Journal of Machine Learning Research 24 (2023) 1-71 Submitted 10/21; Revised 5/23; Published 5/23

Large sample spectral analysis of graph-based multi-manifold clustering

Nicolás García Trillos GARCIATRILLO@WISC.EDU
Department of Statistics
University of Wisconsin
Madison, Wisconsin, USA

Pengfei He HEPENGF1@MSU.EDU
Department of Statistics and Probability
Michigan State University
East Lansing, MI, USA

Chenghui Li CLI539@WISC.EDU
Department of Statistics
University of Wisconsin
Madison, Wisconsin, USA

Editor: Miguel Carreira-Perpiñán

Abstract

In this work we study statistical properties of graph-based algorithms for multi-manifold clustering (MMC). In MMC the goal is to retrieve the multi-manifold structure underlying a given Euclidean data set when this one is assumed to be obtained by sampling a distribution on a union of manifolds $\mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_N$ that may intersect with each other and that may have different dimensions. We investigate sufficient conditions that similarity graphs on data sets must satisfy in order for their corresponding graph Laplacians to capture the right geometric information to solve the MMC problem. Precisely, we provide high probability error bounds for the spectral approximation of a tensorized Laplacian on $\mathcal{M}$ with a suitable graph Laplacian built from the observations; the recovered tensorized Laplacian contains all geometric information of all the individual underlying manifolds. We provide an example of a family of similarity graphs, which we call annular proximity graphs with angle constraints, satisfying these sufficient conditions. We contrast our family of graphs with other constructions in the literature based on the alignment of tangent planes. Extensive numerical experiments expand the insights that our theory provides on the MMC problem.

Keywords: multi-manifold clustering, graph Laplacian, spectral convergence, manifold learning, discrete to continuum limit.

1. Introduction

In this work we study the problem of *multi-manifold clustering* (MMC) from the perspective of spectral geometry. Multi-manifold clustering is the task of identifying the structure of

1. All authors contributed equally to this work. Their names are listed in alphabetical order by last name.

multiple manifolds that underlie an observed data set $X = \{x_1, \dots, x_n\}$, its main challenge being that in general the underlying manifolds may be non-linear, may intersect with each other, and may have different dimensions (see Figures 1-3 and Figures 37-40 for some illustrations). While spectral methods for learning have been analyzed by several authors throughout the past two decades in settings as varied as unsupervised, semi-supervised, and supervised learning, less is known about their theoretical guarantees for the specific multi-manifold clustering problem. We analyze MMC algorithms that are based on the construction of suitable similarity graph representations for the data and in turn on the spectra of their associated graph Laplacians. We provide statistical error guarantees for the identification of the underlying manifolds as well as for the recovery of their individual geometry.

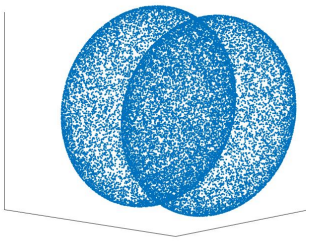

Figure 1

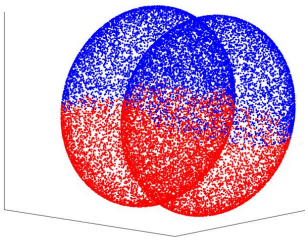

Figure 2

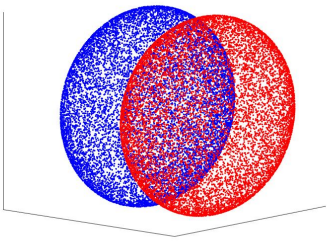

Figure 3

Figure 1 illustrates two intersecting ellipsoids (two dimensional). A *good* multi-manifold clustering algorithm must identify the two underlying ellipsoids. Figure 2 and Figure 3 show the spectral clustering with k-NN graph and annular proximity graph with angle constraint, respectively; see section 3.

As for most spectral approaches to clustering, we are interested in studying spectral properties of graph Laplacian operators of the form

$$\Delta_n u(x_i) := \sum \omega_{ij}(u(x_i) - u(x_j)), \quad x_i \in X. \tag{1.1}$$

Here, the ω_{ij} are appropriately defined symmetric weights that in general depend on the proximity of points x_i, x_j, and, importantly, on a mechanism that detects when points belong to different manifolds even if lying close to each other. Once the graph Laplacian is constructed, we follow the spectral clustering algorithm: the first N eigenvectors of Δ_n (denoted $\psi_1, \dots, \psi_N$) are used to build an embedding of the data set X into $\mathbb{R}^N$:

$$x_i \in X \longmapsto \begin{pmatrix} \psi_1(x_i) \\ \vdots \\ \psi_N(x_i) \end{pmatrix} \in \mathbb{R}^N.$$

In turn, with the aid of a simple clustering algorithm such as k-means the embedded data set is clustered. A successful algorithm will produce clusters that are in agreement with the different manifolds underlying the data set.

As can be imagined, the success of spectral clustering when applied to MMC problems relies strongly on the specific similarity weights ω_{ij} that determine the graph Laplacian, its eigenvectors, and ultimately the partitioning of the data. In the literature, authors have considered different types of mechanisms to discriminate points that lie on different manifolds. Some strategies include the use of local tangent planes from data Arias-Castro et al. (2017); Goh and Vidal (2007); Elhamifar and Vidal (2011); Wang et al. (2011) (whose angles are compared), and the construction of paths between different points (e.g. geodesics) that are considered admissible if they do not exhibit sudden turns (effectively imposing a curvature constraint) Babaeian et al. (2015). All these methods are inspired by heuristics that are meaningful at the continuum level (i.e. the infinite data setting) and use second order geometric information to detect the different intersecting manifolds. While these heuristics provide practical insights, in general they do not guarantee the success of the employed methodologies for MMC at the finite sample level. Part of the motivation for this work is precisely to establish a more concrete and mathematically precise link between the heuristic motivation at the continuum level and the actual methodologies that are used in practice. It is worth highlighting that widely known graph constructions such as ε-proximity graphs or k-NN graphs used for standard data clustering tasks (typically aimed at detecting bottle-necks in data sets) are in general not suitable for MMC. To illustrate this, take for example Figure 2. There, we have used a k-NN graph to build a graph Laplacian whose first non-trivial eigenvector has been used to obtain the partition illustrated in the figure; as can be observed, from the geometry induced by the k-NN graph we are unable to distinguish the two underlying ellipsoids.

To start making the results presented in this paper more precise, let us suppose that the data set X is obtained by sampling a distribution μ supported on a set $\mathcal{M}$ of the form

$$\mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_N, \tag{1.2}$$

where the $\mathcal{M}_l$ are smooth compact connected manifolds with no boundary that for the moment are assumed to have the same dimension m; the manifolds $\mathcal{M}_l$ may have nonempty pairwise intersections, but these are assumed to have measure zero relative to the volume forms of each of the manifolds involved. The distribution μ is assumed to be a mixture model taking the form

$$d\mu = w_1\rho_1 d\mathrm{vol}_{\mathcal{M}_1} + \cdots + w_N\rho_N d\mathrm{vol}_{\mathcal{M}_N},$$

for smooth density functions $\rho_l : \mathcal{M}_l \to \mathbb{R}$ and positive weights w_i that add to one; henceforth we use $d\mathrm{vol}_{\mathcal{M}_l}$ to denote integration with respect to the Riemannian volume form associated to $\mathcal{M}_l$. A *tensorized Laplacian* $\Delta_{\mathcal{M}}$ acting on functions f on $\mathcal{M}$ (which will be written as $f = (f_1, \ldots, f_N)$, where $f_l : \mathcal{M}_l \to \mathbb{R}$) can be defined according to

$$\Delta_{\mathcal{M}} f := (w_1\Delta_{\mathcal{M}_1} f_1, \ldots, w_N\Delta_{\mathcal{M}_N} f_N), \tag{1.3}$$

where $\Delta_{\mathcal{M}_l}$ is a Laplacian operator mapping regular enough functions $f_l : \mathcal{M}_l \to \mathbb{R}$ into functions $\Delta_{\mathcal{M}} f_l : \mathcal{M}_l \to \mathbb{R}$ according to

$$\Delta_{\mathcal{M}_l} f_l = -\frac{1}{\rho_l}\mathrm{div}_{\mathcal{M}_l}\left(\rho_l^2 \nabla_{\mathcal{M}_l} f_l\right).$$

In other words, the operator $\Delta_{\mathcal{M}}$ acts in a coordinatewise fashion, effectively treating each manifold $\mathcal{M}_i$ independently. It is then straightforward to show that eigenfunctions of $\Delta_{\mathcal{M}}$ are spanned by functions of the form

$$(0, \dots, f_l, \dots, 0)$$

for some l, where f_l is an eigenfunction of $\Delta_{\mathcal{M}_l}$. This means that the spectrum of $\Delta_{\mathcal{M}}$ splits the geometries of the $\mathcal{M}_l$. In particular, the different $\mathcal{M}_l$ can be detected by retrieving the eigenfunctions with zero eigenvalue.

Our first main results (**Theorem 6 and Theorem 8**) state that, provided that the weights ω_{ij} defining the graph Laplacian operator Δ_n in (1.1) satisfy two conditions referred to as *full inner connectivity* and *sparse outer connectivity*, the eigenvalues (appropriately scaled) and eigenvectors of Δ_n approximate the eigenvalues and eigenfunctions of the tensorized Laplacian $\Delta_{\mathcal{M}}$; we obtain high probability quantitative bounds for the error of this approximation. The bottom line is that our results imply that the spectral methods studied here are guaranteed, at least for large enough n, to recover the underlying multi-manifold structure of the data; see Figure 3 for an illustration. Our work extends the growing literature of works that study the connection between graph Laplacians on data sets and their continuum analogues. This literature, which we review in section 1.1.1, has mostly focused on the smooth setting where multiple intersecting manifolds are not allowed.

In our second main result (**Theorem 10**), we present some results for the case when the dimensions of the manifolds $\mathcal{M}_i$ do not agree. In this more general setting, the spectrum of the graph Laplacian Δ_n does not recover the tensorized geometry captured by $\Delta_{\mathcal{M}}$ as introduced earlier, but rather, only the tensorized geometry of the manifolds with the *largest* dimension, effectively quotienting out the geometric information of manifolds with dimension strictly smaller than the maximum dimension.

After presenting our general results, we move on to discussing specific examples of graph constructions that satisfy the full inner connectivity and sparse outer connectivity conditions. In particular, we discuss a family of annular proximity graphs with angle constraints (see section 3) that we show satisfies the desired connectivity conditions. In the final section of the paper, we present some insights into the behavior of this graph construction and its ability to tackle the MMC problem in concrete numerical examples, as well as present a performance comparison with other existing spectral-based MMC approaches.

1.1 Related work

In this section we provide an overview of some related works that study spectral clustering and its connection with manifold learning, as well as other works that study the specific multi-manifold clustering problem.

1.1.1 Spectral clustering and manifold learning

In the past two decades, several authors have attempted to establish precise connections between operators such as graph Laplacians built from random data and analogous differential operators defined at the continuum level. To make this connection mathematically precise, one can assume that the data are sampled from some distribution supported on a

certain geometric object $\mathcal{M}$. In the setting where $\mathcal{M}$ is a smooth compact manifold embedded in $\mathbb{R}^d$ that has no boundary, several authors have studied the connection between ε-graph-based Laplacians and weighted versions of Laplace Beltrami operators on $\mathcal{M}$. For pointwise consistency results we refer the reader to Singer (2006); Hein et al. (2005, 2007); Belkin and Niyogi (2005); Ting et al. (2010); Giné and Koltchinskii (2006)). Regarding *spectral convergence* of graph Laplacians, a notion of convergence that is relevant for spectral clustering, the regime $n \to \infty$ and ε constant is studied in von Luxburg et al. (2008) and also in Singer and Wu (2017). The latter analyzes connection Laplacians, which are operators acting on vector fields as opposed to functions. Works that have studied regimes where ε is allowed to decay to zero include Tao and Shi (2020); Burago et al. (2014); García Trillos et al. (2019); Lu (2022); Calder and García Trillos (2022); Dunson et al. (2021); Wormell and Reich (2021). The mathematical theory around graph Laplacians in the smooth manifold setting has developed considerably and even regularity estimates of graph Laplacian eigenvectors are now available (see Calder et al. (2022)).

In the setting of a smooth compact manifold $\mathcal{M}$ *with* boundary, graph Laplacians are seen to behave differently around the manifold's boundary than in their interior. This has been observed in works like Vaughn et al. (2019); Wu and Wu (2018), which study this setting and obtain expansions for graph Laplacians that hold all the way up to the boundary. Earlier works such as García Trillos and Slepčev (2018) use variational methods to provide spectral asymptotic consistency results in this setting but don't obtain convergence rates nor describe the behavior of graph Laplacians close to the boundary. The work Lu (2022) provides rates for spectral convergence in the setting of manifolds with boundary and also considers the case where $\mathcal{M}$ is of the form (1.2). However, in contrast to what we do here, the aim in Lu (2022) is not to analyze graph constructions that guarantee the recovery of the multi-manifold structure of the data, focusing instead on analyzing intrinsic proximity graphs to the union of the intersecting manifolds. Our analysis shares aspects and ideas with this and some of the other works previously mentioned, but to fulfill our goals we must introduce new constructions and estimates not currently available. In addition, to the best of our knowledge, we are the first to present an analysis of the full spectrum of graph Laplacians when data points are supported on a union of intersecting manifolds that have *different* dimensions. Previous work Arias-Castro (2011) had analyzed the null space of a graph Laplacian when the generators (manifolds), although potentially of different dimensions, were assumed to be separated from each other.

From a methodological perspective, it is also worth highlighting several other works that have studied the use of metrics different from the Euclidean one to build proximity graphs for clustering and other unsupervised learning tasks. The idea in those papers is to use the modified metrics to improve the performance of spectral clustering when applied to data sets with some special geometric structure. Examples include: Ahmed et al. (2015); Rosenfeld and Pfaltz (1966); Normand et al. (2011); Fischer et al. (2001); Chang and Yeung (2008); Little et al. (2020, 2022). In a sense, our approach in this paper is in line with the general perspective taken in the previously mentioned works, only that in our case we have a different geometric structure in mind, i.e., we consider multiple intersecting manifolds.

1.1.2 Multi-manifold clustering

In contrast to the graph constructions which are analyzed in most of the works mentioned in section 1.1.1 (i.e. standard ε-graphs and k-NN graphs), graph constructions for multi-manifold clustering must incorporate a mechanism to discriminate between points that lie on different manifolds. One such mechanism relies on the approximation of approximate tangent planes around every point. Pairs of nearby points are then endowed high weights whenever their corresponding tangent planes are aligned, as proposed in Arias-Castro et al. (2017). The recovery of tangent planes from data is a problem that has been studied theoretically in papers such as Aamari and Levrard (2018) (see also references within). The methodology proposed in Singer and Wu (2017), which uses a connection Laplacian, can be considered as a MMC algorithm since it also uses tangent plane information to inform the affinity between points. The LLMC algorithm from Goh and Vidal (2007) is also based on locally fitting planes to points and their nearest neighbors. Sparse Manifold Clustering and Embedding(SMCE) in Elhamifar and Vidal (2011) implicitly attempts to recover tangent planes too; a sparse representation of points in a neighborhood is sought via a local l^1 optimization problem. Another work that considers affinities based on local tangent planes is Wang et al. (2011). In section 3.2 we will discuss some properties of the tangent plane based graphs and their effect on spectral clustering for MMC. Wang et al. (2014, 2015) also consider estimating tangent planes to solve the MMC problem, but now in a generalized setting where the ambient space is a curved manifold and not $\mathbb{R}^d$.

At a high level, all multi-manifold clustering algorithms use curvature information to detect pairs of points that, while close to each other, lie on different manifolds. Measuring the difference of tangent planes is one way to capture curvature, but there are alternative ways. For example, works like Chen and Lerman (2009b,a) use the notion of polar curvature between collections of points to define an algorithm known as spectral curvature clustering (SCC). In Chen and Lerman (2009a) the authors present some theoretical analysis of SCC in the setting where the data are sampled from multiple flats with the same dimension. In Arias-Castro et al. (2011), a localized spectral curvature clustering algorithm is proposed to find local curvature information by constructing similarity graphs that are obtained by aggregating certain alignment score for a collection of data tuples of high enough order. This method is computationally too intensive given that it requires to consider tuples of order larger than the dimension of the manifolds. Besides, from a theoretical perspective, the method scales very poorly with the dimensionality of the underlying manifolds, and the authors indicate that it can only solve the MMC problem in the setting of intersecting curves, i.e. 1d manifolds.

Curvature can also be captured by measuring how quickly paths turn as proposed in Babaeian et al. (2015). Our graph construction from section 3 is inspired by the one proposed in Babaeian et al. (2015), but with some important differences that we will motivate and explain throughout the paper. These differences, in particular, allow us to provide a comprehensive theoretical analysis and provide theoretical guarantees for the success of our algorithms.

To wrap up this brief literature review, it is worth mentioning a special setting where the manifolds $\mathcal{M}_l$ are linear subspaces of the ambient space. In that case, the multi-manifold clustering problem reduces to *subspace clustering* (SubC), a problem that has received

considerable attention in the past decades due to its multiple applications in tasks such as image segmentation, motion segmentation, and image representation (see Vidal (2011)). Many algorithms in SubC rely strongly on the assumed global flat structure of the data and on the fact that the origin is known to lie on the intersection of the spaces. Unfortunately, these approaches can not be used directly for a general multi-manifold clustering task, so we will not discuss them in more detail. Conversely, while it is possible to use general MMC approaches to solve SubC problems, it is clear that the performance of general MMC methods will in general be far from satisfactory when compared to the performance of SubC approaches, which actively target the subspace structure of the manifolds. We refer the reader interested in the SubC problem to the following list of papers and their references: Boult and Brown (1991); Yan and Pollefeys (2006); Zhang et al. (2012); Park et al. (2014); Vidal et al. (2005); Ying Wu et al. (2001); Vidal et al. (2005); Ying Wu et al. (2001); Elhamifar and Vidal (2009, 2010); Oswal and Nowak (2018); Liu et al. (2010).

1.2 Contributions and outline

We summarize our contributions as follows:

- We analyze graph Laplacians on families of proximity graphs when the nodes of the graphs are random data points that are supported on a union of unknown *intersecting* manifolds. The manifolds may all have *different* dimensions.
- We introduce two sufficient conditions that similarity graphs must satisfy in order to recover, from a graph Laplacian operator, the geometric information (as contained in the spectrum of weighted Laplace-Beltrami operators) of the individual smooth manifolds underlying the data set. These conditions are referred to as *full inner connectivity* and *sparse outer connectivity*.
- We introduce and analyze *annular* proximity graphs and their effect on multi-manifold clustering. These are simple extensions of ε-proximity graphs that nonetheless can be shown to be, theoretically and numerically, better than the vanilla ε-graphs for multi-manifold clustering.
- We analyze a family of *annular proximity graphs with angle constraints*. This family is shown to satisfy the full inner connectivity and sparse outer connectivity conditions when their parameters are tuned appropriately. We contrast this construction with other constructions such as those based on local PCA, which in general do not satisfy the full inner connectivity condition.
- Through numerical examples and some heuristic computations, we provide further insights into the use of spectral methods for multi-manifold clustering.

The rest of the paper is organized as follows. Our theoretical framework is presented in section 2, where we formalize the setting for the multi-manifold clustering problem, introduce the definitions of sparsely outer connected and fully inner connected similarity graphs, and state our main theoretical results. In our first results, the ones in section 2.3, we assume that all underlying manifolds have the same dimension, and in section 2.4 we extend

to settings where the dimensions of the underlying manifolds can be different. In section 3 we discuss an example of a graph construction that satisfies the full inner connectivity and sparse outer connectivity conditions. In section 4 we present a series of numerical experiments whose goal is to illustrate the theory developed throughout the paper and highlight some drawbacks of the MMC methods discussed in the paper. In Appendix A we present the proofs of all the results from sections 2.3 and 2.4.

2. Set up and main results

Let $\{\mathcal{M}_l\}_{l=1}^N$ be a collection of N smooth, compact manifolds without boundary embedded in $\mathbb{R}^d$. We denote by m_l the dimension of manifold $\mathcal{M}_l$ and $m = \max_{l=1,\dots,N}\{m_l\}$. Let $\mathcal{M}$ be the union:

$$\mathcal{M} := \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_N.$$

Let $X = \{x_1, \dots, x_n\}$ be i.i.d. samples from a distribution μ on $\mathcal{M}$ of the form:

$$d\mu = \sum_{l=1}^N w_l \rho_l(x) d\text{vol}_{\mathcal{M}_l}(x), \quad \text{where } w_l > 0, \quad \sum_{l=1}^N w_l = 1. \tag{2.1}$$

In the above, for each l, $d\text{vol}_{\mathcal{M}_l}$ is used to denote integration with respect to the Riemannian volume form associated to the manifold $\mathcal{M}_l$, and the probability density $\rho_l : \mathcal{M}_l \to \mathbb{R}$ is assumed to be $C^2(\mathcal{M}_l)$ and satisfy

$$\frac{1}{c_\rho} \leq \rho_l(x) \leq c_\rho, \quad \forall l = 1, \dots, N$$

for some positive constant $c_\rho > 1$. We use μ_l to denote the probability measure $\rho_l d\text{vol}_{\mathcal{M}_l}$. Notice that from (2.1) it follows that the number of data points n_l in manifold $\mathcal{M}_l$ is with very high probability within the interval $[w_l n - t, w_l n + t]$ for some tolerance level t at least in the order of $\sqrt{n}$.

While we will not require the manifolds $\mathcal{M}_l$ to be separated from each other in a distance sense (i.e., we allow manifolds to intersect with each other), we will assume that they are sufficiently "well separated" in an angular sense that we specify below and that we illustrate in Figure 4.

Assumption 1 *For every l, k we assume:*

1. *The intersection $\mathcal{M}_{lk} := \mathcal{M}_l \cap \mathcal{M}_k$ is either the empty set or a smooth manifold of dimension m_{kl} satisfying $0 \leq m_{kl} < \min\{m_l, m_k\}$. In particular, $\mathcal{M}_{lk}$ is of measure zero according to $\text{vol}_{\mathcal{M}_l}$ and $\text{vol}_{\mathcal{M}_k}$.*

2. *For every point x in $\mathcal{M}_l \cap \mathcal{M}_k$ we have:*

$$\sup_{v \in \mathcal{T}_x \mathcal{M}_{lk}^{\perp_l}, \widetilde{v} \in \mathcal{T}_x \mathcal{M}_{lk}^{\perp_k}} |\angle(v, \widetilde{v}) - \frac{\pi}{2}| \leq \beta, \tag{2.2}$$

for some fixed β strictly smaller than $\frac{\pi}{2}$. In the above, $\angle(v, \widetilde{v})$ denotes the angle between vectors $v, \tilde{v}$ (recall that all manifolds are embedded in the ambient space $\mathbb{R}^d$),

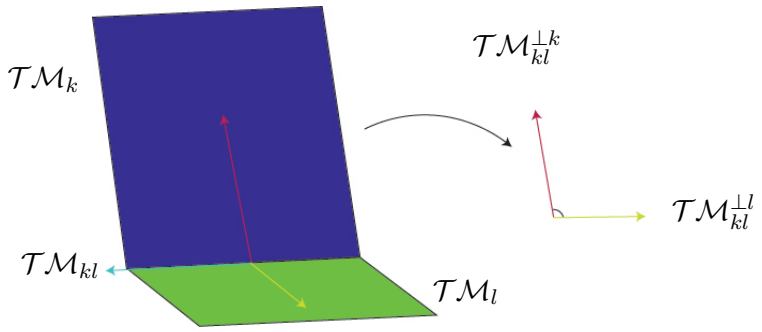

Figure 4: $\mathcal{TM}_{kl}^{\perp k}$ is the orthogonal complement of $\mathcal{TM}_{kl}$ in $\mathcal{TM}_k$. $\mathcal{TM}_{kl}^{\perp l}$ is defined analogously. The second part in Assumption 1 is better satisfied when the angle between these spaces is close to ninety degrees.

and $\mathcal{T}_x\mathcal{M}_{lk}^{\perp_l}$ denotes the orthogonal complement of $\mathcal{T}_x\mathcal{M}_{lk}$ in $\mathcal{T}_x\mathcal{M}_l$, and $\mathcal{T}_x\mathcal{M}_{lk}^{\perp_k}$ is defined analogously.

In the above, and in the remainder, we use $\mathcal{T}_x\mathcal{M}_l$ to denote the tangent plane to $\mathcal{M}_l$ at the point $x \in \mathcal{M}_l$; also, we use $\mathcal{TM}_l$ to denote the tangent bundle associated to $\mathcal{M}_l$. Notice that the second condition in Assumption 1 states that if two manifolds $\mathcal{M}_l$ and $\mathcal{M}_k$ do intersect, they do so in a non-tangential way; see Figure 4 below.

2.1 Fully inner connected, and sparsely outer connected graphs

We endow the data set X with a weighted graph structure (X, ω), where the weights ω are specified by the data. In this section we present the definitions of fully inner connected and sparsely outer connected graphs. The notion of full inner connectivity depends on a prespecified family of base proximity graphs that we introduce next.

Definition 1 *Given $0 \leq \varepsilon_- < \varepsilon_+$ and data points x_i, x_j, we define their $\varepsilon_+, \varepsilon_-$-weight as:*

$$\omega_{ij}^{\varepsilon_+,\varepsilon_-} := \begin{cases} 1 & \text{if } \varepsilon_- \leq |x_i - x_j| \leq \varepsilon_+ \\ 0 & \text{otherwise.} \end{cases}$$

We use $\omega_{ij}^{\varepsilon}$ as shorthand notation for $\omega_{ij}^{\varepsilon,0}$. Notice that with this definition we have the identity:

$$\omega_{yx}^{\varepsilon_+,\varepsilon_-} = \omega_{yx}^{\varepsilon_+} - \omega_{yx}^{\varepsilon_-}.$$

Remark 2 *The above definition extends the notion of ε-proximity graph and in principle allows pairs of points that are too close to each other to have zero weight. While in the literature this annular proximity graphs have not been given any attention, we will see later on that this extended notion is convenient from qualitative and quantitative points of view for the MMC problem; see the discussion surrounding Lemmas 19 and 20.*

Definition 3 (Fully inner Connected graphs) *Let $X = x_1, \dots, x_n$ be samples from μ as defined in (2.1). A weighted graph (X, ω) is said to be fully inner connected relative to the $\varepsilon_+, \varepsilon_-$ weights as $n \to \infty$, if with probability $1 - C_1(n)$, where $C_1(n) \to 0$ as $n \to \infty$, for any pair of points x_i, x_j belonging to the same manifold $\mathcal{M}_k$ we have $\omega_{x_i,x_j} = \omega_{x_i,x_j}^{\varepsilon_+,\varepsilon_-}$.*

Remark 4 *It is possible to generalize the definition of full inner connectivity considered here and adapt it to other base proximity graph constructions like weighted proximity graphs. Indeed, the essential requirement that the full inner connectivity condition imposes on (X, ω) is that it should behave like a graph that can capture the underlying geometry of each individual manifold. We have chosen annular proximity graphs here because 1) they can be used to capture the geometry of the underlying manifolds (as shown by our analysis: just take Theorems 6 and 8 and consider the case where the number of manifolds is equal to one), and 2) because they introduce an extra degree of flexibility that, as we show later on, allows us to prove better recovery guarantees for MMC than what we can show for standard ε-graphs; see Remark 2.*

Next, we introduce the notion of spare outer connectivity.

Definition 5 (Sparsely Outer Connected graphs) *Let $X = x_1, \dots, x_n$ be samples from μ as defined in (2.1), and let (X, ω) be a weighted graph. Let N_{sl} be the number of connections between $x_i \in \mathcal{M}_s$ and $x_j \in \mathcal{M}_l$ such that $\omega_{ij} > 0$, and let*

$$N_0 := \max_{l \neq s} \{N_{ls}\}.$$

The graph is said to be sparsely outer connected relative to ε_+ and ε_- converging to zero as $n \to \infty$ if with probability one $\frac{N_0}{n^2(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})} \to 0$ as $n \to \infty$. We recall that $m = \max_{l=1,\dots,N} m_l$.

The above notions will capture the intuitive desire of giving high weights to pairs of points that are close to each other when they belong to the same manifold (full inner connectivity condition) and to give low weights to pairs of points when they lie on different manifolds (sparse outer connectivity condition). These notions are geometric adaptations to the setting of interest of general notions explored in the literature to describe the feasibility of a clustering problem. We elaborate on this next.

In the setting considered in Ng et al. (2001), for example, given a network (X, W), four conditions on the network (that depend on the Laplacian and degree function of the network) are proposed to ensure that a certain spectral embedding constructed from (X, W) maps the original set of nodes X to points that are close to a set of orthogonal vectors in Euclidean space; notice that the conditions in Ng et al. (2001) do not rely on any specific modelling assumption on the generative process that produces the data set X. Works like Schiebinger et al. (2015) and García Trillos et al. (2021) have taken a different perspective and introduced sufficient geometric conditions on certain families of *generative models* that, at the "ground-truth" level, guarantee the feasibility of a certain ground-truth level clustering problem. These works then show that, when their proposed modelling assumptions hold, the conditions in Ng et al. (2001) are satisfied with high probability by certain network constructions (X, W), where X is a set of samples from the generative

model, and W is a suitable weight matrix over X. The notions of inner connectivity and outer connectivity considered here can be interpreted as conditions that a given weight matrix W built over a data set X sampled from a model like (2.1) must satisfy in order for the network (X, W) to: 1) satisfy the conditions in Ng et al. (2001) and 2) have clusters that are consistent with the underlying manifolds in the generative model (2.1). In this sense, the results that we present in this paper are analogous to those in Schiebinger et al. (2015); García Trillos et al. (2021), except that the geometric structure of the generative models in our paper is substantially different from the ones in those works. Other works in the literature such as Vu (2018) have considered other types of "clusterability" conditions, requiring data points within each cluster to be close to each other and points from different clusters to be far away from each other. These conditions are certainly not satisfied in this paper, mainly because the separation between manifolds can in fact be equal to zero.

Before we finish this section, we remark that the discussion in the upcoming sections 2.2-2.4 will not be restricted to any particular graph construction. In the results presented there, we quantify the error of approximation of the spectra of tensorized Laplacians at the continuum level from the graph Laplacian associated to (X, ω). This approximation error will naturally depend on the quantities $C_1(n)$ and N_0 appearing in the definition of the inner and outer connectivity conditions. In section 3, we provide one example of a family of graphs that satisfies the inner and outer connectivity conditions. The graphs in that family are obtained by pruning an $\varepsilon_+, \varepsilon_-$ graph, removing edges between points for which there is no almost straight path connecting them. In section 3.2, we discuss other popular choices of weights ω that are based on the comparison of local tangent planes, but that, as we will discuss, do not, in general, satisfy the full inner connectivity condition.

