Weighted Graph Laplace Operator under Topological Noise

Tamal K. Dey* Pawas Ranjan* Yusu Wang*

Abstract

Recently, various applications have motivated the study of spectral structures (eigenvalues and eigenfunctions) of the so-called Laplace-Beltrami operator of a manifold and their discrete versions. A popular choice for the discrete version is the so-called Gaussian weighted graph Laplacian which can be applied to point cloud data that samples a manifold. Naturally, the question of stability of the spectrum of this discrete Laplacian under the perturbation of the sampled manifold becomes important for its practical usage. Previous results showed that the spectra of both the manifold Laplacian and discrete Laplacian are stable when the perturbation is "nice" in the sense that it is restricted to a diffeomorphism with minor area distortion. However, this forbids, for example, small topological changes.

We study the stability of the spectrum of the weighted graph Laplacian under more general perturbations. In particular, we allow arbitrary, including topological, changes to the hidden manifold as long as they are localized in the ambient space and the area distortion is small. Manifold Laplacians may change dramatically in this case. Nevertheless, we show that the weighted graph Laplacians computed from two sets of points, uniformly randomly sampled from a manifold and a perturbed version of it, have similar spectra. The distance between the two spectra can be bounded in terms of the size of the perturbation and some intrinsic properties of the original manifold.

1 Introduction

Spectral methods are popular tools in a variety of applications. Various recent works have identified the use of the spectral structures (eigenvalues and eigenfunctions) of the so-called Laplace-Beltrami operator of a manifold, which encodes the intrinsic geometry succinctly. Applications range from data denoising, clustering, and semi-supervised learning in manifold learning (see e.g, [2, 12, 14, 17] and references therein), to geometric optimization and shape editing in graphics (see e.g. surveys [18, 20]).

In practice, the underlying manifold is often approximated by a discrete input, e.g., a point cloud data (PCD) sampled from the manifold. We thus need a discrete version of the Laplace-Beltrami operator. A popular choice in the discrete setting is the so-called Gaussian-weighted graph Laplacian, both for its simplicity and for its convergence to the manifold Laplacian with increasing number of uniformly randomly sampled points. Naturally, the question of robustness of this discrete Laplacian with respect to perturbations of the sampled manifold becomes a practical issue. In particular, estimating the change in spectrum of the discrete Laplacian for PCDs sampled from manifolds M and its perturbed version N becomes important. Previous results show that both manifold and discrete Laplacians are stable when the perturbation is a diffeomorphism between M and N with minor area distortion [9]. This however forbids, for example, small topological changes between M and N.

In this paper, we aim to study the stability of the weighted graph Laplacian under more general perturbations. In particular, we allow arbitrary changes to the hidden manifold as long as they are localized in the ambient space and the area distortion is small. Manifold Laplacians may change dramatically in this case. Nevertheless, we show that the weighted graph Laplacians computed from two sets of points uniformly randomly sampled from M and its perturbed version N maintain a small distance in their spectra (i.e, sequence of eigenvalues). This distance can be bounded in terms of the perturbation and some intrinsic properties of the manifold M. Intuitively, the discrete graph Laplacians with parameters chosen appropriately remain oblivious to perturbations that are localized in the ambient space.

Related work. Several discrete Laplace operators have been developed to compute a Laplace-like operator either from a mesh approximating the hidden manifold [5, 10, 16], or more generally, simply from a set of points sampled from a hidden manifold [1, 6]. For high dimensional data, the most practical and also most popular discrete Laplace operator is perhaps the weighted graph Laplace operator – the reason being its simplicity and practicality. To compute it, one only needs to build the proximity graph from the input points, while other operators require a mesh structure either locally [6] or globally [5, 10, 16], which is expensive for high dimensional data. Furthermore, Belkin and Niyogi showed that, for points uniformly randomly sampled from the hidden manifold, the Gaussian-weighted graph Laplacian converges to the manifold Laplacian both pointwise [4] and in spectrum [3]. These results establish the theoretical foundation for the weighted graph Laplace operator, justifying its practical use.

Previous theoretical study of discrete Laplace operators

*Computer Science and Engineering, The Ohio State University. Emails: *tamaldey,ranjan,yusu@cse.ohio-state.edu*

focused on how well the discrete Laplacian and its spectrum approximate the manifold Laplacian and its spectrum in the limit [4, 5, 13, 19]. From a practical point of view, it is also important to understand the stability of a discrete operator to make sure that it is robust against noise in the sampled manifold. Stability of the so-called mesh-Laplacian [5] under perturbation was studied by Dey et al. [9]. Specifically, they showed that the mesh-Laplacians constructed from meshes approximating two manifolds M and N with a δ-diffeomorphism maintain a spectral distance that is bounded in terms of δ. (Roughly, a δ-diffeomorphism is a diffeomorphism between the two manifolds which maps a point to a nearby point and with small area distortion.) They also showed that as δ tends to zero, the spectrum of the manifold Laplacian of M converges to that of N.

Contributions. The previous stability study in [9] requires that the input manifold is approximated by a mesh structure. The perturbation considered there needs to preserve the manifold topology and be smooth in the first order. In some sense, such a perturbation model is necessary as [9] also bounds changes in the manifold Laplacians. In this paper, we aim to relax on both fronts. First, we focus on the Gaussian-weighted graph Laplacian as it can be built directly from a PCD sampled from the input manifold. Second, we introduce the δ-closeness between two manifolds which generalizes the perturbation model to allow for arbitrary though localized changes, including topological ones. See the figure above for an example where the right hand of the human may either touch or not touch the human body within the circled region, causing a potential topological difference. Since this relaxed perturbation model can alter geodesic distances dramatically, the manifold Laplacian may change significantly under it. However, we show that the weighted graph Laplacian under such perturbations remains stable.

More specifically, we consider two sets of points P and Q uniformly randomly sampled from two δ-close manifolds M and N. We provide an asymptotic upper bound on the distance between the spectra of the weighted graph Laplacians L_P^t and L_Q^t constructed from P and Q respectively. This bound tells us how the spectra distance depends on the size of the perturbation δ, and on the parameter t used to construct the graph Laplacians. This bound is proven for two *uniformly randomly* sampled sets of points. This choice is reasonable since (i) the theoretical guarantee of the weighted graph Laplacian is only established for such PCDs, and (ii) in practice, especially in high dimensional applications, input points are often sampled by certain random processes.

Our result is obtained by a novel way to establish a one-to-one correspondence between the two input sample point sets via the so-called "anchor-regions" that we will introduce later, and the Halls theorem, in the probabilistic setting. In Section 5, we provide further discussion of our results, especially the dependency on parameters involved. In Appendix D, we show some experiments that confirm our theoretical findings.

2 Problem Formulation

Weighted graph Laplace operator. The central object of this paper is the so-called Laplace-Beltrami operator and its discrete counterpart, the weighted graph Laplace operator. Let M be a smooth, compact k-dimensional manifold without boundary which is isometrically embedded in $\mathbb{R}^d$ and thus equipped with a natural Riemannian metric induced from $\mathbb{R}^d$. Given a twice continuously differential function $f \in C^2(\mathsf{M})$, let $\nabla_{\mathsf{M}} f$ denote the gradient vector field of f on M. The Laplace-Beltrami operator Δ_{M} is a second order differential operator defined as the divergence of the gradient, that is, $\Delta_{\mathsf{M}} f = div(\nabla_{\mathsf{M}} f)$. It is a fundamental geometric object encoding intrinsic geometric information about the manifold M.

We use $d_{\mathbb{R}^d}(x, y)$ to denote the Euclidean distance between two points $x, y \in \mathbb{R}^d$ and $d_{\mathsf{M}}(x, y)$ to denote the geodesic distance between x and y on M when x, y are on $\mathsf{M} \subseteq \mathbb{R}^d$. For simplicity, we replace $d_{\mathbb{R}^d}(x, y)$ with $\|x - y\|$ when it appears in the exponent.

Consider a set of discrete sample points $P = \{p_1, \ldots, p_n\} \subset \mathsf{M}$. The (Gaussian-)weighted graph Laplace operator L_P^t is a popular discrete analog of Δ_{M} for such input. Given a function $f : P \to \mathbb{R}$, it is defined w.r.t. a parameter $t > 0$ as

$$\text{(2.1)} \quad \mathsf{L}_P^t f(p_i) = \frac{1}{n} \cdot \frac{1}{(4\pi t)^{k/2} t} \sum_{j=1}^{n} e^{-\frac{\|p_i - p_j\|^2}{4t}} (f(p_i) - f(p_j)),$$

where $n = |P|$ is the number of sample points. Since a discrete function $f : P \to \mathbb{R}$ can be represented as an n-dimensional vector $[f(p_1), f(p_2), \ldots, f(p_n)]^T$, L_P^t is an $n \times n$ matrix where

$$\text{(2.2)} \quad \mathsf{L}_P^t[i][j] = \begin{cases} -\frac{1}{n} \cdot \frac{1}{(4\pi t)^{k/2} t} e^{-\frac{\|p_i - p_j\|^2}{4t}}, & \text{if } i \neq j \\ \frac{1}{n} \cdot \frac{1}{(4\pi t)^{k/2} t} \sum_{l \neq i, l \in [1,n]} e^{-\frac{\|p_i - p_l\|^2}{4t}}, & \text{if } i == j \end{cases}$$

It has been shown [3, 4] that if the set P samples M uniformly randomly according to the volume measure on M, then as n tends to infinity and t tends to 0 at appropriate rates, L_P^t converges to Δ_{M} both pointwise and in spectrum.

We study how robust this discrete Laplacian remains under perturbations of the manifold that points are sampled from. Specifically, we show that, for points uniformly randomly sampled from two "close" manifolds (to be defined shortly), the discrete Laplacians remain close as long as the choice of t is large enough. An explanation of our result is that t indicates the resolution at which we examine the hidden manifold M, and the discrete Laplacian L_P^t remains robust against small perturbations of M not captured at this resolution level.

δ-closeness. Given two smooth manifolds A and B embedded in $\mathbb{R}^d$, a map $f : A \to B$ is a δ-*diffeomorphism* if f is a diffeomorphism, $d_{\mathbb{R}^d}(x, f(x)) \leq \delta$, and $1 - \delta \leq Det(Jf_x) \leq 1 + \delta$ for any $x \in A$, where $Det(Jf_x)$ is the determinant of the Jacobian of the map f at x. Our inputs are two smooth compact k-manifolds M and N embedded in $\mathbb{R}^d$. Manifolds M and N are δ-*close* if there exists two sets of smooth k-dimensional closed submanifolds $\mathcal{X} = \{\mathsf{X}_1, \ldots, \mathsf{X}_m\}$, $\mathsf{X}_i \subset \mathsf{M}$, and $\mathcal{Y} = \{\mathsf{Y}_1, \ldots, \mathsf{Y}_m\}$, $\mathsf{Y}_i \subset \mathsf{N}$ where X_is (and Y_is) are pairwise disjoint and the following conditions hold.

