

Toward understanding complex spaces: graph Laplacians on manifolds with singularities and boundaries

Mikhail Belkin* Qichao Que* Yusu Wang* Xueyuan Zhou†

Abstract

In manifold learning, algorithms based on graph Laplacian constructed from data have received considerable attention both in practical applications and theoretical analysis. Much of the existing work has been done under the assumption that the data is sampled from a manifold without boundaries and singularities or that the functions of interest are evaluated away from such points.

At the same time, it can be argued that singularities and boundaries are an important aspect of the geometry of realistic data. Boundaries occur whenever the process generating data has a bounding constraint; while singularities appear when two different manifolds intersect or if a process undergoes a “phase transition”, changing non-smoothly as a function of a parameter.

In this paper we consider the behavior of graph Laplacians at points at or near boundaries and two main types of other singularities: *intersections*, where different manifolds come together and sharp “*edges*”, where a manifold sharply changes direction. We show that the behavior of graph Laplacian near these singularities is quite different from that in the interior of the manifolds. In fact, a phenomenon somewhat reminiscent of the Gibbs effect in the analysis of Fourier series, can be observed in the behavior of graph Laplacian near such points. Unlike in the interior of the domain, where graph Laplacian converges to the Laplace-Beltrami operator, near singularities graph Laplacian tends to a first-order differential operator, which exhibits different scaling behavior as a function of the kernel width. One important implication is that while points near the singularities occupy only a small part of the total volume, the difference in scaling results in a disproportionately large contribution to the total behavior. Another significant finding is that while the scaling behavior of the operator is the same near different types of singularities, they are very distinct at a more refined level of analysis.

We believe that a comprehensive understanding of these structures in addition to the standard case of a smooth manifold can take us a long way toward better methods for analysis of complex non-linear data and can lead to significant progress in algorithm design.

1 Introduction

Dealing with high-dimensional non-linear data is one of the key challenges in modern data analysis. In recent years a class of methods based on the mathematical notion of a manifold has become popular in machine learning, starting with the papers [24, 26]. The underlying intuition is that a process with a small number of parameters will generate a low-dimensional surface in the potentially very high-dimensional space of features and that this situation is ubiquitous in real-world data. This idea is captured nicely by the technical notion of a smooth embedded Riemannian manifold, which provides the first realistic model for general non-linear data. Still it does not reflect certain important aspects of real data, which can be mathematically understood as singularities and boundaries.

*Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43210, USA. Email: mbelkin, que, yusu@cse.ohio-state.edu

†Department of Computer Science, University of Chicago, USA. Email: zhouxy@uchicago.edu

The most basic and, arguably, most important singularity in real data is an *intersection*, where two different manifolds come together in a region of space. This often happens in classification where two classes with presumably different structure can give rise to similar objects (consider, e.g., the similarity of MNIST digits “7” and “1” on the right). Another important type of singularity is a (co-dimension one) “*edge*”, which is ubiquitous in computer graphics (think of an edge of the surface of a table) and happens whenever the behavior of an underlying process changes rapidly beyond a certain point (a “phase transition”). Finally, *boundaries* occurs when there are bounding constraints on the underlying process (think how poses of the human body are constrained by the range of motion of the joints) or on the representation in the ambient space (e.g., pixels cannot have negative intensity).



In this paper we provide an analysis of these three cases for Laplacian-based learning algorithms. It turns out that all of these singularities result in a behavior near the singularity markedly different from that inside the manifold, a phenomenon somewhat reminiscent of the Gibb’s effect in Fourier analysis. In particular, the scaling of the operator is different near singularities. As a result of this scaling behavior, singularities cannot be ignored, even globally, despite the fact that relatively few points are located near them, since each of these points contributes disproportionately to the total operator. We also find that at a finer level of analysis these singularities have quite different effects, which are discussed below.

We believe that a comprehensive understanding of these structures in addition to the standard case of a smooth manifold can take us a long way toward better methods for analysis of complex non-linear data and can lead to significant progress in algorithm design.

Related work. Methods based on graph Laplacians constructed from data have gained acceptance for a range of inference tasks including clustering [28], semi-supervised learning (e.g. [7, 29]) and dimensionality reduction [3], as well as others. An analysis in [3] provided a mathematical framework for many of these methods by connecting graph Laplacian constructed from data to the Laplace-Beltrami operator of the underlying manifold based on the relation of the Laplacian and the heat equation. That analysis was extended and generalized in [20, 18, 11, 4, 25, 13, 19, 5] providing a detailed understanding of graph Laplacians obtained from manifold data. While boundary effects have been studied in non-parametric kernel smoothing (see, e.g., [16, Chapter 4.4]), they have not been considered in the Laplacian-related literature, although we note that boundary behavior for the infinite data case can be derived from the Taylor series expansion in Lemma 9 of [11]. To the best of our knowledge, other singularities are not considered in that literature. (In fact, the reason causing the different scaling behavior for intersection and edge types of singularities is somewhat different from that for the boundary case.)

In related developments, a considerable amount of recent work has been aimed at understanding the case of intersecting linear spaces using various techniques from algebraic geometry to spectral clustering (e.g., [27, 10, 21]) both in terms of algorithms and theoretical guarantees. In our setting, this is a special case of a singular manifold with intersection singularities. Another related work is [14], where the authors analyze a model based on a mixture of manifolds in the context of semi-supervised learning.

We also note that there has been much recent interest in reconstructing topological invariants of manifolds and other spaces (see, e.g. [8, 9, 22, 23]). The line of work most related to our results is on learning stratified spaces, where multiple submanifolds (strata) are “glued” nicely together [1, 6, 17], which provides an even more general model for a singular space.

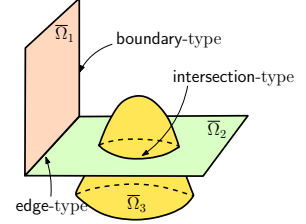
Summary of results. Consider the (appropriately scaled, see Section 2) graph Laplacian $L_{n,t}$ constructed from n data points, using the Gaussian kernel with bandwidth t . It can be shown (see the references above) that as t tends to zero and n tends to infinity at an appropriate rate, $L_{n,t}$ converges to the Laplace-Beltrami operator Δ (a second order differential operator) inside the manifold. In this paper, we show that for singular manifolds $L_{n,t}$ exhibits a very different behavior near the singularity set. Specifically, for a sufficiently

smooth function f and a small t , within distance \sqrt{t} of the singularity set, $L_{n,t}f$ is approximated by $\frac{1}{\sqrt{t}}D$, where D is some first order derivative operator that will be described explicitly. The difference in scaling becomes crucial when t is small as the $\frac{1}{\sqrt{t}}D$ term becomes dominant. To simplify the discussion, we will first assume that the data is infinite, and thus $L_{n,t}f$ can be replaced by $L_t f$. Finite sample bounds and rates will be discussed later. The types of singularities considered in the paper are as follows:

Boundary. At a point x near manifold boundary, L_t can be approximated by $\frac{1}{\sqrt{t}}\phi(\frac{R}{\sqrt{t}})\partial_{\mathbf{n}}$, where R is the distance from x to the boundary and \mathbf{n} is the unit vector toward the nearest boundary point (outward normal at the boundary). $\partial_{\mathbf{n}}$ denotes the directional derivative, while $\phi(z)$ is a scalar function of the form $\phi(z) = Ce^{-z^2}$.

Intersection and edge. The situation is more complicated for other types of singularities, where two different manifolds Ω_1 and Ω_2 intersect or come together.

Given a point $x_1 \in \Omega_1$ consider its projection x_2 onto Ω_2 and its nearest neighbor x_0 in the singularity. Similarly to the boundary case, let \mathbf{n}_1 and \mathbf{n}_2 , be the directions to x_0 from x_1 and x_2 respectively and let R_1 and R_2 be the corresponding distances. $L_t f(x_1)$ can be approximated by $\frac{1}{\sqrt{t}}\phi_1(\frac{R_1}{\sqrt{t}})\partial_{\mathbf{n}_1} f(x_0) + \frac{1}{\sqrt{t}}\phi_2(\frac{R_2}{\sqrt{t}})\partial_{\mathbf{n}_2} f(x_0)$. Importantly, the form of the scalar functions ϕ_1, ϕ_2 is different for different types of singularities: for intersection-type singularity, we have: $\phi_i(z) = A_i z e^{-Cz^2}$, while for edge-type we have¹: $\phi_i(z) = A_i z e^{-Cz^2} + B_i e^{-Cz^2}$. Here the coefficients A_i, B_i and C depend on the angle between the manifolds and can be written explicitly.

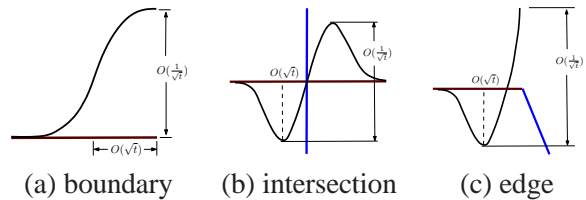


Significant points and observations:

1. Scaling. Within $O(\sqrt{t})$ distance of the singularity, L_t is dominated by a differential operator different from the Laplace-Beltrami operator. Away from the singularity, the higher-valued $(1/\sqrt{t})$ term fades and L_t becomes an approximation to the Laplace-Beltrami operator Δ .

2. Contributions of singular points. Let the dimension of the manifold be d . The volume of points within \sqrt{t} of the singularity is approximately $\sqrt{t} \text{vol}_{d-1}(S)$, where $\text{vol}_{d-1}(S)$ is $d - 1$ dimensional volume of the singular points, assuming that the set of singular points S is $d - 1$ dimensional (co-dimension one), which is always the case with the boundary and “edge” singularities (and may or may not be the case with intersection). Fix a function f . As $t \rightarrow 0$, the integral $L_t f$ over the points within \sqrt{t} of the singularity does not disappear and, in fact, will converge to a constant, depending on the function and the $d - 1$ dimensional volume of the singularity. Moreover, because of the operator scaling, the contribution of those near-singular points to the L_2 norm will tend to infinity. We remark that, while different from a technical point of view, this phenomenon is reminiscent to the Gibbs effect in Fourier series, where effects of discontinuity do not vanish as approximation becomes more and more precise.

3. Shape near the boundary. While the $\frac{1}{\sqrt{t}}$ scaling is common for all singularities, the type of singularity is reflected in the shape of the function ϕ , which is very different for different types of singularities. Somewhat over-simplifying matters, one can think of $\phi = e^{-z^2}$ for the boundary case, $\phi = ze^{-z^2}$ for the intersection singularity and $\phi = e^{-z^2} + ze^{-z^2}$ for the edge (see figures above). These differences lead to quite distinctive patterns, when the operator is applied to a fixed function. For example, the directional derivative terms disappear at the intersection-type singular points. In fact for these points L_t converges to the usual Laplace-Beltrami operator. However, at points within \sqrt{t} distance of these points we expect to see $L_t f$ take both positive and negative values of magnitude \sqrt{t} . We think that this difference may be a key to future algorithms.



¹For compactness, we are simplifying ϕ 's for edge singularities by removing some higher order terms (see Section 3.3).

4. Rates and finite sample bounds. Convergence rates for t as well as finite sample bounds are provided in Section 4. Interestingly, and perhaps counter-intuitively, the scaling implies that fewer data points are required to ensure that $L_{t,n}$ is an accurate (relative) estimate of L_t near the boundary than inside the domain.

5. Impact on eigenfunctions. Eigenfunctions of the graph Laplacian associated to data are used in a large number of applications. While the question of convergence is very subtle (see [5] for the proof in the case of a smooth manifold), our results for fixed functions provide an indication of the expected answer assuming the convergence holds. We provide a brief discussion of it in Section 5.

It turns out that the boundary case automatically leads to Neumann boundary conditions (even though the operator does not "see" the boundary explicitly). Quite surprisingly, the edge singularity appears to have no effect on the limit behavior of eigenfunctions, at least if the singular manifold is isometric to a smooth one (think of bending a sheet of paper). The case of the intersection is more complex, but intersections of co-dimension two or more seem to have no effect on the limit behavior of eigenfunctions (in L_2 norm).

