

# IMPROVED SPECTRAL CONVERGENCE RATES FOR GRAPH LAPLACIANS ON $\varepsilon$ -GRAPHS AND $k$ -NN GRAPHS

JEFF CALDER AND NICOLÁS GARCÍA TRILLOS

**ABSTRACT.** In this paper we improve the spectral convergence rates for graph-based approximations of weighted Laplace-Beltrami operators constructed from random data. We utilize regularity of the continuum eigenfunctions and strong pointwise consistency results to prove that spectral convergence rates are the same as the pointwise consistency rates for graph Laplacians. In particular, for an optimal choice of the graph connectivity  $\varepsilon$ , our results show that the eigenvalues and eigenvectors of the graph Laplacian converge to those of a weighted Laplace-Beltrami operator at a rate of  $O(n^{-1/(m+4)})$ , up to log factors, where  $m$  is the manifold dimension and  $n$  is the number of vertices in the graph. Our approach is general and allows us to analyze a large variety of graph constructions that include  $\varepsilon$ -graphs and  $k$ -NN graphs. We also present the results of numerical experiments analyzing convergence rates on the two dimensional sphere.

## 1. INTRODUCTION

Our work is motivated by applications in machine learning, statistics and artificial intelligence. There, the goal is to learn structure from a given data set  $X = \{x_1, \dots, x_n\}$ . To do this several authors have proposed the use of graphs to endow data sets with some geometric structure, and have utilized graph Laplacians to understand how information propagates on the graph representing the data. Graph Laplacians and their spectra form the basis of algorithms for supervised learning [1, 4, 48, 56], clustering [41, 51] and dimensionality reduction [2, 16]. The works [43, 49, 53] discuss Laplacian regularization in the context of non-parametric regression. Bayesian approaches to learning where graph Laplacians are used to define covariance matrices for Gaussian priors have been proposed in [5, 36, 57].

To better understand algorithms based on graph Laplacians, it has proven useful to study the large sample size asymptotics of graph Laplacians when these capture the closeness of data points in Euclidean space, as is the case in constructions such as  $\varepsilon$ -graphs or  $k$ -NN graphs. In this limit, we pass from discrete graph Laplacians to continuum Laplace-Beltrami operators, or weighted versions thereof, and in particular graph Laplacians are seen as specific discretizations of continuum operators. By analyzing the passage to the limit one effectively studies the consistency of algorithms that utilize said operators. In doing so, one gathers information about allowed choice of parameters, and gets insights about computational stability of algorithms (e.g. [24, 34, 35]). Naturally, in order for the “passage to the continuum” to imply any sort of consistency for a particular machine learning algorithm, it is important to study the convergence in an appropriate sense.

Early work on consistency of graph Laplacians focused on pointwise consistency results for  $\varepsilon$ -graphs (see, for example, [3, 30, 32, 33, 45, 50]). There, as well as hereinafter, the data is assumed to be an *i.i.d.* sample of size  $n$  from a ground truth measure  $\mu$  supported on an  $m$ -dimensional submanifold  $\mathcal{M}$  embedded in a high dimensional Euclidean space  $\mathbb{R}^d$  (i.e., the *manifold assumption* [14]) and pairs of points that are within distance  $\varepsilon$  of each other are given high weights. Pointwise consistency results show that as  $n \rightarrow \infty$  and the connectivity parameter  $\varepsilon \rightarrow 0$  (at a slow enough rate), the graph Laplacian applied to a fixed smooth test function

---

(J. Calder) SCHOOL OF MATHEMATICS, UNIVERSITY OF MINNESOTA  
(N. García Trillo) DEPARTMENT OF STATISTICS, UNIVERSITY OF WISCONSIN-MADISON  
*E-mail address:* [jcalder@umn.edu](mailto:jcalder@umn.edu), [nicolásgarcia@stat.wisc.edu](mailto:nicolasgarcia@stat.wisc.edu).

JC was supported by NSF-DMS grant 1713691. NGT was supported by NSF grant DMS 1912802.

converges to a continuum operator, such as a weighted Laplace-Beltrami operator applied to the test function. Recent work is moving beyond pointwise consistency and studying the sequence of solutions to graph-based learning problems and their continuum limits, using tools like  $\Gamma$ -convergence [11, 18, 28, 47], tools from PDE theory [8, 9, 11, 19, 25, 54] including the maximum principle and viscosity solutions, and more recently random walk and Martingale methods [12]. Regarding spectral convergence of graph Laplacians, the regime  $n \rightarrow \infty$  and  $\varepsilon$  constant was studied in [52], and in [46] which analyzes connection Laplacians. Works that have studied regimes where  $\varepsilon$  is allowed to decay to zero include [27], [44], [7], and [22].

The starting point for our work is the paper [22] which used ideas from [7] in order to obtain what to the best of our knowledge are the state of the art results on spectral convergence of  $\varepsilon$ -graph Laplacians. These results can be summarized as follows. With very high probability the error of approximation of eigenvalues of a continuum elliptic differential operator by the eigenvalues of their graph Laplacian counterpart scales like

$$\varepsilon + \frac{\log(n)^{p_m}}{n^{1/m}\varepsilon},$$

where  $p_m = 1/m$  for  $m \geq 3$ ,  $p_2 = 3/4$ , and  $\varepsilon$  is the length scale for the graph construction. These results suggested that the best rate of convergence is achieved when  $\varepsilon$  is chosen to scale like  $\sqrt{\frac{\log(n)^{p_m}}{n^{1/m}}}$ , in which case the convergence rate for eigenvalues is  $O(n^{-1/2m})$ , up to log factors. For eigenvectors, the error of approximation in the  $L^2$  norm was shown to scale like the square root of the convergence rate of eigenvalues, so  $O(n^{-1/4m})$  up to log factors. In this paper, we improve in several regards the results presented in [22]. Our contributions to the analysis of spectral convergence of graph Laplacians constructed from  $\varepsilon$ -graphs are summarized as follows:

- (1) In the  $\varepsilon$ -graph setting, we show that the eigenvalues of the graph Laplacian converge (with rates), provided that  $\varepsilon$  scales like

$$\frac{\log(n)^{1/m}}{n^{1/m}} \ll \varepsilon \ll 1.$$

This result is valid for all  $m \geq 1$ . This improves the results in [22] by removing an additional logarithmic term. In a sense, the lower bound on the allowed values for  $\varepsilon$  for the convergence to hold is an optimal requirement due to the connectivity threshold results for random geometric graphs [42].

- (2) In the  $\varepsilon$ -graph case, when  $\varepsilon$  scales like

$$(1.1) \quad C \left( \frac{\log(n)}{n} \right)^{\frac{1}{m+4}} \leq \varepsilon \ll 1,$$

we show that the rate of convergence of eigenvalues coincides with the *pointwise* convergence rates of the graph Laplacian (e.g [32]), and in particular with high probability scale linearly in the connectivity length-scale  $\varepsilon$ . If we choose  $\varepsilon = C \left( \frac{\log(n)}{n} \right)^{1/(m+4)}$ , then we obtain convergence rates of  $O(n^{-1/(m+4)})$ , up to log factors, which is sharper than the  $O(n^{-1/2m})$  convergence rate from [22] when  $m \geq 5$ .

- (3) We establish convergence rates for eigenfunctions under  $L^2$ -type distances that will be made explicit later on. In particular, in the same regime for  $\varepsilon$  given in (1.1), we establish that the rate of convergence of eigenvectors scales linearly in  $\varepsilon$ , matching the convergence rate of eigenvalues as well as the pointwise convergence rates. Thus, choosing again  $\varepsilon = C \left( \frac{\log(n)}{n} \right)^{1/(m+4)}$ , we obtain convergence rates for eigenvectors of  $O(n^{-1/(m+4)})$ , which is far sharper than the  $O(n^{-1/4m})$  convergence rates from [22].

A second main contribution of our work is to provide spectral consistency results for graph Laplacians constructed from  $k$ -NN graphs. Our work is the first one to obtain any rates of convergence in such a setting. Moreover, in proving the spectral convergence we also obtain rates for pointwise convergence which to the best of our knowledge are also new in the literature.

There are very few works in the literature that we are aware of that have rigorously addressed consistency for graph Laplacians associated to  $k$ -NN graphs. In [50] pointwise convergence is analyzed (without providing any rates). In [21] asymptotic spectral convergence is discussed, but no rates are provided. In [19], pointwise consistency with rates is established for the game-theoretic  $p$ -Laplacian on  $k$ -NN graphs. Posterior work to the first version of this paper, like that in [15], have considered more general normalizations for  $k$ -NN graphs in order to induce different ways in which data density affects the behavior of data analysis algorithms.

In practical applications,  $k$ -NN graphs are almost always preferred over  $\varepsilon$ -graphs, due to their far better sparsity and connectivity properties (see, e.g., [10, 19] for semi-supervised learning, and [55] for spectral clustering). Since the  $k$ -nearest neighbor relation is *not symmetric*,  $k$ -NN graphs are normally symmetrized in order to ensure the graph Laplacian is self-adjoint and the spectrum real-valued. On a symmetrized  $k$ -NN graph, the local neighborhood is no longer a Euclidean or geodesic ball, and is in fact not even symmetric. This raises technical difficulties in obtaining pointwise consistency results with rates, and makes the analysis far more involved than it is for  $\varepsilon$ -graphs.

Our contributions in this setting are as follows:

- (1) We provide spectral convergence rates for graph Laplacians when the graph is a  $k$ -NN graph, provided  $k$  scales like

$$\log(n) \ll k \ll n.$$

This result is valid for all  $m \geq 1$ . Moreover, we show that the rates of convergence coincide with the pointwise convergence rates (see Theorem 3.7 below) when  $k$  scales like

$$(1.2) \quad C \log(n)^{\frac{m}{m+4}} n^{\frac{4}{m+4}} \log(n)^{\frac{m}{m+4}} \leq k \ll n.$$

- (2) We establish convergence rates for eigenvectors under different topologies of interest that will be discussed later on. Moreover, in the regime

$$C \log(n)^{\frac{m}{m+4}} n^{\frac{4}{m+4}} \log(n)^{\frac{m}{m+4}} \leq k \ll n,$$

the convergence rate for eigenvectors coincides with the convergence rates of eigenvalues and also the pointwise convergence rates from Theorem 3.7.

It is worth mentioning that all our estimates hold with high probability for finite (although possibly large)  $n$ . These results imply a quantitative improvement to a large body of work that has built on previous spectral convergence results. For example the works [22, 23, 26] get directly benefited from our new estimates.

There are two essential steps in our analysis that allow us to improve in several regards the rates presented in [22] for the  $\varepsilon$ -graph case. In the first step, we use a simple modification of the construction of *discretization* and *interpolation* maps introduced in [7] (a construction that was later used in [22], though cast in the language of optimal transport), in order to prove spectral convergence (with rates) for a wider range of scalings of  $\varepsilon$  valid for all dimensions  $m \geq 2$ . A more detailed outline of the construction of these maps and a discussion on what needs to be adjusted from [22] is discussed in Section 2.4 below.

The second step in our analysis makes use of a simple argument for comparing eigenvalues of different *self-adjoint* operators. To illustrate the idea, let  $A, B : H \rightarrow H$  be linear operators on a Hilbert space  $H$ , with  $A$  self-adjoint. Let  $u$  be an eigenfunction of  $A$  with eigenvalue  $\lambda_u$ , and let  $w$  be an eigenfunction of  $B$  with eigenvalue  $\lambda_w$ . We may assume  $\|u\|_H = \|w\|_H = 1$ . Since  $A$  is self-adjoint

$$\lambda_u \langle u, w \rangle_H = \langle Au, w \rangle_H = \langle u, Aw \rangle_H = \lambda_w \langle u, w \rangle_H + \langle u, (A - B)w \rangle_H,$$

and thus

$$|\lambda_u - \lambda_w| \leq \frac{\|Aw - Bw\|_H}{|\langle u, w \rangle_H|}.$$

3

This inequality allows us to convert pointwise estimates on  $\|Aw - Bw\|_H$  into estimates on the spectrum, provided  $\langle u, w \rangle_H$  is bounded away from zero. For graph Laplacians,  $A$  should, say, represent the graph Laplacian, while  $B$  represents the continuum (weighted) Laplace-Beltrami operator (or, more accurately, its restriction to the graph). The key ingredients in our proof are good pointwise estimates, which rely *essentially* on the regularity of the continuum eigenfunctions, and the *a priori* eigenfunction convergence rate from the first step of our analysis, which ensures  $\langle u, w \rangle_H$  is bounded away from zero. The bottom line is that our *a priori* (non-optimal) spectral convergence rates can be bootstrapped to make them coincide with the pointwise consistency rates, provided we are willing to shrink the allowed asymptotic scaling for  $\varepsilon$  slightly. The consistency of eigenfunctions will be a consequence of the *a priori* convergence rate for eigenvalues. This is made explicit by following some of the steps in the proof of the classical Davis-Kahan theorem.

Regarding the results for  $k$ -NN graphs, we first notice that these types of graphs can be thought of intuitively as  $\varepsilon$ -graphs where one allows  $\varepsilon$  to vary in space. Given the inhomogeneity of the natural length scale  $\varepsilon$  (and which intuitively is influenced by data density), the first part of our analysis must rely on the definition of new discretization and interpolation maps that are tailored to the inhomogeneous length-scale setting. After a careful analysis, we are able to provide *a priori* spectral convergence rates analogous to the *a priori* rates obtained for  $\varepsilon$ -graphs. These non-optimal rates can then be bootstrapped to improve them just as in the  $\varepsilon$ -graph case, using the pointwise consistency results that we derive in Theorem 3.7. We note that the pointwise consistency results for graph Laplacians on  $k$ -NN graphs do not follow directly from viewing the graph as an  $\varepsilon$ -graph with  $\varepsilon$  varying in space. Indeed, looking forward to the proof of Theorem 3.7, the local neighborhood on a mutual (or exclusive)  $k$ -NN graph is asymptotically non-symmetric, due to non-uniformity of the data distribution, and so pointwise consistency for  $k$ -NN graph Laplacians requires a far more careful analysis than for  $\varepsilon$ -graphs, where the local neighborhoods are balls.

**1.1. Outline.** The rest of the paper is organized as follows. In Section 2 we give the precise set-up used throughout the paper, state our assumptions, and present our main results. Specifically, Section 2.1 contains the precise definitions of the graph constructions that we study. In Section 2.2 we state our main results regarding convergence of eigenvalues for both  $\varepsilon$ -graphs as well as  $k$ -NN graphs, and in Section 2.3 we present the results regarding convergence of eigenvectors. In Section 2.4 we provide an outline of our proofs. In Section 3 we present the pointwise consistency results of graph Laplacians which will be needed later on. In Section 4 we present the proofs of our main results. More specifically, in Section 4.1 we present the analysis for the  $\varepsilon$ -graph case, and in Section 4.2 for  $k$ -NN graphs. In Section 5 we discuss other modes of convergence for eigenvectors, and in particular the  $TL^2$ -convergence which implies Wasserstein convergence of Laplacian embeddings. In Section 6 we present the results of numerical experiments for convergence rates for  $k$ -NN and  $\varepsilon$ -graph Laplacian spectra to the spherical harmonics on the two dimensional sphere, and we conclude in Section 7.

## 2. SET-UP AND MAIN RESULTS

Let  $\mathcal{M}$  be a compact, connected, orientable, smooth  $m$ -dimensional manifold embedded in  $\mathbb{R}^d$ . We give to  $\mathcal{M}$  the Riemannian structure induced by the ambient space  $\mathbb{R}^d$ . With respect to the induced metric tensor, we let  $\text{Vol}_{\mathcal{M}}$  be  $\mathcal{M}$ 's volume form and we let  $\mu$  be a probability measure supported on  $\mathcal{M}$  with density (with respect to the volume form)  $\rho : \mathcal{M} \rightarrow (0, \infty)$  which we assume is bounded, and bounded away from zero, i.e.

$$0 < \rho_{\min} \leq \rho \leq \rho_{\max} < \infty,$$

and is at least  $C^{2,\vartheta}(\mathcal{M})$ , which here should be interpreted as saying that  $\rho$ , expressed in normal coordinates, has second derivatives that are  $\vartheta$ -Holder continuous. All the previous assumptions on  $\rho$  are not rare in theoretical works on graph-based learning. They form a concrete set of assumptions where tangible connections between machine learning algorithms (like spectral

clustering) and PDEs can be established. Relaxing these assumptions, for example the lower bound on  $\rho$ , is a compelling research direction that we do not pursue in this paper.

Let  $X = \{x_1, \dots, x_n\}$  be a set of i.i.d. samples from  $\mu$ , and let  $\mu_n$  denote the associated empirical measure, i.e.

$$\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{x_i}.$$

In what follows we will use the notation  $L^2(\mu)$  to denote the space of  $L^2$ -functions with respect to the measure  $\mu$ , and by  $L^2(\mu_n)$  the space of functions  $u : X \rightarrow \mathbb{R}$ . We endow  $L^2(\mu)$  with the inner product

$$\langle f, \tilde{f} \rangle_{L^2(\mu)} = \int_{\mathcal{M}} f(x) \tilde{f}(x) d\mu(x), \quad f, \tilde{f} \in L^2(\mu),$$

and  $L^2(\mu_n)$  with the inner product

$$\langle u, \tilde{u} \rangle_{L^2(\mu_n)} := \frac{1}{n} \sum_{i=1}^n u(x_i) \tilde{u}(x_i), \quad u, \tilde{u} \in L^2(\mu_n).$$

**2.1. Graph constructions.** In this section we define two different graph constructions on  $X$  with the purpose of leveraging the geometry of the manifold  $\mathcal{M}$ .

**2.1.1.  $\varepsilon$ -graphs.** Let  $\varepsilon > 0$ . We construct a weighted graph  $G^\varepsilon = (X, w^\varepsilon)$  as follows. First, we put an edge between  $x_i$  and  $x_j$  and between  $x_j$  and  $x_i$  (and write  $x_i \sim x_j$ ) provided that

$$|x_i - x_j| \leq \varepsilon,$$

where in the above,  $|x_i - x_j|$  is the *Euclidean* distance between the points  $x_i, x_j$ . We let  $E = \{(i, j) \in \{1, \dots, n\}^2 : x_i \sim x_j\}$  be the set of such edges. We may endow edges with weights that depend inversely on the distance between the vertices connected by them. For that purpose, let  $\eta : [0, \infty) \rightarrow [0, \infty)$  be a non-increasing function with support on the interval  $[0, 1]$  and whose restriction to  $[0, 1]$  is Lipschitz continuous. For convenience we assume that  $\eta(1/2) > 0$ . We introduce the constant

$$(2.1) \quad \sigma_\eta := \int_{\mathbb{R}^m} |y_1|^2 \eta(|y|) dy,$$

where  $y_1$  represents the first coordinate of the vector  $y \in \mathbb{R}^m$ . Notice that a simple computation using radial coordinates shows that when  $\eta(t) = \mathbb{1}_{[0,1]}(t)$  then  $\sigma_\eta = \frac{\alpha_m}{m+2}$ , where  $\alpha_m$  is the volume of the  $m$ -dimensional Euclidean unit ball.

To every given edge  $(i, j) \in E$  we assign the weight  $w_{x_i x_j}^\varepsilon$  where

$$(2.2) \quad w_{xy}^\varepsilon = \eta\left(\frac{|x - y|}{\varepsilon}\right).$$

and we can consider the weighted graph  $G^\varepsilon(X, w^\varepsilon)$ . In fact, note that if the points  $x_i, x_j$  are not connected by an edge in  $E$  then  $w_{x_i x_j}^\varepsilon = 0$ .

Having introduced the graph  $G^\varepsilon$ , we define an associated graph Laplacian operator  $\mathcal{L}^\varepsilon$  which for a given  $u \in L^2(\mu_n)$  is defined as

$$(2.3) \quad \mathcal{L}^\varepsilon u(x) = \frac{1}{n\varepsilon^{m+2}} \sum_{j=1}^n w_{x_j x}^\varepsilon (u(x) - u(x_j)).$$

Note that in principle  $x$  need not be a vertex of the graph to make sense of the above expression, but unless otherwise stated, in the sequel  $\mathcal{L}^\varepsilon$  will be thought of as an operator  $\mathcal{L}^\varepsilon : L^2(\mu_n) \rightarrow L^2(\mu_n)$ .

It is well known in the literature (e.g. [51]) that  $\mathcal{L}^\varepsilon$  is a positive semi-definite self-adjoint operator with respect to the inner product  $\langle \cdot, \cdot \rangle_{L^2(\mu_n)}$ . In particular, we can list its eigenvalues (repeated according to multiplicity) as

$$0 = \lambda_1^\varepsilon \leq \lambda_2^\varepsilon \leq \dots \leq \lambda_n^\varepsilon,$$

$\lambda_1^\varepsilon$  always being equal to zero. Also, it is possible to find an orthonormal basis of eigenvectors for  $L^2(\mu_n)$ . Moreover, a *graph Dirichlet energy* associated to  $\mathcal{L}^\varepsilon$  defined by

$$(2.4) \quad b_\varepsilon(u) := \frac{1}{n\varepsilon^{m+2}} \sum_{i,j} w_{x_i x_j}^\varepsilon (u(x_i) - u(x_j))^2 = 2 \langle \mathcal{L}^\varepsilon u, u \rangle_{L^2(\mu_n)}, \quad u \in L^2(\mu_n).$$

can be used to define the eigenvalues of  $\mathcal{L}^\varepsilon$  variationally. Namely, the Courant-Fisher minmax principle says that

$$(2.5) \quad \lambda_l^\varepsilon = \frac{1}{2} \min_{S \in \mathfrak{S}_l} \max_{u \in S \setminus \{0\}} \frac{b_\varepsilon(u)}{\|u\|_{L^2(\mu_n)}^2}$$

where in the above  $\mathfrak{S}_l$  denotes the set of all linear subspaces of  $L^2(\mu_n)$  of dimension  $l$ .

In this paper we will restrict our attention to the definition of graph Laplacian in (2.3) (known in the literature as unnormalized Laplacian), but we note that there are several other normalizations that are of high interest in machine learning (e.g. [51]), and we expect we can carry out a similar analysis for other normalizations.

**2.1.2. Undirected  $k$ -NN graph.** A different graph construction on  $X$  proceeds not by fixing a length-scale  $\varepsilon$  but rather by specifying for each point in  $X$  a set of *nearest neighbors*.

**Definition 2.1.** Let  $k \in \mathbb{N}$ . We define a relation  $\sim_k$  on  $X \times X$  by declaring

$$x_i \sim_k x_j,$$

if  $x_j$  is among the  $k$  nearest neighbors (in the Euclidean distance sense) of  $x_i$ .

In this section we symmetrize the relation  $\sim_k$  and place an edge between  $x_i$  and  $x_j$  if  $x_i \sim_k x_j$  or  $x_j \sim_k x_i$ . This is often called the *symmetric*  $k$ -nearest neighbor (or  $k$ -NN) graph [39]. While the graph described above is unweighted, *i.e.*,  $\eta = \mathbb{1}_{[0,1]}$ , we will use a weighted construction given below in (2.9). Another popular construction is the *mutual*  $k$ -NN graph, where we connect  $x_i$  to  $x_j$  if  $x_i \sim_k x_j$  and  $x_j \sim_k x_i$ . While in the sequel we formulate our results for symmetric  $k$ -NN graphs, they apply with minor modifications to mutual  $k$ -NN graphs as well.

To construct the  $k$ -NN graph Laplacian, let

$$(2.6) \quad N_\varepsilon(x) = \sum_{j: 0 < |x_j - x| \leq \varepsilon} 1$$

be the number of random samples in a punctured Euclidean  $\varepsilon$ -neighborhood of  $x$ . Given  $1 \leq k \leq n-1$  define

$$(2.7) \quad \varepsilon_k(x) := \min\{\varepsilon > 0 : N_\varepsilon(x) \geq k\}.$$

The value  $\varepsilon_k(x)$  is the Euclidean distance from  $x$  to the  $k^{\text{th}}$  nearest neighbor of  $x$  from the samples  $x_1, \dots, x_n$ . Thus,  $x_i \sim_k x_j$  if and only if  $|x_i - x_j| \leq \varepsilon_k(x_i)$ . Finally, we define

$$(2.8) \quad r_k(x, y) := \max\{\varepsilon_k(x), \varepsilon_k(y)\}.$$

Notice that  $|x_i - x_j| \leq r_k(x_i, x_j)$  if and only if  $x_i$  and  $x_j$  are connected by an edge in the symmetric  $k$ -NN graph. We would obtain the mutual  $k$ -NN graph by setting  $r_k(x, y) = \min\{\varepsilon_k(x), \varepsilon_k(y)\}$ .

The undirected  $k$ -NN graph Laplacian of an element  $u \in L^2(\mu_n)$  is defined by

$$(2.9) \quad \mathcal{L}^k u(x) = \frac{1}{n} \left( \frac{n\alpha_m}{k} \right)^{1+2/m} \sum_{j=1}^n w_{x_j x}^{r_k(x_j, x)} (u(x) - u(x_j)),$$

where  $w_{x_j x}^{r_k(x_j, x)}$  has the same meaning as in (2.2). As in the  $\varepsilon$ -graph case, we will assume for the most part that  $x$  is one of the elements in  $X$  and hence  $\mathcal{L}^k$  is interpreted as an operator from  $L^2(\mu_n)$  into  $L^2(\mu_n)$ . We list the eigenvalues (repeated according to multiplicity) of  $\mathcal{L}^k$  as

$$0 = \lambda_1^k \leq \lambda_2^k \leq \dots \leq \lambda_n^k,$$

and define an associated graph Dirichlet energy by

$$(2.10) \quad b_k(u) := \frac{1}{n^2} \left( \frac{n\alpha_m}{k} \right)^{1+2/m} \sum_{i,j} w_{x_i x_j}^{r_k(x_i, x_j)} (u(x_i) - u(x_j))^2 = 2 \langle \mathcal{L}^k u, u \rangle_{L^2(\mu_n)}, \quad u \in L^2(\mu_n).$$

This functional can be used to define the eigenvalues of  $\mathcal{L}^k$  via the variational formula

$$(2.11) \quad \lambda_l^k = \frac{1}{2} \min_{S \in \mathfrak{S}_l} \max_{u \in S \setminus \{0\}} \frac{b_k(u)}{\|u\|_{L^2(\mu_n)}^2}.$$

*Remark 2.2.* Notice that the rescaling factor  $\left(\frac{n\alpha_m}{k}\right)^{1+2/m}$  in (2.9) is equal to  $1/r^{m+2}$  where  $r$  is the radius of an  $m$ -dimensional Euclidean ball with volume  $k/n$ . This is the same type of rescaling factor that appears in the definition of the  $\varepsilon$ -graph Laplacian  $\mathcal{L}^\varepsilon$  in (2.3).