(C1) There is a δ-diffeomorphism $\Phi : \mathsf{M} \setminus \mathcal{X} \to \mathsf{N} \setminus \mathcal{Y}$.

(C2) For any $i = 1, \ldots, m$, X_i and Y_i are contained within a *Euclidean* d-ball of radius δ and $(1-\delta)\mathsf{vol}(\mathsf{X}_i) \leq \mathsf{vol}(\mathsf{Y}_i) \leq (1+\delta)\mathsf{vol}(\mathsf{X}_i)$.

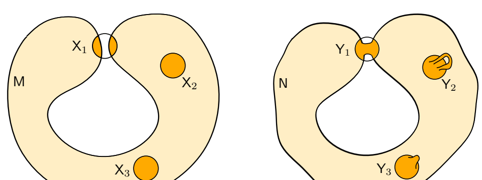

Figure 1: A manifold M and its perturbation N with anomalous regions $\{X_1, X_2, X_3\}$ and $\{Y_1, Y_2, Y_3\}$.

Intuitively, N is a perturbed version of M. The regions in $\mathcal{X}$ and $\mathcal{Y}$ are the *anomalous regions* where arbitrary changes, including topological changes, can happen. See Figure 1 for an illustration, where two types of topological changes happen in Y_1 and Y_2, respectively. Such arbitrary changes can be tolerated as long as they are restricted to a small Euclidean ball, and the volumes of corresponding anomalous regions are similar (i.e, condition C2).

High level framework. Given two δ-close manifolds M and N, let P and Q be two sets of n points, uniformly randomly sampled from M and N according to volume measures. We compute the weighted Graph Laplacians L_P^t and L_Q^t from P and Q respectively, for some parameter t. Our goal is to show that the spectrum of L_P^t is close to that of L_Q^t with high probability. We prove this via the following two steps.

Step 1. We show that, with high probability, there is a one-to-one correspondence $\psi : P \to Q$ such that $d_{\mathbb{R}^d}(p, \psi(p)) = O(\delta)$ for any $p \in P$.

Step 2. Based on the one-to-one correspondence obtained from Step 1, we show that the matrix norm $\|\mathsf{L}_P^t - \mathsf{L}_Q^t\|$ is bounded from above by a function $\mathcal{E}(\delta, t)$ which in turn gives an upper bound on the distance of the spectra for L_P^t and L_Q^t.

Our main result (stated below) follows from these two steps. Let $i(\mathsf{M})$ denote the injectivity radius of M, and μ be the isoperimetric constant of M, see Chavel [7] for definitions of these quantities.

THEOREM 2.1. *Let M be a smooth compact k-dimensional manifold embedded in $\mathbb{R}^d$ with $\mathsf{vol}(\mathsf{M}) = 1$, and N be its δ-close perturbation with $\delta < \min\{\frac{1}{8}, i(\mathsf{M})\}$. Let P and Q be two sets of points uniformly randomly sampled from M and N, respectively, with $|P| = |Q| = n = \Omega(\frac{1}{\delta^{4k}})$. Let L_P^t and L_Q^t be the corresponding Gaussian-weighted graph Laplacians computed from them, with the parameter $t = \Omega(\delta^{2-\varepsilon})$ for any $\varepsilon > 0$. With high probability (at least $1 - e^{-\Omega(n^{\frac{1}{4}})}$), the eigenvalues of L_P^t and L_Q^t satisfy $|\lambda_i(\mathsf{L}_P^t) - \lambda_i(\mathsf{L}_Q^t)| = O(\frac{\delta^{4/5}}{t^{k/2+7/5}})$ for any $i \in [1, n]$. In particular, $|\lambda_i(\mathsf{L}_P^t) - \lambda_i(\mathsf{L}_Q^t)| = O(\delta^{\frac{1}{3}})$ if $t \geq \delta^{\frac{1}{\frac{15}{14}k+3}}$. The big-O and big-$\Omega$ notations hide constants that depend solely on M.*

Isoperimetric constant. Our bounds depend on a concept called the isoperimetric constant. Given a k-dimensional compact Riemannian manifold A without boundary, the *(Cheeger) isoperimetric constant* $\mu(A)$ of A is defined as

$$\mu(A) = \inf_{\Gamma} \frac{\mathsf{vol}(\Gamma)}{\min\{\mathsf{vol}(A_1), \mathsf{vol}(A_2)\}},$$

where Γ ranges over all compact $(k-1)$-dimensional submanifolds of A that divide A into two parts A_1 and A_2. This implies that given any k-dimensional sub-manifold $R \subseteq A$ with piecewise-smooth boundary and with volume at most $\mathsf{vol}(A)/2$, we have $\mathsf{vol}(\partial R) \geq \mu(A)\mathsf{vol}(R)$. The Cheeger isoperimetric constant is an intrinsic quantity of the manifold A and it is closely related to the first non-zero eigenvalue of the Laplace-Beltrami operator of A [8].

3 Step 1: Correspondences

We are given two random samples P and Q, of size n each, from two δ-close manifolds M and N. The goal is to show that, with high probability, there exists a one-to-one correspondence $\psi : P \to Q$ such that corresponding pairs of points are close. To achieve this goal, we construct a bipartite graph $G = (V, E)$ where $V = P \cup Q$ and $E \subseteq P \times Q$ so that $d_{\mathbb{R}^d}(p, q) = O(\delta)$ for each edge $(p, q) \in E$. Given a node $p \in P$, let $Ng(p) \subseteq Q$ denote the set of neighbors of p in Q, and define $Ng(S) = \bigcup_{p \in S} Ng(p)$ for any subset $S \subseteq P$. We then argue that with high probability, we have

$|S| \leq |Ng(S)|$ for all subsets S of P. It then follows from Hall's Theorem that with high probability, there is a perfect bipartite matching of G, inducing a one-to-one correspondence $\psi : P \to Q$ with $d_{\mathbb{R}^d}(p, \psi(p)) = O(\delta)$ for any $p \in P$.

3.1 Bipartite graph construction

Our construction of the bipartite graph for points $P \cup Q$ uses "anchor-nodes" and "anchor-regions" as detailed below. First, an $(\varepsilon_1, \varepsilon_2)$-sample of a manifold M is a set of points $S \subset \mathsf{M}$ such that (i) for any point $x \in \mathsf{M}$, there is a sample point $s \in S$ within ε_1 geodesic distance away from x; and (ii) any two sample points $s_1, s_2 \in S$ are at least ε_2 geodesic distance apart. A set of *anchor-nodes* $\mathcal{A} = \{\mathsf{a}_1, \ldots, \mathsf{a}_r\}$ of M is simply a (δ, δ)-sample of M. It can be easily computed by a standard iterative procedure; see Appendix A. The following observation is straightforward. We include its proof in Appendix A for completeness.

OBSERVATION 3.1. *Assume that $\delta < i(\mathsf{M})$. For a (δ, δ)-sampling $\mathcal{A}$ of M, we have that $|\mathcal{A}| = O(1/\delta^k)$, where the big-O notation hides constants that depend on the intrinsic property of M.*

Anchor-regions. Consider an arbitrary subset of anchor-nodes $A \subseteq \mathcal{A}$. Let $d_{\mathsf{M}}(x, A)$ denote the smallest geodesic distance from x to any point in A. The *anchor-region* $\mathsf{R}_{\mathsf{M}}(A)$ *on* M *induced by* A is the set of points whose geodesic distance to A is at most δ; that is,

$$\mathsf{R}_{\mathsf{M}}(A) = \{x \in \mathsf{M} \mid d_{\mathsf{M}}(x, A) \leq \delta\}.$$

We call A the *defining subset* for $\mathsf{R}_{\mathsf{M}}(A)$. There are $2^{|\mathcal{A}|} = 2^{O(1/\delta^k)}$ number of anchor-regions on M, each defined by one subset of $\mathcal{A}$. Next, we define the set of anchor-regions on the manifold N. Each anchor-region $\mathsf{R}_{\mathsf{M}} = \mathsf{R}_{\mathsf{M}}(A)$ gives rise to one anchor-region $\mathsf{R}^+_{\mathsf{N}} = \mathsf{R}^+_{\mathsf{N}}(A)$ on N which is constructed via an intermediate region $\mathsf{R}^+_{\mathsf{M}}$. We refer to $\mathsf{R}^+_{\mathsf{N}}$ as the *witness anchor-region* of R_{M}. See Figure 2 for an illustration.

First, we construct an intermediate region $\mathsf{R}^+_{\mathsf{M}} \subseteq \mathsf{M}$, which contains all points from M within $\rho\delta$ Euclidean distance to R_{M}; that is,

$$\mathsf{R}^+_{\mathsf{M}} = \{x \in \mathsf{M} \mid d_{\mathbb{R}^d}(x, \mathsf{R}_{\mathsf{M}}) \leq \rho\delta\}.$$

The value of $\rho = O(1)$ will be specified shortly; it depends on the isoperimetric constant of the manifold M. Next, we "map" the region $\mathsf{R}^+_{\mathsf{M}}$ to $\mathsf{R}^+_{\mathsf{N}} \subseteq \mathsf{N}$. Specifically, recall that there is a δ-diffeomorphism $\Phi : \mathsf{M} \setminus \mathcal{X} \to \mathsf{N} \setminus \mathcal{Y}$. We set

$$\mathsf{R}^+_{\mathsf{N}} = \Phi(\mathsf{R}^+_{\mathsf{M}} \setminus \mathcal{X}) \cup \bigcup_{i \in I} \mathsf{Y}_i,$$

where $I = \{i \in [1, m] \mid \mathsf{X}_i \cap \mathsf{R}^+_{\mathsf{M}} \neq \emptyset\}$ is the set of indices of anomalous regions from M that intersect $\mathsf{R}^+_{\mathsf{M}}$. Intuitively, the witness anchor-region $\mathsf{R}^+_{\mathsf{N}}$ on N of an anchor-region $\mathsf{R}_{\mathsf{M}} \subseteq \mathsf{M}$ is obtained by thickening R_{M} by $\rho\delta$ Euclidean distance on M, and then map $\mathsf{R}^+_{\mathsf{M}}$ to $\mathsf{R}^+_{\mathsf{N}}$ on N. Since the diffeomorphism only exists between $\mathsf{M} \setminus \mathcal{X}$ and $\mathsf{N} \setminus \mathcal{X}$, we need to process the intersection of $\mathsf{R}^+_{\mathsf{M}}$ with anomalous regions separately when "mapping" $\mathsf{R}^+_{\mathsf{M}}$ onto N.

