How well behaved is finite dimensional Diffusion Maps embedding?

Wenyu Bo
Cheriton School of Computer Science
University of Waterloo
Waterloo, ON
w3bo@uwaterloo.ca

Marina Meilă
Cheriton School of Computer Science
University of Waterloo
Waterloo, ON
mmp@uwaterloo.ca

Abstract

Diffusion Maps (DM) is a well studied and popular non-linear dimension reduction algorithm. Surprisingly, while the convergence properties of DM are well understood, the *geometric* properties of the embedding, such as reach and smoothness have not yet been studied. Under a set of standard assumptions on a family of submanifolds $\subset \mathbb{R}^D$, we derive a series of geometric properties that are preserved by DM, including almost uniform density, finite polynomial approximation and reach. Leveraging these properties, we establish rigorous bounds on the embedding errors introduced by the DM algorithm, of the order $\left(\frac{\log n}{n}\right)^{\frac{1}{8d+16}}$. These results offer a solid theoretical foundation for understanding the performance and reliability of DM in practical applications.

1 Introduction

Low dimensional embeddings of high-dimensional data are often used for visualization [1], i.e. as the *last step* of a data analysis. We ask whether non-linear dimension reduction, performed in a principled way, can also be useful as a *first step* of the data analysis, just like the classical Principal Component Analysis. Therefore, this paper examines to what extent the Diffusion Maps (DM) embedding[2; 3], a dimensionality reduction technique that captures the geometric structure of data by constructing a diffusion process among data points, obtains "well conditioned" non-linear embeddings, i.e. embeddings whose geometric distortions can be bounded.

Of particular interest is the preservation of the tangent space estimation under the DM embedding. For a smooth function on a manifold M, its gradient (as well as higher order derivatives) live in the tangent space to M. Controlling the error of the tangent space estimation in finite sample settings is a necessary condition to ensure that the gradients of functions of interest are transported unbiasedly to the lower dimensional data representation.

The DM is well studied as a differential geometric object; [4] showed that in the limit of infinite dimension it is an isometric embedding, while [5] showed that almost isometry can be achieved with a finite number m of eigenfunctions, where this m depends on manifold geometric properties. [6] showed that if isometry is not required, then the sufficient embedding dimension by Laplacian eigenfunctions depends on dimension, injectivity radius, Ricci curvature and volume, thus that it can be arbitrarily larger than the Whitney embedding dimension of $2d$.

The DM embedding is widely used for non-linear dimension reduction as the Diffusion Maps algorithm [2; 7; 8; 9], which embeds a sample into m dimensions by the eigenvectors of L_n, a $n \times n$ matrix estimator of the Laplace-Beltrami operator Δ. For instance, DM was used to analyze high-dimensional single-cell RNA sequencing data, revealing cell differentiation trajectories and underlying biological patterns[10], and it can order cells along their differentiation paths, enabling

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

accurate reconstruction of branching developmental processes[11]. In chemistry, DM can extract the dynamical modes of high-dimensional simulation trajectories, furnishing a kinetically low-dimensional framework[12], it can identify the collective coordinates of rare events in molecular transitions[13]. In Astronomy, DM is applied to estimate galaxy redshifts from photometric data, demonstrating comparable accuracy to existing methods[14].

Stimulated in part by the practical applications of DM, the consistency of finite sample estimators of Δ, and of their eigenfunctions has been intensely studied. Many early works discussed the pointwise consistency for empirical operator with smooth functions on M and they derived convergence (with different rates) [15; 16; 17; 18; 19], which provide illuminating results for subsequent studies. Not only Δ itself, the consistency of its embedding are also widely studied, they discussed how the eigenvectors and eigenvalues of the empirical Laplacian with specific graph (K-NN, ϵ-graph) converge to the eigenfunctions and eigenvalues of Δ in the different norm (L^2, L^∞)[20; 21; 22; 23]. There are some other special convergence types studied. For instance, convergence in Lipschitz norm[24], embedding using heat kernel[25; 26]

This paper completes the picture by considering geometric properties of a manifold, such as smoothness, injectivity radius, reach, volume, diameter, and examining to what extent the DM preserves these properties. In other words, if the original manifold $\mathcal{M}$ is "well behaved", what can be said about its DM embedding $\varphi(\mathcal{M})$? Surprisingly, these questions have received less attention in the literature.

Specifically, under a set of mostly standard geometric assumptions on a family of compact manifolds (section 2), in section A.2 we study the smoothness of φ by examining its Sobolev norms (section A.1) and use it to show the existence of local parameterizations, then we derived consistent properties that hold uniformly for all manifolds in the family after DM embedding, like sampling densities and reach (section A). In section 4, we identify the correspondence between the embedding via eigenfunctions and the embedding via eigenvectors of the graph matrix and address the convergence between them, then estimate the error between the eigenfunctions and eigenvectors to quantify the noise in the positions of the embedded points. section 5 presents the experimental results, and section 6 concludes the paper.

2 Background, challenges and assumptions

2.1 Manifolds and geometric preliminaries

For basic definitions readers should consult [27; 28; 29]. Here we consider sub-manifolds $M \subset \mathbb{R}^D$ which we will generically call *manifolds*. Note that the ambient dimension D will not appear throughout the paper, only Euclidean distances $|x - y|$ with $x, y \in \mathbb{R}^D$ will appear, hence D can actually be infinite, and $\mathbb{R}^D$ could become a Hilbert space.

We also assume that M is a closed manifold with smoothness of class $\mathcal{C}^k$, as described in Assumption 3. This assumption ensures the existence of a local polynomial expansion up to order k in the neighborhood of each point on M. Such expansions satisfy specific regularity constraints, enabling precise characterization of the local geometry and supporting rigorous analytical derivations. These properties are fundamental for establishing the theoretical results and ensuring consistency in subsequent computations.

In this work, we focus on several key geometric quantities, including the tangent space, reach, Riemannian metric, and geodesic curves. These quantities play critical roles in understanding the geometry of the manifold and in estimating and reconstructing manifolds. While these concepts are broadly defined in the general framework in Riemannian geometry, considering that our sampling points lie in submanifolds of $\mathbb{R}^D$, we restrict our attention to their specialized formulations for submanifolds embedded in Euclidean space. This setting simplifies their definitions and aligns with the analytical and computational techniques employed in this study. For example, the tangent space is characterized as the hyperplane tangent to M at a given point, and the Riemannian metric is induced from the ambient Euclidean metric.

For definition of **Tangent Space**, **Riemannian Metric**, **Normal Coordinate** and **Laplace-Beltrami Operator**, please refer to [29]. We will talk about other geometric quantities in the following.

Riemannian metric defines a measure on M, called the *Riemannian measure* via the *volume form*. If M is compact, it has a finite volume $\mathrm{Vol}(M)$. We can now have other measures on M, which are absolutely continuous w.r.t. the Riemannian measure, through their densities $f : M \to (0, \infty)$. In statistics/manifold learning, it is assumed that we are given an i.i.d. sample $X_n = \{x_1, \dots x_n\} \subset M$ from f.

Definition 1 (Reach[30])**.** The reach of a subset A of $\mathbb{R}^n$ is the largest τ (possibly ∞) such that if $x \in \mathbb{R}^n$ and the distance from x to A is smaller than τ, then A contains a unique point nearest to x. For more detail and properties, refer to [31].

The reach of a subset $A \subset \mathbb{R}^n$ provides a measure of its local geometric regularity. Specifically, if the reach of A is $\tau > 0$, then every point within a distance less than τ from A has a unique nearest point on A. Geometrically, this implies that A does not exhibit sharp corners, cusps, or regions of high curvature within the specified radius. In particular: For convex subsets of $\mathbb{R}^n$, the reach is infinite, reflecting the absence of curvature bounds or sharp features. For smooth submanifolds, the reach is inversely related to its maximum principal curvature. Intuitively, the reach corresponds to the radius of the smallest osculating ball that fits locally around.

2.2 The Diffusion map

Let M be a smooth, closed manifold of class $\mathcal{C}^3$ embedded in the (possibly high-dimensional) Euclidean space $\mathbb{R}^D$ and $\Delta_{\mathcal{M}}$ be Laplacian operator on M with eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \cdots$. We consider eigenfunctions corresponding to these eigenvalues: $e_0, e_1, \cdots$, and we normalize them such that $\|e_i\|_{L^2} = 1$. It is easy to check $e_0 \equiv \mathrm{const}$. Then we have the following embedding theorem.

Theorem 1. *[5, Thm. 5.1] Let $\mathcal{M}$ be the set of d dimensional, closed Riemannian manifolds whose Ricci curvature is bounded from below by κ, injectivity radius is bounded from below by ι, and the volume is bounded from above by V, we define $\varphi : M \subset \mathbb{R}^D \to \varphi(M) \subset \mathbb{R}^m$ as the following:*

$$\varphi(x) = (2t)^{\frac{d+2}{4}} \sqrt{2}(4\pi)^{\frac{d}{4}} (e^{-\lambda_1 t} e_1(x), e^{-\lambda_2 t} e_2(x), \cdots, e^{-\lambda_m t} e_m(x)) \in \mathbb{R}^m, \tag{1}$$

then there exists a $t_0 = t_0(d, \kappa, \iota, \epsilon)$ such that for all $0 < t < t_0$, there exists a $N_0(d, \kappa, \iota, \epsilon, V, t)$ such that if $m > N_0$, then for all $M \in \mathcal{M}(d, \kappa, \iota, V)$, the map above is an embedding of M into $\mathbb{R}^m$, and $1 - \epsilon < \|d\varphi_p v\| < 1 + \epsilon$ where $\|v\| = 1$.

Lemma 2. *[5, Lem. 3.2] We use the same notation as in the preceding theorem. Then for $\epsilon' > 0$, there exists $N_1 = N_1(d, \kappa, \iota, V, \epsilon', t_0)$ such that when $m \geq N_1$ and $t_0 \leq t' \leq 4$, we have*

$$|K_m(t', p, \cdot) - K(t', p, \cdot)| \leq \epsilon', \tag{2}$$

where $K(t, p, q)$ is the heat kernel: $K(t, p, q) = \sum_{i=0}^{\infty} e^{-\lambda_i t} e_i(p) e_i(q)$, and $K_m(t, p, q)$ is the truncated heat kernel: $K_m(t, p, q) = \sum_{i=0}^{m} e^{-\lambda_i t} e_i(p) e_i(q)$.

We select $t = t_0/2$, $t' = t_0$ (if $t_0 > 4$, we set $t = 2$ and $t' = 4$) and $m = \max\{N_0, N_1\} + 1$ to make the results above both hold.

In addition, $\varphi : M \to \varphi(M)$ is a homeomorphism, so $\dim M = \dim \varphi(M)$, and φ is an embedding, also an immersion, thus φ is a local diffeomorphism, and φ is bijective, so φ is a diffeomorphism. Thus $\varphi(M)$ is a smooth manifold [28].

Given a finite sample $\mathcal{X}_n = \{x_1, \dots, x_n\}$ the DM algorithm constructs a similarity matrix to measure the pairwise relationships between data points, where the commonly used kernel is the Gaussian kernel, defined as $k(x_i, x_j) = \exp(-\|x_i - x_j\|^2/h)$, where h is a scale parameter, and this similarity matrix is then normalized, we denote it as Laplacian graph which is the approximation of Laplace-Beltrami operator. By performing eigenvalue decomposition on normalized Laplacian graph, the algorithm extracts the dominant eigenvalues and eigenvectors. These eigenvectors, scaled by their corresponding eigenvalues, define the diffusion coordinates, providing a low-dimensional embedding of the data. This embedding preserves the global geometry of the dataset while emphasizing its intrinsic structure.

In this paper, we will focus mainly on the geometric properties of the DM $\varphi(M)$, w.r.t. the original manifold M. The results we obtain will be useful in characterizing the output of the DM algorithm in finite sample settings.

[4] have shown that, in the limit of large m and small t, $\varphi(M)$ is isometric with M, and that this is possible approximatively with a finite m for manifolds with bounded diameter and Ricci curvature

bounded from below. In spite of these seemingly encouraging results, the DM can exhibit severe curvature distortions even for apparently "nice" manifolds. The intuitive explanation is the fact that, even if M is compact and smooth to order k, the local interactions between the manifold curvature and the manifold reach (note that these are not independent quantities) can exert a strong influence on the Laplacian eigenfunctions.

2.3 Assumptions

Thus, in predicting smoothness (w.r.t. Sobolev norms) and geometric properties of $\varphi(M)$, one needs to consider $\tau_M, \iota_M, \kappa_M$ in addition to the smoothness of M (made more precise below). We will give the assumptions for the set of manifolds we will perform DM, and then discuss basic properties of M based on these assumptions. We will see all these assumptions intuitively ensure our manifolds have very good shape, which can help us avoid extremely bizarre situations. And we will use them to derive existence of the geometric bound after DM, which are critical in estimating the convergence rate.

Assumption 1 (Reach)**.** The reach τ_M of M is bounded below by $\tau_{\min}$. This assumption is used in Theorem A.6.

Assumption 2 (Volume)**.** The volume of M is bounded below by V_1 and bounded above by V_2. This assumption is used in Theorem 1, Lemma A.1.

Assumption 3 (Smoothness)**.** For k and $\mathbf{L} := (L_\perp, L_3, \ldots, L_k)$, we assume there exists a local one-to-one parameterization Ψ_p for all $p \in M$:

$$\Psi_p : B_{T_pM}(0, r) \to M \quad \text{by} \quad \Psi_p(v) = p + v + N_p(v) \tag{3}$$

for some $r \geq \frac{1}{4L_\perp}$ with $N_p(v) \in \mathcal{C}^k(B_{T_pM}(0, r), \mathbb{R}^D)$ such that

$$N_p(0) = p,\ d_0 N_p = 0,\ \|d_v^2 N_p\| \leq L_\perp,\ \|d_v^i N_p\| \leq L_i \text{ for } i = 3, \ldots, k \tag{4}$$

holds for all $\|v\| \leq \frac{1}{4L_\perp}$. This assumption is used in Lemma A.4

In addition, we make the following more technical regularity assumption that will be used in Lemma A.3, section A.4.2.

Assumption 4 (Christoffel Symbols)**.** In the normal coordinate chart, the derivatives of the Christoffel symbols of order at most $k-2$ (including Christoffel symbol itself) have a uniform upper bound depending only on its order, i.e.

$$\left|\frac{\partial^l \Gamma_{ij}^k}{\partial x_{i_1} \cdots \partial x_{i_l}}\right| \leq C(l) \quad \text{for } 1 \leq i, j, k \leq d \text{ and } l \leq k-2 \tag{5}$$

Assumption 5 (Heat Kernel Bound)**.** The heat kernel coefficients $\Phi_j(p, q)$ and $C(v)$ are uniformly bounded with respect to $d(p, q) \in (s_0, s_1)$. eq. (137).

Let $\mathcal{M}(d, \tau_{\min}, V, k, \mathbf{L}, \Gamma)$ be the set of compact connected submanifolds $M \subset \mathbb{R}^D$, with dimension d satisfying Assumption 1–Assumption 5. For simplicity, in the rest of our paper we always assume that the manifold is d-dimensional and the ambient dimension is D, we also use the abbreviation $\mathcal{M}$ for $\mathcal{M}(d, \tau_{\min}, V, k, \mathbf{L}, \Gamma)$, and sometimes we include some of the parameters above to indicate that the assumptions corresponding to these parameters are satisfied.

We also assume that the sampling density on M does not deviate too much from uniform.

Assumption 6 (Density)**.** Let $\mathcal{P}_{f_{\min}, f_{\max}}$ denote the set of distributions P with support on $M \in \mathcal{M}$, and the density function f of P with respect to Hausdorff measure such that $0 < f_{\min} \leq f \leq f_{\max} < \infty$. This assumption will be used in section A.3.

All the assumptions above, perhaps with the exception of Assumption 3, Assumption 4, and Assumption 5 are generically present in the manifold learning literature.

2.4 Direct Consequences of Assumptions

Here we list some direct results for $M \in \mathcal{M}$ from assumptions which will also be used in our following proof.

Corollary 3 (Complete Manifold)**.** *[27] Any compact Riemannian manifold is geodesically complete according to the Hopf-Rinow theorem.*

Corollary 4 (Curvature)**.** *The sectional curvature of M is bounded by κ, since $|K(u,v)| \leq 2/\tau_M^2 \leq 2/\tau_{\min}^2$ due to the Gauss equation. In addition, it immediately implies Ricci curvature of M is bounded below by $-\kappa(d-1)$ since $\mathrm{Ric}(v,v) = \sum_{i=2}^{d} K(v,x_i)$ where $\{v, x_2, \ldots, x_d\}$ are orthonormal basis. It is used in Theorem 1, Lemma A.1.*

Corollary 5 (The injectivity radius)**.** *[31, Prop. A.1] The injectivity radius ι_M of M is bounded below by $\pi\tau_M$, which implies ι_M is bounded by $\pi\tau_{\min}$ for all $M \in \mathcal{M}$. We use it to make sure DM is an embedding.*

The injectivity radius bounded from below implies we can find a normal coordinate chart with the uniform radius.

Corollary 6 (Diameter)**.** *[32, Lem. 2] For $M \in \mathcal{M}$, we have*

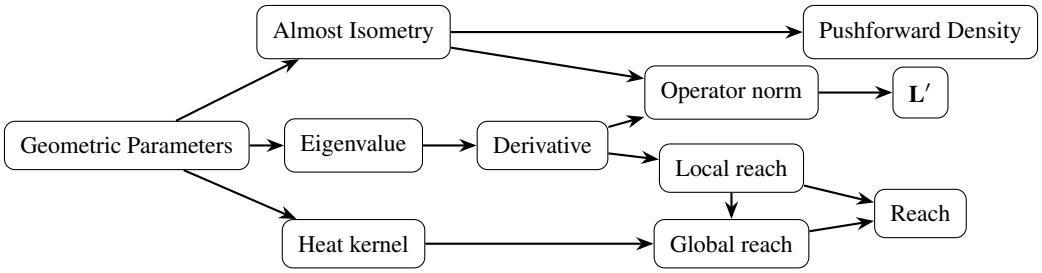
$$\mathrm{diam}(M) \leq \frac{C_d}{\tau_M^{d-1} f_{\min}} \leq \frac{C_d}{\tau_{\min}^{d-1} f_{\min}} \tag{6}$$

where C_d is a constant only depending on d. Thus the diameter of $M \in \mathcal{M}$ has a uniform upper bound. We will use it in Lemma A.1.