## 2.2. Convergence of eigenvalues.

**2.2.1.  $\varepsilon$ -graph.** In our first main result we establish error bounds between the eigenvalues  $\lambda_l^\varepsilon$  of the graph Laplacian  $\mathcal{L}^\varepsilon$ , and the eigenvalues  $\lambda_l$  of a differential operator  $\Delta_\rho$  that for smooth functions  $f : \mathcal{M} \rightarrow \mathbb{R}$  is defined as

$$(2.12) \quad \Delta_\rho f := -\frac{1}{2\rho} \operatorname{div}(\rho^2 \nabla f).$$

In the above  $\operatorname{div}$  stands for the divergence operator on  $\mathcal{M}$ , and  $\nabla$  for the gradient in  $\mathcal{M}$ . It turns out that  $\Delta_\rho$  is a positive semi definite operator with a nice point spectrum so that in particular its eigenvalues can be listed (repeated according to multiplicity) in increasing order as

$$0 \leq \lambda_1 \leq \lambda_2 \leq \dots$$

In addition, each eigenvalue has finite multiplicity, the sequence of eigenvalues grows to infinity, and it is possible to find an orthonormal basis for  $L^2(\mu)$  consisting of eigenfunctions, i.e. functions  $f$  that satisfy the equation  $\Delta_\rho f = \lambda f$ .

The eigenvalues of  $\Delta_\rho$  can be described variationally in terms of the Dirichlet energy

$$D_2(f) := \begin{cases} \int_{\mathcal{M}} |\nabla f(x)|^2 \rho^2(x) d\operatorname{Vol}_{\mathcal{M}}(x), & \text{if } f \in H^1(\mu) \\ +\infty, & \text{otherwise,} \end{cases}$$

where in the above  $H^1(\mu)$  is the space of functions with a weak gradient in  $L^2(\mu)$ . Indeed,

$$(2.13) \quad \lambda_l = \frac{1}{2} \min_{S \in \mathfrak{S}_l} \max_{f \in S \setminus \{0\}} \frac{D_2(f)}{\|f\|_{L^2(\mu)}^2}$$

where in the above  $\mathfrak{S}_l$  is the set of linear subspaces of  $L^2(\mu)$  of dimension  $l$ . The minimum is reached when  $S$  is taken to be the span of the first  $l$  eigenfunctions of  $\Delta_\rho$ . We also notice that when  $f$  is regular enough (in particular if  $f \in H^2(\mu)$ ) then  $D_2(f)$  coincides with  $2 \langle \Delta_\rho f, f \rangle_{L^2(\mu)}$  thanks to the integration by parts formula.

Throughout the paper we will state our probabilistic estimates in terms of parameters  $\varepsilon$  (the graph connectivity),  $n$  (the number of data points), and two parameters  $\theta, \tilde{\delta}$  that describe the error of discrete-to-continuum approximation. At this stage we offer a simple description of the meaning of the parameters  $\tilde{\delta}$  and  $\theta$ . We construct a probability distribution with density  $\tilde{\rho}_n$  mediating between the empirical measure  $\mu_n$  and the ground-truth distribution  $\mu$ . This density is a piecewise constant function defined over regions in  $\mathcal{M}$  of diameter  $\sim \tilde{\delta}$ , and the parameter  $\tilde{\delta}$  will describe how close this intermediate measure is from the empirical measure  $\mu_n$  in a suitable optimal transport sense. Since we will also have to guarantee that  $\tilde{\rho}_n$  is close enough to the original  $\rho$  in a suitable sense, we will not be able to select a very small value of  $\tilde{\delta}$  to get the approximation to  $\rho$  with very high probability. The parameter  $\theta$  will allow us to control the probability that the constructed  $\tilde{\rho}_n$  satisfies the desired properties:  $\tilde{\rho}_n$  is close to  $\mu_n$ , and at the same time it is close to the original  $\rho$  in a uniform sense; see Proposition 2.12 for more details. The introduction of this intermediate measure substitutes the direct discrete

to continuum comparison that other papers in the literature have considered to study similar problems. This modification is what allows for less restrictive assumptions on  $\varepsilon$ , just as in [13]. The following are technical “smallness” assumptions on our parameters that guarantee that we have entered the regime where the statements of our theorems are meaningful.

**Assumption 2.3.** *In the  $\varepsilon$ -graph setting, for  $\theta, \tilde{\delta}$  and  $\varepsilon$  we assume*

(1)  $\varepsilon$  is small enough and in particular satisfies

$$(2.14) \quad 2\varepsilon < \min\{1, i_0, K^{-1/2}, R/2\} =: 2\varepsilon_{\mathcal{M}}.$$

where  $i_0, K, R$  are geometric quantities defined in Section 3.1.

(2)  $\tilde{\delta} \leq \frac{1}{4}\varepsilon$ .

(3)  $\tilde{\delta}$  is larger than  $\frac{1}{n^{1/m}}$ .

(4)  $C(\theta + \tilde{\delta}) \leq \frac{\rho_{\min}}{2}$ , where  $C$  is a constant that depends only on dimension  $m$ ,  $\rho$ , and on geometric quantities of the manifold  $\mathcal{M}$ , such as the injectivity radius, sectional curvature, or reach.

In the  $k$ -NN setting, for  $\theta, \tilde{\delta}$  and  $k$  we assume

(1)  $(k/n)^{1/m}$  is small enough and in particular satisfies

$$2(k/n)^{1/m} < C_{\rho} \min\{1, i_0, K^{-1/2}, R/2\} =: 2\varepsilon_{\mathcal{M}}.$$

where  $C_{\rho}$  is a constant that depends on the density  $\rho$ .

(2)  $\tilde{\delta} \leq c_{\rho}(k/n)^{1/m}$ .

(3)  $\tilde{\delta}$  is larger than  $\frac{1}{n^{1/m}}$ .

(4)  $C(\theta + \tilde{\delta}) \leq \frac{\rho_{\min}}{2}$ , where  $C$  is a constant that depends only on dimension  $m$ ,  $\rho$ , and on geometric quantities of the manifold  $\mathcal{M}$ , such as the injectivity radius, sectional curvature, or reach.

We are ready to state our first main result.

**Theorem 2.4** (Rate of convergence for eigenvalues). *Suppose that  $\mathcal{M}, \mu$  satisfy the assumptions from Section 2. Suppose that the quantities  $\tilde{\delta}, \theta, \varepsilon$  satisfy Assumptions 2.3. Then:*

(1) *For every  $l \in \mathbb{N}$ , there is a constant  $c_l$  such that if*

$$\sqrt{\lambda_l}\varepsilon + C(\theta + \tilde{\delta}) \leq c_l,$$

*then with probability at least  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m)$ ,*

$$|\lambda_l^{\varepsilon} - \sigma_{\eta}\lambda_l| \leq C \left( \varepsilon(\sqrt{\lambda_l} + 1) + \frac{\tilde{\delta}}{\varepsilon} + \theta \right) \lambda_l.$$

(2) *Moreover, for every  $l \in \mathbb{N}$  there is a constant  $c_l$  such that if*

$$\left( \varepsilon\sqrt{\lambda_l} + \frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta \right) \lambda_l \leq c_l$$

*then, with probability greater than  $1 - 2Cn \exp(-cn\varepsilon^{m+4}) - Cn \exp(-cn\theta^2\tilde{\delta}^m)$  we have*

$$|\lambda_l^{\varepsilon} - \sigma_{\eta}\lambda_l| \leq C\varepsilon.$$

Let us pause for a moment and discuss the content of Theorem 2.4. A consequence of the first part of the theorem is that as long as  $\varepsilon \gg \left(\frac{\log(n)}{n}\right)^{1/m}$ , then it is possible to pick  $\tilde{\delta}$  and  $\theta$  according to  $\theta = \sqrt{\frac{(\alpha+1)\log(n)}{Cn\tilde{\delta}^m}}$  for some large enough  $\alpha$  in such a way that as  $n$  grows, both  $\theta$  and  $\tilde{\delta}$  converge to zero, and with probability one

$$|\lambda_l^{\varepsilon} - \sigma_{\eta}\lambda_l| \leq C(\tilde{\delta} + \theta) \rightarrow 0.$$

That is, we can make the error of approximation of eigenvalues converge to zero as  $n \rightarrow \infty$  provided that  $\varepsilon \gg \left(\frac{\log(n)}{n}\right)^{1/m}$ . This result is valid for all  $m \geq 1$ , removing in this way an extra logarithmic factor that was present in the results from [22] when the dimension was  $m = 2$ ; notice also that the results in [22] are stated in the case  $m \geq 2$  only, but as we show throughout our paper there is no reason to exclude the case  $m = 1$  from the convergence results. Of course, the first part of the theorem gives more than just asymptotic convergence and we do get rates. We would like to note that the removal of this logarithmic factor has been done recently in [40] and [13] for other, related, variational problems on graphs. Here we remove the extra logarithmic factors while obtaining rates of convergence. Notice, however, that our estimates would predict that the rates of convergence become quite slow as we get close to the scaling  $\left(\frac{\log(n)}{n}\right)^{1/m}$ , which is the connectivity threshold of the graph. The rates of convergence degrade as we approach the connectivity threshold because the graph is irregular on the smallest length scale of nearest neighbors. However, we expect that on a larger macroscopic scale, the irregularities will average out, and it may be possible to extend the linear  $O(\varepsilon)$  rate to the range  $\left(\frac{\log(n)}{n}\right)^{1/m} \ll \varepsilon \ll \left(\frac{\log(n)}{n}\right)^{1/(m+4)}$ . We expect this to require deeper tools and techniques from the theory of stochastic homogenization, and leave this to future work.

The second part of the theorem in particular implies that if  $\varepsilon$  is larger than  $\left(\frac{\log(n)}{n}\right)^{\frac{1}{m+4}}$ , then with high probability the rate of convergence of eigenvalues and eigenvectors of  $\mathcal{L}^\varepsilon$  towards those of  $\Delta_\rho$  is linear in  $\varepsilon$ . In particular, for any  $p \geq 1$  we can make the choice

$$\varepsilon = c_p \left(\frac{\log(n)}{n}\right)^{\frac{1}{m+4}}$$

for a constant  $c_p$  to obtain that the eigenvalue convergence rate

$$|\lambda_\ell^\varepsilon - \sigma_\eta \lambda_l| \leq C c_p \left(\frac{\log(n)}{n}\right)^{\frac{1}{m+4}},$$

holds with probability at least  $1 - cn^{-p}$ . A similar comment holds for Theorem 2.7 (2). We believe that for eigenvectors the linear convergence rate is the best possible for a generic manifold  $\mathcal{M}$  and kernel  $\eta$ . In numerical analysis of PDEs, one normally sees second order  $O(\varepsilon^2)$  convergence rates for the spectrum of the Laplacian on a uniform grid in Euclidean space, due to using *symmetric* stencils for the Laplacian (see, e.g., [38]), where first order error terms exactly cancel out, and the lack of discretization errors due to the curvature of the manifold. In our manifold setting, the curvature of the manifold arises in the first order error terms and prevents the scheme from attaining second order accuracy. We also note that the numerical scheme provided by the graph Laplacian (2.3) is not a symmetric stencil (i.e., the neighbors of  $x$  are not of the form  $x + v$  and  $x - v$ ), and so the first order terms cancel out only on average, after random fluctuations are accounted for, which necessitates the larger length scale  $\varepsilon \gg \left(\frac{\log(n)}{n}\right)^{1/(m+2)}$  for pointwise consistency. In the flat Euclidean setting, it may be possible to improve convergence rates to  $O(\varepsilon^2)$  with a more severe length scale restriction  $\varepsilon \gg \left(\frac{\log(n)}{n}\right)^{1/(m+6)}$ . Regarding convergence rates for eigenvalues, we actually believe that they may be faster than the convergence rates for eigenfunctions. This is because eigenvalues may be defined by computing the Dirichlet energy of a normalized eigenvector of the graph Laplacian, and thus is computed by taking a double sum which in principle may homogenize better. This is by no means a proof and only reveals a simple intuition.

Finally, we would like to remark that we can trace the dependence of all the constants appearing in Theorem 2.4 on the different parameters of the problem, but we have decided not to state them explicitly to immensely facilitate our presentation.

2.2.2. *k-NN graph.* In the  $k$ -NN graph case we show that the eigenvalues of  $\mathcal{L}^k$  converge towards the eigenvalues of the operator  $\Delta_\rho^{NN}$  defined for smooth functions  $f : \mathcal{M} \rightarrow \mathbb{R}$  according to

$$(2.15) \quad \Delta_\rho^{NN} f := -\frac{1}{2\rho} \operatorname{div}(\rho^{1-2/m} \nabla f).$$

This operator has similar spectral properties as  $\Delta_\rho$ . In particular, if we list its eigenvalues as

$$0 \leq \lambda_1 \leq \lambda_2 \leq \dots,$$

they can be written as

$$\lambda_l = \frac{1}{2} \min_{S \in \mathfrak{S}_l} \max_{f \in S \setminus \{0\}} \frac{D_{1-2/m}(f)}{\|f\|_{L^2(\mu)}^2}$$

where now the associated Dirichlet energy  $D_{1-2/m}$  takes the form

$$(2.16) \quad D_{1-2/m}(f) := \begin{cases} \int_{\mathcal{M}} |\nabla f(x)|^2 \rho^{1-2/m}(x) d\operatorname{Vol}_{\mathcal{M}}(x) & \text{if } f \in H^1(\mu) \\ +\infty & \text{otherwise.} \end{cases}$$

Following a remark in [21], it is possible to show that the spectrum of  $\mathcal{L}^k$  converges as  $n \rightarrow \infty$  towards the spectrum of  $\Delta_\rho^{NN}$  provided that  $k$  scales like

$$(\log(n))^{mpm} \ll k \ll n.$$

No rates are provided in [27] as the convergence is deduced using the notion of  $\Gamma$ -convergence. Besides enlarging the set of admissible values of  $k$  for which we can prove the spectral convergence, we establish the following convergence rates. They are analogous to the ones stated in Theorem 2.4 for  $\varepsilon$ -graphs.

**Theorem 2.5** (Rate of convergence for eigenvalues). *Suppose that  $\mathcal{M}, \mu$  satisfy the assumptions from Section 2. Suppose that the quantities  $\tilde{\delta}, \theta, k$  satisfy Assumptions 2.3. Then:*

(1) *For every  $l \in \mathbb{N}$ , there is a constant  $c_l$  such that if*

$$\sqrt{\lambda_l} (k/n)^{1/m} + C(\theta + \tilde{\delta}) \leq c_l,$$

*then, with probability at least  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m)$ , we have*

$$|\lambda_l^k - \sigma_\eta \lambda_l| \leq +C \left( \left( \frac{k}{n} \right)^{1/m} (\sqrt{\lambda_l} + 1) + \tilde{\delta} \left( \frac{n}{k} \right)^{1/m} + \theta \right) \lambda_l.$$

(2) *Moreover, for every  $l \in \mathbb{N}$  there is a constant  $c_l$  such that if*

$$\left( \left( \frac{k}{n} \right)^{1/m} \sqrt{\lambda_l} + \tilde{\delta} \left( \frac{n}{k} \right)^{1/m} + \theta \right) \lambda_l \leq c_l$$

*then, with probability greater than  $1 - Cn \exp(-c(k/n)^{4/m}k) - Cn \exp(-Cn\theta^2\tilde{\delta}^m)$  we have*

$$|\lambda_l^k - \sigma_\eta \lambda_l| \leq C(k/n)^{1/m}.$$

Let us note that, as in the case of the  $\varepsilon$ -graph, we can simplify the statement of part (2). For any  $p \geq 1$  we can make the choice

$$k = c_p \log(n)^{\frac{m}{m+4}} n^{\frac{4}{m+4}}$$

for a constant  $c_p$  to obtain that the eigenvalue convergence rate

$$|\lambda_l^k - \sigma_\eta \lambda_l| \leq C c_p \left( \frac{\log(n)}{n} \right)^{\frac{1}{m+4}},$$

holds with probability at least  $1 - cn^{-p}$ . A similar comment holds for the convergence rate in Theorem 2.9 (2).

*Remark 2.6.* Theorems 2.4 and 2.5 have a similar form that becomes more evident when in Theorem 2.4 we replace  $\lambda_l^\varepsilon$  by  $\lambda_l^k$ ,  $\lambda_l$  (of  $(\Delta_\rho)$  by  $\lambda_l$  (of  $\Delta_\rho^{NN}$ ), and  $\varepsilon$  by  $(k/n)^{1/m}$  to recover a statement similar to that of Theorem 2.5. Notice that  $(k/n)^{1/m}$  is the natural counterpart of  $\varepsilon$ . Notice, on the other hand, that despite the similarities in these two settings,  $k$ -NN graphs adjust to the data density in a different way than their  $\varepsilon$ -graph counterparts, and in particular, the limiting operators in each case are different. Other works, posterior to the first version of this paper, like that in [15], have considered more general normalizations for  $k$ -NN graphs in order to induce different ways in which data density affects the behavior of data analysis algorithms.

**2.3. Convergence of eigenvectors.** We also establish convergence rates for the eigenvectors of  $\mathcal{L}^\varepsilon$  towards eigenfunctions of  $\Delta_\rho$ , as well as convergence rates of eigenvectors of  $\mathcal{L}^k$  towards eigenfunctions of  $\Delta_\rho^{NN}$ . Naturally, we first need to specify the sense in which the convergence is to take place. To describe this, let us first notice that the eigenfunctions of both  $\Delta_\rho$  and  $\Delta_\rho^{NN}$  are continuous functions, so that pointwise evaluation of eigenfunctions makes perfect sense. In particular, we can restrict a given eigenfunction  $f$  to  $X$  and the quantity  $\|u - f\|_{L^2(\mu_n)}^2$  would be well defined for every  $u \in L^2(\mu_n)$ . Furthermore, if  $f$  solves the elliptic equation

$$\Delta_\rho f = \lambda f$$

for some  $\lambda > 0$ , then, by the regularity theory of elliptic operators [29] (which can be lifted from the Euclidean setting to the curved manifold setting given the assumptions that we made on  $\mathcal{M}$  at the beginning of Section 2) and the regularity assumption on the density  $\rho \in C^2(\mathcal{M})$ , it follows that  $f$  is at least a  $C^3(\mathcal{M})$  function. That is, its first, second and third derivatives are continuous, and given the compactness of  $\mathcal{M}$  also bounded. The same regularity holds for eigenfunctions of  $\Delta_\rho^{NN}$ . Said regularity will be needed in order to apply the pointwise consistency results from Theorem 3.3 (for the  $\varepsilon$ -graph case) and Theorem 3.7 (for the  $k$ -NN case) to a given eigenfunction  $f$ .

We have the following convergence result for  $\varepsilon$ -graphs.

**Theorem 2.7** (Rate of convergence for eigenvectors). *Suppose that  $\mathcal{M}, \mu$  satisfy the assumptions from Section 2. Suppose that the quantities  $\tilde{\delta}, \theta, \varepsilon$  satisfy Assumptions 2.3. Then:*

(1) *For every  $l \in \mathbb{N}$  there is a constant  $c_l$  such that if*

$$\left( \varepsilon \sqrt{\lambda_l} + \frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta \right) \lambda_l \leq c_l,$$

*then with probability at least  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m)$ , for every  $u_l$  normalized eigenvector of  $\mathcal{L}^\varepsilon$  with eigenvalue  $\lambda_l^\varepsilon$ , there is a normalized eigenfunction  $f_l$  of  $\Delta_\rho$  with eigenvalue  $\lambda_l$  such that*

$$\|u_l - f_l\|_{L^2(\mu_n)} \leq \left( C_l \left( \frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta \right) \right)^{1/2} + C_l \tilde{\delta}.$$

(2) *For every  $l \in \mathbb{N}$  there is a constant  $c_l$  such that if*

$$\left( \varepsilon \sqrt{\lambda_l} + \frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta \right) \lambda_l \leq c_l,$$

*then with probability at least  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m) - Cn \exp(-Cn\varepsilon^{m+4})$ , for every  $u_l$  normalized eigenvector of  $\mathcal{L}^\varepsilon$  with eigenvalue  $\lambda_l^\varepsilon$ , there is a normalized eigenfunction  $f_l$  of  $\Delta_\rho$  with eigenvalue  $\lambda_l$  such that*

$$\|u_l - f_l\|_{L^2(\mu_n)} \leq C_l \varepsilon.$$

*Remark 2.8.* The constant  $C_l$  in Theorem 2.7 (ii) depends on the  $C^3$  norm of all normalized eigenfunctions  $f_l$  of  $\Delta_\rho$  with eigenvalue  $\lambda_l$ , since our proof uses the pointwise consistency of graph Laplacians. The constant  $C_l$  is also inversely proportional to the spectral gap  $\gamma_l$  defined as

$$\gamma_l = \min\{|\lambda_l - \lambda_j| : \lambda_l \neq \lambda_j\}.$$

A similar remark holds for the analogous result for  $k$ -nearest neighbor graphs given in Theorem 2.9.

As in Theorem 2.4, the main difference between the error estimates presented in Theorem 2.7 is the regimes of  $\varepsilon$  for which they are meaningful. In general we may use energy arguments as in [7] to obtain convergence rates that scale like the square root of the rates of convergence for eigenvalues. On the other hand, in the regime  $\varepsilon \gg \frac{\log(n)}{n^{1/m+4}}$  we can show that eigenfunctions of the graph Laplacian converge to eigenfunctions of  $\Delta_\rho$  linearly in the connectivity length-scale  $\varepsilon$ , coinciding in this way with the rates for eigenvalues and also with the rates of pointwise convergence (see Theorem 3.3).

The analogous estimates for  $k$ -NN graphs are the following.

**Theorem 2.9** (Rate of convergence for eigenvectors). *Suppose that  $\mathcal{M}, \mu$  satisfy the assumptions from Section 2. Suppose that the quantities  $\tilde{\delta}, \theta, k$  satisfy Assumptions 2.3. Then:*

(1) *For every  $l \in \mathbb{N}$  there is a constant  $c_l$  such that if*

$$((k/n)^{1/m} \sqrt{\lambda_l} + \tilde{\delta}(n/k)^{1/m} + \theta) \lambda_l \leq c_l,$$

*then with probability at least  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m)$ , for every  $u_l$  normalized eigenvector of  $\mathcal{L}^k$  with eigenvalue  $\lambda_l^k$ , there is a normalized eigenfunction  $f_l$  of  $\Delta_\rho^{NN}$  with eigenvalue  $\lambda_l$  such that*

$$\|u_l - f_l\|_{L^2(\mu_n)} \leq \left( C_l \left( \tilde{\delta}(n/k)^{1/m} + (k/n)^{1/m} + \theta \right) \right)^{1/2} + C_l \tilde{\delta}.$$

(2) *For every  $l \in \mathbb{N}$  there is a constant  $c_l$  such that if*

$$((k/n)^{1/m} \sqrt{\lambda_l} + \tilde{\delta}(n/k)^{1/m} + \theta) \lambda_l \leq c_l,$$

*then with probability at least  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m) - Cn \exp(-c(k/n)^{4/m}k)$ , for every  $u_l$  normalized eigenvector of  $\mathcal{L}^k$  with eigenvalue  $\lambda_l^k$ , there is a normalized eigenfunction  $f_l$  of  $\Delta_\rho^{NN}$  with eigenvalue  $\lambda_l$  such that*

$$\|u_l - f_l\|_{L^2(\mu_n)} \leq C_l (k/n)^{1/m}.$$

There are other ways to study the convergence of eigenvectors of graph Laplacians. As we will see in Section 5, Theorems 2.7 and 2.9, and the regularity of eigenfunctions of  $\Delta_\rho$  and  $\Delta_\rho^{NN}$  imply the convergence of eigenvectors in the so called  $TL^2$ -sense (see [22]). This notion of convergence makes explicit a way to compare different spectral embeddings that form the basis of algorithms like spectral clustering [51]. To be more precise let  $L \in \mathbb{N}$ . With the first  $L$  eigenvectors  $u_1, \dots, u_L$  of  $\mathcal{L}^\varepsilon$  (or  $\mathcal{L}^k$ ) one can construct an embedding of the data set

$$F_n : \{x_1, \dots, x_n\} \rightarrow \mathbb{R}^L$$

$$F_n(x_i) = \begin{pmatrix} u_1(x_i) \\ \vdots \\ u_L(x_i) \end{pmatrix}.$$

The resulting set of points can be represented by the measure  $F_{n\sharp}\mu_n$ , i.e., the push-forward of the original empirical measure  $\mu_n$  by the graph Laplacian embedding  $F_n$ . A natural question to ask in this setting is: how far apart is  $F_{n\sharp}\mu_n$  from the measure  $F\sharp\mu$ , where  $F$  is the continuum Laplacian embedding:

$$F : \mathcal{M} \rightarrow \mathbb{R}^L$$

$$F(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_L(x) \end{pmatrix},$$

constructed from eigenfunctions  $f_1, \dots, f_L$  of the continuum operator  $\Delta_\rho$  (or  $\Delta_\rho^{NN}$ )? We answer this question in terms of the Wasserstein distance  $W_2$  between  $F_{n\sharp}\mu_n$  and  $F_\sharp\mu$ , which we recall is defined by

$$(2.17) \quad W_2(F_\sharp\mu, F_{n\sharp}\mu_n) := \min_{\pi \in \Gamma(F_\sharp\mu, F_{n\sharp}\mu_n)} \int_{\mathbb{R}^L \times \mathbb{R}^L} |x - y|^2 d\pi(x, y),$$

where  $\Gamma(F_\sharp\mu, F_{n\sharp}\mu_n)$  denotes the set of couplings or transport plans between  $F_\sharp\mu$  and  $F_{n\sharp}\mu_n$ . This way of comparing the spectral embeddings was proposed in [23].