OBSERVATION 3.2. *(i) Given an anchor-region $\mathsf{R}_{\mathsf{M}}(A)$ induced by $A \subseteq \mathcal{A}$, its witness anchor-region in N satisfies that $\mathsf{R}^+_{\mathsf{N}}(A) = \bigcup_{\mathsf{a} \in A} \mathsf{R}^+_{\mathsf{N}}(\mathsf{a})$. (ii) Given an anchor-node $\mathsf{a} \in \mathcal{A}$, we have $d_{\mathbb{R}^d}(\mathsf{a}, y) \leq (\rho + 2)\delta$ for any point $y \in \mathsf{R}^+_{\mathsf{N}}(\mathsf{a})$.*

Graph construction. We now build a bipartite graph $G = (V, E)$ from input point sets $P \subset \mathsf{M}$ and $Q \subset \mathsf{N}$ as follows: The vertex set is $V = P \cup Q$. For each point $p \in P$, let $\mathsf{a}(p) \in \mathcal{A}$ denote the nearest neighbor of p (in terms of geodesic distance) in the set of anchor-nodes $\mathcal{A}$. We connect p to all points in Q falling inside the region $\mathsf{R}^+_{\mathsf{N}}(\mathsf{a}(p)) \subseteq \mathsf{N}$; that is, $Ng(p) = Q \cap \mathsf{R}^+_{\mathsf{N}}(\mathsf{a}(p))$ in G. For each point $q \in Ng(p)$, by Observation 3.2, $d_{\mathbb{R}^d}(\mathsf{a}(p), q) \leq (\rho + 2)\delta$. Furthermore, since $\mathcal{A}$ is a (δ, δ)-sample of M, we have that $d_{\mathsf{M}}(p, \mathsf{a}(p)) \leq \delta$. We thus have:

CLAIM 3.1. *For any edge (p, q) in G, $p \in P$, $q \in Q$, and $d_{\mathbb{R}^d}(p, q) \leq (\rho + 3)\delta = O(\delta)$.*

CLAIM 3.2. *Given any subset $S \subseteq P$, let $\mathcal{A}_S = \{\mathsf{a}(s) \in \mathcal{A} \mid s \in S\}$ be the union of nearest anchor-nodes to each point in S. Then $S \subseteq P \cap \mathsf{R}_{\mathsf{M}}(\mathcal{A}_S)$ and $Ng(S) = Q \cap \mathsf{R}^+_{\mathsf{N}}(\mathcal{A}_S)$.*

3.2 Bipartite Matching in G

Consider the graph $G = (V, E)$ as constructed above. We wish to show that, for a uniform random sample P of M and Q of N, there exists a perfect bipartite matching in G with high probability.

In what follows, we assume that M has volume 1. By the δ-closeness we have $(1 - \delta)\mathsf{vol}(\mathsf{M}) \leq \mathsf{vol}(\mathsf{N}) \leq (1 + \delta)\mathsf{vol}(\mathsf{M})$; thus $1 - \delta \leq \mathsf{vol}(\mathsf{N}) \leq 1 + \delta$. Let $\mu = \mu(\mathsf{M})$ be the isoperimetric constant for M. We assume that $\delta \leq \min\{\frac{1}{8}, i(\mathsf{M})\}$, and $\rho = \max\{1, \frac{8}{\mu}\}$. Note that by this choice of ρ, we have that $\rho\mu \geq 8$.

Given a region R, let $\#_X(R)$ denote the number of points from a point set X contained inside R. We prove the following key result later in this section.

LEMMA 3.1. *Given two δ-close compact and smooth k-manifolds M and N with $\mathsf{vol}(\mathsf{M}) = 1$ and $\delta < \min\{\frac{1}{8}, i(\mathsf{M})\}$, let P and Q be two sets of $n = \Omega(\frac{1}{\delta^{4k}})$ random samples from M and N respectively. Then, with probability at least $1 - e^{-\Omega(n^{\frac{1}{4}})}$, $\#_P(\mathsf{R}_{\mathsf{M}}) \leq \#_Q(\mathsf{R}^+_{\mathsf{N}})$ for all anchor-regions $\mathsf{R}_{\mathsf{M}} \subseteq \mathsf{M}$ and their witness anchor-regions $\mathsf{R}^+_{\mathsf{N}} \subseteq \mathsf{N}$.*

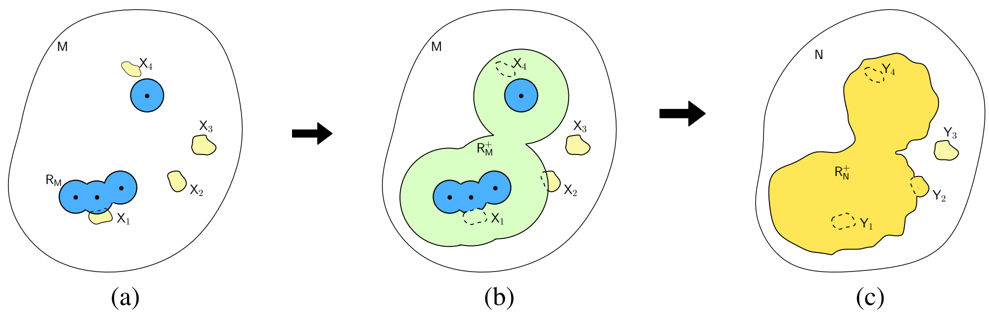

Figure 2: (a) Dark region is the anchor-region $\mathsf{R_M}$ induced by the black anchor-nodes (other anchor-nodes are not shown): $\mathsf{R_M}$ consists of points from M within δ geodesic distance to black points. Light regions are anomalous regions $\mathsf{X}_1, \ldots, \mathsf{X}_4$. (b) The intermediate region $\mathsf{R_M^+} \subset \mathsf{M}$ contains anomalous regions X_1 and X_4 fully, and X_2 partially. (c) The witness anchor-region $\mathsf{R_N^+} \subset \mathsf{N}$ of $\mathsf{R_M}$ includes anomalous regions Y_1, Y_2 and Y_4 fully.

Our main result in this section follows from the above lemma.

THEOREM 3.1. *Let* M *and* N *be two* δ*-close compact and smooth* k*-manifolds embedded in* $\mathbb{R}^d$ *with* $\mathsf{vol}(\mathsf{M}) = 1$*, and* $\delta < \min\{\frac{1}{8}, i(\mathsf{M})\}$*. Let* P *and* Q *be* n *uniform random samples of* M *and* N*, respectively. Then for large enough* $n = \Omega(\frac{1}{\delta^{4k}})$*, with probability higher than* $1 - e^{-\Omega(n^{\frac{1}{4}})}$*, there is a one-to-one correspondence* $\psi : P \to Q$ *such that for any* $p \in P$*,* $d_{\mathbb{R}^d}(p, \psi(p)) \leq (\rho + 3)\delta$ *where* $\rho = \max\{1, \frac{8}{\mu}\}$*.*

Proof. Consider the anchor-regions in M and their witness anchor-regions in N as described in Section 3.1. Then construct the bipartite graph $G = (V, E)$ as described earlier. Now consider the following two events:
(Event-1): $\#_P(\mathsf{R_M}) \leq \#_Q(\mathsf{R_N^+})$ for all anchor-regions $\mathsf{R_M} \subseteq \mathsf{M}$ and their witness anchor-regions $\mathsf{R_N^+} \subseteq \mathsf{N}$; and
(Event-2): $|S| \leq |Ng(S)|$ for all subsets $S \subseteq P$.

By Claim 3.2, for each subset $S \subseteq P$, there is always an anchor-region $\mathsf{R_M}$ such that $|S| \leq \#_P(\mathsf{R_M})$ and $Ng(S) = \#_Q(\mathsf{R_N^+})$. Hence Event-2 must happen if Event-1 happens. This means that $\mathrm{Prob}[\text{Event-1}] \leq \mathrm{Prob}[\text{Event-2}]$. By Lemma 3.1, we have that $\mathrm{Prob}[\text{Event-1}\ ] \geq 1 - e^{-\Omega(n^{\frac{1}{4}})}$, implying that $\mathrm{Prob}[\text{Event-2}] \geq 1 - e^{-\Omega(n^{\frac{1}{4}})}$.

If Event-2 happens, then by Hall's Theorem there is a perfect bipartite matching in the graph G. This provides a one-to-one correspondence $\psi : P \to Q$, where there is an edge in the bipartite graph G between each pair $(p, \psi(p))$. The theorem then follows from Claim 3.1. ∎

3.3 Lemma 3.1

It remains to prove Lemma 3.1. To this end, we first take an arbitrary but fixed anchor-region $\mathsf{R_M} \subseteq \mathsf{M}$ and its witness anchor-region $\mathsf{R_N^+} \subseteq \mathsf{N}$, and argue that $\#_{\mathsf{M}}(\mathsf{R_M}) \leq \#_{\mathsf{N}}(\mathsf{R_N^+})$ with high probability. Let $\mathsf{R_M^+}$ be the thickening of $\mathsf{R_M}$ by $\rho\delta$ in M as used in the construction of $\mathsf{R_N^+}$. We distinguish two cases depending on the volume of $\mathsf{R_M^+}$. Lemma 3.1 then follows from these results and the union-bound. Recall that $\delta \leq \min\{\frac{1}{8}, i(\mathsf{M})\}$ and $\rho = \max\{1, \frac{8}{\mu}\}$.

3.3.1 Case 1: $\mathsf{vol}(\mathsf{R_M^+}) \leq \frac{\mathsf{vol}(\mathsf{M})}{2}$.

First, we need a relation between the volumes of $\mathsf{R_M}$ and $\mathsf{R_N^+}$. To do so, we relate the volumes of $\mathsf{R_M^+}$ and $\mathsf{R_N^+}$, as well as those of $\mathsf{R_M}$ and $\mathsf{R_M^+}$.

CLAIM 3.3. $\mathsf{vol}(\mathsf{R_N^+}) \geq (1-\delta)\mathsf{vol}(\mathsf{R_M^+})$.

