Robust Multi-Manifold Clustering via Simplex Paths

Haoyu Chen   Haoyu.Chen@utah.edu
Department of Mathematics
University of Utah
Salt Lake City, UT 84112, USA

Anna Little   little@math.utah.edu
Department of Mathematics
University of Utah
Salt Lake City, UT 84112, USA

Akil Narayan   akil@sci.utah.edu
Department of Mathematics
Scientific Computing and Imaging (SCI) Institute
University of Utah
Salt Lake City, UT 84112, USA

Editor: My editor

Abstract

This article introduces a novel, geometric approach for multi-manifold clustering (MMC), i.e. for clustering a collection of potentially intersecting, d-dimensional manifolds into the individual manifold components. We first compute a locality graph on d-simplices, using the dihedral angle in between adjacent simplices as the graph weights, and then compute infinity path distances in this simplex graph. This procedure gives a metric on simplices which we refer to as the *largest angle path distance* (LAPD). We analyze the properties of LAPD under random sampling, and prove that with an appropriate denoising procedure, this metric separates the manifold components with high probability. We validate the proposed methodology with extensive numerical experiments on both synthetic and real-world data sets. These experiments demonstrate that the method is robust to noise, curvature, and small intersection angle, and generally out-performs other MMC algorithms. In addition, we provide a highly scalable implementation of the proposed algorithm, which leverages approximation schemes for infinity path distance to achieve quasi-linear computational complexity.

Keywords: Path-based metric, manifold learning, graph theory, clustering

1 Introduction

Analyzing high-dimensional data poses significant challenges, especially when the number of dimensions exceeds the sample size. In such settings, traditional statistical methods often fail, and standard algorithms become computationally impractical. However in many real-world applications, data tends to concentrate near low-dimensional structures, which can be exploited to mitigate the curse of dimensionality. In the simplest case, data may cluster along a single low-dimensional subspace or manifold. A more flexible and realistic model assumes that data lies on multiple low-dimensional structures. By identifying and treating

each structure individually, one can obtain a sparse and efficient representation of the data, and the first step in this process is to assign data points to their respective structures. This leads to the central focus of this article: the multi-manifold clustering (MMC) problem, in which data points are drawn from a collection of manifolds, and the objective is to recover the manifold labels up to permutation.

A closely related problem to MMC is (linear) subspace clustering, which simplifies the general model by assuming that the underlying manifolds are linear subspaces. This problem has been widely studied, with notable methods including Elhamifar and Vidal (2013); Liu et al. (2012); Vidal et al. (2016); Bradley and Mangasarian (2000); Vidal (2011); Ho et al. (2003); Tipping and Bishop (1999); Costeira and Kanade (1998); Wright et al. (2009); Zhang et al. (2015), among others. Many of the most successful approaches rely on the self-expressive property, in which each data point is represented as a linear combination of other data points. Ideally, points from the same subspace have large coefficients in this representation, while points from different subspaces have small coefficients. The resulting coefficient matrix is then transformed into an affinity matrix, and standard clustering algorithms such as K-means or spectral clustering are then applied to obtain the final segmentation. Subspace clustering methods differ in their computational efficiency and in how they enforce self-representation. Broadly, they fall into three main categories: (i) sparse or low-rank representation methods, such as SSC Elhamifar and Vidal (2013) and LRR Liu et al. (2012), which use convex optimization to enforce self-representation; (ii) function or polynomial fitting methods, which cast the problem as learning a parametric model in which algebraic constraints can be used to segment the data (Li et al., 2022a; Vidal et al., 2016); (iii) deep learning-based methods, which employ autoencoders, self-expressive layers, and adversarial training to learn feature representations that reveal subspace structures (Ji et al., 2017; Peng et al., 2020; Zhou et al., 2018).

However, the assumption that data lies on a collection of linear subspaces is often too restrictive for real-world datasets, which frequently exhibit nonlinear structures that are better modeled as multiple manifolds. Such a multi-manifold modeling assumption has been proven useful for motion segmentation and image classification (Souvenir and Pless, 2005; Ye and Zhao; Hastie and Simard, 1997; Basri and Jacobs, 2003; Tomasi and Kanade, 1992), for dynamic pattern identification and optical flow (Gong et al., 2012; Wang et al., 2015), and for classification of text on webpages and in news stories (Zong et al.; Babaeian et al., 2015). Popular manifold clustering methods (Arias-Castro et al., 2017; Souvenir and Pless, 2005; Elhamifar and Vidal, 2013; Babaeian et al., 2015; Wang et al., 2015; Tenenbaum et al., 2000; Belkin and Niyogi, 2003) thus extend the existing algorithms for subspace clustering to handle nonlinear data structures. The recent survey, Abdolali and Gillis (2021), taxonomizes manifold clustering methods into three categories: (i) locality preserving, (ii) kernel based, and (iii) neural network based. Locality preserving methods (Arias-Castro et al., 2017; Elhamifar and Vidal, 2013; Babaeian et al., 2015; Tenenbaum et al., 2000; Belkin and Niyogi, 2003; Trillos et al., 2021) focus on maintaining the local geometric structure of the manifolds when mapping to a lower-dimensional space, ensuring that the new representation reflects the true underlying structure. Kernel based methods leverage kernel functions to map data into a higher-dimensional space where linear subspace methods can be applied (Patel and Vidal, 2014). Neural network-based approaches to manifold clustering (Wu et al., 2022a,b; McConville et al., 2021; Li et al., 2022b) leverage

deep learning models to learn meaningful data embeddings that facilitate clustering. A central technique in this category is autoencoder-based clustering, where an autoencoder compresses the data into a lower-dimensional latent space, and clustering is then performed on this learned representation. Our approach is a locality preserving method which leverages local angle information to define an appropriate distance/affinity.

Naturally, the major challenge in the MMC problem is that the underlying manifolds may intersect (otherwise density-based clustering approaches would be sufficient). Previous locality preserving methods have utilized local tangent plane approximation as well as curvature and angle constraints to mitigate this challenge (Arias-Castro et al., 2017; Babaeian et al., 2015; Trillos et al., 2021; Wang et al., 2015). Our approach avoids explicitly estimating such geometric structures by letting the algorithm itself explore the local information, requiring only knowledge of the intrinsic data dimension d. In particular, we compute a graph of local d-dimensional simplices, weighted according to the dihedral angle in between simplices, and then compute an infinity path distance inside the resulting weighted graph. We note that the computation of infinity path distances is well studied in the literature, where it is encountered under many names, including the maximum capacity path problem, the bottleneck edge query problem, the widest path problem, and the *longest leg path distance* (Little et al., 2020; Hu, 1961; Camerini, 1978; Gabow and Tarjan, 1988). Note Little et al. (2020) analyzed paths of edges weighted by Euclidean distance, instead of simplex paths weighted by angle, but we use the same approximation scheme as Little et al. (2020) to estimate infinity path distances in a scalable manner. Empirical results indicate that clustering based on this angle based metric can recover the manifold components with high accuracy. In addition, our method also has the following advantages over the state-of-the-art approaches (i) it can reliably learn the number of manifolds, which is often a required input by other MMC algorithms; (ii) it can be implemented in computational complexity that is quasi-linear with respect to the sample size, while most other methods are at least quadratic; and (iii) it is more robust to noise and curvature than competing MMC methods. An overview of the proposed method (without analysis) can be found in the conference paper Chen et al. (2023). This article analyzes the proposed metric under random sampling, derives theoretical guarantees for solving MMC, and includes much more extensive numerical experiments (with analysis of real data, not done in Chen et al. (2023)).

The remainder of the paper is organized as follows. Section 2 introduces the problem setup and assumptions and provides an overview of our proposed approach. Section 3 gives a more formal presentation of the methodology including the construction of the LAPD metric from a simplex graph and a novel denoising procedure which enhances manifold separability. Section 4 gives theoretical guarantees that the method can solve MMC with high probability, quantifying the impact of noise and intersection angle (quantifying impact of curvature is left to future work). Section 5 discusses the computational complexity of the algorithm and presents extensive experiments on both synthetic and real data investigating the accuracy, robustness, and scalability of the proposed method. Finally, Section 6 summarizes the article and suggests directions for future research; additional details and technical proofs are provided in the appendices.

2 Problem Setup and Method Overview

Let $\mathcal{M} := \{\mathcal{M}_j\}_{j\in[m]}$ denote a collection of m compact, Riemannian manifolds embedded in $\mathbb{R}^D$, where each $\mathcal{M}_j$ has intrinsic dimension $d \ll D$. We denote by $T_\tau(\mathcal{M}_j) := \{x \in \mathbb{R}^D : \exists z \in \mathcal{M}_j$ such that $\|z - x\| \le \tau\}$ a tube of radius τ around manifold $\mathcal{M}_j$, and $T_\tau(\mathcal{M}) := \cup_{j\in[m]} T_\tau(\mathcal{M}_j)$ the union of these tubes. We think of $T_\tau(\mathcal{M})$ as a collection of noisy manifolds (with noise level τ), and we assume access to a dataset $X = \{x_i\}_{i\in[n]}$ consisting of i.i.d. samples from $T_\tau(\mathcal{M})$. Let $z_i = \arg\min_{y\in\mathcal{M}} \|y - x_i\|$ denote the closest point on $\mathcal{M}$ to data point x_i, i.e. z_i is the noiseless counterpart of x_i. We let $\{\ell_i\}_{i\in[n]}$ denote the underlying manifold labels, i.e. $\ell_i = j$ if the clean sample $z_i \in \mathcal{M}_j$, and our goal is to recover the labels ℓ_i up to permutation from the noisy data X. We are primarily interested in the case when manifolds intersect, and we let $\Theta_{ij} \in (0, \frac{\pi}{2}]$ denote the intersection angle between two intersecting manifolds $\mathcal{M}_i$ and $\mathcal{M}_j$. When the context is clear, we shall omit the manifold index subscript and simply denote the intersection angle between two manifolds as Θ. To simplify the analysis, we make the following assumptions.

Assumption 2.1 (Intersection dimension) *If $\mathcal{M}_i \cap \mathcal{M}_j \ne \emptyset$, then $\mathcal{M}_i \cap \mathcal{M}_j$ has dimension $d - 1$.*

Assumption 2.2 (Uniform sampling) *The samples $\{x_i\}_{i\in[n]}$ are iid draws from the uniform probability measure P on $T_\tau(\mathcal{M})$, i.e. $dP = \frac{dV}{V}$ for dV the natural volume measure on $T_\tau(\mathcal{M})$ and $V = vol(T_\tau(\mathcal{M}))$.*

We remark that Assumptions 2.1 and 2.2 are for convenience, but not fundamentally required by our method. For example, all our theoretical results still hold when sampling from a nonuniform probability measure on $T_\tau(\mathcal{M})$; as long as the corresponding density ρ satisfies $0 < \rho_{\min} \le \rho \le \rho_{\max}$, only constants will be affected. Furthermore, Assumption 2.1 can be thought of as a "worst case" scenario, as we expect success in this regime will guarantee success for intersections with dimension smaller than $d - 1$.