**Corollary 2.10.** *Suppose that  $\mathcal{M}, \mu$  satisfy the assumptions from Section 2.*

- (1) ( $\varepsilon$ -setting) *Suppose that the quantities  $\tilde{\delta}, \theta, \varepsilon$  are small enough as in 2 in Theorem 2.7. Then, with probability at least  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m) - Cn \exp(-Cn\varepsilon^{m+4})$ , we have*

$$W_2(F_\sharp\mu, F_{n\sharp}\mu_n) \leq C\varepsilon + CW_2(\mu, \mu_n),$$

*where  $F_n$  is the graph Laplacian embedding constructed from the first  $L$  eigenvectors of  $\mathcal{L}^\varepsilon$ , and  $F$  is the Laplacian embedding constructed using the eigenfunctions  $f_1, \dots, f_L$  of  $\Delta_\rho$  from Theorem 2.7.*

- (2) *Suppose that the quantities  $\tilde{\delta}, \theta, (k/n)^{1/m}$  are small enough as in 2 in Theorem 2.9. Then, with probability at least  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m) - Cn \exp(-C(k/n)^{4/m}k)$ , we have*

$$W_2(F_\sharp\mu, F_{n\sharp}\mu_n) \leq C(k/n)^{1/m} + CW_2(\mu, \mu_n),$$

*where  $F_n$  is the graph Laplacian embedding constructed from the first  $L$  eigenvectors of  $\mathcal{L}^k$ , and  $F$  is the Laplacian embedding constructed using the eigenfunctions  $f_1, \dots, f_L$  of  $\Delta_\rho^{NN}$  from Theorem 2.9.*

*Remark 2.11.* Probabilistic bounds for

$$W_2^2(\mu, \mu_n) = \min_{\pi \in \Gamma(\gamma, \tilde{\gamma})} \int_{\mathcal{M} \times \mathcal{M}} d\mathcal{M}(x, y)^2 d\pi(x, y)$$

can be easily derived using a localization argument (to deal with the curved manifold) and the concentration inequalities estimating the Wasserstein distance between empirical measures and their ground truth counterparts in the Euclidean setting (see [20] and a more recent treatment with improved constants and better scalability with respect to dimension in [37]). The localization argument is used simply to partition the manifold  $\mathcal{M}$  into a fixed number of regions that are bi-Lipschitz homeomorphic to a bounded domain in  $\mathbb{R}^m$ . One can then lift the concentration results in Euclidean space  $\mathbb{R}^m$  to the manifold. In particular,  $W_2(\mu, \mu_n)$  scales like  $\frac{1}{n^{1/m}}$ .

**2.4. Outline of proofs and discussion.** The proofs of our main theorems follow the same structure in both the  $\varepsilon$ -graph and  $k$ -NN settings. For this reason we outline our proofs only in the  $\varepsilon$ -graph setting and then simply state the modifications needed in order to cover the  $k$ -NN case.

Our first step is to establish some a priori (and non-optimal) rates of convergence for eigenvalues of  $\mathcal{L}^\varepsilon$  towards eigenvalues of  $\Delta_\rho$ . In principle, we could cite the results already proved in [22], but we have decided to prove our own a priori results, so that we have the opportunity to present a simpler approach that works for a larger set of values of  $\varepsilon$  than those covered in [22].

To discuss our construction, it is worth first reviewing the one in [7], which is a main influence for the construction in our work and the work in [22]. There the main idea is to construct maps

$$P: L^2(\mu) \rightarrow L^2(\mu_n), \quad \mathcal{I}: L^2(\mu_n) \rightarrow L^2(\mu)$$

that are almost isometries when restricted to functions of low Dirichlet energy (in discrete and continuum settings), and for which one has estimates of the form

$$b_\varepsilon(Pf) \leq (1 + e)\sigma_\eta D_2(f), \quad \forall f \in L^2(\mu),$$

and

$$(2.18) \quad \sigma_\eta D_2(\mathcal{I}u) \leq (1 + e)b_\varepsilon(u), \quad \forall u \in L^2(\mu_n),$$

where  $e$  is thought of as a small error term. When combined with the variational identities (2.5) and (2.13), these inequalities produce error bounds for the difference between eigenvalues of  $\mathcal{L}^\varepsilon$  and  $\Delta_\rho$ : one can upper bound  $\lambda_l^\varepsilon$  with  $\lambda_l + Ce$  choosing  $S$  in (2.13) to be the image under  $P$  of the span of the first  $l$  eigenvectors of  $\Delta_\rho$ , and likewise, one can upper bound  $\lambda_l$  by  $\lambda_l^\varepsilon + Ce$  choosing  $S$  in (2.5) to be the image under  $\mathcal{I}$  of the span of the first  $l$  eigenvectors of  $\mathcal{L}^\varepsilon$ . Now, the maps  $P$  and  $\mathcal{I}$  introduced in [7] are based on an  $\infty$ -optimal transport map between  $\mu$  and  $\mu_n$ . That is, they use a map  $T : \mathcal{M} \rightarrow \{x_1, \dots, x_n\}$  that pushes forward the measure  $\mu$  into the empirical measure  $\mu_n$ , and does so in such a way that it minimizes the quantity

$$\delta := \sup_{x \in \mathcal{M}} d_{\mathcal{M}}(x, T(x)),$$

over all possible such transport maps.  $T$  can be used to generate a tessellation  $U_1, \dots, U_n$  of  $\mathcal{M}$  (here  $U_i := T^{-1}(\{x_i\})$ ), with the property that all cells  $U_i$  have  $\mu$ -measure equal to  $1/n$  and diameter bounded by  $2\delta$ . The maps  $P$  and  $\mathcal{I}$  are then defined according to

$$(Pf)(x_i) := n \cdot \int_{U_i} f(x) \rho(x) dx, \quad f \in L^2(\mu).$$

and

$$\mathcal{I}u := \Lambda_{\varepsilon-2\delta} P^* u, \quad u \in L^2(\mu_n).$$

where  $P^*$  is the adjoint of  $P$  (i.e. composition with  $T$ ) and  $\Lambda_{\varepsilon-2\delta}$  is a convolution operator with respect to a very carefully chosen kernel (chosen conveniently to guarantee the energy inequality (2.18)) with bandwidth  $\varepsilon - 2\delta$ . Notice that in order to get convergence rates for the spectrum it is crucial that the leading term on the right hand side of (2.18) is  $b_\varepsilon(u)$  (with no constants in front). In this construction it is also important to have  $\varepsilon > 2\delta$ . Now, as proved in [22], the scaling of  $\delta = d_\infty(\mu, \mu_n)$  in terms of  $n$  is

$$(2.19) \quad \delta \sim C \begin{cases} \frac{\log(n)^{3/4}}{n^{1/2}} & m = 2 \\ \frac{\log(n)^{1/m}}{n^{1/m}} & m \geq 3, \end{cases}$$

which means that in dimension  $m = 2$ ,  $\varepsilon$  is forced to be larger than  $\frac{\log(n)^{3/4}}{n^{1/2}}$ . The extra logarithmic factor with respect to the connectivity threshold has been shown to be unnecessary in order to establish consistency of closely related problems as illustrated in [13] and [40]. Here we also remove the extra logarithmic factor, offering also quantitative rates of convergence; our analysis also covers the case  $m = 1$  which was not explicitly included in some of the previous works in the literature. For that purpose, we use modified discretization and interpolation maps  $\tilde{P}, \tilde{\mathcal{I}}$  constructed as before, but based on a map  $\tilde{T}$  pushing forward a conveniently chosen measure  $\tilde{\mu}_n$  with density  $\tilde{\rho}_n$  (uniformly close to  $\rho$ ) into the empirical measure  $\mu_n$ . The idea is that  $\tilde{\mu}_n$  can be chosen to be closer to  $\mu_n$  in the  $\infty$ -OT sense than  $\mu$  itself. More precisely we have the following proposition which is proved in the Appendix.

**Proposition 2.12.** *Suppose that  $\mathcal{M}, \mu$  satisfy the assumptions from Section 2. Suppose that the quantities  $\tilde{\delta}, \theta, \varepsilon$  satisfy Assumptions 2.3. Then, with probability greater than  $1 - n \exp(-Cn\theta^2\tilde{\delta}^m)$ , there exists a probability measure  $\tilde{\mu}_n$  with density function  $\tilde{\rho}_n : \mathcal{M} \rightarrow \mathbb{R}$  such that*

$$\min_{T \sharp \tilde{\mu}_n = \mu_n} \sup_{x \in \mathcal{M}} d_{\mathcal{M}}(x, T(x)) \leq \tilde{\delta},$$

and such that

$$\|\rho - \tilde{\rho}_n\|_{L^\infty(\mu)} \leq C(\theta + \tilde{\delta}).$$

We will use  $\tilde{T}$  to denote an optimal transport map.

*Remark 2.13.* Suppose that in Proposition 2.12 we take  $\theta := \sqrt{\frac{(\alpha+1)\log(n)}{Cn\tilde{\delta}^m}}$  for some  $\alpha > 1$ . Then it follows that with probability at least  $1 - \frac{1}{n^\alpha}$  there exists a measure  $\tilde{\mu}_n$  such that

$$\min_{T \# \tilde{\mu}_n = \mu_n} \sup_{x \in \mathcal{M}} d_{\mathcal{M}}(x, T(x)) \leq \tilde{\delta}$$

and

$$\|\rho - \tilde{\rho}_n\|_{L^\infty(\mu)} \leq C \left( \sqrt{\frac{(\alpha+1)\log(n)}{Cn\tilde{\delta}^m}} + \tilde{\delta} \right).$$

In particular, if  $\tilde{\delta}$  is set to scale like  $\frac{\log(n)^{1/m}}{n^{1/m}} \ll \tilde{\delta} \ll 1$ , we can make  $\|\rho - \tilde{\rho}_n\|_{L^\infty(\mu)}$  arbitrarily small.

After we derive some properties of the new maps  $\tilde{P}$  and  $\tilde{\mathcal{I}}$  we will be able to prove the first part of Theorem 2.4, as well as the first part of Theorem 2.7 following the proof scheme in [22]. We will also be able to prove the second part of Theorem 2.9. For this we follow some of the steps in the proof of the Davis-Kahan theorem. We will need to use *pointwise consistency* for the graph Laplacian together with our a priori estimates for the error of approximation of eigenvalues in order to isolate eigenvalues of  $\Delta_\rho$  and match them with eigenvalues of  $\mathcal{L}^\varepsilon$ . We notice that the pointwise consistency for  $\varepsilon$ -graphs has been established in previous works [32]. For completeness, we give a rather simple proof of consistency in Section 3.3. To apply the pointwise consistency we require regularity of the continuum eigenfunctions (see the discussion at the beginning of Section 2.3), which is an ingredient that has not been utilized in previous works on spectral convergence.

Finally, to improve the rates of convergence for eigenvalues (i.e. to obtain the second part of Theorem 2.4) we will use the error rates for the convergence of eigenvectors as follows. Let us fix  $l \in \mathbb{N}$  and let  $u$  be a normalized eigenvector of  $\mathcal{L}^\varepsilon$  with corresponding eigenvalue  $\lambda_l^\varepsilon$ . We know a priori that we can find  $f \in L^2(\mu)$  a normalized eigenfunction of  $\Delta_\rho$  with eigenvalue  $\lambda_l$ , that is close to  $u$ , so that in particular

$$(2.20) \quad \langle u, f \rangle_{L^2(\mu_n)} \geq c > 0,$$

where  $c$  is independent of  $\varepsilon$ , or  $n$ ; in the above formula we interpret  $f$  as the restriction of  $f : \mathcal{M} \rightarrow \mathbb{R}$  to  $X$  (a well defined operation given the continuity of eigenfunctions of  $\Delta_\rho$ ). We will then arrive to an inequality of the form

$$|\lambda_l^\varepsilon - \sigma_\eta \lambda_l| = \frac{|\langle u, \mathcal{L}^\varepsilon f - \sigma_\eta \Delta_\rho f \rangle|_{L^2(\mu_n)}}{|\langle u, f \rangle|_{L^2(\mu_n)}} \leq \frac{1}{c} \|(\mathcal{L}^\varepsilon - \sigma_\eta \Delta_\rho) f\|_{L^2(\mu_n)}.$$

The above computations show that in order to estimate  $|\lambda_l^\varepsilon - \lambda_l|$  it suffices to use the pointwise consistency of graph Laplacians applied to the function  $f$ .

For the most part, the proof strategy for Theorems 2.5 and 2.9 is similar to the one described before. However, there are two important modifications we need to make. First, the pointwise estimates for  $\mathcal{L}^k f(x_i) - \sigma_\eta \Delta_\rho^{NN} f(x_i)$  require new technical computations which are presented in detail in Section 3.4. These results are new in the literature. Second, regarding the relevant a priori estimates, we must first construct a new interpolation map  $\tilde{\mathcal{I}}$  as now the connectivity length-scale changes in space. Once the main properties of the new map  $\tilde{\mathcal{I}}$  have been established our desired results will follow in the exact same way as in the  $\varepsilon$ -graph case.

We suspect that our eigenvector and eigenvalue rate of  $O(n^{-1/(m+4)})$  is close to optimal. In the general manifold setting, we suspect  $O(n^{-1/m})$  to be the optimal convergence rate for eigenvectors, since this represents the resolution of the point cloud (i.e., the typical inter-point distance). It would be interesting to establish lower bounds for all convergence rates in this paper. We point out that some parts of our proof for the eigenvalue convergence rate can

produce lower bounds. For example, in the  $\varepsilon$ -graph setting, we look forward to (4.17) in the eigenvalue convergence rate proof, which yields

$$|\lambda_l^\varepsilon - \sigma_\eta \lambda_l| \geq C_l \sup_{\substack{f \in S(\Delta_\rho, \lambda_l) \\ u \in S(\mathcal{L}^\varepsilon, \lambda_l^\varepsilon)}} |\langle u, \mathcal{L}^\varepsilon f - \sigma_\eta \Delta_\rho f \rangle_{L^2(\mu_n)}|$$

with probability at least  $1 - 2 \exp(-cn)$ , where  $S(\Delta_\rho, \lambda)$  is the set of unit norm eigenfunctions of  $\Delta_\rho$  with eigenvalue  $\lambda$ , and  $S(\mathcal{L}^\varepsilon, \lambda)$  is the corresponding set for  $\mathcal{L}^\varepsilon$ . Obtaining lower bounds on the right hand side would prove lower bounds on the eigenvalue convergence rate. This appears to be a hard problem that will involve tools from the field of stochastic homogenization, since the pointwise consistency errors may average out against  $u$  to something smaller than the upper bound given by the application of Cauchy-Schwartz that we made above. We plan to explore lower bounds in a future work.

### 3. POINTWISE CONSISTENCY OF GRAPH LAPLACIANS

In this section we prove pointwise consistency with a linear rate for our two constructions of the graph Laplacian. The  $\varepsilon$ -graph Laplacian is considered in Section 3.3, while the undirected  $k$ -NN Laplacian is considered in Section 3.4. For the  $\varepsilon$ -graph Laplacian, the linear rate was established earlier in [32]; we give a simpler proof for completeness. Consistency for  $k$ -NN graph Laplacians was studied in [50], but the methods used were unable to establish any convergence rates.

Before giving the pointwise consistency proofs, we review some differential geometry in Section 3.1, and concentration of measure in Section 3.2.

**3.1. Differential geometry.** We first briefly review some basic results from differential geometry. We refer the reader to [17] for more details. We write  $B_{\mathcal{M}}(x, r) \subset \mathcal{M}$  to denote the geodesic ball in  $\mathcal{M}$  of radius  $r$  centered at  $x$ , while we use  $B(x, r)$  to denote Euclidean balls in  $\mathbb{R}^m$  or in  $\mathbb{R}^d$  depending on context. For each  $x \in \mathcal{M}$ ,  $\exp_x : T_x \mathcal{M} \rightarrow \mathcal{M}$  is the Riemannian exponential map. Let  $K$  be an upper bound on the absolute values of the sectional curvatures, let  $R$  be the reach of  $\mathcal{M}$ , and let  $i_0$  be a lower bound on the injectivity radius of  $\mathcal{M}$ . For any  $0 < r < \min\{i_0, K^{-1/2}\}$ ,  $\exp_x : B(0, r) \rightarrow \mathcal{M}$  is a diffeomorphism between the ball  $B(0, r) \subset T_x \mathcal{M}$  and the geodesic ball  $B_{\mathcal{M}}(x, r) \subset \mathcal{M}$ . Let us denote the Jacobian of  $\exp_x$  at  $v \in B(0, r) \subset T_x \mathcal{M}$  by  $J_x(v)$ . By the Rauch Comparison Theorem

$$(3.1) \quad (1 + CK|v|^2)^{-1} \leq J_x(v) \leq 1 + CK|v|^2.$$

It follows that

$$(3.2) \quad |\text{Vol}_{\mathcal{M}}(B_{\mathcal{M}}(x, r)) - \alpha_m r^m| \leq CKr^{m+2}.$$

We also recall [22, Proposition 2]

$$(3.3) \quad |x - y| \leq d_{\mathcal{M}}(x, y) \leq |x - y| + \frac{8}{R^2}|x - y|^3,$$

provided  $|x - y| \leq R/2$ . In the above  $d_{\mathcal{M}}$  denotes the geodesic distance on  $\mathcal{M}$ .

Throughout this section we always assume  $n$  and  $\varepsilon$  satisfy Assumptions 2.3. We use  $C, c > 0$  to denote arbitrary constants that depend only on dimension  $m$ ,  $\rho$ , and on the geometric properties of the manifold  $\mathcal{M}$ , such as the injectivity radius or sectional curvature. We always have  $0 < c < 1$  and  $C \geq 1$ .

**3.2. Concentration of measure.** We now state a simple concentration inequality, which follows from Bernstein's inequality [6], that is particularly convenient for analysis of graph Laplacians.

**Lemma 3.1.** *Suppose that  $\mathcal{M}, \mu$  satisfy the assumptions from Section 2 and let  $x_1, \dots, x_n$  be samples from  $\mu$ . Let  $\psi : \mathcal{M} \rightarrow \mathbb{R}$  be bounded and Borel measurable. For  $x \in \mathcal{M}$  define*

$$\Psi = \sum_{|x_i - x| \leq \varepsilon} \psi(x_i).$$

Then for any  $\varepsilon^2 \leq \delta \leq 1$

$$(3.4) \quad \mathbb{P}(|\Psi - a| \geq C\rho_{max}\|\psi\|_\infty \delta n\varepsilon^m) \leq 2 \exp(-c\rho_{max}\delta^2 n\varepsilon^m),$$

where  $\|\psi\|_\infty = \|\psi\|_{L^\infty(B_{\mathcal{M}}(x, 2\varepsilon))}$  and

$$(3.5) \quad a = n \int_{B_{\mathcal{M}}(x, \varepsilon)} \psi(y) \rho(y) dVol_{\mathcal{M}}(y).$$

*Remark 3.2.* It is important to point out that  $a \neq \mathbb{E}[\Psi]$ , since the definition of  $\Psi$  uses the metric in the ambient space, while in the definition of  $a$  we integrate over *geodesic* balls. We also point out that the restriction  $\delta \geq \varepsilon^2$  allows us to ignore any effects of the curvature of the manifold.

*Proof.* The argument is similar to [8, Lemma 1], but we include the proof for completeness. Let  $Z_i = \mathbb{1}_{\{|x_i - x| \leq \varepsilon\}} \psi(x_i)$ . The Bernstein inequality applied to  $\sum_{i=1}^n Z_i$  yields

$$(3.6) \quad \mathbb{P}(|\Psi - \mathbb{E}[\Psi]| \geq nt) \leq 2 \exp\left(-\frac{nt^2}{2(\sigma^2 + \frac{1}{3}bt)}\right)$$

for any  $t > 0$ , where  $\sigma^2 = \text{Var}(Z_i)$ ,  $b > 0$  satisfies  $|Z_i - \mathbb{E}[Z_i]| \leq b$  almost surely, and

$$\mathbb{E}[\Psi] = n \int_{B(x, \varepsilon) \cap \mathcal{M}} \psi(y) \rho(y) dVol_{\mathcal{M}}(y).$$

By (3.3) we have

$$(3.7) \quad B_{\mathcal{M}}(x, \varepsilon) \subset B(x, \varepsilon) \cap \mathcal{M} \subset B_{\mathcal{M}}(x, \varepsilon + 8\varepsilon^3/R^2) \subset B_{\mathcal{M}}(x, 2\varepsilon)$$

provided  $8\varepsilon^3/R^2 \leq \varepsilon$ , which is guaranteed by the Assumptions 2.3. Let us write  $\|\psi\|_\infty = \|\psi\|_{L^\infty(B_{\mathcal{M}}(x, 2\varepsilon))}$ . Then by (3.7) and (3.2) we have

$$(3.8) \quad \begin{aligned} |\mathbb{E}[\Psi] - a| &\leq C\rho_{max}\|\psi\|_\infty n Vol_{\mathcal{M}}(B_{\mathcal{M}}(x, \varepsilon + 8\varepsilon^3/R^2) \setminus B_{\mathcal{M}}(x, \varepsilon)) \\ &\leq C\rho_{max}\|\psi\|_\infty n\varepsilon^{m+2}, \end{aligned}$$

for  $\varepsilon$  sufficiently small. We also have by (3.7) that

$$(3.9) \quad \sigma^2 \leq \mathbb{E}[Z_i^2] = \int_{B(x, \varepsilon) \cap \mathcal{M}} \psi(y)^2 \rho(y) dVol(y) \leq C\rho_{max}\|\psi\|_\infty^2 \varepsilon^m,$$

and

$$(3.10) \quad |Z_i - \mathbb{E}[Z_i]| \leq |Z_i| + |\mathbb{E}[Z_i]| \leq |Z_i| + \mathbb{E}[|Z_i|] \leq 2\|\psi\|_\infty,$$

so we can take  $b = 2\|\psi\|_\infty$ . We now set  $t = \rho_{max}\|\psi\|_\infty \varepsilon^m \delta$  in (3.6), for  $\delta > 0$ , and combine this with (3.8) and (3.9) to obtain

$$\mathbb{P}(|\Psi - \mathbb{E}[\Psi]| \geq C\rho_{max}\|\psi\|_\infty(\delta + \varepsilon^2)n\varepsilon^m) \leq 2 \exp\left(-\frac{c\rho_{max}\delta^2 n\varepsilon^m}{1 + \delta}\right)$$

for constants  $C, c > 0$  depending only on  $\mathcal{M}$ . Restricting  $\varepsilon^2 \leq \delta \leq 1$  completes the proof.  $\square$

**3.3.  $\varepsilon$ -graph.** We now turn to proving pointwise consistency of the  $\varepsilon$ -graph Laplacian. Our main result is the following uniform consistency estimate.

**Theorem 3.3** (Consistency for  $\varepsilon$ -graph). *Let  $f \in C^3(\mathcal{M})$ . Then for  $\varepsilon \leq \delta \leq \varepsilon^{-1}$  we have*

$$(3.11) \quad \mathbb{P}\left[\max_{1 \leq i \leq n} |\mathcal{L}^\varepsilon f(x_i) - \sigma_\eta \Delta_\rho f(x_i)| \geq C\delta\right] \leq 2n \exp(-c\delta^2 n\varepsilon^{m+2}),$$

where  $C$  depends on  $\|f\|_{C^3(B_{\mathcal{M}}(x_i, \varepsilon))}$  and  $[f]_{1; B_{\mathcal{M}}(x_i, 2\varepsilon)}$ , where  $[f]_{1; B_{\mathcal{M}}(x_i, 2\varepsilon)}$  is the Lipschitz constant of the function  $f$  when restricted to the ball  $B_{\mathcal{M}}(x_i, 2\varepsilon)$ .

The proof of Theorem 3.3 is split into two lemmas, Lemma 3.4 and Lemma 3.5. Lemma 3.4 controls the fluctuations between the graph Laplacian and the nonlocal operator

$$(3.12) \quad \mathcal{L}_{nl}^\varepsilon f(x) = \frac{1}{\varepsilon^{m+2}} \int_{B_{\mathcal{M}}(x, \varepsilon)} \eta\left(\frac{|x-y|}{\varepsilon}\right) (f(x) - f(y)) \rho(y) dVol_{\mathcal{M}}(y).$$

Lemma 3.5 establishes consistency of the nonlocal operator  $\mathcal{L}_{nl}^\varepsilon$  to  $\Delta_\rho$  as  $\varepsilon \rightarrow 0$ .