2.2 Basic properties of the spectrum of the operator $\Delta_{\mathcal{M}}$

In order to state our main theoretical results in sections 2.3 and 2.4 we first discuss some basic properties of the spectrum of the operator $\Delta_{\mathcal{M}}$ in (1.3) and its relation to the MMC problem.

Let $L^2(\mu)$ be the space of N tuples $(f_1, \dots, f_N)$ where each $f_l \in L^2(\mu_l)$. We endow the space $L^2(\mu)$ with the tensorized inner product:

$$\langle f, g\rangle_{L^2(\mu)} := \sum_{l=1}^{N} w_l \langle f_l, g_l\rangle_{L^2(\mu_l)} = \sum_{l=1}^{N} w_l \int_{\mathcal{M}_l} f_l(x) g_l(x) d\mu_l(x),$$

where $f = (f_1, \dots, f_N) \in L^2(\mu)$ and $g = (g_1, \dots, g_N) \in L^2(\mu)$. A tensorized Sobolev space $H^1(\mu)$ is defined as the space of $f = (f_1, \dots, f_N) \in L^2(\mu)$ for which $f_l \in H^1(\mathcal{M}_l)$ for each $l = 1, \dots, N$. In particular, for elements $f \in H^1(\mu)$ the quantity

$$\sum_{l=1}^{N} w_l^2 \int_{\mathcal{M}_l} |\nabla f_l(x)|^2 \rho_l^2(x) dx$$

is finite. We then define the weighted Dirichlet energy:

$$D(f) := \begin{cases} \sum_{l=1}^{N} w_l^2 \int_{\mathcal{M}_l} |\nabla f_l(x)|^2 \rho_l^2(x) d\mathrm{vol}_{\mathcal{M}_l}(x), & \text{if } f \in H^1(\mu), \\ +\infty, \quad \text{if } f \in L^2(\mu) \setminus H^1(\mu). \end{cases} \tag{2.3}$$

Now, notice that the operator $\Delta_{\mathcal{M}}$ is self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle_{L^2(\mu)}$ simply because each of the operators $\Delta_{\mathcal{M}_l}$ is self-adjoint w.r.t. $\langle \cdot, \cdot \rangle_{L^2(\mu_l)}$ (e.g. see García Trillos and Slepčev (2018)). Given that each $L^2(\mathcal{M}_l, \rho_l)$ admits an orthonormal basis $\{f_l^k\}_{k \in \mathbb{N}}$ of eigenvectors of $\Delta_{\mathcal{M}_l}$, we can see that the set of $f^k \in L^2(\mu)$ of the form

$$f^k = (0, \dots, \frac{1}{\sqrt{w_l}} f_l^k, \dots, 0) \tag{2.4}$$

for $k \in \mathbb{N}$ and $l = 1, \dots, N$ is an orthonormal basis for $L^2(\mu)$. In addition, for such f^k we have

$$\Delta_{\mathcal{M}} f^k = (0, \dots, \frac{w_l}{\sqrt{w_l}} \Delta_{\mathcal{M}_l} f_l^k, \dots, 0) = w_l \lambda (0, \dots, \frac{1}{\sqrt{w_l}} f_l^k, \dots, 0) = w_l \lambda f^k$$

for some eigenvalue λ of $\Delta_{\mathcal{M}_l}$. In conclusion, we can build an orthonormal basis for $L^2(\mu)$ consisting of eigenfunctions of $\Delta_{\mathcal{M}}$ of the form (2.4). From the above we can also conclude that the set of eigenvalues of $\Delta_{\mathcal{M}}$ is the set of numbers of the form $w_l \lambda$ for some l, where λ is an eigenvalue of $\Delta_{\mathcal{M}_l}$. In terms of the Dirichlet energy defined in (2.3), the eigenvalues of $\Delta_{\mathcal{M}}$, arranged in increasing order according to multiplicity, can be written as

$$\lambda_l = \min_{S \in \mathfrak{S}_l} \max_{f \in S \setminus \{0\}} \frac{D(f)}{\|f\|^2_{L^2(\mu)}}. \tag{2.5}$$

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

Regarding the zero eigenvalue of $\Delta_{\mathcal{M}}$, notice that since the manifolds $\mathcal{M}_l$ were assumed to be connected, the multiplicity of the zero eigenvalue for the operator $\Delta_{\mathcal{M}}$ is equal to N. Moreover, an orthonormal basis for this eigenspace is the set of functions of the form $(0, \dots, c_l \mathbb{1}_{\mathcal{M}_l}, \dots, 0)$ where c_l is a normalization constant. This observation is the key property that allows us to think of the multi-manifold clustering problem in terms of the spectrum of the operator $\Delta_{\mathcal{M}}$. However, it should be clear that the tensorized Laplacian has much more information than that needed to solve the MMC problem.

For convenience, we also introduce Dirichlet energies associated to each manifold $\mathcal{M}_l$:

$$D_l(f_l) := \begin{cases} \int_{\mathcal{M}_l} |\nabla f_l(x)|^2 \rho_l^2(x) d\text{vol}_{\mathcal{M}_l}(x), & \text{if } f_l \in H^1(\mu_l), \\ \\ +\infty \quad \text{if} f \in L^2(\mu_l) \setminus H^1(\mu_l). \end{cases} \tag{2.6}$$

2.3 Convergence results in the $m_1 = \cdots = m_N$ case

In this section we establish high probability error bounds between the spectrum of a rescaled version of the graph Laplacian Δ_n defined in (1.1) and the spectrum of $\Delta_{\mathcal{M}}$ under the additional assumption that all manifolds $\mathcal{M}_k$ have the same dimension. The results presented in this section apply to generic weighted graphs (X, ω), but the error estimates are only meaningful when the quantities N_0, $C_1(n)$ and ε_+ from section 2.1 scale appropriately with the number of data points. In section 3 we present a specific construction for (X, ω) where we can make our error estimates concrete.

In what follows we make the following assumptions on the parameters $\varepsilon_+, \varepsilon_-, \widetilde{\delta}, \theta$. Here $\tilde{\delta}$ and θ are small parameters that we use to tune the probabilities of some random events defined in corollary 25.

Assumption 2 *We assume that the quantities $\varepsilon_+, \varepsilon_-, \tilde{\delta}, \theta$ satisfy:*

(1) $\varepsilon_+ \leq \min\{1, \frac{R}{2}, CK^{-1/2}, i_0\}$*, where R, K are uniform upper bounds on the reach and on the absolute values of the sectional curvatures for all the manifolds, i_0 is a lower bound on the injectivity radius of all manifolds, and C is a constant no larger than 1.*

(2) $\varepsilon_- \leq \frac{1}{4}\varepsilon_+$*; the $\frac{1}{4}$ here is an arbitrary number smaller than* 1*.*

(3) $\frac{c}{n^{1/m}} < \widetilde{\delta}$

(4) $C(\widetilde{\delta} + \theta) \leq \frac{1}{2c_\rho}$*, where $\frac{1}{c_\rho}$ is the lower bound of ρ .*

With the above assumptions we can make sure that with the underlying $\varepsilon_+, \varepsilon_-$-weighted graph we can approximate the operators $\Delta_{\mathcal{M}_k}$ in each of the $\mathcal{M}_k$; this part does not rely on the assumption that all manifolds have the same dimension, and only depends on the full inner connectivity property. For the spectrum of the graph Laplacian associated to (X, ω) to successfully recover the spectrum of $\Delta_{\mathcal{M}}$ we need (X, ω) to be fully inner connected and sparsely outer connected as we will make explicit in our first theorem.

Theorem 6 (Convergence rate for eigenvalues) *Let μ be a probability measure on $\mathcal{M}$ as in* (2.1)*. Suppose that the $\mathcal{M}_k$ forming the $\mathcal{M}$ satisfy Assumptions 1, and assume also that $m_1 = \cdots = m_N$. Let $X = \{x_1, \ldots, x_n\}$ be i.i.d. samples from μ. Let (X, ω) be a symmetric weighted graph and let $\mathcal{L}$ be the rescaled graph Laplacian:*

$$\mathcal{L}u(x) := \frac{1}{n^2(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})} \sum_{y \in X} \omega_{xy}(u(x) - u(y)), \quad x \in X, u : X \to \mathbb{R}. \tag{2.7}$$

Suppose that the quantities $\widetilde{\delta}, \theta, \varepsilon_+, \varepsilon_-$ satisfy Assumptions 2. Let $\lambda_k^{\varepsilon_+, \varepsilon_-}$ be the k-th eigenvalue of $\mathcal{L}$ and let λ_k be the k-th eigenvalue of $\Delta_{\mathcal{M}}$, where $\Delta_{\mathcal{M}}$ is the tensorized Laplacian from (1.3)*. Finally, let $t := \frac{n\omega_{\min}}{2}$. Then there exists a constant C (independent of k) such that, with probability at least $1 - \sum_{l=1}^{N}(nw_l + t)\exp\left(-C(nw_l - t)\theta^2\widetilde{\delta}^m\right) - 2N\exp\left(\frac{-2t^2}{n}\right) - C_1(n)$, for every $k \in \mathbb{N}$ for which*

$$C\widetilde{\delta}\sqrt{\lambda_k} + C(\theta + \widetilde{\delta}) < \frac{1}{k},$$

we have:

$$\left|\lambda_k^{\varepsilon_+, \varepsilon_-} - \sigma_\eta \lambda_k\right| \leq e_k + C\left(\varepsilon_+(\sqrt{\lambda_k} + 1) + \theta + \frac{\widetilde{\delta}}{\varepsilon_+}\right)\lambda_k.$$

In the above, $e_k = \frac{CN_0}{n^2(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})}\left(1 + C'(\lambda_k^{m/2+1} + \widetilde{\delta}\sqrt{\lambda_k} + \theta + \widetilde{\delta})\right)$, and $C_1(n)$ and N_0 are introduced in Definition 3 and Definition 5, respectively. The constant σ_η is given by:

$$\sigma_\eta := \int_{\mathbb{R}^m} |y_1|^2 \eta(|y|) dy, \tag{2.8}$$

where $\eta = \mathbb{1}_{r \leq 1}$ and y_1 is the first coordinate of y.

Remark 7 *1. In general, we should expect a trade-off between the quantities $C_1(n)$ and N_0. That is, in general, an attempt at making N_0 smaller (i.e., erase connections between different manifolds) will typically result in a smaller probability of having a graph that is well connected within each manifold $\mathcal{M}_l$.*

2. The benefits that come from taking $\varepsilon_- > 0$ for the MMC problem are not explicit in the error bounds from Theorem 6. However, as we will see later on, by tuning ε_- appropriately, one can substantially eliminate connections between data points in different manifolds when one considers $\varepsilon_- \sim \varepsilon_+$. This means a substantial decrease in N_0. We explain this in Remark 17 for the specific annular graph construction with angle constraints. The fact that we can improve the performance of MMC algorithms by introducing $\varepsilon_+, \varepsilon_-$-graphs motivates the theoretical analysis that we present in the Appendix.

3. The proof of the estimates in Theorem 6 relies on a variational approach that compares Dirichlet energies at discrete and continuum levels. This approach has been used before in Burago et al. (2014); García Trillos et al. (2019); Lu (2022). However, the structure of the $\varepsilon_+, \varepsilon_-$-graph that we consider here forces us to modify the analysis and present new proofs. Even for a single manifold $\mathcal{M} = \mathcal{M}_1$, the analysis of graph Laplacians on $\varepsilon_+, \varepsilon_-$-graphs is a technical contribution of this work. The actual proof of Theorem 6 appears in section A.5 in the Appendix. Several technical preliminary results are established in the preceding sections.

4. The scaling factor relating $\mathcal{L}$ and Δ_n in (2.7) is irrelevant in practice because the eigenvectors of Δ_n are the same as those for $\mathcal{L}$, and the ratio between eigenvalues of Δ_n coincides with the ratio of eigenvalues of $\mathcal{L}$. In other words, in practice we can work directly with Δ_n without having to compute the rescaling factor.

5. If we choose $1 \gg \varepsilon_+ \gg \left(\frac{\log(n)}{n}\right)^{1/m}$, then, with high probability, the error of approximation of eigenvalues scales like:

$$\frac{N_0}{n^2 \varepsilon_+^{m+2}} + \frac{\left(\frac{\log(n)}{n}\right)^{1/m}}{\varepsilon_+} + \varepsilon_+.$$

This result is analogous to results in Burago et al. (2014) and García Trillos et al. (2019), except that now we have the extra $\frac{N_0}{n^2 \varepsilon_+^{m+2}}$ term. In order for this error estimate to converge to zero in the large data limit we thus need to require the graph to satisfy the sparse outer connectivity condition.

Theorem 8 (Convergence rate for eigenvectors) *Under the same setting and assumptions as in Theorem 6, for every $k \in \mathbb{N}$ there is a constant $c_k = c_k(\mathcal{M})$ such that if*

$$e_k + C\left(\varepsilon_+ \sqrt{\lambda_k} + \varepsilon_+^2 + \theta + \frac{\widetilde{\delta}}{\varepsilon_+}\right) \leq c_k,$$

then, with probability at least $1-\sum_{l=1}^{N}(nw_l+t)\exp\left(-C(nw_l-t)\theta^2\widetilde{\delta}^m\right)-2N\exp\left(\frac{-2t^2}{n}\right)-C_1(n)$*, for every* v_k *normalized eigenvector of* $\mathcal{L}$ *with eigenvalue* λ_k*, there is a normalized eigenfunction* f_k *of* $\Delta_{\mathcal{M}}$ *with eigenvalue* λ_k *such that*

$$\|f_k - v_k\|_{L^2(\mu^n)} \leq \left[Ce_k + C\left(\varepsilon_+\sqrt{\lambda_k} + \varepsilon_+^2 + \theta + \frac{\widetilde{\delta}}{\varepsilon_+}\right)\right]^{1/2} + C_{\mathcal{M},\lambda}\widetilde{\delta},$$

where e_k *is the same as in Theorem 6.*

The proof of this theorem is presented in section A.6 in the Appendix.

Remark 9 *The sparse outer connectivity condition* $\frac{N_0}{n^2(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})} \to 0$ *is imposed to guarantee the recovery of the full spectrum of the tensorized Laplacian in the large data limit. However, we highlight that our error estimates continue to be meaningful even if we only impose* $\frac{N_0}{n^2(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})}$ *to be asymptotically smaller than some small tolerance level* c*.*

2.4 Mixed dimensions.

We generalize our results from section 2.3 to a setting where the manifolds $\mathcal{M}_k$ may have different dimensions. For convenience, we introduce some notation first.

Without the loss of generality we can assume that the manifolds $\mathcal{M}_k$ are indexed in decreasing order of dimension, i.e. $m = m_1 \geq m_2 \geq \cdots \geq m_N$. We let $N_{\max}$ be the number of manifolds with the maximum dimension m, i.e. $m_1 = \cdots = m_{N_{\max}} > m_{N_{\max}+1}$. We set $\mathcal{M}_{\max} := \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_{N_{\max}}$ and write $\langle f, g\rangle_{L^2(\mathcal{M}_{\max})}$ to represent:

$$\langle f, g\rangle_{L^2(\mathcal{M}_{\max})} = \sum_{i=1}^{N_{\max}} w_i \langle f_i, g_i\rangle_{L^2(\mu_i)} = \sum_{i=1}^{N_{\max}} w_i \int_{\mathcal{M}_i} f_i(x)g_i(x)d\mu_i(x).$$

We also use $\|f\|^2_{L^2(\mathcal{M}_{\max})} = \langle f, f\rangle_{L^2(\mathcal{M}_{\max})}$.

Notice that with the above inner product we can identify (isometrically) elements in $L^2(\mathcal{M}_{\max})$ with elements in $L^2(\mu)$ that are zero outside of $\mathcal{M}_{\max}$; throughout section A.7 in the Appendix we may use this identification without any further explanation. Finally, we use $\Delta_{\mathcal{M}_{\max}}$ to denote the tensorized Laplacian (1.3) for $\mathcal{M}_{\max}$ (i.e. just as in (1.3) but with only the first $N_{\max}$ coordinates); we use $D_{\max}$ to denote the corresponding Dirichlet energy defined for $L^2(\mathcal{M}_{\max})$ functions.

Theorem 10 *Let* μ *be a probability measure on* $\mathcal{M}$ *as in* (2.1)*. Suppose that the* $\mathcal{M}_k$ *forming* $\mathcal{M}$ *satisfy Assumptions 1, and let* $N_{\max}, \mathcal{M}_{\max}, \Delta_{\mathcal{M}_{\max}}$ *be defined as before. Set* $\lambda_1, \ldots, \lambda_N = 0$ *and let* $\lambda_{N+1} \leq \lambda_{N+2} \leq \ldots$ *be the list of non-zero eigenvalues of* $\Delta_{\mathcal{M}_{\max}}$ *repeated according to multiplicity.*

Let $X = \{x_1, \ldots, x_n\}$ *be i.i.d. samples from* μ*, let* (X, ω) *be a symmetric weighted graph, and let* $\mathcal{L}$ *be the rescaled graph Laplacian from* (2.7)*. Finally, suppose that the quantities* $\widetilde{\delta}, \theta, \varepsilon_+, \varepsilon_-$ *satisfy Assumptions 2.*

Then, for some constant $C = C(\mathcal{M}, \mu)$*, with probability at least*

$$1 - \sum_{l=1}^{N}(nw_l + t)\exp\left(-\mathrm{C}(nw_l - t)\theta^2\widetilde{\delta}^m\right) - 2N\exp\left(\frac{-2t^2}{n}\right) - C_1(n),$$

for every $k \in \mathbb{N}$ for which

$$C\widetilde{\delta}\sqrt{\lambda_k} + C(\theta + \widetilde{\delta}) < \frac{1}{k},$$

we have:

$$\left|\lambda_k^{\varepsilon_+,\varepsilon_-} - \sigma_\eta \lambda_k\right| \leq e_k + C\left(\varepsilon_+(\sqrt{\lambda_k}+1) + \theta + \frac{\widetilde{\delta}}{\varepsilon_+}\right)\lambda_k.$$

In the above, $e_k = \frac{CN_0}{n^2(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})}\left(1 + C'(\lambda_k^{m/2+1} + \widetilde{\delta}\sqrt{\lambda_k} + \theta + \widetilde{\delta})\right)$, and $C_1(n)$ and N_0 are introduced in Definition 3 and Definition 5, respectively.

In addition, there is a constant $c_k = c_k(\mathcal{M})$ such that if

$$e_k + C\left(\varepsilon_+\sqrt{\lambda_k} + \varepsilon_+^2 + \theta + \frac{\widetilde{\delta}}{\varepsilon_+}\right) \leq c_k,$$

then, with probability at least $1 - \sum_{l=1}^N (nw_l + t)\exp\left(-C(nw_l - t)\theta^2\widetilde{\delta}^m\right) - 2N\exp\left(\frac{-2t^2}{n}\right) - C_1(n)$, for every u_k normalized eigenvector of $\mathcal{L}$ with eigenvalue λ_k, there is a normalized eigenfunction f_k of $\Delta_{\mathcal{M}_{\max}}$ with eigenvalue λ_k such that

$$\|f_k - u_k\|_{L^2(\mu^n)} \leq \left[Ce_k + C\left(\varepsilon_+\sqrt{\lambda_k} + \varepsilon_+^2 + \theta + \frac{\widetilde{\delta}}{\varepsilon_+}\right)\right]^{1/2} + C_{\mathcal{M},\lambda}\widetilde{\delta}.$$

In the above, we interpret the functions $f_1, \ldots, f_N$ as an orthonormal basis for $Span\{\mathbb{1}_{\mathcal{M}_1}, \ldots, \mathbb{1}_{\mathcal{M}_N}\}$.

Remark 11 *1. The proof of this theorem appears in section A.7 in the Appendix. We remark that the proof of Theorem 10 is only based on Theorems 6 and 8 and on a few associated preliminary results.*

2. When manifolds have different dimensions, making sure that the sparse outer connectivity condition is satisfied is more difficult because, in general, N_{kl} is much larger when the dimensions of the manifolds $\mathcal{M}_k$ and $\mathcal{M}_l$ are small than when they are large. Indeed, since the number of points in each manifold is in the order of n, the number of points in a neighborhood of size ε around a point on a manifold with small dimension will be larger than when the manifold has larger dimension.

3. Notice that when manifolds do not intersect the outer sparse connectivity condition is trivially satisfied. If in addition we assume the full inner connectivity condition, then we can conclude that the eigenvectors of the graph Laplacian corresponding to non-zero eigenvalues will only recover the spectra of the manifolds with dimension m.

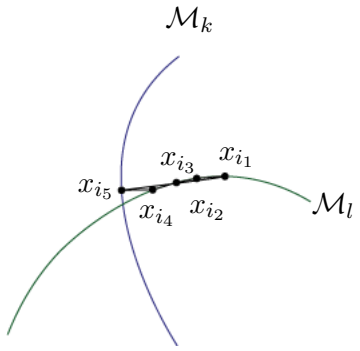

Figure 5: An example of an (α, r)-constrained path where x_{i_1} and x_{i_5} are on different manifolds. To satisfy the constraints, the segments $x_{i_j}x_{i_{j+1}}$ and $x_{i_1}x_{i_m}$ must almost align.

3. Annular proximity graphs with angle constraints

In this section we introduce a graph construction that is both fully inner connected and sparsely outer connected. We start with a definition.

Definition 12 *Let $\alpha \in (0, \pi/2)$ and $r > 0$. We call an ordered sequence of data points*

$$(x_{i_1}, x_{i_2}, ..., x_{i_m})$$

an (α, r)-constrained path between x_{i_1} and x_{i_m} if the following two conditions hold:

1. $\angle(x_{i_m} - x_{i_1}, x_{i_{j+1}} - x_{i_j}) < \alpha, \forall j = 1, 2, ..., m-1.$
2. $|x_{i_j} - x_{i_{j+1}}| < r, \forall j = 1, 2, \ldots, m-1.$

The first condition in the definition of an (α, r)-constrained path requires the path to be almost straight, while the second condition requires consecutive points in the path to be close enough. The example in Figure 5 shows that it is possible to have two points x and y on different manifolds for which there is an (α, r)-constrained path between them. The intuition motivating this definition, however, is that the number of such pairs is small under Assumption 1.

We now introduce the family of *annular proximity graphs with angle constraints* (X, ω) that we study throughout the rest of this section.

Definition 13 (Annular proximity graphs with angle constraints) *Suppose that x_i, x_j are two data points such that $|x_i - x_j| \leq \varepsilon_-$ or $|x_i - x_j| \geq \varepsilon_+$, then we set $\omega_{ij} = 0$. If $\varepsilon_- < |x_i - x_j| < \varepsilon_+$ and there is an (α, r)-constrained path between x_i and x_j, then we set $\omega_{ij} = 1$, otherwise we set $\omega_{ij} = 0$. We refer to this type of graph as an annular proximity graph with angle constraints.*

Algorithm 1 Annular proximity graph with angle constraints

Input: source nodes y_1, y_2; data points $\{y_i\}_{i=3}^{\widetilde{n}}$ such that $|y_i - y_1| \leq \epsilon_+$; parameters $\varepsilon_+, \varepsilon_-, r, \alpha$, where $\varepsilon_+ > \varepsilon_- > r > 0$.
Output: the shortest angle constrained path between y_1 and y_2.

if $|y_1 - y_2| > \varepsilon_+$ or $|y_1 - y_2| < \varepsilon_-$ **then**
  Output $w = 0$.
end if
Construct r-graph E on $\{y_i\}_{i=1}^{\widetilde{n}}$, that is $e_{ij} = \mathbb{1}_{|y_i - y_j| \leq r}$.
for $e_{ij} = 1$ and $\frac{\langle y_j - y_i, y_2 - y_1 \rangle}{|y_j - y_i| \cdot |y_2 - y_1|} < \cos\alpha$ **do**
  Denote $e_{ij} = 0$.
end for
Apply Dijkstra algorithm to find the shortest path between y_1 and y_2 using E.

In section 3.3 we discuss the computational complexity of building annular graphs with angle constraints. As discussed there, ω can be constructed following a simple modification of the constrained Dijkstra's algorithm from Babaeian et al. (2015); see Algorithm 1. On the other hand, from a theoretical perspective, we show that with the right choice of parameters $(\varepsilon_+, \varepsilon_-, \alpha, r)$, these graphs satisfy the full inner connectivity and sparse outer connectivity conditions; the precise statements are contained in Theorem 14 and Theorem 15 below. In particular, in Theorem 15 (see also Remark 17 and an illustration in Figure 10) we quantify the benefits of considering annular graphs with $\varepsilon_- \sim \varepsilon_+$. In section 4.2.1, we revisit the benefits of considering annular graphs for MMC, this time from a numerical perspective.

Theorem 14 *Let n_k be the number of data points in $X \cap \mathcal{M}_k$, let $\alpha \in [0, \pi/4)$ and let $r \leq C\varepsilon_+$. Then, with probability at least*

$$1 - \frac{C_k n^2 \varepsilon_+}{r} \exp\left(-C_k n_k r^{m_k} (\tan\alpha)^{m_k - 1}\right),$$

for any two points $x_i, x_j \in \mathcal{M}_l$ such that $|x_i - x_j| < \varepsilon_+$, there exists an (α, r)-constrained path between x_i and x_j. In the above, C_k is a constant that depends on $\mathcal{M}_k$ and ρ_k.

Theorem 15 *Suppose that Assumption 1 and Assumption 2.1, 2.2 hold, and suppose that $\mathcal{M}_{kl} = \mathcal{M}_k \cap \mathcal{M}_l \neq \emptyset$.*

(1) If $\alpha < \arcsin(\frac{C_{k,l} r}{\varepsilon_+})$, and $r \leq c\varepsilon_+$, then, with probability no less than

$$1 - C_{k,l} n \exp(-C_{k,l} n \min\{r^{m_k - m_{kl}}, r^{m_l - m_{kl}}, \varepsilon_+^{m_k}, \varepsilon_+^{m_l}\}),$$

N_{kl}, the number of connections between points in $X \cap \mathcal{M}_k$ and $X \cap \mathcal{M}_l$, satisfies

$$N_{kl} \leq C_{k,l} n^2 \max\{r^{m_k - m_{kl}} \varepsilon_+^{m_l}, r^{m_l - m_{kl}} \varepsilon_+^{m_k}\}.$$

(2) If in (1) we further assume the lower bound $\varepsilon_- \geq c\varepsilon_+$, then, with probability no less than

$$1 - C_{k,l} n \exp(-C_{k,l} n \min\{(\frac{r^2}{\varepsilon_+} + r\varepsilon_+)^{m_k - m_{kl}}, (\frac{r^2}{\varepsilon_+} + r\varepsilon_+)^{m_k - m_{kl}}, \varepsilon_+^{m_k}, \varepsilon_+^{m_l}\}),$$

we have

$$N_{kl} \leq C_{k,l} n^2 \max\{(\frac{r^2}{\varepsilon_+}+r\varepsilon_+)^{m_k-m_{kl}}\varepsilon_+^{m_l}, (\frac{r^2}{\varepsilon_+}+r\varepsilon_+)^{m_k-m_{kl}}\varepsilon_+^{m_k}\}.$$

Remark 16 *To illustrate our results and obtain concrete error estimates in Theorem 6, for example, consider the case where all manifolds have the same dimension m. Going back to the last point in Remark 7, we need to tune the parameters r and α so that $N_0 \ll n^2\varepsilon_+^{m+2}$ and also, following Theorem 14, we should have $nr^m \tan(\alpha)^{m-1} \gg 1$. Now, from Theorem 15 it follows that we need $r \ll \varepsilon_+^{3/2}$ (assuming the worst case scenario where the dimension $m_{kl} = m-1$, and assuming we impose the lower bound on ε_-, i.e. we treat $\varepsilon_- \sim \varepsilon_+$). On the other hand, if we set $\sin(\alpha) = C\frac{r}{\varepsilon_+}$ we see from Theorem 14 that we need $r^{2m-1}/\varepsilon_+^{m-1} \gg \frac{1}{n}$ (omitting logarithmic terms). Thus, if we take $r \sim \varepsilon_+^2$ and we set $\varepsilon_+ = C\left(\frac{1}{n}\right)^{\frac{1}{3m-1}}$ for large enough constants C we can satisfy all constraints and get from Remark 7 v) the rate (omitting logarithmic terms):*

$$O\left(\frac{1}{n^{1/(3m-1)}}\right)$$

for the convergence of the eigenvalues of the graph Laplacian on an annular path with angle constraints towards the eigenvalues of the tensorized Laplacian on $\mathcal{M}$. For comparison, recall that the convergence rate obtained in García Trillos et al. (2019) for the regular convergence of graph Laplacians was $O\left(\frac{1}{n^{1/(2m)}}\right)$ and the convergence rate in Calder and García Trillos (2022) is $O\left(\frac{1}{n^{1/(m+4)}}\right)$. The extra sample complexity in our setting is induced by the additional mechanism that is needed to collect second order geometric information around the data in order to separate the underlying manifolds.

Remark 17 *From Theorem 15 we can see the quantitative effect of considering a non-zero ε_-. Indeed, when r is taken to be considerably smaller than ε_+, the number of faulty connections in the $\varepsilon_- \sim \varepsilon_+$ setting is much smaller than when $\varepsilon_- = 0$ because $\max\{\frac{r}{\varepsilon_+}, \varepsilon_+\}$ is a small quantity. Intuitively, removing points from a base ε-proximity graph should always reduce the number of faulty connections across different manifolds. However, what Theorem 15 states is that, when combined with the angle constraints , the ratio of faulty connections erased by removing an inner ball with volume comparable to that of the outer ball is actually quite significant. This result motivates the theoretical analysis that we present in the Appendix. In our numerical experiments we illustrate further the superior performance of our MMC algorithm when we set $\varepsilon_- \sim \varepsilon_+$. See an intuitive explanation in Figure 10.*

Remark 18 *It is not difficult to show that in general one can not relax the requirement that $\sin(\alpha) \ll 1$ in order to get sparsely outer connected graphs. Indeed, take for example two flats $\mathcal{M}_k$ and $\mathcal{M}_l$ with dimension 2 that meet perpendicularly at a straight line ℓ. If $\alpha \geq c > 0$ for constant c, it is straightforward to see that there is a small enough constant c_1 (depending on the lower bound for α) such that for all pairs of points $x \in \mathcal{M}_k$ and $y \in \mathcal{M}_l$ for which*

$$dist(x, \mathcal{M}_{kl}) \leq c_1\varepsilon_+, \quad dist(y, \mathcal{M}_{kl}) \leq c_1\varepsilon_+, \quad |x-y| < \varepsilon_+,$$

and for which the angle between $y-x$ and ℓ is smaller than c_1, there is an (α, r) constrained path between x and y. This situation is illustrated on the left panel of Figure 6. In turn, from this one can see that the number of connections that a point $x \in \mathcal{M}_k$ with $dist(x, \mathcal{M}_{kl}) \leq c_1\varepsilon_+$ has with points in $\mathcal{M}_l$ is $O(n\varepsilon_+^2)$ (i.e. the same order as with points in $\mathcal{M}_k$). In that case, $N_0 \sim n^2\varepsilon_+^{2+1}$ and thus $N_0/(n^2\varepsilon_+^{2+2}) \to \infty$.