Next we summarize our approach to solving this problem, postponing many technical details to Section 3. We start by identifying a locality graph $\mathcal{G}_X = (X, E_X)$ on the nodes X, where edges are defined through an annular nearest neighbor computation. The inner radius of the annulus is chosen to lie above the noise level, and the outer radius of the annulus is chosen to maintain a feasible computational complexity. From this node-level graph, we construct a set of d-simplices S, where a simplex $\Delta \in S$ if, among the $d + 1$ vertices of Δ, there is one vertex that is $\mathcal{G}_X$-connected to the remaining d vertices. We remove from S simplices of bad "quality", i.e., those that are very elongated.

From the set of d-simplices, we say that $\Delta_i, \Delta_j \in S$ are *adjacent* if they share a common face, i.e., the number of shared vertices is d, $|\Delta_i \cap \Delta_j| = d$. When two such simplices are adjacent, we can compute a dihedral angle between them, $\theta_{ij} := \theta(\Delta_i, \Delta_j) \in (0, \pi]$. The set of all such intersections and angles from adjacent pairs in S identifies a *simplex-level* graph $\mathcal{G}_S = (S, E_S, W_S)$, where E_S contains pairs of simplices that are adjacent, and the corresponding weights W_S are defined by their intersection angle θ, i.e. $W_S(\Delta_i, \Delta_j) = \pi - \theta_{ij}$. Intuitively, W_S encodes the angle dissimilarity between adjacent simplices, since when the simplices form a flat connection, $\theta_{ij} \approx \pi$ and $W_S(\Delta_i, \Delta_j) \approx 0$. From the simplex graph $\mathcal{G}_S = (S, E_S, W_S)$ created with angle-based weights, we derive an angle-based metric

(a) Simplex construction for $d = 2$.   (b) Two sample paths for $d = 2$.

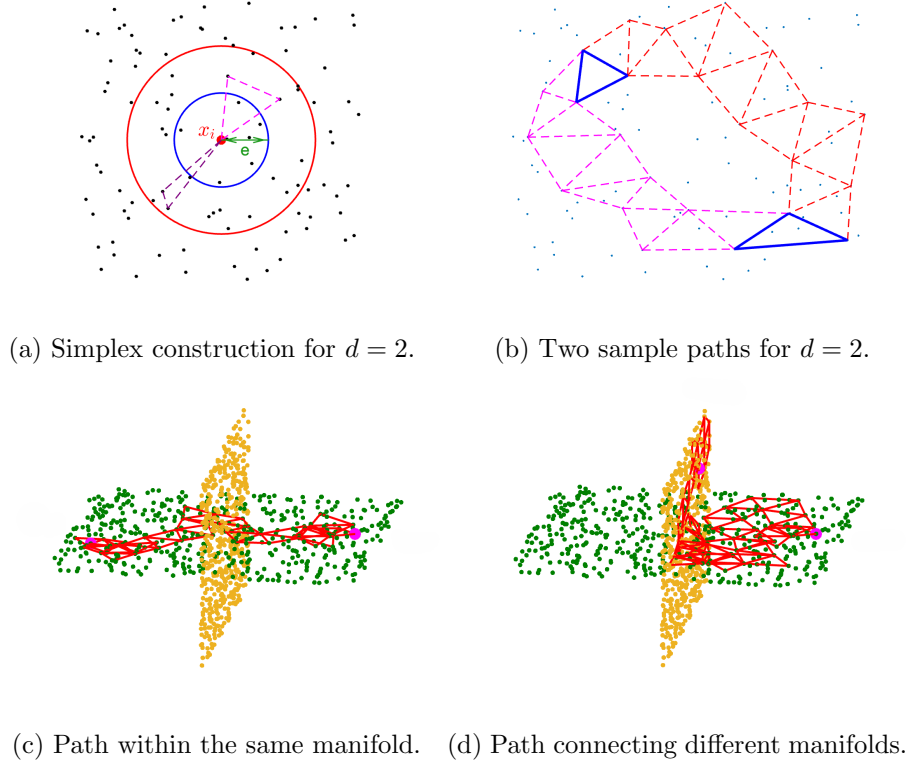(c) Path within the same manifold.   (d) Path connecting different manifolds.

Figure 1: Construction of simplices and simplex paths connected by adjacent simplices over a pair of linear, 2-dimensional intersecting manifolds.

between all simplices in S by computing the infinity shortest path distance in $\mathcal{G}_S$; we refer to this metric as the *largest angle path distance (LAPD)*. After incorporating an appropriate denoising procedure, we perform distance-based clustering with LAPD to recover a partition of S into m clusters (generally the parameter m can also be learned during this step). Finally, the partition of the simplex set S can be extended to a partition of the node set X by a majority vote procedure (i.e. to cluster data point x_i, we assign the most common label in the set of simplices containing x_i).

Figure 1 illustrates the key ideas of our approach. Figure 1a illustrates the construction of the simplex set S: for each data point x_i, we define local simplices by the various subsets formed by combining x_i with d of its annular neighbors (when $d = 2$, one obtains triangles as shown). We then consider paths of connected simplices as shown in Figure 1b: two simplices are generally connected by a large number of different paths, but the best path according to the LAPD metric is the flattest path, i.e. the path which avoids making any sharp turns, since according to LAPD, the length of a path is the single worst angle encountered along the path. Figures 1c and 1d illustrate why the LAPD metric is natural for solving MMC: Figure 1c illustrates that any two simplices from the same manifold can always be connected with a flat path, while Figure 1d shows that every path of simplices between manifolds must always make a sharp turn somewhere along the path. Our theoretical results establish that under reasonable assumptions, the LAPD between simplices on the same manifold is always

small, while the LAPD between simplices on different manifolds is always large, so that the metric can successfully differentiate the manifold components. Informally, the assumptions needed to guarantee success are (1) the noise level τ is not too large, (2) the curvature of the manifolds is not too large, and (3) the intersection angles of the manifolds are not too small; these conditions guarantee that simplices can be constructed at the right scale for solving MMC, i.e. one needs simplices to be large relative to the noise but small relative to the manifold curvature.

3 Methodology

We now describe the details of our proposed algorithm. Section 3.1 describes the construction of the original locality graph and the derived set of local simplices S. Section 3.2 describes the creation of the weighted graph $\mathcal{G}_S$ on S with angle-based weights and the computation of the LAPD shortest path distances in $\mathcal{G}_S$. Section 3.3 discusses a procedure for de-noising the LAPD metric and distance-based clustering with LAPD.

3.1 Construction of Simplices

We first fix a minimal edge length e, which will encode the scale at which the metric is constructed. For each data point x_i, we remove any neighbors within distance e and then connect x_i to its B nearest neighbors among the remaining points. We let $\mathcal{G}_X(e, B) = (X, E_X)$ denote the corresponding unweighted annular locality graph, which for brevity we denote simply as $\mathcal{G}_X$. We then form an initial set of candidate simplices $\mathcal{S}$ by taking subsets of $d + 1$ points where at least one point is $\mathcal{G}_X$ connected to all remaining points, i.e.:

$$\mathcal{S} = \{\Delta : \Delta \subseteq X, |\Delta| = d + 1, \exists x \in \Delta \text{ such that } E_X(x, y) > 0 \, \forall \, y \in \Delta \setminus x\}. \quad (1)$$

As our method is more robust when we use regularly shaped simplices, we define the following measures of simplex distortion:

$$\text{Distortion}_1(\Delta = \{x_1, \ldots, x_{d+1}\}) = \frac{\min_{i \neq j} \|x_i - x_j\|}{\max_{i \neq j} \|x_i - x_j\|},$$

$$\text{Distortion}_2(\Delta = \{x_1, \ldots, x_{d+1}\}) = \frac{\text{vol}_d(\Delta)}{V_0 \min_{i \neq j} \|x_i - x_j\|^d},$$

where $\text{vol}_d(\Delta)$ is the d-volume of simplex Δ and $V_0 := \frac{d+1}{d! \sqrt{2^d}}$ is the d-volume of a regular simplex with unity edge length. Note Distortion_1 measures irregularity of edge lengths while Distortion_2 measures irregularity of volume, i.e. the deviation of the d-volume of any Δ from that of a regular simplex. We then let $0 < q < 1$, $0 < r_0 < 1$ be parameters controlling Distortion_1, Distortion_2, and define the set of *valid simplices* S by

$$S = \{\Delta \in \mathcal{S} : \text{Distortion}_1(\Delta) \geq q, \text{Distortion}_2(\Delta) \geq r_0\}.$$

By construction, the set of valid simplices S satisfies the following assumption:

Assumption 3.1 *(Valid simplices: size and distortion)* *Let S be a set of simplices constructed from samples of $T_\tau(\mathcal{M})$. We assume S satisfies size/distortion constraint encoded*

by (e, q, r_0), *i.e. there exist constants* $e > 0$, $q \in (0, 1)$, *and* $r_0 \in (0, 1)$ *such that every simplex* $\Delta = \{x_1, \ldots, x_{d+1}\} \in S$ *satisfies*

$$e \leq \|x_i - x_j\| \leq \frac{e}{q} \quad \text{for all } i, j \in [d+1], \ i < j, \tag{2}$$

$$r_0 V_0 e^d \leq \text{vol}_d(\Delta) \leq V_0 \left(\frac{e}{q} \right)^d. \tag{3}$$

Thus four key parameters determine the set of valid simplices: e, q, r_0, and B; e constrains the size, while (q, r_0) constrains the shape of allowable simplices. Reasonable and fixed values of q in (2), say $q \approx 1/2$, are not sufficient to guarantee (3) uniformly in d, so we introduce r_0 explicitly to enforce this constraint. Since the cardinality of S is $O(nB^d)$, B constrains the scalability of the proposed algorithm. Figure 1a illustrates the construction of S: e is the radius of the inner circle, B is the number of points in the annulus; the elongated purple simplex is a candidate simplex, but clearly does not survive filtering if for example $q > \frac{1}{2}$. Theorems in Section 4 provide some heuristics for tuning these parameters. Furthermore, our theoretical results will require that the noise level is bounded by the allowable edge length:

Assumption 3.2 *(Small noise) The noise level τ is small compared to the size of valid simplices, i.e. $\tau \leq \frac{e}{q}$.*

3.2 Weighted Graph $\mathcal{G}_S$ and LAPD

Having established a set of valid simplices S, we can proceed to define the weighted graph $\mathcal{G}_S = (S, E_S, W_S)$ by connecting adjacent simplices in S with angle-based weights. Two simplices Δ_i and Δ_j are adjacent if they share a common face, i.e. $|\Delta_i \cap \Delta_j| = d$. For every pair of adjacent simplices Δ_i, Δ_j, we can compute the dihedral angle $\theta_{ij} := \theta(\Delta_i, \Delta_j) \in (0, \pi]$ between them as follows. We let x_i, x_j denote the apex points in Δ_i, Δ_j (i.e. the points which are not shared), and let v_i be the unique vector connecting $\Delta_i \cap \Delta_j$ and x_i which is normal to $\Delta_i \cap \Delta_j$, and similarly for v_j. Then θ_{ij} is simply the angle in between v_i, v_j. More specifically, we fix any basepoint $c \in \Delta_i \cap \Delta_j$ (for example, the centroid) and define:

$$v_i = (x_i - c) - P_{\Delta_i \cap \Delta_j}(x_i - c)$$
$$v_j = (x_j - c) - P_{\Delta_i \cap \Delta_j}(x_j - c)$$
$$\theta_{ij} = \arccos \left(\frac{\langle v_i, v_j \rangle}{\|v_i\| \cdot \|v_j\|} \right),$$

where $P_{\Delta_i \cap \Delta_j}$ denotes projection onto the span of $\{x_s - c : x_s \in \Delta_i \cap \Delta_j\}$, which is $d - 1$ dimensional by Assumption 2.1. That is, $P_{\Delta_i \cap \Delta_j} = QQ^{\mathsf{T}}$ with Q being the orthonormal basis of the shared face $\Delta_i \cap \Delta_j$.