**Lemma 3.4.** *Let  $f \in C^1(\mathcal{M})$ . Then for  $x \in \mathcal{M}$  and  $\varepsilon \leq \delta \leq \varepsilon^{-1}$*

$$(3.13) \quad \mathbb{P} [|\mathcal{L}^\varepsilon f(x) - \mathcal{L}_{nl}^\varepsilon f(x)| \geq C[f]_{1; B_{\mathcal{M}}(x, 2\varepsilon)} \delta] \leq 2 \exp(-c\delta^2 n \varepsilon^{m+2}).$$

*Proof.* The proof is a direct application of Lemma 3.1 with

$$\psi(y) = \frac{1}{n \varepsilon^{m+2}} \eta\left(\frac{|y-x|}{\varepsilon}\right) (f(x) - f(y)),$$

and  $a = \mathcal{L}_{nl}^\varepsilon f(x)$ . We simply need to compute

$$\|\psi\|_{L^\infty(B_{\mathcal{M}}(x, 2\varepsilon))} \leq \frac{C[f]_{1; B_{\mathcal{M}}(x, 2\varepsilon)}}{n \varepsilon^{m+1}}. \quad \square$$

**Lemma 3.5.** *For  $f \in C^3(\mathcal{M})$  and  $x \in \mathcal{M}$*

$$(3.14) \quad |\mathcal{L}_{nl}^\varepsilon f(x) - \sigma_\eta \Delta_\rho f(x)| \leq C(1 + \|f\|_{C^3(B_{\mathcal{M}}(x, \varepsilon))}) \varepsilon.$$

*Proof.* Let us define the intrinsic version of  $\mathcal{L}_{nl}^\varepsilon$  to be

$$(3.15) \quad \mathcal{L}_{nl}^{i,\varepsilon} f(x) := \frac{1}{\varepsilon^{m+2}} \int_{B_{\mathcal{M}}(x, \varepsilon)} \eta\left(\frac{d_{\mathcal{M}}(x, y)}{\varepsilon}\right) (f(x) - f(y)) \rho(y) dVol_{\mathcal{M}}(y).$$

Since  $\eta$  is Lipschitz, it follows from (3.3) that

$$(3.16) \quad |\mathcal{L}_{nl}^{i,\varepsilon} f(x) - \mathcal{L}_{nl}^\varepsilon f(x)| \leq C[f]_{1; B_{\mathcal{M}}(x, \varepsilon)} \varepsilon.$$

Let  $w(v) = f(\exp_x(v))$  and  $p(v) = \rho(\exp_x(v))$ ; that is  $w$  and  $p$  are the functions  $f$  and  $\rho$  expressed in normal Riemannian coordinates. Then we have

$$\begin{aligned} \mathcal{L}_{nl}^{i,\varepsilon} f(x) &= -\frac{1}{\varepsilon^{m+2}} \int_{B(0, \varepsilon) \subset T_x \mathcal{M}} \eta\left(\frac{|v|}{\varepsilon}\right) (w(v) - w(0)) p(v) J_x(v) dv \\ &= -\frac{1}{\varepsilon^2} \int_{B(0, 1)} \eta(|v|) (w(\varepsilon v) - w(0)) p(\varepsilon v) J_x(\varepsilon v) dv. \end{aligned}$$

Using the Taylor expansions  $J_x(\varepsilon v) = 1 + O(\varepsilon^2)$ ,  $p(\varepsilon v) = p(0) + \nabla p(0) \cdot v \varepsilon + O(\varepsilon^2)$ , and

$$w(\varepsilon v) - w(0) = \nabla w(0) \cdot v \varepsilon + \frac{1}{2} v \cdot \nabla^2 w(0) v \varepsilon^2 + O(\|w\|_{C^3(B(0, \varepsilon))} \varepsilon^3),$$

a standard computation yields

$$\mathcal{L}_{nl}^{i,\varepsilon} f(x) = -\sigma_\eta \left( \nabla w(0) \cdot \nabla p(0) + \frac{p(0)}{2} \Delta_{\mathcal{M}} w(0) \right) + O(c_3 \varepsilon) = -\frac{\sigma_\eta}{2p} \text{div}(p^2 \nabla w)|_{v=0} + O(c_3 \varepsilon),$$

where

$$c_3 = 1 + \|w\|_{C^3(B(0, \varepsilon))}.$$

The proof is completed by recalling (3.16) and noting that  $-\frac{\sigma_\eta}{2p} \text{div}(p^2 \nabla w)|_{v=0} = \sigma_\eta \Delta_\rho f(x)$ .  $\square$

**Remark 3.6.** In the previous proof we have implicitly used the fact that the norms  $\|w\|_{C^3(B(0, \varepsilon))}$  and  $\|f\|_{C^3(B_{\mathcal{M}}(x, \varepsilon))}$  are equivalent. This can be justified as follows. Writing the covariant derivatives  $\nabla^1 f$ ,  $\nabla^2 f$  and  $\nabla^3 f$  in normal coordinates around a point (see Section 2 in [31]), and using standard expansions for the Christoffel symbols and metric tensor in these coordinates, it is not difficult to see that:

$$(1 - C\varepsilon) \|w\|_{C^3(B(0, \varepsilon))} \leq \|f\|_{C^3(B_{\mathcal{M}}(x, \varepsilon))} \leq (1 + C\varepsilon) \|w\|_{C^3(B(0, \varepsilon))},$$

from where the equivalence of the norms follows.

We now give the proof of Theorem 3.3.

*Proof of Theorem 3.3.* Combining Lemmas 3.4 and 3.5 we have

$$(3.17) \quad \mathbb{P} [|\mathcal{L}^\varepsilon f(x) - \sigma_\eta \Delta_\rho f(x)| \geq C\delta] \leq 2 \exp(-c\delta^2 n\varepsilon^{m+2})$$

for any  $x \in \mathcal{M}$  and  $\varepsilon \leq \delta \leq \varepsilon^{-1}$ . Conditioning on  $x_i$  and using the law of conditional probability yields

$$\mathbb{P} [|\mathcal{L}^\varepsilon f(x_i) - \sigma_\eta \Delta_\rho f(x_i)| \geq C\delta] \leq 2 \exp(-c\delta^2 n\varepsilon^{m+2})$$

for any  $1 \leq i \leq n$ . The proof is completed by union bounding over  $x_1, \dots, x_n$ .  $\square$

**3.4. Undirected  $k$ NN graph Laplacian.** We now turn to the case of the undirected  $k$ -NN graph Laplacian  $\mathcal{L}^k$ . The consistency proof here is more involved, since the neighbors in the symmetrized  $k$ -NN relation do not fall in a ball (even on average), due to the variability of the distribution  $\rho$ . The additional neighbors added in the symmetrization are in fact important for consistency of the undirected  $k$ -NN graph Laplacian, and when  $\rho$  is not constant the additional points are not symmetrically distributed about  $x$ . This must be accounted for in the consistency results, and introduces an additional drift term in the limiting weighted Laplace-Beltrami operator. The form of the continuum operator  $\Delta_\rho^{NN}$  was established in [50], where the authors proved consistency *without* a convergence rate.

Our main result in this section is the following uniform consistency estimate.

**Theorem 3.7** (Consistency for undirected  $k$ -NN Laplacian). *Let  $f \in C^3(\mathcal{M})$ . For  $1 \leq k \leq cn\varepsilon_{\mathcal{M}}^m$  and  $C(k/n)^{1/m} \leq \delta \leq (k/n)^{-1/m}$  we have*

$$(3.18) \quad \mathbb{P} \left( \max_{1 \leq i \leq n} |\mathcal{L}^k f(x_i) - \sigma_\eta \Delta_\rho^{NN} f(x_i)| \geq C\delta \right) \leq Cn \exp \left( -c\delta^2 (k/n)^{2/m} k \right),$$

where  $C$  depends on  $\|f\|_{C^3(\mathcal{M})}$ .

The proof of Theorem 3.7 follows from Lemmas 3.10 and 3.11 below. Before presenting the lemmas and proofs, we must introduce some notation. We recall  $r_k(x, y) = \max\{\varepsilon_k(x), \varepsilon_k(y)\}$ , as defined in (2.8). For  $x \in \mathcal{M}$  define  $\varepsilon(x)$  by

$$(3.19) \quad k = \alpha_m \rho(x) n \varepsilon(x)^m,$$

and set

$$(3.20) \quad r(x, y) := \max\{\varepsilon(x), \varepsilon(y)\}.$$

**Lemma 3.8.** *Let  $x \in \mathcal{M}$  and suppose that  $\varepsilon < \varepsilon_{\mathcal{M}}$ . Then for  $\varepsilon^2 \leq \delta \leq 1$*

$$(3.21) \quad \mathbb{P} [|N_\varepsilon(x) - \alpha_m \rho(x) n \varepsilon^m| \geq C\delta n \varepsilon^m] \leq 2 \exp(-c\delta^2 n \varepsilon^m).$$

*Proof.* Applying Lemma 3.1 with  $\psi \equiv 1$  yields

$$\mathbb{P} \left[ \left| N_\varepsilon(x) - n \int_{B_{\mathcal{M}}(x, \varepsilon)} \rho(y) dVol_{\mathcal{M}}(y) \right| \geq C\delta n \varepsilon^m \right] \leq 2 \exp(-c\delta^2 n \varepsilon^m).$$

By Taylor expansion we have

$$n \int_{B_{\mathcal{M}}(x, \varepsilon)} \rho(y) dVol_{\mathcal{M}}(y) = \alpha_m \rho(x) n \varepsilon^m + O(n \varepsilon^{m+2}).$$

Noting that  $\varepsilon^2 \leq \delta$  completes the proof.  $\square$

**Lemma 3.9.** *For  $x \in \mathcal{M}$ ,  $1 \leq k \leq cn\varepsilon_{\mathcal{M}}^m$  and  $C(k/n)^{2/m} \leq \delta \leq 1$  we have*

$$(3.22) \quad \mathbb{P} [|\alpha_m \rho(x) n \varepsilon_k(x)^m - k| \geq C\delta k] \leq 4 \exp(-c\delta^2 k).$$

*Proof.* Let  $C'$  be the constant from Lemma 3.8 and define  $\varepsilon$  by

$$\alpha_m \rho(x) n \varepsilon^m = k(1 + C' \alpha_m^{-1} \rho(x)^{-1} \delta).$$

Then we have

$$\begin{aligned} \mathbb{P}(\alpha_m \rho(x) n \varepsilon_k(x)^m > k(1 + C\delta)) &= \mathbb{P}(\varepsilon_k(x) > \varepsilon) \\ &\leq \mathbb{P}(N_\varepsilon(x) < k) \\ &\leq \mathbb{P}(N_\varepsilon(x) - \alpha_m \rho(x) n \varepsilon^m < -C' \alpha_m^{-1} \rho(x)^{-1} k \delta) \\ &\leq \mathbb{P}(|N_\varepsilon(x) - \alpha_m \rho(x) n \varepsilon^m| \geq C' \delta n \varepsilon^m). \end{aligned}$$

Since  $c(k/n)^{1/m} \leq \varepsilon \leq C(k/n)^{1/m}$  the restriction  $\varepsilon < \varepsilon_M$  from Lemma 3.8 is satisfied when  $k \leq cn\varepsilon_M^m$ . Therefore, we can invoke Lemma 3.8 to find that

$$\mathbb{P}(n\alpha_m \rho(x) \varepsilon_k(x)^m > k(1 + C\delta)) \leq 2 \exp(-c\delta^2 k).$$

for  $\varepsilon^2 \leq \delta \leq 1$ . This establishes one direction of (3.22); the proof of the other is similar.  $\square$

We set  $\varepsilon_{max} = \max_{x \in \mathcal{M}} \varepsilon(x)$ , and define the nonlocal operator

$$(3.23) \quad \mathcal{L}_{nl}^k f(x) = \left(\frac{n\alpha_m}{k}\right)^{1+2/m} \int_{B_{\mathcal{M}}(x, \varepsilon_{max})} \eta\left(\frac{|x-y|}{r(x,y)}\right) (f(x) - f(y)) \rho(y) d\text{Vol}_{\mathcal{M}}(y).$$

Our first lemma shows that the nonlocal operator  $\mathcal{L}_{nl}^k$  describes the average behavior of the unweighted  $k$ -nearest neighbor graph Laplacian  $\mathcal{L}^k$ .

**Lemma 3.10.** *Let  $f \in C^1(\mathcal{M})$ . Then for  $x \in \mathcal{M}$  and  $C(k/n)^{1/m} \leq \delta \leq (k/n)^{-1/m}$*

$$(3.24) \quad \mathbb{P}\left[|\mathcal{L}^k f(x) - \mathcal{L}_{nl}^k f(x)| \geq C[f]_{1; B_{\mathcal{M}}(x, 2\varepsilon_{max})} \delta\right] \leq C \exp\left(-c\delta^2 (k/n)^{2/m} k\right).$$

*Proof.* Define

$$(3.25) \quad A = \{j : |x - x_j| \leq r_k(x_j, x)\},$$

and

$$(3.26) \quad A(s) = \{j : |x - x_j| \leq r(x_j, x)(1 + s)\}.$$

By Lemma 3.9, there exists  $\tilde{C} > 0$  such that

$$(3.27) \quad \mathbb{P}(|\varepsilon_k(x)^m - \varepsilon(x)^m| \geq \tilde{C} \delta \varepsilon(x)^m) \leq 6 \exp(-c\delta^2 k)$$

and

$$\mathbb{P}\left[\max_{1 \leq j \leq n} |\varepsilon_k(x_j)^m - \varepsilon(x_j)^m| \geq \tilde{C} \delta \varepsilon(x)^m\right] \leq 6n \exp(-c\delta^2 k)$$

for  $C(k/n)^{2/m} \leq \delta \leq 1$ . Fix such a  $\delta > 0$  and assume that

$$(3.28) \quad |\varepsilon_k(x)^m - \varepsilon(x)^m| \leq \tilde{C} \delta \varepsilon(x)^m$$

and

$$(3.29) \quad \max_{1 \leq j \leq n} |\varepsilon_k(x_j)^m - \varepsilon(x_j)^m| \leq \tilde{C} \delta \varepsilon(x)^m.$$

Then it follows that

$$r_k(x_j, x)^m = \max\{\varepsilon_k(x_j)^m, \varepsilon_k(x)^m\} \leq (1 + \tilde{C}\delta) \max\{\varepsilon(x_j)^m, \varepsilon(x)^m\} = (1 + \tilde{C}\delta) r(x_j, x)^m.$$

Since  $(1 + \tilde{C}\delta)^{1/m} \leq 1 + \tilde{C}\delta$  we obtain

$$r_k(x_j, x) \leq (1 + \tilde{C}\delta) r(x_j, x),$$

and so  $A \subset A(\tilde{C}\delta)$ . A similar argument shows that  $A(-\tilde{C}\delta) \subset A$ .

We define

$$Lf(x) = \frac{1}{n} \left(\frac{n\alpha_m}{k}\right)^{1+2/m} \sum_{j \in A(0)} w_{x_j x}^{r(x_j, x)} (f(x) - f(x_j)),$$

and note, using  $A(-\tilde{C}\delta) \subset A \subset A(\tilde{C}\delta)$ , that

$$\begin{aligned} |Lf(x) - \mathcal{L}^k f(x)| &\leq \frac{C}{n\varepsilon(x)^{m+2}} \left[ \sum_{j \in A(\tilde{C}\delta) \setminus A(-\tilde{C}\delta)} |f(x_j) - f(x)| \right. \\ &\quad \left. + \sum_{j \in A(-\tilde{C}\delta)} |w_{x_j x}^{r(x_j, x)} - w_{x_j x}^{r_k(x_j, x)}| |f(x_j) - f(x)| \right]. \end{aligned}$$

By the Chernoff bounds we have

$$\mathbb{P} \left( \#A(\tilde{C}\delta) - \#A(-\tilde{C}\delta) \geq C\delta n\varepsilon(x)^m \right) \leq 2 \exp(-c\delta^2 k),$$

for  $0 \leq \delta \leq 1$ , and by (3.28) and (2.14) we have

$$\max_{j \in A(-\tilde{C}\delta)} |w_{x_j x}^{r(x_j, x)} - w_{x_j x}^{r_k(x_j, x)}| \leq C\delta.$$

Setting  $\delta = (k/n)^{1/m}t$  we have

$$(3.30) \quad \mathbb{P}(|Lf(x) - \mathcal{L}^k f(x)| \geq C[f]_{1;B_{\mathcal{M}}(x, 2\varepsilon_{max})} t) \leq C \exp(-ct^2(k/n)^{2/m}k)$$

for  $C(k/n)^{1/m} \leq t \leq (k/n)^{-1/m}$ . The proof is completed by invoking Lemma 3.1 to obtain

$$\mathbb{P}(|\mathcal{L}_{nl}^{k,s} f(x) - Lf(x)| \geq C[f]_{1;B_{\mathcal{M}}(x, 2\varepsilon_{max})} t) \leq 2 \exp(-ct^2(k/n)^{2/m}k)$$

for  $C(k/n)^{1/m} \leq t \leq (k/n)^{-1/m}$ , and combining with (3.30).  $\square$

We now establish consistency of the nonlocal operator  $\mathcal{L}_{nl}^k$  with the weighted Laplacian  $\Delta_{\rho}^{NN}$ .

**Lemma 3.11.** *There exists  $c > 0$  depending on  $\rho$ , so that for  $f \in C^3(\mathcal{M})$ ,  $x \in \mathcal{M}$ , and  $k \leq cn$ , we have*

$$(3.31) \quad |\mathcal{L}_{nl}^k f(x) - \sigma_{\eta} \Delta_{\rho}^{NN} f(x)| \leq C(1 + \|f\|_{C^3(B_{\mathcal{M}}(x, \varepsilon_{max}))}) \varepsilon(x).$$

*Proof.* Throughout the proof we write  $\varepsilon = \varepsilon(x)$ , and note that

$$c \left( \frac{k}{n} \right)^{1/m} \leq \varepsilon \leq \varepsilon_{\max} \leq C \left( \frac{k}{n} \right)^{1/m}.$$

We first define the intrinsic version of  $\mathcal{L}_{nl}^{k,s}$ , given by

$$(3.32) \quad Lf(x) := \left( \frac{n\alpha_m}{k} \right)^{1+2/m} \int_{B_{\mathcal{M}}(x, \varepsilon_{max})} \eta \left( \frac{d_{\mathcal{M}}(x, y)}{r(x, y)} \right) (f(x) - f(y)) \rho(y) dVol_{\mathcal{M}}(y).$$

Since  $\eta$  is Lipschitz it follows from (3.3) that

$$(3.33) \quad |Lf(x) - \mathcal{L}_{nl}^k f(x)| \leq C[f]_{1;B_{\mathcal{M}}(x, \varepsilon_{max})} (k/n)^{1/m} \leq C[f]_{1;B_{\mathcal{M}}(x, \varepsilon_{max})} \varepsilon.$$

Let  $w(v) = f(\exp_x(v))$  and  $p(v) = \rho(\exp_x(v))$ . Then we have

$$Lf(x) = - \left( \frac{n\alpha_m}{k} \right)^{1+2/m} \int_{B(0, \varepsilon_{max}) \subset T_x \mathcal{M}} \eta \left( \frac{|v|}{r(x, \exp_x(v))} \right) (w(v) - w(0)) p(v) J_x(v) dv.$$

Let us set  $s(v) = (\rho(\exp_x(v))/\rho(x))^{1/m} = (p(v)/p(0))^{1/m}$ , so that

$$r(x, \exp_x(v)) = \varepsilon \max\{1, s(v)^{-1}\} = \frac{\varepsilon(x)}{\min\{1, s(v)\}}.$$

Then making a change of variables  $v' = v/\varepsilon$ , and renaming  $v'$  as  $v$ , yields

$$(3.34) \quad Lf(x) = - \left( \frac{n\alpha_m}{k} \right)^{2/m} \int_{B(0, C)} \eta(|v| \min\{1, s(\varepsilon v)\}) (w(\varepsilon v) - w(0)) \frac{p(\varepsilon v)}{p(0)} J_x(\varepsilon v) dv,$$

where  $C = \varepsilon^{-1}\varepsilon_{max}$ . We now use the Taylor expansions  $J_x(\varepsilon v) = 1 + O(\varepsilon^2)$ ,  $p(\varepsilon v) = p(0) + \nabla p(0) \cdot v\varepsilon + O(\varepsilon^2)$ ,

$$w(\varepsilon v) - w(0) = \nabla w(0) \cdot v\varepsilon + \frac{1}{2}v \cdot \nabla^2 w(0)v\varepsilon^2 + O(\|w\|_{C^3(B(0,\varepsilon_{max}))}\varepsilon^3),$$

and

$$s(\varepsilon v) = 1 + \frac{1}{m}\nabla \log p(0) \cdot v\varepsilon + O(\varepsilon^2),$$

to obtain

$$\begin{aligned} Lf(x) = -\left(\frac{n\alpha_m}{k}\right)^{2/m} \int_B \eta \left(|v|(1 + \frac{\varepsilon}{m}[\nabla \log p(0) \cdot v]_-)\right) &(\nabla w(0) \cdot v\varepsilon + \frac{1}{2}v \cdot \nabla^2 w(0)v\varepsilon^2) \\ &(1 + \nabla \log p(0) \cdot v\varepsilon) dv + O(\varepsilon), \end{aligned}$$

where  $a_- = \min\{0, a\}$  and

$$B = \{v \in \mathbb{R}^m : |v|(1 + \frac{\varepsilon}{m}[\nabla \log p(0) \cdot v]_-) \leq 1\}.$$

We now make the change of variables

$$z = \Phi(v) := v(1 + \frac{\varepsilon}{m}[\nabla \log p(0) \cdot v]_-).$$

For sufficiently small  $\varepsilon > 0$ ,  $\Phi$  is invertible and

$$v = \Phi^{-1}(z) = z(1 - \frac{\varepsilon}{m}[\nabla \log p(0) \cdot z]_- + O(\varepsilon^2)).$$

Note that for  $v$  with  $\nabla \log p(0) \cdot v > 0$ , we have  $D\Phi(v) = I$  and  $\det(D\Phi(v)) = 1$ . For  $v$  with  $\nabla \log p(0) \cdot v < 0$  we have

$$D\Phi(v) = I + \frac{\varepsilon}{m} ((\nabla \log p(0) \cdot v)I + \nabla \log p(0) \otimes v).$$

Using the Taylor expansion  $\det(I + \varepsilon X) = 1 + \varepsilon \text{Tr}(X) + O(\varepsilon^2)$  we have

$$\begin{aligned} \det(D\Phi(v)) &= 1 + \frac{\varepsilon}{m} ((\nabla \log p(0) \cdot v)m + (\nabla \log p(0) \cdot v)) + O(\varepsilon^2) \\ &= 1 + \varepsilon(1 + \frac{1}{m})(\nabla \log p(0) \cdot v) + O(\varepsilon^2). \end{aligned}$$

Thus, for all  $v$  we have

$$|\det(D\Phi(v))|^{-1} = 1 - \varepsilon(1 + \frac{1}{m})[\nabla \log p(0) \cdot v]_- + O(\varepsilon^2),$$

and so

$$dv = (1 - \varepsilon(1 + \frac{1}{m})[\nabla \log p(0) \cdot v]_- + O(\varepsilon^2))dz.$$

Therefore

$$\begin{aligned}
L f(x) &= - \left( \frac{n\alpha_m}{k} \right)^{2/m} \int_{B(0,1)} \eta(|z|) \left( \nabla w(0) \cdot z \varepsilon - \frac{\varepsilon^2}{m} [\nabla \log p(0) \cdot z]_- \nabla w(0) \cdot z + \frac{1}{2} x \cdot \nabla^2 w(0) x \varepsilon^2 \right) \\
&\quad (1 + \varepsilon \nabla \log p(0) \cdot z) (1 - \varepsilon(1 + \frac{1}{m}) [\nabla \log p(0) \cdot z]_-) dz + O(\varepsilon) \\
&= - \left( \frac{n\alpha_m}{k} \right)^{2/m} \int_{B(0,1)} \eta(|z|) \left( \nabla w(0) \cdot z \varepsilon - \frac{\varepsilon^2}{m} [\nabla \log p(0) \cdot z]_- \nabla w(0) \cdot z + \frac{1}{2} x \cdot \nabla^2 w(0) x \varepsilon^2 \right) \\
&\quad (1 - \varepsilon(1 + \frac{1}{m}) [\nabla \log p(0) \cdot z]_- + \varepsilon \nabla \log p(0) \cdot z) dz + O(\varepsilon) \\
&= -p(0)^{-2/m} \int_{B(0,1)} \eta(|z|) \left( (\nabla \log p(0) \cdot z) (\nabla w(0) \cdot z) + \frac{1}{2} x \cdot \nabla^2 w(0) x \right. \\
&\quad \left. - (1 + \frac{2}{m}) [\nabla \log p(0) \cdot z]_- (\nabla w(0) \cdot z) \right) dz + O(\varepsilon) \\
&= -\sigma_\eta p(0)^{-2/m} (\nabla \log p(0) \cdot \nabla w(0) + \frac{1}{2} \Delta w(0)) \\
&\quad - p(0)^{-2/m} (1 + \frac{2}{m}) \underbrace{\int_{B(0,1)} \eta(|z|) [\nabla \log p(0) \cdot z]_- (\nabla w(0) \cdot z) dz}_I + O(\varepsilon).
\end{aligned}$$

In the final integral above, let  $A$  be an orthogonal transformation so that

$$A \nabla \log p(0) = |\nabla \log p(0)| e_m$$

and make the change of variables  $y = Az$  to deduce

$$\begin{aligned}
I &= |\nabla \log p(0)| \int_{B(0,1)} \eta(|y|) \min\{y_m, 0\} (A \nabla w(0) \cdot y) dy \\
&= |\nabla \log p(0)| \sum_{i=1}^m [A \nabla w(0)]_i \int_{B(0,1)} \eta(|y|) \min\{y_m, 0\} y_i dy \\
&= |\nabla \log p(0)| [A \nabla w(0)]_m \int_{B(0,1)} \eta(|y|) \min\{y_m, 0\} y_m dy \\
&= \frac{1}{2} |\nabla \log p(0)| [A \nabla w(0)]_m \int_{B(0,1)} \eta(|y|) y_m^2 dy \\
&= \frac{\sigma_\eta}{2} |\nabla \log p(0)| [A \nabla w(0)]_m \\
&= \frac{\sigma_\eta}{2} \nabla \log p(0) \cdot \nabla w(0).
\end{aligned}$$

This gives

$$\begin{aligned}
L f(x) &= -\frac{\sigma_\eta}{2} p(0)^{-2/m} (\Delta w(0) + (1 - \frac{2}{m}) \nabla \log p(0) \cdot \nabla w(0)) + O(\varepsilon) \\
&= -\frac{\sigma_\eta}{2} \frac{1}{p(0)} \operatorname{div}(p^{1-2/m} \nabla w) + O(\varepsilon) = \sigma_\eta \Delta_\rho^{NN} f + O(\varepsilon),
\end{aligned}$$

which completes the proof.  $\square$

We now give the proof of Theorem 3.7.

*Proof of Theorem 3.7.* We proceed as in the proof of Theorem 3.3. Combining Lemmas 3.10 and 3.11 we have

$$\mathbb{P} \left( |\mathcal{L}^k f(x) - \sigma_\eta \Delta_\rho^{NN} f(x)| \geq C\delta \right) \leq Cn \exp \left( -c\delta^2 (k/n)^{2/m} k \right),$$

for any  $x \in \mathcal{M}$ ,  $1 \leq k \leq cn\varepsilon_{\mathcal{M}}^m$ , and  $C(k/n)^{1/m} \leq \delta \leq (k/n)^{-1/m}$ . We complete the proof by conditioning on  $x_i$ , using the law of conditional probability, and applying a union bound.  $\square$

#### 4. PROOFS OF MAIN RESULTS

Here we prove all of our main results. The structure of the proofs is exactly the same for the  $\varepsilon$ -graph and the  $k$ -NN settings.