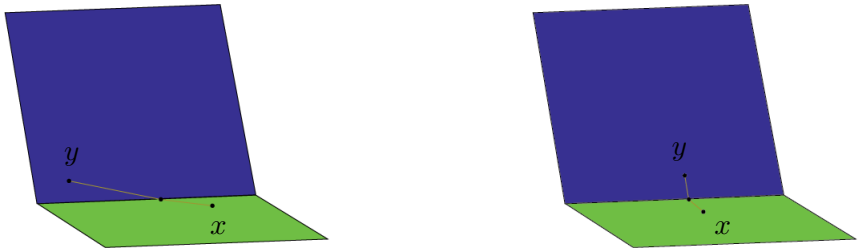

Figure 6: If $\alpha \geq c > 0$, the number of pairs of points in different manifolds that can be connected by paths that are almost tangential to the intersection of the manifolds (as on the left panel) is of the same order as the number of connections between pairs of points on the same manifold that are within distance ε_+ from the intersection of the manifolds. Notice that this situation may arise as soon as $m \geq 2$. For completeness, on the right panel we illustrate the benign situation of a pair of points on different manifolds that are close enough to each other and have no constrained path connecting them.

3.1 Proofs of Theorems 14 and 15

Proof [Proof of Theorem 14] Given that the manifold $\mathcal{M}_k$ is smooth and compact, and given that we only consider connecting two points $x, y \in X \cap \mathcal{M}_k$ when they are within a small distance ε_+ from each other, we can (and will) assume for simplicity that $\mathcal{M}_k$ is a flat of dimension m_k. As will become clear from our argument, the reduction to the flat case is sufficient as all curvature effects only introduce lower order corrections to our estimates.

Consider then the line segment connecting the points x and y and consider also the cylinder in $\mathcal{M}_k$ with axis given by the segment $y-x$ and circular base of radius h_1 centered at x and orthogonal to the segment xy; see Figure 7. We split this bigger cylinder into l parallel smaller cylinders with height h_2. The smaller cylinders are labeled as $\mathcal{C}_1, \ldots, \mathcal{C}_l$. By taking $h_2 = cr$ for small enough constant $(1/4) > c > 0$, $h_1 = \frac{c}{2}\tan(\alpha)r$, and assuming that $|x-y| > 4cr$, we can guarantee that

1. l is an odd number.
2. $4h_1^2 + 9h_2^2 \leq r^2$.
3. $\frac{2h_1}{h_2} \leq \tan(\alpha)$.

With this construction it is clear that if $X \cap \mathcal{C}_s \neq \emptyset$ for every even s, then we can construct an (α, r)-constrained path between x and y; see Figure 7. Notice that if on the other hand $|x-y| \leq 4cr \leq r$, then x and y can be connected directly.

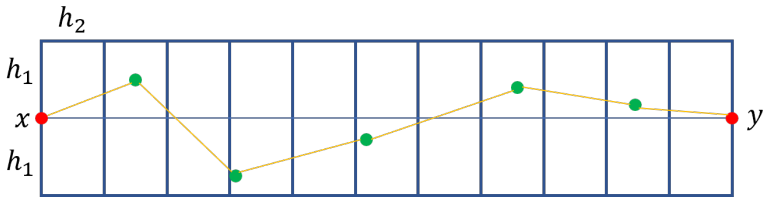

Figure 7: This is a valid path between x and y on the flat.

In the more interesting case $\varepsilon_+ \geq |x-y| > 4cr$, the probability that there is at least one sample from the n_k samples in $\mathcal{M}_k$ in all the cylinders $\mathcal{C}_s$ is no smaller than

$$1 - \frac{C_k \varepsilon_+}{r} \left(1 - C_k r^{m_k} (\tan\alpha)^{m_k - 1}\right)^{n_k} = 1 - \frac{C_k \varepsilon_+}{r} \exp\left(-C_k n_k r^{m_k} (\tan\alpha)^{m_k-1}\right).$$

By the discussion above, we conclude that the probability that there exists an (α, r)-constrained path between x and y is no smaller than the above quantity.

To bound from below the probability that *all* pairs of points $x, y \in X \cap \mathcal{M}_k$ that are within distance ε_+ are connected by (α, r)-constrained paths it is sufficient to take a union bound using the above estimates. ■

Next we study the outer connectivity condition for annular proximity graphs with angle constraints. We start with a result that applies for all choices of $\varepsilon_- \in [0, \frac{1}{4}\varepsilon_+]$ and then refine the estimates for the case $\varepsilon_- = c\varepsilon_+$.

Lemma 19 *Suppose that $\mathcal{M}_l \cap \mathcal{M}_k \neq \emptyset$, and let $x \in \mathcal{M}_l$ and $y \in \mathcal{M}_k$ be such that $|x-y| < \varepsilon_+$. Assume that $\operatorname{dist}(y, \mathcal{M}_l) \geq \operatorname{dist}(x, \mathcal{M}_k) > r$. If y is such that*

$$\frac{C_{k,l} \operatorname{dist}(y, \mathcal{M}_l)}{|x-y|} > \sin(\alpha),$$

then there is no (α, r)-constrained path between x and y; in the above, the constant $C_{k,l}$ depends on the manifolds $\mathcal{M}_k$ and $\mathcal{M}_l$ through their curvature, the quantity β from Assumption 1, and the curvature of $\mathcal{M}_{kl}$. In particular, if we set α to be such that $\sin(\alpha) \leq \frac{C_{k,l} r}{\varepsilon_+}$, then there is no (α, r)-constrained path between points x and y.

In addition, suppose that x, y are such that $\operatorname{dist}(y, \mathcal{M}_l) \geq r > \operatorname{dist}(x, \mathcal{M}_k)$, $\sin(\alpha) \leq \frac{C_{k,l} r}{\varepsilon_+}$ and $|x-y| \leq \varepsilon_+$. Then the first point of any (α, r)-constrained path connecting x and y starting from x must belong to $\mathcal{M}_k$.

Proof We denote the closest point in $\mathcal{M}_{kl}$ to y as O_y and the closest point in $\mathcal{M}_{kl}$ to x as O_x, respectively; uniqueness of these closest points, provided that ε_+ is small enough, is guaranteed by the discussion in Chapter 6 in Lee (2003). Let $\mathcal{TM}_l$ and $\mathcal{TM}_k$ be the tangent planes of $\mathcal{M}_l$ and $\mathcal{M}_k$ at O_x and O_y, respectively. Let x^*, y^* be the closest points from $x \in \mathcal{M}_l$ and $y \in \mathcal{M}_k$ to $\mathcal{TM}_l$ and $\mathcal{TM}_k$, respectively. See an illustration in Figure 8. In the remainder of this proof, we assume for the sake of contradiction that there is an (α, r)-constrained path between x and y.

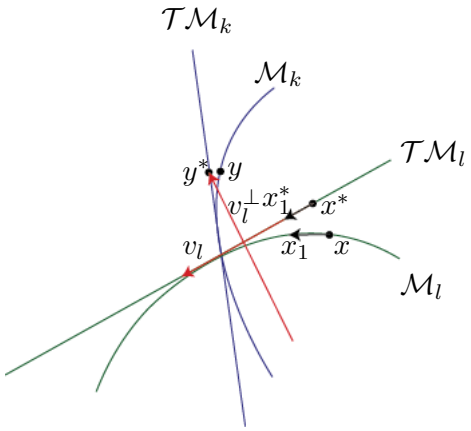

Figure 8: Illustration of proof for Lemma 19

We start by noticing that $y^* - x^*$ can be decomposed as

$$y^* - x^* = y_l^\perp v_l^\perp + y_l v_l, \tag{3.1}$$

for unit vectors $v_l^\perp$ and v_l, where $v_l^\perp$ is vertical to $\mathcal{T}\mathcal{M}_l$ and v_l is tangent to $\mathcal{T}\mathcal{M}_l$; $y_l^\perp$ and y_l are non-negative scalars. On the other hand, given the angle constraint between manifolds in the second part of Assumption 1, it is straightforward to show that

$$|x - O_x| = \text{dist}(x, \mathcal{M}_{kl}) \leq C\text{dist}(x, \mathcal{M}_k), \tag{3.2}$$

for a constant C that depends on β (this constant degenerates as β approaches $\pi/2$). A similar relation holds for $\text{dist}(y, O_y)$.

If there exists an (α, r)-constrained path connecting $x \in \mathcal{M}_l$ and $y \in \mathcal{M}_k$, then the first segment $\bar{x} = x_1 - x$ in this constrained path is such that $x_1 \in \mathcal{M}_l$, because $\text{dist}(x, \mathcal{M}_k) > r$. Denote the closest point to x_1 on $\mathcal{T}\mathcal{M}_l$ as x_1^* and in turn define $\bar{x}^* := x_1^* - x^*$. Notice that $\bar{x}$ satisfies

$$\bar{x} = \bar{x}^* + s_x v_x, \tag{3.3}$$

where v_x is a unit norm vector vertical to $\mathcal{T}\mathcal{M}_l$ and the scalar s_x satisfies

$$s_x \leq C|\bar{x}| \cdot |x - O_x| + C|\bar{x}|^2 \leq C|\bar{x}| \cdot \text{dist}(x, \mathcal{M}_k) + C|\bar{x}|^2 \tag{3.4}$$

because locally around the point O_x the manifold $\mathcal{M}_l$ can be represented by a quadratic function as illustrated in Figure 9. On the other hand, we have

$$y - x = (y - y^*) + (y^* - x^*) + (x^* - x), \tag{3.5}$$

where

$$|y - y^*| \leq C|y - O_y|^2 \leq C\text{dist}(y, \mathcal{M}_l)^2. \tag{3.6}$$

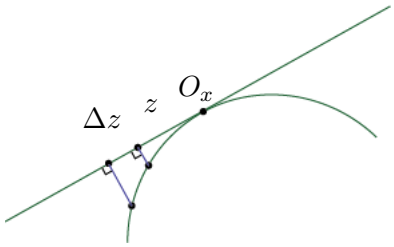

Figure 9: A manifold is locally the graph of a function f which can be approximated by a quadratic function around O_x, thus we have $|f(z)-f(z+\Delta z)| \approx |(z-O_x)^2-(z+\Delta z-O_x)^2| \leq C|z-O_x|\cdot|\Delta z| + C|\Delta z|^2$. Here we have illustrated the case where f is a scalar function, and the general case follows by using the scalar case componentwise.

Indeed, the first inequality follows from the quadratic approximation of the manifold around O_y, the fact that O_y is the closest point to y in $\mathcal{M}_{kl}$, and the angle constraint between manifolds in Assumption 1, which allows us to bound $\text{dist}(y, \mathcal{M}_{kl})$ by a constant times $\text{dist}(y, \mathcal{M}_l)$. Notice that $x^* - x \perp \bar{x}^*$ and that $|x - x^*|$ satisfies a similar relationship to $|y - y^*|$, namely

$$|x - x^*| \leq C|x - O_x|^2 \leq C\text{dist}(x, \mathcal{M}_k)^2. \tag{3.7}$$

Now, since v_x is vertical to $\mathcal{T}\mathcal{M}_l$, we have

$$\langle y - x^*, v_x\rangle = \langle y - y^*, v_x\rangle + \langle y^* - x^*, v_x\rangle \leq |y - y^*| + y_l^\perp \leq C\text{dist}(y, \mathcal{M}_l)^2 + y_l^\perp, \tag{3.8}$$

after using (3.1), $\langle v_l, v_x\rangle = 0$, and (3.6).

Since $\bar{x}$ is a segment in the (α, r)-constrained path, we must have

$$\cos(\alpha) \leq \frac{\langle y - x, \bar{x}\rangle}{|y - x|\cdot|\bar{x}|} = \frac{\langle y - x, \bar{x}^*\rangle}{|y - x|\cdot|\bar{x}|} + \frac{\langle y - x, s_x v_x\rangle}{|y - x|\cdot|\bar{x}|}. \tag{3.9}$$

Using the Cauchy-Schwartz inequality and combining (3.3), (3.5), and the fact that $\bar{x}$ is a segment in the (α, r)-constrained path, we deduce

$$\begin{aligned}\frac{\langle y-x,\bar{x}^*\rangle}{|y-x|\cdot|\bar{x}|} &= \frac{\langle y-y^*+(y^*-x^*)+(x^*-x),\bar{x}^*\rangle}{|y-x|\cdot|\bar{x}|}\\ &= \frac{\langle y-y^*,\bar{x}^*\rangle}{|y-x|\cdot|\bar{x}|}+\frac{\langle y^*-x^*,\bar{x}^*\rangle}{|y-x|\cdot|\bar{x}|}\\ &= \frac{\langle y-y^*,\bar{x}^*\rangle}{|y-x|\cdot|\bar{x}|}+\frac{\langle y_l v_l,\bar{x}^*\rangle}{|y-x|\cdot|\bar{x}|}\\ &\leq \frac{C\mathrm{dist}(y,\mathcal{M}_l)^2}{|y-x|}+\frac{\langle y_l v_l,\bar{x}^*\rangle}{|y-x|\cdot|\bar{x}|}\\ &\leq \frac{C\mathrm{dist}(y,\mathcal{M}_l)^2}{|y-x|}+\frac{|y_l|}{|y-x|}\end{aligned} \tag{3.10}$$

where the second equality is from $\langle x^*-x,\bar{x}^*\rangle=0$, the third equality is from $\langle v_l^\perp,\bar{x}^*\rangle=0$, and the inequalities are from $|\bar{x}^*|\leq|\bar{x}|$, Cauchy-Schwartz inequality, and (3.6). On the other hand, from Cauchy-Schwartz inequality,

$$\begin{aligned}\frac{\langle y-x,s_x v_x\rangle}{|y-x|\cdot|\bar{x}|} &= \frac{\langle y-x^*,s_x v_x\rangle}{|y-x|\cdot|\bar{x}|}+\frac{\langle x^*-x,s_x v_x\rangle}{|y-x|\cdot|\bar{x}|}\\ &\leq \frac{C\left(\mathrm{dist}(y,\mathcal{M}_l)^2+y_l^\perp\right)\cdot\left(\mathrm{dist}(x,\mathcal{M}_k)+|\overline{x}|\right)}{|y-x|}+\frac{C\mathrm{dist}(y,\mathcal{M}_l)^3}{|x-y|}\\ &\leq \frac{C\mathrm{dist}(y,\mathcal{M}_l)^3}{|y-x|}+\frac{Cy_l^\perp\cdot\mathrm{dist}(y,\mathcal{M}_l)}{|y-x|}\end{aligned} \tag{3.11}$$

where the first inequality is from (3.8), (3.7), the assumption $\mathrm{dist}(x,\mathcal{M}_k)\leq\mathrm{dist}(y,\mathcal{M}_l)$, the fact that $|\overline{x}|\leq r\leq\mathrm{dist}(y,\mathcal{M}_l)$, and (3.4); the second inequality is from $\mathrm{dist}(x,\mathcal{M}_k)\leq\mathrm{dist}(y,\mathcal{M}_l)$ and the fact that $|\overline{x}|\leq r\leq\mathrm{dist}(y,\mathcal{M}_l)$.

Therefore, combining (3.9), (3.10) and (3.11), we obtain

$$\cos(\alpha)\leq\frac{|y_l|}{|y-x|}+\frac{Cy_l^\perp}{|y-x|}\mathrm{dist}(y,\mathcal{M}_l)+\frac{C\mathrm{dist}(y,\mathcal{M}_l)^2}{|y-x|}. \tag{3.12}$$

Notice that we have used the fact that $\mathrm{dist}(y,\mathcal{M}_l)^2$ dominates $\mathrm{dist}(y,\mathcal{M}_l)^3$. This is the case because $\mathrm{dist}(y,\mathcal{M}_l)\leq\varepsilon_+\ll 1$.

From the Pythagorean theorem we have $y_l^2+(y_l^\perp)^2=|x^*-y^*|^2$. Therefore, we have

$$\begin{aligned}y_l^2 &= |x^*-y^*|^2-(y_l^\perp)^2\\ &\leq (|x-y|+C\mathrm{dist}(y,\mathcal{M}_l)^2)^2-(y_l^\perp)^2\end{aligned} \tag{3.13}$$

where the inequality is from (3.5), (3.6), (3.7), and the assumption that $\mathrm{dist}(y,\mathcal{M}_l)\geq\mathrm{dist}(x,\mathcal{M}_k)$. Also,

$$\begin{aligned}y_l^\perp &= \mathrm{dist}(y^*,\mathcal{T}\mathcal{M}_l)\geq\mathrm{dist}(y,\mathcal{T}\mathcal{M}_l)-|y-y^*|\\ &\geq \mathrm{dist}(y,\mathcal{M}_l)-|y-y^*|-C|x-y|^2\\ &\geq \mathrm{dist}(y,\mathcal{M}_l)-C|x-y|^2\end{aligned} \tag{3.14}$$

where the first inequality is from the triangle inequality; the third inequality is from (3.6); the second inequality is obtained as follows: denote the closest point to y in $\mathcal{TM}_l$ as $\mathcal{P}_{\mathcal{TM}_l}(y)$, and let $\mathcal{P}^{-1}_{\mathcal{TM}_l}(\mathcal{P}_{\mathcal{TM}_l}(y))$ be the point in $\mathcal{M}_l$ such that the closest point to $\mathcal{P}^{-1}_{\mathcal{TM}_l}(\mathcal{P}_{\mathcal{TM}_l}(y))$ in $\mathcal{TM}_l$ is $\mathcal{P}_{\mathcal{TM}_l}(y)$; the existence of the point $\mathcal{P}^{-1}_{\mathcal{TM}_l}(\mathcal{P}_{\mathcal{TM}_l}(y))$ follows from the local representation of $\mathcal{M}_l$ as a function of points in a neighborhood in $\mathcal{TM}_l$ around O_x. Then,

$$\begin{aligned}
\mathrm{dist}(y,\mathcal{M}_l) - \mathrm{dist}(y,\mathcal{TM}_l) &\le |y - \mathcal{P}^{-1}_{\mathcal{TM}_l}(\mathcal{P}_{\mathcal{TM}_l}(y))| - |y - \mathcal{P}_{\mathcal{TM}_l}(y)| \\
&\le |\mathcal{P}^{-1}_{\mathcal{TM}_l}(\mathcal{P}_{\mathcal{TM}_l}(y)) - \mathcal{P}_{\mathcal{TM}_l}(y)| \\
&\le C|\mathcal{P}_{\mathcal{TM}_l}(y) - O_x|^2 \\
&\le C|\mathcal{P}_{\mathcal{TM}_l}(y) - \mathcal{P}_{\mathcal{TM}_l}(x)|^2 + C|\mathcal{P}_{\mathcal{TM}_l}(x) - O_x|^2 \\
&\le C|y-x|^2 + C|x - O_x|^2 \\
&\le C|x-y|^2.
\end{aligned} \tag{3.15}$$

Similarly, one can derive an upper bound for $y_l^\perp$ of the form:

$$y_l^\perp \le \mathrm{dist}(y,\mathcal{M}_l) + C|x-y|^2. \tag{3.16}$$

Using the condition $1 \gg \varepsilon_+ \ge \mathrm{dist}(y,\mathcal{M}_l) \ge \mathrm{dist}(x,\mathcal{M}_k)$, we infer

$$\begin{aligned}
\sin^2(\alpha) &\ge 1 - \left(\frac{|y_l|}{|y-x|} + \frac{Cy_l^\perp}{|y-x|}\mathrm{dist}(y,\mathcal{M}_l) + \frac{C\mathrm{dist}(y,\mathcal{M}_l)^2}{|y-x|}\right)^2 \\
&\ge 1 - \frac{y_l^2}{|y-x|^2} - \frac{Cy_l y_l^\perp}{|y-x|^2}\mathrm{dist}(y,\mathcal{M}_l) - \frac{C\mathrm{dist}(y,\mathcal{M}_l)^2}{|y-x|} \\
&\ge \frac{(y_l^\perp)^2}{|x-y|^2} - \frac{Cy_l y_l^\perp}{|y-x|^2}\mathrm{dist}(y,\mathcal{M}_l) - \frac{C\mathrm{dist}(y,\mathcal{M}_l)^2}{|y-x|} \\
&\ge \frac{C\mathrm{dist}(y,\mathcal{M}_l)^2}{|y-x|^2} - C\frac{\mathrm{dist}(y,\mathcal{M}_l)^2}{|y-x|},
\end{aligned}$$

where the first inequality is from (3.12); we only keep the leading term in the second inequality; in the third inequality we use (3.13) and $\mathrm{dist}(x,\mathcal{M}_k) \le \mathrm{dist}(y,\mathcal{M}_l)$, and in the last inequality we use (3.14), (3.16), and $\mathrm{dist}(y,\mathcal{M}_l) \le |x-y| \ll 1$. Noticing that $|y-x| \le \varepsilon_+ \ll 1$, we can further simplify the above inequality as

$$\sin(\alpha) \ge \frac{C\mathrm{dist}(y,\mathcal{M}_l)}{|y-x|}.$$

As a consequence, if the above relationship is not satisfied, there can not exist an (α, r)-constrained path between x and y. This completes the proof of the first part.

For the second part, we assume for the sake of contradiction that there is an (α, r)-constrained path between x and y such that the first step (starting from x) in the path belongs to $\mathcal{M}_l$; we call this first step x_1. Since the only condition used for $\mathrm{dist}(x,\mathcal{M}_k)$ in the first part is that $\mathrm{dist}(x,\mathcal{M}_k) \le \mathrm{dist}(y,\mathcal{M}_l)$, and this also holds for $\mathrm{dist}(x,\mathcal{M}_k) < r \le \mathrm{dist}(y,\mathcal{M}_l)$, we can then repeat the same argument as above with $x' = x_1 - x$ to conclude that if $\sin(\alpha) < \frac{C\mathrm{dist}(y,\mathcal{M}_l)}{|y-x|}$, then we would reach a contradiction.

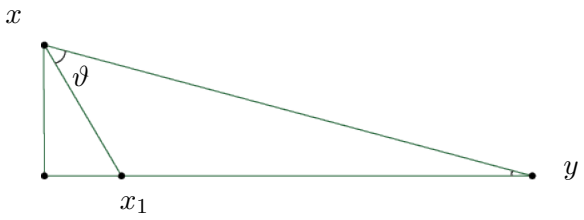

Figure 10: When $\text{dist}(x, \mathcal{M}_k) \leq \text{dist}(y, \mathcal{M}_l)$ and the first step x_1 of a constrained path is on $\mathcal{M}_k$, the angle $\angle x_1 x y$ is larger than the angle $\angle x_1 y x$. This means that the value of $\angle x_1 x y$ dictates whether the path satisfies the angle constraints or not. In turn, we see that for points x, y with a larger value of $|x-y|$ this angle will be larger than when $|x-y|$ is smaller, making it easier for the angle condition to detect that x, y are in different manifolds when their distance is larger.

■

The next lemma helps us justify why, for the multi-manifold clustering problem, choosing ε_- of the same order as ε_+ is better than choosing $\varepsilon_- = 0$ (or in general ε_- much smaller than ε_+). Intuitively, as illustrated in Figure 10, by directly omitting connections between points that are too close to each other we can remove edges between points on different manifolds that the angle constraint condition may not be able to remove. We quantify the gain of considering this step in the next lemma.

Lemma 20 *Suppose that $\mathcal{M}_l \cap \mathcal{M}_k \neq \emptyset$ and let $x \in \mathcal{M}_l$ and $y \in \mathcal{M}_k$ be such that $\varepsilon_- < |x-y| < \varepsilon_+$, where $\varepsilon_- = c\varepsilon_+$. Let $\alpha > 0$ be such that*

$$\sin(\alpha) < \frac{Cr}{\varepsilon_+}.$$

If $\text{dist}(x, \mathcal{M}_k) > C(\frac{r^2}{\varepsilon_+} + r\varepsilon_+)$ and $\text{dist}(y, \mathcal{M}_l) > r$, then there can not exist an (α, r)-constrained path between x and y.

Proof If $\text{dist}(x, \mathcal{M}_k) > r$, then by Lemma 19 there can not exist an (α, r)-constrained path between x, y. Thus, without the loss of generality we can assume that $\text{dist}(x, \mathcal{M}_k) \leq r$.

Assume for the sake of contradiction that there is an (α, r)-constrained path between x and y. Denote the closest point to x in $\mathcal{M}_{kl}$ as O_x, and let $\mathcal{T}\mathcal{M}_k$ be the tangent plane to $\mathcal{M}_k$ at O_x (notice that this definition is different from the one in Lemma 19). Let y^* be the closest point to y in $\mathcal{T}\mathcal{M}_k$, and let z be the closest point to x in $\mathcal{T}\mathcal{M}_k$. Let $t := |x-z|$ and $d := |y^* - x|$; see an illustration in Figure 11. Notice that we have

$$|y^* - x| \geq |y - x| - |y^* - y| > \varepsilon_- - C\varepsilon_+^2 \geq C\varepsilon_+, \tag{3.17}$$

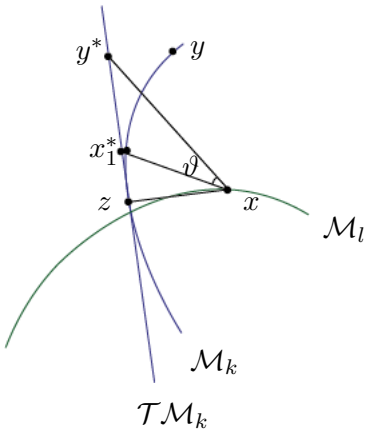

Figure 11: Illustration of proof for Lemma 20

due to the fact that

$$\begin{aligned} |y-y^*| &\le C|y^*-O_x|^2 \le C|y^*-z|^2 + C|z-O_x|^2 \\ &\le C|x-y|^2 + C|x-O_x|^2 \\ &= C|x-y|^2 + C\mathrm{dist}(x,\mathcal{M}_{kl})^2 \\ &\le C|x-y|^2 + C\mathrm{dist}(x,\mathcal{M}_k)^2 \\ &\le C|x-y|^2 \le C\varepsilon_+^2. \end{aligned} \tag{3.18}$$

By the second part of Lemma 19 we know that the first step in the constrained path (starting from x) must lie in $\mathcal{M}_k$; we denote by x_1 this first step and let x_1^* be the closest point to x_1 in $\mathcal{T}\mathcal{M}_k$. We have

$$t = |x-z| = \mathrm{dist}(x,\mathcal{T}\mathcal{M}_k) \le |x-x_1^*| \le |x-x_1| + |x_1-x_1^*| \le r + Cr^2 \le Cr. \tag{3.19}$$

In fact, $|x_1-x_1^*|$ can be bounded by $C|x_1-x|^2$, as it follows from the next computation:

$$\begin{aligned} |x_1-x_1^*| \le C|x_1-O_x|^2 &\le C|x-x_1|^2 + C|x-O_x|^2 \\ &= C|x-x_1|^2 + C\mathrm{dist}(x,\mathcal{M}_{kl})^2 \\ &\le C|x-x_1|^2 + C\mathrm{dist}(x,\mathcal{M}_k)^2 \\ &\le C|x-x_1|^2. \end{aligned} \tag{3.20}$$

In particular, we also have

$$|x_1-x_1^*| \le Cr|x-x_1|. \tag{3.21}$$

We will also use the following inequality:

$$\begin{aligned} |x-x_1^*|^2 = |x_1^*-z|^2 + |x-z|^2 &\le |x_1-z|^2 + 2|x-x_1|^2 + 2|x_1-z|^2 \\ &\le 2|x-x_1|^2 + 3|x_1-x_1^*|^2 + 3|x_1^*-z|^2 \le C|x-x_1|^2 + 3|x_1-x_1^*|^2 \\ &\le C|x-x_1|^2 + C|x-x_1|^4 \le C|x-x_1|^2, \end{aligned} \tag{3.22}$$

where the second to last inequality follows from (3.19) and (3.20).

The angle condition for the constrained path then gives

$$\begin{aligned}
\cos\alpha &\leq \frac{\langle y-x, x_1-x\rangle}{|y-x|\cdot|x_1-x|} \\
&= \frac{\langle y^*-x, x_1^*-x\rangle}{|y-x|\cdot|x_1-x|} + \frac{\langle y-y^*, x_1^*-x\rangle}{|y-x|\cdot|x_1-x|} + \frac{\langle y^*-x, x_1-x_1^*\rangle}{|y-x|\cdot|x_1-x|} + \frac{\langle y-y^*, x_1-x_1^*\rangle}{|y-x|\cdot|x_1-x|} \\
&\leq \frac{|y^*-x|\cdot|x_1^*-x|}{|y-x|\cdot|x_1-x|}\cos\vartheta + \frac{\langle y-y^*, z-x\rangle}{|y-x|\cdot|x_1-x|} + Cr \\
&\leq \frac{|y^*-x|\cdot|x_1^*-x|}{|y-x|\cdot|x_1-x|}\cos\vartheta + C\frac{|y-x|t}{|x-x_1^*|} + Cr \\
&\leq \frac{d}{|y-x|}\cos\vartheta + C\frac{|y-x|t}{|x-x_1^*|} + Cr,
\end{aligned} \tag{3.23}$$

where ϑ is the angle between the vectors y^*-x and x_1^*-x as illustrated in Figure 11. The second inequality is from the fact that $\langle z-x_1^*, y-y^*\rangle = 0$, (3.18), and (3.21). The third inequality is from (3.18) and (3.21); the last inequality is from the following:

$$\begin{aligned}
\left|\frac{|y^*-x|\cdot|x_1^*-x|}{|y-x|\cdot|x_1-x|} - \frac{d}{|x-y|}\right| &= \left|\frac{|y^*-x|(|x_1^*-x|-|x_1-x|)}{|x-y|\cdot|x_1-x|}\right| \\
&\leq \frac{|y^*-x|\cdot|x_1-x_1^*|}{|x-y|\cdot|x_1-x|} \\
&\leq \frac{(|y-x|+|y-y^*|)\,|x_1-x_1^*|}{|x-y|\cdot|x_1-x|} \\
&\leq \frac{C\,(|y-x|+|y-y^*|)\,r}{|x-y|} \leq Cr.
\end{aligned}$$

In turn, we have

$$d-|x-y| \leq |z-y^*|+|z-x|-|x-y| \leq t. \tag{3.24}$$

Combining (3.23) and (3.24), we conclude that

$$\cos\alpha \leq \cos\vartheta + \frac{Ct}{|y-x|} + C\frac{|y-x|t}{|x-x_1^*|} + Cr.$$

This implies

$$\sin^2\alpha \geq \sin^2\vartheta - \frac{Ct}{|y-x|} - C\frac{|y-x|t}{|x-x_1^*|} - Cr, \tag{3.25}$$

where we drop some lower order terms.