When $\theta_{ij} \geq \frac{\pi}{2}$, we connect Δ_i and Δ_j in the graph $\mathcal{G}_S$ and define the edge weight between them by

$$W_S(\Delta_i, \Delta_j) = \pi - \theta_{ij}. \tag{4}$$

Note W_S encodes an angle-based dissimilarity measure on S.[1]

---

1. For numerical implementation, in our code we define $W_S(\Delta_i, \Delta_j) = \max\{\pi - \theta_{ij}, \delta\}$ for a small numerical tolerance parameter $\delta = 10^{-8}$; this ensures $W_S(\Delta_i, \Delta_j) > 0$ if and only if Δ_i, Δ_j are connected in $\mathcal{G}_S$.

When Δ_i, Δ_j form a flat connection, $\theta_{ij} \approx \pi$, and W_S is small; if however the simplices connect at a sharp angle, W_S will be large. We note that some of our experiments utilize a two-sided version of the above weight construction defined in (16). Our core Definition 3.1 of LAPD defines a metric on S by computing the infinity-shortest path distance in $\mathcal{G}_S$, i.e. it uses the local angle dissimilarities encoded in W_S to define a global metric on all of S.

Definition 3.1 (LAPD) *Let* $\mathcal{P} = \mathcal{P}(\Delta_i, \Delta_j)$ *be the set of all simplex paths connecting* Δ_i, Δ_j *in* $\mathcal{G}_S$*, i.e. any* $\{\Gamma_1, \ldots, \Gamma_L\} \in \mathcal{P}$ *satisfies* $\Gamma_1 = \Delta_i, \Gamma_L = \Delta_j$*, and* $E_S(\Gamma_t, \Gamma_{t+1})$ *exists for all* $1 \le t \le L-1$*. Then the largest angle path-distance (LAPD) between* Δ_i, Δ_j *is:*

$$LAPD(\Delta_i, \Delta_j) := \min_{\{\Gamma_1, \ldots, \Gamma_L\} \in \mathcal{P}} \max_{1 \le t \le L-1} W_S(\Gamma_t, \Gamma_{t+1}). \tag{5}$$

We remark that LAPD looks for the single *shortest* path connecting Δ_i, Δ_j in $\mathcal{G}_S$ (thus the minimum over all paths $\mathcal{P}$), where the length of the path is determined by the single largest angle encountered along the path (thus the maximum over the path segments). Thus LAPD is small if there exists a flat path connecting two simplices (this should be the case if Δ_i, Δ_j originate on the same manifold; see Figure 1c), while the LAPD is large if every path connecting Δ_i, Δ_j takes a sharp turn along the way (this should be the case if Δ_i, Δ_j originate on different manifolds; see Figure 1d).

To quantify this observation, we define the maximal within manifold LAPD in Definition 3.2 and the minimal between manifolds LAPD in Definition 3.3. We let $S(\mathcal{M}_j)$ denote the elements $\Delta = \{x_1, \ldots, x_{d+1}\}$ of S for which $z_i \in \mathcal{M}_j$ for all $1 \le i \le d+1$, i.e. the underlying clean samples z_i all lie on the same manifold $\mathcal{M}_j$; we refer to such simplices as *pure* simplices. Those simplices that are not pure, $S \setminus \cup_{j \in [m]} S(\mathcal{M}_j)$, are referred to as *mixed* simplices.

Definition 3.2 (Within LAPD) *Let* S *be the collection of valid simplices constructed from samples of* $T_\tau(\mathcal{M})$*. Then the maximal within manifold LAPD is:*

$$wLAPD := \max_{1 \le j \le m} \max_{\Delta_1, \Delta_2 \in S(\mathcal{M}_j)} LAPD(\Delta_1, \Delta_2). \tag{6}$$

Definition 3.3 (Between LAPD) *Let* S *be the collection of valid simplices constructed from samples of* $T_\tau(\mathcal{M})$*. Then the minimal between manifold LAPD is:*

$$bLAPD := \min_{i \ne j} \min_{\Delta_1 \in S(\mathcal{M}_i), \Delta_2 \in S(\mathcal{M}_j)} LAPD(\Delta_1, \Delta_2). \tag{7}$$

Solving MMC using the LAPD metric becomes tenable precisely when there is a significant gap between the maximal within and minimal between LAPD, i.e. when wLAPD $\ll$ bLAPD. In Section 4 we will derive high probability bounds controlling wLAPD and bLAPD (see Theorems 4.1, 4.4, 4.5), which lead to conditions guaranteeing a good gap between wLAPD and bLAPD (see Theorem 4.7, which is the main theoretical contribution of this article).

Figure 2 illustrates the nature of the LAPD metric for two linear, 2-dimensional noiseless manifolds intersecting at an angle of $\frac{\pi}{2}$. The color corresponds to the LAPD from the base simplex (shown in pink on the horizontal manifold in (a) and (c)) to all of the other simplices. As expected, the distance to all other simplices on the same manifold is small
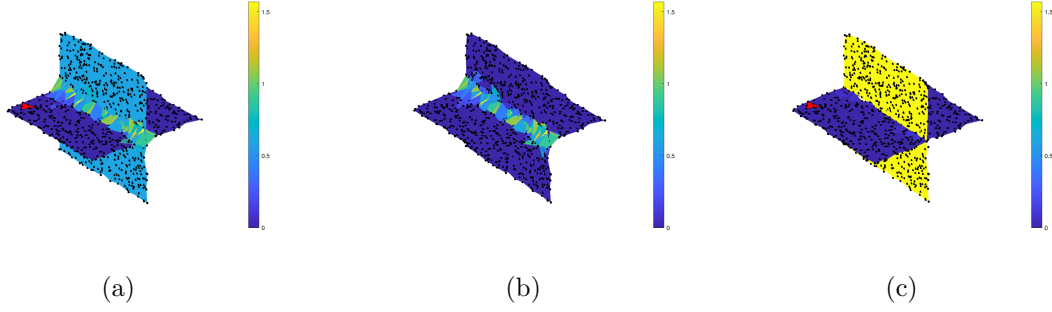
(a)   (b)   (c)

Figure 2: (a) LAPD (indicated by colors) to the target simplex (pink) before de-noising; (b) LAPD of each simplex to its 70$^{\text{th}}$ nearest neighbor; (c) LAPD (indicated by colors) to the target simplex (pink) after de-noising. Black dots represent the data points.

(essentially zero), while the distance to pure simplices on the vertical manifold is large in comparison (approximately 0.6) in Figure 2a. The picture is much less clear for mixed simplices formed from points close to the intersection, which have a highly variable LAPD to the base simplex. In Section 3.3 we will introduce a denoising procedure which significantly increases the LAPD gap by removing most of these mixed simplices. The effectiveness of the denoising procedure can be verified in Figure 2c.

3.3 Denoising and Clustering

As illustrated in Figure 2b, mixed simplices around the intersection decrease the bLAPD and are difficult to cluster, as they can be equidistant from both manifolds. Fortunately, it is generally straightforward to identify these simplices, as they tend to have significantly larger nearest neighbor LAPD distances than the pure simplices. We thus propose the following denoising procedure: we fix a number of nearest neighbors κ and a threshold η, and we discard from S all simplices whose LAPD to their κ^{th} nearest neighbor exceeds η. This procedure is described mathematically in the following definition.

Definition 3.4 (Denoised Simplices) *Given κ and η, we compute the LAPD between a simplex $\Delta \in S$ to its κ-th nearest neighbor as*

$$\kappa NN(\Delta) = \min_{N \subseteq S, |N| = \kappa} \max_{\square \in N} LAPD(\Delta, \square) \tag{8}$$

and define the set of denoised simplices as

$$S_{dns} = \{\Delta \in S : \kappa NN(\Delta) \leq \eta\}. \tag{9}$$

After this denoising procedure, we repeat the LAPD construction outlined in Section 3.2, but simply replace S with S_{dns}, i.e. LAPD is computed in $\mathcal{G}_{S_{\text{dns}}}$. Figures 2b and 2c illustrate the proposed denoising procedure. Figure 2b shows the LAPD of each simplex to its $\kappa = 70^{\text{th}}$ nearest neighbor; this κNN distance is small for the pure simplices, but large for the mixed simplices around the intersection, and thus can be used to identify and remove these problematic simplices. Figure 2c shows that after denoising, the recomputed LAPD

from the base simplex to the vertical manifold is now approximately 1.5, on the same order as the intersection angle $\frac{\pi}{2}$ between the manifolds. We will analyze the impact of denoising mathematically in Theorem 4.5.

The last step of our algorithm is to use the LAPD metric to partition the denoised simplex set S_{dns}, and then infer a partition on the node set X to solve MMC. We let D_{LAPD} denote the $|S_{\mathrm{dns}}| \times |S_{\mathrm{dns}}|$ LAPD distance matrix. Given the LAPD distances, there are many available algorithms for clustering including hierarchical clustering with various linkage functions, density-based clustering, and spectral clustering. However we wish to utilize a method which can avoid explicit construction of the D_{LAPD} matrix, which is quite large $(O(|S_{\mathrm{dns}}|^2))$ and not sparse.