Proof. By construction, $\mathsf{R_N^+} = \Phi(\mathsf{R_M^+} \setminus \mathcal{X}) \cup \bigcup_{i \in I} \mathsf{Y}_i$, where $I = \{i \in [1, m] \mid \mathsf{X}_i \cap \mathsf{R_M^+} \neq \emptyset\}$ is the set of indices of those anomalous regions X_i intersecting $\mathsf{R_M^+}$. Since the map Φ is a δ-diffeomorphism, we have $\mathsf{vol}(\Phi(\mathsf{R_M^+} \setminus \mathcal{X})) \geq (1-\delta)\mathsf{vol}(\mathsf{R_M^+} \setminus \mathcal{X})$. On the other hand, by the δ-closeness between M and N, for each $i \in I$,

$$\mathsf{vol}(\mathsf{Y}_i) \geq (1-\delta)\mathsf{vol}(\mathsf{X}_i) \geq (1-\delta)\mathsf{vol}(\mathsf{R_M^+} \cap \mathsf{X}_i).$$

Putting these two together, we obtain the claim. ∎

CLAIM 3.4. *If* $\mathsf{vol}(\mathsf{R_M^+}) \leq \frac{\mathsf{vol}(\mathsf{M})}{2}$*, then* $\mathsf{vol}(\mathsf{R_M^+}) \geq (1 + 8\delta)\mathsf{vol}(\mathsf{R_M})$*.*

Proof. Let $\mathsf{R_M}(l)$ denote $\{x \in \mathsf{M} \mid d_{\mathsf{M}}(x, \mathsf{R_M}) \leq l\}$, that is, the region expanded from $\mathsf{R_M}$ by *geodesic distance* l. For $l \leq \rho\delta$, we have $\mathsf{R_M} \subseteq \mathsf{R_M}(l) \subseteq \mathsf{R_M}(\rho\delta) \subseteq \mathsf{R_M^+}$ as $\mathsf{R_M^+}$ contains all points within *Euclidean distance* $\rho\delta$ to $\mathsf{R_M}$. Let $\mathrm{bndVol}(l) = \mathsf{vol}(\partial\mathsf{R_M}(l))$. Since $\mathsf{vol}(\mathsf{R_M}(l)) \leq \mathsf{vol}(\mathsf{R_M^+}) \leq \mathsf{vol}(\mathsf{M})/2$, we have that $\mathrm{bndVol}(l) \geq \mu \cdot \mathsf{vol}(\mathsf{R_M}(l)) \geq \mu \cdot \mathsf{vol}(\mathsf{R_M})$. It then follows that

$$\begin{aligned}\mathsf{vol}(\mathsf{R_M^+}) - \mathsf{vol}(\mathsf{R_M}) &\geq \mathsf{vol}(\mathsf{R_M}(\rho\delta)) - \mathsf{vol}(\mathsf{R_M}) \\ &= \int_0^{\rho\delta} \mathrm{bndVol}(l)\,dl \\ &\geq \int_0^{\rho\delta} \mu \cdot \mathsf{vol}(\mathsf{R_M})\,dl = \rho\mu\delta \cdot \mathsf{vol}(\mathsf{R_M}).\end{aligned}$$

Since $\rho\mu \geq 8$, we have $\mathsf{vol}(\mathsf{R}_\mathsf{M}^+) - \mathsf{vol}(\mathsf{R}_\mathsf{M}) \geq 8\delta \cdot \mathsf{vol}(\mathsf{R}_\mathsf{M})$. The claim then follows. ■

Combining the above two claims we obtain the following corollary:

COROLLARY 3.1. *If* $\mathsf{vol}(\mathsf{R}_\mathsf{M}^+) \leq \frac{\mathsf{vol}(\mathsf{M})}{2}$, *then we have* $\mathsf{vol}(\mathsf{R}_\mathsf{N}^+) \geq (1+4\delta)\mathsf{vol}(\mathsf{R}_\mathsf{M})$.

LEMMA 3.2. *If* $\mathsf{vol}(\mathsf{R}_\mathsf{M}^+) \leq \frac{\mathsf{vol}(\mathsf{M})}{2}$, *we have* $\#_P(\mathsf{R}_\mathsf{M}) \leq \#_Q(\mathsf{R}_\mathsf{N}^+)$ *with probability at least* $1 - e^{-\Omega(\delta^{k+2}n)}$. *The big-*$\Omega$ *notation contains constants depending on the intrinsic property of manifold* M *only.*

Proof. Set $r = \mathsf{vol}(\mathsf{R}_\mathsf{M})$. Since P is a uniform random sample of M according to volume measure, $\#_P(\mathsf{R}_\mathsf{M})$ is a random variable and its expected value is rn. Furthermore, by construction, $\mathsf{vol}(\mathsf{R}_\mathsf{M})$ is at least the size of a geodesic ball with radius δ centered at some anchor-node $\mathsf{a} \in \mathcal{A}$. Hence for $\delta < i(\mathsf{M})$, $\mathsf{vol}(\mathsf{R}_\mathsf{M}) \geq C_\mathsf{M}\delta^k$, where k is the dimension of the manifold M, and C_M is a constant depending on the intrinsic curvature of M. It then follows from Chernoff bound (the upper tail) that

$$\mathrm{Prob}[\ \#_P(\mathsf{R}_\mathsf{M}) \geq (1+\delta)rn\ ] \leq e^{-rn\delta^2/4} \leq e^{-C_\mathsf{M}\delta^{k+2}n/4}.$$

On the other hand, the expected number of points from Q falling in R_N^+, i.e, the expected size of $\#_Q(\mathsf{R}_\mathsf{N}^+)$, is $n \cdot \mathsf{vol}(\mathsf{R}_\mathsf{N}^+)/\mathsf{vol}(\mathsf{N})$ which is at least $\frac{1+4\delta}{1+\delta}rn$ by Corollary 3.1 and the bound on $\mathsf{vol}(\mathsf{N})$. Note that since $\delta \leq \frac{1}{8}$, we have that $(1+\delta) \leq \frac{(1-\delta)(1+4\delta)}{(1+\delta)}$. Hence using Chernoff bound (the lower tail), we have that

$$\begin{aligned} & \mathrm{Prob}[\ \#_Q(\mathsf{R}_\mathsf{N}^+) \leq (1+\delta)rn\ ] \\ \leq\ & \mathrm{Prob}[\#_Q(\mathsf{R}_\mathsf{N}^+) \leq (1-\delta)\cdot\frac{(1+4\delta)}{(1+\delta)}\cdot rn] \\ \leq\ & \mathrm{Prob}[\#_Q(\mathsf{R}_\mathsf{N}^+) \leq (1-\delta)Exp[\#_Q(\mathsf{R}_\mathsf{N}^+)]\ ] \\ \leq\ & e^{-Exp[\#_Q(\mathsf{R}_\mathsf{N}^+)]\delta^2/2} = e^{-\Omega(\delta^{k+2}n)}. \end{aligned}$$

The claim then follows from these two inequalities and the union-bound. ■

3.3.2 Case 2: $\mathsf{vol}(\mathsf{R}_\mathsf{M}^+) > \mathsf{vol}(\mathsf{M})/2$.

This case is more complicated to handle than the previous case. First, the relation between $\mathsf{vol}(\mathsf{R}_\mathsf{M})$ and $\mathsf{vol}(\mathsf{R}_\mathsf{N}^+)$ as given in Corollary 3.1 is no longer true. Hence instead of relating the volumes of R_M and of R_N^+, we now need to relate the volumes of their complements in M and N, which are $\widehat{\mathsf{R}_\mathsf{M}} = \mathsf{M} \setminus \mathsf{R}_\mathsf{M}$ and $\widehat{\mathsf{R}_\mathsf{N}^+} = \mathsf{N} \setminus \mathsf{R}_\mathsf{N}^+$, respectively. In what follows, we first show the following lemma, the proof of which can be found in Appendix B. Let $\widehat{\mathsf{R}_\mathsf{M}^+} = \mathsf{M} \setminus \mathsf{R}_\mathsf{M}^+$ be the complement of the intermediate region R_M^+.

LEMMA 3.3. *If* $\mathsf{vol}(\mathsf{R}_\mathsf{M}^+) \geq \frac{\mathsf{vol}(\mathsf{M})}{2}$, *then* $\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{N}^+}) \leq \frac{\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}})}{1+2\delta}$.

The proof of the following observation is simple and is in Appendix C.

OBSERVATION 3.3. *If* $\widehat{\mathsf{R}_\mathsf{N}^+} \neq \emptyset$, *then* $\widehat{\mathsf{R}_\mathsf{M}}$ *contains at least a geodesic ball with radius* $\rho\delta$.

LEMMA 3.4. *If* $\mathsf{vol}(\mathsf{R}_\mathsf{M}^+) \geq \frac{\mathsf{vol}(\mathsf{M})}{2}$, *then* $\#_P(\mathsf{R}_\mathsf{M}) \leq \#_Q(\mathsf{R}_\mathsf{N}^+)$ *with probability at least* $1 - e^{-\Omega(\delta^{k+2}n)}$. *The big-*$\Omega$ *notation contains constants that depend on* M *only.*

Proof. First, if $\widehat{\mathsf{R}_\mathsf{N}^+} = \emptyset$, then the claim holds as $\#_Q(\mathsf{R}_\mathsf{N}^+) = n$. So from now on, we assume that $\widehat{\mathsf{R}_\mathsf{N}^+} \neq \emptyset$. In this case, it follows from Observation 3.3 that $\widehat{\mathsf{R}_\mathsf{M}}$ contains at least one geodesic ball with radius $\rho\delta$. Since $\rho \geq 1$, $\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}}) \geq C_\mathsf{M}\delta^k$ where C_M is a constant that depends on intrinsic curvature of the manifold M.