3 Properties of the diffusion maps embedding $\varphi(\mathcal{M})$

In this section, we will discuss the properties of $\varphi(M)$, here we collect the uniform estimates for the family $\varphi(\mathcal{M})$ and make explicit the logical chain that connects them. We will prove that the family $\varphi(\mathcal{M}) \in \mathcal{M}(\tau_{\varphi(M)}, \mathbf{L}')$ where $\mathbf{L}' := (L'_\perp, \ldots, L'_k)$ and the density function of $\varphi(M) \in \varphi(\mathcal{M})$ admits uniform upper and lower bounds. The proofs are presented in full in section A; here we only state the results and indicate the logical chain.

The dependency chain is



To begin with, we establish the uniform upper bound on the eigenvalues. Using the Li–Yau estimates [33, Thm. 16], the Laplacian eigenvalues grow at a controlled rate uniformly over the family $\mathcal{M}$.

Lemma A.1. There exists a constant C depending only on the uniform geometric parameters $(d, \kappa, \mathrm{diam}, V)$ of the family $\mathcal{M}$, such that

$$\lambda_k \leq C(k+1)^{\frac{2}{d}} \qquad \text{for all } M \in \mathcal{M}. \tag{7}$$

We next derive uniform bounds on the eigenfunction derivatives. Using the eigenvalue estimate in Step 1, together with estimates in [34], we can control the derivatives of the eigenfunctions, and the coordinate derivatives of each component of φ.

Lemma A.2. For every multi–index α with $|\alpha| \leq k$, there exists a constant C_α depending only on α and the uniform geometric parameters of $\mathcal{M}$, such that

$$|\partial^\alpha e_\lambda(x)| \leq C_\alpha (1+\lambda)^{k/2+d/4}, \qquad |\alpha| \leq k. \tag{8}$$

Consequently, since λ_k is bounded above by $C(\mathcal{M})(k+1)^{\frac{2}{d}}$, $\varphi_j = (2t)^{\frac{d+2}{4}}\sqrt{2}(4\pi)^{\frac{d}{4}} e^{-\lambda_1 t} e_j(x)$, then the coordinate derivatives of φ up to order k are uniformly bounded over $M \in \mathcal{M}$.

With the bound on the eigenfunction derivatives, we turn to the uniform bound on the operator norm of $d^l\varphi$. Computing coordinate representation for the covariant derivatives and pluging in the bound from Step 2 with the bound of Christoffel symbols, we obtain the bound on the operator norm of $d^l\varphi$.

Lemma A.3. For every $l \le k$, there exists a constant $A(l)$, depending only on l and the uniform geometric parameters of $\mathcal{M}$, such that

$$\|d^l \varphi\| \le \sqrt{m}\, A(l)\, d^{l/2} \quad \text{for all } M \in \mathcal{M}. \tag{9}$$

We next control the parametrization remainder uniformly. Expanding the remainder of the local parametrization of $\varphi(M)$ through Faà di Bruno's formula, each term is a composition of derivatives of φ (bounded in Step 3), derivatives of the original remainder Ψ_p (bounded by Assumption 3), and derivative of the inverse of φ (bounded due to embedding is almost isometric), hence obtaining a parameter vector $\mathbf{L}' := (L'_\perp, \dots, L'_k)$ controlling the remainder over $\varphi(\mathcal{M})$.

Lemma A.4. There exists $\mathbf{L}' := (L'_\perp, \dots, L'_k)$ depending only on the uniform geometric parameters of $\mathcal{M}$, such that for every $M \in \mathcal{M}$ and every $p' = \varphi(p)$, the remainder satisfies

$$N'_{p'}(0) = 0, d_0 N'_{p'} = 0, \text{ and } \|d^l_{v'} N'_{p'}\| \le L'_l \text{ for } 2 \le l \le k. \tag{10}$$

We then establish the uniform bound on the pushforward density. Since φ is almost isometric ($|\,\|d\varphi\| - 1| < \epsilon$), the Jacobian factor stays close to 1, so the pushforward sampling density g of $\varphi_\# P$ stays comparable to the original density f of the distribution P.

Lemma A.5. Suppose $\epsilon \le \frac{1-4^{-1/d}}{3}$, then for every $M \in \mathcal{M}$ the Jacobian factor of φ lies in $(\frac{1}{2}, 2)$, and consequently the pushforward density g of $\varphi_\# P$ obeys $\frac{f_{\min}}{2} < g < 2 f_{\max}$.

Finally, we derive a uniform lower bound on the reach. Writing $\tau = \min\{\tau_l, \tau_g\}$, we bound the local and global reach separately. The local reach is controlled through the second fundamental form of $\varphi(M)$, whose operator norm is estimated directly from the second-order eigenfunction derivative.

Lemma A.7 (Local reach). There exists a constant $\tau_{l,\varphi(\mathcal{M})} > 0$, depending only on the uniform geometric parameters of $\mathcal{M}$ and on (m, t_0), such that

$$\tau_{l,\varphi(M)} \ge \tau_{l,\varphi(\mathcal{M})} > 0, \quad \text{for all } \varphi(M) \in \varphi(\mathcal{M}). \tag{11}$$

For the global reach, we split the proof according to three ranges of the geodesic distance $d(p, q)$, with $s_0 = 2\sqrt{2}\,\tau_{l,\varphi(M)}$ and a threshold s_1:

(i) $d(p, q) \le s_0/(1+\epsilon)$. We get $d(\varphi(p), \varphi(q)) \le s_0$ since φ is almost isometric, and by local linear approximation (Lemmas A.8 and A.9) shows $\varphi(p), \varphi(q)$ cannot form a bottleneck.

(ii) $d(p, q) > s_1$. Heat-kernel estimates give the global reach condition $|\varphi(p) - \varphi(q)| > r_1 > 0$ (Lemma A.10).

(iii) $s_0/(1+\epsilon) < d(p, q) < s_1$. It is empty whenever $(*)$ holds, where (i) and (ii) already cover all cases. If $(*)$ fails, we verify $d(p, q)$ in this interval. We use a tighter lower bound on s_0, and a short time expansion of the heat kernel that controls $K(t, p, p) - K(t, p, q)$.

Since the three ranges cover all $d(p, q) > 0$, the global reach of $\varphi(M)$ is bounded below by a positive constant uniformly over $\varphi(\mathcal{M})$. Together, the local and global bounds give the reach theorem.

Theorem A.6. There exists a positive constant $\tau_{\varphi(\mathcal{M})}$, depending only on the uniform geometric parameters of $\mathcal{M}$ and on (m, t_0), such that $\tau_{\varphi(M)} \ge \tau_{\varphi(\mathcal{M})} > 0$ for all $M \in \mathcal{M}$.

4 Finite sample error of Diffusion Map φ

In this section, we consider the finite-sample DM. We use a result from [26] to derive an l_∞ simultaneous convergence rate for the DM embedding of the sample points $\mathcal{X}_n = \{x_1, \dots, x_n\} \subset M \in \mathcal{M}$. This uniform bound shows that the finite-sample DM embedding points can be viewed as discrete samples from the differential-geometric DM embedding manifold studied in Section 3, with a vanishing approximation error. Consequently, the finite-sample DM embeddings asymptotically approximate the geometric structure of the DM embedding manifold studied in Section 3.

We consider n sample points $\mathcal{X}_n = \{x_1, \dots, x_n\} \subset M \in \mathcal{M}$, we define graph affinity matrix W and the degree matrix D as $W_{ij} = \frac{k_h(x_i, x_j)}{q_h(x_i) q_h(x_j)}$, $D_{ii} = \sum_{j=1}^n W_{ij}$, where $k_h(x, y) = \exp\left(-\frac{|x-y|^2}{4h^2}\right)$

is the Gaussian kernel and $q_h(x) = \sum_{i=1}^n k_h(x, x_i)$. Then the normalized graph Laplacian L_n is defined as

$$L_n = \frac{D^{-1}W - I}{h^2}. \tag{12}$$

We denote the i-th eigenvalue of $-L_n$ as $\mu_{i,n,h}$ with corresponding eigenvector $\tilde{v}_{i,n,h}$ normalized in l^2 norm. It is easy to verify that $\mu_{0,n,h} = 0$ and $\tilde{v}_{0,n,h} = \mathbb{1}$. Let $\mathbb{N}(i) = |B_h(x_i) \cap \{x_1, \dots, x_n\}|$ which is the cardinality of points in the h-ball of x_i. Then we define the l^2 norm of $\tilde{v}$ with respect to inverse estimate density $1/\hat{p}$ as:

$$\|\tilde{v}\|_{l^2(1/\hat{p})} = \sqrt{\frac{|S^{d-1}|h^d}{d} \sum_{i=1}^n \frac{\tilde{v}^2(i)}{\mathbb{N}(i)}}, \tag{13}$$

and we define $v_{i,n,h} = \frac{\tilde{v}_{i,n,h}}{\|\tilde{v}_{i,n,h}\|_{l^2(1/\hat{p})}}$.

Let Δ be the Laplace-Beltrami operator of M, and $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$ be the eigenvalues of $-\Delta$. Let e_i be the eigenfunctions of $-\Delta$ corresponding to λ_i. Then we have

Theorem 7. *[26, Thm. 2] Let M be a d-dimensional smooth, closed and connected Riemannian manifold embedded in $\mathbb{R}^D$, f be the smooth probability density function on M with infimum $f_{\min} > 0$ and $\mathcal{X}_n = \{x_1, \dots, x_n\}$ be the point cloud sampled following f independently and identically. Suppose eigenvalues of Δ are simple. For fixed $m \in \mathbb{N}$, denote $\Gamma_m = \min_{1 \leq i \leq m} dist(\lambda_i, \sigma(-\Delta) \setminus \{\lambda_i\})$, where $\sigma(-\Delta)$ is the spectrum of $-\Delta$. Suppose*

$$h \leq \mathcal{K}_1 \min\left(\left(\frac{\min(\Gamma_m, 1)}{\mathcal{K}_2 + \lambda_m^{d/2+5}} \right)^2, \frac{1}{(\mathcal{K}_3 + \lambda_m^{(5d+7)/4})^2} \right), \tag{14}$$

where $\mathcal{K}_1$ and $\mathcal{K}_2, \mathcal{K}_3 > 1$ are constants depending on d, $f_{\min}$, $\|f\|_{C^2}$, and the volume, the injectivity radius, the curvature and the second fundamental form of the manifold. Then, when n is sufficiently large so that $h = h(n) \geq (\frac{\log n}{n})^{\frac{1}{4d+13}}$, with probability greater than $1 - n^{-2}$, for all $0 \leq i < m$,

$$|\mu_{i,n,h} - \lambda_i| \leq \Omega_1 h^{3/2}. \tag{15}$$

And when n is sufficiently large so that $h = h(n) \geq (\frac{\log n}{n})^{\frac{1}{4d+8}}$, with probability greater than $1 - n^{-2}$, there are $a_i \in \{1, -1\}$ such that for all $0 \leq i < m$,

$$\max_{x_j \in \mathcal{X}_n} |a_i v_{i,n,\epsilon}(j) - e_i(x_j)| \leq \Omega_2 h^{1/2}. \tag{16}$$

Ω_1 depends on d, the diameter of M, $f_{\min}$, $\|f\|_{C^2}$, and Ω_2 depends on d, the diameter and the volume of M, $f_{\min}$, and $\|f\|_{C^2}$.

Remark 1. In our setting, the sampling is uniform, thus $f \equiv 1/\text{Vol}(M)$, which implies $\|f\|_{C^2} = 1/\text{Vol}(M)$. In addition, d, diameter, volume, injectivity radius, curvature and second order fundamental form are bounded uniformly, which are compact sets, thus $\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3, \Omega_1, \Omega_2$ are bounded uniformly for all $M \in \mathcal{M}$. In the case when the eigenvalues are not simple, the same proof still works by introducing the eigenprojection [35].

Remark 2. If we choose $h = (\frac{\log n}{n})^{\frac{1}{4d+13}}$, then for large enough n, we have

$$|\mu_{i,n,h} - \lambda_i| \leq \Omega_1 (\frac{\log n}{n})^{\frac{3}{8d+26}}, \quad \text{for } 0 \leq i \leq m. \tag{17}$$

Similarly, if we choose $h = (\frac{\log n}{n})^{\frac{1}{4d+8}}$, we have

$$\max_{x_j \in \mathcal{X}_n} |a_i v_{i,n,\epsilon}(j) - e_i(x_j)| \leq \Omega_2 (\frac{\log n}{n})^{\frac{1}{8d+16}}, \quad \text{for } 0 \leq i \leq m. \tag{18}$$

Since $(\frac{\log n}{n})^{\frac{1}{4d+13}} \geq (\frac{\log n}{n})^{\frac{1}{4d+8}}$, the inequalities above hold at the same time if we choose $h = (\frac{\log n}{n})^{\frac{1}{4d+13}}$.

Now, we introduce our algorithm and estimate the approximation error.

Algorithm 1 Diffusion Map

1: **Input:** point cloud $\mathcal{X}_n$, intrinsic dimension d, the lower bound of injectivity radius ι (Corollary 5), the lower bound of Ricci curvature $-\kappa$ (Corollary 4), upper bound of volume V (Assumption 2), norm error ϵ (eq. (84)).

2: Calculate bandwidth $h = (\frac{\log n}{n})^{\frac{1}{4d+13}}$, diffusion time $t = \min\left\{t_0(d,\kappa,\iota,\epsilon), 4, \frac{\iota^2}{4}\right\}$ according to Theorem 1, Lemma 2, Lemma B.4 and embedding dimension $m = \max N_0, N_1 + 1$ according to Theorem 1, Lemma 2.

3: Construct W and D with bandwidth h according to eq. (12).

4: Calculate first m eigenvalues and eigenfunctions $\{\mu_{i,n,h}, \tilde{v}_{i,n,h}\}_{i=1}^m$ of $L_n = \frac{D^{-1}W - I}{h^2}$.

5: For $1 \le i \le m$, calculate

$$\mathbb{N}(i) = |B_h(x_i) \cap \{x_1, \dots, x_n\}|. \tag{19}$$

Normalize

$$v_{i,n,h} = \frac{\tilde{v}_{i,n,h}}{\|\tilde{v}_{i,n,h}\|_{l^2(1/\hat{p})}}, \text{where } \|\tilde{v}\|_{l^2(1/\hat{p})} = \sqrt{\frac{|S^{d-1}|h^d}{d}\sum_{i=1}^n \frac{\tilde{v}^2(i)}{\mathbb{N}(i)}}. \tag{20}$$

6: Embed x_j to $(t)^{\frac{d+2}{4}}\sqrt{2}(4\pi)^{\frac{d}{4}}\left(e^{-\mu_{i,n,h}t/2}v_{i,n,h}(j)\right)_{i=1}^m$

Output: the embedding point clouds $\left\{(t)^{\frac{d+2}{4}}\sqrt{2}(4\pi)^{\frac{d}{4}}\left(e^{-\mu_{i,n,h}t/2}v_{i,n,h}(j)\right)_{i=1}^m\right\}_{i=1}^n \subset \mathbb{R}^m$.

We compare this embedding with embedding in section 2 and estimate the approximation error.

To estimate the error term, we consider the error of i-th component, for large enough n, i.e.

$$(\frac{\log n}{n})^{\frac{1}{4d+13}} = h \le \mathcal{K}_1 \min\left(\left(\frac{\min(\Gamma_m, 1)}{\mathcal{K}_2 + \lambda_m^{d/2+5}}\right)^2, \frac{1}{(\mathcal{K}_3 + \lambda_m^{(5d+7)/4})^2}\right) \tag{21}$$

$$\iff n > \frac{1}{\mathcal{K}}\max\left(\left(\frac{\mathcal{K}_2 + \lambda_K^{d/2+5}}{\min(\Gamma_K, 1)}\right)^{8d+26}, (\mathcal{K}_3 + \lambda_K^{(5d+7)/4})^{8d+26}\right), \tag{22}$$

up to a $\log$ factor. Then adding and subtracting $e^{-\mu_{i,n,h}t/2}e_i(x_j)$ and applying the triangle inequality, we obtain

$$|e^{-\mu_{i,n,h}t/2}v_{i,n,h}(j) - e^{-\lambda_i t/2}e_i(x_j)| \le |v_{i,n,h}(j) - e_i(x_j)| + \frac{t}{2}|\mu_{i,n,h} - \lambda_i|\|e_i\|_{l^\infty}. \tag{23}$$

We also have $\|e_i\|_{l^\infty} \le \|e_i\|_{L^\infty} \le C\lambda_i^{\frac{d-1}{4}}$, where C only depends on injectivity radius and the absolute value of the sectional curvature [36], which are compact, thus C has a uniform upper bound for all $M \in \mathcal{M}$. And we have proved that λ_i are bounded from above for $i \le m$ with fixed m. Consequently, $\|e_i\|_{l^\infty}$ has the uniform bound, we denote it as $C_{\mathcal{M}}$.

Therefore,

$$|e^{-\mu_{i,n,h}t/2}v_{i,n,h}(j) - e^{-\lambda_i t/2}e_i(x_j)| \le \Omega_2(\frac{\log n}{n})^{\frac{1}{8d+16}} + \frac{t}{2}C_{\mathcal{M}}\Omega_1(\frac{\log n}{n})^{\frac{3}{8d+26}} \tag{24}$$

$$= \Omega(\frac{\log n}{n})^{\frac{1}{8d+16}}, \tag{25}$$

hence

$$|\left(e^{-\mu_{i,n,h}t/2}\tilde{v}_{i,n,h}(j)\right)_{i=1}^m - \left(e^{-\lambda_i t/2}e_i(x_j)\right)_{i=1}^m| \le \sqrt{m}\Omega(\frac{\log n}{n})^{\frac{1}{8d+16}}. \tag{26}$$

All prior convergence results control each column of the embedding matrix, i.e. eigenvector $v_{i,n,h}$ as a whole. In contrast, the bound above concerns the *rows* of the embedding matrix: it controls all sample

points x_j simultaneously, which bounds the distance between the finite-sample and infinite-sample embedding uniformly over $j = 1, \ldots, n$. This row-wise guarantee lets us treat the finite-sample embedding as a discrete sample from the DM embedding with controlled pointwise error. It requires the l^∞ control of the eigenvectors, since l^2 convergence only bounds the average error.

Remark 3. Faster convergence rates for eigenvalues and eigenvectors have been obtained in l^2 norm [21; 37; 24; 20]. However, these rates are not directly usable here, our analysis requires pointwise control of the embedding matrix, which is strictly stronger and cannot be derived from l^2 convergence. For comparasion, the exponent for l^2 convergence rate can be $1/(d+4)$, whereas the exponent for l^∞ we use here is $1/(8d+16)$.

5 Experiments

In this section, numerical experiments are presented to show the performance of the proposed DM algorithm in preserving the geometric properties of the underlying manifold, particularly focusing on reach and Riemannian metric. Manifolds are generated by the package in [38]. For details of generating data, please refer to Appendix C. All experiments were conducted with 10000 sample points on a MacBook Air with an Apple M4 processor and standard macOS system configuration, the running time is around 30 minutes for each experiment.