**4.1.  $\varepsilon$ -graph.** Let  $\varepsilon, \tilde{\delta}, \theta$  be positive numbers satisfying Assumptions 2.3. Associated to these numbers we consider the density  $\tilde{\rho}_n$  from Proposition 2.12 (which exists with probability greater than  $1 - n \exp(-Cn\theta^2\tilde{\delta}^2)$ ) and we let  $\tilde{T}$  be an  $\infty$ -OT map between  $\tilde{\mu}_n$  and  $\mu_n$ . Let  $\tilde{U}_1, \dots, \tilde{U}_n$  be defined by

$$\tilde{U}_i := \tilde{T}_n^{-1}(\{x_i\}).$$

We can then define the *contractive discretization* map  $\tilde{P}: L^2(\mu) \rightarrow L^2(\mu_n)$  by

$$(4.1) \quad (\tilde{P}f)(x_i) := n \cdot \int_{\tilde{U}_i} f(x) \tilde{\rho}_n(x) dx, \quad f \in L^2(\mu),$$

and the extension map  $\tilde{P}^*: L^2(\mu_n) \rightarrow L^2(\tilde{\mu}_n)$  by

$$(4.2) \quad (\tilde{P}^*u)(x) := \sum_{i=1}^n u(x_i) \mathbb{1}_{\tilde{U}_i}(x), \quad u \in L^2(\mu_n).$$

We note that  $\tilde{P}^*u$  can be written as  $\tilde{P}^*u = u \circ \tilde{T}$ . We then define the *interpolation* map  $\tilde{\mathcal{I}}: L^2(\mu_n) \rightarrow \text{Lip}(\mathcal{M})$

$$(4.3) \quad \tilde{\mathcal{I}}u := \Lambda_{\varepsilon-2\tilde{\delta}} P^*u$$

where  $\Lambda_{\varepsilon-2\tilde{\delta}}$  is a convolution operator using the kernel  $K_r(\cdot, \cdot)$  (defined below) with bandwidth  $\varepsilon - 2\tilde{\delta}$ . To define the kernel  $K_r$  we let  $\psi: [0, \infty) \rightarrow [0, \infty)$  be given by

$$(4.4) \quad \psi(t) := \frac{1}{\sigma_\eta} \int_t^\infty \eta(s) s ds,$$

and set

$$K_r(x, y) := \frac{1}{r^m} \psi\left(\frac{d_{\mathcal{M}}(x, y)}{r}\right),$$

The operator  $\Lambda_r$  then takes the form

$$\Lambda_r f(x) := \frac{1}{\tau(x)} \int_{\mathcal{M}} K_r(x, y) f(y) d\mu(y),$$

where in the above  $\tau(x)$  is a normalization factor given by

$$\tau(x) := \int_{\mathcal{M}} K_r(x, y) d\mu(y),$$

and serves as normalization constant. Note that in the above we integrate with respect to the density  $\rho$  and not with respect to  $\tilde{\rho}_n$ . This is because  $\tilde{\rho}_n$  is discontinuous, and for some of the estimates that we will use later on, we need to integrate with respect to a smoother density.

In order to prove the first part of Theorem 2.4 we use the following two propositions. These two results contain the fundamental a priori estimates that we use in the sequel.

**Proposition 4.1** (Inequality for Dirichlet energies). *Let  $\varepsilon, \tilde{\delta}$ , and  $\theta$  be fixed but small enough numbers satisfying Assumptions 2.3. Then, with probability greater than  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m)$  we have:*

(1) *For any  $f \in L^2(\mu)$ ,*

$$b_\varepsilon(\tilde{P}f) \leq \left(1 + C \left(\frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta\right)\right) \sigma_\eta D_2(f),$$

(2) For any  $u \in L^2(\mu_n)$ ,

$$\sigma_\eta D_2(\tilde{\mathcal{I}}u) \leq \left(1 + C \left(\frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta\right)\right) b_\varepsilon(u),$$

**Proposition 4.2** (Discretization and interpolation maps are almost isometries). *Let  $\varepsilon$ ,  $\tilde{\delta}$ , and  $\theta$  be fixed but small enough numbers satisfying Assumptions 2.3. Then, with probability at least  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m)$  we have:*

(1) For every  $f \in L^2(\mu)$ ,

$$\left| \|f\|_{L^2(\mu)}^2 - \|\tilde{P}f\|_{L^2(\mu_n)}^2 \right| \leq C\tilde{\delta}\|f\|_{L^2(\mu)}\sqrt{D_2(f)} + C(\theta + \tilde{\delta})\|f\|_{L^2(\mu)}^2.$$

(2) For every  $u \in L^2(\mu_n)$ ,

$$\left| \|u\|_{L^2(\mu_n)}^2 - \|\tilde{\mathcal{I}}u\|_{L^2(\mu)}^2 \right| \leq C\varepsilon\|u\|_{L^2(\mu_n)}\sqrt{b_\varepsilon(u)} + C(\theta + \tilde{\delta})\|u\|_{L^2(\mu_n)}^2.$$

Before proving these two propositions it will be convenient to introduce two intermediate (non-local) Dirichlet energies of interest and establish a connection between them. First, we define the *non-local* energy

$$(4.5) \quad \tilde{E}_r(f) := \int_{\mathcal{M}} \int_{\mathcal{M}} \frac{1}{r^{m+2}} \eta\left(\frac{d_{\mathcal{M}}(x, y)}{r}\right) (f(x) - f(y))^2 \tilde{\rho}_n(x) \tilde{\rho}_n(y) d\text{Vol}_{\mathcal{M}}(x) d\text{Vol}_{\mathcal{M}}(y), \quad f \in L^2(\mathcal{M}, \mu),$$

and the closely related

$$(4.6) \quad E_r(f) := \int_{\mathcal{M}} \int_{\mathcal{M}} \frac{1}{r^{m+2}} \eta\left(\frac{d_{\mathcal{M}}(x, y)}{r}\right) (f(x) - f(y))^2 \rho(x) \rho(y) d\text{Vol}_{\mathcal{M}}(x) d\text{Vol}_{\mathcal{M}}(y), \quad f \in L^2(\mathcal{M}, \mu).$$

Notice that the only difference between  $\tilde{E}_r$  and  $E_r$  is the density we integrate with respect to. Moreover, for the density  $\tilde{\rho}_n$  from Proposition 2.12 we have

$$(4.7) \quad (1 - C(\theta + \tilde{\delta}))E_r(f) \leq \tilde{E}_r(f) \leq (1 + C(\theta + \tilde{\delta}))E_r(f), \quad \forall f \in L^2(\mu),$$

for some constant  $C$ , which follows from the fact that

$$\rho(x) \leq \tilde{\rho}_n(x) + \|\rho - \tilde{\rho}_n\|_{L^\infty(\mu)} \leq \tilde{\rho}_n(x) + \frac{C(\theta + \tilde{\delta})}{\rho_{\min}} \tilde{\rho}_n(x)$$

and

$$\tilde{\rho}_n(x) \leq \rho(x) + \|\rho - \tilde{\rho}_n\|_{L^\infty(\mu)} \leq \rho(x) + \frac{C(\theta + \tilde{\delta})}{\rho_{\min}} \rho(x).$$

We are now ready to prove Proposition 4.1 and Proposition 4.2.

*Proof of Proposition 4.1.* First, we can use Lemma 5 in [22] to conclude that for all  $r > 0$  small enough and all  $f \in L^2(\mu)$ ,

$$E_r(f) \leq (1 + C_\rho r + CmKr^2)\sigma_\eta D_2(f),$$

and Lemma 9 in [22] to get

$$\sigma_\eta D_2(\Lambda_r f) \leq (1 + Cr + CmKr^2) \cdot (1 + C(1 + 1/\sigma_\eta)mKr^2) E_r(f).$$

We can then use 2.12 to obtain

$$(4.8) \quad \tilde{E}_r(f) \leq (1 + Cr + CmKr^2)(1 + C(\theta + \tilde{\delta}))\sigma_\eta D_2(f),$$

as well as

$$(4.9) \quad \sigma_\eta D_2(\Lambda_r f) \leq (1 + Cr + CmKr^2) \cdot (1 + C(1 + 1/\sigma_\eta)mKr^2)(1 + C(\theta + \tilde{\delta}))\tilde{E}_r(f).$$

On the other hand, if we replace  $P, P^*, \delta, T_n$  and  $\rho$  with  $\tilde{P}, \tilde{P}^*, \tilde{\delta}, \tilde{T}_n$  and  $\tilde{\rho}_n$ , (where recall  $\tilde{\delta}$  is an upper bound for the  $\infty$ -OT distance between  $\tilde{\mu}_n$  and  $\mu_n$ ), we can copy word for word the proofs in Lemmas 13 and 14 in [22] to deduce

$$b_\varepsilon(\tilde{P}f) \leq \left(1 + C\frac{\tilde{\delta}}{\varepsilon}\right) \tilde{E}_{\varepsilon+2\tilde{\delta}}(f)$$

as well as

$$(4.10) \quad \tilde{E}_{\varepsilon-2\tilde{\delta}}(\tilde{P}^*u) \leq \left(1 + C\frac{\tilde{\delta}}{\varepsilon}\right) b_\varepsilon(u)$$

for every  $f \in L^2(\mu)$  and  $u \in L^2(\mu_n)$ . Putting together the above estimates with (4.8) and (4.9) with  $r = \varepsilon - 2\tilde{\delta}$ , and using the smallness Assumptions (2.3) on  $\varepsilon, \tilde{\delta}$  and  $\theta$ , we obtain the desired inequalities.  $\square$

*Proof of Proposition 4.2.* The proof follows the same ideas used when proving the analogous results in [22]. The setting is slightly different because the discretization and interpolation maps have changed.

To prove the first assertion, we start by noticing that

$$\|f\|_{L^2(\mu)}^2 = \int_{\mathcal{M}} (f(x))^2 \frac{\rho(x) - \tilde{\rho}_n(x)}{\rho(x)} \rho(x) d\text{Vol}_{\mathcal{M}}(x) + \|f\|_{L^2(\tilde{\mu}_n)}^2.$$

The first term on the right hand side is in absolute value less than

$$\frac{\|\rho - \tilde{\rho}_n\|_{L^\infty(\mu)}}{\rho_{\min}} \|f\|_{L^2(\mu)}^2,$$

and so

$$(4.11) \quad \left| \|f\|_{L^2(\tilde{\mu}_n)}^2 - \|f\|_{L^2(\mu)}^2 \right| \leq C(\theta + \tilde{\delta}) \|f\|_{L^2(\mu)}^2.$$

Notice also that by definition of  $\tilde{P}^*$  we have

$$\|\tilde{P}^* \tilde{P}f\|_{L^2(\tilde{\mu}_n)}^2 = \|\tilde{P}f\|_{L^2(\mu_n)}^2.$$

It follows that,

$$\begin{aligned} \left| \|\tilde{P}f\|_{L^2(\mu_n)}^2 - \|f\|_{L^2(\mu)}^2 \right| &\leq \left| \|\tilde{P}^* \tilde{P}f\|_{L^2(\tilde{\mu}_n)}^2 - \|f\|_{L^2(\tilde{\mu}_n)}^2 \right| + \left| \|f\|_{L^2(\tilde{\mu}_n)}^2 - \|f\|_{L^2(\mu)}^2 \right| \\ &\leq C\|f\|_{L^2(\mu)} \left| \|\tilde{P}^* \tilde{P}f\|_{L^2(\tilde{\mu}_n)} - \|f\|_{L^2(\tilde{\mu}_n)} \right| + C(\theta + \tilde{\delta}) \|f\|_{L^2(\mu)}^2 \\ &\leq C\|f\|_{L^2(\mu)} \|\tilde{P}^* \tilde{P}f - f\|_{L^2(\tilde{\mu}_n)} + C(\theta + \tilde{\delta}) \|f\|_{L^2(\mu)}^2. \end{aligned}$$

On the other hand,

$$\begin{aligned} \|\tilde{P}^* \tilde{P}f - f\|_{L^2(\tilde{\mu}_n)}^2 &= \sum_{i=1}^n \int_{\tilde{U}_i} \left( \tilde{P}^* \tilde{P}f(x) - f(x) \right)^2 \tilde{\rho}_n(x) d\text{Vol}_{\mathcal{M}}(x) \\ &= \sum_{i=1}^n \int_{\tilde{U}_i} \left( \tilde{P}f(x_i) - f(x) \right)^2 \tilde{\rho}_n(x) d\text{Vol}_{\mathcal{M}}(x) \\ &= \sum_{i=1}^n \int_{\tilde{U}_i} \left( n \int_{\tilde{U}_i} f(y) \tilde{\rho}_n(y) d\text{Vol}_{\mathcal{M}}(y) - f(x) \right)^2 \tilde{\rho}_n(x) d\text{Vol}_{\mathcal{M}}(x) \\ &= \sum_{i=1}^n \int_{\tilde{U}_i} \left( n \int_{\tilde{U}_i} (f(y) - f(x)) \tilde{\rho}_n(y) d\text{Vol}_{\mathcal{M}}(y) \right)^2 \tilde{\rho}_n(x) d\text{Vol}_{\mathcal{M}}(x) \\ &\leq n \sum_{i=1}^n \int_{\tilde{U}_i} \int_{\tilde{U}_i} (f(y) - f(x))^2 \tilde{\rho}_n(x) \tilde{\rho}_n(y) d\text{Vol}_{\mathcal{M}}(y) d\text{Vol}_{\mathcal{M}}(x). \end{aligned}$$

Let us now show that the last term in the above chain of inequalities can be controlled by  $C\tilde{\delta}^2 E_{2\tilde{\delta}}(f)$ , where actually  $C$  is a constant that only depends on dimension. For this purpose we use an idea described in Lemma 3.4 in [7] to estimate each of the terms

$$n \int_{\tilde{U}_i} \int_{\tilde{U}_i} (f(y) - f(x))^2 \tilde{\rho}_n(x) \tilde{\rho}_n(y) d\text{Vol}_{\mathcal{M}}(y) d\text{Vol}_{\mathcal{M}}(x).$$

For fixed  $x, y \in \tilde{U}_i$  let  $W := \mathcal{M} \cap B(x, 2\tilde{\delta}) \cap B(y, 2\tilde{\delta})$ , where  $B(x, 2\tilde{\delta})$  is the Euclidean ball of radius  $2\tilde{\delta}$  centered at  $x$ . For every  $z \in W$  we have

$$|f(x) - f(y)|^2 \leq 2|f(x) - f(z)|^2 + 2|f(y) - f(z)|^2,$$

and in particular

$$\begin{aligned} |f(x) - f(y)|^2 &\leq 2 \frac{1}{\text{Vol}_{\mathcal{M}}(W)} \int_W |f(x) - f(z)|^2 d\text{Vol}_{\mathcal{M}}(z) + 2 \frac{1}{\text{Vol}_{\mathcal{M}}(W)} \int_W |f(y) - f(z)|^2 d\text{Vol}_{\mathcal{M}}(z) \\ &\leq \frac{2}{\text{Vol}_{\mathcal{M}}(W)} \int_{\mathcal{M}} \eta\left(\frac{|x-z|}{2\tilde{\delta}}\right) |f(x) - f(z)|^2 d\text{Vol}_{\mathcal{M}}(z) \\ &\quad + \frac{2}{\text{Vol}_{\mathcal{M}}(W)} \int_{\mathcal{M}} \eta\left(\frac{|y-z|}{2\tilde{\delta}}\right) |f(y) - f(z)|^2 d\text{Vol}_{\mathcal{M}}(z) \\ &\leq \frac{C}{\tilde{\delta}^m} \int_{\mathcal{M}} \eta\left(\frac{|x-z|}{2\tilde{\delta}}\right) |f(x) - f(z)|^2 d\text{Vol}_{\mathcal{M}}(z) + \frac{C}{\tilde{\delta}^m} \int_{\mathcal{M}} \eta\left(\frac{|y-z|}{2\tilde{\delta}}\right) |f(y) - f(z)|^2 d\text{Vol}_{\mathcal{M}}(z). \end{aligned}$$

Integrating with respect to  $x$  and  $y$  in both sides of the inequality we get

$$\begin{aligned} n \int_{\tilde{U}_i} \int_{\tilde{U}_i} |f(x) - f(y)|^2 \tilde{\rho}_n(x) \tilde{\rho}_n(y) d\text{Vol}_{\mathcal{M}}(x) d\text{Vol}_{\mathcal{M}}(y) \\ \leq \frac{2C}{\tilde{\delta}^m} \int_{\tilde{U}_i} \int_{\mathcal{M}} \eta\left(\frac{|x-z|}{2\tilde{\delta}}\right) |f(x) - f(z)|^2 \tilde{\rho}_n(x) \tilde{\rho}_n(z) d\text{Vol}_{\mathcal{M}}(x) d\text{Vol}_{\mathcal{M}}(z). \end{aligned}$$

Summing over all  $i = 1, \dots, n$  we deduce from the above expression that

$$\|\tilde{P}^* \tilde{P} f - f\|_{L^2(\tilde{\mu}_n)}^2 \leq C\tilde{\delta}^2 \tilde{E}_{2\tilde{\delta}}(f) \leq C\tilde{\delta}^2 D_2(f),$$

where for the last inequality we have used the same arguments as in the proof of Proposition 4.1.

To show the second assertion we notice that from  $\|u\|_{L^2(\mu_n)} = \|\tilde{P}^* u\|_{L^2(\tilde{\mu}_n)}$  and the triangle inequality we get

$$\begin{aligned} \|\tilde{\mathcal{I}}u\|_{L^2(\tilde{\mu}_n)} - \|u\|_{L^2(\mu_n)} &\leq \|\Lambda_{\varepsilon-2\tilde{\delta}} \tilde{P}^* u - \tilde{P}^* u\|_{L^2(\tilde{\mu}_n)} \\ &\leq \left(1 + \frac{\|\rho - \tilde{\rho}_n\|_{\infty}}{\rho_{\min}}\right) \cdot \|\Lambda_{\varepsilon-2\tilde{\delta}} \tilde{P}^* u - \tilde{P}^* u\|_{L^2(\mu)} \\ &\leq \left(1 + \frac{\|\rho - \tilde{\rho}_n\|_{\infty}}{\rho_{\min}}\right) \cdot C\varepsilon \sqrt{E_{\varepsilon-2\tilde{\delta}}(\tilde{P}^* u)} \\ &\leq C\varepsilon \sqrt{b_{\varepsilon}(u)}, \end{aligned}$$

where for the third inequality we have used Lemma 8 in [22] (notice that our definition of  $E_r$  has an extra factor of  $r^{-2}$  when compared to the definition in [22]), and for the last one we have used Lemma 4.7 and (4.10). Also, notice that

$$\|\tilde{I}u\|_{L^2(\tilde{\mu}_n)} = \|\Lambda_r \tilde{P}^* u\|_{L^2(\tilde{\mu}_n)} \leq C\|\tilde{P}^* u\|_{L^2(\tilde{\mu}_n)} = C\|u\|_{L^2(\mu_n)},$$

where the inequality follows from Lemma 8 in [22] and the fact that by introducing a multiplicative constant we can change integrals with respect to  $\rho$  with integrals with respect to  $\tilde{\rho}_n$

and vice-versa (according to Proposition (2.12)). Therefore,

$$\begin{aligned} \left| \|\tilde{\mathcal{I}}u\|_{L^2(\tilde{\mu}_n)}^2 - \|u\|_{L^2(\mu_n)}^2 \right| &\leq \left| \|\tilde{\mathcal{I}}u\|_{L^2(\tilde{\mu}_n)} - \|u\|_{L^2(\mu_n)} \right| \left( \|\tilde{\mathcal{I}}u\|_{L^2(\tilde{\mu}_n)} + \|u\|_{L^2(\mu_n)} \right) \\ &\leq \left| \|\tilde{\mathcal{I}}u\|_{L^2(\tilde{\mu}_n)} - \|u\|_{L^2(\mu_n)} \right| \left( \|\tilde{\mathcal{I}}u\|_{L^2(\tilde{\mu}_n)} + \|u\|_{L^2(\mu_n)} \right) \\ &\leq C\varepsilon\sqrt{b_\varepsilon(u)}\|u\|_{L^2(\mu_n)}. \end{aligned}$$

Finally, we use (4.11) to compare  $\|\tilde{\mathcal{I}}u\|_{L^2(\tilde{\mu}_n)}^2$  and  $\|\tilde{\mathcal{I}}u\|_{L^2(\mu)}^2$ .  $\square$

With Proposition 4.1 and 4.2 in hand the proof of the first part of Theorem 2.4 now follows from standard arguments.

*Proof of 1) in Theorem 2.4.* Let  $f_1, \dots, f_l$  be an orthonormal set (in  $L^2(\mu)$ ) of eigenfunctions of  $\Delta_\rho$  corresponding to its first  $l$  eigenvalues. For  $i = 1, \dots, l$  let

$$v_i := \tilde{P}f_i.$$

Applying the first part of Proposition 4.2 to functions  $f$  of the form:

$$f := f_i - f_j$$

we can get the bound,

$$|\langle f_i, f_j \rangle_{L^2(\mu)} - \langle v_i, v_j \rangle_{L^2(\mu_n)}| \leq C\sqrt{\lambda_l}\tilde{\delta} + C(\theta + \tilde{\delta}) \leq \frac{1}{2l},$$

where the last inequality is valid thanks to our smallness assumption for  $\tilde{\delta}$  and  $\theta$ . Since  $\langle f_i, f_j \rangle_{L^2(\mu)} = \delta_{ij}$ , the above inequality implies that the vectors  $v_1, \dots, v_l$  are linearly independent, and so the subspace  $S := \text{Span}\{v_1, \dots, v_l\}$  has dimension  $l$ . We can use (2.5) to conclude that

$$\lambda_l^\varepsilon \leq \frac{1}{2} \max_{\substack{v \in S \\ \|v\|_{L^2(\mu_n)}=1}} b_\varepsilon(v).$$

Now, each element  $v$  of  $S$  is of the form

$$v = \sum_{i=1}^l a_i \tilde{P}f_i,$$

for some coefficients  $a_i$ , so that

$$v = \tilde{P}\left(\sum_{i=1}^l a_i f_i\right) =: \tilde{P}(f),$$

and in particular,

$$\frac{1}{2}D_2(f) = \langle \Delta_\rho f, f \rangle_{L^2(\mu)} \leq \lambda_l \|f\|_{L^2(\mu)}^2.$$

By Proposition 4.1 we know that,

$$\frac{1}{2}b_\varepsilon(v) = \frac{1}{2}b_\varepsilon(\tilde{P}f) \leq (1 + C\frac{\tilde{\delta}}{\varepsilon} + C\varepsilon + C\theta + C\tilde{\delta})\frac{\sigma_\eta}{2}D(f) \leq (1 + C\frac{\tilde{\delta}}{\varepsilon} + C\varepsilon + C\theta + C\tilde{\delta})\sigma_\eta\lambda_l \|f\|_{L^2(\mu)}^2.$$

If  $\|v\|_{L^2(\mu_n)} = 1$ , Proposition 4.2 implies

$$\|f\|_{L^2(\mu)}^2 \leq 1 + C\tilde{\delta}\sqrt{\lambda_l} + C(\theta + \tilde{\delta}).$$

Thus,

$$\lambda_l^\varepsilon \leq \sigma_\eta\lambda_l + C(\tilde{\delta}\sqrt{\lambda_l} + \frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta)\lambda_l.$$

This establishes the upper bound for  $\lambda_l^\varepsilon$  in terms of  $\lambda_l$ .

To obtain the lower bound, we let  $u_1, \dots, u_l$  be an orthonormal set of eigenvectors of  $\mathcal{L}^\varepsilon$  corresponding to its first  $l$  eigenvalues. We define functions  $f_i \in L^2(\mu)$ , according to

$$f_i := \tilde{\mathcal{I}}u_i, \quad i = 1, \dots, l.$$

Thanks to the smallness assumption of the parameters  $\tilde{\delta}, \varepsilon, \theta$ , we see that  $S := \text{span}\{f_1, \dots, f_l\}$  has dimension  $l$ . Just as in the proof of the upper bound we can use the second parts of Propositions 4.1 and 4.2 to conclude that

$$\sigma_\eta \lambda_l \leq \lambda_l^\varepsilon + C \left( \varepsilon \sqrt{\lambda_l^\varepsilon} + \frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta \right) \lambda_l^\varepsilon.$$

However, using the upper bound for  $\lambda_l^\varepsilon$ , we can replace all the appearances of  $\lambda_l^\varepsilon$  in the above error terms with  $\lambda_l$ , and obtain the desired inequality.  $\square$

We can now prove Theorem 2.7. We start with its second part.

