda_1 \leq \lambda_2 \leq \dots$. For any i, there exists a sequence $\hat{\lambda}_i^{(N)}$ of eigenvalues of $\mathbf{L}_N$ (or $\mathbf{G}^\top \mathbf{G}$) such that

$$\left|\lambda_i - \hat{\lambda}_i^{(N)}\right| \to 0$$

with high probability as $N \to \infty$.

Convergence of Eigenvalues

Theorem

Let λ_i denote the i-th eigenvalue of Δ_M, enumerated $\lambda_1 \leq \lambda_2 \leq \dots$. For any i, there exists a sequence $\hat{\lambda}_i^{(N)}$ of eigenvalues of $\mathbf{L}_N$ (or $\mathbf{G}^\top \mathbf{G}$) such that

$$\left| \lambda_i - \hat{\lambda}_i^{(N)} \right| \to 0$$

with high probability as $N \to \infty$.

- For the symmetric estimator, we also prove a rate: $O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right)$.

Convergence of Eigenvalues

Theorem

Let λ_i denote the i-th eigenvalue of Δ_M, enumerated $\lambda_1 \leq \lambda_2 \leq \ldots$. For any i, there exists a sequence $\hat{\lambda}_i^{(N)}$ of eigenvalues of $\mathbf{L}_N$ (or $\mathbf{G}^\top \mathbf{G}$) such that

$$\left| \lambda_i - \hat{\lambda}_i^{(N)} \right| \to 0$$

with high probability as $N \to \infty$.

- For the symmetric estimator, we also prove a rate: $O\left(\frac{1}{\sqrt{N}}\right) + O\left(N^{\frac{-2\alpha+(n-d)}{2d}}\right)$.
- Similar results regarding the convergence of eigenvectors were obtained for $\mathbf{G}^\top \mathbf{G}$, but not for $\mathbf{L}_N$.

Numerical Results

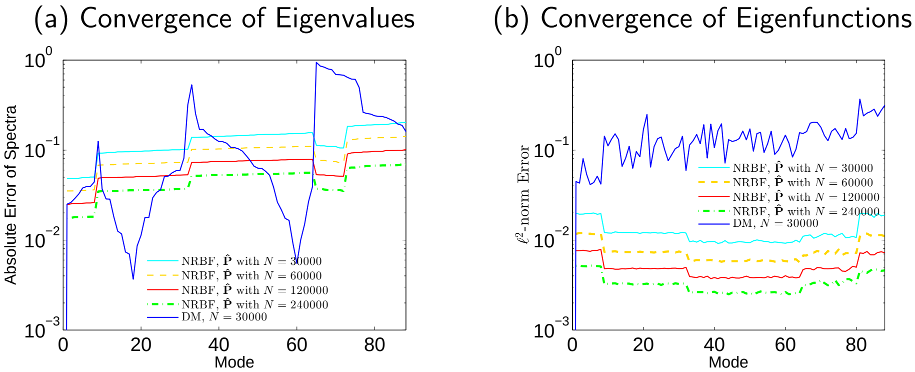


Figure: 4D flat torus in $\mathbb{R}^{16}$. Convergence of (a) NRBF eigenvalues and (b) NRBF eigenfunctions with respect to learned $\hat{\mathbf{P}}$. GA kernel with $s = 0.5$ was fixed for all N. The data points are randomly distributed on the flat torus. The results of DM with $N = 30,000$ are also plotted for comparison.

Numerical Results

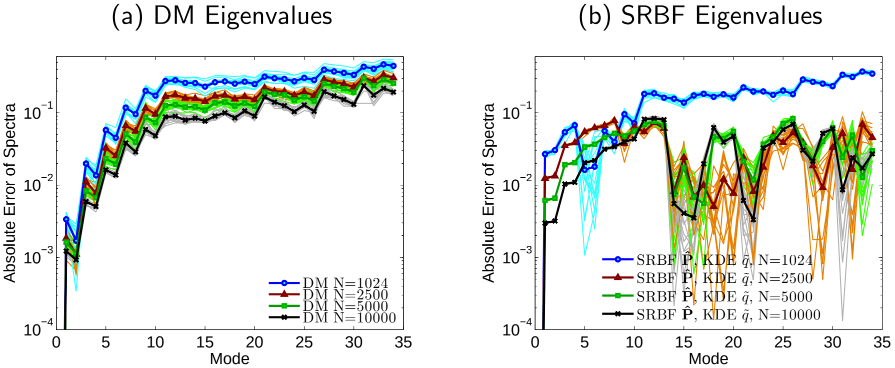


Figure: 2D general torus in $\mathbb{R}^{21}$. Comparison of errors of eigenvalues for (a) DM, (b) SRBF. For each N, 16 independent trials are run and depicted by light color. For each N, the average of all 16 trials are depicted by dark color.