Since $s - Cs^3 \leq \sin(s) \leq s$ when $s \geq 0$ is small enough, we can write (3.25) as,

$$\vartheta - C\vartheta^3 \leq \sqrt{\alpha^2 + \frac{Ct}{|y-x|} + C\frac{|y-x|t}{|x-x_1^*|} + Cr}$$

Let us denote $\sqrt{\alpha^2 + \frac{Ct}{|y-x|} + C\frac{|y-x|t}{|x-x_1^*|} + Cr}$ by α_0.

From a simple geometric observation (just consider the triangles $\triangle xy^*z$ and $\triangle x_1^* xz$), we have

$$\alpha_0 + C\vartheta^3 > \vartheta \geq \arccos\frac{t}{d} - \arccos\frac{t}{|x-x_1^*|} =: f(d,t).$$

When fixing $|x-x_1^*|$, the function $f(d,t)$ is strictly increasing in both coordinates because $\partial_d f(d,t) = \frac{t}{d^2}\cdot\frac{1}{\sqrt{1-\frac{t^2}{d^2}}} > 0$ and $\partial_t f(d,t) = \frac{1}{\sqrt{|x-x_1^*|^2-t^2}} - \frac{1}{\sqrt{d^2-t^2}} > 0$. In particular, $f(d,t) \geq f(C\varepsilon_+, t)$ because of (3.17). We see that if $t > t_0$, where

$$t_0 := \frac{C|x-x_1^*|\varepsilon_+\sin(\alpha_0 + C\vartheta^3)}{\sqrt{(\varepsilon_- - C\varepsilon_+^2)^2 - 2C|x-x_1^*|\varepsilon_+\cos(\alpha_0+C\vartheta^3) + |x-x_1^*|^2}},$$

then

$$\alpha_0 + C\vartheta^3 \geq f(d,t) \geq f(C\varepsilon_+, t) > f(C\varepsilon_+, t_0) = \alpha_0 + C\vartheta^3,$$

and thus we would reach a contradiction. A simpler upper bound for t_0 is the following:

$$t_0 < \frac{C|x-x_1^*|\varepsilon_+\sin(\alpha_0)}{\varepsilon_+\cos(\alpha_0) - |x-x_1^*|} \leq C|x-x_1^*|\sin(\alpha_0).$$

In particular, if $t > C|x-x_1^*|\sin(\alpha_0)$, then there can not be an (α, r)-constrained path between y and x. We can rewrite this condition as

$$t > C|x-x_1^*|(\frac{r}{\varepsilon_+} + \frac{\sqrt{t}}{\sqrt{|y-x|}} + \frac{\sqrt{|y-x|}\sqrt{t}}{\sqrt{|x-x_1^*|}} + \sqrt{r}) \tag{3.26}$$

where we have used the assumption $\sin\alpha \leq C\frac{r}{\varepsilon_+}$. Notice that the right hand side of the inequality is an increasing function with respect to $|x-x_1^*|$, so we can replace $|x-x_1^*|$ with the upper bound Cr as in (3.22). Also, using $\varepsilon_- \leq |y-x| \leq \varepsilon_+$ and $\varepsilon_- = c\varepsilon_+$, we can change (3.26) to

$$t > Cr(\frac{r}{\varepsilon_+} + \frac{\sqrt{t}}{\sqrt{\varepsilon_+}} + \frac{\sqrt{t\varepsilon_+}}{\sqrt{r}} + \sqrt{r}). \tag{3.27}$$

This in turn can be changed to

$$t > C(\frac{r^2}{\varepsilon_+} + r\varepsilon_+), \tag{3.28}$$

after using the fact that $\frac{r^2}{\varepsilon_+} + r\varepsilon_+ \geq 2r\sqrt{r}$. In summary, what we have shown is that if $t = \mathrm{dist}(x, \mathcal{T}\mathcal{M}_k) > C(\frac{r^2}{\varepsilon_+} + r\varepsilon_+)$, then there can not be a constrained path between x and y.

To finalize the proof, we must now find a condition on $\mathrm{dist}(x, \mathcal{M}_k)$ that implies (3.28). For this purpose, we use notation analogous to the one in (3.15) and compute

$$\begin{aligned}\mathrm{dist}(x,\mathcal{M}_k) - \mathrm{dist}(x,\mathcal{T}\mathcal{M}_k) &\leq |x - \mathcal{P}^{-1}_{\mathcal{T}\mathcal{M}_k}(z)| - |x-z| \\ &\leq |\mathcal{P}^{-1}_{\mathcal{T}\mathcal{M}_k}(z) - z| \leq C|z-O_x|^2 \leq Cr^2,\end{aligned}$$

where the last inequality is because $|z - O_x| \leq |x - O_x| \leq C\text{dist}(x, \mathcal{M}_k) \leq Cr$. Given that $r^2 \ll r\varepsilon_+$, we conclude that

$$\text{dist}(x, \mathcal{M}_k) > C(\frac{r^2}{\varepsilon_+} + r\varepsilon_+)$$

implies (3.28), completing in this way the proof.

■

With Theorem 14 and Lemmas 19 and 20 in hand we can now prove the main result in this section.

Proof [Proof Theorem 15]

1. For arbitrary $\varepsilon_- \leq (1/4)\varepsilon_+$ we can use Lemma 19 and a standard concentration bound to see that with probability no less than

$$1 - C_{k,l} n \exp(-C_{k,l} n \min\{r^{m_k - m_{kl}}, r^{m_l - m_{kl}}, \varepsilon_+^{m_k}, \varepsilon_+^{m_l}\},$$

N_{kl}, the number of connections between points in $X \cap \mathcal{M}_k$ and $X \cap \mathcal{M}_l$, satisfies

$$N_{kl} \leq C_{k,l} n^2 \max\{r^{m_k - m_{kl}} \varepsilon_+^{m_l}, r^{m_l - m_{kl}} \varepsilon_+^{m_k}\}.$$

2. When we have the lower bound $\varepsilon_- \geq c\varepsilon_+$ we can proceed as above but now using Lemma 20.

■

Remark 21 *According to Theorem 15, to satisfy the sparse outer connectivity when $\varepsilon_- = c\varepsilon_+$ one needs*

$$\frac{n^2 \max\{(\frac{r^2}{\varepsilon_+} + r\varepsilon_+)^{m_k - m_{kl}} \varepsilon_+^{m_l}, (\frac{r^2}{\varepsilon_+} + r\varepsilon_+)^{m_k - m_{kl}} \varepsilon_+^{m_k}\}}{n^2 \varepsilon_+^{m+2}} \to 0. \tag{3.29}$$

In the worst case for (3.29)*, one requires*

$$r \ll \min\{\varepsilon_+^{m+1-\min_l m_l}, \varepsilon_+^{\frac{m+3-\min_l m_l}{2}}\}, \tag{3.30}$$

which is a quite restrictive condition when there is a large discrepancy between the dimensions of the manifolds, as one would require a very small value of r and in turn a very large number of data points for condition (3.30) *to be satisfied while simultaneously satisfying the inner connectivity condition. The above estimate, however, is quite pessimistic, and in particular assumes that all manifolds intersect with each other. In general, one can replace $\min_l m_l$ with the minimum dimension of manifolds that actually intersect the manifolds with larger dimension. If the gap between m and this restricted minimum is not too large, from moderate number of samples we would expect the spectral clustering algorithm with path constraints to be able to separate the data coming from manifolds with larger dimension from the rest of the data set. At that stage, one can consider a new iteration of the algorithm, this time with a data set with fewer points and with a smaller largest dimension. The exploration of this iterative pruning strategy is beyond the scope of this work.*

3.2 A local PCA approach to MMC

An alternative spectral approach to the multimanifold clustering problem that is popular in the literature (e.g. see Arias-Castro et al. (2017)) is based on building weights ω_{ij} that depend on the level of alignment of local tangent planes around nearby data points. To be precise, as in the path-construction from section 3 we only consider giving an edge to a pair of data points x_i, x_j if $\varepsilon_- < |x_i - x_j| < \varepsilon_+$. If this condition is satisfied, we then set $\omega_{x_i x_j} = 1$ provided that the angle between $\hat{T}_{x_i}$ (a local tangent plane around x_i) and $\hat{T}_{x_j}$ is smaller than a certain threshold, and otherwise we set $\omega_{x_i x_j} = 0$. These local "tangent" planes can be constructed from the observed data using local PCA. Namely, the idea is to run PCA with the data set $X \cap B(x_i, r)$ for some small enough r in order to obtain a collection of principal directions which are then used as generators for the plane $\hat{T}(x_i)$; see Arias-Castro et al. (2017) for more details.

Using the estimates from Arias-Castro et al. (2017) (and some additional computations) it is possible to show that the local PCA graph construction satisfies the sparse outer connectivity condition (with very high probability), provided that the parameter r is tuned appropriately. However, from a theoretical perspective, one should not expect that the full inner connectivity holds with high probability. This is an observation already made in Arias-Castro et al. (2017) (although not with the exact same words). Indeed, let x_i be a point in the manifold $\mathcal{M}_l$ that is close to the intersection of $\mathcal{M}_l$ and $\mathcal{M}_k$ (closer than r). For points $x_j \in \mathcal{M}_l$ within distance ε_+ from x_i and away enough from the intersection $\mathcal{M}_k \cap \mathcal{M}_l$, we expect their PCA-based tangent planes to resemble those of the actual manifold $\mathcal{M}_l$ at those same points (if r has been chosen so that there is consistency in the approximation of tangent planes). However, x_i's empirical tangent plane will be influenced by the presence of the points in $\mathcal{M}_k$ that belong to the ball $B(x_i, r)$ and thus one should not expect this plane to be aligned with the planes of all the other points $x_j \in \mathcal{M}_l$ lying nearby. In contrast, the full inner connectivity for the path-based graph construction from section 3 just depends on the points on each single manifold: having additional points can only help with the full inner connectivity (more points means more possible paths) but never tamper with it.

One of the implications of the above discussion is that the local PCA approach to MMC may in principle produce more clusters than desirable, and for example groups of points that lie close to the intersection of two manifolds may form their own clusters; see the discussion in section 3 in Arias-Castro et al. (2017). It is thus possible that some of the manifolds get split into different components and in particular one may not be able to recover the multi-manifold structure underlying the data without information on the actual location of the manifolds' intersections.

3.3 Computational Complexity of Annular graphs with Angle Constraints

Here we discuss the theoretical computational complexity of building angle-constrained path proximity graphs in their *nearest neighbor version*, where we can more directly quantify the contributions of the different steps in the construction. In this version we substitute all parameters that have a lengthscale interpretation with parameters that specify the number of neighbors to a point. Precisely, instead of fixing the two length scales $\varepsilon_+, \varepsilon_-$, we can alternatively fix two natural numbers $k_+ > k_-$ and substitute the conditions $|y_1 - y_2| > \varepsilon_+$ and $|y_1 - y_2| < \varepsilon_-$ in Algorithm 1 with the conditions "neither y_1 is one of the k_+-nearest

neighbors of y_2, nor y_2 is one of the k_+-nearest neighbors of y_1" and "y_1 is one of the k_- nearest neighbors of y_2 or viceversa", respectively. Likewise, the lengthscale r is substituted with a parameter $\kappa \in \mathbb{N}$, and the second condition in (12) is changed to "x_{i_j} is one of the κ nearest neighbors of $x_{i_{j+1}}$ or viceversa".

Let $|V|$ be the number of neighbors around a point in the base proximity graph, and let $|E|$ be the number of edges among these neighbors. The computational complexity of Algorithm 1 is $\mathcal{O}((|V| + |E|) \log |V|)$, which is essentially the same as the complexity of Dijkstra's algorithm using the Fibonacci heap Fredman and Tarjan (1987). Therefore, the total computational complexity of constructing k_+, k_--graph with angle constraints is $\mathcal{O}\left(n(k_+ - k_-)\kappa k_+ \log(\kappa k_+) + n^2 \log k_+\right)$, where $\mathcal{O}(n^2 \log k_+)$ is the computational cost of constructing the k_+, k_--nearest neighbor base graph. By using the adapted graph construction in section 4.2.3, it is possible to speed up the construction to $\widetilde{\mathcal{O}}\left(n\kappa k_+ + n^2 \log k_+\right)$. If the parameters are chosen as suggested in Remark 16 for the setting of manifolds with the same dimension m, i.e. we use $v_m \varepsilon_+^m = k_+$, $v_m \varepsilon_-^m = k_-$, and $v_m r^m = \kappa$, then the computational complexity of the adapted method and Algorithm 1 are, in terms of the total number of data points, $\widetilde{\mathcal{O}}\left(n^{\frac{6m-3}{3m-1}} + n^2\right)$ and $\widetilde{\mathcal{O}}\left(n^{\frac{8m-4}{3m-1}}\right)$, respectively. In contrast, the computational complexity of building a vanilla k-nearest neighbor graph is $\mathcal{O}(n^2 \log k)$ by using a priority queue structure. Therefore, by using the adapted structure for the angle-constrained path construction, we can build graphs at the same computational complexity as the one for vanilla k-nearest neighbor graphs. It is worth highlighting that once the similarity matrix has been constructed, the computational complexity for the eigendecomposition needed to run spectral clustering will typically depend on the level of sparsity of the input weight matrix ω. In this regard, it is important to notice that the angle-constrained graph will always be sparser than its base proximity graph. Finally, we remark that our algorithm may not be as efficient in practice as the previous theoretical analysis would suggest because, in general, we need to use denser graphs than if no curvature constraints were imposed. In this regard, the use of landmark points in the algorithm by Babaeian et al. (2015) is an alternative to speed up the computation, although at the expense of weaker theoretical consistency guarantees.

4. Numerical experiments

The purpose of this section is twofold. On the one hand, we want to explore the limitations and difficulties that may arise when using the MMC methodologies based on spectral clustering with path-based graphs that we have introduced in section 3. On the other hand, we want to provide further insights into the theoretical results that we have presented throughout the paper. We present a series of numerical experiments aimed at achieving these two goals. In addition, at the end of this section we compare the performance of spectral clustering using path-based graphs with other spectral-based algorithms by testing them on synthetic and real data sets.

In our experiments, we consider our graph construction directly as presented in section 3, or in its nearest neighbor version as discussed in section 3.3. Unless otherwise noted, whenever we use the nearest neighbor version of our algorithm we will select $k_- = 2/3k_+$ and tune k_+ in order to minimize the misclustering rate of the output clusters. In the toy examples where we use the $(\varepsilon_+, \varepsilon_-)$ version of our algorithm, we tune ε_+ and ε_- so that

$v_m\varepsilon_+^m =: k_+ \in \mathbb{N}$ and $v_m\varepsilon_-^m =: k_- \in \mathbb{N}$, where v_m is the volume of the unit ball in $\mathbb{R}^m$. The other parameters in the algorithm, α and κ (or r), are tuned to minimize the misclustering rate. [1]

4.1 Bottlenecks and multiple manifolds

Our theoretical results imply that the spectra of suitable graphs resemble the spectrum of a tensorized Laplacian on the union of smooth manifolds underlying the data set. In particular, when using spectral clustering on finite data sets with a multi-manifold structure, it is possible to obtain a partition of the data into multiple smooth manifolds and/or into regions that are separated by thin bottlenecks. In this section we explore numerically the "confounding" role that bottlenecks may play in MMC.

First, let us consider the bottle and plane example illustrated in Figures 12, 13, and 14. There, data set X is sampled uniformly from the set $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$, where $\mathcal{M}_1$ is a plane and $\mathcal{M}_2$ is a 2-dimensional dumbbell with a bottleneck at its center. A graph (X, ω) has been constructed as in section 3 for appropriate values of $\varepsilon_-, \varepsilon_+, r, \alpha$. Intuitively, this graph should help identify the two manifolds given that they meet perpendicularly (i.e., β is zero in (2.2)). On the other hand, the same graph captures the internal geometry of $\mathcal{M}_2$ and thus should also detect the bottleneck in $\mathcal{M}_2$. Figure 12 shows the sign of the first non-trivial eigenvector of the graph Laplacian, which, as we can observe from the picture, is able to detect the bottleneck. Figure 13 shows the sign of the second non-trivial eigenvector (orthogonal to the first non-trivial eigenvector). In our experiments, our graph Laplacian's first two non-zero eigenvalues are close to zero, and their relative difference is quite small compared to the relative difference between the second and third non-zero eigenvalues. The partition illustrated in Figure 13 is not directly interpretable. However, when considering a suitable linear combination of the first and second non-trivial eigenvectors, we recover the partition illustrated in Figure 14 which correctly separates the two manifolds. This linear combination is obtained by minimizing the *Ratio cut* functional (see Von Luxburg (2007) for a definition) among all the partitions induced by norm one linear combination of the first two non-trivial eigenvectors. In this case, it is a simple one-dimensional search.

This example illustrates that bottlenecks are indeed confounders for MMC when using spectral methods. Still, even in the presence of competitor bottlenecks, we see that the graph Laplacian's spectrum possesses the information needed to recover the desired partition of the data, and the combination of spectral clustering with Ratio cut minimization is shown to help in the detection of the desired partition. Warm start initialization for balanced cut minimization using spectral clustering has been considered in the literature before (e.g. Bresson and Laurent (2012); Bresson et al. (2013a,b, 2012a,b)).

Another example where multiple manifolds and bottlenecks are present is the one illustrated in Figures 15 and 16 which we will refer to as the *dollar sign* example. We again build the graph Laplacian as in section 3. Figure 15 shows the sign of the first non-trivial eigenvector, which, as we can observe, can detect the "bottleneck" at the center of the dollar sign shape. Figure 16 shows the sign of the second non-trivial eigenvector. Notice that the multi-manifold structure is identified correctly using this eigenvector. In this exam-

1. The implementation of our algorithm can be found in github.com/chl781/manifold-clustering

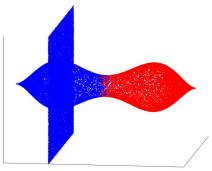

Figure 12: Ratio cut: 0.129

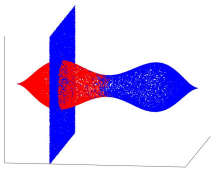

Figure 13: Ratio cut: 0.240

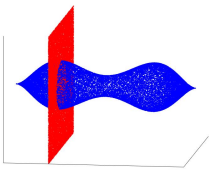

Figure 14: Ratio cut: 0.083

ple, the partition induced by the second non-trivial eigenvector is a minimizer of the Ratio cut functional among partitions induced by linear combinations of the first two non-trivial eigenvectors. For comparison, in Figures 17 and 18 we illustrate the partitions induced by the first and second non-trivial eigenvectors of a graph Laplacian on a standard ε-graph with no path constraints. We can see that with that graph construction we can not retrieve the desired multimanifold structure.

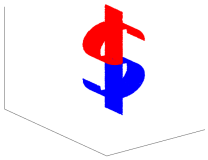

Figure 15

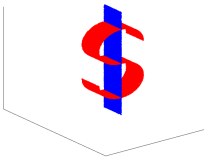

Figure 16

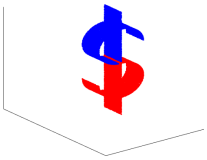

Figure 17

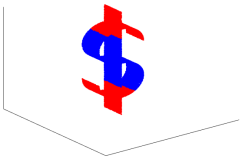

Figure 18

In Figures 15 and 16 2nd and 3th eigenvectors from annular graph with angle constraints. In Figures 17 and 18 2nd and 3th eigenvectors from standard ε-proximity graph.

Remark 22 *When using ratio cut for MMC the important energy to consider is the cut functional/total variation functional:*

$$TV_n(u) \sim \frac{1}{n^2\varepsilon_+^{m+1}} \sum_{i,j} \omega_{ij}|u(x_i) - u(x_j)|.$$

Notice that the correct scaling factor $\frac{1}{n^2\varepsilon_+^{m+1}}$ *for* TV_n *is different from the one for the graph Dirichlet energy which scales like* $\frac{1}{n^2\varepsilon_+^{m+2}}$ *(see* (A.2) *in the Appendix). In general, this discrepancy in scaling factors explains the superior performance of Ratio cut over spectral clustering on MMC problems. To see this, we return to the discussion in Remark 18, and notice that in order to create a very cheap balanced cut that captures the multi-manifold structure underlying the data it is sufficient to choose the angle* α *to be small enough without making it arbitrarily small.*

While in general ratio cut minimization is expected to be superior to spectral clustering from a theoretical perspective, it is at the algorithmic level that ratio cut is less appealing.

In general, using spectral clustering as a warm start for ratio cut minimization is a reasonable strategy to consider as we have illustrated in the dumbbell-and-plane and dollar sign examples.

4.2 Comparison of different proximity graphs

We consider the setting illustrated in Figures 19 and 20 where we have generated 6500 points on the horizontal line and 2000 points on the vertical line, i.e., an uneven setting. We can see that when we add angle constraints to a k_+, k_--NN base graph, we do not recover the two lines as we do when we use the $\varepsilon_+, \varepsilon_-$-graph setting. The outcomes illustrated here are markedly different from the ones in the even case where the NN approach provides more stable results, as is the case with vanilla spectral clustering using a standard k-NN graph. The reason is that around the intersection of the two lines, a k-NN neighborhood of a point in the vertical line mostly picks points in the horizontal line (since there is a higher density of points there), which results in very few connections with other points on the vertical line.

In general, we expect the NN setting to struggle in settings where the densities of points are different around the intersections of the manifolds, as illustrated by the "inclusion problem" in the setting from Figures 21 and 22. In Figure 21 we illustrate the clusters output by spectral clustering in the path constrained NN setting. We see that the part of the plane contained inside the cone has been merged with the cone despite the fact that a strong angle constraint was used. The points in the inner part of the plane around the boundary have many more nearest neighbors in the cone than in the external portion of the plane, thus effectively discarding connections that would otherwise keep the plane better connected (as it is the case with the $\varepsilon_+, \varepsilon_-$ construction as shown in Figure 22).

We remark that the effect of density in clustering can be reduced by considering suitable normalized versions of proximity graphs as in Coifman and Lafon (2006), where in particular one can take the random walk Laplacian associated to a new set of weights ω_{ij}^{α} of the form:

$$\omega_{ij}^{\alpha} = \frac{\omega_{ij}}{d_i^{\alpha} d_j^{\alpha}},$$

for some $\alpha \in (0, 1]$, where d_i and d_j are the degrees of x_i and x_j relative to the original weight matrix ω. It is possible to show that with the choice $\alpha = 1$ one can effectively remove the effect of density in clustering. We notice that in the multi-manifold clustering setting, when manifolds have different dimensions, the role of density is more severe than when manifolds have the same dimension. This is because we are assuming that the number of points in each manifold is roughly the same, and so, densities on smaller dimensional objects tend to be considerably larger than densities on larger dimensional objects. The use of appropriate normalized Laplacians may thus help considerably with multi-manifold clustering problems.

4.2.1 Role of ε_- in annular graphs

Here we illustrate the effect of ε_- in the performance of spectral clustering. We focus on two possible choices: $\varepsilon_- = 0$ Vs $\varepsilon_- \sim \varepsilon_+$. We consider points uniformly sampled from two

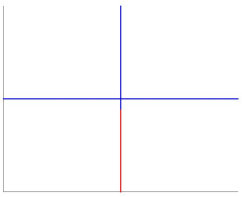

Figure 19: k_+, k_- NN with angle constraint

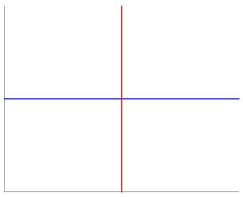

Figure 20: $\varepsilon_+, \varepsilon_-$ graph with angle constraint

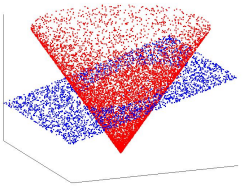

Figure 21: k_+, k_- NN setting with angle constraint

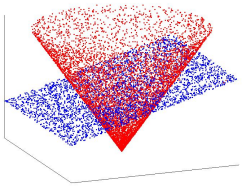

Figure 22: $\varepsilon_+, \varepsilon_-$ graph with angle constraint

intersecting 2-dimensional spheres with radius 1 and distance between their centers equal to 0.6 as illustrated in Figures 24 and 23. We can observe the discrepancy between the clusters obtained in both settings and how when we set $\varepsilon_- \sim \varepsilon_+$, the two manifolds are correctly identified. ε_+ is the same in both cases.

Notice that our theory shows that, in principle, any choice of ε_- (not too close to ε_+) can provide correct identification of the manifolds as long as the number of samples is large enough. On the other hand, our theory also suggests that a non-zero ε_- can reduce the error of approximation (see Remark 17) as more faulty connections can be removed between points that are too close to the intersection. Our numerical experiments complement our theoretical findings.

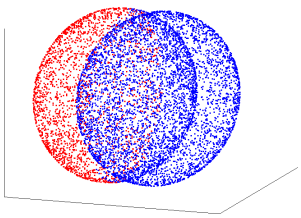

Figure 23: $(k_+, 0)$-graph with angle constraint

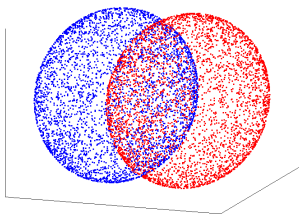

Figure 24: $(k_+, \frac{2k_+}{3})$-graph with angle constraint

4.2.2 Path constraint graphs Vs other proximity graphs: self-intersections

In general, the use of fully inner connected and sparsely outer connected graphs on data sets imposes a specific geometric structure on the set $\mathcal{M}$ that is not necessarily inherited from the ambient space $\mathbb{R}^d$. This is true in the multi-manifold setting or even in the case of a single self-intersecting manifold (a setting not considered in our theoretical results). Take for example the self-intersecting manifold illustrated in Figures 25-27. When running spectral clustering with the annular graph with angle constraints, we get a partition of the data corresponding to the one we would have obtained when clustering a one-dimensional curve with no self-intersections. This is illustrated in Figure 25. Figures 26 and 27, on the other hand, show the clusters obtained when running spectral clustering based on a standard k-NN graph and a standard ε-proximity graph respectively. As can be observed, these partitions are markedly different from the one in Figure 25. Notice that the MMC method can detect the self-intersection point in the manifold from this example.

Figures 25-27 illustrate the effect of different graphs on the output clusters. Likewise, different graphs capture the underlying manifold differently when using higher eigenmodes to summarize additional geometric content (a specific geometric content) of the self-intersecting manifold.

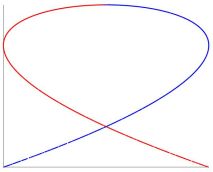

Figure 25: $\varepsilon_+, \varepsilon_-$-graph with angle constraint

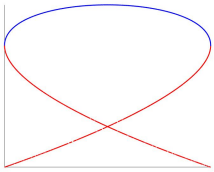

Figure 26: standard k-NN

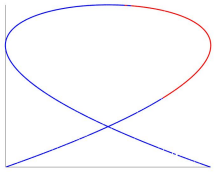

Figure 27: standard ε-neighbor

4.2.3 Other path algorithms

The path-based similarity weights we study in this paper are inspired by an algorithm proposed in Babaeian et al. (2015). There, a less stringent notion of a "smooth" discrete path is used to construct a proximity graph on the data set. In our construction, we force discrete paths to satisfy that every line segment in the path must be aligned with the segment connecting the first and last point in the path (i.e., essentially requiring a straight path). In contrast, in Babaeian et al. (2015) the constraint is that any two consecutive segments in the path must be aligned (i.e., a path that does not turn too quickly). It is straightforward to see that when two manifolds with a dimension larger than two intersect, it is possible to construct paths connecting points in the two manifolds that meet the criterion in Babaeian et al. (2015) but not our criterion. In summary, the more stringent constraint we impose helps remove more connections (faulty and correct). The removal of faulty connections seems more significant, and overall, our path algorithm outperforms the one in Babaeian et al. (2015).

Another sensible path algorithm to build graphs for MMC is to directly find geodesic paths along the graph using Dijkstra's algorithm and then check whether they satisfy the angle constraints or not. The theoretical analysis for this approach is more involved since one needs to check that geodesics do satisfy the angle constraint in the cases where one expects them to (i.e., when connecting two points on the same manifold). Still, in practice this graph construction behaves comparably to the path algorithm we analyze mathematically. Our path construction and the geodesic based one are two examples of a more general procedure where one seeks a path that connects a pair of points satisfying the angle constraints *and* whose length is no larger than a constant parameter times the geodesic distance along the path between the two points. This construction can be analyzed by combining ideas similar to the ones we have presented in section 3 with some analysis of the geodesic distance in a proximity graph.

4.3 Sensitivity of theoretical assumptions

4.3.1 Angles

We test the performance of spectral clustering on an annular graph with angle constraints when trying to separate manifolds as their angles of intersection decrease (i.e. β in (2.2) grows). In our experiment, we consider the simple setting of two intersecting planes. We see

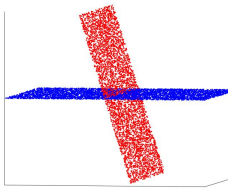

Figure 28: 75°

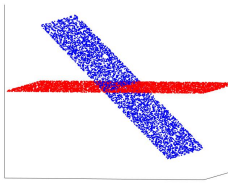

Figure 29: 50°

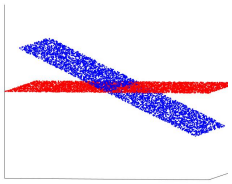

Figure 30: 30°

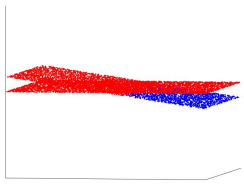

Figure 31: 10°

that in Figures 28-30 we recover the two planes, while in Figure 31 we do not. The results here are reasonable because when the angle of intersection is too small, a much smaller threshold value for the angle constraint is needed to discriminate different manifolds at the expense of removing connections between points that should have been connected otherwise. For these experiments we have used the NN version of our algorithm.

4.3.2 Orthogonal noise

In our theoretical results, we assumed data points to lie exactly on top of a set of the form $\mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_K$. However, a natural question is whether spectral clustering with the similarity graph constructed with the path algorithm continues to perform well when orthogonal noise is added to the data. Figures 32 and 33 show two examples of data sets contaminated by orthogonal noise. In both cases, the multi-manifold structure is readily apparent: three intersecting lines at a single point. However, in the setting depicted in Figure 33, where the noise level is large, we see that the path algorithm does not recover the

multi-manifold structure correctly. This suggests that the path algorithm is quite sensitive to noise. For these experiments we have used the NN version of our algorithm.

We can use the number of connections to see how much the noise affects the algorithm. For example, in Figure 33, where we exhibit the clean data, the total number of connections between data points is 579208, while the number of faulty connections is 1126. When noise is added, the number of total connections is 193426, while the number of faulty connections is 5414. That is, in general, we expect noise to worsen both inner and outer connectivities.