*Proof of 2) in Theorem 2.7.* A given eigenvalue  $\lambda > 0$  of  $\Delta_\rho$  is equal to  $\lambda_{i+1}, \dots, \lambda_{i+k}$  for some  $i$  and some  $k$ , where  $k$  is the multiplicity of  $\lambda$ . Associated to  $\lambda$  we define the gap  $\gamma_\lambda$  according to:

$$(4.12) \quad \gamma_\lambda := \frac{1}{2} \min\{|\lambda - \lambda_i|, |\lambda - \lambda_{i+k+1}|\}.$$

Now, for every  $N \in \mathbb{N}$ , we can pick  $\varepsilon, \theta$  and  $\tilde{\delta}$  to be small enough so that for every  $l = 1, \dots, N$  we have

$$C \left( \varepsilon \sqrt{\lambda_l} + \frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta \right) \lambda_l \leq \gamma_{\lambda_l}.$$

In particular, for such choice of parameters and according to Theorem 2.4, with probability greater than  $1 - Cn \exp(-n\theta^2\tilde{\delta}^m)$  we have

$$(4.13) \quad |\lambda_l^\varepsilon - \sigma_\eta \lambda_l| \leq \gamma_{\lambda_l}, \quad \forall l = 1, \dots, N.$$

Let  $\lambda$  be one of  $\lambda_1, \dots, \lambda_N$ . By making  $N$  slightly larger if necessary, we can assume that  $i+k \leq N$  (where  $i$  and  $k$  are as defined earlier for  $\lambda$ ). Let  $S$  be a subspace of  $L^2(\mu_n)$  spanned by eigenvectors of  $\mathcal{L}^\varepsilon$  with corresponding eigenvalues  $\lambda_{i+1}^\varepsilon, \dots, \lambda_{i+k}^\varepsilon$  and let us denote by  $P_S$  the orthogonal projection onto  $S$ , and by  $P_S^\perp$  the orthogonal projection onto the orthogonal complement of  $S$ . We know that an arbitrary unit norm eigenfunction  $f$  of  $\Delta_\rho$  with eigenvalue  $\lambda$  is at least  $C^3(\mathcal{M})$ , and since  $\Delta_\rho u = \lambda u$ , we obviously have that  $\Delta_\rho u$  is also  $C^3(\mathcal{M})$ . Restricting  $f$  and  $\Delta_\rho f$  to the point cloud  $X$ , we can view both of these functions as elements in  $L^2(\mu_n)$ , and we can also see that

$$P_S^\perp \Delta_\rho f = \lambda P_S^\perp f = \lambda \sum_{j \neq i+1, \dots, i+k} \langle f, \psi_j^\varepsilon \rangle_{L^2(\mu_n)} \psi_j^\varepsilon$$

where  $\psi_1^\varepsilon, \dots, \psi_n^\varepsilon$  is an orthonormal basis of eigenvectors of  $\mathcal{L}^\varepsilon$  with corresponding eigenvalues  $\lambda_1^\varepsilon, \dots, \lambda_n^\varepsilon$ . Likewise, we can compute  $\mathcal{L}^\varepsilon f$  (again thinking of  $f$  as its restriction to  $X$ ) and see that

$$P_S^\perp \mathcal{L}^\varepsilon f = \sum_{j \neq i+1, \dots, i+k} \lambda_j^\varepsilon \langle f, \psi_j^\varepsilon \rangle_{L^2(\mu_n)} \psi_j^\varepsilon.$$

Subtracting these two expressions and using the orthogonality of the  $\psi_j^\varepsilon$  we obtain

$$\min\{|\lambda_i^\varepsilon - \lambda|, |\lambda_{i+k+1}^\varepsilon - \lambda|\} \|P_S^\perp f\|_{L^2(\mu_n)} \leq \|P_S^\perp (\mathcal{L}^\varepsilon f - \Delta_\rho f)\|_{L^2(\mu_n)} \leq \|\mathcal{L}^\varepsilon f - \Delta_\rho f\|_{L^2(\mu_n)}.$$

However, from (4.13) we have

$$\gamma_\lambda \leq \min\{|\lambda_i^\varepsilon - \lambda|, |\lambda_{i+k+1}^\varepsilon - \lambda|\},$$

and so

$$\|P_S^\perp f\|_{L^2(\mu_n)} \leq \frac{1}{\gamma_\lambda} \|\mathcal{L}^\varepsilon f - \Delta_\rho f\|_{L^2(\mu_n)}.$$

Naturally, since  $P_S^\perp f = f - P_S f$  we obtain

$$(4.14) \quad \|f - P_S f\|_{L^2(\mu_n)} \leq \frac{1}{\gamma_\lambda} \|\mathcal{L}^\varepsilon f - \Delta_\rho f\|_{L^2(\mu_n)}.$$

On the other hand, from the pointwise consistency results from Theorem 3.3 it follows that if  $f_1, \dots, f_k$  is an orthonormal basis for the eigenspace of eigenfunctions of  $\Delta_\rho$  with eigenvalue  $\lambda$ , then with probability greater than  $1 - 2kn \exp(-cn\varepsilon^{m+4})$  we have

$$\|\mathcal{L}^\varepsilon f_j - \Delta_\rho f_j\|_{L^2(\mu_n)} \leq C\varepsilon, \quad \forall j = 1, \dots, k.$$

Combining the above with (4.14) we conclude that with probability greater than  $1 - Cn \exp(-n\theta^2\tilde{\delta}^m) - 2kn \exp(-cn\varepsilon^{m+4})$  we can find an orthonormal set  $v_1, \dots, v_k$  spanning  $S$  such that

$$\|f_j - v_j\|_{L^2(\mu_n)} \leq C\varepsilon, \quad \forall j = 1, \dots, k.$$

In turn this implies that if  $u_1, \dots, u_k$  is a family of orthonormal eigenfunctions of  $\mathcal{L}^\varepsilon$  with corresponding eigenvalues  $\lambda_{i+1}^\varepsilon, \dots, \lambda_{i+k}^\varepsilon$ , then there exists an orthonormal set  $\tilde{f}_1, \dots, \tilde{f}_k$  of eigenfunctions of  $\Delta_\rho$  with eigenvalue  $\lambda$  such that

$$\|u_i - \tilde{f}_i\|_{L^2(\mu_n)} \leq C\varepsilon.$$

This implies the desired result.  $\square$

*Proof of 1) in Theorem 2.7.* In general it will not be meaningful to use the pointwise consistency results from Theorem 3.3 because  $\varepsilon$  here can be smaller than what the probabilistic estimates in Theorem 3.3 allow it to be. Thus, we use an energy estimate based on Propositions 4.1 and 4.2 just as in [22]. Since this argument has been shown in detail in [22], here we only present the proof for the first non-trivial eigenvectors.

The first non-trivial eigenvalue  $\lambda$  of  $\Delta_\rho$  is equal to  $\lambda_2, \dots, \lambda_{k+1}$  where  $k$  is its multiplicity. Let  $f$  be an eigenfunction of  $\Delta_\rho$  with eigenvalue  $\lambda$  and let  $u$  be equal to  $u = \tilde{P}f$  where  $\tilde{P}$  is as in (4.1). Consider now the span of a set of orthonormal eigenvectors of  $\mathcal{L}^\varepsilon$  with eigenvalues  $\lambda_2^\varepsilon, \dots, \lambda_{k+1}^\varepsilon$ , and denote this linear subspace by  $S$ .

We see from Proposition 4.1 that

$$(4.15) \quad \begin{aligned} \left(1 + C \left(\frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta\right)\right) \sigma_\eta \lambda_2 &= \left(1 + C \left(\frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta\right)\right) \sigma_\eta D_2(f) \geq b_\varepsilon(u) = \langle \mathcal{L}^\varepsilon u, u \rangle_{L^2(\mu_n)} \\ &\geq \lambda_2^\varepsilon \|P_S u\|_{L^2(\mu_n)}^2 + \lambda_{k+2}^\varepsilon \|u - P_S u\|_{L^2(\mu_n)}^2. \\ &= \lambda_2^\varepsilon (\|u\|_{L^2(\mu_n)}^2 - \|u - P_S u\|_{L^2(\mu_n)}^2) + \lambda_{k+2}^\varepsilon \|u - P_S u\|_{L^2(\mu_n)}^2. \end{aligned}$$

Using the estimates from the first part of Theorem 2.4 and Proposition 4.2 we have that as long as  $\tilde{\delta}$  and  $\theta$  are small enough, then with probability greater than  $1 - Cn \exp(-cn\theta^2\tilde{\delta}^m)$ ,

$$\begin{aligned} |\sigma_\eta \lambda_2 - \lambda_2^\varepsilon| &\leq C \left(\frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta\right) \leq \frac{\gamma_\lambda}{2} \\ |\sigma_\eta \lambda_{k+2} - \lambda_{k+2}^\varepsilon| &\leq C \left(\frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta\right) \leq \frac{\gamma_\lambda}{2} \end{aligned}$$

and

$$|1 - \|u\|_{L^2(\mu_n)}^2| \leq C(\theta + \tilde{\delta}),$$

where  $C$  is some constant that may depend on  $\lambda$  and where  $\gamma_\lambda$  is the spectral gap defined in (4.12). Combining the above inequalities with (4.15) we conclude that

$$\|u - P_S u\|_{L^2(\mu_n)} \leq \left(\frac{C}{\gamma_\lambda} \left(\frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta\right)\right)^{1/2}.$$

Now, by definition

$$u(x_i) - f(x_i) = \tilde{P}f(x_i) - f(x_i) = n \int_{\tilde{U}_i} (f(x) - f(x_i)) \tilde{\rho}_n(x) d\text{Vol}_{\mathcal{M}}(x).$$

We notice that the last term is in absolute value less than

$$\|\nabla f\|_{L^\infty(\mu)} \tilde{\delta},$$

and that  $\|\nabla f\|_{L^\infty(\mu)}$  is finite because  $f$  is actually  $C^3(\mathcal{M})$ . In particular,

$$\|f - u\|_{L^2(\mu_n)} \leq C_{\mathcal{M}} \tilde{\delta},$$

Therefore,

$$\|f - P_S \tilde{P}f\|_{L^2(\mu_n)} \leq \left( \frac{C}{\gamma_\lambda} \left( \frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta \right) \right)^{1/2} + C_{\mathcal{M}, \lambda} \tilde{\delta}.$$

From this it is straightforward to see that if  $f_1, \dots, f_k$  form an orthonormal basis for the eigenspace of eigenfunctions of  $\Delta_\rho$  with eigenvalue  $\lambda$ , then we can find an orthonormal set  $v_1, \dots, v_k$  spanning  $S$  such that

$$\|f_j - v_j\|_{L^2(\mu_n)} \leq \left( \frac{C}{\gamma_\lambda} \left( \frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta \right) \right)^{1/2} + C_{\mathcal{M}, \lambda} \tilde{\delta}, \quad \forall j = 1, \dots, k.$$

In turn this implies that if  $u_1, \dots, u_k$  is a family of orthonormal eigenfunctions of  $\mathcal{L}^\varepsilon$  with corresponding eigenvalues  $\lambda_2^\varepsilon, \dots, \lambda_{k+1}^\varepsilon$ , then there exists an orthonormal set  $\tilde{f}_1, \dots, \tilde{f}_k$  of eigenfunctions of  $\Delta_\rho$  with eigenvalue  $\lambda$  such that

$$\|u_j - \tilde{f}_j\|_{L^2(\mu_n)} \leq \left( \frac{C}{\gamma_\lambda} \left( \frac{\tilde{\delta}}{\varepsilon} + \varepsilon + \theta \right) \right)^{1/2} + C_{\mathcal{M}, \lambda} \tilde{\delta}, \quad \forall j = 1, \dots, k.$$

This implies the desired result. □

We are ready to prove 2) in Theorem 2.4.

*Proof of 2) in Theorem 2.4.* First notice that if  $\lambda > 0$  is an eigenvalue of  $\Delta_\rho$  and  $\lambda^\varepsilon > 0$  is an eigenvalue of the graph Laplacian  $\mathcal{L}^\varepsilon$ , then,

$$(4.16) \quad |\sigma_\eta \lambda - \lambda^\varepsilon| \leq \inf_{\substack{f \in S(\Delta_\rho, \lambda) \\ u \in S(\mathcal{L}^\varepsilon, \lambda^\varepsilon)}} \frac{\|\mathcal{L}^\varepsilon f - \sigma_\eta \Delta_\rho f\|_{L^2(\mu_n)}}{|\langle u, f \rangle_{L^2(\mu_n)}|}$$

where in the above  $S(\Delta_\rho, \lambda)$  is the set of unit norm eigenfunctions of  $\Delta_\rho$  with eigenvalue  $\lambda$ , and  $S(\mathcal{L}^\varepsilon, \lambda^\varepsilon)$  is the set of eigenvectors of  $\mathcal{L}^\varepsilon$  with eigenvalue  $\lambda^\varepsilon$ . To see this, let  $f \in S(\Delta_\rho, \lambda)$  and  $u \in S(\mathcal{L}^\varepsilon, \lambda^\varepsilon)$ . Then, we can restrict  $f$  and  $\Delta_\rho f$  to the point cloud and get

$$\begin{aligned} \lambda^\varepsilon \langle u, f \rangle_{L^2(\mu_n)} &= \langle \mathcal{L}^\varepsilon u, f \rangle_{L^2(\mu_n)} \\ &= \langle u, \mathcal{L}^\varepsilon f \rangle_{L^2(\mu_n)} \\ &= \langle u, \mathcal{L}^\varepsilon f - \sigma_\eta \lambda f + \sigma_\eta \lambda f \rangle_{L^2(\mu_n)} \\ &= \sigma_\eta \lambda \langle u, f \rangle_{L^2(\mu_n)} + \langle u, \mathcal{L}^\varepsilon f - \sigma_\eta \Delta_\rho f \rangle_{L^2(\mu_n)}. \end{aligned}$$

Rearranging the terms we obtain

$$(4.17) \quad \lambda^\varepsilon - \sigma_\eta \lambda = \frac{\langle u, \mathcal{L}^\varepsilon f - \sigma_\eta \Delta_\rho f \rangle_{L^2(\mu_n)}}{\langle u, f \rangle_{L^2(\mu_n)}},$$

provided  $\langle u, f \rangle_{L^2(\mu_n)} \neq 0$ . Applying the Cauchy-Schwarz inequality we obtain

$$|\lambda^\varepsilon - \sigma_\eta \lambda| \leq \frac{\|\mathcal{L}^\varepsilon f - \sigma_\eta \Delta_\rho f\|_{L^2(\mu_n)}}{|\langle u, f \rangle_{L^2(\mu_n)}|}$$

Take  $\lambda = \lambda_l$  and  $\lambda^\varepsilon = \lambda_l^\varepsilon$  in the above formula. From Theorem 2.7 we know that with probability at least  $1 - n \exp(-n\theta^2\tilde{\delta}^m) - Cn \exp(-cn\varepsilon^{m+4})$  there are unit norm eigenvectors  $u$  and  $f$  of  $\mathcal{L}^\varepsilon$  and  $\Delta_\rho$  with corresponding eigenvalues  $\lambda_l^\varepsilon$  and  $\lambda_l$ , such that

$$\|u_l - f_l\|_{L^2(\mu_n)} \leq C\varepsilon,$$

so that in particular

$$|\langle u, f \rangle_{L^2(\mu_n)}| \geq 1 - C\varepsilon \geq \frac{1}{2},$$

when  $\varepsilon$  is small enough. Also, from Theorem 3.11 we know that with probability at least  $1 - Cn \exp(-cn\varepsilon^{m+4})$  we have

$$\|\mathcal{L}^\varepsilon f - \sigma_\eta \Delta_\rho f\|_{L^2(\mu_n)} \leq C\varepsilon.$$

The result now follows.  $\square$

**4.2. Undirected  $k$ -NN graph.** After defining appropriate interpolation and discretization maps, and establishing their relevant properties, we will be able to follow the exact same proof as the one we presented in the  $\varepsilon$ -graph case. For this reason we focus on establishing Proposition 4.4 and 4.5 below and skip the rest of the details.

In what follows, for simplicity we restrict our analysis to the kernel  $\eta$  given by

$$\eta(t) = \begin{cases} 1 & t < 1 \\ 0 & t > 1, \end{cases}$$

as it already captures the main differences between the  $k$ -NN and  $\varepsilon$  graph settings. Only minor, although slightly more cumbersome, modifications are needed to consider the general case. Let  $\psi$  be defined as (4.4). Notice that in this case we have

$$(4.18) \quad \psi(t) \leq \frac{1}{\sigma_\eta} \eta(t), \quad \forall t > 0.$$

For a function  $r : \mathcal{M} \rightarrow (0, \infty)$ , and for a function  $f \in L^2(\mathcal{M})$  we define

$$\Lambda_r f(x) := \frac{1}{\tau(x)} \int_{\mathcal{M}} K_{r(y)}(x, y) f(y) d\mu(y),$$

as in Section 4.1 where the only difference is that now the bandwidth  $r$  is allowed to change in space. The normalization constant continues to be of the form

$$\tau(x) := \int_{\mathcal{M}} K_{r(y)}(x, y) d\mu(y).$$

It is important to notice that in the definition of  $\Lambda_r f$ , the function  $r$  is evaluated at  $y$  (and not at  $x$ ) so as to have an expression for the gradient of  $\Lambda_r f$  which does not depend on derivatives of  $r$  (which we may even take it to be discontinuous).

*Remark 4.3.* In what follows we set  $\varepsilon$  to be the function defined as in (3.19). Notice that from the fact that  $0 < \rho_{min} \leq \rho \leq \rho_{max}$  we see that  $\varepsilon(x)$  and  $(\frac{k}{n})^{1/m}$  are of the same order for every  $x \in \mathcal{M}$ . Also, since  $k$  has been assumed to be much smaller than  $n$ , then  $\varepsilon(x)^2$ , which is of the same order as  $(k/n)^{2/m}$ , is much smaller than  $\varepsilon(x)$ . We will use  $\varepsilon_{max}$  to denote the maximum of  $\varepsilon$ . In particular,

$$c \left( \frac{k}{n} \right)^{1/m} \leq \varepsilon_{min} \leq \varepsilon_{max} \leq C \left( \frac{k}{n} \right)^{1/m}.$$

**Proposition 4.4** (Inequality for Dirichlet energies). *Let  $k$ ,  $\tilde{\delta}$ , and  $\theta$  be fixed numbers satisfying Assumptions 2.3. Then, with probability greater than  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m)$  we have:*

(1) For any  $f \in L^2(\mu)$ ,

$$b_k(\tilde{P}f) \leq (1 + C \left(\frac{k}{n}\right)^{1/m} + C\tilde{\delta} \left(\frac{n}{k}\right)^{1/m} + C\theta) \sigma_\eta D_{1-2/m}(f)$$

(2) For any  $u \in L^2(\mu_n)$ ,

$$\sigma_\eta D_{1-2/m}(\tilde{\mathcal{I}}u) \leq (1 + C \left(\frac{k}{n}\right)^{1/m} + C\tilde{\delta} \left(\frac{n}{k}\right)^{1/m} + C\theta) b_k(u).$$

**Proposition 4.5** (Discretization and interpolation maps are almost isometries). *Suppose  $\theta, \tilde{\delta}$  and  $k$  satisfy Assumptions 2.3. Then, with probability at least  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m)$  we have:*

(1) For every  $f \in L^2(\mu)$ ,

$$\left| \|f\|_{L^2(\mu)}^2 - \|\tilde{P}f\|_{L^2(\mu_n)}^2 \right| \leq C\tilde{\delta} \|f\|_{L^2(\mu)} \sqrt{D_{1-2/m}(f)} + C(\theta + \tilde{\delta}) \|f\|_{L^2(\mu)}^2.$$

(2) For every  $u \in L^2(\mu_n)$ ,

$$\left| \|u\|_{L^2(\mu_n)}^2 - \|\tilde{\mathcal{I}}u\|_{L^2(\mu)}^2 \right| \leq C \left(\frac{k}{n}\right)^{1/m} \|u\|_{L^2(\mu)} \sqrt{b_k(u)} + C(\theta + \tilde{\delta}) \|u\|_{L^2(\mu_n)}^2.$$

In order to show the above propositions we will first need to introduce some non-local energies  $E_r$  and  $\tilde{E}_r$  (we will use the same notation for simplicity as in the  $\varepsilon$ -graph case) defined in terms of the spatially varying length-scale  $r : \mathcal{M} \rightarrow (0, \infty)$  by

$$E_r(f) := \int_{\mathcal{M}} \int_{\mathcal{M}} \eta \left( \frac{d_{\mathcal{M}}(x, y)}{r(y)} \right) \frac{(f(x) - f(y))^2}{r(y)^{m+2} \rho(y)^{2/m-1}} d\text{Vol}_{\mathcal{M}}(x) d\text{Vol}_{\mathcal{M}}(y), \quad f \in L^2(\mu)$$

$$\tilde{E}_r(f) := \int_{\mathcal{M}} \int_{\mathcal{M}} \eta \left( \frac{d_{\mathcal{M}}(x, y)}{r(y)} \right) \frac{(f(x) - f(y))^2}{r(y)^{m+2} \tilde{\rho}_n(y)^{2/m-1}} d\text{Vol}_{\mathcal{M}}(x) d\text{Vol}_{\mathcal{M}}(y), \quad f \in L^2(\mu),$$

where  $\tilde{\rho}_n$  is the density from Proposition 2.12. Notice that the  $L^\infty$  bound for the difference between  $\rho$  and  $\tilde{\rho}_n$  implies

$$(4.19) \quad (1 - C(\theta + \tilde{\delta})) E_r(f) \leq \tilde{E}_r(f) \leq (1 + C(\theta + \tilde{\delta})) E_r(f), \quad \forall f \in L^2(\mu),$$

These inequalities are analogous to the ones in (4.7).

The following three lemmas will be the main tools for proving Propositions 4.4 and 4.5.

**Lemma 4.6.** *Suppose that  $k, \tilde{\delta}$  and  $\theta$  satisfy Assumptions 2.3, and let  $\varepsilon : \mathcal{M} \rightarrow (0, \infty)$  be as in (2.14). Then, with probability at least  $1 - Cn \exp(-Cn\theta^2\tilde{\delta}^m)$  we have for all  $i = 1, \dots, n$*

$$(4.20) \quad \varepsilon(x_i) - \tilde{C}(\varepsilon_{\max}^2 + \tilde{\delta}\varepsilon_{\max} + \theta\varepsilon_{\max}) \leq \varepsilon_k(x_i) \leq \varepsilon(x_i) + \tilde{C}(\varepsilon_{\max}^2 + \tilde{\delta}\varepsilon_{\max} + \theta\varepsilon_{\max}).$$

*Proof.* We prove the upper bound. The lower bound is proved similarly.

Let  $\tilde{\rho}_n$  be the density from Proposition 2.12, which exists with probability at least  $1 - Cn \exp(-n\theta^2\tilde{\delta}^m)$ . First of all for  $x \in \mathcal{M}$  and  $0 < \gamma_1 < \gamma_2$  small enough define

$$A(x, \gamma_1, \gamma_2) := B_{\mathcal{M}}(x, \gamma_2) \setminus B_{\mathcal{M}}(x, \gamma_1).$$

Then, we see that

$$C_1 \gamma_1^{m-1} (\gamma_2 - \gamma_1) \leq \tilde{\mu}_n(A(x, \gamma_1, \gamma_2)) \leq C_2 \gamma_1^{m-1} (\gamma_2 - \gamma_1)$$

where  $C_1, C_2$  do not depend on  $x$  nor  $\gamma_1, \gamma_2$  as long as these numbers are small enough.

Pick  $\gamma_2 = \varepsilon(x) + \tilde{C}(\varepsilon(x)^2 + \tilde{\delta} + \theta)$  for a large enough constant  $\tilde{C}$  that will be chosen soon. We also let  $\gamma_1 = \varepsilon(x)$ . Then,

$$\begin{aligned}
\tilde{\mu}_n(B_{\mathcal{M}}(x_i, \gamma_2)) &\geq \int_{B_{\mathcal{M}}(x_i, \varepsilon(x_i))} \tilde{\rho}_n(x) d\text{Vol}_{\mathcal{M}}(x) + \tilde{\mu}_n(A(x_i, \gamma_1, \gamma_2)) \\
&\geq (1 - C\varepsilon(x_i))(1 - C(\theta + \delta))\rho(x_i) \text{Vol}_{\mathcal{M}}(B_{\mathcal{M}}(x_i, \varepsilon(x_i))) + \tilde{\mu}_n(A(x_i, \gamma_1, \gamma_2)) \\
&\geq (1 - C\varepsilon(x_i))(1 - C(\theta + \delta))(1 - CK\varepsilon(x_i)^{m+2})\rho(x_i)\alpha_m(\varepsilon(x_i))^m + \tilde{\mu}_n(A(x_i, \gamma_1, \gamma_2)) \\
&\geq (1 - C\varepsilon(x_i))(1 - C(\theta + \delta))(1 - CK\varepsilon(x_i)^{m+2})\frac{k}{n} + \tilde{\mu}_n(A(x_i, \gamma_1, \gamma_2)) \\
&\geq \frac{k}{n} - C(\varepsilon(x_i) + \tilde{\delta} + \theta)\varepsilon(x_i)^m + \tilde{\mu}_n(A(x_i, \gamma_1, \gamma_2)) \\
&\geq \frac{k}{n} - C(\varepsilon(x_i) + \tilde{\delta} + \theta)\varepsilon(x_i)^m + C_1\varepsilon(x_i)^{m-1}(\tilde{C}(\varepsilon(x_i)^2 + \tilde{\delta}\varepsilon(x_i) + \theta\varepsilon(x_i))).
\end{aligned}$$

We pick  $\tilde{C}$  precisely so that the above is greater than  $k/n$ . Finally, we see that

$$\mu_n(B_{\mathcal{M}}(x_i, \gamma_2 + \tilde{\delta})) = \tilde{\mu}_n(\tilde{T}^{-1}(B_{\mathcal{M}}(x_i, \gamma_2 + \tilde{\delta}))) \geq \tilde{\mu}_n(B_{\mathcal{M}}(x_i, \gamma_2)) \geq k/n,$$

from where it now follows that  $\varepsilon_k(x_i) \leq \gamma_2 + \tilde{\delta}$ .  $\square$

**Lemma 4.7.** *Let  $r : \mathcal{M} \rightarrow (0, \infty)$  be an arbitrary function which is bounded away from zero and for which  $r_{max}$  is sufficiently small. Then, for any  $f \in L^2(\mu)$  we have*

$$E_r(f) \leq (1 + Cr_{max})(1 + \beta_r)^{2m+4}\sigma_{\eta}D_{1-2/m}(f).$$

Where in the above  $\beta_r$  is defined as

$$(4.21) \quad \beta_r := \sup_{x, y \in \mathcal{M}, d_{\mathcal{M}}(x, y) \leq r_{max}} \frac{r(x) - r(y)}{r_{min}}$$

As we will later see, in our setting we will work with  $r$  for which  $r_{max}, \beta_r \ll 1$ , so that Lemma 4.7 implies that  $E_r(f)$  is smaller than a quantity that up to leading order is equal to  $\sigma_{\eta}D_{1-2/m}(f)$ .