We thus utilize the approach developed in Little et al. (2020), which offers a fast, multi-scale procedure to approximate single linkage hierarchical clustering of infinity path distances with a controlled error. Briefly, given a sequence of thresholds $t_1 < t_2 < \cdots < t_k$, we build approximating graphs $\mathcal{G}(t_1), \ldots, \mathcal{G}(t_k)$ such that each $\mathcal{G}(t_i)$ for $i \in [k]$ is constructed by discarding edge weights of magnitude greater than t_i in $\mathcal{G}_{S_{\mathrm{dns}}}$. LAPD can then be approximated simply by computing the connected components in this sequence of graphs, since if Δ_1, Δ_2 are in different connected components at scale t_i but the same connected component at scale t_{i+1}, then the true LAPD distance between them is between t_i and t_{i+1}. The connected components of the multi-scale graphs thus provide a collection of k slices of the full LAPD single linkage dendrogram; see Little et al. (2020) for more details, including an efficient implementation of spectral clustering on infinity path distances (for simplicity we simply utilize hierarchical clustering in this article). Table 1 illustrates this multi-scale procedure on a toy example of 16 simplices; 8 scales (appearing in the first row) are used to approximate the metric, and the table shows the indices of the connected components at each scale. The simplices start in distinct connected components but eventually merge into a single connected component at the last scale, provided the graph is fullly connected.

An important question when performing distance-based clustering is how to choose the number of clusters m (indeed, we see different numbers of clusters at different scales). When this parameter is unknown, we approximate the number of clusters using a persistence approach, i.e. letting $m(t_1), \ldots, m(t_k)$ denote the number of nontrivial connected components of $\mathcal{G}(t_1), \ldots, \mathcal{G}(t_k)$, we define

$$c_j = \sum_{i=1}^{k} \mathbf{1}\{m(t_i) = j\}, \quad \hat{m} = \arg\max_{j} c_j. \tag{10}$$

For example, we would infer $\hat{m} = 2$ in the toy example shown in Table 1. When m is known, we simply return the hierarchical clustering corresponding to that m. This heuristic of identifying the number of clusters with the longest branches is well known (Belyadi et al., 2021; Gere, 2023) and aligns with the theory of the gap statistic (Tibshirani et al., 2002) and the variance ratio criterion (Caliński and JA, 1974). One advantage of our proposed method is that it can usually learn the correct number of manifolds m in the mixture, while the methods we compare with require this parameter as an input (see the experiments in Section 5).

Finally, we extend the simplex partition to a node partition by a majority vote procedure. More specifically, letting $\ell(\Delta) \in [m]$ denote the simplex partition and $S_{\mathrm{dns}}(x_i)$ the subset

scales simplices	1e-8	3e-7	1.4e-5	6.3e-4	0.028	0.3418	0.8866	1.5707
Δ_1	1	1	1	1	1	1	1	1
Δ_2	2	1	1	1	1	1	1	1
Δ_3	3	1	1	1	1	1	1	1
Δ_4	4	2	2	2	1	1	1	1
Δ_5	5	2	2	2	1	1	1	1
Δ_6	6	3	3	2	1	1	1	1
Δ_7	7	3	3	2	1	1	1	1
Δ_8	8	3	3	2	1	1	1	1
Δ_9	9	4	4	3	2	2	2	1
Δ_{10}	10	4	4	3	2	2	2	1
Δ_{11}	11	5	4	3	2	2	2	1
Δ_{12}	12	5	4	3	2	2	2	1
Δ_{13}	13	5	4	3	2	2	2	1
Δ_{14}	14	6	5	3	2	2	2	1
Δ_{15}	15	6	5	3	2	2	2	1
Δ_{16}	16	7	5	3	2	2	2	1

Table 1: The merging procedure in Little et al. (2020) to approximate hierarchical clustering. A toy example of 16 simplices and 8 scales are used here to demonstrate the idea. Each cell contains the index of the connected component at that scale. Simplices start in distinct components but eventually merge into a single connected component at the last scale.

of simplices containing x_i, we define the label assigned to a node x_i as

$$\ell(x_i) = \arg\max_{j \in [m]} \sum_{\Delta \in S_{\mathrm{dns}}(x_i)} \mathbf{1}\{\ell(\Delta) = j\}. \tag{11}$$

Alternatively, one could define a node-level distance by incorporating an additional minimum over all pairs of simplices containing the nodes x_i, x_j; namely,

$$\mathrm{nLAPD}(x_i, x_j) := \min_{\Delta_1 \in S_{\mathrm{dns}}(x_i), \, \Delta_2 \in S_{\mathrm{dns}}(x_j)} \mathrm{LAPD}(\Delta_1, \Delta_2), \tag{12}$$

and then cluster the nodes using nLAPD. Experiments indicate that clustering the nodes with nLAPD is quite similar to clustering the simplices with LAPD and then inferring a partition on the nodes. However, the additional minimum takes an $O(n^2)$ search and is thus not computationally feasible; we thus utilize the simplex level LAPD for clustering.

4 Main Results

In this section we analyze the proposed LAPD metric, bound wLAPD and bLAPD with high probability (Theorems 4.1, 4.4, and 4.5), and derive conditions guaranteeing a good LAPD gap (wLAPD $\ll$ bLAPD), so that the metric can successfully solve MMC with

high probability (Theorem 4.7). To simplify the analysis, we consider the following two components, linear model.

Assumption 4.1 *(Linear manifolds)* $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$ *for two linear manifolds* $\mathcal{M}_1, \mathcal{M}_2$ *intersecting at angle* Θ.

We discuss extensions to the general case at the end of this section.

We start by deriving an upper bound on wLAPD in terms of the noise level τ and simplex edge length e; Theorem 4.1 is in fact not a probabilistic statement, but derived from a deterministic "worst case" bound on how much bending can occur for pure simplices, i.e. simplices constructed from a single manifold. The key idea is that if the noise level τ is small relative to the simplex size e, then all pure simplices will be nearly parallel to the underlying manifold, and wLAPD will be small.

Theorem 4.1 (wLAPD upper bound) *Let Assumptions 2.1, 2.2, 3.1, 3.2 and 4.1 hold. Assume also the construction outlined in Section 3 with either weight construction* (4) *or* (16)*, and that* $\mathcal{G}_S$ *is connected. Then:*

$$wLAPD \lesssim \frac{\tau}{e}. \tag{13}$$

Proof See Appendix A. ∎

Remark 4.2 *The upper bound in* (13) *is independent of the ambient dimension D; in particular,* $wLAPD \le C_{d,q}\frac{\tau}{e}$ *for a constant $C_{d,q}$ depending on the intrinsic dimension d and simplex quality q.*

Remark 4.3 *Our goal is for $wLAPD \lesssim \Theta$, i.e. for the within manifold distance to be small relative to the manifold intersection angle, as this will be necessary to achieve a good LAPD gap;* (13) *guarantees this will be the case as long as $e \gtrsim \frac{\tau}{\Theta}$, i.e. we choose a scale e for the simplex construction which is large relative to the noise level τ and the inverse of the intersection angle.*

Next we establish an upper bound on bLAPD, the between manifold distance. Theorem 4.4 is derived from probabilistic arguments that when sampling randomly it is unlikely to obtain a long chain of mixed simplices near the intersection which all connect at a flat angle, i.e. the formation of a flat bridge in between manifolds is unlikely. We introduce the following non-restrictive assumption in order to give a precise bound on this distance.

Assumption 4.2 *There exists a constant $R_{\max}$ such that for all $R \in [\tau, R_{\max}]$*

$$vol_D(\mathcal{T}_R) \in [C_1, C_2]R\tau^{D-d}, \tag{14}$$

where $\mathcal{T}_R = T_R(\mathcal{M}_1 \cap \mathcal{M}_2) \cap T_\tau(\mathcal{M})$ denotes the subset of the data support within distance R of $\mathcal{M}_1 \cap \mathcal{M}_2$.

The constants C_1, C_2 encode the elongation of $\mathcal{M}$ in the direction of the intersection; in particular, C_1, C_2 will scale as $\text{vol}_{d-1}(\mathcal{M}_1 \cap \mathcal{M}_2)$. Describing $\text{vol}_D(\mathcal{T}_R)$ with an interval is convenient, since if the boundaries of $\mathcal{M}_1, \mathcal{M}_2$ curve, $\text{vol}_D(\mathcal{T}_R)/R$ will have small variations as R changes.

Theorem 4.4 (bLAPD lower bound) *Let Assumptions 2.1, 2.2, 3.1, 3.2, 4.1, and 4.2 hold and assume the construction outlined in Section 3 with either weight construction (4) or (16). If e satisfies $ne^d \geq \frac{2(d+1)}{C_5 \Theta \log 2} \log n$ and $e \leq 1$, then for n large enough*

$$bLAPD \geq (2C_5 ne^d)^{-1}$$

with probability at least $1 - 3C_4 n^{-(d+1)}$, where $C_4 = \left(\frac{2C_2}{q \sin \Theta} \right)^{d+1}$, $C_5 = \frac{2C_2 C_{\Theta,q,d}}{C_1}$, C_1, C_2 are the constants from (14), and C_{Θ,q,d,r_0} is the constant from Proposition B.4.

Proof See Appendix B. A key technical step is the single link volume estimate given in Proposition B.4, which is proved in Appendix D. ∎

Theorem 4.4 guarantees that when there is no noise, wLAPD $= 0 \ll (\log n)^{-1} \lesssim$ bLAPD, and the metric can differentiate the manifolds; this is obtained by choosing e as small as possible subject to the lower bound $ne^d \gtrsim \log n$. However when observations are corrupted by noise and the wLAPD can become as large as $O(\frac{\tau}{e})$ as quantified in Theorem 4.1, the metric would only be able to differentiate the manifolds for very small values of τ, namely $\tau \lesssim (\log n)^{-\frac{(d-1)}{d}} n^{-\frac{1}{d}}$. Fortunately, if we incorporate an appropriate denoising procedure, we can increase the bLAPD significantly without compromising the wLAPD. In particular, by incorporating the denoising procedure from Definition 3.4, one can obtain a bLAPD scaling like the intersection angle Θ; despite the much stronger conclusion, the proof of Theorem 4.5 is a minor modification of the proof of Theorem 4.4.

Theorem 4.5 (bLAPD lower bound with denoising) *Let Assumptions 2.1, 2.2, 3.1, 3.2, 4.1, and 4.2 hold and assume the construction outlined in Section 3 with either weight construction (4) or (16). Assume also that e satisfies $ne^d \geq \frac{8(d+1)}{C_5 \Theta \log 2} \log n$, $e \leq 1$ and that we remove all simplices whose LAPD to their $\kappa = \frac{\Theta}{4} C_5 ne^d$-th nearest neighbor exceeds $\eta = (2C_5 ne^d)^{-1}$. Then for n large enough, if LAPD is recomputed after denoising,*

$$bLAPD_{dns} \geq \frac{\Theta}{4}$$

with probability at least $1 - 3C_4 n^{-(d+1)}$, where constants C_4, C_5 are as in Theorem 4.4.

Proof See Appendix C. ∎

Remark 4.6 *Again, the best possible result is obtained by choosing $e \sim \left(\frac{\log n}{n} \right)^{\frac{1}{d}}$ so that $\kappa \sim \log n$ and $\eta \sim (\log n)^{-1}$; in this setting the bLAPD is proportional to Θ with high probability.*

Finally, we present the main result of the article, which combines Theorems 4.1 and 4.5 to obtain a condition guaranteeing that LAPD can solve MMC. We remark that the denoising procedure will not impact the wLAPD upper bound computed in Theorem 4.1 as long as the graph stays connected.