Some experimental results based on our theoretical findings can be found in Appendix C.

2 Problem Setup

Let $\overline{\Omega}$ be the set of k smooth compact Riemannian submanifolds of intrinsic dimension d embedded in \mathbb{R}^N . For each smooth component $\overline{\Omega}_i$, let Ω_i denote its interior and $\partial\Omega_i$ its boundary of Ω_i . We assume the boundaries $\partial\Omega_i$ satisfy the *necessary smoothness conditions*². Let Ω be the union of Ω_i , and $\partial\Omega$ be the union of $\partial\Omega_i$. $\overline{\Omega}_i$ and $\overline{\Omega}_j$ may not be disjoint, and we will consider the following two ways that they can associate with each other: they intersect in their interior (i.e., $\Omega_i \cap \Omega_j \neq \emptyset$) or they are "glued" along (part of) their boundaries. More precisely, given a point $x \in \overline{\Omega}$, we say that x is *regular* if it falls in the interior of *exactly one* submanifold Ω_i ; that is, $x \in \Omega_i, x \notin \Omega_j, j \neq i$. Otherwise, x is a *singular point*. There are several possible singularities, and this paper considers the following *three general classes*³.

1. *boundary-type*: Boundary points that belong to exactly one smooth submanifold Ω_i : that is, $x \in \partial\Omega_i, x \notin \overline{\Omega}_j, j \neq i$;
2. *intersection-type*: Points at the intersection of two smooth submanifolds. For simplicity of exposition, we assume that the intersection happens between the interior of the submanifolds; that is, $x \in \Omega_i \cap \Omega_j, i \neq j$;
3. *edge-type*: Boundary points that belong to the boundaries of two smooth submanifolds. That is, $x \in \partial\Omega_i \cap \partial\Omega_j, i \neq j$.

For technical reasons, we will assume that the set of singular points is also a smooth manifold of lower dimension, at least locally. Moreover, for edge and intersection singularities the tangent space at each point of the singular manifold is the intersection of the tangent spaces to each "piece". We will call a union of smooth manifolds $\overline{\Omega}$ with the above types of singularities a *singular manifold*. Notice that for a regular point x , the manifold is smooth at x , and the tangent space at x is homeomorphic to \mathbb{R}^d .

For a singular manifold $\overline{\Omega}$, we consider the following *piecewise-smooth function* $f : \overline{\Omega} \mapsto \mathbb{R}$. In particular, set $f_i := f|_{\overline{\Omega}_i}$, $i = 1, \dots, k$, be f restricted to the submanifold $\overline{\Omega}_i$. We require that each f_i is C^2 -continuous on interior points. let $p(x)$ be a piecewise smooth probability density function on $\overline{\Omega}$ and we assume that $0 < a \leq p_i(x) \leq b < \infty$.

Graph Laplacian. Given n random samples $X = \{X_1, \dots, X_n\}$ drawn i.i.d. from a distribution with density $p(x)$ on $\overline{\Omega}$, we can build a weighted graph $G(V, E)$ by mapping each sample point X_i to vertex v_i

²We require that the derivatives of the first two orders of the transition map exist, i.e., \mathcal{M} is a C^2 -manifold, see [12], page 50.

³There are other cases, which can be considered as a mixture of these three general cases.

and assigning a weight w_{ij} to edge e_{ij} . One typical weight function is the Gaussian, which is used in this paper and defined as follows:

$$w_{n,t}(X_i, X_j) = \frac{1}{nt} K_t(X_i, X_j) = \frac{1}{n} \frac{1}{t^{d/2+1}} e^{-\frac{\|X_i - X_j\|_{\mathbb{R}^N}^2}{t}}.$$

Notice that in this Gaussian weight function, the Euclidean distance is used. The normalization by $\frac{1}{nt^{d/2+1}}$ is for the convenience of limit analysis.

Let $W_{n,t}$ be the edge weight matrix of graph G with $W_{n,t}(i, j) = w_t(X_i, X_j)$, and $D_{n,t}$ be a diagonal matrix such that $D_{n,t}(i, i) = \sum_j w_t(X_i, X_j)$, then the unnormalized graph Laplacian is defined as the $n \times n$ matrix $L_{n,t} = D_{n,t} - W_{n,t}$. Now for a fixed smooth function $f(x)$, and any point $x \in \overline{\Omega}$, the graph Laplacian applied to this function f is thus

$$L_{n,t}f(x) = \frac{1}{nt} \sum_{j=1}^n K_t(x, X_j)[f(x) - f(X_j)].$$

Limit behavior of graph Laplacian. The limit study of graph Laplacians primarily involves the limits of two parameters, sample size n and weight function bandwidth t . As n increases, one typically decreases t to let the graph Laplacian capture progressively a finer local structure. With a proper rate as a function of n and t , the limit of $L_{n,t}$ and its various aspects at regular points, including the finite sample analysis, are studied in [2, 20, 18, 25, 13, 19, 4, 5]. The basic result is that the limit of $L_{n,t}f(x)$ at a regular point x is (up to a constant) $L_{n,t}f(x) \xrightarrow{p} -p(x)\Delta_{p^2}f(x) = -p(x)\{\frac{1}{p^2}\text{div}[p^2\text{grad}f(x)]\}$, where Δ_{p^2} is called the *weighted Laplacian*, see [15]. See [19] for the limit analysis of other versions of graph Laplacians at regular points.

3 Limit Analysis of Graph Laplacian on Singular Manifolds

The limit analysis of graph Laplacians typically involves limits of two parameters, the sample size n and the (Gaussian) kernel bandwidth t . This is usually done in two steps, first analyzing the limit $t \rightarrow 0$ for infinite data and then obtaining finite sample results and rates as $n \rightarrow \infty$ using concentration inequalities.

In this section we analyze the behavior of the infinite graph Laplacian (i.e., $L_{n,t}$, $n = \infty$), $L_t f(x)$, when x is on or near a singular point, t is small and the function $f : \overline{\Omega} \rightarrow \mathbb{R}$, is fixed. Finite sample results and rates are given in Section 4.

For a fixed t we define L_t as the limit of $L_{n,t}$ as the amount of data tends to infinity:

$$L_t f(x) = L_{\infty,t} f(x) = \mathbb{E}_{p(X)}[L_{n,t} f(x)] f(x) = \frac{1}{t} \int_{\overline{\Omega}} K_t(x, y)(f(x) - f(y))p(y)dy. \quad (1)$$

Local coordinate system: The above integral is defined on the manifold $\overline{\Omega}$. In order to study the behavior of this integral for a small t , we need to introduce a local coordinate systems at x . The most convenient coordinate system for our purposes⁴ is obtained by a local projection from the manifold to its tangent space $\pi_x : \overline{\Omega} \rightarrow T_x$. This projection is one-to-one and smooth for points sufficiently close to x . Hence its inverse, π_x^{-1} exists as well. Note that on an intersection or edge singularity the projection is onto the union of tangent spaces. However, it is still one-to-one as long as we restrict the projection of points from each piece to their respective tangent space.

In what follows, we often use change of variables to convert an integral over a manifold to an integral over the tangent space at a specific point. The following bounds are used throughout the paper. For two points $x, y \in \overline{\Omega}_i$, let $y' = \pi_x(y)$ be the projection of y in the tangent space T_x of $\overline{\Omega}_i$ at x . Let $J\pi_x|_y$ (resp.

⁴We note that another possibility is to use the coordinate system given by the exponential map (as in [2, 11]). This has an advantage of being independent of the embedding, but requires more subtle analysis in the presence of singularities.

$J\pi^{-1}|_{y'}$) denote the Jacobian of the map π_x at point $y \in \bar{\Omega}_i$ (resp. of the inverse map π_x^{-1} at $y' \in T_x$). For y sufficiently close to x , we have (e.g., [22])

$$\|x - y\| = \|x - y'\| + O(\|x - y'\|^3) \Rightarrow \|x - y\|^2 = \|x - y'\|^2 + O(\|x - y'\|^4), \quad (2)$$

$$|J\pi_x|_y - 1| = O(\|x - y\|^2) \text{ and } |J\pi^{-1}|_{y'} - 1| = O(\|x - y'\|^2). \quad (3)$$

3.1 Boundary-type: Manifold with Boundary

We first analyze the limit of $L_t f(x)$ when x is near a *smooth* boundary of a single smooth manifold $\bar{\Omega}$ as the results for the boundary case are simpler to state.

Theorem 1. *Let $f \in C^2(\bar{\Omega})$, $p(x) \in C^\infty(\bar{\Omega})$ with $0 < a \leq p(x) \leq b < \infty$, and $\partial\Omega$ be the smooth boundary of Ω . Given a point x near the boundary, let x_0 be its nearest neighbor on the boundary $\partial\Omega$, and \mathbf{n} the inward normal direction of Ω at x_0 . Put $\|x - x_0\| = r\sqrt{t}$. For t sufficiently small we have*

$$L_t f(x) = -\frac{1}{\sqrt{t}} \frac{\pi^{(d-1)/2}}{2} e^{-r^2} p(x_0) \partial_{\mathbf{n}} f(x_0) + o\left(\frac{1}{\sqrt{t}}\right). \quad (4)$$

Consequently, if x is on the boundary, i.e., $r = 0$, then $x = x_0$ and

$$L_t f(x) = -\frac{1}{\sqrt{t}} \frac{\pi^{(d-1)/2}}{2} p(x) \partial_{\mathbf{n}} f(x) + o\left(\frac{1}{\sqrt{t}}\right).$$

The proof of the theorem can be found in Appendix A.1. For comparison, the corresponding result for interior points is as follows:

$$L_t f(x) = -\frac{1}{2} \pi^{d/2} p(x) \Delta_{p^2} f(x) + o(1), \quad (5)$$

Hence the graph Laplacian has a different behavior on or near a boundary point from interior points. The values of $L_t f(x)$ are of different order, $O(1/\sqrt{t})$ near the boundary as opposed to $O(1)$ in the interior. Intuitively, the scaling difference stems from integrating over a half-plane for boundary points versus integrating over a (high-dimensional) plane for points away from boundary, which causes the integration of the first-order derivative terms non-vanishing in the former case.

In practice, we do not know the boundary, and apply the same *global* normalization for all $x \in \bar{\Omega}$. The result is that the large values of $L_t f(x)$ are likely to correspond to points near the boundary (or other singular set, see below).

We think that these observation could lead to useful techniques for data analysis and help to design better algorithms that respect singularities. See Appendix C.2 for some preliminary experiments on the MNIST dataset.

3.2 Intersection-type: Intersection of Manifolds

We now present results on the behavior of the graph Laplacian near the intersection of two d -manifolds $\bar{\Omega}_1$ and $\bar{\Omega}_2$ embedded in \mathbb{R}^N with $N > d$. Note that we do not assume that the intersection is of codimension one. We also remark that while the boundary behavior is caused by the integration over a half-disk (which causes asymmetry in the integration of first-order terms) for boundary points, this is not the case for points around intersection singularity and edge-type singularity. The behavior around these latter two types of singularities are more involved and are more subtle to analyze.