*Proof.* By density of smooth functions in  $H^1(\mu)$  it is enough to prove the result for  $f$  smooth.

In what follows we will use the geodesic flow  $\Phi_t : T\mathcal{M} \rightarrow T\mathcal{M}$ , which maps a point  $(x, v) \in \mathcal{M} \times T_x\mathcal{M}$  in  $\mathcal{M}$ 's tangent bundle into the point  $\Phi_t(x, v) = (\exp_x(tv), d\exp_x(v))$  (i.e. flow  $x$  for  $t$  seconds along the geodesic emanating from it with initial velocity  $v$ ). We will use  $\xi = (x, v)$  to represent a generic point in the tangent bundle and abuse notation slightly to write things like  $g(\xi) = g(x)$ , when  $g$  is a real valued function on  $\mathcal{M}$ . We will also use  $df$  (i.e. the differential of  $f$ ) which is the 1-form that when acting on a tangent vector  $v$  returns the directional derivative of  $f$  in the direction  $v$ , and will write things like  $df(\xi)$  to denote the directional derivative of  $f$  at the point  $x$  in the direction  $v$ .

With this notation in hand, and following the proof of Lemma 3.3 in [7], we obtain for every  $x \in \mathcal{M}$

$$\int_{B_{\mathcal{M}}(x, r(x))} |f(y) - f(x)|^2 d\text{Vol}_{\mathcal{M}}(y) \leq \int_{B(0, r(x)) \subseteq T_x\mathcal{M}} \int_0^1 |df(\Phi_t(x, v))|^2 dt dv.$$

Dividing by  $r(x)^{m+2}\rho(x)^{2/m-1}$  and integrating over  $x$  we get:

$$\begin{aligned}
& (4.22) \quad \int_{\mathcal{M}} \int_{B_{\mathcal{M}}(y, r(y))} \frac{|f(y) - f(x)|^2}{r(x)^{m+2}\rho(x)^{2/m-1}} d\text{Vol}_{\mathcal{M}}(y) d\text{Vol}_{\mathcal{M}}(x) \\
& \leq (1 + \beta_r)^{m+2}(1 + Cr_{max}) \int_{\mathcal{M}} \int_{B_m(0, r(x)) \subseteq T_x \mathcal{M}} \int_0^1 \frac{|df(\Phi_t(y, v))|^2}{r(\Phi_t(x, v))^{m+2}\rho(\Phi_t(x, v))^{2/m-1}} dt dv d\text{Vol}_{\mathcal{M}}(x) \\
& = (1 + \beta_r)^{m+2}(1 + Cr_{max}) \int_0^1 \int_{\mathcal{B}_r} \frac{|df(\Phi_t(\xi))|^2}{r(\Phi_t(\xi))^{m+2}\rho(\Phi_t(\xi))^{2/m-1}} d\text{Vol}_{T\mathcal{M}}(\xi) dt,
\end{aligned}$$

where for the first inequality we have used the definition of  $\beta_r$  and the fact that for every  $y$  with  $d_{\mathcal{M}}(x, y) \leq r_{max}$  we have

$$\left( \frac{\rho(y)}{\rho(x)} \right)^{2/m-1} = \left( 1 + \frac{\rho(y) - \rho(x)}{\rho(x)} \right)^{2/m-1} \leq 1 + Cr_{max}.$$

Finally, in the last line we use  $\text{Vol}_{T\mathcal{M}}$  to denote the volume form on  $T\mathcal{M}$ , and

$$\mathcal{B}_r := \{(x, v) \in T\mathcal{M} : x \in \mathcal{M}, v \in B_m(0, r(x)) \subseteq T_x \mathcal{M}\}.$$

Now let  $t \in (0, 1)$  and let  $(x, v) \in \mathcal{B}_r$ . Define  $\tilde{x} := \exp_x(-tv)$  and  $\tilde{v} := d\exp_x(-v)$ . It is straightforward to see that

$$\Phi_t(\tilde{x}, -\tilde{v}) = (x, v).$$

Moreover,

$$\|-\tilde{v}\|_{\tilde{x}} = \|d\exp_x(-v)\|_{\tilde{x}} = \|-v\|_x \leq r(x) = r(\tilde{x}) + (r(x) - \tilde{r}(x)) \leq \hat{r}(\tilde{x}),$$

where

$$\hat{r}(\tilde{x}) := (1 + \alpha_r)r(\tilde{x}).$$

We have shown that if  $(x, v) \in \mathcal{B}_r$ , then for every  $t \in (0, 1)$   $(x, v) \in \Phi_t(\mathcal{B}_{\hat{r}})$ . That is,

$$\mathcal{B}_r \subseteq \Phi_t(\mathcal{B}_{\hat{r}}).$$

From this we deduce that for all  $t \in (0, 1)$  we have

$$\int_{\mathcal{B}_r} \frac{|df(\Phi_t(\xi))|^2}{r(\Phi_t(\xi))^{m+2}\rho(\Phi_t(\xi))^{2/m-1}} d\text{Vol}_{T\mathcal{M}}(\xi) \leq \int_{\Phi_t(\mathcal{B}_{\hat{r}})} \frac{|df(\Phi_t(\xi))|^2}{r(\Phi_t(\xi))^{m+2}\rho(\Phi_t(\xi))^{2/m-1}} d\text{Vol}_{T\mathcal{M}}(\xi)$$

In turn, the right hand side is equal to

$$\int_{\mathcal{B}_{\hat{r}}} \frac{|df(\Phi_t(\xi))|^2}{r(\xi)^{m+2}\rho(\xi)^{2/m-1}} d\text{Vol}_{T\mathcal{M}}(\xi)$$

which follows from the well known fact that  $\Phi_t$  preserves  $\text{Vol}_{T\mathcal{M}}$  (i.e. it pushes forward  $\text{Vol}_{T\mathcal{M}}$  into itself). Integrating over  $t$  and using (4.22) we deduce that

$$E_r(f) \leq (1 + \beta_r)^{m+2}(1 + Cr_{max}) \int_{\mathcal{M}} \frac{1}{r(x)^{m+2}\rho(x)^{2/m-1}} \left( \int_{B_m(0, \hat{r}(x))} |df(x, v)|^2 dv \right) d\text{Vol}_{\mathcal{M}}(x).$$

Finally, a simple computation shows that

$$\int_{B_m(0, \hat{r}(y))} |df(x, v)|^2 dv = \int_{B_m(0, \hat{r}(y))} |\langle \nabla f(x), v \rangle|^2 dv = \sigma_{\eta} |\nabla f(x)|^2 \hat{r}(x)^{m+2},$$

and so

$$\begin{aligned}
E_r(f) & \leq (1 + \beta_r)^{m+2}(1 + Cr_{max}) \sigma_{\eta} \int_{\mathcal{M}} \left( \frac{\hat{r}(x)}{r(x)} \right)^{m+2} |\nabla f(x)|^2 \rho(x)^{1-2/m} d\text{Vol}_{\mathcal{M}}(x) \\
& \leq (1 + \beta_r)^{2m+4}(1 + Cr_{max}) \sigma_{\eta} D_{1-2/m}(f).
\end{aligned}$$

This concludes the proof. □

**Lemma 4.8.** Let  $r : \mathcal{M} \rightarrow (0, \infty)$  be an arbitrary function which is bounded away from zero and for which  $r_{max}$  and  $\beta_r$  (as defined in (4.21)) are sufficiently small. Then, for any  $f \in L^2(\mu)$  we have

$$\sigma_\eta D_{1-2/m}(\Lambda_r f) \leq (1 + C\beta_r^{2m+4} + Cr_{max}^2) E_{\hat{r}}(f),$$

where the length-scale  $\hat{r}$  in  $E_{\hat{r}}(f)$  is of the form

$$\hat{r}(x) = (1 + \beta_r)r(x).$$

*Proof.* We first notice that for a fixed function  $f \in L^2(\mu)$ , the gradient of the function  $\Lambda_r f$  can be written as

$$\nabla \Lambda_r f(x) = \tau^{-1}(x) A_1(x) + A_2(x),$$

where

$$A_1(x) := \int_{\mathcal{A}(x)} \nabla K_r(\cdot, y)(x)(f(y) - f(x)) d\text{Vol}_{\mathcal{M}}(y),$$

$$A_2(x) := \nabla(\tau^{-1})(x) \int_{\mathcal{A}(x)} K_{r(y)}(x, y)(f(y) - f(x)) d\text{Vol}_{\mathcal{M}}(y),$$

and the region  $\mathcal{A}(x)$  is defined as

$$\mathcal{A}(x) := \{y \in \mathcal{M} : |x - y| \leq r(y)\}.$$

First  $|A_1(x)| = \langle A_1(x), w \rangle$  for some unit vector  $w \in T_x \mathcal{M}$ . Therefore,

$$\begin{aligned} |A_1(x)| &= |\langle A_1(x), w \rangle| \\ &= \left| \int_{\mathcal{A}(x)} \frac{1}{\sigma_\eta r^{m+2}(y)} \eta\left(\frac{d_{\mathcal{M}}(x, y)}{r(y)}\right) (f(y) - f(x)) \langle \exp_x^{-1}(y), w \rangle d\text{Vol}_{\mathcal{M}}(y) \right| \\ &= \left| \int_{\mathcal{A}} \frac{1}{\sigma_\eta r^{m+2}(v)} \eta\left(\frac{|v|}{r(v)}\right) \varphi(v) \langle v, w \rangle J_x(v) dv \right| \\ &\leq \int_{\mathcal{A}} \frac{1}{\sigma_\eta r^{m+2}(v)} \eta\left(\frac{|v|}{r(v)}\right) |\varphi(v)| |\langle v, w \rangle| J_x(v) dv \end{aligned}$$

where  $\varphi(v) := f(\exp_x(v)) - f(x)$ ,  $\mathcal{A} := \{v \in T_x \mathcal{M} : |x - \exp_x(v)| \leq r(v)\}$  and with a slight abuse of notation we write  $r(v) = r(\exp_x(v))$  and also  $\hat{r}(v) = \hat{r}(\exp_x(v))$ . Notice that by definition of  $\hat{r}$  we have

$$\hat{r}_0 := \hat{r}(0) \geq r(v)$$

and so

$$|A_1(x)| \leq (1 + \beta_r)^{m+2} \int_{B(0, \hat{r}_0)} \frac{1}{\sigma_\eta \hat{r}_0^{m+2}} \eta\left(\frac{|v|}{\hat{r}_0}\right) |\varphi(v)| |\langle v, w \rangle| J_x(v) dv.$$

By the Cauchy-Schwartz inequality,

$$\begin{aligned} \frac{|A_1(x)|^2}{(1 + \beta_r)^{2m+4}} &\leq \left( \frac{1}{\sigma_\eta \hat{r}_0^{m+2}} \right)^2 \left( \int_{B(0, \hat{r}_0)} |\varphi(v)|^2 J_x(v)^2 \eta\left(\frac{|v|}{\hat{r}_0}\right) dv \right) \left( \int_{B(0, \hat{r}_0)} \langle v, w \rangle^2 \eta\left(\frac{|v|}{\hat{r}_0}\right) dv \right) \\ &= \frac{1}{\sigma_\eta \hat{r}_0^{m+2}} \left( \int_{B(0, \hat{r}_0)} |\varphi(v)|^2 J_x(v)^2 \eta\left(\frac{|v|}{\hat{r}_0}\right) dv \right) \end{aligned}$$

where, in the last step, we used radial symmetry to conclude that

$$\int_{B(0, \hat{r}_0)} \langle v, w \rangle^2 \eta\left(\frac{|v|}{\hat{r}_0}\right) dv = \hat{r}_0^{m+2} \int_{B(0, 1)} v_1^2 \eta(|v|) dv = \hat{r}_0^{m+2} \sigma_\eta.$$

We obtain,

$$\begin{aligned} \frac{|A_1(x)|^2}{(1 + \beta_r)^{2m+4}} &\leq \frac{1 + C(r_{max}(1 + \beta_r))^2}{\sigma_\eta \hat{r}_0^{m+2}} \int_{B(0, \hat{r}_0)} |\varphi(v)|^2 \eta\left(\frac{|v|}{\hat{r}_0}\right) J_x(v) dv \\ &= \frac{1 + C(r_{max}(1 + \beta_r))^2}{\sigma_\eta \hat{r}(x)^{m+2}} \int_{\mathcal{M}} \eta\left(\frac{d_{\mathcal{M}}(x, y)}{\hat{r}(x)}\right) (f(y) - f(x))^2 d\text{Vol}_{\mathcal{M}}(y) \end{aligned}$$

Integrating this inequality with respect to  $\rho^{1-2/m}(x) d\text{Vol}_{\mathcal{M}}(x)$  and using the Lipschitz continuity of  $\rho$ , we obtain

$$\begin{aligned} (4.23) \quad \|A_1\|_{L^2(\mathcal{M}, \rho^{1-2/m} \text{Vol}_{\mathcal{M}})}^2 &\leq (1 + \beta_r)^{2m+4} (1 + C(r_{max}(1 + \beta_r))^2) \\ &\quad \cdot \frac{1}{\sigma_\eta} \int_{\mathcal{M}} \int_{\mathcal{M}} \eta\left(\frac{d_{\mathcal{M}}(x, y)}{\hat{r}(x)}\right) \frac{|f(y) - f(x)|^2}{\hat{r}(x)^{m+2} \rho(x)^{2/m-1}} d\text{Vol}_{\mathcal{M}}(y) d\text{Vol}_{\mathcal{M}}(x) \\ &= (1 + \beta_r)^{2m+4} (1 + C(r_{max}(1 + \beta_r))^2) \frac{1}{\sigma_\eta} E_{\hat{r}}(f). \end{aligned}$$

Regarding  $A_2$ , first note that by Lemma 5.1 in [7] we have

$$(4.24) \quad (1 + Cr_{max}^2)^{-1} \leq \tau(x) \leq (1 + Cr_{max}^2)$$

and also

$$|\nabla(\tau^{-1})| \leq Cr_{max}$$

Therefore,

$$\begin{aligned} |A_2(x)|^2 &\leq |\nabla(\tau^{-1})(x)|^2 \tau(x) \int_{\mathcal{M}} |f(y) - f(x)|^2 K_{r(y)}(x, y) d\text{Vol}_{\mathcal{M}}(y) \\ &\leq r_{max}^2 \int_{\mathcal{M}} \frac{1}{r(y)^m} \eta\left(\frac{d_{\mathcal{M}}(x, y)}{r(y)}\right) |f(y) - f(x)|^2 d\text{Vol}_{\mathcal{M}}(y), \end{aligned}$$

where the first inequality follows from Cauchy-Schwartz inequality and in the second one we used

$$\psi(s) \leq \frac{1}{\sigma_\eta} \eta(s), \quad \forall s > 0.$$

Integrating the estimate for  $|A_2(x)|^2$  with respect to  $\rho^{1-2/m}(x) d\text{Vol}_{\mathcal{M}}(x)$  while using the Lipschitz continuity of  $\rho$  we obtain

$$\|A_2\|_{L^2(\mathcal{M}, \rho^{1-2/m} \text{Vol}_{\mathcal{M}})}^2 \leq C(1 + r_{max})^2 r_{max}^4 E_r(f) \leq C(1 + \beta_r)^{m+2} (1 + r_{max})^2 r_{max}^4 E_{\hat{r}}(f).$$

By combining the above with (4.23) and (4.24) estimates and the lower bound for  $\tau$  we obtain the desired inequality.  $\square$

We will apply Lemma 4.8 with  $r$  for which  $r_{max}, \beta_r \ll 1$ . In particular  $E_{\hat{r}}(f)$  (where  $\hat{r}$  is to the first order equal to  $r$ ) is greater than a quantity that up to leading order is equal to  $\sigma_\eta D_{1-2/m}(f)$ . This is the reverse inequality to the one in Lemma 4.7.

*Proof of Proposition 4.4.* To prove the first inequality, let us start by recalling that  $r_k(x_i, x_j) = \max\{\varepsilon_k(x_i), \varepsilon_k(x_j)\}$  where  $\varepsilon_k$  is defined as in (2.7). Recall also that for arbitrary  $u \in L^2(\mu_n)$  we can write

$$b_k(u) := \frac{1}{n^2} \left(\frac{n\alpha_m}{k}\right)^{1+2/m} \sum_{i,j} \eta\left(\frac{|x_i - x_j|}{r_k(x_i, x_j)}\right) (u(x_i) - u(x_j))^2.$$

Using the  $\infty$ -OT map  $\tilde{T}$  between  $\tilde{\mu}_n$  and  $\mu_n$  we can apply the change of variables formula to rewrite  $b_k(u)$  as

$$\begin{aligned} b_k(u) &= \left(\frac{n\alpha_m}{k}\right)^{1+2/m} \int_{\mathcal{M}} \int_{\mathcal{M}} \eta\left(\frac{|\tilde{T}(x) - \tilde{T}(y)|}{\max\{\varepsilon_k(\tilde{T}(x)), \varepsilon_k(\tilde{T}(y))\}}\right) |\tilde{P}^*u(x) - \tilde{P}^*u(y)|^2 d\tilde{\mu}_n(y) d\tilde{\mu}_n(x) \\ &= \int_{\mathcal{M}} \int_{\mathcal{M}} \frac{1}{\rho(\tilde{T}(y))^{1+2/m} \varepsilon(\tilde{T}(y))^{m+2}} \eta\left(\frac{|\tilde{T}(x) - \tilde{T}(y)|}{\max\{\varepsilon_k(\tilde{T}(x)), \varepsilon_k(\tilde{T}(y))\}}\right) |\tilde{P}^*u(x) - \tilde{P}^*u(y)|^2 d\tilde{\mu}_n(y) d\tilde{\mu}_n(x). \end{aligned}$$

where for the last line we have used the definition of  $\varepsilon$  in (2.14). From Lemma 4.6 we know that

$$\varepsilon_k(\tilde{T}(x)) \leq \varepsilon(\tilde{T}(x)) + C\varepsilon_{max}^2 + C\varepsilon_{max}\tilde{\delta} + C\varepsilon_{max}\theta.$$

Moreover, using the smoothness of  $\rho$  and the definition of  $\varepsilon$  for  $x, y$  with  $|x - y| \leq \varepsilon_{max}$  we have

$$\varepsilon(\tilde{T}x) \leq \varepsilon(\tilde{T}y) + C(\varepsilon_{max}^2 + \tilde{\delta}\varepsilon_{max}).$$

Thus, if  $|\tilde{T}(x) - \tilde{T}(y)| \leq r_k(\tilde{T}(x), \tilde{T}(y))$ , we have  $|\tilde{T}(x) - \tilde{T}(y)| \leq \varepsilon(\tilde{T}y) + C(\varepsilon_{max}^2 + \tilde{\delta}\varepsilon_{max})$  and also

$$|x - y| \leq d_{\mathcal{M}}(x, y) \leq \varepsilon(\tilde{T}y) + C(\varepsilon_{max}^2 + \tilde{\delta}\varepsilon_{max} + C\theta\varepsilon_{max}) + 2\tilde{\delta} =: r(y).$$

It follows that

$$\begin{aligned} b_k(u) &\leq \int_{\mathcal{M}} \int_{\mathcal{M}} \frac{1}{\rho(\tilde{T}(y))^{1+2/m} \varepsilon(\tilde{T}(y))^{m+2}} \eta\left(\frac{|\tilde{T}(x) - \tilde{T}(y)|}{r(y)}\right) |\tilde{P}^*u(x) - \tilde{P}^*u(y)|^2 d\tilde{\mu}_n(y) d\tilde{\mu}_n(x) \\ &\leq (1 + C\tilde{\delta})(1 + C(\varepsilon_{max}^2 + \tilde{\delta} + C\theta\varepsilon_{max})) \\ &\quad \cdot \int_{\mathcal{M}} \int_{\mathcal{M}} \frac{1}{\rho(y)^{2/m-1} r(y)^{m+2}} \eta\left(\frac{|\tilde{T}(x) - \tilde{T}(y)|}{r(y)}\right) |\tilde{P}^*u(x) - \tilde{P}^*u(y)|^2 d\text{Vol}_{\mathcal{M}}(y) d\text{Vol}_{\mathcal{M}}(x) \\ &= (1 + C\tilde{\delta})(1 + C(\varepsilon_{max} + \frac{\tilde{\delta}}{\varepsilon_{min}} + \theta)) E_r(\tilde{P}^*u). \end{aligned}$$

Let  $f \in L^2(\mu)$  and let  $u := \tilde{P}f$  in the above estimate to obtain

$$b_k(\tilde{P}f) \leq (1 + C\tilde{\delta})(1 + C(\varepsilon_{max} + \frac{\tilde{\delta}}{\varepsilon_{min}} + \theta)) E_r(\tilde{P}^*\tilde{P}f) \leq (1 + C\tilde{\delta})(1 + C(\varepsilon_{max} + \frac{\tilde{\delta}}{\varepsilon_{min}} + \theta)) E_r(f),$$

where in the last line we have used Jensen's inequality to relate  $E_r(\tilde{P}^*\tilde{P}f)$  and  $E_r(f)$ . On the other hand, from Lemma 4.7 we have

$$E_r(f) \leq (1 + Cr_{max})(1 + \beta_r)^{2m+4} \sigma_{\eta} D_{1-2/m}(f).$$

However for the function  $r$  defined above the quantity  $\beta_r$  satisfies

$$\beta_r \leq C\varepsilon_{max},$$

so

$$b_k(\tilde{P}f) \leq (1 + C(\varepsilon_{max} + \frac{\tilde{\delta}}{\varepsilon_{min}} + \theta)) \sigma_{\eta} D_{1-2/m}(f),$$

proving in this way the desired inequality thanks to Remark 4.3.

The second inequality is proved very similarly using the lower bound in Lemma 4.8, which allows us to show that

$$(4.25) \quad E_r(\tilde{P}^*u) \leq (1 + C(\varepsilon_{max} + \frac{\tilde{\delta}}{\varepsilon_{min}} + \theta)) b_k(u), \quad \forall u \in L^2(\mu_n),$$

for  $r$  of the form  $r(y) = \varepsilon(y) - C\varepsilon_{max}^2 - C\tilde{\delta} - C\theta\varepsilon_{max}$ . Lemma 4.8 in turn allows us to bound  $\sigma_{\eta} D_{1-2/m}(\tilde{I}u)$  with  $E_r(\tilde{P}^*u)$  (up to leading order).  $\square$

*Proof of Proposition 4.5.* The first part follows directly from 1) in Proposition 4.1 since

$$D_2(f) \leq CD_{1-2/m}(f),$$

given that we have assumed that  $\rho$  is bounded and also bounded away from zero.

For the second part, notice that for every  $f \in L^2(\mu)$  by Jensen's inequality we have

$$\begin{aligned} (\Lambda_r f(x) - f(x))^2 &= \left( \frac{1}{\tau(x)} \int_{\mathcal{M}} \frac{1}{r(y)^m} \psi \left( \frac{|x-y|}{r(y)} \right) (f(y) - f(x)) \rho(y) d\text{Vol}_{\mathcal{M}}(y) \right)^2 \\ &\leq \frac{1}{\tau(x)} \int_{\mathcal{M}} \frac{1}{r(y)^m} \psi \left( \frac{|x-y|}{r(y)} \right) (f(y) - f(x))^2 \rho(y) d\text{Vol}_{\mathcal{M}}(y) \\ &\leq C \int_{\mathcal{M}} \frac{1}{r(y)^m} \psi \left( \frac{|x-y|}{r(y)} \right) (f(y) - f(x))^2 \rho(y) d\text{Vol}_{\mathcal{M}}(y) \\ &\leq C \int_{\mathcal{M}} \frac{1}{r(y)^m} \eta \left( \frac{|x-y|}{r(y)} \right) (f(y) - f(x))^2 \rho(y) d\text{Vol}_{\mathcal{M}}(y) \end{aligned}$$

where the last inequality follows from (4.18). Multiplying by  $\rho(x)$  and integrating over all  $x \in \mathcal{M}$  we get

$$(4.26) \quad \|\Lambda_r f - f\|_{L^2(\mu)}^2 \leq Cr_{\max}^2 E_r(f),$$

where we have used again the fact that  $\rho$  is bounded and bounded away from zero. Taking  $r$  to be

$$r(y) := \varepsilon(y) - 2\tilde{\delta}$$

we continue as in the second part of the proof of (4.2) to obtain

$$\begin{aligned} \left| \|\tilde{\mathcal{I}}u\|_{L^2(\tilde{\mu}_n)} - \|u\|_{L^2(\mu_n)} \right| &\leq \|\Lambda_{\varepsilon-2\tilde{\delta}} \tilde{P}^* u - \tilde{P}^* u\|_{L^2(\tilde{\mu}_n)} \\ &\leq \left( 1 + \frac{\|\rho - \tilde{\rho}_n\|_{\infty}}{\rho_{\min}} \right) \cdot \|\Lambda_{\varepsilon-2\tilde{\delta}} \tilde{P}^* u - \tilde{P}^* u\|_{L^2(\mu)} \\ &\leq \left( 1 + \frac{\|\rho - \tilde{\rho}_n\|_{\infty}}{\rho_{\min}} \right) \cdot C\varepsilon_{\max} \sqrt{E_{\varepsilon-2\tilde{\delta}}(\tilde{P}^* u)} \\ &\leq C\varepsilon_{\max} \sqrt{b_k(u)}, \end{aligned}$$

where for the third inequality we have used (4.26), and for the last one we have used (4.25). Also, from (4.26)

$$\|\tilde{\mathcal{I}}u\|_{L^2(\tilde{\mu}_n)} = \|\Lambda_r \tilde{P}^* u\|_{L^2(\tilde{\mu}_n)} \leq C\|\tilde{P}^* u\|_{L^2(\tilde{\mu}_n)} = C\|u\|_{L^2(\mu_n)},$$

and hence

$$\left| \|\tilde{\mathcal{I}}u\|_{L^2(\tilde{\mu}_n)}^2 - \|u\|_{L^2(\mu_n)}^2 \right| \leq C\varepsilon_{\max} \|u\|_{L^2(\mu_n)} \sqrt{b_k(u)}.$$

Finally, we use (4.11) to compare  $\|\tilde{\mathcal{I}}u\|_{L^2(\tilde{\mu}_n)}^2$  and  $\|\tilde{\mathcal{I}}u\|_{L^2(\mu)}^2$ . □

With Proposition 4.4 and 4.5 in hand, the proofs of Theorems 2.5 and 2.9 are exactly as in the  $\varepsilon$ -setting and so we skip the details.