Numerical Results

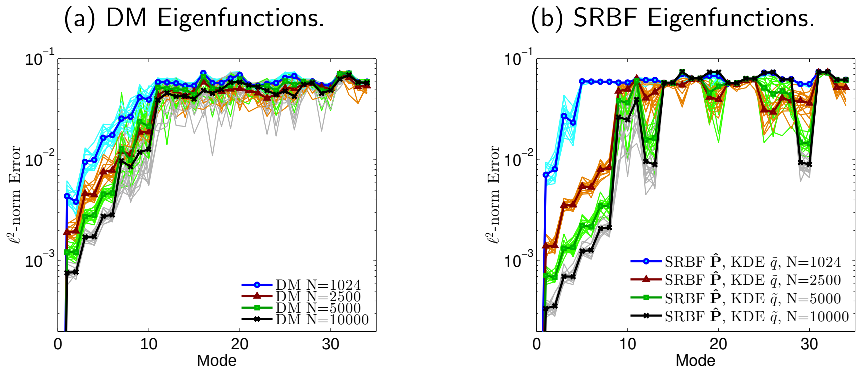


Figure: 2D general torus in $\mathbb{R}^{21}$. Comparison of errors of eigenfunctions for (a) DM, (b) SRBF. For each N, 16 independent trials are run and depicted by light color. For each N, the average of all 16 trials are depicted by dark color.

Matrix Estimator of Bochner Laplacian

Definition

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

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

Matrix Estimator of Bochner Laplacian

Definition

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

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

- Here $\mathrm{grad}_g u$ is the gradient of a vector field, and results in a $(2, 0)$ tensor field.

Matrix Estimator of Bochner Laplacian

Definition

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

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

- Here $\mathrm{grad}_g u$ is the gradient of a vector field, and results in a $(2, 0)$ tensor field.
- In local coordinates,

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

Spectral Convergence Result for Bochner Laplacian

- Similar to before, one can approximate grad_g acting on vector fields with a matrix $\mathbf{H} : \mathbb{R}^{nN} \to \mathbb{R}^{Nn \times n}$.

Spectral Convergence Result for Bochner Laplacian

- Similar to before, one can approximate grad_g acting on vector fields with a matrix $\mathbf{H} : \mathbb{R}^{nN} \to \mathbb{R}^{Nn \times n}$.
- This gives rise to a symmetric estimator $\mathbf{H}^\top \mathbf{H}$ of Δ_B.

Spectral Convergence Result for Bochner Laplacian

- Similar to before, one can approximate grad_g acting on vector fields with a matrix $\mathbf{H} : \mathbb{R}^{nN} \to \mathbb{R}^{Nn \times n}$.
- This gives rise to a symmetric estimator $\mathbf{H}^\top \mathbf{H}$ of Δ_B.
- Similar to before, there is also a non-symmetric estimator $\mathbf{L}_B$ of Δ_B.

Spectral Convergence Result for Bochner Laplacian

- Similar to before, one can approximate grad_g acting on vector fields with a matrix $\mathbf{H} : \mathbb{R}^{nN} \to \mathbb{R}^{Nn \times n}$.
- This gives rise to a symmetric estimator $\mathbf{H}^\top \mathbf{H}$ of Δ_B.
- Similar to before, there is also a non-symmetric estimator $\mathbf{L}_B$ of Δ_B.