5.1 Common Manifolds

The first experiment is testing the reach preservation and Riemannian metric (Rmetric) preservation on synthetic datasets sampled from known manifolds. The process is the following:

1. Adopting the method in [39; 40], choosing the best h in eq. (12) and neighborhood radius r using grid search based on empirical validation for estimating Rmetric. In this step, Rmetric (baseline) and reach are estimated. If the manifold has a boundary, then we only consider estimation at the inner points (Appendix C) to avoid the bias caused by boundary points.
2. Selecting best t for the DM algorithm according to [39; 40], and embed the data into $\mathbb{R}^3$, then estimate the reach and pushforward Riemannian metric for the embedded manifold.

Rmetirc–error$_{\text{baseline}}$ (R–error$_{\text{base}}$) is computed directly on the original manifold and measures the intrinsic estimation error of the metric estimation procedure. Since the ground-truth pushforward metric is equal to 1, the baseline error is defined as the deviation of the estimated metric from 1. R–error$_{\text{emb}}$ is computed on the embedded manifold and measures how well the DM embedding preserves the Riemannian metric. It is defined as the relative error with respect to the baseline estimate. The reach is estimated using definitions from [31]. The table below summarizes the results.

Table 1: Results of the experiments on synthetic datasets.

Manifolds	R–error$_{\text{base}}$	reach	R–error$_{\text{emb}}$	reach$_{\text{emb}}$	h	r	t
Sphere	0.0125	0.9999	0.0203	0.9441	0.5198	1.578	0.3360
Swiss Roll	0.1085	0.4468	0.0933	0.3459	0.9161	2.689	1.880
Torus	0.1564	0.3162	0.1562	0.1731	0.9190	2.812	1.788

The first manifold is S^2 since S^2 is well-studied and has analytical eigenvalues and eigenfunctions of the Laplace–Beltrami operator, we can compare the estimated results with true values easily. The second and third manifolds are Swiss Roll and Torus (Standardized), respectively, which are more complex manifolds with non-trivial geometry.

The results show that the DM algorithm preserves the reach and Riemannian metric well for the sphere, with only a slight increase in R–error and a small decrease in reach. For the Swiss Roll and Torus, original reach is accurately estimated, where the true values of reach are approximately $\pi/6.97 = 0.4507$ and $0.668^2/1.409 = 0.3167$ (constants are related to standardization scaling factors), respectively. However, the Rmetric cannot be captured as well as the sphere, since these manifolds have more complex geometry and the algorithm cannot handle them effectively, where larger sample sizes are needed to decrease the error. And an increase in R–error on the original manifold is associated with a corresponding increase in R–error in the embedded manifold, along with a larger deviation in reach.

5.2 Random Manifolds

The DM algorithm is also tested on 2 randomly generated complex manifolds (for more details see appendix C), and the results are summarized in the table below.

Table 2: Results of the experiments on randomly generated manifolds, where the baseline is the R–error computed on the original manifold, and other columns are the relative R–error computed on the embedded manifold with different embedding dimensions.

Manifold \ Dim	baseline	10	20	30	40	h	r
Manifold 1	0.2312	0.5528	0.4272	0.3530	0.3276	0.1718	0.3938
Manifold 2	0.3286	0.5945	0.3909	0.2797	0.2440	0.3006	0.6789

The results suggest that as the embedding dimension increases, the R–error decreases, indicating better preservation of the Riemannian metric. And the decreasing rate is slower when the embedding dimension is larger, which is expected since the eigenvalues increase exponentially, and the contribution of higher order eigenvectors to the embedding is weakened by the eigenvalues.

However, even when the embedding dimension is 40, the R–error remains relatively large. One possible reason is that the algorithm for estimating Riemannian metric cannot perform well on complex manifolds (as presented by the baseline error), so that even almost isometric embeddings can have large error when estimating from point clouds.

6 Conclusion

Limitations: The theory relies on several strong assumptions, including smoothness, reach and strong regularity of the manifold, which ensures the preservation of geometric properties under the DM embedding. Additionally, the convergence analysis assumes i.i.d. and noiseless sampling, under which the convergence of eigenfunction and eigenvector estimates is guaranteed. Violations of these assumptions may lead to larger estimation errors or even loss of convergence. Moreover, in practical applications, the choice of parameters can be challenging and manifold properties are intricate, thus empirical validation and application are limited.

Conclusion: Nevertheless, under bounded curvature, reach bounded from below, finite volume, smoothness and regular condition expressed by Assumptions 1 to 6, choosing appropriate parameters: bandwidth h, embedding time t and embedding dimension m, the finite dimensional Diffusion Map embedding admits a controlled deviation from isometry and the smoothness, reach and sampling density are preserved. Moreover, the finite sample approximation can approach the correct embedding manifold with error rate $O\left((\frac{\log n}{n})^{\frac{1}{8d+16}}\right)$.

Acknowledgments and Disclosure of Funding

This research was supported by the Canada Research Chairs Program, by the Vector Institute for AI, through the CIFAR Pan-Canadian AI Strategy, the Government of Ontario, and leading industry sponsors from across the Canadian economy, and by the Cheriton School of Computer Science, University of Waterloo. Parts of this research were conducted while the authors were with the Department of Statistics, University of Washington. The authors thank the anonymous reviewers for their helpful comments and suggestions. M.M. gratefully acknowledges the DataShape Group at INRIA Saclay and the Institute for Mathematical and Statistical Innovation (IMSI) for hospitality while a portion this research was carried out.

References

[1] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.

[2] R. R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5–30, 2006. Special Issue: Diffusion Maps and Wavelets.

[3] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[4] P. Bérard, G. Besson, and S. Gallot, "Embedding riemannian manifolds by their heat kernel," *Geometric & Functional Analysis GAFA*, vol. 4, pp. 373–398, 1994.

[5] J. W. Portegies, "Embeddings of riemannian manifolds with heat kernels and eigenfunctions," *Communications on Pure and Applied Mathematics*, vol. 69, no. 3, pp. 478–518, 2016.

[6] J. Bates, "The embedding dimension of laplacian eigenfunction maps," *Applied and Computational Harmonic Analysis*, vol. 37, no. 3, pp. 516–530, 2014.

[7] S. Lafon and A. B. Lee, "Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 9, pp. 1393–1403, 2006.

[8] R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler, "Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems," *Multiscale Modeling & Simulation*, vol. 7, no. 2, pp. 842–864, 2008.

[9] C. M. Arvizu and A. R. Messina, "Dimensionality reduction in transient simulations: A diffusion maps approach," *IEEE Transactions on Power Delivery*, vol. 31, no. 5, pp. 2379–2389, 2016.

[10] L. Haghverdi, F. Buettner, and F. J. Theis, "Diffusion maps for high-dimensional single-cell analysis of differentiation data," *Bioinformatics*, vol. 31, no. 18, pp. 2989–2998, 2015.

[11] L. Haghverdi, M. Büttner, F. A. Wolf, F. Buettner, and F. J. Theis, "Diffusion pseudotime robustly reconstructs lineage branching," *Nature methods*, vol. 13, no. 10, pp. 845–848, 2016.

[12] A. L. Ferguson, A. Z. Panagiotopoulos, I. G. Kevrekidis, and P. G. Debenedetti, "Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach," *Chemical Physics Letters*, vol. 509, no. 1-3, pp. 1–11, 2011.

[13] L. Boninsegna, G. Gobbo, F. Noé, and C. Clementi, "Investigating molecular kinetics by variationally optimized diffusion maps," *Journal of chemical theory and computation*, vol. 11, no. 12, pp. 5947–5960, 2015.

[14] P. Freeman, J. Newman, A. Lee, J. Richards, and C. Schafer, "Photometric redshift estimation using spectral connectivity analysis," *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 4, pp. 2012–2021, 2009.

[15] U. Von Luxburg, M. Belkin, and O. Bousquet, "Consistency of spectral clustering," *The Annals of Statistics*, pp. 555–586, 2008.

[16] M. Belkin and P. Niyogi, "Towards a theoretical foundation for laplacian-based manifold methods," *Journal of Computer and System Sciences*, vol. 74, no. 8, pp. 1289–1308, 2008.

[17] E. Giné and V. Koltchinskii, "Empirical graph laplacian approximation of laplace-beltrami operators: large sample results," *Lecture Notes-Monograph Series*, pp. 238–259, 2006.

[18] D. Ting, L. Huang, and M. I. Jordan, "An analysis of the convergence of graph laplacians," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, (Madison, WI, USA), p. 1079–1086, Omnipress, 2010.

[19] M. Hein, J.-Y. Audibert, and U. v. Luxburg, "Graph laplacians and their convergence on random neighborhood graphs.," *Journal of Machine Learning Research*, vol. 8, no. 6, 2007.

[20] X. Cheng and N. Wu, "Eigen-convergence of gaussian kernelized graph laplacian by manifold heat interpolation," *Applied and Computational Harmonic Analysis*, vol. 61, pp. 132–190, 2022.

[21] N. García Trillos, M. Gerlach, M. Hein, and D. Slepčev, "Error estimates for spectral convergence of the graph laplacian on random geometric graphs toward the laplace–beltrami operator," *Foundations of Computational Mathematics*, vol. 20, no. 4, pp. 827–887, 2020.

[22] X. Wang, "Spectral convergence rate of graph laplacian," *arXiv preprint arXiv:1510.08110*, 2015.

[23] M. Belkin and P. Niyogi, "Convergence of laplacian eigenmaps," *Advances in neural information processing systems*, vol. 19, 2006.

[24] J. Calder, N. Garcia Trillos, and M. Lewicka, "Lipschitz regularity of graph laplacians on random data clouds," *SIAM Journal on Mathematical Analysis*, vol. 54, no. 1, pp. 1169–1222, 2022.

[25] P. W. Jones, M. Maggioni, and R. Schul, "Manifold parametrizations by eigenfunctions of the laplacian and heat kernels," *Proceedings of the National Academy of Sciences*, vol. 105, no. 6, pp. 1803–1808, 2008.

[26] D. B. Dunson, H.-T. Wu, and N. Wu, "Spectral convergence of graph laplacian and heat kernel reconstruction in l^∞ from random samples," *Applied and Computational Harmonic Analysis*, vol. 55, pp. 282–336, 2021.

[27] M. P. Do Carmo and J. Flaherty Francis, *Riemannian geometry*, vol. 2. Springer, 1992.

[28] J. Lee, *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics, Springer, 2003.

[29] J. M. Lee, *Introduction to Riemannian manifolds*, vol. 2. Springer, 2018.

[30] H. Federer, "Curvature measures," *Transactions of the American Mathematical Society*, vol. 93, no. 3, pp. 418–491, 1959.

[31] E. Aamari, J. Kim, F. Chazal, B. Michel, A. Rinaldo, and L. Wasserman, "Estimating the reach of a manifold," *Electronic Journal of Statistics*, vol. 13, no. 1, pp. 1359 – 1399, 2019.

[32] E. Aamari and C. Levrard, "Stability and minimax optimality of tangential delaunay complexes for manifold reconstruction," *Discrete & Computational Geometry*, vol. 59, pp. 923–971, 2018.

[33] P. Li and S. Yau, "Estimates of eigenvalues of a compact riemannian manifold," *Proc. Sympos. PureMath.*, vol. 36, 01 1980.

[34] B. Xu, "Derivatives of the spectral function and sobolev norms of eigenfunctions on a closed riemannian manifold," *Annals of Global Analysis and Geometry - ANN GLOB ANAL GEOM*, vol. 26, pp. 231–252, 10 2004.

[35] F. Chatelin, *Spectral Approximation of Linear Operators*. Society for Industrial and Applied Mathematics, 2011.

[36] H. Donnelly, "Eigenfunctions of the Laplacian on Compact Riemannian Manifolds," *Asian Journal of Mathematics*, vol. 10, no. 1, pp. 115 – 126, 2006.

[37] J. Calder and N. Garcia Trillos, "Improved spectral convergence rates for graph laplacians on ε-graphs and k-nn graphs," *Applied and Computational Harmonic Analysis*, vol. 60, pp. 123–175, 2022.

[38] S. J. Koelle, H. Zhang, M. Meila, and Y.-C. Chen, "Manifold coordinates with physical meaning," *Journal of Machine Learning Research*, vol. 23, no. 133, pp. 1–57, 2022.

[39] D. Perraul-Joncas and M. Meila, "Non-linear dimensionality reduction: Riemannian metric estimation and the problem of geometric discovery," 2013.

[40] D. Joncas, M. Meila, and J. McQueen, "Improved graph laplacian via geometric self-consistency," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[41] Y. Shi and B. Xu, "Gradient estimate of an eigenfunction on a compact riemannian manifold without boundary," *Annals of Global Analysis and Geometry*, vol. 38, 05 2009.

[42] B. Xu, "Asymptotic behavior of l^2-normalized eigenfunctions of the laplace-beltrami operator on a closed riemannian manifold," *arXiv preprint math/0509061*, 2005.

[43] A. Hassannezhad, G. Kokarev, and I. Polterovich, "Eigenvalue inequalities on riemannian manifolds with a lower ricci curvature bound," *Journal of Spectral Theory*, vol. 6, no. 4, pp. 807–835, 2016.

[44] D. Li and D. B. Dunson, "Geodesic distance estimation with spherelets," *arXiv preprint arXiv:1907.00296*, 2019.

[45] F.-Y. Wang, "Sharp explicit lower bounds of heat kernels," *The Annals of Probability*, vol. 25, no. 4, pp. 1995–2006, 1997.

[46] M. Ludewig, "Strong short-time asymptotics and convolution approximation of the heat kernel," *Annals of Global Analysis and Geometry*, vol. 55, no. 2, pp. 371–394, 2019.

[47] H. Federer, *Geometric Measure Theory*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Springer, 1969.

[48] P. Li and S. Yau, "On the parabolic kernel of the schödinger operator," *Acta Mathematica*, vol. 156, pp. 153–201, 07 1986.

[49] C. B. Croke, "Some isoperimetric inequalities and eigenvalue estimates," in *Annales scientifiques de l'École normale supérieure*, vol. 13, pp. 419–435, 1980.

[50] S.-Y. Cheng and P. Li, "Heat kernel estimates and lower bound of eigenvalues.," *Commentarii mathematici Helvetici*, vol. 56, pp. 327–338, 1981.

A Properties of the diffusion maps embedding $\varphi(\mathcal{M})$

In this section, we will discuss the properties of $\varphi(M)$, and we will first give some uniform estimate of geometric quantities for all $M \in \mathcal{M}$, and we will prove that the family $\varphi(\mathcal{M}) \in \mathcal{M}(\tau_{\varphi(M)}, \mathbf{L}')$ where $\mathbf{L}' := (L'_\perp, \dots, L'_k)$ and the density function of $\varphi(M) \in \varphi(\mathcal{M})$ admits an uniform upper and lower bounds.

A.1 Bounds on λ_m and the (higher) derivatives of ϕ

Estimating and uniformly bounding the k-th derivative of ϕ, $\|d^k\varphi\|$, plays an important role in controlling other geometric quantities such as τ_{min} and proving existence of local one–to–one parametrization of $\varphi(\mathcal{M})$. Recall that k is the smoothness of $\mathcal{M}$. Because the i–th component of map $\varphi : M \to \mathbb{R}^m$ is e_i multiple of $e^{-\lambda_i t_0/2}$ which is a bounded scalar when t_0 is fixed, we can ignore it when we estimate the upper bound.

We notice that $\varphi(M)$ is a d–dimensional manifold embedded in $\mathbb{R}^m$, then $d\varphi_p$ is a map such that

$$d\varphi_p : T_pM \to T_{\varphi(p)}\varphi(M) \hookrightarrow \mathbb{R}^m \tag{27}$$

and we can treat the tangent vector in $\mathbb{R}^m$ as a tangent vector in $T_\varphi(p)\varphi(M)$ through the natural isomorphism. This isomorphism maps local basis to local basis, thus this map is also isometric. In the following, we will consider $T_{\varphi(p)}\varphi(M) \cong \mathbb{R}^d$ and $T_{\varphi(p)}\varphi(M) \subset \mathbb{R}^m$ as equivalent through this mapping.

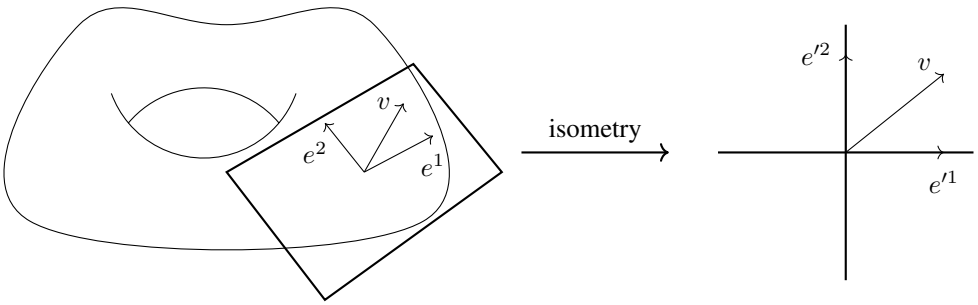


Figure 1: Isometry

A.1.1 Upper Bound of λ_m

The most classical result for estimating eigenvalues is Weyl's law:

$$N(\lambda) \sim \frac{\omega(d)Vol(M)\lambda^{\frac{d}{2}}}{(2\pi)^d}, \tag{28}$$

where $N(\lambda)$ is the number of eigenvalues less than or equal to λ.

If we ask $\lambda = \lambda_k$, the above is equivalent to

$$\lambda_k^{d/2} \sim \frac{(2\pi)^d k}{\omega(n)Vol(M)} \tag{29}$$

Weyl's law provides an asymptotic expression for the eigenvalues of the Laplace–Beltrami operator, offering profound insights into their growth rates. However, in our paper, strict control for eigenvalues is essential. This requires not only asymptotic estimates but also rigorous upper bounds that hold universally. A notable result for the upper bound was established by Li–Yau [33, Thm. 16], who derived explicit upper bounds for the eigenvalues under geometric constraints, which align with the assumptions in our study.