A potential remedy is to pre-process the data set by running a denoiser. However, some naive denoising methods, including the centering method or projected PCA, do not improve the performance. In Figure 34 we illustrate the outcome of spectral clustering on an annular proximity graph with angle constraints on the denoised data set. Specifically, using the centering method, the total number of connections was 215260, and the number of faulty connections was 5418. For the projected PCA method, the total number of connections is 271498, and the number of faulty connections is 6956. Roughly speaking, these methods can improve the inner connectivity while worsening the outer connectivity. How to implement a good denoising strategy in the MMC setting is an interesting direction to explore.

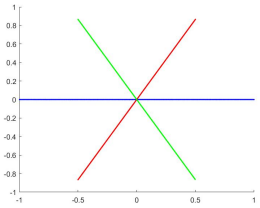

Figure 32: Small Perturbation

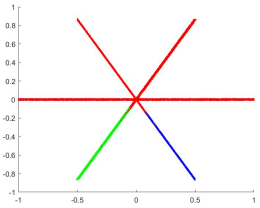

Figure 33: Large Perturbation

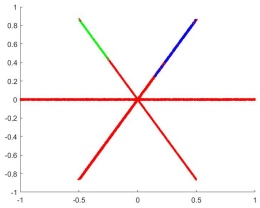

Figure 34: After denoising

4.3.3 Small Vs large number of data points

In this section, we consider data sets supported on the union of three intersecting planes as illustrated in Figures 35 and 36. In both figures, the underlying planes are the same, and the only thing that changes from one figure to the other is the sample size.

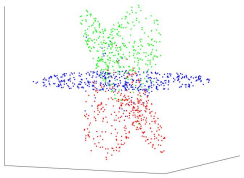

Figure 35: n data points

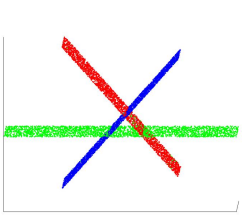

Figure 36: $2n$ data points

As can be observed, the three planes are not appropriately identified in the small sample size regime from Figure 35. In contrast, when we duplicate the amount of data as in Figure 36 the three planes are identified correctly.

This simple example illustrates some crucial drawbacks of the MMC methods based on spectral clustering discussed throughout the paper. In order to correctly construct local paths (or local tangent planes) to, in turn, detect the underlying manifolds, one needs to consider a large enough neighborhood around every point containing enough samples for the variance of the estimation to be small at the expense of increasing the bias considerably. Building MMC methods that can operate at smaller sample sizes is an interesting direction to explore in future research. For example, one could attempt to design a hybrid method that uses both path-based and local tangent plane information to make the method more robust to lower sample size; this is motivated by the fact that local PCA approaches can more accurately operate at smaller sample sizes when considering points that are far away from the intersection of manifolds.

4.4 Different dimensions

In section 2.4 we presented a series of theoretical results for multi-manifold clustering when $\mathcal{M}$ is the union of smooth manifolds with different dimensions. We now illustrate these results with a few simple numerical examples.

4.4.1 Planes and lines

We consider a data set uniformly sampled from the union of two planes and two lines that meet orthogonally, as illustrated in Figures 37-40. We run spectral clustering with $K = 2, 3, 4, 5$ to understand how the geometries of the manifolds get captured.

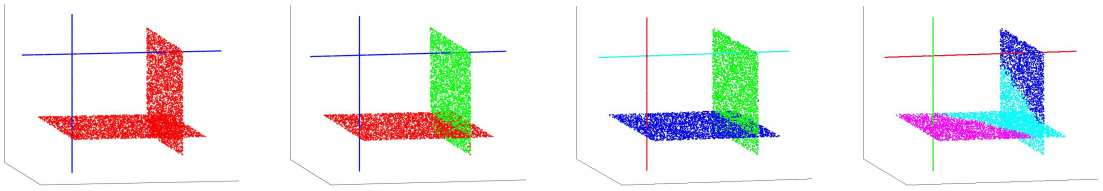

Figure 37: 2 Clusters Figure 38: 3 Clusters Figure 39: 4 Clusters Figure 40: 5 Clusters

In Figure 37, when we consider $K = 2$, the whole data set splits into two parts: lines and planes, indicating that manifolds with different dimensions are separated, and manifolds with the same dimension are put into the same cluster. In Figure 38, when we try $K = 3$ clusters, the two planes get separated perfectly while the lines are clustered as one; this is supported by our theory which indeed suggests that the geometry of the higher dimensional objects is detected first. When $K = 4$, lines get separated as shown in Figure 39. The case $K = 5$ illustrates the theory developed in this paper quite well. It shows how the internal geometry of the higher dimensional manifolds (in this case, the planes) is detected because the internal geometry of lines is more expensive than planes.

Another illustration of the behavior of spectral clustering with constrained annular proximity graphs is presented in Figures 41- 43. Here the data set is supported in the union of a 2-dimensional sphere and three lines that connect at one point. The same observations we made in the planes and lines example also apply to this setting.

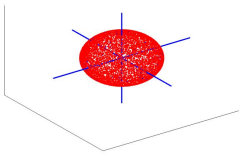

Figure 41: 2 clusters

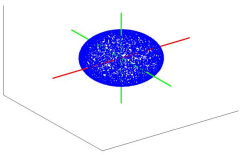

Figure 42: 3 clusters

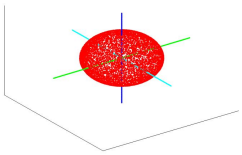

Figure 43: 4 clusters

4.5 Comparison with other MMC approaches

In this section, we compare the performances of spectral clustering using annular proximity graphs with angle constraints, SMCE (Sparse Manifold Clustering and Embedding) Elhamifar and Vidal (2011), and SC with local PCA Arias-Castro et al. (2017); both SMCE and SC with local PCA have been designed for MMC tasks. For SMCE, we follow the parameter choices in Elhamifar and Vidal (2011) and grid search for optimal parameters. For local PCA, we follow the parameter setting as in Arias-Castro et al. (2017) and tune the radius parameter and dimension by running a grid search to get the lowest misclustering rate. Since the algorithm in Arias-Castro et al. (2017) is a randomized algorithm, we run the algorithm 100 times and report the average misclustering rate. We use both k-NN and ε graphs in the setting of Arias-Castro et al. (2017) and report the results of the best performing settings. For SC with the constrained path algorithm, we use its nearest neighbor version as discussed at the beginning of section 4. We first compare all algorithms when we run them on synthetic data sets and then conduct a comparison when running them on the MNIST data set.

4.5.1 Synthetic data sets

We generate data points from five different settings of intersecting manifolds; see Table 1. We see that SC with the angle-constrained path algorithm achieves, overall, the lowest misclustering rate, outperforming the competing algorithms. We can see that SC with local PCA can work well for the settings of 2 spheres and 1 sphere with 1 plane; in those settings, most misclustered points are points close to the intersections of manifolds. For the 3 planes example from section 4.3.3, i.e. three planes intersecting at the same line, we sample 3000 points. Local PCA has particular difficulty distinguishing points close to the intersection, and only until the sample size has been increased considerably we recover the correct clustering with that algorithm. The 2 planes with 1 line and 1 sphere with 1 line examples are used to evaluate the performance of the algorithms when manifolds have different dimensions, a setting that is not the original target of Arias-Castro et al. (2017), where a dimension parameter must be chosen. On the other hand, SMCE is not particularly

designed to handle intersecting manifolds, and we see the overall low performance in most of the experiments run.

Algorithm	3planes	2spheres	2planes 1line	1sphere 1line	1sphere 1plane
path	14.72%	0.49%	0.38%	0.22%	0.78%
SMCE	48.9%	16.2%	31.1%	26.8%	42.7%
local PCA	43.66%	3.1%	32.17%	16.43%	1.19%

Table 1: Misclustering rate

4.5.2 MNIST

In this section we compare misclustering rates when we test algorithms on subsets of the MNIST data set consisting of different pairs of digits. Following the same preprocessing step as in Babaeian (2018), we first utilize the SURF feature of the Bag of words model to represent the features of each image. The original feature vectors have a size of 500. Then, for some pairs of digits, we use PCA to reduce the dimension of the image vector to 10. In the final step we apply the unsupervised algorithms to the data sets. We only present the results for some examples of pairs of digits for brevity (see Table 2), but similar observations to the ones that we discuss below can be drawn from other choices of digits. Like in the synthetic data experiments, we grid search the optimal parameters for every algorithm and for every task. In all the tasks considered in this section we also run vanilla SC algorithm with a standard kNN graph, tuning k to achieve the best misclustering rate. In contrast to the experiments in section 4.5.1, here it is not clear that the considered data sets possess an underlying multi-manifold geometric structure.

SC with the angle-constrained path algorithm can be seen as a generalization to vanilla SC, and we can see that it improves SC significantly in some tasks, such as clustering between digits [0,2], and at least behaves comparably to vanilla SC in other tasks; presumably, the improved performance over vanilla SC is manifested when the data manifolds corresponding to different digits do intersect. Notice that for the [0,7] digits SC with path algorithm and vanilla SC fail, while SMCE does perform very well. Overall, local PCA performs poorly for the tasks discussed here.

We want to highlight that the performance of the MMC algorithms that we have compared in these experiments may strongly depend on the manifold assumption (which may not hold on first place), and thus, if one was to stick to the theoretical assumptions discussed in this paper, one would need to guarantee, for example, that the data embedding methods in the preprocessing steps preserve or enhance these assumptions. The experiments that we have considered here are thus not meant to suggest that one algorithm is always better than the others. Instead, we wanted to evaluate the performance of algorithms with theoretical guarantees such as SC using path-based graphs on real data sets to test their capabilities and highlight that other methods used in the literature may underperform in some standard real data tasks.

Algorithm	[0,1]	[0,2]	[0,3]	[0,4]	[0,5]	[0,6]	[0,7]	[0,8]	[0,9]
path	14.0%	5.6%	1.9%	1.8%	2.6%	7.7%	46.4%	9.7%	1.9%
local PCA	6.4%	25.9%	30.0%	45.5%	34.8%	34.5%	34.1%	26.6%	25.1%
SMCE	20.0%	25.5%	6.9%	9.2%	24.1%	12.1%	2.9%	17.8%	3.8%
SC	18.8%	12.8%	1.8%	2.2%	2.6%	10.0%	46.4%	11.8%	2.3%

Table 2: Misclustering rates for some subsets of MNIST

Acknowledgments

NGT was supported by NSF-DMS grant 2005797. Support for this research was provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation. The authors would like to thank the IFDS at UW-Madison and NSF through TRIPODS grant 2023239 for their support.

References

Eddie Aamari and Clément Levrard. Stability and minimax optimality of tangential delaunay complexes for manifold reconstruction. Discrete & Computational Geometry, 59(4): 923–971, 2018.

Mahmuda Ahmed, Brittany Terese Fasy, Kyle S Hickmann, and Carola Wenk. A path-based distance for street map comparison. ACM Transactions on Spatial Algorithms and Systems (TSAS), 1(1):1–28, 2015.

Ery Arias-Castro. Clustering based on pairwise distances when the data is of mixed dimensions. IEEE Transactions on Information Theory, 57(3):1692–1706, 2011.

Ery Arias-Castro, Guangliang Chen, Gilad Lerman, et al. Spectral clustering based on local linear approximations. Electronic Journal of Statistics, 5:1537–1587, 2011.

Ery Arias-Castro, Gilad Lerman, and Teng Zhang. Spectral clustering based on local pca. The Journal of Machine Learning Research, 18(1):253–309, 2017.

Thierry Aubin. Some nonlinear problems in Riemannian geometry. Springer Science & Business Media, 1998.

Amir Babaeian. Multiple manifold clustering using curvature constrained path. arXiv preprint arXiv:1812.02327, 2018.

Amir Babaeian, Mohammadreaza Babaee, Alireza Bayestehtashk, and Mojtaba Bandarabadi. Nonlinear subspace clustering using curvature constrained distances. Pattern Recognition Letters, 68:118–125, 2015.

Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. In International Conference on Computational Learning Theory, pages 486–500. Springer, 2005.

Terrance E Boult and Lisa Gottesfeld Brown. Factorization-based segmentation of motions. In Proceedings of the IEEE workshop on visual motion, pages 179–186. IEEE, 1991.

Xavier Bresson and Thomas Laurent. Asymmetric cheeger cut and application to multi-class unsupervised clustering. CAM report, pages 12–27, 2012.

Xavier Bresson, Thomas Laurent, David Uminsky, and James von Brecht. Convergence and energy landscape for cheeger cut clustering. Advances in Neural Information Processing Systems, 25, 2012a.

Xavier Bresson, Xue-Cheng Tai, Tony F Chan, and Arthur Szlam. Multi-class transductive learning based on l1 relaxations of Cheeger cut and mumford-shah-potts model. UCLA CAM Report, pages 12–03, 2012b.

Xavier Bresson, Thomas Laurent, David Uminsky, and James von Brecht. Multiclass total variation clustering. In Advances in Neural Information Processing Systems, pages 1421–1429, 2013a.

Xavier Bresson, Thomas Laurent, David Uminsky, and James von Brecht. An adaptive total variation algorithm for computing the balanced cut of a graph. preprint arXiv:1302.2717, 2013b.

Dmitri Burago, Sergei Ivanov, and Yaroslav Kurylev. A graph discretization of the Laplace-Beltrami operator. Journal of Spectral Theory, 4(4):675–714, 2014.

Jeff Calder and Nicolás García Trillos. Improved spectral convergence rates for graph laplacians on ε-graphs and k-nn graphs. Applied and Computational Harmonic Analysis, 60:123–175, 2022.

Jeff Calder, Nicolás García Trillos, and Marta Lewicka. Lipschitz regularity of graph laplacians on random data clouds. SIAM Journal on Mathematical Analysis, 54(1):1169–1222, 2022.

Hong Chang and Dit Yan Yeung. Robust path-based spectral clustering. Pattern Recognition, 41(1):191–203, 2008.

Guangliang Chen and Gilad Lerman. Foundations of a multi-way spectral clustering framework for hybrid linear modeling. Foundations of Computational Mathematics, 9(5):517–558, 2009a.

Guangliang Chen and Gilad Lerman. Spectral curvature clustering (scc). International Journal of Computer Vision, 81(3):317–330, 2009b.

Ronald R Coifman and Stéphane Lafon. Diffusion maps. Applied and computational harmonic analysis, 21(1):5–30, 2006.

David B Dunson, Hau Tieng Wu, and Nan Wu. Spectral convergence of graph laplacian and heat kernel reconstruction in l^∞ from random samples. Applied and Computational Harmonic Analysis, 55:282–336, 2021.

Ehsan Elhamifar and René Vidal. Sparse subspace clustering. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 2790–2797. IEEE, 2009.

Ehsan Elhamifar and René Vidal. Clustering disjoint subspaces via sparse representation. In 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, pages 1926–1929. IEEE, 2010.

Ehsan Elhamifar and René Vidal. Sparse manifold clustering and embedding. Advances in neural information processing systems, 24, 2011.

Lawrence C. Evans. Partial differential equations, volume 19 of Graduate Studies in Mathematics. American Mathematical Soc., second edition, 2010.

Herbert Federer. Curvature measures. Transactions of the American Mathematical Society, 93(3):418–491, 1959.

Bernd Fischer, Thomas Zöller, and Joachim M Buhmann. Path based pairwise data clustering with application to texture segmentation. In International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, pages 235–250. Springer, 2001.

Michael L Fredman and Robert Endre Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. Journal of the ACM (JACM), 34(3):596–615, 1987.

Nicolás García Trillos and Dejan Slepčev. A variational approach to the consistency of spectral clustering. Applied and Computational Harmonic Analysis, 45(2):239–281, 2018.

Nicolás García Trillos, Moritz Gerlach, Matthias Hein, and Dejan Slepčev. Error estimates for spectral convergence of the graph laplacian on random geometric graphs toward the laplace–beltrami operator. Foundations of Computational Mathematics, pages 1–61, 2019.

Nicolás García Trillos, Franca Hoffmann, and Bamdad Hosseini. Geometric structure of graph laplacian embeddings. Journal of Machine Learning Research, 22:63–1, 2021.

Evarist Giné and Vladimir Koltchinskii. Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results. In High dimensional probability, volume 51, pages 238–259. JSTOR, 2006.

Alvina Goh and René Vidal. Segmenting motions of different types by unsupervised manifold clustering. In 2007 IEEE Conference on Computer Vision and Pattern Recognition, pages 1–6. IEEE, 2007.

Matthias Hein, Jean-Yves Audibert, and Ulrike 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.

Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. Graph laplacians and their convergence on random neighborhood graphs. Journal of Machine Learning Research, 8 (Jun):1325–1368, 2007.

John M. Lee. Introduction to Smooth Manifolds. Graduate Texts in Mathematics. Springer, 2003.

Anna Little, Mauro Maggioni, and James M. Murphy. Path-based spectral clustering: Guarantees, robustness to outliers, and fast algorithms. Journal of Machine Learning Research, 21(6):1–66, 2020.

Anna Little, Daniel McKenzie, and James M. Murphy. Balancing geometry and density: Path distances on high-dimensional data. SIAM Journal on Mathematics of Data Science, 4(1):72–99, 2022.

Guangcan Liu, Zhouchen Lin, Yong Yu, et al. Robust subspace segmentation by low-rank representation. In International Conference on Machine Learning, volume 1, page 8. Citeseer, 2010.

Jinpeng Lu. Graph approximations to the laplacian spectra. Journal of Topology and Analysis, 14(01):111–145, 2022.

Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. Advances in neural information processing systems, 14, 2001.

Nicolas Normand, Robin Strand, Pierre Evenou, and Aurore Arlicot. Path-based distance with varying weights and neighborhood sequences. In Discrete Geometry for Computer Imagery, pages 199–210, 2011.

Urvashi Oswal and Robert Nowak. Scalable sparse subspace clustering via ordered weighted l_1 regression. In 2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pages 305–312. IEEE, 2018.

Dohyung Park, Constantine Caramanis, and Sujay Sanghavi. Greedy subspace clustering. In Advances in neural information processing systems, pages 2753–2761, 2014.

Azriel Rosenfeld and John L. Pfaltz. Sequential operations in digital picture processing. The Journal of the ACM, page 471–494, 1966.

Geoffrey Schiebinger, Martin J Wainwright, and Bin Yu. The geometry of kernelized spectral clustering. The Annals of Statistics, 43(2):819–846, 2015.

Amit Singer. From graph to manifold laplacian: The convergence rate. Applied and Computational Harmonic Analysis, 21(1):128–134, 2006.

Amit Singer and Hau Tieng Wu. Spectral convergence of the connection Laplacian from random samples. Information and Inference: A Journal of the IMA, 6(1):58–123, 2017.

Wenqi Tao and Zuoqiang Shi. Convergence of laplacian spectra from random samples. Journal of Computational Mathematics, 38(6):952–984, 2020.

Daniel Ting, Ling Huang, and Michael I. Jordan. An analysis of the convergence of graph laplacians. In Proceedings of the 27th International Conference on International Conference on Machine Learning, page 1079–1086, 2010.

Ryan Vaughn, Tyrus Berry, and Harbir Antil. Diffusion maps for embedded manifolds with boundary with applications to pdes. arXiv preprint arXiv:1912.01391, 2019.

René Vidal. Subspace clustering. IEEE Signal Processing Magazine, 28(2):52–68, 2011.

Rene Vidal, Yi Ma, and Shankar Sastry. Generalized principal component analysis (gpca). IEEE transactions on pattern analysis and machine intelligence, 27(12):1945–1959, 2005.

Ulrike Von Luxburg. A tutorial on spectral clustering. Statistics and computing, 17(4): 395–416, 2007.

Ulrike von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. The Annals of Statistics, 36(2):555–586, 2008.

Van Vu. A simple svd algorithm for finding hidden partitions. Combinatorics, Probability and Computing, 27(1):124–140, 2018.

Xu Wang, Konstantinos Slavakis, and Gilad Lerman. Riemannian multi-manifold modeling. arXiv preprint arXiv:1410.0095, 2014.

Xu Wang, Konstantinos Slavakis, and Gilad Lerman. Multi-Manifold Modeling in Non-Euclidean spaces. In Guy Lebanon and S. V. N. Vishwanathan, editors, Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, volume 38 of Proceedings of Machine Learning Research, pages 1023–1032, San Diego, California, USA, 09–12 May 2015. PMLR.

Yong Wang, Yuan Jiang, Yi Wu, and Zhi-Hua Zhou. Spectral clustering on multiple manifolds. IEEE Transactions on Neural Networks, 22(7):1149–1161, 2011.

Caroline L Wormell and Sebastian Reich. Spectral convergence of diffusion maps: Improved error bounds and an alternative normalization. SIAM Journal on Numerical Analysis, 59 (3):1687–1734, 2021.

Hau Tieng Wu and Nan Wu. When locally linear embedding hits boundary. arXiv preprint arXiv:1811.04423, 2018.

Jingyu Yan and Marc Pollefeys. A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In Computer Vision – ECCV 2006, pages 94–106, 2006.

Ying Wu, Zhengyou Zhang, T. S. Huang, and J. Y. Lin. Multibody grouping via orthogonal subspace decomposition. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, volume 2, pages II–II. IEEE, 2001.

Teng Zhang, Arthur Szlam, Yi Wang, and Gilad Lerman. Hybrid linear modeling via local best-fit flats. International journal of computer vision, 100(3):217–240, 2012.

Appendix A. Proofs of main results

A.1 Discrete Dirichlet energies

It is well known that an operator like $\mathcal{L}^{\varepsilon_+,\varepsilon_-}$ (defined in (2.7)) is positive semi-definite with respect to $\langle\cdot,\cdot\rangle_{L^2(\mu^n)}$ (e.g. Von Luxburg (2007)); here and in the remainder we use μ^n to denote the empirical measure of X. Notice that $\mathcal{L}^{\varepsilon_+,\varepsilon_-}$'s eigenvalues, labeled in ascending order as

$$0 = \lambda_1^{\varepsilon_+,\varepsilon_-} \leq \lambda_2^{\varepsilon_+,\varepsilon_-} \leq \lambda_3^{\varepsilon_+,\varepsilon_-} \leq \cdots \leq \lambda_n^{\varepsilon_+,\varepsilon_-},$$

can be characterized variationally according to the Courant-Fisher minmax principle:

$$\lambda_l^{\varepsilon_+,\varepsilon_-} = \min_{S\in\mathcal{G}_l} \max_{u\in S\setminus\{0\}} \frac{b^{\varepsilon_+,\varepsilon_-}(u)}{\|u\|^2_{L^2(\mu^n)}}, \tag{A.1}$$

where $\mathcal{G}_l$ denotes the set of all linear subspaces of $L^2(\mu^n)$ of dimension l. Here, $b^{\varepsilon_+,\varepsilon_-}$ is the Dirichlet energy:

$$b^{\varepsilon_+,\varepsilon_-}(u) := \frac{1}{n^2(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})} \sum_{x_i,x_j\in\mathcal{X}_n} \omega_{x_ix_j}(u(x_i)-u(x_j))^2 = \frac{1}{2}\langle \mathcal{L}^{\varepsilon_+,\varepsilon_-}u, u\rangle_{L^2(\mu^n)} \tag{A.2}$$

where $u \in L^2(\mu^n)$.

We introduce *inner* and *outer* weights associated to the ω defined as $\omega^I_{x_ix_j} = \omega_{x_ix_j}$ and $\omega^O_{x_ix_j} = 0$ when x_i, x_j belong to the same manifold, and $\omega^I_{x_ix_j} = 0$ and $\omega^O_{x_ix_j} = \omega_{x_ix_j}$ otherwise. With this notation in place, we can introduce outer and inner Dirichlet energies associated to $\mathcal{L}^{\varepsilon_+,\varepsilon_-}$ according to:

$$\begin{aligned} b_O^{\varepsilon_+,\varepsilon_-}(u) &:= \frac{1}{n^2(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})} \sum_{x_i,x_j\in\mathcal{X}_n} \omega^O_{x_ix_j}(u(x_i)-u(x_j))^2, \\ b_I^{\varepsilon_+,\varepsilon_-}(u) &:= \frac{1}{n^2(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})} \sum_{x_i,x_j\in\mathcal{X}_n} \omega^I_{x_ix_j}(u(x_i)-u(x_j))^2. \end{aligned} \tag{A.3}$$

Clearly $b^{\varepsilon_+,\varepsilon_-} = b_O^{\varepsilon_+,\varepsilon_-} + b_I^{\varepsilon_+,\varepsilon_-}$.

It will be convenient for our analysis to decompose $b_I^{\varepsilon_+,\varepsilon_-}$ further and write it as the sum of Dirichlet energies associated to each of the manifolds $\mathcal{M}_k$. For that purpose we split the data set X into disjoint sets $X = \bigcup_{k=1}^N X_k$, where each of the X_k can be taken to be, without the loss of generality, equal to $X_k = X \cap \mathcal{M}_k$ (this is due to the first condition in Assumption 1 which implies that with probability one no x_i belongs to two or more of the $\mathcal{M}_k$). It is worth highlighting that the previous partitioning of the data makes sense mathematically even if it is not meaningful in practice (because we do not know the manifolds $\mathcal{M}_k$). In what follows and whenever needed we list the points in X_k as $\{x_{1k}, x_{2k}, \cdots, x_{n_kk}\}$ and use μ_k^n to denote their associated empirical probability measure. The number of data points in $\mathcal{M}_k$, i.e. n_k, is easily seen to satisfy $\mathbb{E}n_l = nw_l$. Moreover, the following concentration estimate holds.

Proposition 23 *With probability no less than* $1 - 2\exp\left(\frac{-2t^2}{n}\right)$, *we have*

$$nw_i - t < n_i < nw_i + t. \tag{A.4}$$

The graph Dirichlet energy associated to an individual manifold is defined by

$$b_l^{\varepsilon_+,\varepsilon_-}(u_l) := \frac{1}{n_l^2(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})} \sum_{x_i,x_j \in \chi_n^l} \omega_{x_i x_j}(u_l(x_i) - u_l(x_j))^2, \quad u_l \in L^2(\mu_l^n). \tag{A.5}$$

It follows that

$$b_I^{\varepsilon_+,\varepsilon_-}(u) = \sum_{l=1}^{N} \left(\frac{n_l^2}{n^2}\right) b_l^{\varepsilon_+,\varepsilon_-}(u_l), \quad u \in L^2(\mu^n),$$

where in the above and in the remainder we identify a function $u : X \to \mathbb{R}$ with a tuple $(u_1, \ldots, u_N)$ where each of the u_k is a function from X_k into $\mathbb{R}$.

Remark 24 *The local discrete Dirichlet energies $b_k^{\varepsilon_+,\varepsilon_-}$ are similar to discrete Dirichlet energies that have been studied in the literature under the smooth manifold assumption. There is however an important difference. Indeed, although the weight matrix ω is assumed to satisfy the full inner connectivity condition, i.e. with high probability the weights $\omega_{x_i x_j}$ can be thought of as those coming from a proximity graph, the type of proximity graph that we consider here is not standard since it is built with a kernel that has annular geometric structure. This type of kernel has not been considered nor analyzed before. As observed intuitively, as well as in our experiments, the idea of removing connections between points that are too close to each other significantly helps in reducing the number of connections between points in different manifolds, a feature that is useful for the multi-manifold clustering problem.*

A.2 Discretization and interpolation maps

Our first goal is to find a quantitative relationship between the Dirichlet energies D and $b^{\varepsilon_+,\varepsilon_-}$ via two conveniently chosen maps $P : L^2(\mu) \to L^2(\mu^n)$ and $\mathcal{I} : L^2(\mu^n) \to L^2(\mu)$. We look forward to obtaining inequalities of the form:

$$\sigma_\eta D(Iu) \leq (1 + e_1) b^{\varepsilon_+,\varepsilon_-}(u); \quad b^{\varepsilon_+,\varepsilon_-}(Pf) \leq (1 + e_2)\sigma_\eta D(f) + e_3 \tag{A.6}$$

where e_1, e_2, e_3 are small error terms depending on the problem's parameters, and σ_η is the constant in (2.8).

We start by combining Proposition 2.11 in Calder and García Trillos (2022) with Proposition (23) to obtain the probabilistic estimates that we use in the remainder to connect graph-based energies with their continuum counterparts.

Corollary 25 *With probability at least $1 - \sum_{l=1}^{N}(nw_l + t)\exp\left(-\mathrm{C}(nw_l - t)\theta^2\widetilde{\delta}^{m_l}\right) - 2N\exp\left(\frac{-2t^2}{n}\right)$, there exist:*

1. *probability density functions $\widetilde{\rho}_l^n : \mathcal{M}_l \to \mathbb{R}$ satisfying:*

$$\|\rho_l - \widetilde{\rho}_l^n\|_{L^\infty(\mathcal{M}_l)} \leq C(\theta + \widetilde{\delta})$$

for each $l = 1, \ldots, N$, and also

2. *maps* $\widetilde{T}_1, \dots, \widetilde{T}_N$ *such that for each* l, $\widetilde{T}_l : \mathcal{M}_l \to X_l$ *is the* ∞*-OT map between* $\widetilde{\rho}_l^n d\text{vol}_{\mathcal{M}_l}$ *and* μ_l^n*, and*

$$\sup_{x \in \mathcal{M}_l} d_{\mathcal{M}_l}(x, \widetilde{T}_l(x)) \leq \widetilde{\delta}.$$

Each of the maps $\widetilde{T}_l$ in the above corollary induces a partition $\widetilde{U}_{1l}, \dots, \widetilde{U}_{n_l l}$ of $\mathcal{M}_l$, where:

$$\widetilde{U}_{il} := \widetilde{T}_l^{-1}\left(\{x_{il}\}\right).$$

For each $l = 1, \dots, N$, a (local) *discretization* map $\widetilde{P}_l : L^2(\mu_l) \to L^2\left(\mu_l^n\right)$ is defined as

$$\left(\widetilde{P}_l f_l\right)(x_{il}) := n_l \cdot \int_{\widetilde{U}_{il}} f(x)\widetilde{\rho}_l^n(x) d\text{vol}_{\mathcal{M}_l}(x), \quad f_l \in L^2(\mu_l), \tag{A.7}$$

and an associated (local) *extension* map $\widetilde{P}_l^* : L^2\left(\mu_l^n\right) \to L^2\left(\widetilde{\mu}_l^n\right)$ defined as

$$\widetilde{P}_l^* u = u \circ \widetilde{T}_l.$$

The (global) discretization map $P : L^2(\mu) \to L^2(\mu^n)$ can now be defined according to

$$Pf := (P_1 f_1, \dots, P_N f_N)$$

where $f = (f_1, \dots, f_N) \in L^2(\mu)$. In other words, P acts on f according to the coordinatewise action of the P_l on the f_l. Likewise, we may define $\widetilde{P}^* : L^2(\mu^n) \to L^2(\mu)$ according to:

$$P^* u = (P_1^* u_1, \dots, P_N^* u_N).$$

We now introduce the interpolation map $\mathcal{I} : L^2(\mu^n) \to L^2(\mu)$. This map takes the form $\mathcal{I} = \Lambda \widetilde{P}^*$, i.e. it is the composition of the extension map $\widetilde{P}^*$ and a smoothening operator that acts coordinatewise. The smoothening operator is chosen conveniently so as to make the error in the first inequality in (A.6) as small as possible; the first work to our knowledge that attempted to do something similar when analyzing graph Laplacians is Burago et al. (2014). To conduct the analysis in our setting we must introduce new constructions and prove new results given the annular geometry of the kernel used to build the data graph.