Theorem

Let λ_i denote the i-th eigenvalue of Δ_B, enumerated $\lambda_1 \leq \lambda_2 \leq \dots$. For any i, there exists a sequence of $\hat{\lambda}_i^{(N)}$ of eigenvalues of $\mathbf{L}_B$ (or $\mathbf{H}^\top \mathbf{H}$) such that

$$\left| \lambda_i - \hat{\lambda}_i^{(N)} \right| \to 0$$

with high probability as $N \to \infty$.

Numerical Results

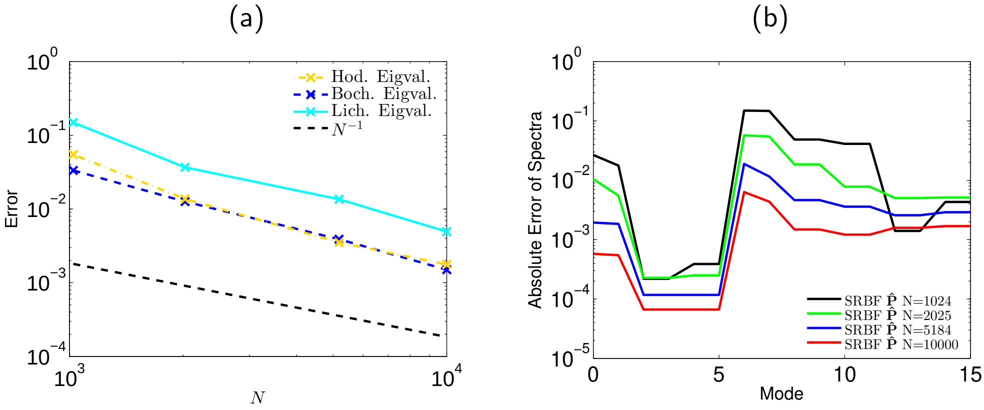


Figure: 2-Sphere in $\mathbb{R}^3$. (a) Mean absolute error of the leading 16 modes for Bochner and Hodge Laplacians and 20 modes for the Lichnerowicz Laplacian, plotted against N. (b) Absolute error between the eigenvalues of the Bochner Laplacian and its approximation, over the leading 15 modes.

Numerical Results

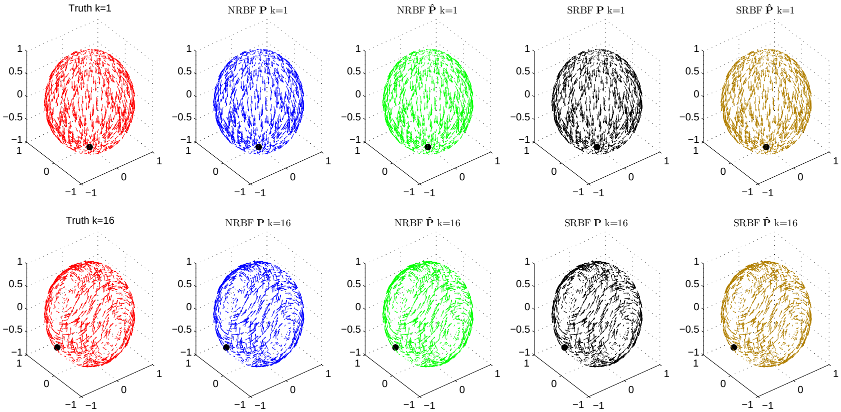


Figure: **2D Sphere in $\mathbb{R}^3$.** Comparison of eigen-vector fields of Bochner Laplacian for $k = 1, 16$. For NRBF, GA kernel with $s = 1.0$ is used, and for SRBF, IQ kernel with $s = 0.5$ is used. The $N = 1024$ data points are randomly distributed on the manifold.

Questions?

References

Edward Fuselier and Grady B Wright (2012)
Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates.
SIAM Journal on Numerical Analysis 50(3):1753–1776.

John Harlim, et al (2022)
Radial Basis Approximation of Tensor Fields on Manifolds: from Operator Estimation to Manifold Learning
A Preprint

Hemant Tyagi, et al (2013)
Tangent space estimation for smooth embeddings of riemannian manifolds.
Information and Inference: A Journal of the IMA 2(1):69–114.

Zhenyue Zhang and Hongyuan Zha (2004)
Principal manifolds and nonlinear dimensionality reduction via tangent space alignment.
SIAM journal on scientific computing 26(1):313–338.

The End