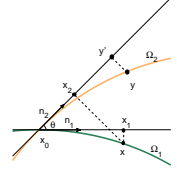
Theorem 2. *Let $\bar{\Omega}_1$ and $\bar{\Omega}_2$ be two d -dimensional smooth manifolds in \mathbb{R}^N potentially with boundaries, and their intersection $\Omega_1 \cap \Omega_2$ is a smooth manifold of dimension $l (\leq d - 1)$. Let f be a continuous function*

over $\bar{\Omega} = \bar{\Omega}_1 \cup \bar{\Omega}_2$ whose restriction $f_i := f|_{\bar{\Omega}_i}$ on $\bar{\Omega}_i$, $i = 1, 2$, is C^2 -continuous. Given a point $x \in \Omega_1$ near the intersection, let x_0 be its nearest neighbor in $\Omega_1 \cap \Omega_2$, and x_1 (resp. x_2) be its projection in the tangent space of $\bar{\Omega}_1$ at x_0 (resp. in the tangent space of $\bar{\Omega}_2$ at x_0). Put $\|x - x_0\| = r\sqrt{t}$. For a sufficiently small t , we have

$$L_t f(x) = \frac{1}{\sqrt{t}} \pi^{d/2} r e^{-r^2 \sin^2 \theta} p(x_0) (\partial_{\mathbf{n}_1} f_1(x_0) + \cos \theta \partial_{\mathbf{n}_2} f_2(x_0)) + o\left(\frac{1}{\sqrt{t}}\right), \quad (6)$$

where \mathbf{n}_1 and \mathbf{n}_2 are the unit vectors in the direction of $x_0 - x_1$ and $x_0 - x_2$, respectively, and θ is the angle between \mathbf{n}_1 and \mathbf{n}_2 .

Proof. Given $x \in \Omega_1$ and its nearest neighbor x_0 in $\Omega_1 \cap \Omega_2$, recall that x_1 and x_2 are the orthogonal projection of x onto the tangent space T_{x_0, Ω_1} of Ω_1 at x_0 , and onto the tangent space T_{x_0, Ω_2} of Ω_2 at x_0 , respectively. For any $y \in \Omega_2$, let y' denote its orthogonal projection onto the tangent space T_{x_0, Ω_2} . See the right figure for an illustration. By the definition of L_t , we have $L_t f(x) = \frac{1}{t} \int_{\bar{\Omega}} K_t(x, y) (f(x) - f(y)) p(y) dy$. Since $\bar{\Omega} = \bar{\Omega}_1 \cup \bar{\Omega}_2$, this can then be decomposed as



$$L_t f(x) = \frac{1}{t} \int_{\bar{\Omega}_1} K_t(x, y) (f_1(x) - f_1(y)) p(y) dy + \frac{1}{t} \int_{\bar{\Omega}_2} K_t(x, y) (f_1(x) - f_2(y)) p(y) dy \quad (7)$$

The first term above, which is the integral over $\bar{\Omega}_1$, is exactly the graph Laplacian of f_1 at an interior point x of $\bar{\Omega}_1$. Thus, it is bounded by $O(1)$. We now focus on the second integral over $\bar{\Omega}_2$. Let $\mathbb{B}(x)$ denote the ball in \mathbb{R}^N center around x with radius $t^{\frac{1}{2}-\varepsilon}$ for a sufficiently small constant $\varepsilon > 0$. In the derivation below, we will approximate the integral over the manifold $\bar{\Omega}_2$ by an integral over the region $\mathbb{B}(x) \cap \bar{\Omega}_2 \subseteq \bar{\Omega}_2$. The error term induced is $O(e^{-t^{-\varepsilon}}) = o(1)$ for small t ; see Appendix B in [11]. Using the fact that $f_1(x_0) = f_2(x_0)$, we have:

$$\begin{aligned} \frac{1}{t} \int_{\bar{\Omega}_2} K_t(x, y) (f_1(x) - f_2(y)) p(y) dy &= \frac{1}{t} \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x, y) (f_1(x) - f_2(y)) p(y) dy + o(1) \\ &= \frac{1}{t} \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x, y) (f_1(x) - f_1(x_0)) p(y) dy + \frac{1}{t} \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x, y) (f_2(x_0) - f_2(y)) p(y) dy + o(1) \\ &= \frac{1}{t} (f(x) - f(x_0)) \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x, y) p(y) dy + \frac{1}{t} \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x, y) (f(x_0) - f(y)) p(y) dy + o(1). \end{aligned} \quad (8)$$

Since $\|x - x_0\| = r\sqrt{t}$, it follows from Eqn (2) that $\|x_0 - x_1\| = r\sqrt{t} + O(t^{3/2})$. It can then be shown that $\|x - x_2\| = r\sqrt{t} \cdot \sin \theta + O(t)$ and $\|x_0 - x_2\| = r\sqrt{t} \cdot |\cos \theta| + O(t)$. For simplicity, let $\tilde{O}(t^\beta)$ denote $O(t^{\beta-\varepsilon})$ for a sufficiently small positive constant ε . For a point $y \in \mathbb{B}(x) \cap \bar{\Omega}_2$ and its projection y' in T_{x_0, Ω_2} , we have that $\|x - y\| = \tilde{O}(t^{\frac{1}{2}})$ and $\|x_0 - y\| \leq \|x_0 - x\| + \|x - y\| = \tilde{O}(t^{\frac{1}{2}})$. This implies that $\|y - y'\| = \tilde{O}(t)$ and $\|x - y'\| = \tilde{O}(t^{\frac{1}{2}})$ by Eqn (2). Using these distance bounds, we have:

$$\begin{aligned} K_t(x, y) &= \frac{1}{t^{d/2}} e^{-\|x-y\|^2/t} = \frac{1}{t^{d/2}} e^{-\|x-y'+y'-y\|^2/t} = \frac{1}{t^{d/2}} e^{-\|x-y'\|^2/t} \cdot e^{-\|y'-y\|^2/t} \cdot e^{\frac{O(\|x-y'\| \cdot \|y'-y\|)}{t}} \\ &= \frac{1}{t^{d/2}} e^{-\|x-y'\|^2/t} \cdot (1 + \tilde{O}(t)) \cdot (1 + \tilde{O}(t^{1/2})) = \frac{1}{t^{d/2}} e^{-\|x-y'\|^2/t} \cdot (1 + \tilde{O}(t^{1/2})) \\ &= \frac{1}{t^{d/2}} e^{-\frac{\|x-x_2\|^2}{t} + \frac{\|x_2-y'\|^2}{t}} \cdot (1 + \tilde{O}(t^{1/2})) = \frac{1}{t^{d/2}} (e^{-r^2 \sin^2 \theta} \cdot e^{O(t^{\frac{1}{2}})}) e^{-\frac{\|x_2-y'\|^2}{t}} \cdot (1 + \tilde{O}(t^{1/2})) \\ &= \frac{1}{t^{d/2}} e^{-r^2 \sin^2 \theta} e^{-\|x_2-y'\|^2/t} (1 + \tilde{O}(t^{1/2})) = e^{-r^2 \sin^2 \theta} K_t(x_2, y') (1 + \tilde{O}(t^{1/2})). \end{aligned} \quad (9)$$

Now consider the first term in Eqn (8). Let π_{x_0} denote the projection map from Ω_2 onto T_{x_0, Ω_2} and π^{-1} its inverse. Denote by $B := \pi_{x_0}(\mathbb{B}(x) \cap T_{x_0, \Omega_2})$ to be the projection of $\mathbb{B}(x) \cap T_{x_0, \Omega_2}$ on the tangent space T_{x_0, Ω_2} . By applying Taylor expansion to $f(x)$ at x_0 and $p(y)$ at x_0 , and combining Eqn 2 and 3, we have that:

$$\begin{aligned}
& \frac{1}{t}(f(x) - f(x_0)) \int_{\mathbb{B}(x) \cap \overline{\Omega_2}} K_t(x, y) p(y) dy = \frac{1}{t} (\langle \nabla f(x_0), x_1 - x_0 \rangle + \tilde{O}(t)) \int_{\mathbb{B}(x) \cap \overline{\Omega_2}} K_t(x, y) p(y) dy \\
&= \frac{1}{t^{\frac{d+1}{2}}} (r \cdot \langle \nabla f(x_0), \mathbf{n}_1 \rangle + \tilde{O}(t^{\frac{1}{2}})) \int_B \left[(e^{-r^2 \sin^2 \theta} e^{-\frac{\|x_2 - y'\|^2}{t}} (1 + \tilde{O}(t^{\frac{1}{2}}))) (p(x_0) + \tilde{O}(t^{\frac{1}{2}})) \right] J\pi^{-1}|_{y'} dy' \\
&= \frac{1}{t^{\frac{d+1}{2}}} (r \partial_{\mathbf{n}_1} f_1(x_0) + \tilde{O}(t^{\frac{1}{2}})) e^{-r^2 \sin^2 \theta} (1 + \tilde{O}(t^{\frac{1}{2}})) (p(x_0) + \tilde{O}(t^{\frac{1}{2}})) \int_B e^{-\|x_2 - y'\|^2/t} dy' \\
&= \frac{1}{t^{\frac{d+1}{2}}} (r \partial_{\mathbf{n}_1} f_1(x_0) + \tilde{O}(t^{\frac{1}{2}})) e^{-r^2 \sin^2 \theta} (1 + \tilde{O}(t^{\frac{1}{2}})) (p(x_0) + \tilde{O}(t^{\frac{1}{2}})) \left(\int_{T_{x_0, \Omega_2}} e^{-\frac{\|x_2 - y'\|^2}{t}} dy' + O(1) \right) \\
&= \frac{1}{\sqrt{t}} C_3 p(x_0) r e^{-r^2 \sin^2 \theta} \partial_{\mathbf{n}_1} f_1(x_0) + o\left(\frac{1}{\sqrt{t}}\right), \tag{10}
\end{aligned}$$

where $C_3 = \int_{T_{x_0, \Omega_2}} e^{-\|u\|^2} du = \int_{\mathbb{R}^d} e^{-\|u\|^2} du = \pi^{d/2}$. Note that we use the fact that $\frac{1}{\sqrt{t}} \tilde{O}(\sqrt{t}) = O\left(\frac{1}{t^\varepsilon}\right) = o\left(\frac{1}{\sqrt{t}}\right)$ in the above derivation. From the first line to the second line we perform a change of variable. From the third line to the fourth line in the above derivation, we relax $\int_B e^{-\|x_2 - y'\|^2/t} dy'$ to be $\int_{T_{x_0, \Omega_2}} e^{-\|x_2 - y'\|^2/t} dy' + O(e^{-t^{-\varepsilon}}) = \int_{T_{x_0, \Omega_2}} e^{-\|x_2 - y'\|^2/t} dy' + O(1)$ based on the result from Appendix B in [11].

For the second item in Eqn (8), we have

$$\begin{aligned}
& \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega_2}} K_t(x, y) (f(x_0) - f(y)) p(y) dy \\
&= \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega_2}} K_t(x, y) (f(x_0) - f(x_2) + f(x_2) - f(y)) p(y) dy \\
&= \frac{1}{t} [f(x_0) - f(x_2)] \int_{\mathbb{B}(x) \cap \overline{\Omega_2}} K_t(x, y) dy + \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega_2}} K_t(x, y) (f(x_2) - f(y)) dy \\
&= \frac{1}{\sqrt{t}} C_3 p(x_0) r \cos \theta \cdot e^{-r^2 \sin^2 \theta} \partial_{\mathbf{n}_2} f_2(x_0) + o\left(\frac{1}{\sqrt{t}}\right) \tag{11}
\end{aligned}$$

Intuitively, the first term in the third line is bounded by the quantity in the last line by a similar argument as the one carried out in Eqn (10) above. For the second term in the third line, observe that if we replace the kernel $K_t(x, y)$ to be $K_t(x_2, y)$, then $\frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega_2}} K_t(x_2, y) (f(x_2) - f(y)) dy$ roughly corresponds to the standard weighted graph Laplacian of function f_2 at an interior point x_2 in $\overline{\Omega_2}$. Furthermore, by a derivation similarly to the one in Eqn (9), replacing $K_t(x, y)$ by $K_t(x_2, y)$ only changes the integral by a factor of $e^{-r^2 \sin^2 \theta}$. Hence the second term can be bounded by $o\left(\frac{1}{\sqrt{t}}\right)$. See Appendix A.2 for details ⁵.

The theorem then follows from Eqn 10 and 11. ■

From the theorem, we can see that for a point x on the intersection or near intersection, $L_t f(x)$ is of the form $\frac{1}{\sqrt{t}} C r e^{-r^2}$ where the coefficient C is determined by the derivatives of f and the position of x . Furthermore, for a point x on the intersection, we have $r = 0$. Hence it follows that the order of $L_t f(x)$ at an intersection point is *the same* as those at regular points, i.e., $O(1)$, instead of order $O(1/\sqrt{t})$ as for points near singularities.