Theorem 4.7 (LAPD gap) *Let Assumptions 2.1, 2.2, 3.1, 3.2, 4.1, and 4.2 hold and assume the construction outlined in Section 3 with either weight construction (4) or (16). In particular, assume S_{dns} is obtained by removing all simplices whose κ^{th} nearest neighbor LAPD exceeds η for $\kappa \sim ne^d$, $\eta \sim (ne^d)^{-1}$, and that $\mathcal{G}_{S_{dns}}$ is connected. For n large enough, if e satisfies*

$$\underbrace{\frac{\tau}{\Theta}}_{wLAPD} \vee \underbrace{\left(\frac{\log n}{n}\right)^{\frac{1}{d}}}_{bLAPD} \lesssim e \leq 1, \tag{15}$$

then wLAPD $\lesssim$ bLAPD with probability $1 - O(n^{-(d+1)})$.

Proof By Theorem 4.1, wLAPD $\lesssim \frac{\tau}{e}$. Applying Theorem 4.5, bLAPD $\gtrsim \Theta$ as long as $ne^d \geq \frac{8(d+1)}{C_5 \Theta \log 2} \log n$ holds, i.e. for $e \gtrsim \left(\frac{\log n}{n}\right)^{\frac{1}{d}}$ which is guaranteed by (15). We thus obtain wLAPD $\lesssim \frac{\tau}{e} \lesssim \Theta \lesssim$ bLAPD. ∎

A few observations are in order: Theorem 4.7 illustrates that for LAPD to successfully differentiate the manifolds, simplices must be contructed at a lengthscale which is above the noise, and a small intersection angle Θ leads to a smaller range of parameters where the approach succeeds. The upper bound $e \leq 1$ is quite non-restrictive (assumed for convenience in the proof of Theorem 4.4), as throughout the article we consider e as a local scaling. Nonetheless when manifolds are curved, the curvature will require a much more restrictive upper bound on the simplex size e than the one appearing in Theorem 4.7. When the curvature is large, the wLAPD can become large where the manifold bends rapidly, and simplices must be kept small to mitigate this. We do not compute this curvature-based upper bound explicitly, as a precise characterization of the volumes arising in our chaining arguments is already cumbersome even in the linear case (see Proposition B.4). From an algorithmic perspective, the strategy is already clear: take e as small as possible while still satisfying (15) will allow the algorithm to tolerate as much curvature as possible.

5 Numerical Implementation & Experiments

This section provides implementation details and experimental results. In particular, Subsection 5.1 presents pseudo-code for the LAPD algorithm, analyzes the computational complexity, and discusses a variant of the weight definition in (4) which enhances performance in some settings. We then conduct numerous experiments investigating the performance of LAPD and compare against two benchmark subspace clustering methods SSC-ADMM (Elhamifar and Vidal, 2011) and EKSS (Lipor et al., 2020) and three benchmark manifold clustering methods LocPCA (Arias-Castro et al., 2017), PBC (Babaeian et al., 2015), and DCV (Wu et al., 2022b). Subsection 5.2 reports results for subspace clustering on synthetic

(a) Default η for a noiseless dataset.

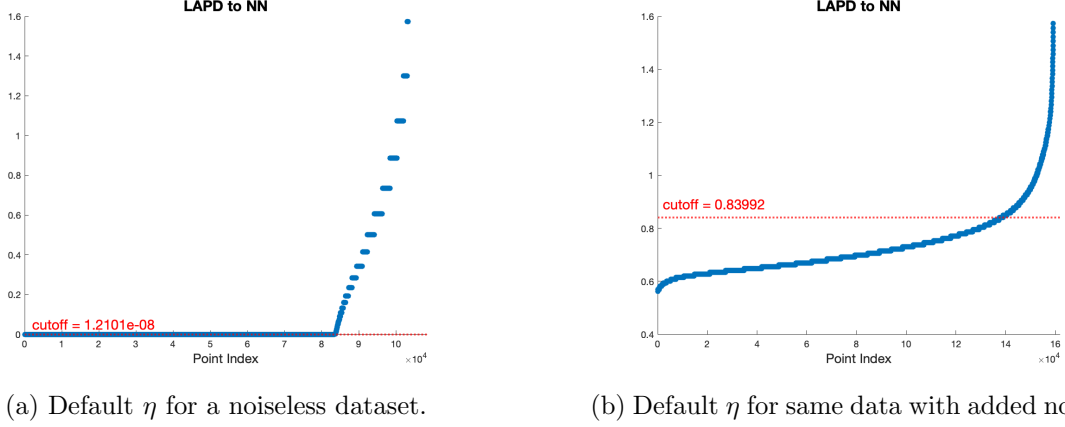(b) Default η for same data with added noise.

Figure 3: Default η (cutoff) from the elbow point criteria; figure on the left is for noiseless data, and figure on the right is for the same data with added noise.

data, Subsection 5.3 reports results for nonlinear multimanifold clustering on synthetic data, and Subsection 5.4 reports performance on real data.

5.1 Numerical Implementation

The pseudo-code of the LAPD algorithm is given in Algorithm 1. Our implementation[2] provides default values to the parameters ($B = 25$, $e = \sqrt{2}\tau$, $q = \frac{1}{1.25+0.15(d-2)}$, $\kappa = 10\log n$, $k = 100$) to make the code easy to use, and estimates the intrinsic dimension d and noise level τ using Little et al. (2017) if not provided by the user. Note that the choice of B, k is to keep computation and memory use feasible, the choice of e is to ensure the simplex size to be above the noise level, and the choice of q and κ is empirical. The default value of η is set to be the elbow point when all the κNN LAPDs of the simplices are sorted in a non-decreasing order. An illustration is given in Figure 3. Careful readers may find that the parameter set of Algorithm 1 does not include r_0, the one that constraints the volume of the simplices. The reason is quite simple: keeping track of the volume of each single simplex is computationally unfeasible and empirically, we observe that characterizing the quality of simplices with e, q is sufficient to guarantee a good performance of the algorithm.

Based on the pseudo-code, the total complexity of LAPD as an algorithm for MMC can be calculated as follows: (i) searching and finding nearest neighbors on dataset X takes $O(Dn\log(n))$ using KD-Tree; (ii) constructing the weighted simplex graph takes $nB^d\frac{d(d+1)}{2}D = O(Dnd^2B^d)$; (iii) using the multiscaled graph procedure from Little et al. (2020) to approximate a pruned single linkage dendrogram takes

$$O\left(ndB^{d+1}\left(t \wedge \log\left(ndB^{d+1}\right)\right) + nB^dt\log\left(nB^d\right)\right)$$

(iv) denoising takes $(nB^d)\kappa = O(dB^dn\log(n))$. Putting all the pieces together gives the total complexity of the LAPD algorthm

$$O\left(\left(D + dB^d\right)n\log n + \left(Dd^2 + dB\left(k \wedge \log\left(ndB^{d+1}\right)\right)\right)nB^d\right) = O\left(\left(D + B^d\right)n\log n + DnB^d\right),$$

---

2. https://github.com/HYfromLA/LAPD

Algorithm 1: LAPD for Multi-Manifolds Clustering.

Input: Data $X = \{x_1, \ldots, x_n\}$; Optional parameters $d, \tau, B, e, q, \kappa, \eta, k, m$.
Output: Clustering labels on X.

1 Estimate intrinsic dimension d and noise level τ using the algorithm from Little et al. (2017) if not provided by user.

2 Construct annular locality graph $\mathcal{G}_X(e, B) = (X, E_X)$ on X and candidate simplices via (1).

3 Filter simplices with quality parameter q via (2) to obtain simplex set S.

4 Construct an angle-based weighted graph on S: $\mathcal{G}_S \leftarrow (S, E_S, W_S)$ with $W_S(\Delta_i, \Delta_j)$ from (4) or (16) dependent on data distribution.

5 Approximate LAPD in $\mathcal{G}_S$ from k scales using the multi-scale procedure in Little et al. (2020).

6 Denoise S according to (8) and (9) with thresholds κ, η to obtain S_{dns} and re-compute LAPD in $\mathcal{G}_{S_{\mathrm{dns}}}$.

7 Use the multi-scale procedure Little et al. (2020) to approximate hierarchical clustering on S_{dns}.

8 If unknown, estimate number of clusters m via (10) (persistence) and extract corresponding labels.

9 Extend the labels assigned to S_{dns} to labels on X via a majority vote (see (11)).

---

assuming $d, k, B = O(1)$, i.e. the LAPD algorithm is linear in the ambient dimension D and quasi linear in the dataset size n. We verify these computational dependencies with experiments (see Figure 5), while most of the benchmark algorithms are at least quadratic. On the other hand, the complexity of LAPD is exponential in the intrinsic dimension d, both memory-wise and computation-wise, which makes it hard to use for datasets of large intrinsic dimension (normally when $d > 5$).

We also introduce the following weight definition:

$$W_S^2(\Delta_i, \Delta_j) = \min\{\pi - \theta(\Delta_i, \Delta_j), \theta(\Delta_i, \Delta_j)\} \tag{16}$$

as an alternative to (4). We use the superscript 2 to distinguish this second version of weight definition from (4). Indeed, (16) allows "backflipping", i.e. if two simplices connect with a dihedral angle close to 0, this is also deemed a flat connection, while (4) requires that the dihedral angle be close to π for a flat connection. If we refer to the LAPD defined with (4) as LAPD^1 and respectively the LAPD defined with (16) as LAPD^2, then it is straightforward to see that

$$\mathrm{LAPD}^1(\Delta_i, \Delta_j) \geq \mathrm{LAPD}^2(\Delta_i, \Delta_j), \tag{17}$$

since all simplices are more connected (closer) under LAPD^2. Using LAPD^2 thus decreases the wLAPD (better within-manifold cohesiveness) at the cost of also decreasing the bLAPD (worse between manifold separability), and it is a priori unclear what is most advantageous for multi-manifold clustering, i.e. which metric will lead to the largest LAPD gap. Although our theoretical guarantees cover both constructions, we found LAPD^1 to be preferable for the synthetic data sets considered in Sections 5.2 and 5.3 while LAPD^2 was preferable for the real data sets considered in Section 5.4. We conjecture that this is because in our synthetic data sets, the data density is high at the intersections, while for the real data sets the data density near class intersections tends to be lower (clusters are more separated or even disjoint); when the intersections are of high density, the

(a) $d = 1, \Theta = \frac{\pi}{2}$ (b) $d = 1, \Theta = \frac{\pi}{4}$ (c) $d = 1, \Theta = \frac{\pi}{6}$