We now show that $\#_P(\widehat{\mathsf{R}_\mathsf{M}}) \geq \#_Q(\widehat{\mathsf{R}_\mathsf{N}^+})$ with high probability, which will imply the lemma. Since P is a uniform random sample of M with respect to the volume measure, we have that $Exp[\#_P(\widehat{\mathsf{R}_\mathsf{M}})] = \mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}}) \cdot n$. By Chernoff bound (the lower tail), we have

$$\begin{aligned} \text{(3.3)}\qquad & \mathrm{Prob}[\#_P(\widehat{\mathsf{R}_\mathsf{M}}) \leq (1-\tfrac{\delta}{4})\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}})n] \\ & \leq\ e^{-\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}})n\delta^2/32} \leq e^{-C_\mathsf{M}\delta^{k+2}n/32}. \end{aligned}$$

Next, we aim to bound $m = \#_Q(\widehat{\mathsf{R}_\mathsf{N}^+})$. Since Q is a uniform random sample of N, $Exp[m] = \frac{\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{N}^+})\cdot n}{\mathsf{vol}(\mathsf{N})}$ which gives $\frac{\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{N}^+})\cdot n}{1-\delta} \geq Exp[m] \geq \frac{\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{N}^+})\cdot n}{1+\delta}$. We first assume that $\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{N}^+}) \geq \frac{C_\mathsf{M}\delta^k}{1+2\delta}$. In this case,

$$Exp[m] \geq \frac{C_\mathsf{M}\delta^k n}{(1+\delta)(1+2\delta)} \geq C_\mathsf{M}\delta^k n/2.$$

Chernoff bound (the upper tail) provides:

$$\begin{aligned} \text{(3.4)}\quad & \mathrm{Prob}[m \geq (1-\frac{\delta}{4})\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}})n] \\ & \leq\ \mathrm{Prob}[\ m \geq (1-\frac{\delta}{4})(1+2\delta)\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{N}^+})n] \\ & \leq\ \mathrm{Prob}[\ m \geq (1-\frac{\delta}{4})(1+2\delta)(1-\delta)Exp[m]\ ] \\ & \leq\ \mathrm{Prob}[\ m \geq (1+\frac{\delta}{4})Exp[m]\ ] \\ & \leq\ e^{-\delta^2 Exp[m]/64}\ \leq\ e^{-\Omega(\delta^{k+2}n)}. \end{aligned}$$

Combining Eqn (3.4) with Eqn (3.3), we have that $\#_Q(\widehat{\mathsf{R}_\mathsf{N}^+}) \leq \#_P(\widehat{\mathsf{R}_\mathsf{M}})$ and hence $\#_P(\mathsf{R}_\mathsf{M}) \leq \#_Q(\mathsf{R}_\mathsf{N}^+)$ with probability at least $1 - e^{-\Omega(\delta^{k+2}n)}$ when $\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{N}^+}) \geq \frac{C_\mathsf{M}\delta^k}{1+2\delta}$.

What remains is to prove our lemma for the case when $\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{N}^+}) < \frac{C_\mathsf{M}\delta^k}{1+2\delta}$. In this case, we cannot directly apply the previous argument, because we cannot bound $Exp[m]$ from below. However, it turns out that we can use a different quantity $s \geq m$ which we will define shortly, and show that with high probability s is smaller than $\#_P(\widehat{\mathsf{R}_\mathsf{M}})$. It follows that $m \leq s \leq \#_P(\widehat{\mathsf{R}_\mathsf{M}})$ with high probability, which proves the lemma.

Specifically, consider any region R' such that $\widehat{\mathsf{R}_\mathsf{N}^+} \subset R'$ and $\mathsf{vol}(R') = \frac{C_\mathsf{M}\delta^k}{1+2\delta}$. Set $s := \#_Q(R')$. Since R' contains $\widehat{\mathsf{R}_\mathsf{N}^+}$, we have that $s \geq m$. The expected value of s is

$$Exp[s] = \frac{\mathsf{vol}(R')n}{\mathsf{vol}(\mathsf{N})} \leq \frac{C_\mathsf{M}\delta^k n}{(1+2\delta)(1-\delta)} \leq \frac{\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}})n}{(1+2\delta)(1-\delta)}.$$

We thus have:

$$\begin{aligned}
(3.5)\qquad & \mathrm{Prob}[m \geq (1-\frac{\delta}{4})\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}})n] \\
&\leq \mathrm{Prob}[s \geq (1-\frac{\delta}{4})\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}})n] \\
&\leq \mathrm{Prob}[\, s \geq (1-\frac{\delta}{4})(1+2\delta)(1-\delta)Exp[s]\,] \\
&\leq \mathrm{Prob}[\, s \geq (1+\frac{\delta}{4})Exp[s]\,] \\
&\leq e^{-\delta^2 Exp[s]/64} \leq e^{-\Omega(\delta^{k+2}n)}.
\end{aligned}$$

The last inequality holds as $Exp[s] = \frac{\mathsf{vol}(R')n}{\mathsf{vol}(\mathsf{N})} \geq C_\mathsf{M}\delta^k n/2$. Finally, combining Eqn (3.5) with Eqn (3.3), we have that $m \leq \#_P(\widehat{\mathsf{R}_\mathsf{M}})$ with probability at least $1 - e^{-\Omega(\delta^{k+2}n)}$ for the case when $\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{N}^+}) < \frac{C_\mathsf{M}\delta^k}{1+2\delta}$ as well. ■

Proof of Lemma 3.1. By Observation 3.1, there are $2^{\Theta(1/\delta^k)}$ number of anchor-regions in M and in N. By Lemma 3.2 and 3.4, for each anchor-region $\mathsf{R}_\mathsf{M} \subset \mathsf{M}$ and its witness anchor-region $\mathsf{R}_\mathsf{N}^+ \subset \mathsf{N}$, $\#_P(\mathsf{R}_\mathsf{M}) \leq \#_Q(\mathsf{R}_\mathsf{N}^+)$ with probability at least $1 - e^{-\Omega(\delta^{k+2}n)}$. It follows from union bound that $\#_P(\mathsf{R}_\mathsf{M}) \leq \#_Q(\mathsf{R}_\mathsf{N}^+)$ for all anchor-region R_M simultaneously with probability at least

$$\begin{aligned}
1 - e^{-\Omega(\delta^{k+2}n)} \cdot 2^{\Theta(\frac{1}{\delta^k})} &\geq 1 - e^{-\Omega(\delta^{k+2}n)+\Theta(\frac{1}{\delta^k})} \\
&\geq 1 - e^{-\Omega(n^{\frac{1}{4}}\cdot\frac{\delta^{k+2}}{\delta^{3k}})+\Theta(\frac{1}{\delta^k})} \geq 1 - e^{-\Omega(n^{\frac{1}{4}})+\Theta(\frac{1}{\delta^k})} \\
&= 1 - e^{-\Omega(n^{\frac{1}{4}})};
\end{aligned}$$

for sufficiently large n, say $n = \Omega(\frac{1}{\delta^{4k}})$. This proves Lemma 3.1.

4 Step 2: Bounding Spectra Distance

We now assume that we are given two sets of n points $P = \{p_1, \ldots, p_n\}$ and $Q = \{q_1, \ldots, q_n\}$ sampled from hidden k-manifolds M and N, respectively, such that $d_{\mathbb{R}^d}(p_i, q_i) = \|p_i - q_i\| \leq (\rho+3)\delta = O(\delta)$ for any $i \in [1,n]$, where $\rho = \max\{1, \frac{8}{\mu}\}$ is a constant depending on the isoperimetric constant μ of the manifold M. Notice that Theorem 3.1 implies that we have such inputs with high probability when P and Q are uniformly randomly sampled from M and N according to their volume measures. Now consider the weighted graph Laplacians L_P^t and L_Q^t constructed from P and Q, respectively. Our goal is to show that the spectra of these two graph Laplacians are close. We achieve this by showing that the matrix 2-norm $\|\mathsf{L}_P^t - \mathsf{L}_Q^t\|$ of the matrix $\mathsf{L}_P^t - \mathsf{L}_Q^t$ is bounded, which further bounds the spectra distance of L_P^t and L_Q^t by Weyl's Theorem for eigenvalue perturbations [15].

For simplicity, set $G_{ij} := \frac{1}{(4\pi)^{\frac{k}{2}} t^{\frac{k}{2}+1}} \cdot e^{\frac{-\|p_i-p_j\|^2}{4t}}$ and $\tilde{G}_{ij} := \frac{1}{(4\pi)^{\frac{k}{2}} t^{\frac{k}{2}+1}} \cdot e^{\frac{-\|q_i-q_j\|^2}{4t}}$ for any $i,j \in [1,n]$. Notice that $\mathsf{L}_P^t[i][j] = -\frac{1}{n}G_{ij}$ and $\mathsf{L}_Q^t[i][j] = -\frac{1}{n}\tilde{G}_{ij}$ for $i \neq j$.

First, we need the following key result. This lemma bounds $|G_{ij} - \tilde{G}_{ij}|$, which is then used to bound $\|\mathsf{L}_P^t - \mathsf{L}_Q^t\|$.

LEMMA 4.1. *$|G_{ij} - \tilde{G}_{ij}| = O(\frac{\delta^{4/5}}{t^{k/2+7/5}})$ if $t = \Omega(\delta^{2-\varepsilon})$ for any positive real $\varepsilon > 0$. In particular, $|G_{ij} - \tilde{G}_{ij}| = O(\delta^{1/3})$ for $t \geq \delta^{\frac{1}{\frac{15}{14}k+3}}$. The big-O and big-$\Omega$ notations hide constants depending on the isoperimetric constant μ of the manifold* M.

Proof. For simplicity, denote $C_\mu := \rho + 3$. By the one-to-one closeness between points in P and in Q, we have that $d_{\mathbb{R}^d}(p_i, q_i) = \|p_i - q_i\| \leq C_\mu\delta$ for constant $C_\mu \geq 6$. For every $i,j \in [1,n]$, the triangle inequality imply that $\|p_i - p_j\| - 2C_\mu\delta \leq \|q_i - q_j\| \leq \|p_i - p_j\| + 2C_\mu\delta$.

We distinguish two cases: $\|p_i - p_j\| \leq \tau$ and $\|p_i - p_j\| > \tau$ for some parameter $\tau \geq 4C_\mu\delta$ whose value is to be specified later.

Case 1: $\|p_i - p_j\| \leq \tau$.