We estimate the upper bound of eigenvalues. If the lower bound of the Ricci curvature κ_M of M is less than 0, we have

$$\lambda_k \leq \begin{cases} \frac{(2\beta+1)^2}{4}(-\kappa_M) + 4(1+2^\beta)^2\pi^2 \left(\frac{\sinh\sqrt{-\kappa_M}\text{diam}}{\sqrt{-\kappa_M}\text{diam}}\right)^{\frac{2d-2}{d}} \\ \qquad\qquad \times \left((k+1)^{\frac{\alpha(d-1)}{d}}\frac{1}{V}\right)^{\frac{2}{d}} \\ \text{when } d = 2(\beta+1), \beta = 0,1,2\ldots \\ \frac{(2\beta+2)^2}{4}(-\kappa_M) + 4(1+\pi^2)(1+2^{2\beta})^2 \left(\frac{\sinh\sqrt{-\kappa_M}\text{diam}}{\sqrt{-\kappa_M}\text{diam}}\right)^{\frac{2d-2}{d}} \\ \qquad\qquad \times \left((k+1)^{\frac{\alpha(d-1)}{d}}\frac{1}{V}\right)^{\frac{2}{d}} \\ \text{when } d = 2\beta+3, \beta = 0,1,2\ldots \end{cases}, \tag{30}$$

if M has non–negative Ricci curvature κ_M, we have

$$\lambda_k \leq (d+4)d^{1-\frac{2}{d}}\left(\frac{k+1}{V}\alpha(d-1)\right)^{\frac{2}{d}}, \tag{31}$$

where d is the dimension of M, V is the volume of M and $\alpha(n)$ is the volume of S^n in $\mathbb{R}^{n+1}$.[33]

Since for $M \in \mathcal{M}$, the volume is greater that V_1 and the lower bound of the Ricci curvature is greater than κ, so $\frac{1}{V} \leq \frac{1}{V_1}$ and $-\kappa_M \leq -\kappa$, thus we have

$$\lambda_k \leq \begin{cases} C(d,V)(k+1)^{\frac{2}{d}} & \text{if } \kappa_M \text{ is non-negative} \\ C(d,\kappa,\text{diam},V)(k+1)^{\frac{2}{d}} & \text{if } \kappa_M \text{ is negative} \end{cases}. \tag{32}$$

Summarizing above, we have

Lemma A.1 (Uniform upper bound on λ_k)**.** *There exists a constant $C(\mathcal{M})$, depending only on the uniform geometric parameters $(d, \kappa, \mathrm{diam}, V)$ of the family $\mathcal{M}$, such that*

$$\lambda_k \leq C(\mathcal{M})(k+1)^{\frac{2}{d}} \qquad \textit{for all } M \in \mathcal{M}. \tag{33}$$

A.1.2 Estimate of Higher Order Derivatives

In this subsection, we will estimate the L^∞ norm of higher order derivatives for eigenfunctions to control the operator norm of dN' in the next subsection.

One similar result is Bernstein inequality, for eigenfunction e_λ of Δ: $\Delta e_\lambda = \lambda e_\lambda$, we have[41]:

$$\|\nabla e_\lambda\|_{L^\infty} \leq C\lambda^{\frac{1}{2}}\|e_\lambda\|_{L^\infty} \tag{34}$$

[42] proved a more general estiamte, for eigenfunction e_λ of Δ: $\Delta e_\lambda = \lambda e_\lambda$, we have

$$\|e_\lambda\|_{H^\infty_k} \sim \lambda^{\frac{k}{2}}\|e_\lambda\|_{L_\infty} \quad \text{as } \lambda \to \infty \tag{35}$$

where $\sim$ means the equivalence between two norms, and $\|\cdot\|_{H^k_\infty}$ is Sobolev norm. If we revise this result to have a upper bound (the author says we only need a little effort to do so), we can control $\|d^k\varphi\|$ since we have had the upper bound of λ_m which only depend on M.

[34] also gave a better estimate:

$$\|e_\lambda(x)\|_{H^\infty_k} \leq C\lambda^{k/2+(d-1)/4}\|e_\lambda(x)\|_{L^2} \tag{36}$$

however, C may not be an uniform constant for $M \in \mathcal{M}$.

To estimate the L^∞ norm of derivatives for eigenfunctions, we first choose the coordinate chart to be normal coordinate chart to make our derivatives well defined. We take derivative of $e(x) : M \to \mathbb{R}$ in the following sense: Let $x \in U$ be the normal coordinate chart $(U, \mathbf{x})$ and $\hat{U} = \mathbf{x}(U)$, then $\hat{e}(x) = e \circ \mathbf{x}^{-1} : \hat{U} \to R$ be the coordinate representation of $e(x)$. We treat $\partial_i e(x)$ as $\partial_i \hat{e}(x)$, and for higher order derivatives as well.

Lemma A.2 (Uniform bound on the higher derivatives of the eigenfunctions)**.** *For every multi–index α with $|\alpha| \leq k$ there exists a constant C_α, depending only on α and the uniform geometric parameters of $\mathcal{M}$, such that the eigenfunction derivatives satisfy the pointwise bound*

$$|\partial^\alpha e_\lambda(x)| \leq C_\alpha(1+\lambda)^{k/2+d/4}, \qquad |\alpha| \leq k. \tag{37}$$

Consequently, since λ_k is bounded above by $C(\mathcal{M})(k+1)^{\frac{2}{d}}$, $\varphi_j = (2t)^{\frac{d+2}{4}}\sqrt{2}(4\pi)^{\frac{d}{4}}e^{-\lambda_1 t}e_j(x)$, then the coordinate derivatives of φ up to order k are uniformly bounded over $M \in \mathcal{M}$.

Proof of Lemma A.2. Let γ be a $2d$ multi–index with $|\gamma| := \sum_i \gamma_i \leq k$, and $x, y \in U$ be a geodesic normal coordinate of M. We denote $\sum_{\lambda_j \leq \lambda} e_j(x)e_j(y)$ as $e(x, y, \lambda)$. Let $\gamma = (\gamma_1, \ldots, \gamma_{2d})$, then the γ order derivative for $e_j(x)e_j(y)$ is

$$\partial^\gamma_{x,y} e_j(x)e_j(y) = \frac{\partial^{|\gamma_1+\cdots+\gamma_d|}e_j(x)}{\partial x_1^{\gamma_1}\cdots\partial x_d^{\gamma_d}}\frac{\partial^{|\gamma_{d+1}+\cdots+\gamma_{2d}|}e_j(y)}{\partial y_1^{\gamma_{d+1}}\cdots\partial y_d^{\gamma_{2d}}} \tag{38}$$

We have [34, Lem. 2.4]

$$|\partial^\gamma_{x,y} e(x, y, \lambda)| \leq C_\gamma(1+\lambda)^{\frac{d+|\gamma|}{2}}. \tag{39}$$

To have the estimate for derivatives of eigenfunctions, let α be a d dimension multi–index, and using the inequality above with $\gamma = (\alpha, \alpha)$, then we have

$$|\partial^\alpha e_\lambda(x)|^2 \leq \sum_{\lambda_j \leq \lambda} |\partial^\alpha e_j(x)|^2 \leq C_\alpha(1+\lambda)^{d/2+|\alpha|} \tag{40}$$

so

$$\|e_\lambda\|_{H^\infty_k} = \max_{|\alpha|\leq k} |\partial^\alpha e_\lambda(x)| \leq \sup_{|\alpha|\leq k}\{C_\alpha\}(1+\lambda)^{k/2+d/4} \tag{41}$$

Since λ_k is bounded above by $C(\mathcal{M})(k+1)^{\frac{2}{d}}$, which implies that $\|\varphi_j\|_{H^\infty_k} \leq C(l)$ for $\varphi_j : M \to \mathbb{R}$. Therefore, $|\frac{\partial^l \varphi_j}{\partial x_{i_1}\cdots\partial x_{i_l}}| \leq C(l)$ for $l \leq k$ with respect to normal coordinate. □

After deriving the uniform bound for the norm of the derivatives of ϕ, we now establish the bound of the operator norm of the covariant derivatives.

Lemma A.3 (Uniform bound on the higher derivatives of the embedding map φ)**.** *For every $l \leq k$ there exists a constant $A(l)$, depending only on l and the uniform geometric parameters of $\mathcal{M}$, such that the operator norm of the l-th derivative of the embedding map is bounded uniformly over $\mathcal{M}$:*

$$\|d^l\varphi\| \leq \sqrt{m}\, A(l)\, d^{l/2} \quad \textit{for all } M \in \mathcal{M}. \tag{42}$$

Proof of Lemma A.3. By Lemma A.2, the coordinate derivatives of φ up to order k are uniformly bounded over $M \in \mathcal{M}$; we use this together with the uniform bounds on the Christoffel symbols and their derivatives to control $\|d^l\varphi\|$.

To estimate the norm of $d\varphi$, we can derive its local coordinate representation and compute the covariant derivatives in the normal chart since choosing different chart does not have the influence on the norm. Now we denote φ as j–th component of φ to simplify the notation and let $(U, \mathbf{x})$ be the normal chart.

For the first order covariant derivative, we have

$$\nabla\varphi(X) = \nabla_X\varphi = X\varphi, \tag{43}$$

thus

$$\nabla\varphi = \varphi_i dx^i \quad \text{with } \varphi_i = \partial_i\varphi, \tag{44}$$

so for the first order covariant derivative, the local coordinate representation is the first order derivative of local representation of φ.

For second order covariant derivative, which is also known as covariant Hessian. We have

$$\nabla^2\varphi(X,Y) = \nabla^2_{Y,X} = \nabla_Y(\nabla_X\varphi) - \nabla_{\nabla_Y X}\varphi = X(Y\varphi) - (\nabla_X Y)\varphi \tag{45}$$

thus in local coordinates,

$$\nabla^2\varphi = \varphi_{ij} dx^i \otimes dx^j \quad \text{with } \varphi_{ij} = \partial^2_{ij}\varphi - \Gamma^k_{ij}\partial_k\varphi, \tag{46}$$

with $i, j, k \in \{1, 2, \dots d\}$. For the third order coordinate derivative, we have

$$\begin{aligned}
\nabla^3\varphi(X,Y,Z) &= (\nabla_Z\nabla^2\varphi)(X,Y) && (47)\\
&= Z(\nabla^2\varphi(X,Y)) - \nabla^2\varphi(\nabla_Z X, Y) - \nabla^2\varphi(X, \nabla_Z Y) && (48)\\
&= Z(X(Y\varphi) - (\nabla_X Y)\varphi) - \nabla_Z X(Y\varphi) + (\nabla_{\nabla_Z X}Y)\varphi \\
&\qquad - X(\nabla_Z Y\varphi) + (\nabla_X\nabla_Z Y)\varphi, && (49)
\end{aligned}$$

thus in the local coordinate,

$$\nabla^3\varphi = \varphi_{ijk} dx^i \otimes dx^j \otimes dx^k \tag{50}$$

$$\begin{aligned}
\text{with} \quad \varphi_{ijk} = \partial^3_{ijk}\varphi - \partial_k\Gamma^l_{ij}\partial_l\varphi - \Gamma^l_{ij}\partial^2_{kl}\varphi - \Gamma^m_{ik}\partial^2_{mj}\varphi + \Gamma^l_{ik}\Gamma^m_{lj}\partial_m\varphi - \partial_i\Gamma^l_{jk}\partial_l\varphi \\
- \Gamma^l_{jk}\partial^2_{il}\varphi + \Gamma^m_{il}\Gamma^l_{jk}\partial_m\varphi. \qquad (51)
\end{aligned}$$

The local coordinate representation is combined with derivative of Christoffel symbol and third order derivative of φ. And we can prove easily by induction that the local coordinate representation of $\nabla^l\varphi$ is $\varphi_{i_1\cdots i_l} dx^{i_1} \otimes \cdots \otimes dx^{i_l}$ with $\varphi_{i_1\cdots i_l}$ can be represented by derivative of φ not higher than order l and derivative of Christoffel symbol not higher than order $l-2$. To prove this, we only need to notice that

$$\begin{aligned}
\nabla^{l+1}\varphi(X_1, \cdots, X_l, X) &= \nabla(\nabla^l\varphi)(X_1, \cdots, X_l, X) && (52)\\
&= (\nabla_X\nabla^l\varphi)(X_1, \cdots, X_l) && (53)\\
&= X(\nabla^l\varphi(X_1, \cdots, X_l)) - \sum_{i=1}^{l} \nabla^l\varphi(X_1, \cdots, \nabla_X X_i, \cdots, X_l) && (54)
\end{aligned}$$

Since we assume for any $M \in \mathcal{M}$, the local coordinate representation in the normal chart of the Christoffel symbols and their derivatives with order not above $k-2$ have an uniform bound only

depends on its order, and we have claimed derivatives of φ which order is not higher than k have an uniform bound, and we notice that the local representation of $\nabla^k\varphi$ is combined with them in the same pattern among all $M \in \mathcal{M}$. Therefore, $\varphi_{i_1\cdots i_l}$ is bounded uniformly for all $l \leq k$, i.e.

$$|\varphi_{i_1\cdots i_l}| \leq A(l) \quad \text{for } l \leq k \tag{55}$$

With this control of upper bound, we can estimate the norm of higher derivatives of φ:

$$\begin{aligned}
\|d^l\varphi\| &= \sup_{\|v_1\|=1,\ldots,\|v_l\|=1} \|\nabla^l\varphi(v_1,\ldots,v_l)\| && (56)\\
&= \sup_{\|v_1\|=1,\ldots,\|v_l\|=1} \|\nabla^l\varphi(v_1\otimes\cdots\otimes v_l)\| && (57)\\
&= |\varphi_{i_1\cdots i_l} dx^{i_1}\otimes\cdots\otimes dx^{i_l}(v_1^{i_1}\frac{\partial}{\partial x^{i_1}}\otimes\cdots\otimes v_l^{i_l}\frac{\partial}{\partial x^{i_l}})| && (58)\\
&= |\varphi_{i_1\cdots i_l} v_1^{i_1}\cdots v_l^{i_l}| && (59)\\
&\leq A(l)\left|\sum_{i_1,i_2,\ldots i_l=1}^{d} v_1^{i_1}\cdots v_l^{i_l}\right| && (60)\\
&= A(l)\left|\sum_{i_1=1}^{d} v_1^{i_1}\sum_{i_2,\ldots i_l=1}^{d} v_2^{i_2}\cdots v_l^{i_l}\right| && (61)\\
\text{(AM-–GM)} &\leq A(l)\left|\sum_{i_2,\ldots i_l=1}^{d} v_2^{i_2}\cdots v_l^{i_l}\right|\sqrt{d} && (62)\\
&\vdots && (63)\\
&\leq A(l)d^{l/2}. && (64)
\end{aligned}$$

This inequality holds for all $l \leq k$. Here $\varphi : M \to \mathbb{R}$, thus $d^l\varphi(v_1,\ldots,v_l)$ gives a vector in $\mathbb{R}$, and then $d^l\varphi(v_1,\ldots,v_l) = \big(d^l\varphi_1(v_1,\ldots,v_l),\ldots,d^l\varphi_m(v_1,\ldots,v_l)\big)$ is a vector in $\mathbb{R}^m$, which is also a tangent vector in $T_{\varphi(p)}\varphi(M)$ through the natural isometry. Since $d^l\varphi(v_1,\ldots,v_l)$ has m components, it is easy to check the operator norm of $d^l\varphi$ has the upper bound $\sqrt{m}A(l)d^{l/2}$. Therefore, $\|d^l\varphi\|$ are bounded uniformly for $l \leq k$. □

A.2 Uniform upper bounds $\{L_\perp, L_3, \ldots, L_k\}$ on the higher derivatives of the local parametrization reminder

In this section, we will verify that there exists an common parameter set $\mathbf{L}' := \{L'_\perp, L'_3, \ldots, L'_k\}$ such that the operator norm of the derivatives of remainder $N_{\varphi(p)}$ defined in Assumption 3 can be controlled uniformly over $\varphi(\mathcal{M})$, which is builded on the uniform bound on $\|d^l\varphi\|$ from Lemma A.3.

Lemma A.4 (Uniform control of the parametrization remainder)**.** *There exists a common parameter vector $\boldsymbol{L}' := (L'_\perp, \ldots, L'_k)$, depending only on the uniform geometric parameters of $\mathcal{M}$, such that for every $M \in \mathcal{M}$ and every $p' = \varphi(p) \in \varphi(M)$ the local parametrization remainder $N'_{p'}$ satisfies $N'_{p'}(0) = 0$, $d_0N'_{p'} = 0$, and*

$$\|d^l_{v'}N'_{p'}\| \leq L'_l \qquad \textit{for } 2 \leq l \leq k. \tag{65}$$

Proof of Lemma A.4. We fix a point $p' = \varphi(p) \in \varphi(M)$, for $v' \in T_{p'}\varphi(M) = d\varphi_p(T_pM)$, we define

$$\Psi'_{p'}(v') = \varphi(\Psi_p(d(\varphi^{-1})_{p'}(v'))) = p' + v' + N'_{p'}(v') \tag{66}$$

from definition of Ψ_p, we know $\Psi'_{p'}(v')$ and $N'_{p'}(v')$ are well defined when $\|d(\varphi^{-1})_{p'}(v')\| \leq \frac{1}{4L_\perp}$.

We estimate $\|d\varphi_{p'}^{-1}\|$: We can treat T_pM as a d dimensional subspace of $\mathbb{R}^D$ and treat $T_{p'}\varphi(M)$ is a d dimensional subspace of $\mathbb{R}^m$, and these inclusions are canonical and isometric, so they are both

closed subspaces. And $d\varphi_p$ is homeomorphism, so graph of $d\varphi_{p'}^{-1}$ is closed, then $d\varphi_{p'}^{-1}$ is bounded, i.e. $\|d\varphi_{p'}^{-1}\|$ exists.

$$\|d\varphi_{p'}^{-1}\| = \sup\left\{\|d\varphi_{p'}^{-1}v'\| : v' \in Im(d\varphi_p), \|v'\| = 1\right\} \tag{67}$$

$$= \sup\left\{\frac{\|v\|}{\|d\varphi_p v\|} : v' \in Im(d\varphi_p), \|v'\| = 1, v = d\varphi_{p'}^{-1}v'\right\} \tag{68}$$

$$= \sup\left\{\frac{1}{\|d\varphi_p v\|} : \|v\| = 1\right\} \tag{69}$$

$$\leq \frac{1}{1-\epsilon} \tag{70}$$

so when $\|v'\| \leq \frac{1}{4L_\perp(1-\epsilon)}$, $\|d(\varphi^{-1})_{p'}(v')\| \leq \frac{1}{4L_\perp}$, then $\Psi'_{p'}(v')$ and $N'_{p'}(v')$ are well defined.