(d) $d = 2, \Theta = \frac{\pi}{2}$ (e) $d = 2, \Theta = \frac{\pi}{4}$ (f) $d = 2, \Theta = \frac{\pi}{6}$
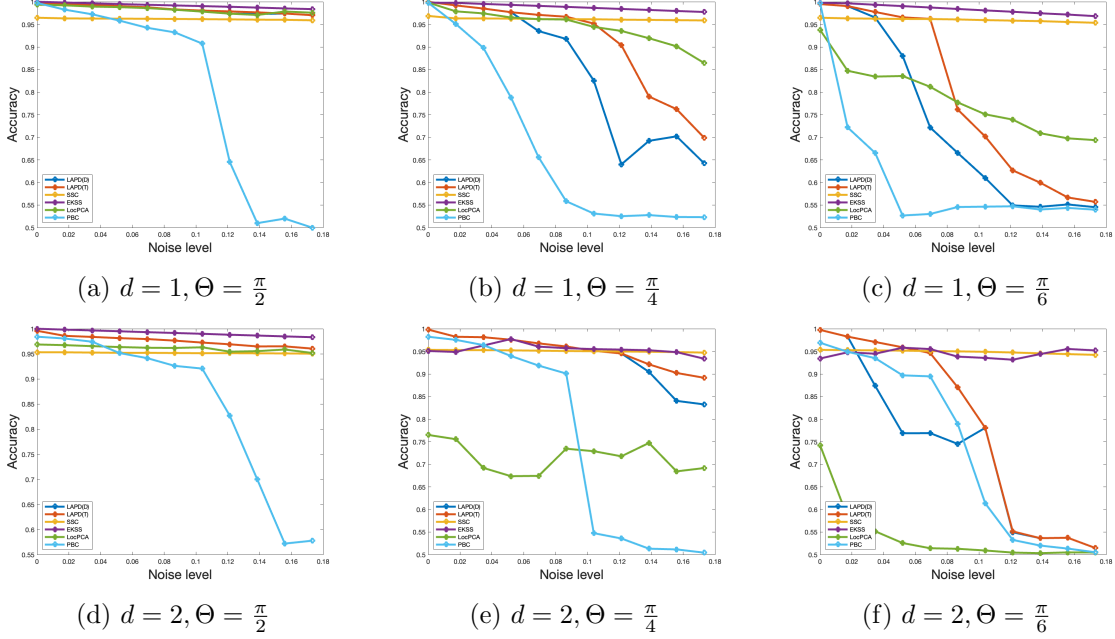
Figure 4: Subspace clustering accuracy.

risk of a bad path forming is greatly increased by allowing backflipping. Note if no superscript is specified, LAPD refers to the default LAPD[1].

5.2 Subspace Clustering

We first evaluate LAPD on a linear subspace clustering task using synthetic data. Each dataset consists of a pair of intersecting unit hypercubes: for $d = 1$ or $d = 2$, the data represent intersecting unit vectors or squares, respectively. We fix the dataset size at $n = 6,000$, with 3,000 points drawn from each subspace, and set the ambient dimension to $D = 100$. To introduce noise, we add a uniform random variable in each coordinate from the interval $[-\frac{\sqrt{3}\,\sigma}{\sqrt{D-d}}, \frac{\sqrt{3}\,\sigma}{\sqrt{D-d}}]$ resulting in a global noise variance of σ^2 and a global noise level $\tau = \sqrt{3}\sigma$. The experiment proceeds as follows: for $d \in \{1, 2\}$, we test intersection angles $\Theta \in \left\{ \frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{6} \right\}$. For each pair (d, Θ), we vary σ from 0 to 0.1 in increments of 0.01, and so equivalently the global noise level τ from 0 to $0.1\sqrt{3}$ in increments of $0.01\sqrt{3}$.

For LAPD, we record the clustering accuracy both from using the default simplex size parameter e (LAPD-D) and from using manually tuned e (LAPD-T) in Figure 4 (all other parameters were set to the default values). We compare against SSC-ADMM (Elhamifar and Vidal, 2011) and EKSS (Lipor et al., 2020), methods designed for subspace clustering, as well as LocPCA (Arias-Castro et al., 2017), which projects neighborhoods of data points onto a tangent subspace, and PBC (Babaeian et al., 2015), a path-based method; we omit comparison with the deep learning method DCV (Wu et al., 2022b) in this section since it fails even without noise. Note all comparison algorithms require the ground truth number of clusters m. To make the comparison fair, we provide m to LAPD as well. We shall explore LAPD's capability of automatically learning the number of clusters in Subsection 5.3.

Comparisons for the subspace clustering tasks are reported in Figure 4. As expected, SSC-ADMM and EKSS achieve the highest clustering accuracy in this context as they are specifically

17

designed for subspace clustering. Among the manifold clustering methods, LAPD maintains the most reliable performance across various noise levels and intersection angles. In particular, LocPCA achieves higher clustering accuracy than LAPD for $d = 1$ but performs poorly for $d = 2$. LocPCA's deteriorating performance is observed for larger d's as well, even with a large intersection angle of $\Theta = \frac{\pi}{2}$. PBC, another path based clustering method, is much less robust to noise and intersection angles. LAPD-D (all default parameters) performs quite well when the noise is moderate and intersection angle not too small; however manual tuning of the edge length parameter e allows for LAPD-T to tolerate higher noise levels. For comparison, we remark that we manually tuned many parameters for all competing methods, and generally performance was sensitive to these parameters. Overall, these results support our theoretical result in (15) as we empirically observe that increasing τ and decreasing θ results in a more limited range of parameter values where the method can succeed, since the edge length e must still remain a local construction.

We now examine how runtime (in seconds) varies with the sample size n and the ambient dimension D. These experiments use a dataset of two intersecting unit hypercubes with $d = 2$. The intersection angle Θ is set to $\frac{\pi}{4}$ and σ fixed at 0.05 (noise level τ fixed at $0.05\sqrt{3}$). In theory, the computational complexities of the benchmark methods are as follows: SSC-ADMM is $O(n^3)$, EKSS is $O(n^2)$, and PBC is $O(n \log n)$, assuming the number of source points is $O(1)$. We did not find a formal complexity analysis for LocPCA in Arias-Castro et al. (2017), but in our experiments, it proved highly scalable.

To investigate runtime dependence on n, we fix $D = 100$ and take a grid of n values ranging from 1000 to 35000. As shown in Figure 5a, LAPD demonstrates near-linear scalability with respect to n. Note we denote by LAPD(NE) our proposed method *not estimating* the intrinsic dimension d and noise τ, while LAPD(E) includes those estimation steps. SSC and EKSS appear to grow at least quadratically, whereas LocPCA and PBC display performance comparable to LAPD.

To examine how runtime depends on D, we fix the dataset size $n = 6,000$ and take a grid of D values ranging from 25 to 2500. Figure 5b plots the measured runtimes against D. Overall, we observe a linear relationship between the LAPD runtime and D. By contrast, EKSS, LocPCA, and PBC exhibit rapid growth as D increases, indicating they suffer from the "curse of dimensionality." Among the compared methods, only SSC outperforms LAPD; however, its apparent independence from D arises because the $O(n^3)$ term dominates the dependence on the ambient dimension, which is $O(Dn^2)$. Taken together, these findings show that LAPD is highly scalable for large datasets in high-dimensional settings.

5.3 Manifold Clustering

We next evaluate LAPD on manifold clustering using 11 synthetic datasets whose underlying manifolds may be nonlinear (see Figure 6). The first four datasets have intrinsic dimension $d = 1$, and the next four have $d = 2$. We refer to the $d = 1$ datasets as *Dollar Sign* (DS), *Olympic Rings* (OR), *Three Curves* (TC), and *Rose Circle* (RC), and the $d = 2$ datasets as *Two 2-Spheres* (2S), *Two Triangles* (TT), *Three Planes* (TP), and *Swiss Roll* (SR). To explore performance in even higher dimensions, we also run experiments on unit d-spheres ($d \in \{3, 4, 5\}$), referred to as 3S, 4S, and 5S for *two 3-spheres*, *two 4-spheres*, and *two 5-spheres*, respectively. We fix the ambient dimension at $D = 100$, the dataset size at $n = 6000$, and the overall noise level at $\tau = 0.05\sqrt{3}$. Note that the noise is evenly distributed across the D ambient dimensions and so only a small portion of the noise is visible in Figure 6 We compare LAPD against the manifold clustering methods LocPCA, PBC, and DCV from the previous section, with clustering results shown in Table 2. For LAPD, we report clustering accuracy both when the ground truth number of clusters m is known (denoted as LAPD(m)) and when m is estimated by LAPD (denoted as LAPD($\hat{m}$)). The runtime remains the same in both cases.

Table 2 shows that LAPD achieves the highest clustering accuracy for all datasets under default parameters and oracle m. The only dataset where it performs poorly is TP, and in fact manual
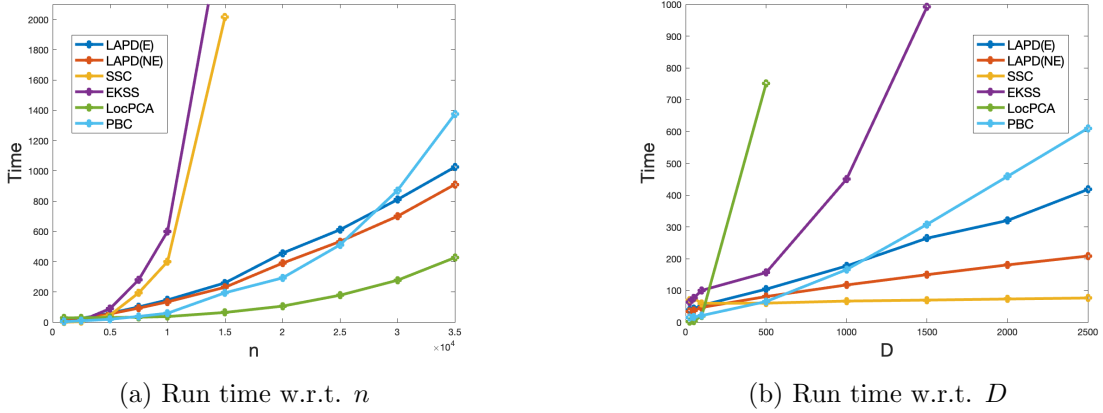
(a) Run time w.r.t. n (b) Run time w.r.t. D

Figure 5: Runtime (in seconds) of various clustering methods. (a) Runtime against the datasize n; (b) Runtime against the ambient dimension D. LAPD(NE) records the runtime of LAPD only, whereas LAPD(E) includes the time for estimating the intrinsic dimension d and noise level τ using the algorithm from Little et al. (2017).

parameter tuning of e can increase the accuracy of LAPD(m) to .938 on the TP dataset. Competing methods may succeed on simpler or more regular manifolds—such as those with lower curvature (DS, OR, TT), larger intersection angles (DS), or smoothly varying shapes (2S, TT, 3S, 4S, 5S)—but they struggle with more challenging cases. For instance, TP presents vastly different intersection angles $\Theta_1 = \frac{\pi}{2}$ and $\Theta_2 = \frac{\pi}{5}$, SR has continuously changing curvature, and RC features multiple clusters lying close to each other. Only LAPD reliably handles such complexities. LocPCA, in particular, experiences the same difficulty here as in subspace clustering: its performance drops when $d = 2$ compared to $d = 1$. Although LocPCA handles higher-dimensional spheres (3S, 4S, 5S) reasonably well, its behavior on more complex shapes warrants further study. PBC achieves an accuracy level similar to LocPCA but often exhibits larger standard deviations, suggesting less consistent cluster assignments. While the runtime of these other methods may increase more slowly with the intrinsic dimension d, extensive parameter tuning is required for each, offsetting any practical advantage over LAPD for higher dimensions.