Since $\tau \geq 4C_\mu\delta$, we have that $\delta\tau = \Omega(\delta^2)$. We can then bound $\tilde{G}_{ij}$ from below as follows as long as $\frac{{C_\mu}^2\delta^2}{t} < 1$ and $\frac{C_\mu\delta\tau}{t} < 1$:

$$\begin{aligned}
\tilde{G}_{ij} &= \frac{1}{(4\pi)^{\frac{k}{2}} t^{\frac{k}{2}+1}} \cdot e^{\frac{-(\|q_i-q_j\|)^2}{4t}} \\
&\geq \frac{1}{(4\pi)^{\frac{k}{2}} t^{\frac{k}{2}+1}} \cdot e^{\frac{-(\|p_i-p_j\|+2C_\mu\delta)^2}{4t}} \\
&\geq G_{ij} \cdot e^{\frac{-{C_\mu}^2\delta^2}{t}} \cdot e^{\frac{-C_\mu\delta\|p_i-p_j\|}{t}} \\
&\geq G_{ij} \cdot e^{\frac{-{C_\mu}^2\delta^2}{t}} \cdot e^{\frac{-C_\mu\delta\tau}{t}} \\
&\geq G_{ij}(1 - \frac{O(\delta^2)}{t})(1 - \frac{O(\delta\tau)}{t}) \geq G_{ij} \cdot (1 - \frac{O(\delta\tau)}{t}).
\end{aligned}$$

Now we bound $\tilde{G}_{ij}$ from above. First, assume that $\tau \geq \|p_i - p_j\| \geq 2C_\mu\delta$. In this case we have that $\|q_i - q_j\| \geq \|p_i - p_j\| - 2C_\mu\delta \geq 0$, implying that $\|q_i - q_j\|^2 \geq (\|p_i - p_j\| - 2C_\mu\delta)^2$. Hence for $\frac{C_\mu\delta\tau}{t} < 1$,

$$\begin{aligned}\tilde{G}_{ij} &\leq \frac{1}{(4\pi)^{\frac{k}{2}} t^{\frac{k}{2}+1}} \cdot e^{\frac{-(\|p_i-p_j\|-2C_\mu\delta)^2}{4t}} \\ &\leq G_{ij} \cdot e^{\frac{-C_\mu{}^2\delta^2}{t}} \cdot e^{\frac{C_\mu\delta\|p_i-p_j\|}{t}} \leq G_{ij} \cdot e^{\frac{C_\mu\delta\|p_i-p_j\|}{t}} \\ &\leq G_{ij} \cdot (1 + O(\frac{\delta\tau}{t})).\end{aligned}$$

Otherwise, $\|p_i - p_j\| \leq 2C_\mu\delta$. Since $e^{-\frac{\|q_i-q_j\|^2}{4t}} \leq 1$, we have $\tilde{G}_{ij} \leq \frac{1}{(4\pi)^{\frac{k}{2}} t^{\frac{k}{2}+1}}$. It then follows that, for $\frac{C_\mu{}^2\delta^2}{t} < 1$, we have

$$\begin{aligned}\tilde{G}_{ij} &\leq \frac{1}{(4\pi)^{\frac{k}{2}} t^{\frac{k}{2}+1}} = G_{ij}/e^{\frac{-(\|p_i-p_j\|)^2}{4t}} \\ &= G_{ij} \cdot e^{\frac{(\|p_i-p_j\|)^2}{4t}} \leq G_{ij} \cdot e^{\frac{C_\mu{}^2\delta^2}{t}} \leq G_{ij} \cdot (1 + O(\frac{\delta^2}{t})).\end{aligned}$$

Putting the above inequalities together, and using the fact that $G_{ij} = O(\frac{1}{t^{k/2+1}})$, we have

$$\begin{aligned}E_\leq &:= |G_{ij} - \tilde{G}_{ij}| = G_{ij} \cdot O(\frac{\delta\tau}{t}) = O(\delta\tau/t^{\frac{k+4}{2}}) \\ &\quad \text{when } \|p_i - p_j\| \leq \tau.\end{aligned}$$

Case 2: $\|p_i - p_j\| > \tau$.

Recall that $\|q_i - q_j\| \geq \|p_i - p_j\| - 2C_\mu\delta$. If $\|p_i - p_j\| > \tau$, and $\tau \geq 4C_\mu\delta$, then $\|q_i - q_j\| > \tau - \tau/2 > \tau/2$. It then follows

$$\tilde{G}_{ij} = \frac{1}{(4\pi)^{\frac{k}{2}} t^{\frac{k}{2}+1}} \cdot e^{\frac{-\|q_i-q_j\|^2}{4t}} \leq \frac{1}{(4\pi)^{\frac{k}{2}} t^{\frac{k}{2}+1}} \cdot e^{\frac{-\tau^2}{16t}}.$$

Similarly, $G_{ij} \leq \frac{1}{(4\pi)^{\frac{k}{2}} t^{\frac{k}{2}+1}} \cdot e^{\frac{-\tau^2}{4t}}$ as $\|p_i - p_j\| > \tau$. Since $e^{-\frac{1}{x}} \leq x^2$ for any $x > 0$, we have

$$\begin{aligned}E_> &:= |\tilde{G}_{ij} - G_{ij}| \leq \max\{G_{ij}, \tilde{G}_{ij}\} \\ &\leq \frac{1}{(4\pi)^{\frac{k}{2}} t^{\frac{k}{2}+1}} \cdot O((\frac{t}{\tau^2})^2) = O(\frac{1}{\tau^4 t^{\frac{k}{2}-1}}) \\ &\quad \text{when } \|p_i - p_j\| > \tau.\end{aligned}$$

We balance the two error terms $E_\leq$ and $E_>$ by choosing $\tau = \frac{t^{3/5}}{\delta^{1/5}}$ so that $E_\leq = E_> = \frac{\delta^{4/5}}{t^{k/2+7/5}}$. The condition $\frac{C_\mu{}^2\delta^2}{t} < 1$, $\frac{C_\mu\delta\tau}{t} < 1$, and $\tau \geq 4C_\mu\delta$ can be satisfied as long as $t = \Omega(\delta^{2-\varepsilon})$ for any $\varepsilon > 0$. Finally, if $t > \delta^{\frac{1}{\frac{15k}{14}+3}}$, we have that $E_\leq = E_> = O(\delta^{1/3})$. The lemma follows. ■

Given a matrix (operator) D, let $\lambda_i(D)$ denote its i-th smallest eigenvalue. We have the following result.

THEOREM 4.1. *Let $P = \{p_1, \ldots, p_n\}$ and $Q = \{q_1, \ldots, q_n\}$ be two sets of n points such that $\|p_i - q_i\| = O(\delta)$ for every $i \in [1, n]$. Let L_P^t and L_Q^t be the corresponding Gaussian-weighted graph Laplacians computed from P and Q, respectively. The eigenvalues of L_P^t and L_Q^t satisfy $|\lambda_i(\mathsf{L}_P^t) - \lambda_i(\mathsf{L}_Q^t)| = O(\mathcal{E}(\delta, t))$, where $\mathcal{E}(\delta, t) = \frac{\delta^{4/5}}{t^{k/2+7/5}}$. In particular, $\mathcal{E}(\delta, t) = O(\delta^{\frac{1}{3}})$ if $t \geq \delta^{\frac{1}{\frac{15}{14}k+3}}$.*

Proof. First, notice that the Gaussian-weighted graph Laplace matrix is symmetric. Set matrix D to be the difference between L_P^t and L_Q^t; that is $D[i][j] = \mathsf{L}_P^t[i][j] - \mathsf{L}_Q^t[i][j]$ for all $i, j \in [1, n]$. Consider the definition of $\mathsf{L}_P^t[i][j]$ in Eqn (2.2). By Lemma 4.1,

$$|D[i][j]| = \frac{1}{n} \cdot |G_{ij} - \tilde{G}_{ij}| = O(\mathcal{E}(\delta, t)/n), \quad \text{for each } \ i \neq j.$$

For diagonal entries, we have

$$|D[i][i]| \leq \sum_{j=1}^{n} \frac{1}{n} \cdot |G_{ij} - \tilde{G}_{ij}| = O(\mathcal{E}(\delta, t)), \quad i \in [1, n].$$

Hence the (induced) matrix 1-norm of D satisfies

$$\|D\|_1 = \max_{j=1}^{n} \sum_{i=1}^{n} |D[i][j]| = O(\mathcal{E}(\delta, t)),$$

and the matrix ∞-norm of D satisfies

$$\|D\|_\infty = \max_{i=1}^{n} \sum_{j=1}^{n} |D[i][j]| = O(\mathcal{E}(\delta, t)).$$

Since the matrix 2-norm $\|D\|_2$ satisfies $\|D\|_2 \leq \sqrt{\|D\|_1 \|D\|_\infty}$ (see e.g., [11]), we have that $\|D\|_2 = O(\mathcal{E}(\delta, t))$. By Weyl's theorem for eigenvalue perturbation of Hermitian matrices (see e.g., [15]), the distance between corresponding eigenvalues of the matrix L_Q^t and $\mathsf{L}_P^t = \mathsf{L}_Q^t + D$ is bounded by the matrix 2-norm of the difference matrix D, which is $O(\mathcal{E}(\delta, t))$. ■

Our main result, Theorem 2.1 stated earlier, then follows from Theorems 3.1 and 4.1.

5 Discussions

We now provide further discussions of our results to elucidate some subtle points. Experimental results that substantiate our theory are presented in Appendix D.

Dependency of perturbation bound. Consider the bound on spectrum perturbation $\mathcal{E}(\delta, t) = \frac{\delta^{4/5}}{t^{k/2+7/5}}$. Notice that, $\delta^{\frac{4}{5}}$ and $\delta^{\frac{1}{3}}$ are larger than δ for $\delta < 1$. Hence, the bound on the spectrum perturbation is larger than the perturbation of the underlying manifold when t is small. Furthermore, the spectrum perturbation bound $\mathcal{E}(\delta, t)$ decreases as t increases. Intuitively, this is because large t has the effect of smoothing out entries in the weighted graph Laplacian as $\sqrt{t}$ is the bandwidth of the Gaussian weight. Hence, the discrete Laplacian (and its spectrum) becomes less sensitive to the differences in $\|p_i - p_j\|$ and $\|q_i - q_j\|$ for $i, j \in [1, n]$, that is, less sensitive to the manifold perturbation, for large t value. In the limit, as t tends to infinity, the weighted graph Laplacians L_P^t and L_Q^t both become identity matrix (up to scaling), and the spectra difference becomes $\mathcal{E}(\delta, t) = 0$.

Dependency on intrinsic properties of M. We note that, the choice of parameters depends only on M, instead of its perturbed version N. This is desirable as one would expect the manifold M typically to be much nicer than its perturbed version. The constants in big-O and big-Ω notations in our results depend on two quantities intrinsic to M: (i) the constant C_{M} where $C_{\mathsf{M}}\delta^k$ is the lower-bound on the volume of any geodesic ball of radius $\delta < i(\mathsf{M})$ on M; and (ii) the isoperimetric constant $\mu = \mu(\mathsf{M})$.

Bounding manifold Laplacians of M and N? It has been shown in [3, 4] that as n tends to infinity, and t tends to 0 at an appropriate rate, the weighted graph Laplacian converges to manifold Laplacian both pointwise and in spectrum. One may wonder if Theorem 2.1 combined with these results implies that the spectrum distance between manifold Laplacians Δ_{M} and Δ_{N} is also bounded via the spectrum distance between their discrete approximations L_P^t and L_Q^t. The answer is negative. Roughly speaking, the reason is that the choice of δ and t can be such that they are small enough with respect to the intrinsic property of M (so that results from [3, 4] can apply), but not of N. In fact, under the δ-close perturbation that we allow, there is no reason to believe that manifold Laplacians for M and N are close because the geodesic distances (the intrinsic metric) can be altered significantly for large parts of the manifolds.