Now, We check $N'_{p'}(0) = 0$ and $d_0 N'_{p'} = 0$:

$$N'_{p'}(0) + p' = \varphi(\Psi_p(d(\varphi^{-1})_{p'}(0))) = \varphi(\Psi_p(0)) = \varphi(p + 0 + N_p(0)) = \varphi(p) = p' \tag{71}$$

$$d_0 N'_{p'} = d\varphi_p d_0\Psi_p d(\varphi)_{p'}^{-1} - I_m = d\varphi_p(I_D + d_0 N_p)d(\varphi)_{p'}^{-1} - I_m = 0 \tag{72}$$

We assume $\|v'\| \leq \frac{1}{4L_\perp(1-\epsilon)}$ and denote $d\varphi_{p'}^{-1}v'$ as v, and $c(v') = p + v + N_p(v)$, then for any unit vector $w' \in T_{p'}\varphi(\mathcal{M})$, using Faà di Bruno's formula, we have

$$\left\|d^k_{v'}N'_{p'}(w'^{\otimes k})\right\| = \left\|\sum_{\pi\in P(k)} d^l_{c(v')}\varphi \circ \left(d^{j_1}_v\Psi_p\left(\left\{d_{p'}\varphi^{-1}w'\right\}^{\otimes j_1}\right), \ldots, d^{j_l}_v\Psi_p\left(\left\{d_{p'}\varphi^{-1}w'\right\}^{\otimes j_l}\right)\right)\right\|, \tag{73}$$

where $P(k)$ is a partition of k with l parts such that $j_1 + \cdots + j_l = k$.

Since $\Psi_p(v) = p + v + N_p(v)$, so $d_v\Psi_p = I + d_v N_p$, $d^i_v\Psi_p = d^i_v N_p$, thus

$$\|d_v\Psi_p\| = \|I + d_v N_p\| \leq 1 + L_\perp\|v\| \tag{74}$$

$$\|d^i_v\Psi_p\| = \|d^i_v N_p\| \leq L_i \tag{75}$$

therefore

$$\|d^k_{v'}N'_{p'}(w'^{\otimes k})\| \leq \sum_{\pi\in P(k)} \|d^l\varphi\| \prod_{i=1}^{l} \|d^{j_i}\Psi_p\| \|d_{p'}\varphi^{-1}w'\|^{j_i} \tag{76}$$

$$\leq \sum_{\pi\in P(k)} \|d^l\varphi\| \frac{1}{(1-\epsilon)^l} \prod_{i=1}^{l} \|d^{j_i}\Psi_p\| \tag{77}$$

We have proved $\|d^j\varphi\|$ and $\|d^j_v\Psi_p\|$ have the uniform upper bound for $j = 2, \ldots, k$ in Lemma A.3, so we can select $\mathbf{L}' := (L'_\perp, \ldots, L'_k)$ as the uniform upper bound of $\|d^2_{v'}N'_{p'}\|, \ldots, \|d^k_{v'}N'_{p'}\|$. □

A.3 Bounds on the pushforward density

In this subsection, we estimate the pushforward density of the sampling process. By establishing both lower and upper bounds for the pushforward density, we ensure that the sampling over $\varphi(M)$ remains approximately uniform. This guarantees that there are always some sampling points in any nonzero measure region with nonzero probability, thereby allowing the manifold to be effectively approximated using the sampled points.

Lemma A.5 (Almost-uniform pushforward density)**.** *Suppose $\epsilon \leq \frac{1-4^{-1/d}}{3}$. Then for every $M \in \mathcal{M}$ the Jacobian factor of φ satisfies*

$$\sqrt{\det\left(\pi_{T_pM} \circ d\varphi_p^T \circ d\varphi_p|_{T_pM}\right)} \in \left(\tfrac{1}{2}, 2\right), \tag{78}$$

and consequently the pushforward density g of $\varphi_\# P$ obeys $\frac{f_{\min}}{2} < g < 2f_{\max}$.

Proof of Lemma A.5. We denote $d\varphi_p : T_pM \to T_{p'}N$ as A, then for $\|v\| = 1$, $1-\epsilon < \|Av\| < 1+\epsilon$, which means $\|A\| < 1+\epsilon$, so $\sqrt{\lambda_{\max}(A^TA)} < 1+\epsilon$. We know A^TA is positive-definite symmetric matrix, we consider an unit eigenvector v of it with eigenvalue λ, then

$$\lambda = \|v^T\lambda v\| = \|v^TA^TAv\| = \|Av\|^2 \in ((1-\epsilon)^2, (1+\epsilon)^2). \tag{79}$$

Since A^TA is symmetric, then $A^TA = Q\Lambda Q^T$, where Q is orthogonal matrix and Λ is diagonal matrix whose diagonal elements are eigenvalues of A^TA and all of them are close to 1.

$$\begin{aligned} |\|\pi_{T_pM}A^TAv\| - \|v\|| &\le \|\pi_{T_pM}(A^TA - I)v\| && (80)\\ &= \|\pi_{T_pM}Q(\Lambda - I)Q^Tv\| && (81)\\ &\le \|Q(\Lambda - I)Q^Tv\| && (82)\\ &\le 3\epsilon, && (83) \end{aligned}$$

the last inequality is due to orthogonal matrix does not change norm of vector and diagonal elements in $\Lambda - I$ are smaller than 3ϵ.

Thus we have $1 - 3\epsilon < \|\pi_{T_pM}A^TAv\| < 1 + 3\epsilon$, so the eigenvalues of $\pi_{T_pM}A^TA|_T$ is greater than $1 - 3\epsilon$ and less than $1 + 3\epsilon$. Therefore, if

$$\epsilon \le \min\left\{\frac{4^{1/d}-1}{3}, \frac{1-\frac{1}{4^{1/d}}}{3}\right\} = \frac{1-\frac{1}{4^{1/d}}}{3}, \tag{84}$$

then

$$\det(\pi_{T_pM}A^TA|_T) = \prod \lambda_i \in ((1-3\epsilon)^d, (1+3\epsilon)^d) \subset (\frac{1}{4}, 4), \tag{85}$$

which implies

$$\sqrt{\det\left(\pi_{T_pM} \circ d\varphi_p^T \circ d\varphi_p|_{T_pM}\right)} \in (\frac{1}{2}, 2). \tag{86}$$

According to Lemma B.2, we have the pushforward density of $\varphi_\# P$ is

$$g(p') = f(p)/\sqrt{\det\left(\pi_{T_pM} \circ d\varphi_p^T \circ d\varphi_p|_{T_pM}\right)}, \tag{87}$$

where $p' = \varphi(p)$. Therefore, $\frac{f_{min}}{2} < g < 2f_{\max}$. □

In our setting, our sampleing is uniform, thus $f \equiv \frac{1}{Vol(M)}$, which is bounded from both and above, which implies the density of $\varphi(\mathcal{M})$ is also bounded from below by $\min_{M\in\mathcal{M}} \{Vol(M)\}/2$ and bounded above by $2\max_{M\in\mathcal{M}} \{Vol(M)\}$.

A.4 Estimation of $\tau_{\min}$

If M has reach τ, then at least one of the following cases holds:

I (Global case) M has a bottleneck, i.e. there exist $p, q \in M$, such that $(p+q)/2 \in \text{Med}(M)$ and $\|p - q\| = 2\tau$.

II (Local case) There exists $p \in M$, and an arc-length parametrized geodesic γ such that $\gamma(0) = p$ and $|\gamma''(0)| = 1/\tau$.

where $\text{Med}(M) = \left\{z \in \mathbb{R}^D : \exists p \ne q \in M, \|z-p\| = \|z-q\| = d_E(z, M)\right\}$, $d_E(x, M)$ is the distance between x and M.

We denote global reach and local reach as τ_g and τ_l, respectively. Thus $\tau = \min\{\tau_g, \tau_l\}$

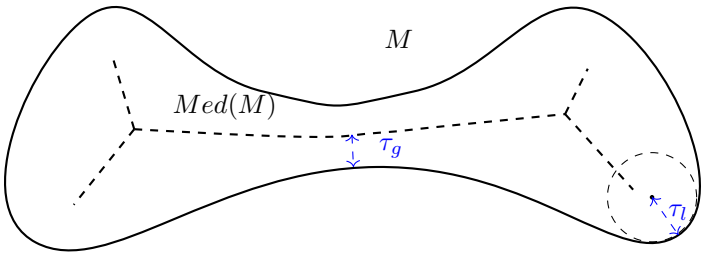


Figure 2: Reach

We can define τ_l easily by

$$\tau_l := \inf_{p \in M} \left\{ \frac{1}{\|\mathrm{II}_p\|} \right\}. \tag{88}$$

We will show that there exists a uniform positive lower bound on the reach of $\varphi(M)$ for all $\varphi(M) \in \varphi(\mathcal{M})$:

Theorem A.6 (Uniform lower bound on the reach of $\varphi(\mathcal{M})$)**.** *There exists a positive constant $\tau_{\varphi(\mathcal{M})}$, depending only on the uniform geometric parameters of $\mathcal{M}$ and on the DM parameters (m, t_0), such that*

$$\tau_{\varphi(M)} \geq \tau_{\varphi(\mathcal{M})} > 0 \qquad \textit{for all } M \in \mathcal{M}. \tag{89}$$

The bound is obtained by writing $\tau = \min\{\tau_l, \tau_g\}$ and bounding each term separately: the local reach is controlled through the second fundamental form of $\varphi(M)$ (Lemma A.7), and the global reach is controlled in section A.4.2.

A.4.1 Local Reach

Local reach is a quantity measuring the "curvature" locally, controlling of local reach will avoid some extremely weird manifold. Now we estimate the lower bound of local reach for $\varphi(\mathcal{M})$.

To control the local reach, we will derive the local coordinate representation of $\varphi(M)$ and then use the boundedness of derivative of eigenfunctions to obtain the upper bound of the second fundamental form of $\varphi(M)$ in $\mathbb{R}^m$.

Lemma A.7 (Uniform lower bound on the local reach)**.** *The local reach of $\varphi(M)$ admits the uniform lower bound: there exists a constant $\tau_{l,\varphi(\mathcal{M})} > 0$, depending only on the uniform geometric parameters of $\mathcal{M}$ and on (m, t_0), such that*

$$\tau_{l,\varphi(M)} \geq \tau_{l,\varphi(\mathcal{M})} > 0, \quad \textit{for all } \varphi(M) \in \varphi(\mathcal{M}). \tag{90}$$

Proof of Lemma A.7. For any $p \in M$, we can choose neighborhoods $U \subset M$ and $V \subset \varphi(M)$ s.t. $\varphi(U) = V$, and $(U, \mathbf{x})$ is a normal coordinate chart of p, where $\mathbf{x}^{-1} = \exp_p \circ E$, E is the isomorphism from $\mathbb{R}^d$ to T_pM, and we denote $\mathbf{x}(U)$ as $\hat{U} \subset \mathbb{R}^d$ is open.

We have a parametrization of $U \subset M$:

$$\mathbf{x}^{-1} : \hat{U} \subset \mathbb{R}^d \to U \subset M \tag{91}$$

$$(x_1, \ldots, x_d) \to \mathbf{x}^{-1}(x_1, \ldots, x_d) = (\mathbf{x}_1^{-1}(x_1, \ldots, x_d), \ldots, \mathbf{x}_D^{-1}(x_1, \ldots, x_d)) \tag{92}$$

Using this map, we have a coordinate representation of φ, i.e. $\hat{\varphi} := \varphi \circ \mathbf{x}^{-1}$ of $V \subset \varphi(M)$ by $\hat{U} \overset{\mathbf{x}^{-1}}{\to} U \overset{\varphi}{\to} V \subset \mathbb{R}^m$:

$$(x_1, \ldots, x_d) \overset{\mathbf{x}^{-1}}{\to} \mathbf{x}^{-1}(x_1, \ldots, x_d) \overset{\varphi}{\to} (\varphi_1 \circ \mathbf{x}^{-1}, \ldots, \varphi_m \circ \mathbf{x}^{-1})(x_1, \ldots, x_d) \tag{93}$$

Let $r = \hat{\varphi}$ be the parameterization of $\varphi(M)$, we compute its k-th tangent vector

$$r_i = \frac{\partial \hat{\varphi}}{\partial x_i} = \left(\frac{\partial \hat{\varphi}_1}{\partial x_i}, \ldots, \frac{\partial \hat{\varphi}_m}{\partial x_i} \right) \quad i = 1, \ldots, d, \tag{94}$$

and r's second derivative is

$$r_{ij} = \left(\frac{\partial^2 \hat{\varphi}_1}{\partial x_i \partial x_j}, \ldots, \frac{\partial^2 \hat{\varphi}_m}{\partial x_i \partial x_j} \right) \quad i, j = 1, \ldots, d. \tag{95}$$

For $\varphi(p) \in \varphi(M)$, we have the orthogonal decomposition:

$$\mathbb{R}^m = T_{\varphi(p)}\varphi(M) \oplus N_{\varphi(p)}\varphi(M), \tag{96}$$

where $T_{\varphi(p)}\varphi(M)$ is spanned by $\{r_1, \ldots, r_d\}$ and $N_{\varphi(p)}\varphi(M)$ is the $m - d$ subspace orthogonal to $T_{\varphi(p)}\varphi(M)$. Let π be the orthogonal projection from $\mathbb{R}^m$ onto $N_{\varphi(p)}\varphi(M)$, we can derive the second fundamental form on $\varphi(M)$, which is a symmetric 2-tensor field given by

$$\mathrm{II}_{ij} dx^i \otimes dx^j \quad \text{with} \quad \mathrm{II}_{ij} = \pi(r_{ij}) \in N_{\varphi(p)}\varphi(M). \tag{97}$$

Then we can verify that

$$\|\mathrm{II}\|_{op} \leq d \max\{\pi(r_{ij})\} \leq d \max\{|r_{ij}|\}. \tag{98}$$

We compute the module of $r_{ij} = \left(\frac{\partial^2 \hat{\varphi}_1}{\partial x_i \partial x_j}, \ldots, \frac{\partial^2 \hat{\varphi}_m}{\partial x_i \partial x_j} \right)$ in more detail. We have $\hat{\varphi}_i = (t_0)^{\frac{d+2}{4}} \sqrt{2} (4\pi)^{\frac{d}{4}} e^{-\lambda_i t_0/2} \hat{e}_i$, thus

$$\left| \frac{\partial^2 \hat{\varphi}_k}{\partial x_i \partial x_j} \right| = (t_0)^{\frac{d+2}{4}} \sqrt{2} (4\pi)^{\frac{d}{4}} e^{-\lambda_k t_0/2} \left| \frac{\partial^2 \hat{e}_k}{\partial x_i \partial x_j} \right| \tag{99}$$

$$\leq (t_0)^{\frac{d+2}{4}} \sqrt{2} (4\pi)^{\frac{d}{4}} e^{-\lambda_k t_0/2} C_2 (1 + \lambda_k)^{1+\frac{d}{4}}, \tag{100}$$

where the inequality is from eq. (41) in section A.1.2.

We also have the lower bound of λ_k,

$$\lambda_k \geq C_1^{1+\mathrm{diam}\sqrt{\kappa}} \mathrm{diam}^{-2} k^{2/d}, \tag{101}$$

where C_1 is constant only depending on d [43, Thm. 1.2.1].

Therefore,

$$|r_{ij}| \leq (t_0)^{\frac{d+2}{4}} \sqrt{2} (4\pi)^{\frac{d}{4}} C_2 \sqrt{\sum_{k=1}^{m} e^{-\lambda_k t_0} (1 + \lambda_k)^{2+\frac{d}{2}}}. \tag{102}$$

It is easy to see that $e^{-xt_0}(1+x)^{4+d}$ is increasing when $x < \frac{4+d}{2t_0} - 1$ and decreasing when $x > \frac{4+d}{2t_0} - 1$, thus we can use

$$C_1^{1+\mathrm{diam}\sqrt{\kappa}} \mathrm{diam}^{-2} k^{2/d} \leq \lambda_k \leq C + C' k^{2/d}, \tag{103}$$

which implies $c_1 k^{2/d} \leq \lambda_k \leq c_2 k^{2/d}$, where c_1, c_2 are constants depending on κ, diam, V, d. Then we have $|r_{ij}|$ is bounded from above, thus $\|\mathrm{II}\|_{op}$ is bounded from above, which implies that $\tau_{l,\varphi(M)}$ is bounded from below. □

A.4.2 Global Reach

In this subsection, we estimate the lower bound of the global reach of $\varphi(\mathcal{M})$. The global reach constrains the overall shape of the manifold by controlling the Euclidean distance to "separate" different parts of the manifold.

The proof is organized by considering three different cases depending on different ranges of the geodesic distance $d(p,q)$. When $d(p,q)$ is small, a local linear approximation excludes the global case directly (Lemma A.9); when $d(p,q)$ is large, heat-kernel estimates give a positive lower bound on $|\varphi(p)-\varphi(q)|$ (Lemma A.10); the remaining is closed either by the condition $(*)$ or by a refined heat-kernel analysis if $(*)$ fails.

Under Assumptions 1, when p is close to q, the geodesic distance $d(p,q)$ can be controlled by the Euclidean distance $|p-q|$.

Lemma A.8. *[44, Prop. 1] If $d(p,q)=s$, then*

$$s-\frac{s^3}{24r_0^2}\le |p-q|\le s, \tag{104}$$

where $\frac{1}{r_0}=\sup\{|\gamma''(s)|\}$ and γ varies among all geodesics on M in arc length parameter.

We consider the left inequality of Lemma A.8, it is easy to see if $s\le 2\sqrt{2}r_0$, then

$$\frac{2}{3}s\le s-\frac{s^3}{24r_0^2}\le |p-q|, \tag{105}$$

which implies

$$|p-q|\le d(p,q)\le \frac{3}{2}|p-q|. \tag{106}$$

We can verify easily that $1/r_0=\sup\{|\gamma''(s)|\}=1/\tau_l$, and this local linear approximation can exclude the global reach case when $d(p,q)<2\sqrt{2}\tau_l$, and $\frac{3}{2}$ plays an important role here, it will lead to the contradiction to the global reach case.

Lemma A.9. *If $d(p,q)\le s_0$ where $s_0=2\sqrt{2}\tau_l$, then p,q cannot satisfy global reach case.*

Proof of Lemma A.9. If not, we assume there exists $p,q\in M$ such that $d(p,q)\le s_0$ and p,q satisfy global reach case, i.e.

$$\frac{p+q}{2}\in Med(M), \tag{107}$$

$$\pi(\frac{p+q}{2})=p,q \quad \text{and} \quad d_E(p,M)=d_E(q,M)=|p-q|/2 \tag{108}$$

Since $d(p,q)\le s_0$, we have $d(p,q)\le\frac{3}{2}|p,q|$.