⁵Note x_2 is in T_{x_0, Ω_2} . In Eqn (11), to illustrate the intuition behind the derivation, we abuse the notation slightly and use $f(x_2)$ to refer to $f(\pi^{-1}(x_2))$, where $\pi^{-1}(x_2)$ is the point from $\overline{\Omega_2}$ whose projection in T_{x_0, Ω_2} is x_2 .

3.3 Edge-type: Gluing Boundaries

We now consider the case where two manifolds are glued along (a part of) their boundaries. The behavior of graph Laplacian for edge-type points is in some sense the combination of that for boundary-type and for intersection-type points: locally, a point can both “see” the boundary (thus will have terms arising from boundary effect) and the other manifold (which has effects similar to those produced by Eqn (8)). See Appendix A.3 for the proof of the following theorem.

Theorem 3. *Let $\overline{\Omega}_i$ ($i = 1, 2$) be two d -dimensional smooth manifolds in \mathbb{R}^N with interior Ω_i and nonempty boundary $\partial\Omega_i$. Assume their shared boundary $\partial\Omega_1 \cap \partial\Omega_2$ is a $d - 1$ -dimensional smooth manifold. Let f be a continuous function over $\overline{\Omega} = \overline{\Omega}_1 \cup \overline{\Omega}_2$ whose restriction $f_i := f|_{\overline{\Omega}_i}$, $i = 1, 2$ is C^2 -continuous. Given a point $x \in \Omega_1$ near the glued boundary, let x_0 be its nearest neighbor in $\partial\Omega_1 \cap \partial\Omega_2$, and put $\|x - x_0\| = r\sqrt{t}$. Then for a sufficiently small t , we have that*

$$L_t f(x) = -\frac{1}{\sqrt{t}} p(x_0) [\alpha(r, \theta) \partial_{\mathbf{n}_1} f(x_0) + \beta(r, \theta) \partial_{\mathbf{n}_2} f(x_0)] + o\left(\frac{1}{\sqrt{t}}\right), \quad (12)$$

where θ is the angle between the two inward boundary normal \mathbf{n}_1 and \mathbf{n}_2 of $\partial\Omega_1$ and $\partial\Omega_2$ at x_0 respectively,

$$\alpha(r, \theta) = \frac{1}{2} \pi^{(d-1)/2} e^{-r^2} - r \pi^{d/2} \Phi(\sqrt{2}r \cos \theta) e^{-r^2 \sin^2 \theta},$$

$$\beta(r, \theta) = \frac{1}{2} \pi^{(d-1)/2} e^{-r^2} + r \pi^{d/2} \Phi(\sqrt{2}r \cos \theta) \cos \theta e^{-r^2 \sin^2 \theta},$$

and $\Phi(x)$ is the cumulative distribution function of the standard normal distribution.

Furthermore, if x is on the edge, i.e., $r = 0$, then we have

$$L_t f(x) = -\frac{1}{2\sqrt{t}} p(x) \pi^{(d-1)/2} [\partial_{\mathbf{n}_1} f(x) + \partial_{\mathbf{n}_2} f(x)] + o\left(\frac{1}{\sqrt{t}}\right).$$

The theorem indicates that for a point x near $\partial\Omega_1 \cap \partial\Omega_2$, the limit of $L_t f(x)$ is close to the weighted sum of two normal gradients. As x moves to the set $\partial\Omega_1 \cap \partial\Omega_2$, the weights for the two normal gradients will be equal. Finally, for a point x on the boundary $\partial\Omega_1 \cap \partial\Omega_2$, the behavior $L_t f(x)$ is equivalent to the addition of the two boundary effects from $\overline{\Omega}_1$ and $\overline{\Omega}_2$ together.

Intersection-type singularity again. Notice that for the case of two manifolds Ω_1 and Ω_2 intersect at $\Omega_1 \cap \Omega_2$ of co-dimension 1, we can actually regard this scenario as four pieces of manifolds, $\Omega_1^+, \Omega_1^-, \Omega_2^+$ and Ω_2^- , glued together by $\Omega_1 \cap \Omega_2$. Here, we use Ω_i^+ and Ω_i^- to denote the two pieces of Ω_i but on the different side of $\Omega_1 \cap \Omega_2$. The advantage of taking this view is that we allow the functions on manifolds Ω_1 and Ω_2 to be only C^0 -continuous at the intersection $\Omega_1 \cap \Omega_2$, instead of C^2 -continuous as required in Theorem 2. (However, note that Theorem 2 also holds for the cases when the co-dimension of the intersection is higher than 1.) Using the same argument as in the proof of Theorem 3 to this case of gluing 4 d -manifolds along a common $d - 1$ -boundary, we obtain the following corollary.

Corollary 4. *Let $\overline{\Omega}_1$ and $\overline{\Omega}_2$ be two d -dimensional smooth manifolds in \mathbb{R}^N potentially with boundaries, and their intersection $\Omega_1 \cap \Omega_2$ is a smooth manifold of dimension 1. Let f be a continuous function over $\overline{\Omega} = \overline{\Omega}_1 \cup \overline{\Omega}_2$ whose restriction $f_i := f|_{\overline{\Omega}_i}$ on $\overline{\Omega}_i$, $i = 1, 2$, is C^2 -continuous at any regular point and C^0 -continuous at points in $\Omega_1 \cap \Omega_2$. Given a point $x \in \Omega_1$ near the intersection, let x_0 be its nearest neighbor in $\Omega_1 \cap \Omega_2$, and x_1 (resp. x_2) be its projection in the tangent space of $\overline{\Omega}_1$ at x_0 (resp. in the tangent space of $\overline{\Omega}_2$ at x_0). For a sufficiently small t , we have*

$$L_t f(x) = -\frac{\pi^{(d-1)/2}}{2} \frac{1}{\sqrt{t}} p(x) [\partial_{\mathbf{n}_1} f_1^+(x_0) + \partial_{(-\mathbf{n}_2)} f_1^-(x_0) + \partial_{\mathbf{n}_2} f_2^+(x_0) + \partial_{(-\mathbf{n}_1)} f_2^-(x_0)] + o\left(\frac{1}{\sqrt{t}}\right) \quad (13)$$

where $\partial_{\mathbf{n}_1} f_1^+(x_0), \partial_{(-\mathbf{n}_1)} f_1^-(x_0)$ are the directional derivatives of f_1 on two sides of $\overline{\Omega}_1 \cap \overline{\Omega}_2$, same for $\partial_{\mathbf{n}_2} f_2^+(x_0), \partial_{\mathbf{n}_1} f_2^-(x_0)$.

Interestingly, it appears the eigenfunctions for a singular manifold with intersection-type singularity can be only C^0 -continuous across the intersection (instead of C^2 -continuous) within each piece of manifold. Hence the above Corollary will be useful in analysing the eigenfunctions around intersection singularities (or co-dimension 1).

4 Finite Sample Complexity and Convergence Rate

Theorem 5. *Let x_1, \dots, x_n , be i.i.d. random variables in \mathbb{R}^N sampled from a probability distribution on $\overline{\Omega}$ with the intrinsic dimension d with density $p(x)$, $0 < a \leq p(x) \leq b < \infty$, defined on the manifold on $\overline{\Omega}$ with the intrinsic dimension d . Let f be a bounded function, $|f(x)| < M$. Then for any $x \in \overline{\Omega}$ we have*

$$P(|L_{n,t}f(x) - L_t f(x)| > \epsilon) \leq 2n \exp\left(-\frac{nt^{d/2+2}\epsilon^2}{2C_v + 2C_m\epsilon t/3}\right) \quad (14)$$

where C_v and C_m are constants depending on the manifold.

The proof is based on an application of the Bernstein inequality and the union bound and can be found in Appendix B.

The immediate corollary is that to get an asymptotic convergence for an arbitrary fixed point (that is an error of the order of $o(1)$), we need to select t so that $nt^{d/2+2}/\log(n) \rightarrow \infty$. That is, we can choose $t = (\log(n)/n)^{\frac{2}{d+4}}g(n)$, where $g(n)$ is an arbitrary function such that $\lim_{n \rightarrow \infty} g(n) = \infty$.

On the other hand, for points near the singular set the scaling of the operator $L_{n,t}$ changes. While inside the domain $L_{n,t}f(x) = O(1)$ for a smooth (fixed) function f , for x on the singular set or sufficiently close to it (within \sqrt{t} distance), $L_{n,t}f(x) = O(\frac{1}{\sqrt{t}})$. Thus an accurate estimate for the appropriately rescaled operator requires that $nt^{d/2+1}/\log(n) \rightarrow \infty$ leading to $t = (\log(n)/n)^{\frac{2}{d+2}}g(n)$. This may seem counter-intuitive as fewer points are required on the singularity than inside the domain to obtain an accurate estimate. One explanation is that near a singularity we are effectively estimating a degree one differential operator, while inside the domain we are estimating the Laplace-Beltrami operator, which is degree two.

5 Discussion of Impact on Laplacian Eigenfunctions

Eigenfunctions of graph Laplacians obtained from data play an important role in a variety of applications from spectral clustering to dimensionality reduction and semi-supervised learning. While full proof of their convergence is likely to be very subtle and is beyond the scope of this paper (see [5] for the proof of eigenfunction convergence for the smooth case), we would like to discuss the implications of our results for eigenfunctions under the assumption that such convergence takes place. Let ϕ_t be an eigenfunction of L_t (normalized to norm 1 in L_2). We put $\lambda = \lim_{t \rightarrow 0} \lambda_t$ and $\phi = \lim_{t \rightarrow 0} \phi_t$ (with the functions converging in L_2 norm). We will also assume that the necessary derivatives exist.

Below, we briefly discuss the impact of each type of singularity. Some numerical experiments are provided in Appendix C.3. For simplicity, we will assume the samples to be infinite and the multiplicity of each eigenfunction to be one.

Boundary singularity: For a point x on or near (within \sqrt{t} of) the boundary our results indicate

$$\lambda_t \phi_t(x) = L_t \phi_t(x) = C \frac{1}{\sqrt{t}} \partial_{\mathbf{n}} \phi(x) + o\left(\frac{1}{\sqrt{t}}\right)$$

where C is a constant independent of t . Thus, for small values of t , and $\lambda_t \phi_t(x) = O(\frac{1}{\sqrt{t}}) \partial_{\mathbf{n}} \phi(x)$, which is clearly impossible unless $\partial_{\mathbf{n}} \phi_t(x) = O(\sqrt{t})$. Passing to the limit⁶ we see that for x on the boundary, $\partial_{\mathbf{n}} \phi(x) = 0$, meaning that the limit eigenfunctions of the graph Laplacian automatically satisfy the Neumann boundary conditions. The experimental results in Appendix C.3 are consistent with this finding.

Edge singularity: Consider now a point x at the edge where two "sheets" Ω_1 and Ω_2 come together. From Theorem 3 we have

$$L_t \phi_t(x) = \frac{C}{\sqrt{t}} (\partial_{\mathbf{n}_1} \phi_t(x) + \partial_{\mathbf{n}_2} \phi_t(x)) + o\left(\frac{1}{\sqrt{t}}\right).$$

where \mathbf{n}_1 and \mathbf{n}_2 are the two directions normal to the singular set at x . By an argument following the one above, we have $\partial_{\mathbf{n}_1} \phi_t(x) + \partial_{\mathbf{n}_2} \phi_t(x) = O(\sqrt{t})$. and hence

$$\partial_{\mathbf{n}_1} \phi(x) + \partial_{\mathbf{n}_2} \phi(x) = 0$$

In other words the directional derivatives for the limit eigenfunction must cancel each other. This is quite surprising for the following reason: suppose, that the edge singularity is obtained by folding a smooth manifold (imagine folding a sheet of paper, creating an edge). Then the above condition means that the eigenfunctions of the original surface are invariant under this "folding", despite the fact that the graph Laplacian operator is very different near the singular set! In other words if a singular surface is isometric to a smooth manifolds, the spectral structure appears to be preserved, even though the distances computed for points near the edge are very far from the intrinsic distances on smooth and the operator near the edge is not the Laplace-Beltrami operator.