Future directions. In this paper we introduced the notion of δ-closeness for analysis of localized topological perturbation for a given manifold M. It will be interesting to explore whether one can achieve similar results for even more general perturbation models such as perturbing M to N within bounded Wasserstein distance; or allowing P and Q to be two uniform samples drawn from the same "thickening" of a hidden manifold.

Our current argument requires that the two sets of points P and Q have the same cardinality. It would be interesting to see whether this constraint can be removed. Furthermore, right now, we bound the distance between all eigenvalues of L_P^t and L_Q^t. Can better error bounds be obtained if we consider only the first few eigenvalues? Finally, it will be interesting to also provide stability results for eigenfunctions of the weighted graph Laplacians under perturbation.

Acknowledgment. The authors would like to thank anonymous reviewers for helpful comments. This work is supported by National Science Foundation under grants CCF-0747082 and CCF-1116258.

References

[1] M. Belkin and P. Niyogi. Laplacian Eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[2] M. Belkin and P. Niyogi. Semi-supervised learning on riemannian manifolds. *Machine Learning*, 56(1-3):209–239, 2004.

[3] M. Belkin and P. Niyogi. Convergence of laplacian eigenmaps. In *NIPS*, pages 129–136, 2006.

[4] M. Belkin and P. Niyogi. Towards a theoretical foundation for laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289–1308, 2008.

[5] M. Belkin, J. Sun, and Y. Wang. Discrete Laplace operator on meshed surfaces. In *Proc. 24st Annu. ACM Sympos. Comput. Geom.*, pages 278–287, New York, NY, USA, 2008. ACM.

[6] M. Belkin, J. Sun, and Y. Wang. Constructing Laplace operator from point clouds in $\mathbb{R}^d$. In *Proc. 20th ACM-SIAM Sympos. Discrete Algorithms*, pages 1031–1040, 2009.

[7] I. Chavel. *Riemannian Geometry: A Modern Introduction*. Cambridge University Press, second edition, 2006.

[8] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. *Problems in Analysis, A Symposium in Honor of Salomon Bochner*, pages 195–199, 1970.

[9] T. K. Dey, P. Ranjan, and Y. Wang. Convergence, stability, and discrete approximation of Laplace spectra. In *Proc. 21st ACM-SIAM Sympos. Discrete Algorithms*, pages 650–663, 2010.

[10] J. Dodziuk. Finite-difference approach to the hodge theory of harmonic forms. *American Journal of Mathematics*, 98(1):79–104, 1978.

[11] G. H. Golub and C. F. van Van Loan. *Matrix Computations*. Johns Hopkins Studies in Mathematical Sciences. The Johns Hopkins University Press, 3rd edition, 1996.

[12] M. Hein and M. Maier. Manifold denoising. In *NIPS*, pages 561–568, 2006.

[13] K. Hildebrandt, K. Polthier, and M. Wardetzky. On the convergence of metric and geometric properties of polyhedral surfaces. *Geometriae Dedicata*, 123(1):89–112, December 2006.

[14] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17:395–416, December 2007.

[15] Y. Nakatsukasa. Absolute and relative Weyl theorems for generalized eigenvalue problems. *Linear Algebra and its Applications*, 432(1):242–248, 2010.

[16] U. Pinkall and K. Polthier. Computing discrete minimal surfaces and their conjugates. *Experimental Mathematics*, 2(1):15–36, 1993.

[17] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000.

[18] O. Sorkine. Differential representations for mesh processing. *Computer Graphics Forum*, 25(4):789–807, 2006.

[19] M. Wardetzky. *Discrete Differential Operators on Polyhedral Surfaces – Convergence and Approximation*. PhD thesis, Freie Universität Berlin, 2006.

[20] H. Zhang, O. van Kaick, and R. Dyer. Spectral mesh processing. *Comp. Gra. Forum*, 29(6):1865–1894, 2010.

A Constructing Anchor-Nodes

Constructing an (δ,δ)-sample. We can compute a set of anchor-nodes $\mathcal{A} = \{\mathsf{a}_1,\ldots,\mathsf{a}_r\} \subset \mathsf{M}$, which is simply a (δ,δ)-sample of M, using the following standard iterative procedure.

Initialize $\mathcal{A}$ with an arbitrary point $\mathsf{a}_1 \in \mathsf{M}$. In the i-th round after constructing $\mathcal{A}_{i-1} := \{\mathsf{a}_1,\ldots,\mathsf{a}_{i-1}\}$, identify the point from M that is furtherest away, in terms of geodesic distance, from points in $\mathcal{A}_{i-1}$ and set it as a_i if this distance is at least δ. We stop when this distance is smaller than δ. The procedure creates a (δ,δ)-sample. Indeed, no point in M can be further than δ away from its nearest neighbor in $\mathcal{A}$ (otherwise, the procedure will continue), and by construction, no two a_i and a_j with $j < i$ are within δ distance since a_j is at least δ away from a_i.

Proof of Observation 3.1. Compute the geodesic Voronoi diagram of $\mathcal{A}$ on M. We claim that the Voronoi cell $\mathrm{Vor}(\mathsf{a}_i)$ for each point $\mathsf{a}_i \in \mathcal{A}$ contains the geodesic ball of radius $\delta/2$ centered at a_i. Indeed, if this is not the case, then there exists some point y on the boundary of $\mathrm{Vor}(\mathsf{a}_i)$ such that $d_{\mathsf{M}}(y,\mathsf{a}_i) < \delta/2$. Since y is on the boundary of $\mathrm{Vor}(\mathsf{a}_i)$, there is another site, say $\mathsf{a}_j \in \mathcal{A}$ such that $d_{\mathsf{M}}(y,\mathsf{a}_j) = d_{\mathsf{M}}(y,\mathsf{a}_i)$ and thus $d_{\mathsf{M}}(\mathsf{a}_i,\mathsf{a}_j) < \delta$ by triangle inequality. This contradicts the fact that $\mathcal{A}$ is a (δ,δ)-sample. The volume of a geodesic ball with radius $\delta/2$ is $C_M\delta^k$ where C_M is a constant that depends on the intrinsic curvature of M and $\delta/2$ is smaller than the injectivity radius. It follows that $\mathsf{vol}(\mathrm{Vor}(\mathsf{a}_i)) = \Omega(\delta^k)$ implying $|\mathcal{A}| = O(1/\delta^k)$ by a packing argument.

B Proof of Lemma 3.3

Here, the aim is to obtain a relation between the volumes of $\widehat{\mathsf{R}_\mathsf{M}}$ and $\widehat{\mathsf{R}^+_\mathsf{N}}$. To do so, we first relate the volumes of $\widehat{\mathsf{R}^+_\mathsf{M}}$ and $\widehat{\mathsf{R}^+_\mathsf{N}}$, and then those of $\widehat{\mathsf{R}_\mathsf{M}}$ and $\widehat{\mathsf{R}^+_\mathsf{M}}$.

Claim B.1. $\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{N}}) \le (1+\delta)\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}})$.

Proof. Set $I = \{i \in [1,m] \mid \mathsf{X}_i \cap \mathsf{R}^+_\mathsf{M} \neq \emptyset\}$ to be the set of indices of anomalous regions from M intersecting the anchor-region R^+_M. Given the δ-diffeomorphism Φ, it is easy to see that

$$\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{N}} \setminus \mathcal{Y}) = \mathsf{vol}(\Phi(\widehat{\mathsf{R}^+_\mathsf{M}} \setminus \mathcal{X})) \le (1+\delta)\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}} \setminus \mathcal{X}).$$

Furthermore, since $\mathsf{vol}(\mathsf{R}^+_\mathsf{M}) \le \mathsf{vol}(\mathsf{R}^+_\mathsf{M} \setminus \mathcal{X}) + \bigcup_{i\in I}\mathsf{vol}(\mathsf{X}_i)$, we have that $\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}}) \ge \mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}} \setminus \mathcal{X}) + \bigcup_{i\notin I}\mathsf{vol}(\mathsf{X}_i)$. It then follows that

$$\begin{aligned}\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{N}}) &= \mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{N}} \setminus \mathcal{Y}) + \bigcup_{i\notin I}\mathsf{vol}(\mathsf{Y}_i) \\ &\le (1+\delta)\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}} \setminus \mathcal{X}) + (1+\delta)\bigcup_{i\notin I}\mathsf{vol}(\mathsf{X}_i) \\ &\le (1+\delta)\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}}).\end{aligned}$$

■

Claim B.2. *If* $\mathsf{vol}(\mathsf{R}^+_\mathsf{M}) \ge \frac{\mathsf{vol}(\mathsf{M})}{2}$*, then* $\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}}) \le \frac{\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}})}{(1+4\delta)}$.

Proof. Since $\mathsf{vol}(\mathsf{R}^+_\mathsf{M}) \ge \mathsf{vol}(\mathsf{M})/2$, we have that $\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}}) \le \mathsf{vol}(\mathsf{M})/2$. First, suppose $\mathsf{vol}(\mathsf{R}_\mathsf{M}) \le (1-4\delta)\mathsf{vol}(\mathsf{M})/2$. The claim then follows in this case as we have:

$$\begin{aligned}\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}}) &\ge \mathsf{vol}(\mathsf{M}) - \frac{(1-4\delta)\mathsf{vol}(\mathsf{M})}{2} \ge \frac{(1+4\delta)\mathsf{vol}(\mathsf{M})}{2} \\ &\ge (1+4\delta)\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}}).\end{aligned}$$

Hence we now consider the remaining case where $\mathsf{vol}(\mathsf{R}_\mathsf{M}) \ge (1-4\delta)\cdot\frac{\mathsf{vol}(\mathsf{M})}{2}$. Similar to Section 3.3.1, let $\mathsf{R}_\mathsf{M}(l) := \{x \in \mathsf{M} \mid d_\mathsf{M}(x,\mathsf{R}_\mathsf{M}) \le l\}$. Set $\widehat{\mathsf{R}_\mathsf{M}}(l) = \mathsf{M} \setminus \mathsf{R}_\mathsf{M}(l)$, which is the region shrunk from $\widehat{\mathsf{R}_\mathsf{M}}$ by geodesic distance l. Since R^+_M includes all points within $\rho\delta$ Euclidean distance to R_M, we have $\mathsf{R}_\mathsf{M} \subseteq \mathsf{R}_\mathsf{M}(l) \subseteq \mathsf{R}_\mathsf{M}(\rho\delta) \subseteq \mathsf{R}^+_\mathsf{M}$ for any $l \le \rho\delta$, implying that

$$\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}}) \le \mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}}(\rho\delta)) \le \mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}}(l)) \le \mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}}).$$