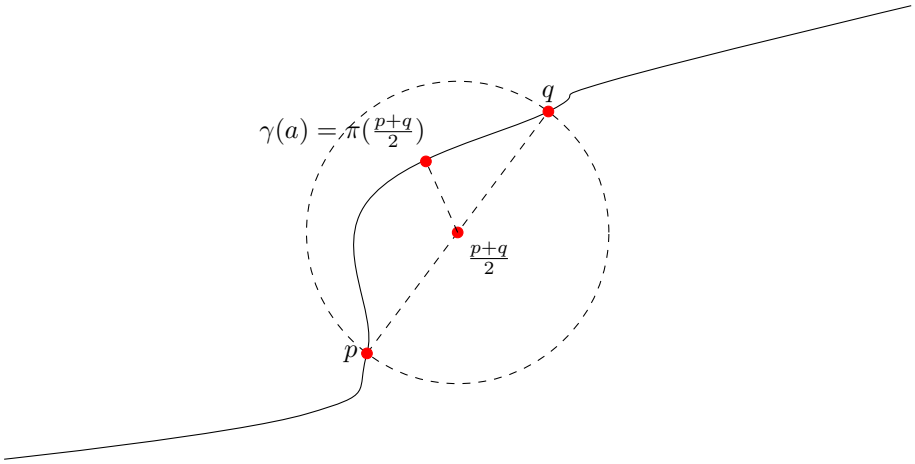


Figure 3: Local Estimate

The inequality above implies the geodesic $\gamma(s)$ connecting p and q must be partly inside the ball $B(\frac{p+q}{2},\frac{|p-q|}{2})$, since $\overline{pq}$ is the diameter of this ball and

$$|\gamma(s)|=d(p,q)\le\frac{3}{2}|p-q|<\frac{\pi}{2}|p-q|, \tag{109}$$

which implies there exist $a \in (0, s)$ such that

$$|\gamma(a) - \frac{p+q}{2}| < \frac{|p-q|}{2}, \tag{110}$$

which contradicts with $\varphi(p), \varphi(q)$ satisfy global reach case eq. (108). □

Since φ is almost isometry, i.e. $|\|d\varphi\| - 1| < \epsilon$, thus $d(p,q) \leq s/(1+\epsilon)$ implies $d(\varphi(p), \varphi(q)) \leq s$. Since we have proved that the local reach of $\varphi(M)$ admits a uniform lower bound $\tau_{l,\varphi(\mathcal{M})}$ depending only on the geometric parameters of $\mathcal{M}$, then we choose $s_0 = 2\sqrt{2}\tau_{l,\varphi(\mathcal{M})}$, then $\varphi(p), \varphi(q)$ cannot satisfy global reach case for any $\varphi(M) \in \varphi(\mathcal{M})$ Consequently, only when $d(p,q) > \frac{s_0}{1+\epsilon}$, $\varphi(p), \varphi(q)$ can satisfy global reach case.

And in the following, we will claim that if the geodesic distance of two points $p, q \in M$ is large enough, then the Euclidean distance of $\varphi(p), \varphi(q) \in \varphi(M)$ is bounded from below, i.e.

Lemma A.10. *For large enough $s_1 > 0$, there exists $r_1 > 0$ such that*

$$d(p,q) > s_1 \implies |\varphi(p) - \varphi(q)| > r_1. \tag{111}$$

We call this the global reach condition.

To bound the Euclidean distance from below, we need to bound the heat kernel $K(t,p,q)$ from above and $K(t,p,p)$ from below to control this Euclidean distance. For upper bound, we have

Lemma A.11. *Let M be a complete Riemannian manifold of dimension d with Ricci curvature is greater than $-\kappa(d-1)$ for some $\kappa \geq 0$, then heat kernel satisfies:*

$$K(t,p,q) \leq \frac{C_1(d)}{t^{d/2}} \exp\left(C_2(d)\kappa t - \frac{2d(p,q)^2}{9t}\right) \tag{112}$$

.

This lemma follows from Theorem B.3 of Li-Yau and Lemma B.4 of Croke with $\alpha_1 = \frac{3}{2}$, $\alpha_2 = \frac{1}{2}$ and $t_0 \leq \iota^2/4$, and Theorem B.3, Lemma B.4 are in the appendix.

For lower bound, we only need on-diagonal lower estimate:

Lemma A.12. *[45, Thm. 3.1] Let M be a complete Riemannian manifold of dimension d with Ricci curvature is greater than $-\kappa(d-1)$ for some $\kappa \geq 0$. For any $t > 0$ and $p \in M$, we have*

$$K(t,p,p) \geq (4\pi t)^{-d/2} \exp\left(-\frac{\beta^2}{4}t - \frac{2\sqrt{3d}\beta}{3}\sqrt{t}\right). \tag{113}$$

where $\beta = \sqrt{\kappa}(d-1)$.

This follows Theorem B.5 of Wang with $\sigma^2 = \frac{3\beta^2}{8d}$ and $p = q$.

Proof of Lemma A.10. To show this, we will compute the distance between $\varphi(p)$ and $\varphi(q)$ in $\mathbb{R}^m$ directly and then estimate it using geodesic distance of p, q.

For any $p \in M$, $\varphi(p) = (t_0)^{\frac{d+2}{4}}\sqrt{2}(4\pi)^{\frac{d}{4}}\left(e^{-\lambda_1 t_0/2}e_1(p), \cdots, e^{-\lambda_m t_0/2}e_m(p)\right) \in \mathbb{R}^m$, thus the Euclidean distance between $\varphi(p)$ and $\varphi(q)$ is

$$|\varphi(p) - \varphi(q)| = (t_0)^{\frac{d+2}{4}}\sqrt{2}(4\pi)^{\frac{d}{4}}\sqrt{\sum_{i=1}^{m} e^{-\lambda_i t_0}(e_i(p) - e_i(q))^2} \tag{114}$$

$$= (t_0)^{\frac{d+2}{4}}\sqrt{2}(4\pi)^{\frac{d}{4}}\sqrt{\sum_{i=1}^{m} e^{-\lambda_i t_0}(e_i(p)^2 + e_i(q)^2 - 2e_i(p)e_i(q))} \tag{115}$$

$$= (t_0)^{\frac{d+2}{4}}\sqrt{2}(4\pi)^{\frac{d}{4}}\sqrt{\sum_{i=0}^{m} (e^{-\lambda_i t_0}e_i(p)^2 + e^{-\lambda_i t_0}e_i(q)^2 - 2e^{-\lambda_i t_0}e_i(p)e_i(q))}, \tag{116}$$

the last equality holds since e_0 is a constant function.

We have the representation of heat kernel

$$K(t,p,q) = \sum_{i=0}^{\infty} e^{-\lambda_i t} e_i(p) e_i(q), \tag{117}$$

and the truncated heat kernel K_m is

$$K_m(t,p,q) = \sum_{i=0}^{m} e^{-\lambda_i t} e_i(p) e_i(q), \tag{118}$$

and we have $\|K_m(t_0,p,\cdot) - K(t_0,p,\cdot)\|_\infty \le \epsilon'$, thus

$$|\varphi(p) - \varphi(q)| \ge (t_0)^{\frac{d+2}{4}} \sqrt{2} (4\pi)^{\frac{d}{4}} \sqrt{K_m(t_0,p,p) + K_m(t_0,q,q) - 2K_m(t_0,p,q)} \tag{119}$$

$$\ge (t_0)^{\frac{d+2}{4}} \sqrt{2} (4\pi)^{\frac{d}{4}} \sqrt{K(t_0,p,p) + K(t_0,q,q) - 2K(t_0,p,q) - 4\epsilon'}. \tag{120}$$

Combining two inequities in Lemma A.11 and Lemma A.12, we have

$$|\varphi(p) - \varphi(q)| \ge (t_0)^{\frac{d+2}{4}} 2(4\pi)^{\frac{d}{4}} \times \sqrt{(4\pi t_0)^{-\frac{d}{2}} \exp\left(-\frac{\beta^2 t_0}{4} - \frac{2\sqrt{3dt_0}\beta}{3}\right) - \frac{C_1(d)}{t_0^{d/2}} \exp\left(C_2(d)\kappa t_0 - \frac{2d(p,q)^2}{9t_0}\right)} - 2\epsilon'. \tag{121}$$

In our case, t_0, m are fixed when embedding M into $\mathbb{R}^m$. The first two items in the square root are only depends on the geometric properties of $\mathcal{M}$, and ϵ' is chosen when we perform DM, we can choose appropriate ϵ' based on our geometric setting such that

$$2\epsilon' < (4\pi t_0)^{-\frac{d}{2}} \exp\left(-\frac{\beta^2 t_0}{4} - \frac{2\sqrt{3dt_0}\beta}{3}\right) \tag{122}$$

Therefore, for large enough s_1 with $d(p,q) \ge s_1$, we have $|\varphi(p) - \varphi(q)| \ge r_1$, where r_1, s_1 depend on our setting of $\mathcal{M}$. □

The proof above only claimed the existence of s_1 and r_1. For the uniformity claim we now make them explicit, choosing ϵ', s_1 and r_1 so that they depend only on the geometric parameters of $\mathcal{M}$, we consider the item under square root in eq. (121) which is positive, that is

$$(4\pi t_0)^{-\frac{d}{2}} \exp\left(-\frac{\beta^2 t_0}{4} - \frac{2\sqrt{3dt_0}\beta}{3}\right) - \frac{C_1(d)}{t_0^{d/2}} \exp\left(C_2(d)\kappa t_0 - \frac{2s_1^2}{9t_0}\right) - 2\epsilon' > 0 \tag{123}$$

$$\iff \exp\left(C_2(d)\kappa t_0 - \frac{2s_1^2}{9t_0}\right) < \frac{(4\pi)^{-\frac{d}{2}}}{C_1(d)} \exp\left(-\frac{\beta^2 t_0}{4} - \frac{2\sqrt{3dt_0}\beta}{3}\right) - \frac{2(t_0)^{\frac{d}{2}}\epsilon'}{C_1(d)}. \tag{124}$$

We denote the right side $\frac{(4\pi)^{-\frac{d}{2}}}{C_1(d)} \exp\left(-\frac{\beta^2 t_0}{4} - \frac{2\sqrt{3dt_0}\beta}{3}\right) - \frac{2(t_0)^{\frac{d}{2}}\epsilon'}{C_1(d)}$ as F, we need it be a fixed constant, thus we can select

$$\epsilon' = \frac{(4\pi t_0)^{-\frac{d}{2}}}{8} \exp\left(-\frac{\beta^2 t_0}{4} - \frac{2\sqrt{3dt_0}\beta}{3}\right), \tag{125}$$

which implies

$$F = \frac{3(4\pi)^{-\frac{d}{2}}}{4C_1(d)} \exp\left(-\frac{\beta^2 t_0}{4} - \frac{2\sqrt{3dt_0}\beta}{3}\right), \tag{126}$$

Therefore, ϵ' is only depending on the geometric properties of $\mathcal{M}$. Choosing ϵ' will affect $m = \max\left\{N_0(d,\kappa,\iota,\epsilon,V,\frac{t_0}{2}), N_1(d,\kappa,\iota,V,\epsilon',t_0)\right\} + 1$ and is independent with choice of $t_0 = t_0(d,\kappa,\iota,\epsilon)$ when perform DM. We need set appropriate ϵ' based on geometric setting and derive m and t_0 at the beginning, but until now we know how to choose them.

Then we ask $\exp\left(C_2(d)\kappa t_0 - \frac{2s_1^2}{9t_0}\right) < \frac{2}{3}F$, which is equivalent to

$$s_1^2 > \frac{9t_0}{2}\left(C_2(d)\kappa t_0 + \frac{\beta^2 t_0}{4} + \frac{2\sqrt{3dt_0}\beta}{3} + \log\left(2(4\pi)^{\frac{d}{2}}C_1(d)\right)\right), \tag{127}$$

Intuitively, t_0 is small, so s_1 is not very large.

When s_1 is the square root of eq. (127) and $\epsilon' = \frac{(4\pi t_0)^{-\frac{d}{2}}}{8}\exp\left(-\frac{\beta^2 t_0}{4} - \frac{2\sqrt{3dt_0}\beta}{3}\right)$, we have

$$r = r_1 = \sqrt{t_0}\exp\left(-\frac{\beta^2 t_0}{8} - \frac{\sqrt{3dt_0}\beta}{3}\right) \tag{128}$$

We now split $d(p,q)$ into three cases according to the range, which together cover all pairs $p, q \in M$:

(i) $d(p,q) \le s_0/(1+\epsilon)$. Since $|\|d\varphi\| - 1| < \epsilon$, we have $d(\varphi(p),\varphi(q)) \le s_0$, then by Lemma A.9, $\varphi(p), \varphi(q)$ cannot be the global reach case.

(ii) $d(p,q) > s_1$. It implies $|\varphi(p) - \varphi(q)| > r_1$ by Lemma A.10, we have the reach is bounded from the below.

(iii) $s_0/(1+\epsilon) < d(p,q) < s_1$. If there is no gap between $s_0/1+\epsilon$ and s_1, that is

$$s_0^2 > \frac{9(1+\epsilon)^2 t_0}{2}\left(C_2(d)\kappa t_0 + \frac{\beta^2 t_0}{4} + \frac{2\sqrt{3dt_0}\beta}{3} + \log\left(2(4\pi)^{\frac{d}{2}}C_1(d)\right)\right). \tag{129}$$

Combining $s_0 = 2\sqrt{2}r_0$ and $r_0 = \tau_{l,\varphi(\mathcal{M})}$. The condition making $s_0/1+\epsilon \ge s_1$ is

$$\begin{aligned} 8\tau_{l,\varphi(\mathcal{M})}^2 &\ge 8/\left(d^2(t_0)^{\frac{d+2}{2}}2(4\pi)^{\frac{d}{2}}C_2^2\sum_{k=1}^{m}e^{-\lambda_k t_0}(1+\lambda_k)^{2+\frac{d}{2}}\right) \\ &\ge \frac{9(1+\epsilon)^2 t_0}{2}\left(C_2(d)\kappa t_0 + \frac{\beta^2 t_0}{4} + \frac{2\sqrt{3dt_0}\beta}{3} + \log\left(2(4\pi)^{\frac{d}{2}}C_1(d)\right)\right), \quad (*) \end{aligned}$$

which is an assumption only based on geometric properties since $t_0 = t_0(d,\kappa,\iota,\epsilon)$, $m = \max\{N_0, N_1\} + 1$ and so on. If $(*)$ fails, there is an uncovered interval $(s_0/(1+\epsilon),\ s_1)$, we need to verify $d(p,q)$ in it. To do this, we need a tighter lower bound on s_0, and a short time expansion of the heat kernel that controls $K(t,p,p) - K(t,p,q)$ in this interval.

Now we consider (iii) when there is a gap. First we estimate $\tau_{l,\varphi(M)}$ in more details. From now, we ignore some constants and merge them as C to make our computation clearly. We have $\|II\|_{op} \le C(t_0)^{\frac{d+2}{4}}\sqrt{\sum_{k=1}^{m}e^{-\lambda_k t_0}(1+\lambda_k)^{2+\frac{d}{2}}}$, combining the estimate for eigenvalues

$$C_1^{1+\text{diam}\sqrt{\kappa}}\text{diam}^{-2}k^{2/d} \le \lambda_k \le C + C'k^{2/d}, \tag{130}$$

we can derive

$$\|II\|_{op} \le C(t_0)^{\frac{d+2}{4}}\sqrt{\sum_{k=1}^{m}e^{-k^{2/d}t_0}k^{1+\frac{4}{d}}} < C(t_0)^{\frac{d+2}{4}}\sqrt{\sum_{k=1}^{\infty}e^{-k^{2/d}t_0}k^{1+\frac{4}{d}}}. \tag{131}$$

Now we estimate $\sum_{k=1}^{\infty}e^{-k^{2/d}t_0}k^{1+\frac{4}{d}}$ by integral. We have

$$\sum_{k=1}^{\infty}e^{-k^{2/d}t_0}k^{1+\frac{4}{d}} \le \int_0^{\infty}e^{-x^{2/d}t_0}x^{1+\frac{4}{d}}dx + C \tag{132}$$

let $u = x^{2/d}t_0$, then $x^{1+\frac{4}{d}} = (\frac{u}{t_0})^{2+\frac{d}{2}}$ and $dx = \frac{u^{\frac{d}{2}-1}d}{2t_0^{d/2}}du$, then

$$\int_0^{\infty}e^{-x^{2/d}t_0}x^{1+\frac{4}{d}}dx = \frac{d}{2t_0^{2+d}}\int_0^{\infty}e^{-u}u^{d+1}du = \frac{d}{2t_0^{2+d}}\Gamma(d+2) \tag{133}$$

Therefore, $\|II\|_{op}^2 \leq Ct_0^{\frac{d+2}{2}}/t_0^{d+2} = Ct_0^{-(1+\frac{d}{2})}$, and $s_0^2 = 8\tau_{l,\varphi(M)} \geq Ct_0^{1+\frac{d}{2}}$.

Now we try to estimate the diffusion distance carefully using the asymptotic expression, which is due to Varadhan:

$$\lim_{t\to 0} -4t\log K(t,p,q) = d(p,q)^2 \tag{134}$$

and thus

$$\lim_{t\to 0} -4t\log \frac{K(t,p,q)}{K(t,p,p)} = d(p,q)^2 \tag{135}$$

which implies $K(t,p,q) \approx K(t,p,p)e^{-\frac{d(p,q)^2}{4t}}$ for small enough t. Then

$$K(t,p,p) - K(t,p,q) \approx (1-e^{-\frac{d(p,q)^2}{4t}})K(t,p,p). \tag{136}$$

To make it more acurate, we have

$$\left| K(t,p,q) - \frac{e^{-d(p,q)^2/4t}}{(4\pi t)^{d/2}} \sum_{j=0}^{v} t^j \frac{\Phi_j(p,q)}{j!} \right| \leq C(v)t^{v+1}\frac{e^{-d(p,q)^2/4t}}{(4\pi t)^{d/2}} \tag{137}$$

holds for any $v \in \mathbb{N}$, all (x,y) not in the cut locus and t is less than one specified T, and $\Phi_j(p,q)$ are the heat kernel coefficients [46].