Some numerical illustrations of this phenomenon are given in the Appendix C.3.

Intersection singularity: For the intersection singularity the analysis follows that of the edge case, but the implications for the shape of eigenfunctions are not completely clear, and warrents further investigation. One intriguing observation is that if the co-dimension of the singularity is greater than one, it has only a local effect on the shape of eigenfunctions. The situation is analogous to the disjoint union of manifolds. The reason is easy to see – the volume of \sqrt{t} -neighborhood B around the singular set of co-dimension at least two, will be at most $O((\sqrt{t})^2) = O(t)$. Recall that an eigenfunction minimizes the quantity $\langle L_t \phi_t, \phi_t \rangle_{L_2(p)}$ under orthogonality conditions. We see that the contribution of the points around the singular set is bounded by $O(t)O(\frac{1}{\sqrt{t}}) = O(\sqrt{t})$ and vanishes as $t \rightarrow 0$.

⁶There are additional subtleties related to convergence in L_2 , which are ignored since we are not aiming to provide a full proof of convergence.

References

- [1] M. Aanjaneya, F. Chazal, D. Chen, M. Glisse, L. Guibas, and D. Morozov. Metric graph reconstruction from noisy data. In *Proc. 27th Sympos. Comput. Geom.*, pages 37–46, 2011.
- [2] M. Belkin. *Problems of learning on manifold*. PhD thesis, University of Chicago, 2003.
- [3] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comp*, 15(6):1373–1396, 2003.
- [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] Mikhail Belkin and Partha Niyogi. Convergence of laplacian eigenmaps. *preprint, short version appeared in NIPS 2007*, 2008.
- [6] Paul Bendich, Bei Wang, and Sayan Mukherjee. Local homology transfer and stratification learning. In *Proceedings of ACM-SIAM Symposium on Discrete Algorithms*, 2012.
- [7] O. Chapelle, B. Schölkopf, and A. Zien. *Semi-supervised Learning*. MIT Press, 2006.
- [8] Frédéric Chazal, David Cohen-Steiner, and André Lieutier. A sampling theory for compact sets in euclidean space. *Discrete & Computational Geometry*, 41(3):461–479, 2009.
- [9] Frédéric Chazal and Steve Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proc. 24th ACM Sympos. on Comput. Geom.*, pages 232–241, 2008.
- [10] G. Chen and G. Lerman. Foundations of a multi-way spectral clustering framework for hybrid linear modeling. *Foundations of Computational Mathematics*, 9(5):517–558, 2009.
- [11] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.
- [12] Lawrence Conlon. *Differentiable manifold*. Birkhäuser Boston, 2008.
- [13] Evarist Giné and Vladimir Koltchinskii. Empirical graph laplacian approximation of laplace-beltrami operators: Large sample results. 51:238–259, 2006.
- [14] A. Goldberg, X. Zhu, A. Singh, Z. Xu, , and R. Nowak. Multi-manifold semi-supervised learning. In *Artificial Intelligence and Statistics, AISTATS*, 2009.
- [15] A. Grigor’yan. Heat kernels on weighted manifolds and applications. *Cont. Math.*, 398:93–191, 2006.
- [16] Wolfgang Härdle. *Applied Nonparametric Regression*. Cambridge Univeristy Press, 1992.
- [17] Gloria Haro, Gregory Randall, and Guillermo Sapiro. Translated poisson mixture model for stratification learning. *International Journal of Computer Vision*, 80(3):358–374, 2008.
- [18] M. Hein. *Geometrical aspects of statistical learning theory*. PhD thesis, Wissenschaftlicher Mitarbeiter am Max-Planck-Institut für biologische Kybernetik in Tübingen in der Abteilung, 2005.
- [19] Matthias Hein, Jean yves Audibert, and Ulrike Von Luxburg. Graph laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8:1325–1368, 2007.
- [20] S. Lafon. *Diffusion Maps and Geodesic Harmonics*. PhD thesis, Yale University, 2004.

- [21] Gilad Lerman and Teng Zhang. Probabilistic recovery of multiple subspaces in point clouds by geometric l_p minimization, 2010.
- [22] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.
- [23] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. A topological view of unsupervised learning from noisy data. *SIAM J. Comput.*, 40(3):646–663, 2011.
- [24] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
- [25] A. Singer. From graph to manifold laplacian: The convergence rate. *Appl. Comput. Harmon. Anal.*, 21:128–134, 2006.
- [26] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- [27] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (gpca). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1945–1959, 2005.
- [28] U. von Luxburg. A tutorial on spectral clustering. In *Statistics and Computing*, volume 17, pages 395–416, 2007.
- [29] X. Zhu. Semi-supervised learning literature survey. *Computer Science, University of Wisconsin-Madison*, 2006.

A Proofs for Theorems on Graph Laplacian Limit

A.1 Sketch of Proof for Theorem 1

For a sufficiently small t , let Ω_{bd} be the set of points that are within distance $t^{1/2-\epsilon}$ from the boundary $\partial\Omega$ (a thin layer of “shell”), where ϵ is any sufficiently small positive constant. Set $\Omega_{in} = \overline{\Omega}/\Omega_{bd}$. We will show that for a small t , $L_t f(x)$ is approximated by two different terms on Ω_{db} and Ω_{in} , and more importantly they have different orders of t .

Points away from boundary. Given a point $x \in \Omega$, let $\mathbb{B}(x)$ denote the ball of radius $t^{1/2-\epsilon}$ centered at x , where ϵ is any sufficiently small positive constant. First, we approximate the integral in Eqn (1) when the integral is taken inside the ball $\mathbb{B}(x)$. By results from Appendix B of [11], the error induced by constraining the integral to within $\mathbb{B}(x) \cap \overline{\Omega}$ is only $O(e^{-t^{-\epsilon}})$.

Now given a point $y \in \mathbb{B}(x)$, let u be the projection of y onto the tangent space $T_{x,\Omega}$ of $\overline{\Omega}$ at x . Choose x to be the origin of $T_{x,\Omega}$. We have the following approximations for each of the three terms in the integral in Eqn (1) (see e.g, [2, Chapter 4.2] and [11, Appendix B]). Here $K(a)$ denotes $K(a) = e^{-\frac{a}{t}}$. N and d are the dimension of the ambient space and of the tangent space, respectively.

$$\begin{aligned} K(\|x - y\|_{\mathbb{R}^N}^2) &= K(\|u\|_{\mathbb{R}^d}^2) + O(\|u\|_{\mathbb{R}^d}^4) \\ f(x) - f(y) &= -u^T \nabla f(x) - \frac{1}{2} u^T H(x) u + O(\|u\|_{\mathbb{R}^d}^3) \\ p(y) &= p(x) + u^T \nabla p(x) + O(\|u\|_{\mathbb{R}^d}^2) \end{aligned} \quad (15)$$

where $H(x)$ is the Hessian of $f(x)$ at x .

Now for simplicity let $\tilde{O}(t^\beta)$ denote $O(t^{\beta-\epsilon})$ for any sufficiently small positive constant $\epsilon > 0$. Let $\pi_x : \overline{\Omega} \rightarrow T_{x,\Omega}$ be the projection from $\overline{\Omega}$ onto the tangent space $T_{x,\Omega}$, and set $R := \pi_x(\mathbb{B}(x) \cap \overline{\Omega})$. Obviously, $R \subset T_{x,\Omega}$. Combing the above approximations together with the fact that $y \in \mathbb{B}(x)$ and thus $\|x - y\| = \tilde{O}(t)$, and using a change of variable $u \rightarrow \sqrt{t}w$, we obtain the following:

$$\begin{aligned} L_t f(x) &= \frac{1}{t} \int_{\overline{\Omega}} K_t(x, y) (f(x) - f(y)) p(y) dy \\ &= \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}} K_t(x, y) (f(x) - f(y)) p(y) dy + O(e^{-t^{-\epsilon}}) \\ &= -\frac{1}{t^{d/2}} \int_R \frac{1}{t} K(\|u\|_{\mathbb{R}^d}^2) [(\sqrt{t}u^T \nabla f(x) + \frac{t}{2}u^T H(x)u) \times \\ &\quad (p(x) + \sqrt{t}u^T \nabla p(x))] t^{d/2} J\pi^{-1}|_u du + \frac{1}{t} \cdot \tilde{O}(t^{1/2}) \\ &= -\int_{T_{x,\Omega}} e^{-\|w\|_{\mathbb{R}^d}^2} \left\{ \frac{1}{\sqrt{t}} [p(x)(w^T \nabla f(x))] + \right. \\ &\quad \left. [w^T \nabla f(x) \times w^T \nabla p(x) + \frac{1}{2} p(x) w^T H(x) w] \right\} dw + o\left(\frac{1}{\sqrt{t}}\right) \end{aligned} \quad (16)$$

From the first line to the second line, we replace the integral over $\overline{\Omega}$ with ball $\mathbb{B}(x)$, generating an exponentially small error for sufficiently small t [11, Appendix B]. From the second line to the third line, by changing the variable from y to $\pi_x(y) = u$, this integral can be rewritten as an integral over the region $R \subset T_{x,\Omega}$. In the fourth line, we relax this integral over R to be the integral over the entire tangent space $T_{x,\Omega}$, plus another exponentially small error $O(e^{-t^{-\epsilon}})$ which is consumed by $o(\frac{1}{\sqrt{t}})$.

For an interior point $x \in \Omega_{in}$, $T_{x,\Omega}$ is a d -dimensional linear space, and the function $e^{-\|w\|_{\mathbb{R}^d}^2}$ is an even function on $T_{x,\Omega}$. When taking the integral, the first term in the last integral in Eqn (16) with order $O\left(\frac{1}{\sqrt{t}}\right)$ is odd and therefore vanishes. The remaining three terms in the last line of Eqn (16) are of order $O(1)$ inside

the integral, and they are exactly the weighted Laplacian at x as previously studied in the literature, and hence $L_t f(x) = o(\frac{1}{\sqrt{t}})$. The kernel is however not symmetric for points on or near the boundary, which are studied next.

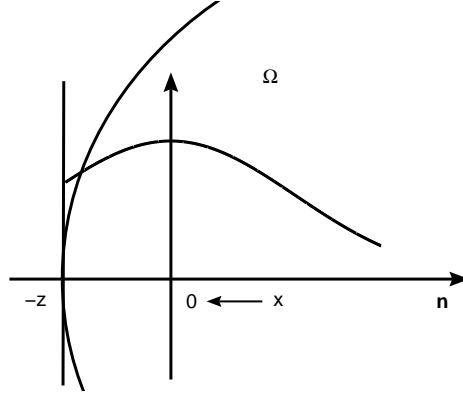


Figure 1: Gaussian weight at x near the boundary.

Points near boundary. We now consider $x \in \Omega_{bd}$ (the “shell”) near the boundary. See Figure 1, where we show the local neighborhood around x projected in $T_{x,\Omega}$. Let x_0 be the nearest boundary point on $\partial\Omega$ to x . Choose a local coordinate system of $T_{x,\Omega}$ with x at the origin, and assume the projection of x_0 in $T_{x,\Omega}$ is at $-z$. Let \mathbf{n} the unit direction along $x - z$. When x is close to the boundary, the kernel $K_t(x, y)$ is no longer symmetric. In particular, consider an orthonormal coordinate system around x , the Gaussian kernel is symmetric in all coordinate-axis other than along $x - z$. Along the direction \mathbf{n} , the Gaussian convolution is from $-z/\sqrt{t}$ (as we have changed variable from $u \rightarrow \sqrt{t}w$) to $+\infty$, which is not symmetric. Therefore, $e^{-\|w\|_{\mathbb{R}^d}^2}$ is not an even function in the normal direction \mathbf{n} , and the integral of the highest order term (i.e. the first term of $O(\frac{1}{\sqrt{t}})$) in the last integral of Eqn (16) will not vanish.