5.4 Real-world Data

Finally, we evaluate LAPD on four widely used real-world datasets: COIL20 (Nene et al., 1996), USPS (Hull, 1994), MNIST-test (Deng, 2012), and MNIST-full (Deng, 2012). For COIL20, we set $d = 1$ as the intrinsic dimension is clear from the data construction. For the remaining data sets we use $d = 2$, which generally yielded good performance. These datasets are popular in subspace clustering (Peng et al., 2020), but their underlying manifolds generally have low curvature, and most classes are well-separated (Wu et al., 2022b,a; McConville et al., 2021). Thus, these benchmarks do not fully expose the advantages of our geometry-based approach, which is designed to handle more pronounced manifold intersections.

Despite the overall separability of most classes, certain subsets do have strong similarities. For instance, COIL20 classes 2, 5, and 18 (cars) are nearly identical, and some MNIST digit classes overlap (e.g., 4, 7, and 9). To give a more nuanced view, for each real-world dataset we analyze 3 subsets: the first uses classes that are relatively distinct, the second introduces more difficult classes, and the third includes all classes. Detailed results appear in Table 3.
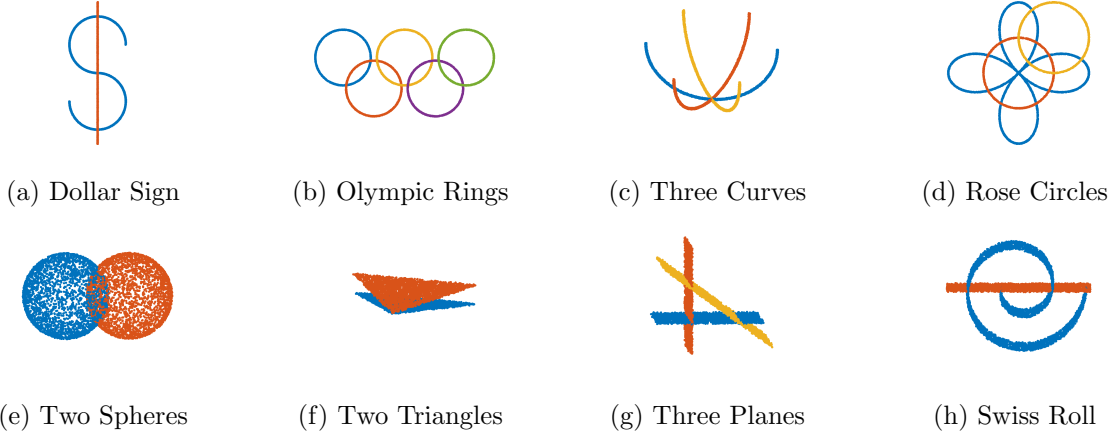
(a) Dollar Sign   (b) Olympic Rings   (c) Three Curves   (d) Rose Circles

(e) Two Spheres   (f) Two Triangles   (g) Three Planes   (h) Swiss Roll

Figure 6: Synthetic manifold learning datasets. Objects on the first row have intrinsic dimension $d = 1$; objects on the second row have intrinsic dimension $d = 2$. All the objects have ambient dimension $D = 100$. For plotting purpose, the figures display the datasets' projection on the $d + 1$ ($d + 2$ for TC) dimensional space.

Methods / Objects	LAPD (m)			LAPD $(\hat{m})$			LocPCA			PBC			DCV		
DS	.992	.002	12.9	.992	.002	.984	.003	11.4	.956	.007	10.5	.510	.006	175.6
OR	.982	.002	11.4	.982	.002	.881	.043	54.6	.838	.121	7.0	.597	.007	213.7
TC	.989	.001	12.8	.989	.001	.984	.003	9.6	.694	.115	26.8	.475	.006	295.3
RC	.976	.003	12.9	.917	.092	.533	.039	50.1	.740	.017	1.5	.482	.005	233.9
2S	.982	.003	16.4	.982	.003	.965	.003	85.0	.920	.014	13.4	.851	.006	227.5
TT	.988	.003	35.4	.988	.003	.547	.050	38.0	.980	.009	6.9	.516	.004	234.6
TP	.733	.174	41.3	.675	.078	.542	.032	88.6	.640	.056	16.3	.413	.003	227.1
SR	.975	.002	26.7	.975	.002	.700	.064	39.4	.803	.114	15.8	.608	.005	234.8
3S	.976	.005	42.5	.976	.005	.946	.004	121.2	.937	.010	23.1	.929	.004	220.3
4S	.971	.003	180.2	.971	.003	.954	.002	194.2	.925	.006	35.2	.960	.003	193.2
5S	.960	.005	3352.9	.960	.005	.963	.002	204.5	.941	.006	34.0	.973	.002	247.7

Table 2: Manifolds clustering results. The three sub-columns under each method record the average clustering accuracy, standard deviation, and runtime in seconds. LAPD(m) denotes clustering with oracle cluster numbers m; LAPD($\hat{m}$) denotes when m is estimated by LAPD. The two share about the same runtime.

A few observations stand out. First of all, LAPD^2 typically outperforms LAPD^1; on these datasets, utilizing the two-sided weight (16) enhances within-class connections without substantially increasing inter-class connections, but we emphasize this was not the case for the synthetic data sets analyzed in Sections 5.2 and 5.3. Secondly, LocPCA and PBC generally yield lower accuracy than LAPD and can be sensitive to parameter tuning. Interestingly, DCV often matches LAPD's high accuracy on these well-separated data sets, but the method fails on the majority of the synthetic data in Section 5.3 where manifolds intersect more substantially. Moreover, DCV's deep-learning framework has many parameters, which can be challenging to optimize in an unsupervised setting. Finally, we note EKSS requires a large ensemble size (up to 1,000), additional preprocessing (e.g., whitening for COIL20), and carefully chosen parameters (e.g., intrinsic dimensions), and these factors

	LAPD[1]	LAPD[2]	time	SSC	time	EKSS[3]	time	LocPCA	time	PBC	time	DCV	time
COIL [0:6,8:17,19]	.997	.982	4.1	.876	24.6	.799	95.3	.473	165.7	.386	112.7	.957	101.4
COIL [0:17,19]	.913	.931	4.1	.863	28.4	.833	110.7	.464	199.5	.405	126.8	.963	96.8
COIL [0:19]	.857	.904	4.8	.836	31.2	.806	132.8	.360	180.9	.437	134.2	.900	104.6
MNIST-T [0:7]	.952	.961	84.9	.737	403.8	.869	4029.7	.715	101.3	.452	1658.0	.967	400.6
MNIST-T [0:8]	.757	.958	97.5	.797	650.5	.791	8752.5	.557	387.6	.414	3422.5	.952	451.2
MNIST-T [0:9]	.688	.795	112.1	.472	1081.3	.685	13,675.7	.489	604.0	.358	5521.8	.855	544.4
MNIST-F [0:6]	.948	.975	836.2	NA	NA	NA	NA	.618	1320.4	NA	NA	NA	NA
MNIST-F [0:8]	.751	.957	1204.9	NA	NA	NA	NA	.562	1803.6	NA	NA	NA	NA
MNIST-F [0:9]	.718	.859	1338.4	NA	NA	NA	NA	.538	2095.2	NA	NA	NA	NA
USPS [0:6]	.944	.984	40.7	.665	356.5	.823	347.8	.653	1.41	.440	64.6	.980	214.5
USPS [0:6,8:9]	.724	.942	77.1	.679	1037.8	.782	573.4	.635	55.5	.424	13.3	.967	300.7
USPS [0:9]	.748	.935	87.5	.774	1400.4	.778	1218.2	.583	63.7	.378	70.0	.950	340.3

Table 3: Clustering accuracy and runtime on real-world datasets. Four sections are each for COIL20, MNIST-test, MNIST-full, and USPS in order. NA indicates the method either encounters an out-of-memory issue or consumes a very long runtime on that particular dataset. For EKSS, we did not incorporate the pre-processing steps so our results are different from those reported in their original paper (Lipor et al., 2020).

can make its workflow more complex. Overall, LAPD provides consistently strong performance on these benchmarks with relatively few parameters, reinforcing its utility as a practical method for manifold clustering.

6 Conclusion

We propose and investigate a new algorithm for MMC, using angles between d-simplices as an ingredient for computing with the LAPD metric. Highlights of the algorithm include: (i) it works extremely well for clustering intersecting manifolds; (ii) it works with nonlinear manifolds and is robust to noise; (iii) it can generally give an estimate of the number of manifolds; and (iv) it is scalable with quasi-linear computational complexity. Our results demonstrate that LAPD is competitive with or superior to alternative MMC methods on both synthetic and real-world datasets.

In future work we hope to extend the method to collections of manifolds with varying intrinsic dimension; such an extension would be non-trivial but potentially possible through an iterative procedure. In addition, in this article we have explored a completely angle-based path distance (LAPD) instead of a density-based path distance (LLPD). In practice, it could be helpful to define a metric combining both of these components, since as observed in the real data experiments class intersections typically consist of low-density regions.

Acknowledgments and Disclosure of Funding

Appendix A. Proof of Theorem 4.1

Theorem 4.1 (wLAPD upper bound) *Let Assumptions 2.1, 2.2, 3.1, 3.2 and 4.1 hold. Assume also the construction outlined in Section 3 with either weight construction (4) or (16), and that $\mathcal{G}_S$ is connected. Then:*

$$wLAPD \lesssim \frac{\tau}{e}. \tag{13}$$

Recall our setting that the noise polluted data $X = \{x_i\}_{i \in [n]}$ is sampled from a tubular region $T_\tau(\mathcal{M}) = \cup_{j \in [m]} T_\tau(\mathcal{M}_j)$ around the set of underlying manifolds. Note that when $\tau = 0$, the result is trivial since wLAPD $= 0$; we thus assume $\tau > 0$. We first bound the wLAPD for a fixed manifold component $\mathcal{M}_1$:

$$\text{wLAPD}(\mathcal{M}_1) = \max_{\Delta_+, \Delta_- \in S(\mathcal{M}_1)} \text{LAPD}(\Delta_+, \Delta_-),$$

with $S(\mathcal{M}_1)$ denoting the set of *valid, pure* simplices of $\mathcal{M}_1$ constructed from noisy samples from $T_\tau(\mathcal{M}_1)$. Refer to Section 3 for the definition of valid and pure simplices. For the rest of the section, since we are focusing on the fixed manifold $\mathcal{M}_1$, we are only dealing with the pure simplices of $\mathcal{M}_1$ and we will not specify this fact as long as the context is clear. We will also write the dihedral angle θ_{ij} for $\theta(\Delta_i, \Delta_j)$ or $\theta_\pm$ for $\theta(\Delta_-, \Delta_+)$.