## 5. CONVERGENCE OF EIGENVECTORS IN $TL^2$ AND CONVERGENCE OF GRAPH LAPLACIAN EMBEDDINGS

To establish Theorem 2.10 it will be convenient to recall the definition of the  $TL^2$  space presented in [27]. Here we consider  $\mathbb{R}^L$ -valued functions.

We define the set

$$TL^2(\mathcal{M}; \mathbb{R}^L) := \{(\gamma, H) : \gamma \in \mathcal{P}(\mathcal{M}), H \in L^2(\gamma; \mathbb{R}^L)\},$$

and the metric

$$\left( d_{TL^2}((\gamma, H), (\tilde{\gamma}, \tilde{H})) \right)^2 := \min_{\pi \in \Gamma(\gamma, \tilde{\gamma})} \int_{\mathcal{M} \times \mathcal{M}} d_{\mathcal{M}}(x, y)^2 d\pi(x, y) + \int_{\mathbb{R}^L \times \mathbb{R}^L} |H(x) - \tilde{H}(y)|^2 d\pi(x, y).$$

In the above  $\mathcal{P}(\mathcal{M})$  denotes the space of Borel probability measures on  $\mathcal{M}$ , and for  $\gamma \in \mathcal{P}(\mathcal{M})$ ,  $L^2(\gamma; \mathbb{R}^L)$  denotes the  $\mathbb{R}^L$ -valued  $L^2$  functions with respect to  $\gamma$ .  $\Gamma(\gamma, \tilde{\gamma})$  denotes the set of couplings or transport plans between  $\gamma$  and  $\tilde{\gamma}$ . Notice that if we remove the second term on the right hand side of the definition of  $d_{TL^2}$  we get,

$$W_{\mathcal{M}}^2(\gamma, \tilde{\gamma}) := \min_{\pi \in \Gamma(\gamma, \tilde{\gamma})} \int_{\mathcal{M} \times \mathcal{M}} d_{\mathcal{M}}(x, y)^2 d\pi(x, y)$$

which is the square of the 2-Wasserstein distance between  $\gamma$  and  $\tilde{\gamma}$  (using the intrinsic geometry of  $\mathcal{M}$ ).

For our purposes, the key property of the  $TL^2$ -metric is that it gives an upper bound for the Wasserstein distance between induced embeddings.

**Lemma 5.1.** *Let  $(\gamma, H), (\tilde{\gamma}, \tilde{H}) \in TL^2(\mathcal{M}; \mathbb{R}^L)$ . Then for all  $\pi \in \Gamma(\gamma, \tilde{\gamma})$  we have*

$$W_2(H_{\sharp}\gamma, \tilde{H}_{\sharp}\tilde{\gamma}) \leq \int_{\mathbb{R}^L \times \mathbb{R}^L} |H(x) - \tilde{H}(y)|^2 d\pi(x, y),$$

where  $W_2$  is the 2-Wasserstein distance (w.r.t. to the Euclidean metric) between probability measures on  $\mathbb{R}^L$ , and  $W_2^2$  its square. In particular, it follows that

$$W_2(H_{\sharp}\mu, \tilde{H}_{\sharp}\tilde{\gamma}) \leq d_{TL^2}((\gamma, H), (\tilde{\gamma}, \tilde{H})).$$

*Proof.* To see this let  $\pi \in \Gamma(\gamma, \tilde{\gamma})$  and define

$$\hat{\pi} := (H \times \tilde{H})_{\sharp}\pi,$$

where  $H \times \tilde{H}$  is the map

$$H \times \tilde{H} : (x, y) \in \mathcal{M} \times \mathcal{M} \mapsto (\hat{x}, \hat{y}) = (H(x), \tilde{H}(y)) \in \mathbb{R}^L \times \mathbb{R}^L.$$

It is straightforward to see that  $\hat{\pi} \in \Gamma(H_{\sharp}\gamma, \tilde{H}_{\sharp}\tilde{\gamma})$ . From the change of variables formula we see that

$$W_2^2(H_{\sharp}\gamma, \tilde{H}_{\sharp}\tilde{\gamma}) \leq \int_{\mathbb{R}^L \times \mathbb{R}^L} |\hat{x} - \hat{y}|^2 d\hat{\pi}(\hat{x}, \hat{y}) = \int_{\mathcal{M} \times \mathcal{M}} |H(x) - \tilde{H}(y)|^2 d\pi(x, y),$$

which implies the desired result. □

*Proof of Theorem 2.10.* Let  $\pi^*$  be an optimal transport plan between  $\mu$  and  $\mu_n$ , that is,  $\pi^*$  is such that

$$W_2^2(F_{\sharp}\mu, F_{n\sharp}\mu_n) = \int_{\mathcal{M} \times \mathcal{M}} d_{\mathcal{M}}(x, y)^2 d\pi^*(x, y).$$

From Lemma 5.1 we know that

$$W_2^2(F_{\sharp}\mu, \tilde{F}_{n\sharp}\mu_n) \leq \int_{\mathcal{M} \times \mathcal{M}} |F(x) - F_n(y)|^2 d\pi^*(x, y) = \sum_{l=1}^L \int_{\mathcal{M} \times \mathcal{M}} |u_l(x) - f_l(y)|^2 d\pi^*(x, y).$$

Using the regularity of the eigenfunctions  $f_l$  we can bound the above with

$$\begin{aligned} & 2 \sum_{l=1}^L \int_{\mathcal{M} \times \mathcal{M}} |u_l(x) - f_l(y)|^2 d\pi^*(x, y) + 2C \sum_{l=1}^L \int_{\mathcal{M} \times \mathcal{M}} d_{\mathcal{M}}(x, y)^2 d\pi^*(x, y) \\ & \leq 2 \sum_{l=1}^L \|u_l - f_l\|_{L^2(\mu_n)}^2 + 2CLW_2^2(\mu, \mu_n). \end{aligned}$$

Combining with the results from Theorem 2.7 we get the desired inequality. □

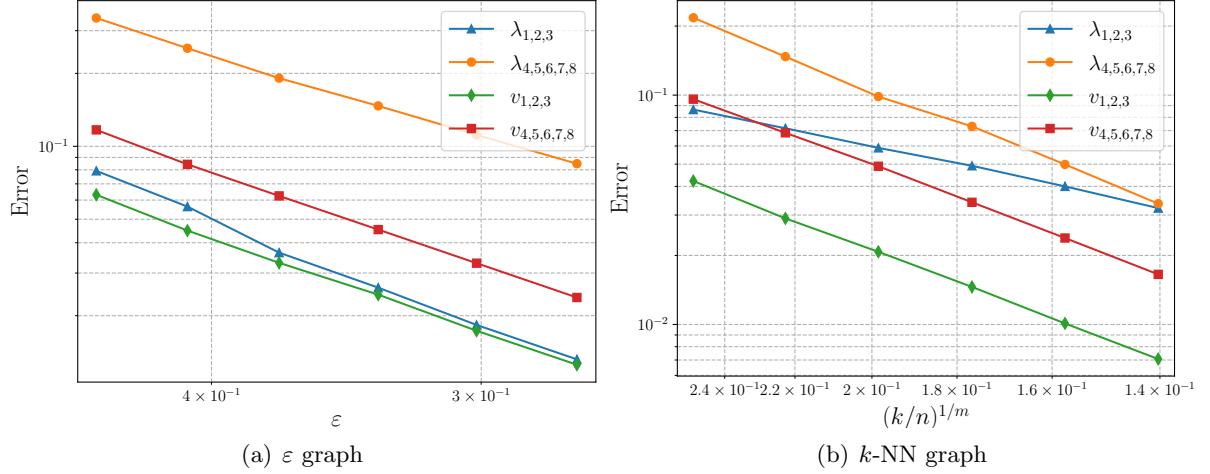


FIGURE 1. Observed rates of convergence from numerical experiments on the 2-dimensional sphere. We tested the first 8 non-trivial eigenvalue/eigenvector pairs, and averaged the errors over the two corresponding eigenspaces. The convergence rates for the eigenvalues and eigenvectors (in  $L^2$ ) appear to be between second and third order in  $\varepsilon$  and  $(k/n)^{1/m}$  in (a) and (b), respectively.

## 6. NUMERICAL EXPERIMENTS

To test the convergence rates in our main results, we ran some numerical experiments on the two dimensional sphere.<sup>1</sup> The density is thus  $\rho = 1/((m+1)\alpha_{m+1})$ . The experiments used  $n = 2^{12}$  up to  $n = 2^{17} = 131072$  points independently and uniformly distributed on the sphere, with the errors averaged over 100 trials. For the  $k$ -NN graph Laplacian, see (2.9), we set

$$k = n^{\frac{4}{m+4}},$$

as required in (1.2) for our main convergence rates on  $k$ -NN graphs. For the  $\varepsilon$ -graphs we chose  $\varepsilon$  to be the minimum distance to the  $k^{\text{th}}$  nearest neighbor. This value for  $\varepsilon$  has the same scaling as the lower bound for our convergence rates for  $\varepsilon$ -ball graph Laplacians given in (1.1). We used a characteristic function kernel  $\eta = \mathbb{1}_{[0,1]}$ , which is to say we used an unweighted graph. In this case  $\sigma_\eta = \alpha_m/(m+2)$ .

We tested the convergence of the first 8 non-trivial eigenfunctions of the Laplace-Beltrami operator, which in this case are the spherical harmonics, given unnormalized by  $v_1(x) = x_1$ ,  $v_2(x) = x_2$ ,  $v_3(x) = x_3$ ,  $v_4(x) = x_1^2 - x_2^2$ ,  $v_5(x) = x_1x_2$ ,  $v_6(x) = x_1x_3$ ,  $v_7(x) = x_2x_3$ , and  $v_8(x) = 3x_3^2 - 1$ , with corresponding eigenvalues  $\lambda_1 = \lambda_2 = \lambda_3 = 2$  and  $\lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 6$ . We compared the absolute difference between the graph and continuum eigenvalues, and the  $L^2$  norm of the difference between the spherical harmonics and graph Laplacian eigenvectors over the graph. The eigenvector error was computed by restricting the spherical harmonics to the graph, using a QR decomposition to find orthonormal bases for the subspaces spanned by  $\{v_1, v_2, v_3\}$  and  $\{v_4, v_5, v_6, v_7, v_8\}$ , and then projecting the graph Laplacian eigenvectors to the corresponding subspaces and computing the  $L^2$  norm of the residual. We then averaged the errors over each eigenspace to simplify the presentation

Figure 1 shows the experimental convergence rates. A loglog regression reports experimental convergence rates of  $O(\varepsilon^{3.1})$ ,  $O(\varepsilon^{2.4})$ ,  $O(\varepsilon^{2.8})$ , and  $O(\varepsilon^{2.8})$  for the  $\varepsilon$  graph in Figure 1(a), and  $O((k/n)^{1.7/m})$ ,  $O((k/n)^{3.2/m})$ ,  $O((k/n)^{3.1/m})$ , and  $O((k/n)^{3.1/m})$  for the  $k$ -NN graph in Figure 1(b). These convergence rates are better than the linear convergence rates given in

<sup>1</sup>The code for the experiments is available online at <https://github.com/jwcalder/kNNSpectralRates>.

our main results. We expect the uniform density on the sphere is a highly special case, and convergence rates may be significantly worse in general.

## 7. CONCLUSION

In this paper we have obtained new results on the spectral convergence of graph Laplacian operators built from random data towards weighted Laplace-Beltrami operators on smooth compact manifolds without boundary. Our results contribute to the growing manifold learning literature in several regards. First, we improve existing spectral convergence rates for Laplacians based on  $\varepsilon$ -graphs in the regime  $\left(\frac{\log(n)}{n}\right)^{1/(m+4)} \ll \varepsilon \ll 1$ , showing that the spectral convergence rate scales like  $\varepsilon$  with very high probability. Second, we analyze Laplacians based on  $k$ -NN graphs, a setting where neither pointwise convergence rates nor spectral rates existed in the literature.

Some directions that we believe are worth studying in the future and that stem from this work include: 1) the improvement of the convergence rates in the regime  $\left(\frac{\log(n)}{n}\right)^{1/m} \ll \varepsilon \leq \left(\frac{\log(n)}{n}\right)^{1/(m+4)}$  where our proof techniques fail to work; 2) the analysis of graph Laplacians under less restrictive assumptions on the probability distribution underlying the data set. In this direction it is of interest to extend the analysis to more general geometric settings (not just compact, smooth manifolds without boundary) and to more general densities (for example, relaxing the lower bound assumption).

## APPENDIX A. PROOF OF PROPOSITION 2.12

*Proof.* First we notice that for small enough but macroscopic (i.e. fixed)  $r > 0$  we can find a partition of  $\mathcal{M}$  (up to overlaps of  $\text{Vol}_{\mathcal{M}}$  measure zero) into closed sets  $V_1, \dots, V_L$  ( $L$  depends on  $r$ ) for which:

- For every  $l = 1, \dots, L$ , there is a bi-Lipschitz homeomorphism  $\Phi_l : V_l \rightarrow \overline{B(0, r/2)} \subseteq \mathbb{R}^m$  with bi-Lipschitz constant less than 18.

Such partition can be constructed using a covering of  $\mathcal{M}$  with balls of radius  $r$ . The centers of the balls can then be used to construct a Voronoi tessellation of  $\mathcal{M}$ , inducing in this way the sets  $V_i$ . This construction is presented in detail in Propositions 2 and 3 in [22]. Since the regions  $V_i$  are bi-Lipschitz homeomorphic to  $B(0, r/2)$ , we can also construct a bi-Lipschitz homeomorphism

$$\Phi_l : V_l \rightarrow [0, r]^m$$

with bi-Lipschitz constant less than  $C$  (which only depends on  $m$ ).

Let  $\tilde{\delta} < \frac{r}{4}$  and let us consider a partition of  $[0, 1]^m$  into a collection of ( $m$ -dimensional) rectangles  $\tilde{Q}$  with diameter less than  $\tilde{\delta}$  and with aspect ratio no larger than 2. We construct subsets  $Q$  of  $\mathcal{M}$  by letting

$$Q := \Phi_l^{-1}(\tilde{Q})$$

for some  $l = 1, \dots, L$  and some  $\tilde{Q}$ . From the properties of the maps  $\Phi_l$  it follows that the collection of cells  $Q$  is a partition of  $\mathcal{M}$  (the different cells are disjoint up to sets of  $\text{Vol}_{\mathcal{M}}$  measure zero) and satisfy:

- (1)  $\text{diam}(Q) \leq C\tilde{\delta}$ .
- (2)  $\tilde{\delta}^m \leq C \text{Vol}_{\mathcal{M}}(Q) \leq C\mu(Q)$ .

We can then let  $\tilde{\rho}_n : \mathcal{M} \rightarrow$  be the piecewise constant density defined by

$$\tilde{\rho}_n(x) = \frac{\mu_n(Q)}{\text{Vol}_{\mathcal{M}}(Q)}, \quad x \in Q.$$

It is clear that  $\tilde{\rho}_n$  is indeed a density and moreover, for every  $x \in Q$  we have

$$|\rho(x) - \tilde{\rho}_n(x)| \leq \left| \frac{\mu_n(Q) - \mu(Q)}{\text{Vol}_M(Q)} \right| + \frac{1}{\text{Vol}_M(Q)} \left| \int_Q (\rho(y) - \rho(x)) d\text{Vol}_M(y) \right|$$

In particular,

$$\|\rho - \tilde{\rho}_n\|_{L^\infty(\mu)} \leq \frac{C}{\tilde{\delta}^m} \sup_Q |\mu_n(Q) - \mu(Q)| + C \text{Lip}(\rho) \tilde{\delta}.$$

From Chernoff's bound one can get for all  $\theta$  small enough,

$$\begin{aligned} \mathbb{P}(\mu_n(Q) \geq (1 + \theta)\mu(Q)) &\leq \exp\left(-\frac{n\theta^2\mu(Q)}{3}\right) \\ &\leq \exp\left(-\frac{Cn\theta^2\tilde{\delta}^m}{3}\right) \end{aligned}$$

where in the last line we have used the fact that

$$\rho_{\min} \text{Vol}_M(Q) \leq \mu(Q)$$

and that

$$C\tilde{\delta}^m \leq \text{Vol}_M(Q).$$

Likewise we can obtain

$$\mathbb{P}(\mu_n(Q) \leq (1 - \theta)\mu(Q)) \leq \exp\left(-\frac{Cn\theta^2\tilde{\delta}^m}{3}\right),$$

and so

$$\mathbb{P}(|\mu_n(Q) - \mu(Q)| \geq \theta\mu(Q)) \leq 2 \exp\left(-\frac{Cn\theta^2\tilde{\delta}^m}{3}\right).$$

Taking a union bound, we deduce that

$$\mathbb{P}(|\mu(Q) - \mu_n(Q)| \geq \theta \text{Vol}_M(Q) \quad \forall Q) \leq \frac{L}{\theta^m} \exp\left(-Cn\theta^2\tilde{\delta}^m\right).$$

and as long as  $\theta \geq \frac{1}{n^{1/m}}$  we can replace  $L/\theta^m$  above with  $n$ .

□

## REFERENCES

- [1] R. K. Ando and T. Zhang. Learning on graph with Laplacian regularization. In *Proceedings of the 19th International Conference on Neural Information Processing Systems*, NIPS'06, pages 25–32, Cambridge, MA, USA, 2006. MIT Press.
- [2] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in neural information processing systems*, pages 585–591, 2002.
- [3] M. Belkin and P. Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. In *International Conference on Computational Learning Theory*, pages 486–500. Springer, 2005.
- [4] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of machine learning research*, 7(Nov):2399–2434, 2006.
- [5] A. Bertozzi, X. Luo, A. Stuart, and K. Zygalakis. Uncertainty quantification in graph-based classification of high dimensional data. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):568–595, 2018.
- [6] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.
- [7] D. Burago, S. Ivanov, and Y. Kurylev. A graph discretization of the Laplace-Beltrami operator. *J. Spectr. Theory*, 4(4):675–714, 2014.
- [8] J. Calder. The game theoretic p-Laplacian and semi-supervised learning with few labels. *Nonlinearity*, 32(1), 2018.
- [9] J. Calder. Consistency of Lipschitz learning with infinite unlabeled data and finite labeled data. *SIAM Journal on Mathematics of Data Science*, 1(4):780–812, 2019.
- [10] J. Calder, B. Cook, M. Thorpe, and D. Slepčev. Poisson Learning: Graph Based semi-supervised learning at very low label rates. *To appear in International Conference on Machine Learning (ICML)*, 2020.
- [11] J. Calder and D. Slepčev. Properly-weighted graph Laplacian for semi-supervised learning. *Applied Mathematics and Optimization: Special Issue on Optimization in Data Science*, 2019.

- [12] J. Calder, D. Slepčev, and M. Thorpe. Rates of convergence for Laplacian semi-supervised learning with low labelling rates. *arXiv:2006.02765*, 2020.
- [13] M. Caroccia, A. Chambolle, and D. Slepčev. Mumford-shah functionals on graphs and their asymptotics. [arxiv.org/abs/1906.09521](https://arxiv.org/abs/1906.09521), 2019.
- [14] O. Chapelle, B. Scholkopf, and A. Zien. *Semi-supervised learning*. MIT, 2006.
- [15] X. Cheng and H.-T. Wu. Convergence of graph laplacian with knn self-tuned kernels. 2020.
- [16] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences*, 102(21):7426–7431, 2005.
- [17] M. P. do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.
- [18] M. M. Dunlop, D. Slepčev, A. M. Stuart, and M. Thorpe. Large data and zero noise limits of graph-based semi-supervised learning algorithms. *Applied and Computational Harmonic Analysis*, 2019.
- [19] M. Flores, J. Calder, and G. Lerman. Algorithms for  $L_p$ -based semi-supervised learning on graphs. *arXiv:1901.05031*, 2019.
- [20] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, 162(3-4):707–738, 2015.
- [21] N. García Trillo. Variational limits of k-nn graph-based functionals on data clouds. *SIMODS*, 1(1):93–120, 2019.
- [22] N. García Trillo, M. Gerlach, M. Hein, and D. Slepčev. Spectral convergence of the graph Laplacian on random geometric graphs towards the Laplace Beltrami operator. *To appear in FOCM*, 2019.
- [23] N. García Trillo, F. Hoffmann, and B. Hosseini. Geometric structure of graph Laplacian embeddings. *Preprint*, 2019.
- [24] N. García Trillo, Z. Kaplan, T. Samakhoana, and D. Sanz-Alonso. On the consistency of graph-based bayesian learning and the scalability of sampling algorithms. [arxiv.org/abs/1710.07702](https://arxiv.org/abs/1710.07702), 2017.
- [25] N. García Trillo and R. Murray. A maximum principle argument for the uniform convergence of graph Laplacian regressors. [arxiv.org/abs/1901.10089](https://arxiv.org/abs/1901.10089), 2019.
- [26] N. García Trillo, D. Sanz-Alonso, and R. Yang. Local regularization of noisy point clouds: Improved global geometric estimates and data analysis. *Journal of Machine Learning Research*, 20(136):1–37, 2019.
- [27] N. García Trillo and D. Slepčev. A variational approach to the consistency of spectral clustering. *Applied and Computational Harmonic Analysis*, 45(2):239–281, 2018.
- [28] N. García Trillo, D. Slepčev, J. Von Brecht, T. Laurent, and X. Bresson. Consistency of cheeger and ratio graph cuts. *The Journal of Machine Learning Research*, 17(1):6268–6313, 2016.
- [29] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Springer, 2015.
- [30] E. Giné and V. Koltchinskii. Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results. In *High dimensional probability*, volume 51 of *IMS Lecture Notes Monogr. Ser.*, pages 238–259. Inst. Math. Statist., Beachwood, OH, 2006.
- [31] E. Hebey. *Sobolev spaces on Riemannian manifolds*. Berlin ; New York : Springer, [1996] ©1996, [1996]. Includes bibliographical references (pages [106]-113) and indexes.
- [32] M. Hein, J.-Y. Audibert, and U. v. Luxburg. Graph Laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8(Jun):1325–1368, 2007.
- [33] M. Hein, J.-Y. Audibert, and U. Von Luxburg. From graphs to manifolds—weak and strong pointwise consistency of graph Laplacians. In *International Conference on Computational Learning Theory*, pages 470–485. Springer, 2005.
- [34] F. Hoffmann, B. Hosseini, A. A. Oberai, and A. M. Stuart. Spectral analysis of weighted Laplacians arising in data clustering. [arxiv.org/abs/1909.06389](https://arxiv.org/abs/1909.06389), 2019.
- [35] F. Hoffmann, B. Hosseini, Z. Ren, and A. M. Stuart. Consistency of semi-supervised learning algorithms on graphs: Probit and one-hot methods. [arxiv.org/abs/1906.07658](https://arxiv.org/abs/1906.07658), 2019.
- [36] A. Kirichenko and H. van Zanten. Estimating a smooth function on a large graph by Bayesian Laplacian regularisation. *Electron. J. Statist.*, 11(1):891–915, 2017.
- [37] J. Lei. Convergence and concentration of empirical measures under wasserstein distance in unbounded functional spaces. [arxiv.org/pdf/1804.10556.pdf](https://arxiv.org/pdf/1804.10556.pdf), 2018.
- [38] R. J. LeVeque. *Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems*. SIAM, 2007.
- [39] M. Maier, M. Hein, and U. von Luxburg. Optimal construction of k-nearest-neighbor graphs for identifying noisy clusters. *Theoretical Computer Science*, 410(19):1749–1764, 2009.
- [40] T. Müller and M. D. Penrose. Optimal cheeger cuts and bisections of random geometric graphs. To appear in *Annals of Applied Probability*, 2019.
- [41] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002.
- [42] M. Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.

- [43] L. Rosasco, S. Villa, S. Mosci, M. Santoro, and A. Verri. Nonparametric sparsity and regularization. *The Journal of Machine Learning Research*, 14(1):1665–1714, 2013.
- [44] Z. Shi. Convergence of Laplacian spectra from random samples. arXiv preprint arXiv:1507.00151, 2015.
- [45] A. Singer. From graph to manifold Laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21(1):128–134, 2006.
- [46] A. Singer and H.-T. Wu. Spectral convergence of the connection Laplacian from random samples. *Information and Inference: A Journal of the IMA*, 6(1):58–123, 2017.
- [47] D. Slepcev and M. Thorpe. Analysis of p-Laplacian regularization in semisupervised learning. *SIAM Journal on Mathematical Analysis*, 51(3):2085–2120, 2019.
- [48] A. J. Smola and R. Kondor. Kernels and regularization on graphs. In *Learning theory and kernel machines*, pages 144–158. Springer, 2003.
- [49] R. J. Tibshirani. Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, 42(1):285–323, 2014.
- [50] D. Ting, L. Huang, and M. Jordan. An analysis of the convergence of graph Laplacians. *International Conference on Machine Learning (ICML)*, 2010.
- [51] U. von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.
- [52] U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *Ann. Statist.*, 36(2):555–586, 2008.
- [53] Y.-X. Wang, J. Sharpnack, A. J. Smola, and R. J. Tibshirani. Trend filtering on graphs. *The Journal of Machine Learning Research*, 17(1):3651–3691, 2016.
- [54] A. Yuan, J. Calder, and B. Oeding. A continuum limit for the PageRank algorithm. *arXiv:2001.08973*, 2020.
- [55] L. Zelnik-Manor and P. Perona. Self-tuning spectral clustering. In *Advances in neural information processing systems*, pages 1601–1608, 2005.
- [56] X. Zhu, Z. Ghahramani, and J. D. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *Proceedings of the 20th International conference on Machine learning (ICML-03)*, pages 912–919, 2003.
- [57] X. Zhu, J. D. Lafferty, and Z. Ghahramani. Semi-supervised learning: From Gaussian fields to Gaussian processes. *Technical Report: Carnegie Mellon University*, 2003.