Specifically, let $u = (u_1, u_2, \dots, u_d)$ be the coordinate of u in the aforementioned coordinate system of $T_{x,\Omega}$. Assume that u_1 is the coordinate along the normal direction \mathbf{n} . Recall that we have applied the change of variable $u \rightarrow \sqrt{t}w$. So let $w = (w_1, w_2, \dots, w_d)$ be the coordinate of w . Since we have assumed in the theorem that $\|x - x_0\| = r\sqrt{t}$, we have that $\frac{z}{\sqrt{t}} = r + O(r^3)$. When we integrate the last integral in Eqn (16), all the odd terms of u_i still vanish in all directions except the normal direction \mathbf{n} , and the most important point is that the leading term along the normal direction is of order $O(\frac{1}{\sqrt{t}})$, different from $O(1)$ for the interior points. In particular, we have:

$$\begin{aligned} L_t f(x) &= -\frac{1}{\sqrt{t}} p(x) \partial_{\mathbf{n}} f(x) \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \int_{-z/\sqrt{t}}^{\infty} e^{-\|w\|_{\mathbb{R}^d}^2} w_1 dw_1 dw_2 \dots dw_d + o\left(\frac{1}{\sqrt{t}}\right) \\ &= -\frac{1}{\sqrt{t}} \cdot \frac{1}{2} \pi^{\frac{d-1}{2}} e^{-r^2} \partial_{\mathbf{n}} f(x) + o\left(\frac{1}{\sqrt{t}}\right). \end{aligned}$$

The theorem follows from further performing the Taylor expansion of $f(x)$ at x_0 .

A.2 Details for the Intersection Case

Our goal is to bound the integral $\frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y) (f(x_0) - f(y)) p(y) dy$ as in Eqn (11). First, recall that $\pi_{x_0} : \overline{\Omega}_2 \rightarrow T_{x_0, \Omega_2}$ is the projection map from $\overline{\Omega}_2$ to the tangent space of $\overline{\Omega}_2$ at x_0 . For points on $\overline{\Omega}_2$

sufficiently close to x_0 , this map is in fact a bijection and its inverse π^{-1} exists. Hence for t sufficiently small (and thus x_2 is close enough to x), $\pi^{-1}(x_2)$ exists and we set it to be $x_3 := \pi^{-1}(x_2) \in \bar{\Omega}_2$. Since $\|x_0 - x_2\| = r \cos \theta \sqrt{t} + O(t)$, we have $\|x_2 - x_3\| = O(t)$ by Eqn (2). It follows from triangle inequality that $\|x - x_3\| = \|x - x_2\| + O(t) = r \sin \theta \sqrt{t} + O(t)$. We can then apply a similar derivation as in Eqn (9) to show that

$$K_t(x, y) = e^{-r^2 \sin^2 \theta} K_t(x_3, y) \cdot (1 + \tilde{O}(\sqrt{t})); \text{ and } K_t(x, y) = e^{-r^2 \sin^2 \theta} K_t(x_3, y') \cdot (1 + \tilde{O}(\sqrt{t})), \quad (17)$$

where y' is the projection of y on T_{x_0, Ω_2} as defined in Section 3.2. We now have

$$\begin{aligned} & \frac{1}{t} \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x, y) (f(x_0) - f(y)) p(y) dy \\ &= \frac{1}{t} (f(x_0) - f(x_3)) \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x, y) p(y) dy + \frac{1}{t} \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x, y) (f(x_3) - f(y)) p(y) dy \end{aligned} \quad (18)$$

Following the same derivation as in Eqn (10), and combining with $\|x_0 - x_2\| = r \cos \theta \sqrt{t} + O(t)$, the first term can be bound by:

$$\begin{aligned} & \frac{1}{t} (f(x_0) - f(x_3)) \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x, y) p(y) dy \\ &= \frac{1}{\sqrt{t}} (\|x_0 - x_2\| \partial_{\mathbf{n}_2} f_2(x_0) + \tilde{O}(t^{\frac{1}{2}})) e^{-r^2 \sin^2 \theta} (1 + \tilde{O}(t^{\frac{1}{2}})) (p(x_0) + \tilde{O}(t^{\frac{1}{2}})) \int_{T_{x_0, \Omega_2}} K_t(x_2, y') dy' \\ &= \frac{1}{\sqrt{t}} C_3 p(x_0) r \cos \theta \cdot e^{-r^2 \sin^2 \theta} \partial_{\mathbf{n}_2} f_2(x_0) + o\left(\frac{1}{\sqrt{t}}\right). \end{aligned} \quad (19)$$

By using Eqn (17), the second term in Eqn (18) is:

$$\begin{aligned} & \frac{1}{t} \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x, y) (f(x_3) - f(y)) p(y) dy \\ &= \frac{1}{t} \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} e^{-r^2 \sin^2 \theta} (1 + \tilde{O}(\sqrt{t})) \cdot K_t(x_3, y) (f(x_3) - f(y)) p(y) dy \\ &= e^{-r^2 \sin^2 \theta} (1 + \tilde{O}(\sqrt{t})) \cdot \frac{1}{t} \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x_3, y) (f(x_3) - f(y)) p(y) dy + o\left(\frac{1}{\sqrt{t}}\right). \end{aligned} \quad (20)$$

Now observe that since the radius of $\mathbb{B}(x)$ is $\Theta(t^{\frac{1}{2}-\varepsilon})$ for a sufficiently small positive ε , and since the distance from x to x_3 is $O(t^{1/2})$, we have that there exists some ball $\mathbb{B}'(x_3)$ of radius still in the asymptotic order of $\Theta(t^{\frac{1}{2}-\varepsilon})$ around x_3 such that $\mathbb{B}'(x_3)$ is contained inside $\mathbb{B}(x)$. This implies that $\mathbb{B}'(x_3) \cap \bar{\Omega}_2 \subseteq \mathbb{B}(x) \cap \bar{\Omega}_2$. Hence by results from Appendix B of [11],

$$\int_{\mathbb{B}'(x_3) \cap \bar{\Omega}_2} K_t(x_3, y) (f(x_3) - f(y)) p(y) dy = \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x_3, y) (f(x_3) - f(y)) p(y) dy + O(e^{-t^{-\varepsilon}}).$$

On the other hand, $\frac{1}{t} \int_{\mathbb{B}'(x_3) \cap \bar{\Omega}_2} K_t(x_3, y) (f_2(x_3) - f_2(y)) p(y) dy$ is simply an approximation of the functional Laplacian $L_t f_2(x_3)$ of the function f_2 at an interior point x_3 over a manifold $\bar{\Omega}_2$. Hence by Eqn (5),

$$\frac{1}{t} \int_{\mathbb{B}'(x_3) \cap \bar{\Omega}_2} K_t(x_3, y) (f(x_3) - f(y)) p(y) dy = -\frac{1}{2} \pi^{d/2} p(x_3) \Delta_{p^2} f_2(x_3) + o(1).$$

It then follows from Eqn (20) that

$$\begin{aligned}
& \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y)(f(x_3) - f(y))p(y)dy \\
&= e^{-r^2 \sin^2 \theta} \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x_3, y)(f(x_3) - f(y))p(y)dy + o\left(\frac{1}{\sqrt{t}}\right) \\
&= -\frac{1}{2} \pi^{d/2} p(x_3) \Delta_{p^2} f_2(x_3) + o\left(\frac{1}{\sqrt{t}}\right) = o\left(\frac{1}{\sqrt{t}}\right). \tag{21}
\end{aligned}$$

Putting Eqn (18), (19) and (21) together, we conclude that

$$\frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y)(f(x_0) - f(y))p(y)dy = \frac{1}{\sqrt{t}} C_3 p(x_0) r \cos \theta \cdot e^{-r^2 \sin^2 \theta} \partial_{\mathbf{n}_2} f_2(x_0) + o\left(\frac{1}{\sqrt{t}}\right),$$

which is what we claimed in Eqn (11).

A.3 Proof for Edge-type Points

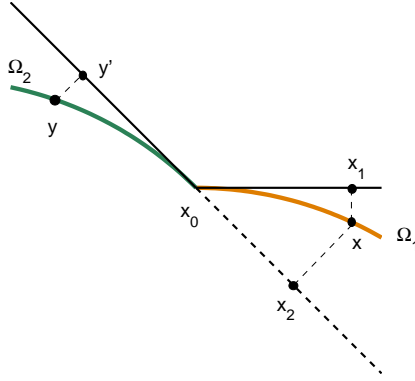


Figure 2: Edge case.

Suppose we have two manifolds forming the “edge”, denote them by $\overline{\Omega}_1$ and $\overline{\Omega}_2$. We use the same notation as in the proof for the intersection case; see the illustration in Figure 2. In particular, let x_0 denote the closest point from $\partial\overline{\Omega}_1 \cap \partial\overline{\Omega}_2$ to a point $x \in \Omega_1$. Let x_1 and x_2 be the projection of x on the tangent space T_{x_0, Ω_1} and T_{x_0, Ω_2} at x_0 , respectively. Note that $\|x - x_0\| = r\sqrt{t}$. Let f_1, f_2 be the restriction of f on Ω_1 and Ω_2 respectively; we know that $f_1(x_0) = f_2(x_0)$. Again, $\mathbb{B}(x)$ denotes the ball centered at x with radius $t^{\frac{1}{2}-\varepsilon}$. Then we have

$$\frac{1}{t} \int_{\overline{\Omega}} K_t(x, y)(f(x) - f(y))p(y)dy = \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}} K_t(x, y)(f(x) - f(y))p(y)dy + o(1);$$

and

$$\begin{aligned}
& \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}} K_t(x, y)(f(x) - f(y))p(y)dy \\
&= \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_1} K_t(x, y)(f_1(x) - f_1(y))p(y)dy + \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y)(f_1(x) - f_2(y))p(y)dy \tag{22}
\end{aligned}$$

The first integral in Eqn (22) on $\mathbb{B}(x) \cap \overline{\Omega}_1$ is exactly what we had before for the boundary singularity (recall the second line in Eqn (16)), which we have shown to be the following in Appendix A.1:

$$-\frac{1}{\sqrt{t}} \cdot \frac{1}{2} \pi^{(d-1)/2} p_1(x_0) e^{-r^2} \langle \nabla f_1, \mathbf{n}_1 \rangle + o\left(\frac{1}{\sqrt{t}}\right). \quad (23)$$

Now consider the second integral in Eqn (22) on $\mathbb{B}(x) \cap \overline{\Omega}_2$:

$$\begin{aligned} & \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y) (f_1(x) - f_2(y)) p(y) dy \\ &= \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y) (f_1(x) - f_2(x_0) + f_2(x_0) - f_2(y)) p(y) dy \\ &= \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y) (f_1(x) - f_1(x_0)) p(y) dy + \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y) (f_2(x_0) - f_2(y)) p(y) dy \\ &= \frac{1}{t} (f_1(x) - f_1(x_0)) \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y) p(y) dy + \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y) (f_2(x_0) - f_2(y)) p(y) dy \end{aligned} \quad (24)$$

Let π_x denote the projection from $\overline{\Omega}_2$ to T_{x_0, Ω_2} , and set $y' = \pi_x(y)$ for any point $y \in \overline{\Omega}_2$. Since the notation is the same with intersection case, we have by Eqn (9)

$$K_t(x, y) = \frac{1}{t^{d/2}} (e^{-r^2 \sin^2 \theta} e^{-\|x_2 - y'\|^2/t} (1 + \tilde{O}(t^{1/2}))).$$