Then we have

$$\frac{K(t,p,q)}{K(t,p,p)} \leq \frac{\frac{e^{-d(p,q)^2/4t}}{(4\pi t)^{d/2}}(\sum_{j=0}^{v} t^j \frac{\Phi_j(p,q)}{j!} + C(v)t^{v+1})}{\frac{1}{(4\pi t)^{d/2}}(\sum_{j=0}^{v} t^j \frac{\Phi_j(p,p)}{j!} - C(v)t^{v+1})} \tag{138}$$

thus

$$K(t,p,p) - K(t,p,q) \geq \left(1 - \frac{\sum_{j=0}^{v} t^j \frac{\Phi_j(p,q)}{j!} + C(v)t^{v+1}}{\sum_{j=0}^{v} t^j \frac{\Phi_j(p,p)}{j!} - C(v)t^{v+1}} e^{-d(p,q)^2/4t}\right) K(t,p,p). \tag{139}$$

We assume that we can find t and v such that $\frac{\sum_{j=0}^{v} t^j \frac{\Phi_j(p,q)}{j!} + C(v)t^{v+1}}{\sum_{j=0}^{v} t^j \frac{\Phi_j(p,p)}{j!} - C(v)t^{v+1}} \leq e^{d(p,q)^2/8t}$ when $d(p,q)^2 \in (s_0^2, s_1^2)$. We also have

$$K(t,p,p) \geq (4\pi t)^{-d/2}\exp\left(-\frac{\beta^2}{4}t - \frac{2\sqrt{3d}\beta}{3}\sqrt{t}\right). \tag{140}$$

thus

$$K(t,p,p) - K(t,p,q) \geq \frac{1-e^{-\frac{d(p,q)^2}{8t}}}{(4\pi t)^{d/2}}\exp\left(-\frac{\beta^2}{4}t - \frac{2\sqrt{3d}\beta}{3}\sqrt{t}\right). \tag{141}$$

when t is small, $\exp\left(-\frac{\beta^2}{4}t - \frac{2\sqrt{3d}\beta}{3}\sqrt{t}\right)$ is obviously bounded from below, we denote the lower bound as C_l. The remaining part increases monotonically with $d(p,q)^2$ in (s_0^2, s_1^2), so

$$K(t,p,p) - K(t,p,q) \geq \frac{1-e^{-\frac{C}{8}t^{d/2}}}{(4\pi t)^{d/2}} C_l \tag{142}$$

which is converge to a positive constant as $t \to 0$. This implies $K(t,p,p) - K(t,p,q)$ has a positive lower bound with an appropriate t, also $|\varphi(p) - \varphi(q)|$ has the lower bound. So when $d(p,q)^2 \geq s_0^2 = Ct^{1+\frac{d}{2}}$, we can always find the lower bound of $|\varphi(p) - \varphi(q)|$. Hence the global reach are bounded from below if we choose appropriate t.

Therefore, the local reach and global reach both have the uniform lower bound, thus $\tau_{\varphi(M)} \geq \tau_{\varphi(\mathcal{M})} > 0$ for some fixed constant.

B Supplemental material

B.1 Pushforward Density

Here we list the lemma and theorem for estimating pushforward density.

Lemma B.1. *(Area Formula)[47] If $f : \mathbb{R}^m \to \mathbb{R}^n$ is Lipschitzian and $m \leq n$, then*

$$\int_A g(f(x))J_f(x)d\mathcal{L}^m x = \int_{\mathbb{R}^n} g(y)N(f|A, y)d\mathcal{H}^m y$$

where A is an lebesgue measurable set, $J_f(x)$ is the Jacobian $\sqrt{\det(d_x f^T d_x f)}$ and $g : \mathbb{R}^n \to \mathbb{R}$ and $N(f|A, y) < \infty$ for $\mathcal{H}^m$ almost all y.

Lemma B.2. *If P is distribution on $\mathcal{M}$ with density f with respect to the d-dimensional Hausdorff measure and φ is a diffeomorphism, then the density $g(p')$ with respect to d-dimensional Hausdorff measure of $P' := \varphi_{\#}P$ is*

$$g(p') = f(p)/\sqrt{\det\left(\pi_{T_pM} \circ d\varphi_p^T \circ d\varphi_p|_{T_pM}\right)}$$

where $p = \varphi^{-1}(p')$.

Proof of Lemma B.2. Let $p \in \mathcal{M}$ be a fixed point and choose r small enough such that exponential map $\Psi_p : T_p\mathcal{M} \to \mathbb{R}^D$ is an injection onto $B(p, r) \cap \mathcal{M}$.

Choose $A \subset B(p, r) \cap \mathcal{M}$, by definition of pushforward measure, we have

$$\int_{\varphi(A)} dP' = \int_A dP = \int_A f(y)d\mathcal{H}^d y$$

Since Ψ_p is an injection, so $N(\Psi_p|\Psi_p^{-1}(A), y) = 1$ for all $y \in A$ and 0 otherwise. By Lemma B.1, we have

$$\int_A f(y)d\mathcal{H}^d y = \int_{\Psi_p^{-1}(A)} f(\Psi_p(x))J_{\Psi_p}(x)d\mathcal{L}^d x$$

And φ is diffeomorphism, so $h = \varphi \circ \Psi_p$ is also injective, so

$$\begin{aligned}\int_{\Psi_p^{-1}(A)} f(\Psi_p(x))J_{\Psi_p}(x)d\mathcal{L}^d x &= \int_{\Psi_p^{-1}(A)} f(\varphi^{-1}(h(x)))\frac{J_{\Psi_p}(h^{-1}(h(x)))}{J_{\varphi\circ\Psi_p}(h^{-1}(h(x)))}J_{\varphi\circ\Psi_p}(x))d\mathcal{L}^d x \\ &= \int_{\varphi(A)} f(\varphi^{-1}(z))\frac{J_{\Psi_p}(h^{-1}(z))}{J_{\varphi\circ\Psi_p}(h^{-1}(z))}d\mathcal{H}^d z\end{aligned}$$

Thus

$$\int_{\varphi(A)} dP' = \int_{\varphi(A)} f(\varphi^{-1}(z))\frac{J_{\Psi_p}(h^{-1}(z))}{J_{\varphi\circ\Psi_p}(h^{-1}(z))}d\mathcal{H}^d z$$

$T_p\mathcal{M}$ is a subspace of $\mathbb{R}^D$ with dimension d, so we can choose basis of $T_p\mathcal{M}$ such that all elements in diagonal of transformation matrix are 1. We notice that $h^{-1}(z) = \Psi^{-1}_{\varphi^{-1}(z)} \circ \varphi^{-1}(z) = 0$, and $d_0\Psi_p = I_D + d_0N_p = I_D$ is inclusion map. Thus

$$f(\varphi^{-1}(z))\frac{J_{\Psi_p}(h^{-1}(z))}{J_{\varphi\circ\Psi_p}(h^{-1}(z))} = f(p)/\sqrt{\det\left(\pi_{T_pM} \circ d\varphi_p^T \circ d\varphi_p|_{T_pM}\right)}$$

where $p = \varphi^{-1}(z)$. □

B.2 Heat Kernel Estimate

Theorem B.3 (Upper Bound for The Heat Kernel)**.** *[48, Cor. 3.1] Let M be a complete manifold without boundary. If the Ricci curvature of M is bounded from below by $-\kappa$ for some constant $\kappa \geq 0$, then for any $1 < \alpha_1 < 2$ and $0 < \alpha_2 < 1$, the heat kernel satisfies:*

$$K(t, p, q) \leq C(\alpha_2)^{\alpha_1}V^{-1/2}(B_p(\sqrt{t}))V^{-1/2}(B_q(\sqrt{t}))\exp\left(C(d)\alpha_2(\alpha_1 - 1)^{-1}\kappa t - \frac{d(p, q)^2}{(4 + \alpha_2)t}\right), \tag{143}$$

where $B_x(r)$ is geodesic ball centered at p with radius r, $C(\alpha_2)$ depends on α_2 with $C(\alpha_2) \to \infty$ as $\alpha_2 \to 0$.

Lemma B.4 (Estimate for Volume of Geodesic Ball)**.** *[49, Prop. 14] Let M be a complete manifold without boundary. then for $r \leq \iota_M/2$, we have*

$$Vol(B_p(r)) \geq C'(d) r^d, \tag{144}$$

where $C'(d) = \frac{2^{d-1}\omega(d-1)^d}{d^d \omega(d)^{d-1}}$, and $\omega(d) = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}$ is the volume of the unit d-dimensional sphere . Thus $C'(d) = \frac{2^d \Gamma(\frac{d}{2})^{d-1}}{d^d \Gamma(\frac{d-1}{2})^d}$.

Theorem B.5 (Lower Bound for Heat Kernel)**.** *[45, Thm. 3.1] Let M be a complete Riemannian manifold of dimension d with Ricci curvature is greater than $-\kappa(d-1)$ for some $\kappa \geq 0$. For any $t, \sigma > 0$ and $p, q \in M$, we have*

$$K(t,p,q) \geq (4\pi t)^{-d/2} \exp\left(- \left(\frac{1}{4t} + \frac{\sigma}{3\sqrt{2t}} \right) d(p,q)^2 - \frac{\beta^2}{4} t - (\frac{\beta^2}{4\sigma} + \frac{2d\sigma}{3})\sqrt{2t} \right), \tag{145}$$

where $\beta = \sqrt{\kappa}(d-1)$.

B.3 Selection of t_0, m

B.3.1 Selection of t_0

For t_0, see theorem 4.4 in [5]. We define

$$\Gamma(s,x,y) := r^d K(sr^2, u^{-1}(xr), u^{-1}(yr)), \tag{146}$$

where $u : B_r(p) \to \mathbb{R}^d$ is the harmonic coordinates.

We select $R_1 = R_1(d,\epsilon)$ such that, $\frac{1}{2} \leq s \leq 2$,

$$\int_{\mathbb{R}^d \setminus B_{R_1}(0)} |\nabla \Gamma(s,0,y)|^2 dy < \epsilon. \tag{147}$$

We set $\alpha = \frac{1}{2}$ and $Q > 1$ such that

$$2(Q-1)C(d,\alpha) \leq \sigma, \tag{148}$$

where $C(d,\alpha)$ is the constant in Lemma 6.1 of [5] and $\sigma \leq \sigma_1(d,\epsilon)$ such that $C(d,\alpha)\sigma |B_{R_1}(0)| < \epsilon$.

We also set $R_0 = R(d,\alpha,C(d),Q)$ is the radius in Lemma 6.1, where $C(d)$ is the constant as below:

$$|\nabla \Gamma(s,x,y)| \leq \frac{C(d)}{s^{(d+1)/2}} \exp\left(-\frac{|x-y|^2}{8} \right). \tag{149}$$

Let $r_h := r_h(d,\kappa,\iota,\alpha,Q)$ be the harmonic radius, and set $r_3 = r_3(d,\kappa,\iota,\epsilon) < r_h/2$ such that for $t < 2r_3^2$, then

$$(2t)^{\frac{d+2}{2}} \int_{M \setminus B_{r_h/2}(p)} |\nabla K(t,p,q)|^2 dq < \epsilon. \tag{150}$$

Now we set

$$r_0 = \min\left\{ \frac{r_h}{R_0}, \frac{r_h}{R_1}, r_3 \right\}, \tag{151}$$

then $t_0 = r_0^2/2$.

B.3.2 Selection of N_0

According to theorem 4.4 in [5],

$$\mathcal{H}(p)(q) := (2t)^{\frac{d+2}{4}} \sqrt{2} (4\pi)^{d/4} K(t,p,q) \tag{152}$$

is an embedding of M into $L^2(M)$ with

$$1 - \epsilon < \|(d\mathcal{H})_p\| < 1 + \epsilon. \tag{153}$$

In addition, there is an isometry from $L^2(M)$ to l^2 by

$$U(f)_j = \int_M f(q)e_j(q)dq. \tag{154}$$

Then,

$$p \in M \xrightarrow{\mathcal{H}} \mathcal{H}(p) \in L^2(M) \xrightarrow{U} U(\mathcal{H}(p)) \tag{155}$$

is an embedding, i.e.

$$\mathcal{F}(p) := (2t)^{\frac{d+2}{4}}\sqrt{2}(4\pi)^{d/4}\left(e^{-\lambda_1 t}e_1(p), e^{-\lambda_2 t}e_2(p), \dots\right) \tag{156}$$

is an embedding. We ignore the e_0 since it is the constant.

Therefore, for any $v \in T_pM$ with $|v| = 1$, we have

$$2(2t)^{\frac{d+2}{2}}(4\pi)^{d/2}\sum_{i=1}^{\infty} e^{-2\lambda_i t}(\nabla e_i(v))^2 = \|(d\mathcal{F})_p\|^2 = \|(d\mathcal{H})_p\|^2 \in \left((1-\epsilon)^2, (1+\epsilon)^2\right). \tag{157}$$

In section A.1.2, we derived

$$\|\nabla e_j\| \leq C\lambda_j^{\frac{d+2}{4}}, \tag{158}$$

thus

$$2(2t)^{\frac{d+2}{2}}(4\pi)^{d/2}\sum_{i=1}^{\infty} e^{-2\lambda_i t}(\nabla e_i(v))^2 \leq 2C(2t)^{\frac{d+2}{2}}(4\pi)^{d/2}\sum_{i=1}^{\infty} e^{-2\lambda_i t}\lambda_j^{\frac{d+2}{2}}. \tag{159}$$

Furthermore, λ_i is bounded from below by

$$\lambda_i^{d/2} \geq \alpha(d)\frac{i}{V}\left(\frac{V}{\int_0^{\text{diam}} F_\kappa(r)dr}\right)^{n(n+1)/(n-1)}, \tag{160}$$

where $F_\kappa(r) = \begin{cases} (-\kappa)^{-1/2}(\sinh\sqrt{-\kappa}r)^{n-1} & \text{if } \kappa < 0, \\ r^{n-1} & \text{if } \kappa = 0, \\ (\kappa)^{-1/2}(\sinh\sqrt{\kappa}r)^{n-1} & \text{if } \kappa < 0. \end{cases}$ and $\kappa(n-1)$ is the lower bound of Ricci curvature of M[50]. Therefore, $\lambda_i \geq ci^{2/d}$ for some uniform constant c, when i is large enough, we have

$$e^{-2\lambda_i t}\lambda_i^{\frac{d+2}{2}} \leq e^{-\lambda_i t} \leq e^{-ci^{2/d}t}, \tag{161}$$

then

$$2C(2t)^{\frac{d+2}{2}}(4\pi)^{d/2}\sum_{i=N_0}^{\infty} e^{-2\lambda_i t}\lambda_j^{\frac{d+2}{2}} \leq C'\sum_{i=N_0}^{\infty} e^{-ci^{2/d}t}. \tag{162}$$

It is easy to verify that the tail can be controlled to arbitrary small, thus we can truncate $\mathcal{F}$ by $\mathcal{F}^{N_0}$:

$$\mathcal{F}^{N_0}(p) := (2t)^{\frac{d+2}{4}}\sqrt{2}(4\pi)^{d/4}\left(e^{-\lambda_1 t}e_1(p), \dots, e^{-\lambda_{N_1} t}e_{N_1}(p)\right) \tag{163}$$

such that $1 - \epsilon < \|\mathcal{F}^{N_0}\| < 1 + \epsilon$.

B.3.3 Selection of N_1

We need find $N_1 = N_1(d, \kappa, \iota, V, \epsilon', t_0)$ such that $|K_m(t', p, \cdot) - K(t', p, \cdot)| \leq \epsilon'$ holds for any $m \geq N_1$ and any t' with $t_0 \leq t' \leq 4$.

In fact,

$$|K_m(t', p, q) - K(t', p, q)| = |\sum_{i=m}^{\infty} e^{-\lambda_i t'} e_i(p) e_i(q)| \tag{164}$$

$$\leq \sum_{i=N_1}^{\infty} e^{-\lambda_i t'} |e_i(p)||e_i(q)|. \tag{165}$$

We also have the estimate for L-infinity norm of eigenfunction [36]:

$$\|e_i(p)\|_{L^\infty} \leq C(d, \kappa, \iota) \lambda_i^{\frac{d-1}{4}} \|e_i(p)\|_{L^2}. \tag{166}$$

Therefore, we have

$$|K_m(t', p, q) - K(t', p, q)| \leq C(d, \kappa, \iota) \sum_{i=N_1}^{\infty} e^{-\lambda_i t'} \lambda_i^{\frac{d-1}{2}}. \tag{167}$$

Similarly, $\lambda_i \geq 2ci^{2/d}$ for some uniform constant c, when i is large enough, we have

$$e^{-\lambda_i t'} \lambda_i^{\frac{d-1}{2}} \leq e^{-\lambda_i t'/2} \leq e^{-ci^{2/d} t'}, \tag{168}$$

thus

$$|K_m(t', p, q) - K(t', p, q)| \leq C(d, \kappa, \iota) \sum_{i=N_1}^{\infty} e^{-ci^{2/d} t'}. \tag{169}$$

It is obvious that $\sum_{i=1}^{\infty} e^{-ci^{2/d} t'}$ converges, so there exists $N_1 = N_1(d, \kappa, \iota, V, \epsilon', t_0)$ such that $|K_m(t', p, \cdot) - K(t', p, \cdot)| \leq \epsilon'$.

We notice that more precise estimate of eigenvalues and eigenfunction can make N_1 smaller, here we only list one possible method for selection.

B.4 Example of S^2

In this section, we consider the special case of S^2 with radius 1 to verify the eq. (*).

For S^2, the intrinsic dimension $d = 2$, the injectivity radius is π, the sectional curvature and the Ricci curvature are both 1, thus negative bound $\kappa = 0$, $\beta = \sqrt{\kappa}(d-1) = 0$.

The eigenvalues of S^2 is $l(l+1)$, with multiplicity $2l+1$, and the complex-valued eigenfunctions are

$$Y_l^m(\theta, \phi) = N_l^m P_l^m(\cos\theta) e^{im\phi} \quad l \in \mathbb{N}, m = 0, \pm 1, \ldots, \pm l, \tag{170}$$

where N_l^m is the normalization factor, P_l^m is the associated Legendre polynomials, then the real valued eigenfunctions are

$$Y_{lm} = \begin{cases} \sqrt{2}(-1)^m \Re(Y_l^m) & \text{if } m > 0, \\ Y_l^0 & \text{if } m = 0, \\ \sqrt{2}(-1)^m \Im(Y_l^{-m}) & \text{if } m < 0. \end{cases} \tag{171}$$

$$= \begin{cases} (-1)^m \sqrt{2} \sqrt{\frac{2l+1}{2\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) \cos(m\phi) & \text{if } m > 0, \\ \sqrt{\frac{2l+1}{4\pi}} P_l^m(\cos\theta) & \text{if } m = 0, \\ (-1)^m \sqrt{2} \sqrt{\frac{2l+1}{2\pi} \frac{(l-|m|)!}{(l+|m|)!}} P_l^{|m|}(\cos\theta) \sin(|m|\phi) & \text{if } m < 0. \end{cases} \tag{172}$$

We first consider embedding S^2 into $L^2(S^2)$ as the following:

$$f : p \to (4t)\sqrt{2\pi}K(t, p, \cdot). \tag{173}$$

Its operator norm is

$$\|df_p\|^2 = \sup_{\|v\|=1} 32\pi t^2 \int_{S^2} |\nabla_v K(t, p, q)|^2 dq. \tag{174}$$

We need find t_0 such that for any $0 < t < t_0$,

$$(1-\epsilon)^2 < \sup_{\|v\|=1} 32\pi t^2 \int_{S^2} |\nabla_v K(t, p, q)|^2 dq < (1+\epsilon)^2. \tag{175}$$

The coordinate represent of S^2 is

$$r(\theta, \phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \quad \theta \in (0, \pi), \phi \in (0, 2\pi), \tag{176}$$

and thus we can select unnormalized coordinate basis as the following

$$\frac{\partial}{\partial\theta} = (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta), \tag{177}$$

$$\frac{\partial}{\partial\phi} = (-\sin\theta\sin\phi, \sin\theta\cos\phi, 0). \tag{178}$$

Therefore, $g_{\theta\theta} = 1, g_{\theta\phi} = 0, g_{\phi\phi} = \sin^2\theta$. We normalize basis and then we have the gradient of f

$$\nabla f = \frac{\partial f}{\partial\theta} e_\theta + \frac{\partial f}{\partial\phi}\frac{1}{\sin\theta} e_\phi, \tag{179}$$

where $e_\theta = \frac{\partial}{\partial\theta}$, $e_\phi = \frac{1}{\sin\theta}\frac{\partial}{\partial\phi}$.