Let $\mathrm{bndVol}(l)$ denote the volume of the boundary of $\mathsf{R}_\mathsf{M}(l)$ which is also the volume of the boundary of $\widehat{\mathsf{R}_\mathsf{M}}(l)$; that is, $\mathrm{bndVol}(l) = \mathsf{vol}(\partial\mathsf{R}_\mathsf{M}(l)) = \mathsf{vol}(\partial\widehat{\mathsf{R}_\mathsf{M}}(l))$. If $\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}}(l)) \le \frac{\mathsf{vol}(\mathsf{M})}{2}$, then we have

$$\mathrm{bndVol}(l) \ge \mu\cdot\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}}(l)) \ge \mu\cdot\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}}). \quad \text{(B.1)}$$

If $\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}}(l)) \ge \frac{\mathsf{vol}(\mathsf{M})}{2}$, then we have that $\mathrm{bndVol}(l) \ge \mu\cdot\mathsf{vol}(\mathsf{R}_\mathsf{M}(l)) \ge \mu\cdot\mathsf{vol}(\mathsf{R}_\mathsf{M})$. Since we have assumed that $\mathsf{vol}(\mathsf{R}_\mathsf{M}) \ge (1-4\delta)\cdot\frac{\mathsf{vol}(\mathsf{M})}{2}$, this implies that

$$\begin{aligned}\mathrm{bndVol}(l) &\ge \mu(1-4\delta)\cdot\mathsf{vol}(\mathsf{M})/2 \\ &\ge \mu(1-4\delta)\cdot\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}}).\end{aligned} \quad \text{(B.2)}$$

Putting Eqn (B.1) and (B.2) together, and observing that $\rho\mu \ge 8$ and $\delta \le \frac{1}{8}$, we have

$$\begin{aligned}\mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}}) - \mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}}) &\ge \mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}}) - \mathsf{vol}(\widehat{\mathsf{R}_\mathsf{M}}(\rho\delta)) \\ &= \int_0^{\rho\delta}\mathrm{bndVol}(l)dl \\ &\ge \rho\delta\mu(1-4\delta)\cdot\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}}) \ge 4\delta\cdot\mathsf{vol}(\widehat{\mathsf{R}^+_\mathsf{M}}).\end{aligned}$$

The claim then follows. ■

Combining the above two claims, the lemma follows.

C Proof of Observation 3.3

Let $\tilde{u}$ be an arbitrary point in $\widehat{\mathsf{R}_\mathsf{N}^+} \subseteq \mathsf{N}$. We now identify a "pre-image" u of $\tilde{u}$ in M. If $\tilde{u} \notin \mathcal{Y}$ then simply set $u = \Phi^{-1}(\tilde{u}) \in \widehat{\mathsf{R}_\mathsf{M}^+}$ to be the pre-image of $\tilde{u}$ under the diffeomorphism $\Phi : \mathsf{M} \setminus \mathcal{X} \to \mathsf{N} \setminus \mathcal{Y}$. Otherwise, assume that $\tilde{u} \in \mathsf{Y}_j$. Then we pick an arbitrary point from X_j as u; note, $\mathsf{X}_j \cap \mathsf{R}_\mathsf{M}^+ = \emptyset$ as $\tilde{u} \notin \mathsf{R}_\mathsf{N}^+$. Hence in either case, the "pre-image" u falls in the region $\widehat{\mathsf{R}_\mathsf{M}^+}$. Therefore its nearest Euclidean and geodesic distance to R_M is at least $\rho\delta$, implying the claim.

D Experiments

In this section, we show through experiments that the discrete (weighted-graph) Laplace operator is indeed stable against small topological changes. We also show that the weighted-graph Laplace operator changes smoothly with the size of the region of topological change for a fixed t. In our experiments, we use the first 300 eigenvalues.

First, we show the effect of a small topological change on the eigenvalue, and how this effect changes with the size of the region of change. Figure 3 shows two tori connected with a very thin bridge that is enlarged successively. The graph depicts the top 300 eigenvalues at different stages of this process for $t^2 = 0.00001$. We can see that for smaller changes, the eigenvalues are similar, since the region of change is small when compared to our choice of t. However, for a large change, the eigenvalues expectedly deviate more from the original surface.

Figure 4 shows another example of a torus with another small torus attached to it incrementing its genus, and hence changing its topology. Then, we gradually increase the size of this attached torus until it becomes as large as the original torus. Here, we fix t to be much larger ($t^2 = 0.001$). Note that the eigenvalues still differ a lot since the region of change is large in all the cases.

Finally, figure 5 shows the effect of multiple topological changes for the armadillo model. We fix $t^2 = 0.0001$. Here, the eigenvalues are similar despite multiple topological changes in different regions of the model. For a large change (top-right model), however, the shift in eigenvalues is noticeable.

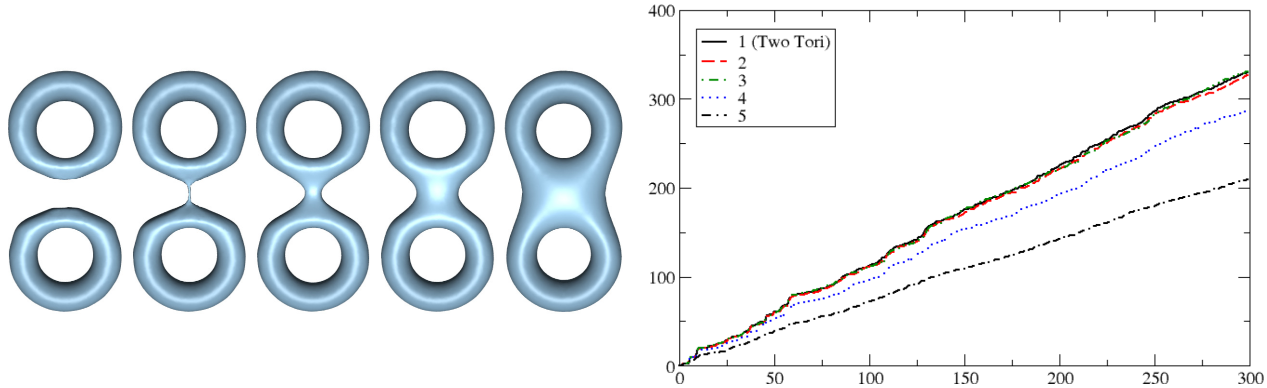

Figure 3: Left: Connecting two tori with increasingly larger bridges. Right: Comparison of eigenvalues

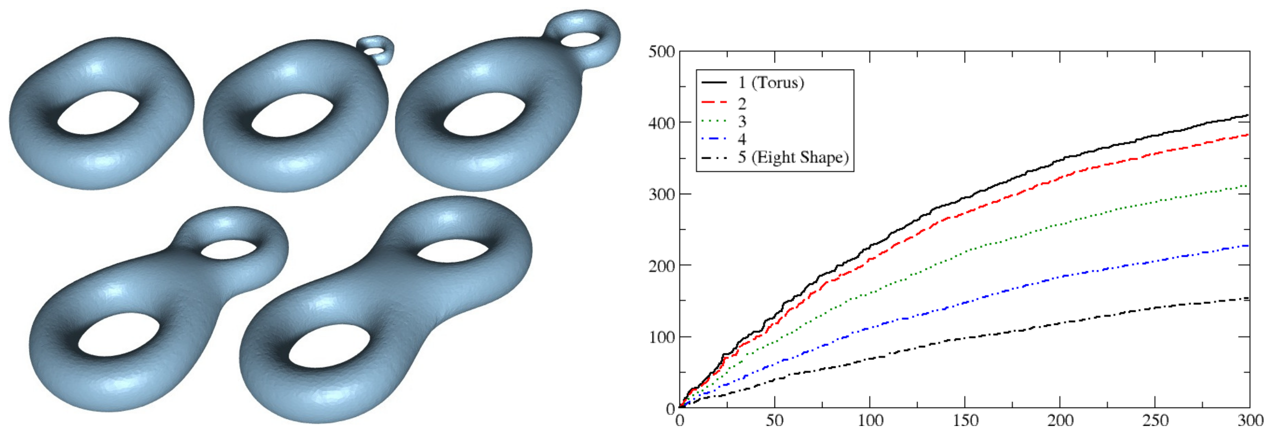

Figure 4: Left: Increasingly large region of topological change on a torus. Right: Comparison of eigenvalues

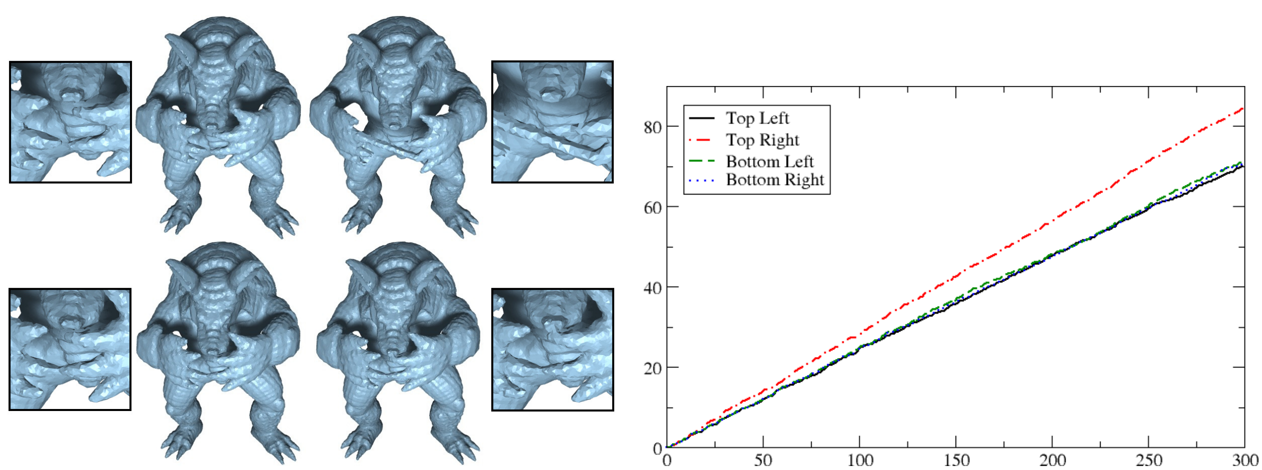

Figure 5: Topological changes on the Armadillo. Top Row: Original Armadillo model; and a variation with two fingers joined. Bottom Row: Another variation with two fingers touching; and a model with two fingers touching and another finger touching the nose. Right: Comparison of eigenvalues for $t^2 = 0.0001$