Set $B := \pi_x(\mathbb{B}(x) \cap \overline{\Omega}_2)$ as the image of $\mathbb{B}(x) \cap \overline{\Omega}_2$ in the tangent space T_{x_0, Ω_2} . Let T_{x_0, Ω_2}^+ denote the half-space of the tangent space T_{x_0, Ω_2} that contains R . Combing the bounds in Eqn (2) and (3), for the first integral in Eqn (24), we have

$$\begin{aligned} & \frac{1}{t} (f_1(x) - f_1(x_0)) \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y) p(y) dy = \frac{1}{t} (f_1(x) - f_1(x_0)) \int_B K_t(x, y) p(y) J \pi_x^{-1}|_y dy' \\ &= \frac{1}{t^{d/2+1}} (f_1(x) - f_1(x_0)) e^{-r^2 \sin^2 \theta} \int_B e^{-\|x_2 - y'\|^2/t} (1 + \tilde{O}(t^{1/2})) (p(x_0) + \tilde{O}(t^{1/2})) dy' \\ &= \frac{1}{t^{d/2+1}} (f_1(x) - f_1(x_0)) e^{-r^2 \sin^2 \theta} (1 + \tilde{O}(t^{1/2})) (p(x_0) + \tilde{O}(t^{1/2})) \left(\int_{T_{x_0, \Omega_2}^+} e^{-\|x_2 - y'\|^2/t} dy' + O(e^{-t^{-\varepsilon}}) \right) \\ &= C(r, \theta) \frac{1}{\sqrt{t}} p_2(x_0) r e^{-r^2 \sin^2 \theta} \langle \nabla f_1(x_0), \mathbf{n}_1 \rangle + o\left(\frac{1}{\sqrt{t}}\right), \end{aligned} \quad (25)$$

where $C(r, \theta) = \int_{T_{x_0, \Omega_2}^+} e^{-\|x_2 - y'\|^2/t} dy' = \int_{\mathbb{R}^d, y_d > -r \cos(\theta)} e^{-\|y\|^2} dy = \pi^{d/2} \Phi(\sqrt{2}r \cos \theta)$, and y_d is the d -th entry of coordinate of y in the d -dimensional tangent space. The last step above also involves applying the following formula:

$$f_1(x) - f_1(x_0) = \langle \nabla f_1(x_0), x_1 - x_0 \rangle + \tilde{O}(\|x - x_0\|^2) = r\sqrt{t} \langle \nabla f_1(x_0), \mathbf{n}_1 \rangle + \tilde{O}(t).$$

Now for the second integral in Eqn (24), we have:

$$\begin{aligned} & \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y) (f_2(x_0) - f_2(y)) p(y) dy \\ &= \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y) \langle \nabla f_2|_{x_0}, x_0 - y \rangle p(y) dy \\ &= \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y) \langle \nabla f_2|_{x_0}, x_0 - x_2 \rangle p(y) dy + \frac{1}{t} \int_{\mathbb{B}(x) \cap \overline{\Omega}_2} K_t(x, y) \langle \nabla f_2|_{x_0}, x_2 - y \rangle p(y) dy \end{aligned} \quad (26)$$

The second term from the above equation is exactly the same as if we were considering a point x_2 nears the boundary of $\bar{\Omega}_2$. Hence using the results from Appendix A.1, we have that

$$\begin{aligned}
& \frac{1}{t} \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x, y) \langle \nabla f_2|_{x_0}, x_2 - y \rangle p(y) dy \\
&= \frac{1}{t^{\frac{d}{2}+1}} e^{-r^2 \sin^2 \theta} \int_B e^{-\|x_2 - y'\|^2/t} \langle \nabla f_2|_{x_0}, x_2 - y \rangle (p(x_0) + \tilde{O}(t^{1/2})) dy \\
&= \frac{1}{t^{\frac{d}{2}+1}} e^{-r^2 \sin^2 \theta} (p(x_0) + \tilde{O}(t^{1/2})) \langle \nabla f_2|_{x_0}, \int_B e^{-\|x_2 - y'\|^2/t} (x_2 - y) dy \rangle \\
&= \frac{1}{t^{\frac{d}{2}+1}} e^{-r^2 \sin^2 \theta} (p(x_0) + \tilde{O}(t^{1/2})) \langle \nabla f_2|_{x_0}, \int_{T_{x_0, \Omega_2}^+} e^{-\|x_2 - y'\|^2/t} (x_2 - y) dy \rangle \\
&= \frac{1}{t^{\frac{d}{2}+1}} e^{-r^2 \sin^2 \theta} (p(x_0) + \tilde{O}(t^{1/2})) \left(-\frac{t^{d/2}}{2} \pi^{\frac{d-1}{2}} e^{-r^2 \cos^2 \theta} \partial_{\mathbf{n}_2} f_2 \right) \\
&= -\frac{1}{\sqrt{t}} \cdot \frac{1}{2} \pi^{\frac{d-1}{2}} e^{-r^2} \partial_{\mathbf{n}_2} f_2(x_0) + o\left(\frac{1}{\sqrt{t}}\right). \tag{27}
\end{aligned}$$

For the first term from Eqn (26), we have:

$$\begin{aligned}
& \frac{1}{t} \int_{\mathbb{B}(x) \cap \bar{\Omega}_2} K_t(x, y) \langle \nabla f_2|_{x_0}, x_0 - x_2 \rangle p(y) dy \\
&= \frac{1}{t^{\frac{d}{2}+1}} e^{-r^2 \sin^2 \theta} \int_B e^{-\|x_2 - y'\|^2/t} (1 + \tilde{O}(t^{1/2})) \langle \nabla f_2(x_0), x_0 - x_2 \rangle (p(x_0) + \tilde{O}(t^{1/2})) dy' \\
&= \frac{1}{t^{\frac{d}{2}+1}} e^{-r^2 \sin^2 \theta} (1 + \tilde{O}(t^{1/2})) (p(x_0) + \tilde{O}(\sqrt{t})) \int_B e^{-\|x_2 - y'\|^2/t} \langle \nabla f_2(x_0), x_0 - x_2 \rangle dy' \\
&= \frac{1}{t^{\frac{d}{2}+1}} e^{-r^2 \sin^2 \theta} p(x_0) \langle \nabla f_2(x_0), x_0 - x_2 \rangle \int_R e^{-\|x_2 - y'\|^2/t} dy' \\
&= \frac{1}{t^{\frac{d}{2}+1}} e^{-r^2 \sin^2 \theta} p(x_0) \langle \nabla f_2, x_0 - x_2 \rangle \left(\int_{T_{x_0, \Omega_2}^+} e^{-\|x_2 - y'\|^2/t} dy' + O(e^{-t^{-\epsilon}}) \right) \\
&= \frac{1}{\sqrt{t}} p(x_0) C(r, \theta) \cdot r \cos \theta e^{-r^2 \sin^2 \theta} \langle \nabla f_2(x_0), \mathbf{n}_2 \rangle + o\left(\frac{1}{\sqrt{t}}\right), \tag{28}
\end{aligned}$$

where $C(r, \theta)$ is defined earlier. Let x' be the projection of x_1 on the tangent space T_{x_0, Ω_2} . The last step in Eqn (28) uses the fact that $\|x_0 - x_2\| = r\sqrt{t} \cdot |\cos \theta| + O(t)$. Combining all the results in (23), (25), (26), (27), and (28), we prove Theorem 3.

B Proof for Convergence Rate

Lemma 6. *Let X_1, \dots, X_n, Z be i.i.d. random variables of \mathbb{R}^N from density $p(x) \in C^\infty(\bar{\Omega})$, $0 < a \leq p(x) \leq b < \infty$, $K_t(x, y)$ be a Gaussian weight function and $|f| < M$, then for any $x \in \bar{\Omega}$*

$$P \left(\left| \frac{1}{n} \sum_{i=1}^n K_t(x, X_i) f(X_i) - \mathbb{E}_Z [K_t(x, Z) f(Z)] \right| > \epsilon \right) \leq 2 \exp \left(-\frac{nt^{d/2} \epsilon^2}{2C_v + 2C_m \epsilon / 3} \right) \tag{29}$$

where C_v and C_m only depend on d and $K(\cdot, \cdot)$, $f(\cdot)$, $p(\cdot)$.

Proof. Let $W_i(x) = K_t(x, X_i)f(X_i)$, and $Y_i(x) = W_i(x) - \mathbb{E}_{X_i}[W_i(x)]$. We have $\mathbb{E}_{X_i}[Y_i(x)] = 0$, and

$$\begin{aligned} |Y_i(x)| &\leq |W_i(x)| + |\mathbb{E}_{X_i}[W_i(x)]| \\ &= |K_t(x, X_i)f(X_i)| + \left| \int_{\bar{\Omega}} K_t(x, y)f(y)p(y)dy \right| \\ &\leq \frac{1}{t^{d/2}}M + Mb \int_{\bar{\Omega}} K_t(x, y)dy \\ &\leq \frac{M}{t^{d/2}} + MbC_g = (C_m + MbC_g t^{d/2})/t^{d/2} \end{aligned} \quad (30)$$

where $C_g = \int_{\bar{\Omega}} K_t(x, y)dy < \infty$, and $C_m = M$. For $\text{Var}[Y_i(x)]$,

$$\begin{aligned} \text{Var}[Y_i(x)] &= \mathbb{E}_{X_i}[W_i^2(x)] - \{\mathbb{E}_{X_i}[W_i(x)]\}^2 \\ &\leq \int_{\bar{\Omega}} K_t^2(x, y)f^2(y)p(y)dy + \left[\int_{\bar{\Omega}} K_t(x, y)f(y)p(y)dy \right]^2 \\ &\leq \frac{1}{t^{d/2}}M^2b \int_{\bar{\Omega}} K_t(x, y)dy + M^2b^2C_g^2 \\ &\leq \frac{1}{t^{d/2}}M^2bC_g + M^2b^2C_g^2 = (C_v + M^2b^2C_g^2 t^{d/2})/t^{d/2} \end{aligned} \quad (31)$$

where $C_v = M^2bC_g$. By the Bernstein's inequality

$$P \left(\left| \frac{1}{n} \sum_{i=1}^n Y_i(x) \right| > \epsilon \right) \leq 2 \exp \left(- \frac{nt^{d/2}\epsilon^2/2}{(C_v + M^2b^2C_g^2 t^{d/2}) + (C_m + MbC_g t^{d/2})\epsilon/3} \right) \quad (32)$$

□

Proof of theorem 5: For $x \in \bar{\Omega}$, we put $f(x) - f(X_i)$ into Lemma 6, obtaining:

$$\begin{aligned} &P (|L_{n,t}f(x) - L_t f(x)| > \epsilon) \\ &= P \left(\left| \frac{1}{n} \sum_{i=1}^n K_t(x, X_i)(f(x) - f(X_i)) - \mathbb{E}_Z[K_t(x, Z)(f(x) - f(Z))] \right| > t\epsilon \right) \\ &\leq 2 \exp \left(- \frac{nt^{(d+4)/2}\epsilon^2}{2C_v + 2C_m t\epsilon/3} \right) \end{aligned}$$

The proof of theorem 5 is done.

C Experimental results

Below we provide some numerical results validating theoretical analysis, some experiments with real datasets and several numerical experiments in support of our conjectures on the eigenfunction behavior.

C.1 A numerical example

In this section, we illustrate the behavior of graph Laplacian on or near the singularity points through a simple numerical example. In order to explore different cases, we consider the union of 1-dimensional manifolds $\bar{\Omega}$ as shown in Figure 3 (a), which is a combination of three linear intervals, $\bar{\Omega}_1$, $\bar{\Omega}_2$, and $\bar{\Omega}_3$ in \mathbb{R}^2 . We take 2500 point uniformly spaced over the singular manifold and choose $f(x_1, x_2) = (x_1 + 0.2)^2 + x_2^2$ (restricted to the manifold).

Shape and Scaling Behavior: The value of $L_{n,t}f(x)$ over all points in $\bar{\Omega}_1$ is shown in Figure 3 (b). We can see that $L_{n,t}f(x)$ near the boundary is approximately half a Gaussian function, while the function near the intersection is similar to a function of the form $\varphi(x) = axe^{-bx^2}$, as predicted by Theorem 2. The shape of $L_{n,t}f(x)$ near the corner is also what we expect from Theorem 3.