Then for $v = v^1 e_\theta + v^2 e_\phi$, we have

$$\nabla_v f = v(f) = df(v) = \langle \nabla f, v\rangle = \frac{\partial f}{\partial\theta} v^1 + \frac{\partial f}{\partial\phi}\frac{v^2}{\sin\theta}. \tag{180}$$

Now we compute $\int_{S^2} |\nabla_v K(t, p, q)|^2 dq$, without loss of generality, we can assume $v = \frac{\sqrt{2}}{2} e_\theta + \frac{\sqrt{2}}{2} e_\phi$ and $p = (\pi/2, 0)$ ($\pi/2$ can simplify Riemannian metric.) due to symmetry of S^2.

We first prove $\int_{S^2} \frac{\partial K}{\partial\theta}\frac{\partial K}{\partial\phi} dq = 0$. We know

$$K(t, p, q) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} e^{-l(l+1)t} Y_l^m(p), \bar{Y}_l^m(q) \tag{181}$$

$$= \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} e^{-l(l+1)t} P_l(\cos d), \tag{182}$$

where the equality comes from spherical harmonic addition theorem, d is the geodesic distance between p and q and $\cos d = \cos\theta_p\cos\theta_q + \sin\theta_p\sin\theta_q\cos(\phi_p - \phi_q)$.

Thus

$$\frac{\partial K}{\partial\theta} = \sum_{l=0}^{\infty} c_l P_l'(\cos d)\left(-\sin\theta_p\cos\theta_q + \cos\theta_p\sin\theta_q\cos(\phi_p - \phi_q)\right), \tag{183}$$

$$\frac{\partial K}{\partial\phi} = \sum_{l=0}^{\infty} c_l P_l'(\cos d)\left(-\sin\theta_p\sin\theta_q\sin(\phi_p - \phi_q)\right). \tag{184}$$

At $p = (\pi/2, 0)$, they are

$$\frac{\partial K}{\partial\theta} = \sum_{l=0}^{\infty} c_l P_l'(\sin\theta_q\cos\phi_q)\left(-\cos\theta_q\right), \tag{185}$$

$$\frac{\partial K}{\partial\phi} = \sum_{l=0}^{\infty} c_l P_l'(\sin\theta_q\cos\phi_q)\left(\sin\theta_q\sin\phi_q\right). \tag{186}$$

We only need to check

$$\int_{S^2} P_l'(\sin\theta_q\cos\phi_q)P_{l'}'(\sin\theta_q\cos\phi_q)\cos\theta_q\sin\theta_q\sin\phi_q dq = 0. \tag{187}$$

We notice that P_l is the Legendre polynomial, so P_l' is a polynomial. All terms are looks like $(\sin\theta_q\cos\phi_q)^k$.

It is easy to verify

$$\int_{S^2} (\sin\theta_q\cos\phi_q)^k\cos\theta_q\sin\theta_q\sin\phi_q dq \tag{188}$$

$$= \int_0^{2\pi}\int_0^{\pi} \sin^{k+2}(\theta_q)\cos\theta_q\cos^k(\phi_q)\sin\phi_q d\phi d\theta \tag{189}$$

$$= \int_0^{2\pi} \cos^k(\phi_q)\sin\phi_q d\phi \int_0^{\pi} \sin^{k+2}(\theta_q)\cos\theta_q d\theta_q = 0. \tag{190}$$

Therefore, add them up and we get the result we need. And immediately, we have

$$\int_{S^2} |\nabla_v K(t,p,q)|^2 dq = \int_{S^2} \left|\frac{\sqrt{2}}{2}\frac{\partial K}{\partial\theta} + \frac{\sqrt{2}}{2}\frac{\partial K}{\partial\phi}\right|^2 dq \tag{191}$$

$$= \frac{1}{2}\int_{S^2} \left(\frac{\partial K}{\partial\theta}\right)^2 + \left(\frac{\partial K}{\partial\phi}\right)^2 + 2\frac{\partial K}{\partial\theta}\frac{\partial K}{\partial\phi} dq \tag{192}$$

$$= \frac{1}{2}\int_{S^2} \langle\nabla K, \nabla K\rangle\, dq \tag{193}$$

Then by Green Identity,

$$\int_{S^2} \langle\nabla K, \nabla K\rangle\, dq = -\int_{S^2} K(t,p,q)\Delta K(t,p,q) dq, \tag{194}$$

and $K(t,p,q)$ is the heat kernel, thus

$$\Delta_p K(t,p,q) = \frac{\partial}{\partial t}K(t,p,q) \tag{195}$$

$$= -\frac{1}{4\pi}\sum_{l=0}^{\infty} l(l+1)(2l+1)e^{-l(l+1)t}P_l(\cos d). \tag{196}$$

Therefore,

$$-\int_{S^2} K(t,p,q)\Delta K(t,p,q) dq \tag{197}$$

$$= \frac{1}{(4\pi)^2}\sum_{l,l'=0}^{\infty} l(l+1)(2l+1)(2l'+1)e^{-\left(l(l+1)+l'(l'+1)\right)t}\int_{S^2} P_l(\cos d)P_{l'}(\cos d) dq. \tag{198}$$

To compute the last integral, we can always assume p is the north pole by symmetry, so the geodesic distance between p and q is naturally the polar angle of q, that is

$$\int_0^{2\pi}\int_0^{\pi} P_l(\cos\theta)P_{l'}(\cos\theta)\sin\theta d\theta d\phi \tag{199}$$

$$= 2\pi\int_{-1}^{1} P_l(x)P_{l'}(x)dx = \frac{4\pi}{2l+1}\delta_l^{l'}. \tag{200}$$

hence

$$\int_{S^2} \langle\nabla K, \nabla K\rangle\, dq = \int_{S^2} K(t,p,q)\Delta K(t,p,q) dq \tag{201}$$

$$= \frac{1}{4\pi}\sum_{l=0}^{\infty} l(l+1)(2l+1)e^{-2l(l+1)t}. \tag{202}$$

Therefore,

$$32\pi t^2 \int_{S^2} |\nabla_v K(t,p,q)|^2 dq = 4t^2 \sum_{l=0}^{\infty} l(l+1)(2l+1)e^{-2l(l+1)t}. \tag{203}$$

By the isometry from $L^2(M)$ to l^2

$$U(f)_j = \int_M f(q)e_j(q)dq, \tag{204}$$

we have

$$32\pi t^2 \sum_{i=1}^{\infty} e^{-2\lambda_i t}(\nabla e_i(v))^2 \in \left((1-\epsilon)^2, (1+\epsilon)^2\right), \tag{205}$$

We can truncate it at N_0 such that

$$32\pi t^2 \sum_{i=N_0}^{\infty} e^{-2\lambda_i t}(\nabla e_i(v))^2 < \epsilon^2. \tag{206}$$

We consider the first 8 eigenfunctions: $\frac{1}{2}\sqrt{\frac{3}{\pi}}\cos\theta$, $\frac{1}{2}\sqrt{\frac{3}{\pi}}\sin\theta\cos\phi$, $\frac{1}{2}\sqrt{\frac{3}{\pi}}\sin\theta\sin\phi\ldots$, we can obtain easily that the embedding norm, that is

$$32\pi t^2 \sum_{i=1}^{8} e^{-2\lambda_i t}(\nabla e_i(v))^2 = 4t^2 \sum_{l=1}^{2} l(l+1)(2l+1)e^{-2l(l+1)t}. \tag{207}$$

We set $\epsilon = 0.05$ and $t = 0.25$, then this embedding norm $\|df\| \in (0.95, 1.05)$.

In addition, to control the error between the truncated heat kernel and the heat kernel, we need

$$\|K_{N_1}(t,p,q) - K(t,p,q)\| = |\sum_{i=N_1}^{\infty} e^{-\lambda_i t}e_i(p)e_i(q)| < \epsilon', \tag{208}$$

where $\epsilon' = \frac{1}{32\pi t}\exp\left(-\frac{\beta^2 t}{4} - \frac{2\sqrt{3dt}\beta}{3}\right) = \frac{1}{32\pi t}$.

By spherical harmonic addition theorem, we have

$$\sum_{i=N_1}^{\infty} e^{-\lambda_i t}e_i(p)e_i(q) = \frac{1}{4\pi}\sum_{l=l_1}^{\infty} e^{-l(l+1)t}(2l+1)P_l(\cos d). \tag{209}$$

We can compute ϵ' and the summation above numerically, and we find that $l_1 = 3$ makes truncation error is smaller than ϵ'.

As a result, our embedding setting is $t = 0.25$ and $m = 8$.

B.4.1 Verification for Inequality (*)

We compute the second fundamental form of $\varphi(M)$ with ambient manifold $\mathbb{R}^m$. Through the numerical computation, the local reach for $\varphi(S^2)$ is approximately 0.646924.

Substituting coefficients into the eq. (*), it is

$$8\tau^2_{l,\varphi(S^2)} \geq \frac{9(1+0.05)^2 \times 0.25}{2}\left(\log\left(8\pi C_1(S^2)\right)\right), \tag{210}$$

by numerical computation, we have $C_1(S^2)$ is approximately 0.408912. After computation, the left side is approximately 3.34809, the right side is approximately 2.88982. Therefore, there exist manifolds ensures that eq. (*) holds.

C Details about Experiments

Here we discuss more about experiments. The generating process is the following:

1. Generate n samples from the uniform distribution on $[-1,1]^d$. Inner points are the ones that are sampled from $[-0.8,0.8]^d$.
2. Apply the transformation f to these points to get the samples on the manifold.
3. Use high dimensional rotation matrix to lift manifold into high dimension and add noise to the samples if needed.

that is

$$x \in \mathbb{R}^2 \overset{Ax+b}{\longrightarrow} h \in \mathbb{R}^2 \overset{f}{\longrightarrow} y_{small} \in \mathbb{R}^d \overset{Ratation}{\longrightarrow} y \in \mathbb{R}^{48} \tag{211}$$

For example, setting for swiss roll is

$$A = \begin{pmatrix} (\alpha-\beta)/2 & 0 \\ 0 & h/2 \end{pmatrix}, b = ((\alpha+\beta)/2, h/2), f = \begin{pmatrix} h_1 \cos h_1 \\ h_2 \\ h_1 \sin h_1 \end{pmatrix}, \tag{212}$$

the setting for torus is

$$A = \begin{pmatrix} \pi & 0 \\ 0 & \pi \end{pmatrix}, b = (\pi, \pi), f = \begin{pmatrix} (R + r\cos h_2)\cos h_1 \\ (R + r\cos h_2)\sin h_1 \\ r \sin h_2 \end{pmatrix}, \tag{213}$$

the setting for random manifold 1 is

$$A = \begin{pmatrix} -2.224, & 0.073 \\ -0.134 & -1.006 \end{pmatrix}, b = (0.124, 1.403), f = \begin{pmatrix} 1/(e^{h_1}+1) \\ \cos h_2 \\ \sin 2h_2 \end{pmatrix}, \tag{214}$$

the setting for random manifold 2 is

$$A = \begin{pmatrix} -1.885 & 0.111 \\ 0.006 & 1.266 \end{pmatrix}, b = (0.370, -0.457), f^1 = \begin{pmatrix} 1/(e^{h_1}+1) \\ e^{h_2} \end{pmatrix}, f^2 = \begin{pmatrix} f_1^1 + 0.5 \\ 2f_2^1 \end{pmatrix} \tag{215}$$

$$f^3 = \begin{pmatrix} f_1^2 \cos f_2^2 \\ f_1^2 \sin f_2^2 \end{pmatrix}, f = \begin{pmatrix} f^1 \\ f^3 \end{pmatrix} \tag{216}$$

These manifold data are sampled from non uniform distribution on the manifold, but the density is bounded from above and below, which satisfies our assumption.

C.1 More Experiments about Diffusion Maps

We frist discuss the DM on the unbounded linear space. For simplicity, we consider 2 dimensional linear subspace in $\mathbb{R}^3$, i.e. xOy plane, and the Laplacian operator Δ on it is $-\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2}$. Then for any $a, b \in \mathbb{R}$, $\cos(ax+by)$ is an eigenfunction of Δ with eigenvalue a^2+b^2. Therefore, its spectrum is continuous, thus we cannot enumerate eigenvalues and eigenfunctions, so DM is not well defined.

If we consider a bounded linear space, i.e. a square, we observe the horseshoe phenomenon, as shown in Figure 4. Selecting the first two coordinates of the DM exhibits the horseshoe phenomenon. However, removing the repeated eigenvector (φ_2 or y–axis) can recover the manifold correctly.

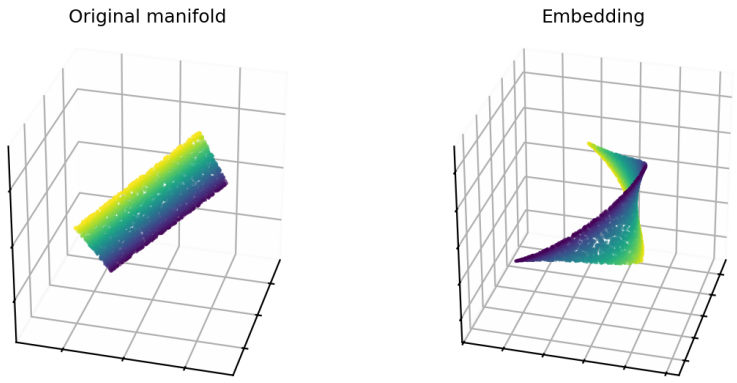


Figure 4: Diffusion maps on a bounded flat manifold exhibit the horseshoe phenomenon. Left: samples from a 2D linear subspace in $\mathbb{R}^3$. Right: the 3D diffusion-map embedding (ϕ_1, ϕ_2, ϕ_3), which bends the flat square into a horseshoe shape rather than recovering the plane.

C.2 Experiments about n

Our main theorems are for a family of manifolds, where the geometric quantities are defined independently of any sampling. In practice, a finite point cloud is used and it is only a numerical discretization of such a continuous object. The purpose of this experiment is to show that our estimates exhibit a convergence behavior: as the number of samples n increases, the estimation error decreases steadily toward zero. Such convergence indicates that the estimated quantities genuinely approximate the underlying geometric quantities, which makes the measurements reported in our main experiments meaningful.

We illustrate this on two quantities. For the Riemannian metric, we sample n points from the unit sphere S^2 and report the deviation of the estimated metric's singular values σ from 1, measured both on average ($\text{mean}\,|\sigma - 1|$) and in the worst case ($\max\,|\sigma - 1|$). For the reach, we sample n points from a torus with reach $\tau = 1.5$ and compare the estimate $\hat{\tau}$ against this ground truth. We vary $n \in \{500, 1000, 2000, 5000, 10000\}$, repeat each setting three times, and report the mean together with the min–max range.

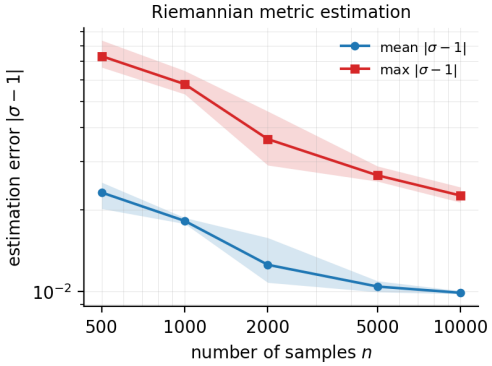


Figure 5: Metric estimation error on S^2 decreases as n grows

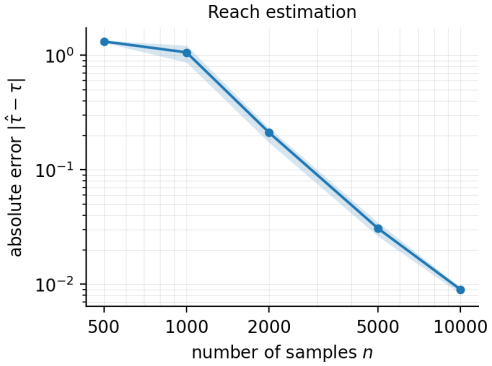


Figure 6: Reach estimation error on a torus ($\tau = 1.5$) decreases as n grows.

As shown in Figure 5 and Figure 6, the estimation error decreases steadily as n grows for both the Riemannian metric and the reach. This convergence confirms that our estimators are consistent in practice and that the measurements reported in our main experiments are meaningful.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Yes. Abstract and Introduction include our claims.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Yes. We discussed the limitation of our work in the Conclusion section.

   Our regularity condition is very rigorous, to avoid embedding manifolds through DM having vanishing global reach. We have verified inequality related to this assumption along with all the other assumption holds for some well-behaved manifolds (see section B.4), but it is strict.

   This paper includes only a few experiments for the following reasons:

   (a) People do not have any method to compute eigenvalues and eigenfunctions for general manifolds, thus it is difficult or even impossible to derive the true tangent space and compute the estimating error.

   (b) Our algorithm is designed for a family of manifolds $\mathcal{M}$ whose cardinality equals that of the continuum and whose members can vary drastically in their geometric structures. As a result, it is not feasible to conduct experiments over all possible manifolds, nor to reliably evaluate the algorithm based on a finite collection of representative manifolds.

   (c) The constants involved in our theorem are merely asserted to exist and are not given in an explicit form. It is difficult to derive them through numerical simulations and then perform experiments.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.

- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Yes. We list all assumptions in section 2. For theorem in main part, see section 3 and section A. For lemma and theorem in appendix, we show proof or give reference.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Yes. We provide all algorithm needed and their references to reproduce experiments. We also provide code and parameters in the supplemental material.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example

(a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
(b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Yes. We provide code and parameters in the supplemental material.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We specify the generating process and we only have very few parameters to choose, so grid search is used.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: In the manifold learning setting, the estimation error is not Gaussian and the error bar is not well defined. Besides, large sample size vanishes the variance of the estimation error, which is the case of this paper.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We specify the experimental platform and running time in Experiment section.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We followed the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.

- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This research is not tied to particular applications and to our knowledge there are no direct paths to negative applications.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no safeguards risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

Justification: The paper does not use existing assets.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: The paper does not release new assets.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.