Let $\eta : [0, \infty) \to \mathbb{R}$ and $\psi : [0, \infty) \to [0, \infty)$ be the functions given by

$$\eta(t) := \begin{cases} 1 & 0 \leq t \leq 1 \\ 0 & t > 1, \end{cases} \qquad \psi(t) := \frac{1}{\sigma_\eta} \int_t^\infty \eta(s) s ds, \tag{A.8}$$

where recall σ_η was defined in (2.8).

For every r_1, r_2 such that $r_1 > r_2$ we define the function:

$$\mathcal{K}^l_{r_1, r_2}(x, y) := \left(\frac{r_1^2}{r_1^{m+2} - r_2^{m+2}} \psi\left(\frac{d_{\mathcal{M}_l}(x, y)}{r_1} \right) - \frac{r_2^2}{r_1^{m+2} - r_2^{m+2}} \psi\left(\frac{d_{\mathcal{M}_l}(x, y)}{r_2} \right) \right), \quad x, y \in \mathcal{M}_l$$

which serves as "kernel" and induces the convolution operator:

$$\Lambda^l_{r_1, r_2} f(x) := \frac{1}{\tau_l(x)} \int_{\mathcal{M}_l} \mathcal{K}^l_{r_1, r_2}(x, y) f_l(y) d\text{vol}_{\mathcal{M}_l}(y),$$

which acts on functions $f_l : \mathcal{M}_l \to \mathbb{R}$. In the above, $\tau_l(x)$ is a normalization factor given by

$$\tau_l(x) := \int_{\mathcal{M}_l} \mathcal{K}^l_{r_1,r_2}(x,y) d\mathrm{vol}_{\mathcal{M}_l}(y);$$

notice that $\mathcal{K}^l_{r_1,r_2}$ is non-negative.

We can put together the action of each convolution operator on each of the manifolds and define:

$$\Lambda_{r_1,r_2} f := (\Lambda^1_{r_1,r_2} f_1, \ldots, \Lambda^N_{r_1,r_2} f_N), \quad f = (f_1, \ldots, f_N) \in L^2(\mu).$$

Our global *interpolation operator* takes the form:

$$\mathcal{I} u := \Lambda \widetilde{P}^* u, \quad u \in L^2(\mu^n);$$

In the remainder it will be convenient to write the above in coordinates as:

$$\mathcal{I} u = (\mathcal{I}_1 u_1, \ldots, \mathcal{I}_N u_N).$$

Having defined the maps P and $\mathcal{I}$ we are now ready to state precisely the connection between the Dirichlet energies D and b.

Proposition 26 (Inequality for Dirichlet energies) *Let $\varepsilon_+, \varepsilon_-, \widetilde{\delta}$, and θ be fixed but small enough numbers satisfying Assumptions 2. Let b be the Dirichlet energy associated to the weighted graph (X, ω) defined in* (A.2) *and D the Dirichlet energy defined in* (2.3).

Then, with probability greater than $1 - \sum_{l=1}^N (nw_l + t) \exp\left(-\mathrm{C}(nw_l - t)\theta^2 \widetilde{\delta}^m\right) - 2N \exp\left(\frac{-2t^2}{n}\right) - C_1(n)$, *we have:*

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

$$\sigma_\eta D(\widetilde{\mathcal{I}} u) \leq \left(1 + C(\varepsilon_+ + \frac{\widetilde{\delta}}{\varepsilon_+} + \theta + \widetilde{\delta})\right) b^{\varepsilon_+,\varepsilon_-}(u)$$

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

$$b_I^{\varepsilon_+,\varepsilon_-}(\widetilde{P} f) \leq \left(1 + C(\varepsilon_+ + \frac{\widetilde{\delta}}{\varepsilon_+} + \theta + \widetilde{\delta})\right) \sigma_\eta D(f)$$

In addition, if f is in the span of $\Delta_{\mathcal{M}}$'s eigenfunctions with corresponding eigenvalue less than λ, then:

$$b_O^{\varepsilon_+,\varepsilon_-}(\widetilde{P} f) \leq \frac{CNN_0}{w_{min}^2 n^2 (\varepsilon_+^{m+2} - \varepsilon_-^{m+2})} \left(1 + \lambda^{m/2+2}\right) \|f\|^2_{L^2(\mathcal{M})}$$

We recall that the quantity N_0 was introduced in section 2.1 in Definition 5 and it represents the largest number of connections in the graph (X, ω) between two distinct manifolds. The following estimates complement Proposition 26 and essentially state that the maps $\mathcal{I}$ and P are almost isometries.

Proposition 27 (Discretization and interpolation maps are almost isometries) *Let $\varepsilon_+, \varepsilon_-, \widetilde{\delta}$, and θ be fixed but small enough numbers satisfying Assumptions 2. Then, with probability at least $1 - \sum_{l=1}^{N}(nw_l + t)\exp\left(-\mathrm{C}(nw_l - t)w_l^2\widetilde{\delta}^m\right) - 2N\exp\left(\frac{-2t^2}{n}\right) - C_1(n)$, we have:*

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

$$\left|\|f\|^2_{L^2(\mu)} - \|Pf\|^2_{L^2(\mu^n)}\right| \leq C\widetilde{\delta}\|f\|_{L^2(\mu)}\sqrt{D(f)} + C(\theta + \widetilde{\delta})\|f\|^2_{L^2(\mu)}$$

(2) For every $u \in L^2(\mu^n)$

$$\left|\|u\|^2_{L^2(\mu^n)} - \|\mathcal{I}u\|^2_{L^2(\mu)}\right| \leq C\varepsilon_+\|u\|_{L^2(\mu^n)}\sqrt{b^{\varepsilon_+,\varepsilon_-}(u)} + C(\theta + \widetilde{\delta})\|u\|^2_{L^2(\mu^n)}$$

A.3 Preliminary local energy estimates

In order to prove the above results we first establish a sequence of preliminary estimates on each of the individual manifolds $\mathcal{M}_l$. The results presented in this subsection are independent of the fact that all manifolds forming our model (1.2) have the same dimension or not.

Lemma 28 *Suppose $0 < r_2 < \frac{1}{4}r_1$ are small enough, in particular smaller than half the injectivity radius of the manifold $\mathcal{M}_l$. Then, there exists an absolute constant $C > 0$ such that*

$$(1 + Cm_lK_lr_1^2)^{-1} \leq \tau_l(x) \leq 1 + Cm_lK_lr_1^2, \quad \text{and} \quad |\nabla\tau_l(x)| \leq \frac{Cm_lK_lr_1}{\sigma_\eta},$$

for all $x \in \mathcal{M}_l$. Here K_l is a uniform bound on the absolute value of sectional curvatures.

Proof First, notice that:

$$\begin{aligned}\tau_l(x) &= \frac{r_1^2}{r_1^{m_l+2} - r_2^{m_l+2}}\int_{\mathcal{M}_l}\psi\left(\frac{d_{\mathcal{M}_l}(x,y)}{r_1}\right)d\mathrm{vol}_{\mathcal{M}_l}(y) - \frac{r_2^2}{r_1^{m_l+2} - r_2^{m_l+2}}\int_{\mathcal{M}_l}\psi\left(\frac{d_{\mathcal{M}_l}(x,y)}{r_2}\right)d\mathrm{vol}_{\mathcal{M}_l}(y)\\ &= \frac{r_1^2}{r_1^{m_l+2} - r_2^{m_l+2}}\int_{B_{m_l}(0,r_1)}\psi\left(\frac{|v|}{r_1}\right)J_x(v)dv - \frac{r_2^2}{r_1^{m_l+2} - r_2^{m_l+2}}\int_{B_{m_l}(0,r_2)}\psi\left(\frac{|v|}{r_2}\right)J_x(v)dv,\end{aligned}$$

where in the above J_x denotes the Jacobian of the exponential map $\exp_x : B_{m_l}(0,\iota_l) \to B_{\mathcal{M}_l}(x,\iota_l)$. Using a standard estimate for the Jacobian, namely

$$(1 + Cm_lK_l|v|^2)^{-1} \leq J_x(v) \leq 1 + Cm_lK_l|v|^2, \quad \forall v \in B_{m_l}(0,\iota_l/2), \tag{A.9}$$

we can see that $\tau_l(x)$ satisfies $(1 + Cm_lK_lr_1^2)^{-1}C_\alpha \leq \tau_l(x) \leq (1 + Cm_lK_lr_1^2)C_\alpha$ for some constant C, and for

$$C_\alpha := \frac{r_1^2}{r_1^{m_l+2} - r_2^{m_l+2}}\int_{B_{m_l}(0,r_1)}\psi\left(\frac{|v|}{r_1}\right)dv - \frac{r_2^2}{r_1^{m_l+2} - r_2^{m_l+2}}\int_{B_{m_l}(0,r_2)}\psi\left(\frac{|v|}{r_2}\right)dv.$$

A direct computation using polar coordinates and integration by parts reveals that C_α is actually equal to one. This establishes the first assertion.

To obtain the estimate for the gradient of $\tau_l(x)$, we notice that from the the definition of ψ in (A.8), the chain rule, and the fact that $\nabla d_{\mathcal{M}_l}(\cdot,y)(x) = -\frac{1}{d_{\mathcal{M}_l}(x,y)}\exp_x^{-1}(y)$, it follows:

$$\begin{aligned}
|\nabla\tau_l(x)| &= \frac{1}{\sigma_\eta(r_1^{m_l+2}-r_2^{m_l+2})}\left|\int_{B_{\mathcal{M}_l}(x,r_1)}\left(\eta\left(\frac{d_{\mathcal{M}_l}(x,y)}{r_1}\right)-\eta\left(\frac{d_{\mathcal{M}_l}(x,y)}{r_2}\right)\right)\exp_x^{-1}(y)d\mathrm{vol}_{\mathcal{M}_l}(y)\right| \\
&= \frac{1}{\sigma_\eta(r_1^{m_l+2}-r_2^{m_l+2})}\left|\int_{B_{m_l}(r_1)}\eta\left(\frac{|v|}{r_1}\right)vJ_x(v)dv-\int_{B_{m_l}(r_1)}\eta\left(\frac{|v|}{r_2}\right)vJ_x(v)dv\right| \\
&\leq \frac{Cm_lK_lr_1^2}{\sigma_\eta(r_1^{m_l+2}-r_2^{m_l+2})}\left(\int_{B_{m_l}(r_1)}\eta\left(\frac{|v|}{r_1}\right)|v|dv+\int_{B_{m_l}(r_2)}\eta\left(\frac{|v|}{r_2}\right)|v|dv\right) \\
&\leq \frac{Cm_lK_lr_1^2(r_1^{m_l+1}+r_2^{m_l+1})}{\sigma_\eta(r_1^{m_l+2}-r_2^{m_l+2})} \leq \frac{Cm_lK_lr_1}{\sigma_\eta}.
\end{aligned}$$

Notice that in the first inequality we have used (A.9) and the radial symmetry of the integrands (which induces a cancellation). In the last step we used $0 < r_2 \leq \frac{1}{4}r_1$. ■

The next definitions is used in the subsequent lemmas. For every $l = 1,\dots,N$ we define

$$\begin{aligned}
\widetilde{D}_{NL,l}^{\varepsilon_+,\varepsilon_-}(f_l) := \frac{1}{\varepsilon_+^{m+2}-\varepsilon_-^{m+2}}\int_{\mathcal{M}_l}\int_{\mathcal{M}_l}\left[\eta\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_+}\right)-\eta\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_-}\right)\right](f_l(x)-f_l(y))^2 \\
\widetilde{\rho}_l^n(x)\widetilde{\rho}_l^n(y)d\mathrm{vol}_{\mathcal{M}_l}(x)d\mathrm{vol}_{\mathcal{M}_l}(y),
\end{aligned} \tag{A.10}$$

where we recall that the densities $\widetilde{\rho}_l^n$ are defined in Corollary 25. We also consider:

$$E_l^r(f_l) := \int_{\mathcal{M}_l}\int_{\mathcal{M}_l}\eta\left(\frac{d_{\mathcal{M}}(x,y)}{r}\right)(f_l(x)-f_l(y))^2\widetilde{\rho}_l^n(x)\widetilde{\rho}_l^n(y)d\mathrm{vol}_{\mathcal{M}_l}(x)d\mathrm{vol}_{\mathcal{M}_l}(y), \quad f_l\in L^2(\mu_l) \tag{A.11}$$

Notice that for every $r_1 > r_2 > 0$ we have:

$$(r_1^{m+2}-r_2^{m+2})D_{NL,l}^{r_1,r_2}(f_l) = E_l^{r_1}(f_l)-E_l^{r_2}(f_l), \quad f_l\in L^2(\mu_l). \tag{A.12}$$

Lemma 29 *Suppose $\varepsilon_+,\varepsilon_-$ satisfy Assumptions 2. Then, there exists a universal constant $C>0$ such that for every $0<\varepsilon_-<\frac{1}{4}\varepsilon_+$ and every $f_l\in L^2(\mu_l)$*

$$\frac{1}{\varepsilon_+^{m+2}-\varepsilon_-^{m+2}}E_l^{\varepsilon_+}(f_l) \leq C\left(1+c_\rho L_{\rho_l}\right)D_{NL,l}^{\varepsilon_+,\varepsilon_-}(f_l),$$

where L_{ρ_l} is a constant depending on ρ_l.

Proof

The proof is very similar to the one in Lemma 4 in García Trillos et al. (2019). In that Lemma the idea is to cover a larger ball with smaller balls and use the triangle inequality. Here the only difference is that we want to cover a larger ball with a collection of annuli.

■

Lemma 30 *Suppose that $\varepsilon_+, \varepsilon_-$ satisfy Assumptions 2. Then, there exists a universal constant $C > 0$ such that*

$$D_l(\Lambda^l_{\varepsilon_+,\varepsilon_-} f_l) \leq (1 + c_\rho L_{\rho_l} \varepsilon_+) \left[1 + C m_l K_l \varepsilon_+^2 \left(1 + \frac{\sqrt{1 + c_\rho L_p}}{\sigma_\eta}\right)\right] \frac{1}{\sigma_\eta} D^{\varepsilon_+,\varepsilon_-}_{NL,l}(f_l), \quad \forall f_l \in L^2(\mathcal{M}_l, \rho_l).$$

We recall that D_l was defined in (2.6) *and $D^{\varepsilon_+,\varepsilon_-}_{NL,l}$ in* (A.10).

Proof We can write $\nabla(\Lambda^l_{\varepsilon_+,\varepsilon_-} f_l)$ as

$$\nabla(\Lambda^l_{\varepsilon_+,\varepsilon_-} f_l) = \frac{1}{\tau_l(x)} A^l_1(x) + A^l_2(x),$$

where

$$A^l_1(x) := \int_{R_{\mathcal{M}_l}(x,\varepsilon_+,\varepsilon_-)} \nabla \mathcal{K}^l_{\varepsilon_+,\varepsilon_-}(\cdot, y)(x) \left(f_l(y) - f_l(x)\right) d\mathrm{vol}_{\mathcal{M}_l}(y)$$

and

$$A^l_2(x) = \nabla\left(\frac{1}{\tau_l(x)}\right) \int_{R_{\mathcal{M}_l}(x,\varepsilon_+,\varepsilon_-)} \mathcal{K}^l_{\varepsilon_+,\varepsilon_-}(x, y) \left(f_l(y) - f_l(x)\right) d\mathrm{vol}_{\mathcal{M}_l}(y);$$

here $R_{\mathcal{M}_l}(x, \varepsilon_+, \varepsilon_-) := \{\tilde{x} \in \mathcal{M}_l \,:\, \varepsilon_- < d_{\mathcal{M}_l}(x, \tilde{x}) < \varepsilon_+\}$.

We find a bound for $|A^l_1(x)|^2$; notice that $\frac{1}{\tau_l(x)} \leq 1 + C m_l K_l \varepsilon_+^2$ by (28). First, notice that:

$$\begin{aligned}
\nabla \mathcal{K}^l_{\varepsilon_+,\varepsilon_-}(\cdot, y)(x) &= \nabla\left[\frac{\varepsilon_+^2}{\varepsilon_+^{m+2} - \varepsilon_-^{m+2}} \psi\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_+}\right) - \frac{\varepsilon_-^2}{\varepsilon_+^{m+2} - \varepsilon_-^{m+2}} \psi\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_-}\right)\right] \\
&= -\frac{1}{\varepsilon_+^{m+2} - \varepsilon_-^{m+2}} \left(\varepsilon_+ \psi'\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_+}\right) - \varepsilon_- \psi'\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_-}\right)\right) \frac{\exp_x^{-1}(y)}{d_{\mathcal{M}_l}(x,y)} \\
&= \frac{\exp_x^{-1}(y)}{\sigma_\eta(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})} \left[\eta\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_+}\right) - \eta\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_-}\right)\right].
\end{aligned}$$

Since for $A^l_1(x)$ we have $|A^l_1(x)| = \langle A^l_1(x), w\rangle$ for some unit vector $w \in T_x\mathcal{M}_l$, we can combine with the inequality above to obtain:

$$\begin{aligned}
|A^l_1(x)| &= \langle A^l_1(x), w\rangle \\
&= \frac{1}{\sigma_\eta(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})} \int_{R_{\mathcal{M}_l}(x,\varepsilon_+,\varepsilon_-))} \left[\eta\left(\frac{d(x,y)}{\varepsilon_+}\right) - \eta\left(\frac{d(x,y)}{\varepsilon_-}\right)\right] \\
&\quad \cdot \left(f_l(y) - f_l(x)\right) \langle \exp_x^{-1}(y), w\rangle d\mathrm{vol}_{\mathcal{M}_l}(y) \\
&= \frac{1}{\sigma_\eta(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})} \int_{R(\varepsilon_+,\varepsilon_-)} \left[\eta\left(\frac{|u|}{\varepsilon_+}\right) - \eta\left(\frac{|u|}{\varepsilon_-}\right)\right] \phi(u) \langle u, w\rangle J_x(u) du,
\end{aligned}$$

where $\phi(u) = f_l\left(\exp_x(u)\right) - f_l(x)$ and $R(\varepsilon_+, \varepsilon_-) := \{u \in \mathbb{R}^{m_l} \,:\, \varepsilon_- < |u| < \varepsilon_+\}$. By the Cauchy-Schwartz inequality,

$$
\begin{aligned}
|A_1^l(x)|^2 &\le \frac{1}{\sigma_\eta^2(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})^2}\int_{R(\varepsilon_+,\varepsilon_-)}|\phi(u)|^2 J_x(u)^2\left[\eta\left(\frac{|u|}{\varepsilon_+}\right)-\eta\left(\frac{|u|}{\varepsilon_-}\right)\right]du \\
&\quad\cdot\int_{R(\varepsilon_+,\varepsilon_-)}\langle u,w\rangle^2\left[\eta\left(\frac{|u|}{\varepsilon_+}\right)-\eta\left(\frac{|u|}{\varepsilon_-}\right)\right]du \\
&= \frac{1}{\sigma_\eta(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})}\int_{R(\varepsilon_+,\varepsilon_-)}|\phi(u)|^2 J_x(u)^2\left[\eta\left(\frac{|u|}{\varepsilon_+}\right)-\eta\left(\frac{|u|}{\varepsilon_-}\right)\right]du \\
&\le \frac{1+Cm_lK_l\varepsilon_+^2}{\sigma_\eta(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})}\int_{\mathcal{M}_l}\left[\eta\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_+}\right)-\eta\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_-}\right)\right](f_l(y)-f_l(x))^2\,d\mathrm{vol}_{\mathcal{M}_l}(y),
\end{aligned}
\tag{A.13}
$$

where the equality comes from

$$
\int_{R(\varepsilon_+,\varepsilon_-)}\langle u,w\rangle^2\left[\eta\left(\frac{|u|}{\varepsilon_+}\right)-\eta\left(\frac{|u|}{\varepsilon_-}\right)\right]du=\int_{B_m(\varepsilon_+)}\langle u,w\rangle^2du-\int_{B_m(\varepsilon_-)}\langle u,w\rangle^2du=\sigma_\eta(\varepsilon_+^{m+2}-\varepsilon_-^{m+2}),
$$

and the last inequality from the bound (A.9). Integrating (A.13) against $\rho_l^2 d\mathrm{vol}_{\mathcal{M}_l}$ and using the Lipschitz continuity of ρ_l we deduce

$$
\begin{aligned}
\left\|\frac{A_1^l}{\tau_l}\right\|^2_{L^2(\mathcal{M}_l,\rho_l^2\mathrm{vol}_{\mathcal{M}_l})} &\le \frac{(1+Cm_lK_l\varepsilon_+^2)(1+c_\rho L_{\rho_l}\varepsilon_+)}{\sigma_\eta(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})} \\
&\quad\cdot\int_{\mathcal{M}_l}\int_{B_{\mathcal{M}_l}(x,\varepsilon_+,\varepsilon_-)}\left[\eta\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_+}\right)-\eta\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_-}\right)\right]|f_l(y)-f_l(x)|^2 d\mu_l(x)d\mu_l(y) \\
&\le \frac{(1+Cm_lK_l\varepsilon_+^2)(1+c_\rho L_{\rho_l}\varepsilon_+)}{\sigma_\eta}D_{NL,l}^{\varepsilon_+,\varepsilon_-}(f_l).
\end{aligned}
$$

Now we analyze the term $A_2^l(x)$. First, recall that $|\nabla(\tau_l^{-1})|\le\frac{Cm_lK_l\varepsilon_+}{\sigma_\eta}$ and $\tau_l\le 1+Cm_lK_l\varepsilon_+^2$ by Lemma 28. Using the mean value theorem, it is straightforward to show that

$$
\mathcal{K}^l_{\varepsilon_+,\varepsilon_-}(x,y)\le\frac{\varepsilon_+^2}{\varepsilon_+^{m+2}-\varepsilon_-^{m+2}}\psi(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_+})\le\frac{\varepsilon_+^2\eta(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_+})}{\sigma_\eta(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})}.
\tag{A.14}
$$

Thus, by the Cauchy-Schwartz inequality and (A.14), we have

$$
\begin{aligned}
|A_2^l(x)|^2 &\le |\nabla(\frac{1}{\tau_l(x)})|^2\int_{\mathcal{M}_l}\mathcal{K}^l_{\varepsilon_+,\varepsilon_-}(x,y)d\mu(y)\int_{\mathcal{M}_l}|f_l(x)-f_l(y)|^2\mathcal{K}^l_{\varepsilon_+,\varepsilon_-}(x,y)d\mu_l(y) \\
&\le \frac{C^2m^2K_l^2\varepsilon_+^2}{\sigma_\eta^3(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})}\int_{\mathcal{M}_l}\varepsilon_+^2\eta\left(\frac{d_{\mathcal{M}_l}(x,y)}{\varepsilon_+}\right)|f_l(x)-f_l(y)|^2d\mu_l(y).
\end{aligned}
$$

Integrating both sides of the above inequality with respect to $\rho_l^2 d\mathrm{vol}_{\mathcal{M}_l}$, using the Lipschitz continuity of ρ_l, and using Lemma 29, we conclude that

$$
\|A_2^l\|_{L^2(\mathcal{M}_l,\rho_l^2\mathrm{vol}_{\mathcal{M}_l})}\le\frac{Cm_lK_l\varepsilon_+^2(1+c_\rho L_{\rho_l}\varepsilon_+)\sqrt{1+c_\rho L_{\rho_l}}}{\sigma_\eta}\sqrt{\frac{1}{\sigma_\eta}D_{NL,l}^{\varepsilon_+,\varepsilon_-}(f_l)},
$$

for some universal constant C. Combining the estimates for $\left\| \frac{A_1^l}{\tau_l} \right\|^2_{L^2(\mathcal{M}_l, \rho_l^2 \mathrm{vol}_{\mathcal{M}_l})}$ and $\|A_2^l\|_{L^2(\mathcal{M}_l, \rho_l^2 \mathrm{vol}_{\mathcal{M}_l})}$ we finally obtain:

$$\left(D_l(\Lambda_{\varepsilon_+,\varepsilon_-} f_l)\right)^{1/2} \leq (1 + c_\rho L_{\rho_l} \varepsilon_+) \left[1 + C m_l K_l \varepsilon_+^2 \left(1 + \frac{\sqrt{1 + c_\rho L_{\rho_l}}}{\sigma_\eta}\right)\right] \sqrt{\frac{1}{\sigma_\eta} D_{NL,l}^{\varepsilon_+,\varepsilon_-}(f_l)}.$$

■

Lemma 31 *Suppose $\varepsilon_+, \varepsilon_-$ satisfy Assumptions 2. Then, there exists a universal constant $C > 0$ such that*

$$D_{\varepsilon_+,\varepsilon_-}^{NL,l}(f_l) \leq (1 + c_\rho L_{\rho_l} \varepsilon_+) \left(1 + C m_l K_l \varepsilon_+^2\right) \sigma_\eta D_l(f_l), \quad \forall f_l \in L(\mu_l).$$

Proof By a density argument we may assume without the loss of generality that f_l is smooth. Now, for every $x \in \mathcal{M}_l$,

$$\int_{R_{\mathcal{M}_l}(x,\varepsilon_+,\varepsilon_-)} |f_l(x) - f_l(y)|^2 d\mu(y) = \int_{R(\varepsilon_+,\varepsilon_-)} |f_l(\exp_x(v)) - f_l(x)|^2 \rho_l(\exp_x(v)) J_x(v) dv,$$

where $R_{\mathcal{M}_l}(x, \varepsilon_+, \varepsilon_-)$ and $R(\varepsilon_+, \varepsilon_-)$ are as defined in the proof of Lemma 30, and J_x is the Jacobian of the exponential map at x. From the Fundamental Theorem of Calculus it follows that

$$|f_l(\exp_x(v)) - f_l(x)|^2 \leq \int_0^1 |\frac{d}{dt} f_l\left(\exp_x(tv)\right)|^2 dt = \int_0^1 |df_l(\Phi_t(x,v)_2)|^2 dt,$$

where Φ_t denotes the time t geodesic flow on $\mathcal{M}_l$'s tangent bundle $\mathcal{TM}_l$: that is, $\Phi_t(x,v) = (\varphi_{x,v}(t), \varphi'_{x,v}(t)) \in \mathcal{TM}_l$, where $\varphi_{x,v}(t) := \exp_x(tv)$ and $\varphi'_{x,v}(t)$ is obtained by parallel transporting the vector $v \in \mathcal{T}_x\mathcal{M}_l$ along the geodesic connecting x and $\exp_x(tv)$; $\Phi_t(x,v)_2$ denotes the second coordinate of $\Phi_t(x,v)$. We can then obtain:

$$\begin{aligned}
&\int_{\mathcal{M}_l} \int_{R(\varepsilon_+,\varepsilon_-)} |f_l(\exp_x(v)) - f_l(x)|^2 \rho_l\left(\exp_x(v)\right) dv \rho_l(x) d\mathrm{vol}_{\mathcal{M}_l}(x) \\
&\leq \int_0^1 \int_{\mathcal{M}} \int_{R(\varepsilon_+,\varepsilon_-)} |df_l\left(\Phi_t(x,v)\right)|^2 \rho_l\left(\Phi_1(x,v)_1\right) \rho_l\left(\Phi_0(x,v)_1\right) dv d\mathrm{vol}_{\mathcal{M}_l}(x) dt,
\end{aligned}$$

where we use $\Phi_0(x,v)_1$ and $\Phi_1(x,v)_1$ to denote the first coordinates of $\Phi_0(x,v)$ and $\Phi_1(x,v)$ respectively. From the Lipschitz continuity of ρ_l it follows $\rho_l(x) \leq (1 + c_\rho L_{\rho_l} \varepsilon_+) \rho(y)$ for all $x, y \in \mathcal{M}_l$ satisfying $d(x,y) \leq \varepsilon_+$. Combining the fact that Φ_t preserves the canonical volume form $\mathrm{vol}_{T\mathcal{M}_l}$ on $T\mathcal{M}_l$ and that

$$\mathcal{R}(\varepsilon_+, \varepsilon_-) := \{\xi = (x,v) \in T\mathcal{M} : \varepsilon_- \leq |v| \leq \varepsilon_+\}; \quad \mathcal{B}_r := \{\xi = (x,v) \in T\mathcal{M} : |v| \leq r\}$$

are invariant under Φ_t, we obtain

$$\begin{aligned}
&\int_{\mathcal{M}}\int_{R(\varepsilon_+,\varepsilon_-)} |f_l(\exp_x(v)) - f_l(x)|^2 \rho_l\left(\exp_x(v)\right) dv \rho_l(x) d\mathrm{vol}_{\mathcal{M}_l}(x) \\
&\leq (1 + c_\rho L_{\rho_l}\varepsilon_+)\int_0^1 \int_{\mathcal{R}(\varepsilon_+,\varepsilon_-)} |df_l\left(\Phi_t(\xi)_2\right)|^2 \rho_l^2\left(\Phi_t(\xi_1)\right) d\mathrm{vol}_{T\mathcal{M}_l}(\xi) dt \\
&= (1 + c_\rho L_{\rho_l}\varepsilon_+)\left[\int_{\mathcal{B}_{\varepsilon_+}} |df_l(\xi_2)|^2 \rho_l^2(\xi_1) d\mathrm{vol}_{T\mathcal{M}_l}(\xi) - \int_{\mathcal{B}_{\varepsilon_-}} |df_l(\xi_2)|^2 \rho_l^2(\xi_1) d\mathrm{vol}_{T\mathcal{M}_l}(\xi)\right] \\
&= (1 + c_\rho L_{\rho_l}\varepsilon_+)\sigma_\eta(\varepsilon_+^{m+2} - \varepsilon_-^{m+2}) \int_{\mathcal{M}} |\nabla f_l|^2 \rho_l^2(x) d\mathrm{vol}_{\mathcal{M}_l}(x).
\end{aligned}$$

Therefore,

$$\begin{aligned}
D_{NL,l}^{\varepsilon_+,\varepsilon_-}(f_l) &\leq \left(1 + C m_l K_l \varepsilon_+^2\right) \cdot \\
&\quad \cdot \frac{1}{\varepsilon_+^{m+2} - \varepsilon_-^{m+2}} \int_{\mathcal{M}_l}\int_{R(\varepsilon_+,\varepsilon_-)} |f_l(\exp_x(v)) - f_l(x)|^2 \rho_l\left(\exp_x(v)\right) dv \rho(x) d\mathrm{vol}_{\mathcal{M}_l}(x) \\
&\leq \left(1 + C m_l K_l \varepsilon_+^2\right) \cdot (1 + c_\rho L_{\rho_l}\varepsilon_+)\sigma_\eta D_l(f_l).
\end{aligned}$$

■

The following is an adaptation of Lemma 14 in García Trillos et al. (2019) to the kernel with annular geometry that we consider in this paper.