To explore the scaling effects, in Figure 3 (c) we plot values of $\log |L_{n,t}f(x)|$ against $\log(t)$ for different values of t . Each graph corresponds to a fixed point x chosen near a boundary, intersection or edge singularity. Our theoretical results predict a linear curve with slope $-\frac{1}{2}$ in log-log coordinates (corresponding to $1/\sqrt{t}$, which is consistent with our experimental results as shown in (c).

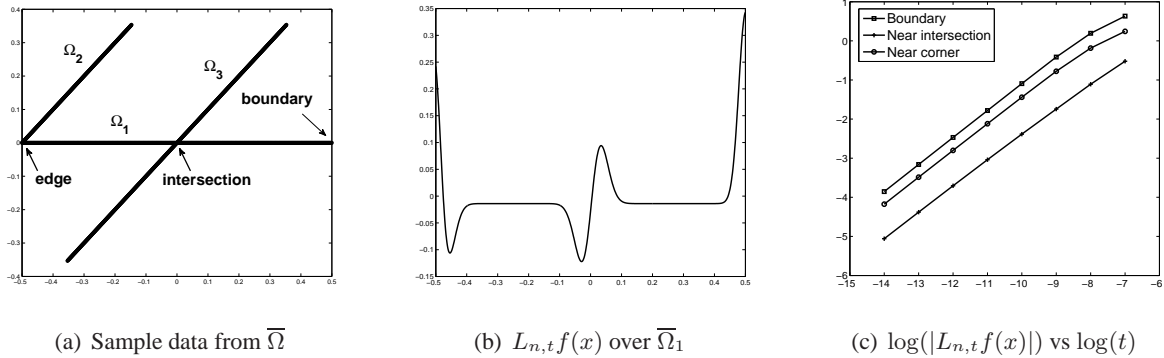
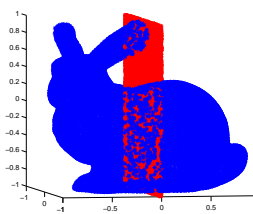


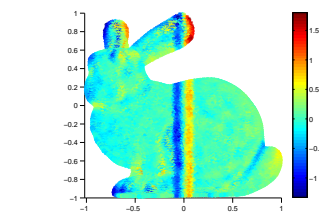
Figure 3: $L_{n,t}f(x)$ with $f(x_1, x_2) = (x_1 + 0.2)^2 + x_2^2$ over $\bar{\Omega}$. In (c), x axis is $\log(|L_{n,t}f(x)|)$ and y axis is $\log t$. The three curves are for three points near each type of singularity.

C.2 Singularity Detection and Estimation

Bunny Model: In the following example, we use the union of the Stanford bunny dataset (in blue) and a plane $y = 0$ (in red), which intersects the bunny in the middle as shown in panel (a). We have taken 10000 points for both the bunny and the plane in the point cloud, and choose the function $f = y$. The values of $L_{n,t}f$ on the bunny are shown in panel (b). We can



(a) Data set



(b) Values of $L_{n,t}f$ on Bunny when $f(x, y, z) = y$.

see that the values of $L_{n,t}f$ at the intersection are 0 and values near the intersection go up (in yellow) and down (in blue) as expected from the theoretical analysis of the intersection singularity.

MNIST Data: From the scaling behavior of the graph Laplacian near singularities, it is potentially possible to detect certain such singularities on real world data sets. We use the example of the MNIST digit images. The image of each digit can be thought of as a sample from a low-dimensional manifold embedded in the high-dimensional pixel space. We choose function $f(x)$ to be the summation of the pixel intensity of each image. We take the images corresponding to the top 2% of the highest value of $L_{n,t}f(x)$ as the “near singularity/boundary” images, which are compared to the “average” images of each digit. The results are shown in Figure (7). We see that the images on the right are significantly larger (and systematically different) from average, suggesting that they may belong to the part of the boundary, which can be detected through the sum-of-pixel-intensity function.

C.3 Laplacian Eigenfunctions on Singular Manifolds

In this section, we will provide two numerical examples to validate our conjectures about the Laplacian eigenfunctions on the singular manifolds.

Folded Rectangle: In this example, we consider two manifolds: (1) the rectangular region Ω_1 , $(-0.3, 0.3) \times (-0.5, 0.5) \times \{0\}$, in \mathbb{R}^3 and (2) the “folded” rectangle Ω_2 , obtained by transforming the positive y part of

Ω_1 , i.e. $\{(x, y, z) \in \Omega_1 : y > 0\}$, by applying the linear transformation given by the matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\pi/4) & -\sin(\pi/4) \\ 0 & \sin(\pi/4) & \cos(\pi/4) \end{bmatrix}$$

and keeping the rest of Ω_1 fixed as showing in Fig. 4. We see Ω_1 and Ω_2 are intrinsically isometric, yet Ω_2 has an edge singularity.

We choose 6000 uniformly-spaced points in both Ω_1 and Ω_2 .

With this two data sets, we construct their graph Laplace matrices respectively, and calculate their eigenvalues and eigenvectors using the Gaussian kernel with $t = 10^{-4}$. Figure 5 shows the first 16 nontrivial eigenvectors of two graph Laplace matrices. We can see that the corresponding eigenvectors match nearly perfectly. It can be easily verify that they also match eigenfunctions of the rectangle with Neumann boundary conditions.

To give some precise numerical results we measure the relative difference between the sets of eigenvalues. Let λ_k^1 and λ_k^2 be the two vectors of first k eigenvalues for Ω_1 and Ω_2 respectively. We define the (relative) difference as

$$\text{diff}_k = \frac{\|\lambda_k^1 - \lambda_k^2\|}{\|\lambda_k^1\|}$$

The table below shows the difference for various values of k . We observe that the approximation appears quite precise with difference of the order of 0.1%, thus providing further evidence in support of our conjecture on the behavior of eigenfunctions.

k	10	50	100
diff _k	0.0014	0.0012	0.0011

Sliced and flipped sphere: To give a more interesting non-flat example, we consider two spaces shown in Fig. 6: the standard unit sphere in \mathbb{R}^3 and the "sliced and flipped" sphere, where the top of the sphere is sliced off and glued in the opposite direction. More precisely, (1) $S_1 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1.\}$; (2) S_2 is the set of points in S_1 except that the points above the plane $z = 0.75$ are moved to their mirror reflection on the other side of the plane $z = 0.75$. We see that S_1 and S_2 are intrinsically isometric. To construct the Laplacian we sample 10000 points from each space and choose $t = 5 \times 10^{-4}$. In this example, it is not easy to visualize the eigenvectors due the multiplicity of eigenvalues. However, from the table below we observe that the spectra of the graph Laplacian on those manifolds are very similar.

k	10	50	100
diff _k	0.0099	0.0182	0.0148

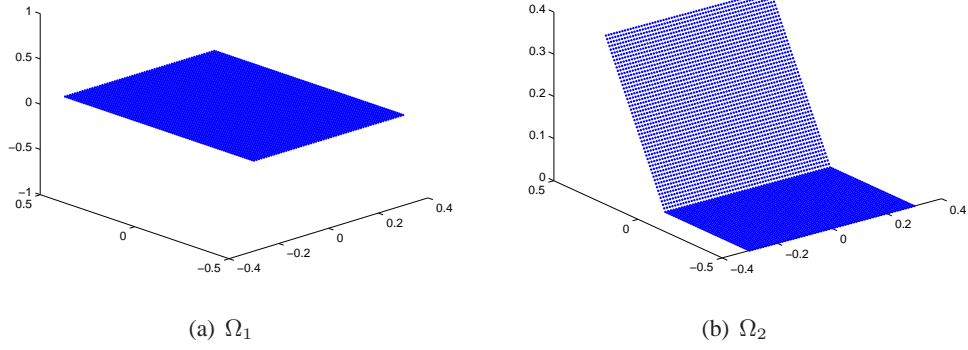


Figure 4: Rectangle and folded rectangle Ω_1 and Ω_2 .

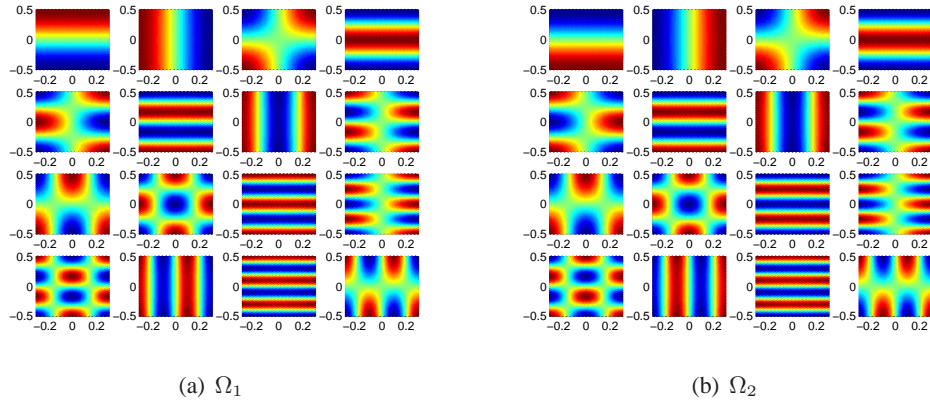


Figure 5: 2nd-17th eigenvectors of the graph Laplace matrices of the two manifolds Ω_1 and Ω_2 .

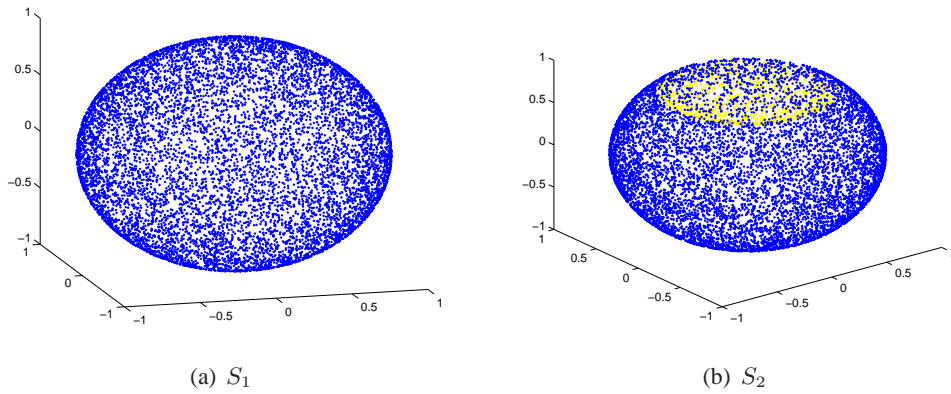
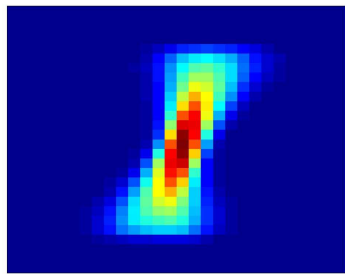
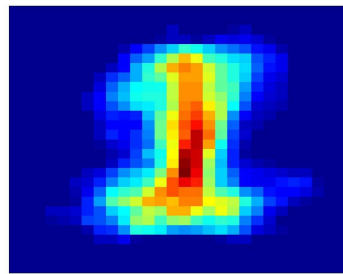


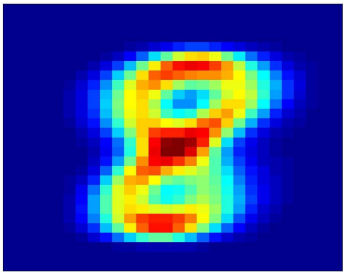
Figure 6: Unit sphere S_1 (left), and unit sphere with the top “sliced and flipped” S_2 (right).



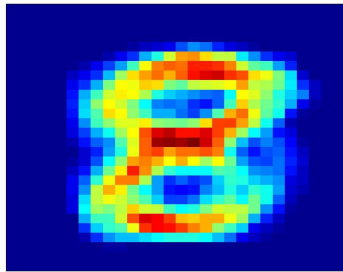
(a)



(b)



(c)



(d)

Figure 7: An averaged image of digits "1" and "8" (left panels) vs an averaged image with a large value of graph Laplacian constructed from the data set of images, applied to a function defined by the sum of pixel intensities.