Lemma 32 *Suppose $\widetilde{\delta}, \varepsilon_+, \varepsilon_-$ satisfy Assumptions 2. Then, with probability at least $1 - \sum_{l=1}^N (nw_l + t)\exp\left(-\mathrm{C}(nw_l - t)\theta^2\widetilde{\delta}^m\right) - 2N\exp\left(\frac{-2t^2}{n}\right)$ and for a universal constant $C > 0$, the following statements hold:*

(1) For every $u_l : \mathcal{X}_n^l \to \mathbb{R}$ we have

$$D_{NL,l}^{\varepsilon'_+,\varepsilon'_-}(P_l^* u_l) \leq \left(1 + C(\theta + \widetilde{\delta})\right)\left(1 + C(\frac{\widetilde{\delta}}{\varepsilon_+} + \varepsilon_-^2)\right) b_l^{\varepsilon_+,\varepsilon_-}(u_l),$$

where $\varepsilon'_+ := \varepsilon_+ - 2\widetilde{\delta}$ and $\varepsilon'_- := \varepsilon_- + 2\widetilde{\delta} + \frac{8\varepsilon_-'^3}{R^2}$. Applying Assumption 2, $C\frac{\widetilde{\delta}\varepsilon_+^{m+1} + \varepsilon_-^{m+4}}{\varepsilon_+^{m+2} - \varepsilon_-^{m+2}}$ can be simplified to $C_1\frac{\widetilde{\delta}}{\varepsilon_+} + C_2\varepsilon_-^2$, where $C_1 \leq \frac{16C}{15}, C_2 \leq \frac{C}{15}$; or just $C(\frac{\widetilde{\delta}}{\varepsilon_+} + \varepsilon_-^2)$.

(2) For every $f_l \in H^1(\mathcal{M}_l)$

$$b_l^{\varepsilon_+,\varepsilon_-}(P_l f_l) \leq \left(1 + C(\theta + \widetilde{\delta})\right)\left(1 + C(\frac{\widetilde{\delta}}{\varepsilon_+} + \varepsilon_-^2)\right) D_{NL,l}^{\varepsilon''_+,\varepsilon''_-}(f_l), \tag{A.15}$$

where $\varepsilon''_+ := \varepsilon_+ + \frac{8\varepsilon_+''^3}{R^2} + 2\widetilde{\delta}$ and $\varepsilon''_- := \varepsilon_- - 2\widetilde{\delta}$. We recall that b^l was introduced in (A.3).

Proof

We first recall a well known relation between the geodesic distance in $\mathcal{M}_l$ and the Euclidean distance in the ambient space $\mathbb{R}^d$. Namely,

$$|x-y| \leq d_{\mathcal{M}_l}(x,y) \leq |x-y| + \frac{8}{R_l^2}|x-y|^3, \quad x,y \in \mathcal{M}_l, \tag{A.16}$$

where R_l is the reach of the manifold $\mathcal{M}_l$ (see Federer (1959) for a definition of reach).

To show (1), notice that if $|x-y| < \varepsilon_-$, then from (A.16) we get

$$|x-y| \leq d_{\mathcal{M}_l}(x,y) \leq |x-y| + \frac{8\varepsilon_-^3}{R_l^2}. \tag{A.17}$$

We now use the map $\widetilde{T}_l$, the density $\widetilde{\rho}_l^n$ from Corollary 25, and the induced partition $\{U_{1l}, \dots, U_{n_l l}\}$ on $\mathcal{M}_l$ of the form $U_{il} = \widetilde{T}_l^{-1}(x_{il})$, where $x_{il} \in X_l$, to write

$$\begin{aligned}
&n_l^2(\varepsilon_+^{m+2} - \varepsilon_-^{m+2}) b_l^{\varepsilon_+,\varepsilon_-}(u_l) \\
&\quad = \sum_{i,j} \int_{U_{il}} \int_{U_{jl}} \left[\eta\left(\frac{|\widetilde{T}_l(x) - \widetilde{T}_l(y)|}{\varepsilon_+} \right) - \eta\left(\frac{|\widetilde{T}_l(x) - \widetilde{T}_l(y)|}{\varepsilon_-} \right) \right] \\
&\qquad \cdot |(P^* u_l)(x) - (P^* u_l)(y)|^2 \, \widetilde{\rho}_l^n(x) \widetilde{\rho}_l^n(y) d\mathrm{vol}_{\mathcal{M}_l}(y) d\mathrm{vol}_{\mathcal{M}_l}(x) \\
&\quad \geq \left(1 - C(\theta + \widetilde{\delta})\right) \int_{\mathcal{M}_l} \int_{\mathcal{M}_l} \left[\eta\left(\frac{d(\widetilde{T}_l(x), \widetilde{T}_l(y))}{\varepsilon_+} \right) - \eta\left(\frac{\left[d(\widetilde{T}_l(x), \widetilde{T}_l(y)) - \frac{8\varepsilon_-^3}{R^2} \right]_+}{\varepsilon_-} \right) \right] \\
&\qquad \cdot |(P^* u_l)(x) - (P^* u_l)(y)|^2 \, d\mu_l(y) d\mu_l(x) \\
&\quad \geq \left(1 - C(\theta + \widetilde{\delta})\right) \int_{\mathcal{M}_l} \int_{\mathcal{M}_l} \left[\eta\left(\frac{d(x,y) + 2\widetilde{\delta}}{\varepsilon_+} \right) - \eta\left(\frac{\left[d(x,y) - \frac{8\varepsilon_-^3}{R^2} - 2\widetilde{\delta} \right]_+}{\varepsilon_-} \right) \right] \\
&\qquad \cdot |(P^* u_l)(x) - (P^* u_l)(y)|^2 \, d\mu_l(y) d\mu_l(x) \\
&\quad \geq \left(1 - C(\theta + \widetilde{\delta})\right) (\varepsilon_+'^{m+2} - \varepsilon_-'^{m+2}) D_{NL,l}^{\varepsilon_+', \varepsilon_-'}(P_l^* u_l),
\end{aligned}$$

where in the first inequality we use i) in Corollary 25 and (A.17), and in the second inequality we use ii) in Corollary (25). Combining the above inequality with Assumptions 2 we conclude that

$$\left(1 + C(\theta + \widetilde{\delta})\right) \left(1 + C \frac{\widetilde{\delta}\varepsilon_+^{m+1} + \varepsilon_-^{m+4}}{\varepsilon_+^{m+2} - \varepsilon_-^{m+2}}\right) b_l^{\varepsilon_+,\varepsilon_-}(u_l) \geq D_{NL,l}^{\varepsilon_+', \varepsilon_-'}(P^* u_l).$$

For (2), we proceed similarly as in the proof of (1) to deduce

$$\begin{aligned}
b_l^{\varepsilon_+,\varepsilon_-}(Pf_l) &\leq \frac{1+C(\theta+\widetilde{\delta})}{\sigma_\eta(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})}\sum_i\sum_j\int_{U_{il}}\int_{U_{jl}}\left[\eta\left(\frac{|\widetilde{T}(x)-\widetilde{T}(y)|}{\varepsilon_+}\right)-\eta\left(\frac{|\widetilde{T}(x)-\widetilde{T}(y)|}{\varepsilon_-}\right)\right]\\
&\quad\cdot|f_l(y)-f_l(x)|^2d\mu_l(y)d\mu_l(x)\\
&\leq \frac{1+C(\theta+\widetilde{\delta})}{\sigma_\eta(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})}\int_{\mathcal{M}_l}\int_{\mathcal{M}_l}\left[\eta\left(\frac{\left[d(x,y)-\frac{8\varepsilon_+^3}{R^2}-2\widetilde{\delta}\right]_+}{\varepsilon_+}\right)-\eta\left(\frac{d(x,y)+2\widetilde{\delta}}{\varepsilon_-}\right)\right]\\
&\quad\cdot|f_l(y)-f_l(x)|^2d\mu_l(y)d\mu_l(x)\\
&\leq \frac{1+C(\theta+\widetilde{\delta})}{\sigma_\eta(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})}\left[E_l^{\varepsilon_+''}(f_l)-E_l^{\varepsilon_-''}(f_l)\right]\\
&\leq \left(1+C(\theta+\widetilde{\delta})\right)\left(1+C\frac{\varepsilon_-^{m+4}+\varepsilon_+^{m+1}\widetilde{\delta}}{\varepsilon_+^{m+2}-\varepsilon_-^{m+2}}\right)D_{NL,l}^{\varepsilon_+'',\varepsilon_-''}(f_l),
\end{aligned}$$

where in the last line we have used (A.11) and Assumptions 2. ■

We are ready to prove Proposition 26.

A.4 Proofs of Propositions 26 and 27

Proof [Proof of Proposition 26]

(1): Let $u\in L^2(\mu^n)$. We write u in coordinates as $u=(u_1,\dots,u_N)$. We combine Lemmas 30 and 32 to obtain for every $l=1,\dots,N$:

$$\sigma_\eta D_l(\mathcal{I}_l u_l)\leq\left(1+C(\theta+\widetilde{\delta})\right)(1+C\frac{\widetilde{\delta}}{\varepsilon_+})(1+c_\rho L_{\rho_l}\varepsilon_+)\left[1+Cm_lK_l\varepsilon_+^2\left(1+\frac{\sqrt{1+c_\rho L_\rho}}{\sigma_\eta}\right)\right]b_l^{\varepsilon_+,\varepsilon_-}(u_l).$$

From the above we deduce that $D(\mathcal{I}u)=\sum_{l=1}^N w_l^2D_l(\mathcal{I}_lu_l)$ is smaller than:

$$\left(1+C(\theta+\widetilde{\delta})\right)(1+C\frac{\widetilde{\delta}}{\varepsilon_+})(1+c_\rho L_\rho\varepsilon_+)\left[1+CmK\varepsilon_+^2\left(1+\frac{\sqrt{1+c_\rho L_\rho}}{\sigma_\eta}\right)\right]\sum_{l=1}^N w_l^2b_l^{\varepsilon_+,\varepsilon_-}(u_l).$$

In turn, Proposition 23 implies that with probability at least $1-2N\exp\left(\frac{-2t^2}{n}\right)$ we have

$$\sum_{l=1}^N w_l^2b_l^{\varepsilon_+,\varepsilon_-}(u_l)\leq(1+t)\sum_{l=1}^N\left(\frac{n_l}{n}\right)^2b_l^{\varepsilon_+,\varepsilon_-}(u_l)=(1+t)b_I^{\varepsilon_+,\varepsilon_-}(u)\leq(1+t)b^{\varepsilon_+,\varepsilon_-}(u).$$

Putting together the above inequalities we obtain the desired estimate. Here it is worth highlighting that the last inequality in the above expression comes from the fact that the discrete Dirichlet energy $b^{\varepsilon_+,\varepsilon_-}$ is the sum of $b_I^{\varepsilon_+,\varepsilon_-}$ and $b_O^{\varepsilon_+,\varepsilon_-}$. As we will see below, in order to obtain a reverse inequality between $b^{\varepsilon_+,\varepsilon_-}$ and D one needs to control $b_O^{\varepsilon_+,\varepsilon_-}(Pf)$ using regularity estimates of f in each of the $\mathcal{M}_l$. We will be able to obtain this control

when f is in the span of the eigenfunctions of $\Delta_{\mathcal{M}}$ smaller than a certain value (which is all we need in the remainder).

(2): Similarly to **(1)**, we may combine Lemma 31 and Lemma 32, to deduce:

$$\begin{aligned} b_I^{\varepsilon_+,\varepsilon_-}(Pf) &\leq \left(1 + c_\rho L_{\rho_l} \varepsilon_+''\right)\left(1 + C m_l K_l \varepsilon_+''^2\right)\left(1 + C\left(\frac{\varepsilon_+^{m+4} + \varepsilon_+^{m+1}\widetilde{\delta}}{\varepsilon_+^{m+2} - \varepsilon_-^{m+2}} + \theta + \widetilde{\delta}\right)\right)\sigma_\eta D(f) \\ &\leq \left(1 + C(\varepsilon_+'' + \frac{\widetilde{\delta}}{\varepsilon_+} + \theta + \widetilde{\delta}) + t\right)\sigma_\eta D(f). \end{aligned} \tag{A.18}$$

where the last step we used Assumption 2.

Let $f \in L^2(\mathcal{M})$ belong to the span of $\Delta_{\mathcal{M}}$'s eigenfunctions with corresponding eigenvalue less than λ. Then, f can be written as $f = \sum_{l=1}^N f^l$ where each f^l has support on $\mathcal{M}_l$, and where, abusing notation slightly, each f^l has the form $f^l = \sum_q b_{ql} f_q^l$ for an orthonormal basis of eigenfunctions of $\Delta_{\mathcal{M}_l}$, $\{f_q^l\}$, with corresponding eigenvalues smaller than λ. It is straightforward to see that:

$$\begin{aligned} b_O^{\varepsilon_+,\varepsilon_-}(\widetilde{P}f) &= \frac{1}{n^2(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})} \sum_{x_i, x_j \in \mathcal{X}_n} \omega_{x_i x_j}^O (\widetilde{P}f(x_i) - \widetilde{P}f(x_j))^2 \\ &\leq \frac{2}{\varepsilon_+^{m+2} - \varepsilon_-^{m+2}} \sum_{x_i \in \mathcal{X}_n} \sum_{x_j \in \mathcal{X}_n} \omega_{x_i x_j}^O (\widetilde{P}f(x_i))^2 \\ &= \frac{2}{\varepsilon_+^{m+2} - \varepsilon_-^{m+2}} \sum_{l=1}^N \sum_{s:s\neq l} \sum_{x_i \in \mathcal{M}_l} \sum_{x_j \in \mathcal{M}_s} \omega_{x_i x_j}^O \left| \int_{U_{il}} f^l(x) \widetilde{p}_l^n(x) d\mathrm{vol}_{\mathcal{M}_l}(x) \right|^2 \\ &= \frac{2}{\varepsilon_+^{m+2} - \varepsilon_-^{m+2}} \sum_{l=1}^N \frac{\|f^l\|_{L^\infty(\mathcal{M}_l)}^2}{n_l^2} \sum_{s:s\neq l} \sum_{x_i \in \mathcal{M}_l} \sum_{x_j \in \mathcal{M}_s} \omega_{x_i x_j}^O \\ &= \frac{2N \cdot N_0}{\varepsilon_+^{m+2} - \varepsilon_-^{m+2}} \sum_{l=1}^N \frac{\|f^l\|_{L^\infty(\mathcal{M}_l)}^2}{n_l^2} \\ &\leq \frac{2N \cdot N_0(1+t)}{w_{min}^2 n^2(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})} \sum_{l=1}^N \|f^l\|_{L^\infty(\mathcal{M}_l)}^2. \end{aligned} \tag{A.19}$$

In the above, the last inequality follows with high probability according to Proposition (23). To complete the proof we find estimates for each of the terms $\|f^l\|_{L^\infty(\mathcal{M}_l)}^2$ and to do this we adapt the argument in Lemma 3.3 of Lu (2022). From standard higher order elliptic regularity results (e.g. Theorem 2 in Evans (2010)) it follows that for every $s \in \mathbb{N}$

$$\left\|f^l\right\|_{H^{2s}(\mathcal{M}_l)}^2 \leq C(\mathcal{M}_l, \rho_l, s)\left(\left\|\Delta_{\mathcal{M}_l}^s f^l\right\|_{L^2(\mathcal{M}_l)}^2 + \left\|f^l\right\|_{L^2(\mathcal{M}_l)}^2\right) \quad ,$$

where in the above $H^{2s}(\mathcal{M}_l)$ is the Sobolev space of functions on $\mathcal{M}_l$ with square-integrable derivatives of order up to $2s$; it is at this stage that we use the smoothness of the manifold $\mathcal{M}_l$ and the density ρ_l. Moreover, by the Sobolev embedding theorem on compact manifolds

(e.g. Theorem 2.20 in Aubin (1998)), we have $H^{2s}(\mathcal{M}_k) \subset C^1(\mathcal{M}_k)$ as long as $2s > m/2+1$. Choosing $2s = m/2+2$ we obtain:

$$\left\|f^l\right\|_{L^\infty(\mathcal{M}_l)} \leq \left\|f^l\right\|_{C^1(\mathcal{M}_l)} \leq C\left(\mathcal{M}_l, \rho_l\right) \left\|f^l\right\|_{H^{m/2+2}(\mathcal{M}_l)} \leq C\left(\mathcal{M}_l, \rho_l\right) \left(\lambda^{m/4+1}+1\right) \left\|f^l\right\|_{L^2(\mathcal{M}_l)}. \tag{A.20}$$

Recalling that f^l has the form $\sum_q b_{ql} f_q^l$, where the f_q^l are orthonormal in $L^2(\mathcal{M}_l, \rho_l)$ and are eigenfunctions of $\Delta_{\mathcal{M}_l}$ with eigenvalues λ_q^l smaller than λ, we can see that

$$\begin{aligned}\left\|\Delta_{\mathcal{M}_l}^s f^l\right\|_{L^2(\mathcal{M}_l,\rho_l)}^2 &= \left\|\sum_q b_{ql}(\lambda_q^l)^s f_q^l\right\|_{L^2(\mathcal{M}_l,\rho_l)}^2 \\ &= \sum_q b_{ql}^2(\lambda_q^l)^{2s}\left\|f_q^l\right\|_{L^2(\mathcal{M}_l,\rho_l)}^2 \leq \lambda^{2s}\|f^l\|_{L^2(\mathcal{M}_l)}^2.\end{aligned}$$

Putting the above estimates together and combining with (A.19) gives us the desired result. ■

Before proving Proposition 27 we need one last preliminary estimate.

Lemma 33 *Suppose $\varepsilon_+, \varepsilon_-$ satisfy Assumptions 2. Then, there exists a universal constant $C > 0$ such that*

$$\|\Lambda_{\varepsilon_+,\varepsilon_-} f\|_{L^2(\mathcal{M},\rho)}^2 \leq (1 + Cc_\rho L_\rho \varepsilon_+)(1 + CmK\varepsilon_+^2)\|f\|_{L^2(\mathcal{M},\rho)}^2,$$

and

$$\|\Lambda_{\varepsilon_+,\varepsilon_-} f - f\|_{L^2(\mathcal{M},\rho)}^2 \leq \frac{Cc_\rho^2\varepsilon_+^2}{\sigma_\eta}\sum_{l=1}^N w_l D_{NL,l}^{\varepsilon_+,\varepsilon_-}(f_l) \leq \frac{Cc_\rho^2\varepsilon_+^2}{\sigma_\eta w_{\min}} D(f).$$

for all $f \in L^2(\mathcal{M}, \rho)$. In the above, $w_{\min} := \min_{l=1,\ldots,N} w_l$.

Proof

Since $\Lambda_{\varepsilon_+,\varepsilon_-}$ acts on f coordinatewise, we get

$$\begin{aligned}\int_{\mathcal{M}} (\Lambda_{\varepsilon_+,\varepsilon_-} f(x))^2 d\mu(x) &= \sum_{l=1}^N w_l \int_{\mathcal{M}_l} (\Lambda_{\varepsilon_+,\varepsilon_-}^l f_l(x))^2 \rho_l(x) d\mathrm{vol}_{\mathcal{M}_l}(x) \\ &\leq \sum_{l=1}^N w_l \int_{\mathcal{M}_l}\int_{\mathcal{M}_l} \frac{\mathcal{K}_{\varepsilon_+,\varepsilon_-}^l(x,y)}{\tau_l(x)} (f_l(y))^2 \rho_l(x) d\mathrm{vol}_{\mathcal{M}_l}(y) d\mathrm{vol}_{\mathcal{M}_l}(x) \\ &\leq (1 + Cc_\rho L_\rho\varepsilon_+)(1 + CmK\varepsilon_+^2)\sum_{l=1}^N w_l \int_{\mathcal{M}_l} (f_l(y))^2 \rho_l(y) d\mathrm{vol}_{\mathcal{M}_l}(y) \\ &= (1 + Cc_\rho L_\rho\varepsilon_+)(1 + CmK\varepsilon_+^2)\int_{\mathcal{M}} (f(x))^2 d\mu(x),\end{aligned}$$

where the first inequality follows from Jensen's inequality, and the second inequality follows from Lemma 28 and the properties of the density functions ρ_l.

For the second inequality, we first calculate the difference between $\Lambda^l_{\varepsilon_+,\varepsilon_-} f_l(x)$ and $f_l(x)$:

$$\begin{aligned}
|\Lambda_l^{\varepsilon_+,\varepsilon_-} f_l(x) - f_l(x)|^2 &= \left(\frac{1}{\tau_l(x)} \int_{\mathcal{M}_l} \mathcal{K}^l_{\varepsilon_+,\varepsilon_-}(x,y)(f_l(y) - f_l(x)) d\mu_l(y)\right)^2 \\
&\leq \frac{1}{\tau_l(x)^2} \int_{\mathcal{M}_l} \mathcal{K}^l_{\varepsilon_+,\varepsilon_-}(x,y) d\mu_l(y) \int_{\mathcal{M}_l} \mathcal{K}^l_{\varepsilon_+,\varepsilon_-}(x,y)(f_l(x) - f_l(y))^2 d\mu_l(y) \\
&= \frac{1}{\tau_l(x)} \int_{\mathcal{M}_l} \mathcal{K}^l_{\varepsilon_+,\varepsilon_-}(x,y)(f_l(x) - f_l(y))^2 d\mu_l(y).
\end{aligned}$$

Then we integrate with respect to $\rho_l(x) d\mathrm{vol}_{\mathcal{M}_l}(x)$ to get:

$$\begin{aligned}
\|\Lambda_l^{\varepsilon_+,\varepsilon_-} f_l - f_l\|^2_{L^2(\mathcal{M}_l,\rho_l)} &\leq (1 + C m_l K_l \varepsilon_+^2) \int_{\mathcal{M}_l} \int_{\mathcal{M}_l} \mathcal{K}^l_{\varepsilon_+,\varepsilon_-}(x,y)(f(x) - f(y))^2 d\mu_l(y) d\mu_l(x) \\
&\leq \frac{(1 + C m_l K_l \varepsilon_+^2)\varepsilon_+^2}{\sigma_\eta(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})} \int_{\mathcal{M}_l} \int_{\mathcal{M}_l} \eta\left(\frac{d(x,y)}{\varepsilon_+}\right)(f(x) - f(y))^2 d\mu_l(y) d\mu_l(x) \\
&\leq \frac{C c_\rho^2 \varepsilon_+^2}{\sigma_\eta} D^{\varepsilon_+,\varepsilon_-}_{NL,l}(f_l),
\end{aligned}$$

where the second inequality follows from the fact that $\eta \leq \frac{1}{\sigma_\eta}\psi$ (recall (A.8)), and the third inequality follows from Lemma 29. Multiplying the above by w_l, adding over l, and using Lemma 31 we get the desired result. ■

Proof [Proof of Proposition 27] **(1):** For each $l = 1, \ldots, N$, we use estimates proved in Calder and García Trillos (2022) (appearing in Pages 24-25 in the proof of Proposition 4.2) to conclude that there is a constant C for which

$$\begin{aligned}
&\left| \|f_l\|^2_{L^2(\widetilde{\mu}_l^n)} - \|f_l\|^2_{L^2(\mu_l)} \right| \leq C(\theta + \widetilde{\delta}) \|f_l\|^2_{L^2(\mu^l)} \\
&\left| \|\widetilde{P}_l f_l\|^2_{L^2(\mu_l^n)} - \|f_l\|^2_{L^2(\mu^l)} \right| \leq C \|f_l\|_{L^2(\mu^l)} \|\widetilde{P}_l^* \widetilde{P}_l f_l - f_l\|_{L^2(\widetilde{\mu}_n^l)} + C(\theta + \widetilde{\delta}) \|f_l\|^2_{L^2(\mu^l)} \\
&\|\widetilde{P}_l^* \widetilde{P}_l f_l - f_l\|^2_{L^2(\widetilde{\mu}_n^l)} \leq C \widetilde{\delta}^2 D_l(f_l),
\end{aligned}$$

for every $f_l \in L^2(\mu_l)$; in the above we use $\widetilde{\mu}_n^l$ to denote the measure $\widetilde{\rho}_n^l d\mathrm{vol}_{\mathcal{M}_l}(x)$. Combining the previous inequalities, we deduce that:

$$\begin{aligned}
\left| \|\widetilde{P} f\|^2_{L^2(\mu)} - \|f\|^2_{L^2(\mu)} \right| &\leq \sum_{l=1}^N w_l \left| \|\widetilde{P}_l f_l\|^2_{L^2(\mu_l^n)} - \|f_l\|^2_{L^2(\mu^l)} \right| \\
&\leq C \sum_{l=1}^N w_l \|f_l\|_{L^2(\mu^l)} \|\widetilde{P}_l^* \widetilde{P}_l f_l - f_l\|_{L^2(\widetilde{\mu}_n^l)} + C(\theta + \widetilde{\delta}) \|f\|^2_{L^2(\mu)}.
\end{aligned}$$

Now, the first term in the last inequality above is controlled by $C\widetilde{\delta} \|f\|_{L^2(\mu)} \sqrt{D(f)}$. Indeed, this follows from Cauchy-Schwartz inequality:

$$\begin{aligned}
\left(\sum_{l=1}^N w_l \|f_l\|_{L^2(\mu^l)} \|\widetilde{P}_l^* \widetilde{P}_l f_l - f_l\|_{L^2(\widetilde{\mu}_n^l)} \right)^2 &\leq \left(\sum_{l=1}^N w_l \|f_l\|^2_{L^2(\mu^l)} \right) \left(\sum_{l=1}^N w_l \|\widetilde{P}_l^* \widetilde{P}_l f_l - f_l\|^2_{L^2(\widetilde{\mu}_n^l)} \right) \\
&\leq C \frac{\widetilde{\delta}^2}{w_{\min}} \|f\|^2_{L^2(\mu)} D(f).
\end{aligned}$$

Putting things together we finally deduce

$$\left|\|\widetilde{P}f\|^2_{L^2(\mu^n)}-\|f\|^2_{L^2(\mu)}\right| \leq C\frac{\widetilde{\delta}}{\sqrt{w_{\min}}}\|f\|_{L^2(\mu)}\sqrt{D(f)}+C(\theta+\widetilde{\delta})\|f\|^2_{L^2(\mu)}.$$

(2): From the identity $\|u_l\|_{L^2(\mu_l^n)}=\|\widetilde{P}_l^{\,*}u_l\|_{L^2(\widetilde{\mu}_l^n)}$ (which follows automatically from the fact that the map $\widetilde{T}_l$ is a transport map between $\widetilde{\mu}_l^n$ and μ_l^n) and the triangle inequality we get

$$\begin{aligned}\left|\|\mathcal{I}_l u_l\|_{L^2(\widetilde{\mu}_l^n)}-\|u_l\|_{L^2(\mu_l^n)}\right| &\leq \|\Lambda_l^{\varepsilon_+,\varepsilon_-}\widetilde{P}_l^* u_l-\widetilde{P}_l^* u_l\|_{L^2(\widetilde{\mu}_l^n)}\\ &\leq \left(1+c_\rho\|\rho_l-\widetilde{\rho}_l^n\|_{L^\infty(\mathcal{M}_l)}\right)\cdot\|\Lambda_l^{\varepsilon_+,\varepsilon_-}\widetilde{P}_l^* u_l-\widetilde{P}_l^* u_l\|_{L^2(\mu_l)}\\ &\leq \left(1+c_\rho\|\rho_l-\widetilde{\rho}_l^n\|_{L^\infty(\mathcal{M}_l)}\right)\cdot C\varepsilon_+\sqrt{D^{\varepsilon_+,\varepsilon_-}_{NL,l}(\widetilde{P}_l^* u_l)}\\ &\leq C\varepsilon_+\sqrt{b_l^{\varepsilon_+,\varepsilon_-}(u_l)},\end{aligned}$$

where the third inequality comes from Lemma 33 and the last one follows from Lemma 32. Also, notice that

$$\|\mathcal{I}_l u_l\|_{L^2(\widetilde{\mu}_l^n)}=\|\Lambda_l^{\varepsilon_+,\varepsilon_-}\widetilde{P}_l^* u_l\|_{L^2(\widetilde{\mu}_l^n)}\leq C\|\widetilde{P}_l^* u_l\|_{L^2(\widetilde{\mu}_l^n)}=C\|u_l\|_{L^2(\mu_l^n)}.$$

This inequality is also a consequence of Lemma 33. So far we have proved that

$$\left|\|\mathcal{I}_l u_l\|^2_{L^2(\widetilde{\mu}_l^n)}-\|u_l\|^2_{L^2(\mu_l^n)}\right| \leq C\varepsilon_+\sqrt{b_l^{\varepsilon_+,\varepsilon_-}(u)}\|u_l\|_{L^2(\mu_l^n)}.$$

Next, we compare $\|\mathcal{I}_l u_l\|^2_{L^2(\widetilde{\mu}_l^n)}$ and $\|\mathcal{I}_l u_l\|^2_{L^2(\mu_l)}$, bounding their difference with

$$\left|\|\widetilde{\mathcal{I}}_l u_l\|^2_{L^2(\widetilde{\mu}_l^n)}-\|\widetilde{\mathcal{I}}_l u_l\|^2_{L^2(\mu_l)}\right| \leq C(\theta+\widetilde{\delta})\|\widetilde{\mathcal{I}}_l u_l\|^2_{L^2(\widetilde{\mu}_l^n)}\leq Cc_\rho(\theta+\widetilde{\delta})\|u_l\|^2_{L^2(\mu_l^n)},$$

as it follows from the fact that the difference between ρ_l and $\widetilde{\rho}_l^n$ is uniformly controlled with very high probability, i.e. i) in Corollary 25.

Finally, we obtain

$$\begin{aligned}\left|\|u_l\|^2_{L^2(\mu_l^n)}-\|\widetilde{\mathcal{I}}_l u_l\|^2_{L^2(\mu_l)}\right| &\leq \left|\|\widetilde{\mathcal{I}}_l u_l\|^2_{L^2(\widetilde{\mu}_l^n)}-\|u_l\|^2_{L^2(\mu_l^n)}\right|+\left|\|\widetilde{\mathcal{I}}_l u_l\|^2_{L^2(\widetilde{\mu}_l^n)}-\|\widetilde{\mathcal{I}}_l u_l\|^2_{L^2(\mu_l)}\right|\\ &\leq C\varepsilon_+\sqrt{b_l^{\varepsilon_+,\varepsilon_-}(u_l)}\|u_l\|_{L^2(\mu_l^n)}+Cc_\rho(\theta+\widetilde{\delta})\|u_l\|^2_{L^2(\mu_l^n)}.\end{aligned}$$

Adding over all $l=1,\dots,N$ and using Cauchy-Schwarz inequality we obtain the desired estimate:

$$\begin{aligned}\left|\|u\|^2_{L^2(\mu^n)}-\|\widetilde{\mathcal{I}}u\|^2_{L^2(\mu)}\right| &= \left|\sum_{l=1}^N\|u_l\|^2_{L^2(\mu_l^n)}-\sum_{l=1}^N w_l\|\widetilde{\mathcal{I}}_l u_l\|^2_{L^2(\mu_l)}\right|\\ &\leq C\varepsilon_+\sum_{l=1}^N w_l\sqrt{b_l^{\varepsilon_+,\varepsilon_-}(u)}\|u_l\|_{L^2(\mu_l^n)}+Cc_\rho(\theta+\widetilde{\delta})\sum_{l=1}^N w_l\|u_l\|^2_{L^2(\mu_l^n)}\\ &\leq C\varepsilon_+\|u\|_{L^2(\mu^n)}\sqrt{b^{\varepsilon_+,\varepsilon_-}(u)}+Cc_\rho(\theta+\widetilde{\delta})\|u\|^2_{L^2(\mu^n)}.\end{aligned}$$

■

A.5 Proof of Theorem 6

Proof [Proof of Theorem 6] With the aid of Propositions 26 and 27 we can now compare $\lambda_k^{\varepsilon_+,\varepsilon_-}$ and λ_k, the k-th eigenvalues of $\mathcal{L}$ and $\Delta_{\mathcal{M}}$ (listed according to multiplicity) respectively.

First, to find an upper bound for $\lambda_k^{\varepsilon_+,\varepsilon_-}$ in terms of λ_k, let $f^1,\dots,f^k$ be an orthonormal set (w.r.t. $L^2(\mu)$) consisting of eigenfunctions of $\Delta_{\mathcal{M}}$ corresponding to its first k eigenvalues (and let us label them $\lambda_1 \leq \cdots \leq \lambda_k$). Let

$$v_i := \widetilde{P} f_i, \forall i = 1,\dots,k.$$

Applying Proposition 27 to every f of the form

$$f := f_i - f_j,$$

we deduce that

$$|\langle f_i, f_j\rangle_{L^2(\mu)} - \langle v_i, v_j\rangle_{L^2(\mu^n)}| \leq C\widetilde{\delta}\sqrt{\lambda_k} + C(\theta + \widetilde{\delta}) < \frac{1}{k}.$$

We can then conclude that $v_1,\dots,v_k$ are linearly independent and that the subspace $S := Span\{v_1,\dots,v_k\}$ has dimension k. From (A.1) we deduce that

$$\lambda_k^{\varepsilon_+,\varepsilon_-} \leq \max_{v\in S, \|v\|_{L^2(\mu^n)}=1} b^{\varepsilon_+,\varepsilon_-}(v).$$

For $v \in S$, written as $v = \sum_{i=1}^k a_i v_i = \sum_{i=1}^k a_i \widetilde{P} f_i$, we can write $v := \widetilde{P} f$ where $f = \sum_{i=1}^k a_i f_i$. This f satisfies:

$$D(f) = \langle \Delta_{\mathcal{M}} f, f\rangle_{L^2(\mu)} \leq \lambda_k \|f\|^2_{L^2(\mu)}$$

according to the spectral decomposition of $\Delta_{\mathcal{M}}$. Applying part (2) of Proposition (26) we obtain:

$$\begin{aligned}
&b^{\varepsilon_+,\varepsilon_-}(v)\\
&\leq \frac{CNN_0}{w_{min}^2 n^2(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})}\left(1 + \lambda_k^{m/2+2}\right)\|f\|^2_{L^2(\mathcal{M})} + \left(1 + C(\varepsilon_+'' + \frac{\widetilde{\delta}}{\varepsilon_+})\right)\sigma_\eta D(f)\\
&\leq \frac{CNN_0}{w_{min}^2 n^2(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})}\left(1 + \lambda_k^{m/2+2}\right)\|f\|^2_{L^2(\mathcal{M})} + \left(1 + C(\varepsilon_+'' + \frac{\widetilde{\delta}}{\varepsilon_+})\right)\lambda_k\sigma_\eta \cdot \|f\|^2_{L^2(\mu)}.
\end{aligned}$$

Finally, from Proposition 27 applied to a $v \in S$ with norm one, we deduce that v's corresponding f satisfies:

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

From this we conclude that

$$\begin{aligned}
\frac{1}{\sigma_\eta}\lambda_k^{\varepsilon_+,\varepsilon_-} \leq{}& \frac{CNN_0}{w_{min}^2 n^2(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})}\left(1 + C'(\lambda_k^{m/2+2} + \widetilde{\delta}\sqrt{\lambda_k} + \theta + \widetilde{\delta})\right)\\
&+ \left(1 + C(\varepsilon_+'' + \widetilde{\delta}\sqrt{\lambda_k} + \theta + \frac{\widetilde{\delta}}{\varepsilon_+})\right)\lambda_k.
\end{aligned}$$

This establishes the upper bound for $\lambda_k^{\varepsilon_+,\varepsilon_-}$ in terms of λ_k.

For the lower bound, we follow completely analogous arguments as the ones above, relating functions $u \in L^2(\mu^n)$ with $f \in L^2(\mu)$ via the map $\mathcal{I}$ and applying Propositions 26 and 27.

■

A.6 Proof of Theorem 8

Proof [Proof of Theorem 8] We use an energy estimate based on Proposition 26 to find a relationship between eigenvectors of $\mathcal{L}^{\varepsilon_+,\varepsilon_-}$ and eigenfunctions of $\Delta_{\mathcal{M}}$. We follow a similar strategy to the one in García Trillos et al. (2019).

Let λ be an eigenvalue of $\Delta_{\mathcal{M}}$ and let $k \in \mathbb{N}$ be the first integer for which $\lambda = \lambda_k$ (here λ_k is as in (2.5)). Let l be the multiplicity of λ so that $\lambda = \lambda_k = \cdots = \lambda_{k+l-1} < \lambda_{k+l}$. The gap γ_λ associated to λ is given by:

$$\gamma_\lambda := \frac{1}{2}\min\{|\lambda - \lambda_{k-1}|, |\lambda - \lambda_{k+l}|\} \tag{A.21}$$

if $\lambda > 0$, and $\gamma_\lambda := \lambda_{l+1} = \lambda_{N+1}$ otherwise.

Now, we can pick $\varepsilon_+, \varepsilon_-, \theta, \widetilde{\delta}$ to be small enough so that

$$e + C\left(\varepsilon_+ + \theta + \widetilde{\delta}\right)\lambda \leq \gamma_\lambda$$

Then, for these choices of parameters, we know from Theorem 6 that

$$|\lambda^{\varepsilon_+,\varepsilon_-} - \sigma_\eta \lambda| \leq \gamma_\lambda \tag{A.22}$$

Let S be a subspace of $L^2(\mu^n)$ spanned by all eigenvectors of $\mathcal{L}^{\varepsilon_+,\varepsilon_-}$ with eigenvalues

$$\lambda_k^{\varepsilon_+,\varepsilon_-}, \cdots, \lambda_{k+l-1}^{\varepsilon_+,\varepsilon_-},$$

and let us denote the orthogonal projection onto S as P_S, the orthogonal projection onto the span of the eigenvectors of $\mathcal{L}$ with eigenvalue strictly smaller than $\lambda_k^{\varepsilon_+,\varepsilon_-}$ as P_{S_-}, and the orthogonal projection onto the span of the eigenvectors of $\mathcal{L}$ with eigenvalue strictly larger than $\lambda_{k+l-1}^{\varepsilon_+,\varepsilon_-}$ as P_{S_+}.

Let f be a normalized (w.r.t $L^2(\mu)$) eigenfunction of $\Delta_{\mathcal{M}}$ with eigenvalue λ and let $u = \widetilde{P}f$. Notice that we can assume without the loss of generality that f takes the form in (2.4) for one of the manifolds $\mathcal{M}_k$ (in particular the support of f is $\mathcal{M}_k$). Based on Proposition 26 and its proof (specifically the bound (A.20)) we have:

$$\begin{aligned}
e + \sigma_\eta\left[1 + C\left(\varepsilon_+ + \theta + \widetilde{\delta}\right)\right]\lambda &\geq e + \sigma_\eta\left[1 + C\left(\varepsilon_+ + \theta + \widetilde{\delta}\right)\right] D(f) \\
&\geq b^{\varepsilon_+,\varepsilon_-}(u) = \langle \mathcal{L}^{\varepsilon_+,\varepsilon_-} u, u\rangle \\
&\geq \lambda_k^{\varepsilon_+,\varepsilon_-}\|P_S u\|^2_{L^2(\mu^n)} + \lambda_{k+l}^{\varepsilon_+,\varepsilon_-}\|P_{S_+} u\|^2_{L^2(\mu^n)} \\
&\geq \lambda_k^{\varepsilon_+,\varepsilon_-}\left(\|u\|^2_{L^2(\mu^n)} - \|u - P_S u\|^2_{L^2(\mu^n)}\right) \\
&+ \lambda_{k+l}^{\varepsilon_+,\varepsilon_-}\left(\|u - P_S u\|^2_{L^2(\mu^n)} - \|P_{S_-} u\|^2_{L^2(\mu^n)}\right).
\end{aligned} \tag{A.23}$$

Using the results about γ_λ we obtained above and Proposition 27 we deduce

$$|\sigma_\eta\lambda_2 - \lambda_2^{\varepsilon_+,\varepsilon_-}| \leq e + C\sigma_\eta\left(\varepsilon_+ + \theta + \widetilde{\delta}\right)\lambda \leq \gamma_\lambda; \quad |1 - \|u\|^2_{L^2(\mu^n)}| \leq C(\theta + \widetilde{\delta})$$

Here C is some constant may correspond to λ. Combining the above inequalities with (A.23), we obtain:

$$\begin{aligned}
&e + \sigma_\eta\left[1 + C\sigma_\eta\left(\varepsilon_+ + \theta + \widetilde{\delta}\right)\right]\lambda \\
&\geq \sigma_\eta\lambda + (\lambda_2^{\varepsilon_+,\varepsilon_-} - \sigma_\eta\lambda_2) + \lambda_2^{\varepsilon_+,\varepsilon_-}(\|u\|^2_{L^2(\mu^n)} - 1) + (\lambda_{k+2}^{\varepsilon_+,\varepsilon_-} - \lambda_2^{\varepsilon_+,\varepsilon_-})\|u - P_S u\|^2_{L^2(\mu^n)} \\
&- \lambda_{k+l}^{\varepsilon_+,\varepsilon_-}\|P_{S_-}u\|^2_{L^2(\mu^n)} \\
&\geq \sigma_\eta\lambda - e - C\left(\varepsilon_+ + \theta + \widetilde{\delta}\right) + 2\gamma_\lambda\|u - P_S u\|^2_{L^2(\mu^n)} - \lambda_{k+l}^{\varepsilon_+,\varepsilon_-}\|P_{S_-}u\|^2_{L^2(\mu^n)}.
\end{aligned}$$

From this and the upper bound for $\lambda_{k+l}^{\varepsilon_-,\varepsilon_+}$ in terms of λ_{k+l}:

$$\|u - P_S u\|_{L^2(\mu^n)} \leq \left[\frac{e}{\gamma_\lambda} + \frac{C}{\gamma_\lambda}(\varepsilon_+ + \theta + \widetilde{\delta})\right]^{1/2} + \sqrt{\lambda_{k+l}}\|P_{S_-}u\|_{L^2(\mu^n)}.$$

We now compare the functions u and f at the data points x_i. Notice that for every data point $x_i \in \mathcal{M}_k$ we have:

$$|u(x_i) - f(x_i)| = |\widetilde{P}f(x_i) - f(x_i)| \leq n\int_{\widetilde{U}_{ik}} |f(x) - f(x_i)|\widetilde{\rho}_n(x)d\mathrm{vol}_{\mathcal{M}_k}(x) \leq \|\nabla f\|_{L^\infty(\mu_k)}\widetilde{\delta}.$$

Also, due to (A.20) we know that $\|\nabla f\|_{L^\infty(\mu_k)} \leq \sqrt{w_k}C(\mathcal{M}_k, \rho_k)(\lambda^{m/4+1}+1) \leq C(\mathcal{M}, \rho)(\lambda^{m/4+1}+1)$. Thus,

$$|u(x_i) - f(x_i)| \leq C(\mathcal{M}, \rho)(\lambda^{m/4+1} + 1)\widetilde{\delta}, \quad \forall x_i \in \mathcal{M}_k.$$

Notice that on the other hand, $u(x_i) = f(x_i) = 0$ for $x_i \in \mathcal{M} \setminus \mathcal{M}_k$ by definition of $\widetilde{P}$ and the fact that f is zero outside of $\mathcal{M}_k$. We conclude that:

$$\|u - f\|_{L^2(\mu^n)} \leq C_{\mathcal{M},\rho}(\lambda^{m/4+1} + 1)\widetilde{\delta},$$

and in turn

$$\|f - P_S\widetilde{P}f\|_{L^2(\mu^n)} \leq \left[\frac{e}{\gamma_\lambda} + \frac{C}{\gamma_\lambda}(\varepsilon_+ + \theta + \widetilde{\delta})\right]^{1/2} + C_{\mathcal{M},\rho}(\lambda^{m/4+1}+1)\widetilde{\delta} + \sqrt{\lambda_{k+l}}\|P_{S_-}\widetilde{P}(f)\|_{L^2(\mu^n)}. \tag{A.24}$$

From this point on the idea is to use an inductive argument. We describe in detail the base case and outline the inductive step. **Base Case:** When $\lambda = 0$ (and $\lambda_1 = \cdots = \lambda_N = 0 < \lambda_{N+1}$) we have $\|P_{S_-}\|_{L^2(\mu^n)} = 0$ and thus we can drop the last term in (A.24). This means that if $f_1, \cdots, f_l$ form an orthonormal basis for the space of eigenfunctions of $\Delta_{\mathcal{M}}$ with eigenvalue λ, then we can find an orthonormal set $v_1, \cdots, v_l$ spanning S such that

$$\|f_i - v_i\|_{L^2(\mu^n)} \leq \left[\frac{e}{\gamma_\lambda} + \frac{C}{\gamma_\lambda}(\varepsilon_+ + \theta + \widetilde{\delta})\right]^{1/2} + C(\mathcal{M}, \rho)\widetilde{\delta}.$$

In turn, this also implies that if $u_1, \cdots, u_l$ form an orthonormal basis of $\mathcal{L}^{\varepsilon_+,\varepsilon_-}$ with corresponding eigenvalues $\lambda_2^{\varepsilon_+,\varepsilon_-}, \cdots, \lambda_{l+1}^{\varepsilon_+,\varepsilon_-}$, then there exists an orthonormal set $\widetilde{f}_1, \cdots, \widetilde{f}_l$ for Δ_{ρ_l} with eigenvalue λ satisfying the same inequality above with f_i replaced with $\tilde{f}_i$ and v_i replaced with u_i.

Inductive step: having found the desired relationship for the eigenvectors and eigenfunctions associated to the first portion of the spectrum of $\Delta_{\mathcal{M}}$, we return to (A.24) and notice that by Proposition 27 we can conclude that the term $\|P_{S_-}\widetilde{P}f\|_{L^2(\mu^n)}$ is smaller than

$$C(\mathcal{M},\rho)((\lambda^{(1/4)}+1)\sqrt{\widetilde{\delta}}+\sqrt{\theta}).$$

We can plug this estimate in (A.24) and then proceed as in the base case to obtain the desired result.

■

A.7 Different dimensions: Proof of Theorem 10

We start by writing the discrete Dirichlet form $b^{\varepsilon_+,\varepsilon_-}$ (A.2) as the sum of three terms:

$$b^{\varepsilon_+,\varepsilon_-}(u) = b_{max}(u_{max}) + b_S(u_S) + b_O(u),$$

where

$$b_{max}(v) := \frac{1}{n^2(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})} \sum_{x_i,x_j\in\mathcal{X}_n\cap\mathcal{M}_{max}} \omega_{x_ix_j}(v(x_i)-v(x_j))^2, \quad v\in L^2(\mathcal{X}_n\cap\mathcal{M}_{max}),$$

$$b_S(u_S) := \frac{1}{n^2(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})} \sum_{k=N_{max}+1}^{N} \sum_{x_i,x_j\in\mathcal{X}_n\cap\mathcal{M}_k} \omega_{x_ix_j}(u_k(x_i)-u_k(x_j))^2, \quad u_S=(u_{N_{max}+1},\ldots,u_N),$$

and lastly,

$$b_O(u) := b^{\varepsilon_+,\varepsilon_-}(u) - b_{max}(u_{max}) - b_S(u_S).$$

Notice that b_{max} captures all interactions between points that belong to the manifolds with the maximum dimension m. For this energy we can use all the results presented in section 2.3 and in particular relate it to the Dirichlet form:

$$D_{max}(f) := \begin{cases} \sum_{i=1}^{N_{max}} w_i^2 \int_{\mathcal{M}_i} |\nabla f_i(x)|^2 \rho_i^2(x) d\mathrm{vol}_{\mathcal{M}_i}(x), & \text{if } f\in H^1(\mathcal{M}_{max}) \\ +\infty, \quad \text{if } f\in L^2(\mathcal{M}_{max})\setminus H^1(\mathcal{M}_{max}). \end{cases}$$

The energy b_S, on the other hand, captures the interactions between points that are on the same manifold when this manifold is not one of the ones with the largest dimension m. Using (A.5), we can write b_S as:

$$b_S(u_S) = \sum_{k=N_{max}+1}^{N} \left(\frac{n_k}{n}\right)^2 \cdot \left(\frac{\varepsilon_+^{m_k+2}-\varepsilon_-^{m_k+2}}{\varepsilon_+^{m+2}-\varepsilon_-^{m+2}}\right) \cdot b_k(u_k).$$

Finally, the term $b_O(u)$ accounts for all interactions between points in two different manifolds when the two manifolds are among the ones with dimension smaller than m, or when one of them has dimension m and the other one does not. In short, b_O accounts for all interactions not accounted for by the terms b_{max} and b_S and is thus a non-negative term.

We let $\widetilde{\mathcal{I}}_{max} : L^2(\mathcal{X}_n \cap \mathcal{M}_{max}) \to L^2(\mathcal{M}_{max})$ and $P_{max} : L^2(\mathcal{M}_{max}) \to L^2(\mathcal{X}_n \cap \mathcal{M}_{max})$ be the maps constructed in section A.2 applied to the data set $\mathcal{X}_n \cap \mathcal{M}_{max}$ and $\mathcal{M}_{max}$, i.e. the union of manifolds with the same dimension m. We also consider the following maps:

$$\mathcal{I}' : L^2(\mathcal{X}_n) \to L^2(\mathcal{X}_n \cap \mathcal{M}_{max})$$

$$\mathcal{I}' : u \longmapsto u_{max},$$

$$P' : L^2(\mathcal{X}_n \cap \mathcal{M}_{max}) \to L^2(\mathcal{X}_n)$$

$$P' : v \longmapsto u = (v, 0),$$

where by $u = (v, 0)$ we mean that u coincides with v for data points in $\mathcal{M}_{max}$ and $u = 0$ for data points in $\mathcal{M} \setminus \mathcal{M}_{max}$.

It will be convenient to introduce the norms:

$$\|u_k\|^2_{L^2(\mathcal{X}_n \cap \mathcal{M}_k)} := \frac{1}{n} \sum_{x_i \in \mathcal{X}_n \cap \mathcal{M}_k} (u_k(x_i))^2, \quad u_k \in L^2(\mathcal{X}_n \cap \mathcal{M}_k),$$

and

$$\|v\|^2_{L^2(\mathcal{X}_n \cap \mathcal{M}_{max})} := \frac{1}{n} \sum_{x_i \in \mathcal{X}_n \cap \mathcal{M}_{max}} (v(x_i))^2, \quad v \in L^2(\mathcal{X}_n \cap \mathcal{M}_{max}),$$

as well as the discrete Laplacians:

$$\mathcal{L}_k u_k(x) := \frac{1}{n_k(\varepsilon_+^{m_k+2} - \varepsilon_-^{m_k+2})} \sum_{y \in \mathcal{X}_n \cap \mathcal{M}_k} \omega_{xy}(u(x) - u(y)), \quad x \in X \cap \mathcal{M}_k, \quad u : X \to \mathbb{R}.$$

We use $\lambda_{2,k}^{\varepsilon_+,\varepsilon_-}$ to denote the second eigenvalue of $\mathcal{L}_k$.

Proof [Proof of Theorem 10]

Following the structure of the proofs of Theorems 6 and 8 we see that we can obtain our desired estimates if we can obtain similar inequalities to the ones in Propositions 26 and 27 where now we use the maps $\mathcal{I}_{max} \circ \mathcal{I}'$ and $P' \circ P_{max}$ as interpolation and discretization maps respectively. There is only one small caveat in the almost isometry property of $\mathcal{I}_{max} \circ \mathcal{I}'$ as we explain below.

We start by noticing that from the above definitions we have:

$$b_{max}(\mathcal{I}'u) \leq b^{\varepsilon_+,\varepsilon_-}(u), \quad \forall u \in L^2(\mathcal{X}_n),$$

and by Proposition 26

$$\sigma_\eta D_{max}(\mathcal{I}_{max} \circ \mathcal{I}'u) \leq \left(1 + C(\varepsilon_+ + \frac{\widetilde{\delta}}{\varepsilon_+} + \theta + \widetilde{\delta})\right) b_{max}(\mathcal{I}'u),$$

so that

$$\sigma_\eta D_{max}(\mathcal{I}_{max}\circ\mathcal{I}'u)\leq\left(1+C(\varepsilon_+ +\frac{\widetilde{\delta}}{\varepsilon_+}+\theta+\widetilde{\delta})\right)b^{\varepsilon_+,\varepsilon_-}(u),\quad\forall u\in L^2(\mathcal{X}_n). \tag{A.25}$$

The above occurs with probability at least $1-\sum_{l=1}^{N}(nw_l+t)\exp\left(-\mathrm{C}(nw_l-t)\theta^2\widetilde{\delta}^m\right)-2N\exp\left(\frac{-2t^2}{n}\right)-C_1(n)$.

On the other hand, for arbitrary $f\in L^2(\mathcal{M}_{max})$ we have

$$\begin{aligned}b^{\varepsilon_+,\varepsilon_-}(P'\circ P_{max}f)&=b_{max}(P_{max}f)+b_O(P'\circ P_{max}f)\\&\leq\left(1+C(\varepsilon_+ +\frac{\widetilde{\delta}}{\varepsilon_+}+\theta+\widetilde{\delta})\right)D_{max}(f)+b_O(P'\circ P_{max}f),\end{aligned} \tag{A.26}$$

whereas

$$b_O(P'\circ\widetilde{P}_{max}f)\leq\frac{CNN_0}{w_{min}^2n^2(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})}\left(1+\lambda^{m/2+2}\right)\|f\|^2_{L^2(\mathcal{M}_{max})}, \tag{A.27}$$

for f an element in the span of $\Delta_{\mathcal{M}_{max}}$'s eigenfunctions with corresponding eigenvalue less than λ, as it follows from a completely analogous computation to the one in (A.19); this holds in the same event of very high probability where (A.25) holds.

We consider now the norm distortion of the maps $\mathcal{I}\circ\mathcal{I}'$ and $P'\circ P_{max}$. First, notice that by definition, for $v\in L^2(\mathcal{X}_n\cap\mathcal{M}_{max})$ we have

$$\|P'v\|^2_{L^2(\mathcal{X}_n)}=\|v\|^2_{L^2(\mathcal{X}_n\cap\mathcal{M}_{max})},$$

and thus combining with 1) in Proposition (27) we obtain:

$$\left|\|P'\circ P_{max}f\|^2_{L^2(\mathcal{X}_n)}-\|f\|^2_{L^2(\mathcal{M}_{max})}\right|\leq C\widetilde{\delta}\|f\|_{L^2(\mathcal{M}_{max})}\sqrt{D_{max}(f)}+C(\theta+\widetilde{\delta})\|f\|^2_{L^2(\mathcal{M}_{max})}. \tag{A.28}$$

Now, for a given $u\in L^2(\mathcal{X}_n)$ we have:

$$\left|\|\mathcal{I}'u\|^2_{L^2(\mathcal{X}_n\cap\mathcal{M}_{max})}-\|u\|^2_{L^2(\mathcal{X}_n)}\right|=\sum_{k=N_{max}+1}^{N}\|u_k\|^2_{L^2(\mathcal{X}_n\cap\mathcal{M}_k)}.$$

Also, if we let $\overline{u}_k$ represent the average of u_k in $\mathcal{M}_k\cap\mathcal{X}_n$ we see that

$$\|u_k-\overline{u}_k\|^2_{L^2(\mathcal{X}_n\cap\mathcal{M}_k)}\leq\frac{1}{\lambda_{2,k}^{\varepsilon_+,\varepsilon_-}}\langle\mathcal{L}_ku_k,u_k\rangle_{L^2(\mathcal{M}_k\cap\mathcal{X}_n)}=\frac{1}{\lambda_{2,k}^{\varepsilon_+,\varepsilon_-}}\frac{n_k^2}{n^2}b_k(u_k)\leq C(\mathcal{M}_k,w_k,\rho_k)\varepsilon_+^{m-m_k}b(u),$$

for all $k=N_{max}+1,\dots,N$, where the last inequality holds with very high probability. Indeed, notice that by Theorem 6 applied to a single manifold $\mathcal{M}_k$ we can find a lower bound for $\lambda_{2,k}^{\varepsilon_+,\varepsilon_-}$ in terms of the first non-trivial eigenvalue for $w_k\Delta_{\mathcal{M}_k}$. We have also used the fact that $b_k(u_k)\leq(n/n_k)^2\varepsilon_+^{m-m_k}b(u)$. This means that

$$\left|\|\mathcal{I}'u\|^2_{L^2(\mathcal{X}_n\cap\mathcal{M}_{max})}-\|u\|^2_{L^2(\mathcal{X}_n)}\right|\leq C(\mathcal{M},\mu)\varepsilon_+^{m-m_{N_{max}+1}}b^{\varepsilon_+,\varepsilon_-}(u)+\sum_{k=N_{max}+1}^{N}(\overline{u}_k)^2.$$

Combining with Proposition 27 and using the triangle inequality we deduce that

$$\begin{aligned}\left|\|\mathcal{I}_{max}\circ\mathcal{I}'u\|^2_{L^2(\mathcal{M}_{max})}-\|u\|^2_{L^2(\mathcal{X}_n)}\right| &\leq C\varepsilon_+\|u\|_{L^2(\mu^n)}\sqrt{b^{\varepsilon_+,\varepsilon_-}(u)}\\ &\quad + C(\theta+\widetilde{\delta})\|u\|^2_{L^2(\mu^n)} + C(\mathcal{M},\mu)\sum_{k=N_{max}+1}^{n}\varepsilon_+^{m-m_k}b^{\varepsilon_+,\varepsilon_-}(u)\\ &\quad + \sum_{k=N_{max}+1}^{N}(\overline{u}_k)^2.\end{aligned} \tag{A.29}$$

Notice that the right hand side in the above expression is small for a u with low Dirichlet energy only when u is close to the orthogonal complement of $\text{Span}\{\mathbb{1}_{\mathcal{M}_{N_{max}+1}},\dots,\mathbb{1}_{\mathcal{M}_N}\}$ (i.e. the $\overline{u}_k$ are small). Because of this, we will only be able to proceed as in the proofs of Theorems 6 and 8 to obtain all our estimates if first we show that the top N eigenvectors of $\mathcal{L}$ are close to the indicator functions of $\mathcal{M}_1\cap\mathcal{X}_n,\dots,\mathcal{M}_N\cap\mathcal{X}_n$. However, this is straightforward from the following observations:

1. We can obtain an upper bound for the first N eigenvalues of $\mathcal{L}$ following the representation (A.1) and computing the graph Dirichlet energy of the indicator functions of the sets $\mathcal{M}_k\cap\mathcal{X}_n$. Namely, we have:

$$\lambda_k^{\varepsilon_+,\varepsilon_-}\leq\frac{CNN_0}{w_{min}^2n^2(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})},\quad k=1,\dots,N.$$

2. Using the alternative representation:

$$\lambda_{N+1}^{\varepsilon_+,\varepsilon_-}=\max_{S\in\mathcal{G}_N}\min_{u\in S^\perp\setminus\{0\}}\frac{b^{\varepsilon_+,\varepsilon_-}(u)}{\|u\|^2_{L^2(\mu^n)}}$$

   we can obtain the lower bound

$$\lambda_{N+1}^{\varepsilon_+,\varepsilon_-}\geq\frac{1}{2}\sigma_\eta\lambda_{N+1},$$

   with very high probability. Indeed, taking $S=\text{Span}\{\mathbb{1}_{\mathcal{M}_1\cap\mathcal{X}_n},\dots,\mathbb{1}_{\mathcal{M}_N\cap\mathcal{X}_n}\}$ and a unit norm $u\in S$ (in particular $\overline{u}_k=0$ for all $k=N_{max}+1,\dots,N$) we see from (A.25) and (A.29) that

$$b^{\varepsilon_+,\varepsilon_-}(u)\geq\sigma_\eta\lambda_{N+1}\left(1-C(\varepsilon_++\varepsilon_+\sqrt{\lambda_{N+1}}+\varepsilon_+^{m-m_{N_{max}+1}}\lambda_{N+1}+\theta+\frac{\widetilde{\delta}}{\varepsilon_+})\right)\geq\frac{\sigma_\eta}{2}\lambda_{N+1}.$$

3. Combining the previous steps we get an order one lower bound for the gap between $\lambda_N^{\varepsilon_+,\varepsilon_-}$ and $\lambda_{N+1}^{\varepsilon_+,\varepsilon_-}$. We can then follow the proof of Theorem 8 to show that there exists an orthonormal set $v^1,\dots,v^N$ consisting of eigenvectors of $\mathcal{L}$ corresponding to $\mathcal{L}$'s first N eigenvalues such that

$$\|\sqrt{\frac{n}{n_k}}\mathbb{1}_{\mathcal{M}_k\cap\mathcal{X}_n}-v^k\|^2_{L^2(\mathcal{X}_n)}\leq\frac{C(\mathcal{M},\mu)N_0}{n^2(\varepsilon_+^{m+2}-\varepsilon_-^{m+2})}.$$

We deduce that if u belongs to the orthogonal complement of $\mathrm{Span}\{v^1, \ldots, v^N\}$, then

$$(\overline{u}_k)^2 \leq \frac{C(\mathcal{M}, \mu) N_0}{n^2(\varepsilon_+^{m+2} - \varepsilon_-^{m+2})},$$

with very high probability.

As discussed above, with the above estimates in hand we can now proceed as in the proofs of Theorems 6 and 8.

■