We derive a (nonrandom) upper bound by replacing the max over the discrete set $S(\mathcal{M}_1)$ with a max over $V(\mathcal{M}_1)$, the set of all (e, q)-simplices (equation 2) which can be constructed from points in a D-dimensional tube of radius τ about $\mathcal{M}_1$. Specifically

$$V(\mathcal{M}_1) = \{\Delta = \{x_1, \ldots, x_{d+1}\} : \|x_i - P_{\mathcal{M}_1} x_i\| \le \tau, \, e \le \|x_i - x_j\| \le \frac{e}{q} \, \forall i, j \in [d+1], \, i < j\},$$

where $P_{\mathcal{M}_1}$ is projection onto $\mathcal{M}_1$; note $S(\mathcal{M}_1) \subseteq V(\mathcal{M}_1)$ because $V(\mathcal{M}_1)$ contains *all* valid simplices, while $S(\mathcal{M}_1)$ contains just those arising from our discrete samples. We prove that Theorem 4.1 holds for both weight constructions considered in our paper, i.e. for $W_S^1(\Delta_i, \Delta_j) = \pi - \theta_{ij}$ when $\theta_{ij} \ge \frac{\pi}{2}$ (else not connected; see (4)) and for $W_S^2(\Delta_i, \Delta_j) = \min\{\pi - \theta_{ij}, \theta_{ij}\}$ (see (16)). Note that since $\mathcal{G}_S$ is connected, all distances are finite, and:

$$\text{wLAPD}(\mathcal{M}_1, W_S^1) \le \max_{\Delta_i, \Delta_j \in V(\mathcal{M}_1), \, \theta_{ij} \ge \frac{\pi}{2}} \pi - \theta_{ij} := (A)$$

$$\text{wLAPD}(\mathcal{M}_1, W_S^2) \le \max_{\Delta_i, \Delta_j \in V(\mathcal{M}_1)} \min\{\pi - \theta_{ij}, \theta_{ij}\} := (B)$$

Since $(A) \le (B)$, wLAPD$(\mathcal{M}_1) \le (B)$ under both weight constructions, and it suffices to bound (B). Letting $V(\mathcal{M}_1, D)$ denote the dependence on the ambient dimension of the simplex set construction, we first establish that (B) is independent of the ambient dimension D. Specifically, since Δ_i, Δ_j must always live in a $(d+1)$-dimensional space by construction, we have:

$$\max_{\Delta_i, \Delta_j \in V(\mathcal{M}_1, D)} \min\{\pi - \theta_{ij}, \theta_{ij}\} = \max_{\substack{\text{subspaces } S, \, \Delta_i, \Delta_j \in V(\mathcal{M}_1, D) \cap S \\ \dim(S) = d+1}} \min\{\pi - \theta_{ij}, \theta_{ij}\}$$

$$= \max_{\Delta_i, \Delta_j \in V(\mathcal{M}_1, d+1)} \min\{\pi - \theta_{ij}, \theta_{ij}\}.$$

This is verified with numerical results from optimization in Figure 9a. Thus moving forward without loss of generality we can consider $D = d + 1$.

The proof consists of the following steps:

1. Assuming $q = 1$, i.e. all simplices being regular, the optimal configuration of a pair of simplices which generates (B) occurs when the shared face lies on one side of the boundary of $T_\tau(\mathcal{M}_1)$ and the two apex vertices lie on the opposite side of the boundary; see Subsubsection A.0.1.

2. LAPD upper bound corresponding to this configuration is $(B) = \sqrt{\frac{32d}{d+1}}\frac{\tau}{e} \sim O(\frac{\tau}{e})$; see Sub-subsection A.0.2.

3. When $q \neq 1$, the optimal configuration still occurs when this configuration happens, and thus (B) will depend on q in a continuous fashion; we provide numerical evidence supporting this claim in Subsubsection A.0.3. Analytically computing the exact dependence on q is difficult and not particularly insightful, but clearly we still have $(B) = O\left(\frac{\tau}{e}\right)$.

Finally, since an identical argument shows $\text{wLAPD}(\mathcal{M}_2) = O\left(\frac{\tau}{e}\right)$, we can conclude $\text{wLAPD} = O\left(\frac{\tau}{e}\right)$, which proves Theorem 4.1.

A.0.1 OPTIMAL SIMPLEX CONFIGURATION

We denote an arbitrary pair of adjacent d-simplices in a $(d+1)$-dimensional ambient space as $\Delta_- = [x_1, \ldots, x_d, x_-]$ and $\Delta_+ = [x_1, \ldots, x_d, x_+]$ with $x_i, x_\pm \in \mathbb{R}^{d+1}$ for $i \in [d]$, i.e. the first d vertices $[x_1, \ldots, x_d]$ denote the shared face and $x_\pm$ denote the apex vertices from the pair of adjacent simplices $\Delta_\pm$. We choose a coordinate system so that $\mathcal{M}_1 \subseteq \text{span}\{e_1, \ldots, e_d\}$, i.e. the manifold lives in the first d dimensions and the noise in dimension $d+1$, so that $|x_i^{(d+1)}| \leq \tau$ and similarly for $x_\pm$. The following proposition establishes the optimal configuration generating (B) occurs when the shared face lies on the boundary of the tubular noise region and the two non-shared apex vertices lie on the other side of the boundary.

Proposition A.1 *Assume $q = 1$ and let $\Delta_+, \Delta_- \in V(\mathcal{M}_1, d+1)$ be two adjacent simplices maximizing $B(\Delta_+, \Delta_-) = \max_{\Delta_+, \Delta_- \in V(\mathcal{M}_1, d+1)} \min\{\pi - \theta_\pm, \theta_\pm\}$. Then either*

$$x_i^{(d+1)} = -\tau, \ 1 \leq i \leq d \ \text{and} \ x_\pm^{(d+1)} = \tau$$

or

$$x_i^{(d+1)} = \tau, \ 1 \leq i \leq d \ \text{and} \ x_\pm^{(d+1)} = -\tau.$$

Proof We first present a proof for $d = 1, 2$. For larger d's, a rigorous proof becomes challenging and we will present an empirical result from optimization to support our conclusion.

Consider the optimization problem:

$$\max_{\Delta_+, \Delta_- \in V(\mathcal{M}_1, d+1)} \min\{\pi - \theta_\pm, \theta_\pm\}. \tag{18}$$

For $d = 1$, we make the shared node x_1 of the pair of adjacent vectors $\overrightarrow{x_1 x_-}$ and $\overrightarrow{x_1 x_+}$ the origin of e_1, i.e. $x_1 = (0, x_1^{(2)})^\intercal$. Using the polar coordinate system gives a simplex expression for x_-, x_+ that satisfies the edge length constraint, $x_- = (e\cos\theta_-, x_1^{(2)} + e\sin\theta_-)^\intercal$ and $x_+ = (e\cos\theta_+, x_1^{(2)} + e\sin\theta_+)^\intercal$ where θ_- and θ_+ are the angle made by the vector $\overrightarrow{v}_- := \overrightarrow{x_1 x_-} = (e\cos\theta_-, e\sin\theta_-)^\intercal$ and $\overrightarrow{v}_+ := \overrightarrow{x_1 x_+} = (e\cos\theta_+, e\sin\theta_+)^\intercal$ from e_1. W.l.o.g, we let $\theta_- \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $|\theta_+ - \theta_-| = \theta$ for $\theta \in [0, \pi]$. The case for $\theta_- \in [\frac{\pi}{2}, \frac{3\pi}{2}]$ can be proved with the same argument and thus will lead to the same conclusion. It follows that the optimization problem (18) for $d = 1$, along with the appropriate constraints, can be reformulated as:

$$\max_{\theta_-, \theta_+, x_1^{(2)}} \min\{\pi - \arccos\frac{v_- \cdot v_+}{\|v_-\|\|v_+\|}, \arccos\frac{v_- \cdot v_+}{\|v_-\|\|v_+\|}\}$$

$$\text{s.t. } |x_1^{(2)}| \leq \tau, \ |x_1^{(2)} + e\sin\theta_-| \leq \tau, \ |x_1^{(2)} + e\sin\theta_+| \leq \tau$$

whose objective function can be furthered simplified to

$$\max_{\theta \in [0, \pi]} \min\{\pi - \theta, \theta\}.$$

23

Here two cases need be considered: (i) $\theta \in [0, \frac{\pi}{2}]$ and (ii) $\theta \in [\frac{\pi}{2}, \pi]$. For (i) the problem simplifies to

$$\max_{\theta \in [0, \frac{\pi}{2}]} \quad \theta$$

$$\text{s.t. } |x_1^{(2)}| \leq \tau, \quad \theta_-, \theta_+ \in \arcsin \left[\frac{-\tau - x_1^{(2)}}{e}, \frac{\tau - x_1^{(2)}}{e} \right]$$

Because arcsin is a strictly increasing function on its domain, the maximum is achieved when θ_-, θ_+ take the two ends of their constraint interval. It follows that

$$\max \theta = \arcsin \left(\frac{\tau - x_1^{(2)}}{e} \right) - \arcsin \left(\frac{-\tau - x_1^{(2)}}{e} \right)$$

$$= \arcsin \left(\frac{\tau - x_1^{(2)}}{e} \right) + \arcsin \left(\frac{\tau + x_1^{(2)}}{e} \right)$$

$$= \arcsin(a - b) + \arcsin(a + b) =: f(b)$$

where $a = \frac{\tau}{e}$ is a constant and $b = \frac{x_1^{(2)}}{e}$ is a variable in $[-a, a]$. Note that $f(b)$ is an even function and the maximum is achieved at the boundary points $b = \pm a$. Equivalently, we must have $x_1^{(2)} = \pm \tau$. We list all the possible configurations that achieve the maximum

- $x_1^{(2)} = \tau, x_-^{(2)} = \tau, x_+^{(2)} = -\tau$.

- $x_1^{(2)} = \tau, x_-^{(2)} = -\tau, x_+^{(2)} = \tau$.

- $x_1^{(2)} = -\tau, x_-^{(2)} = \tau, x_+^{(2)} = -\tau$.

- $x_1^{(2)} = -\tau, x_-^{(2)} = -\tau, x_+^{(2)} = \tau$.

In words, the optimal configuration for the $\theta \in [0, \frac{\pi}{2}]$ case occurs when the shared face x_1 lies on one side of the boundary of $T_\tau(\mathcal{M}_1)$, along with one of the two apex vertices $\{x_-, x_+\}$ lying on the same side of the boundary; The other apex vertex lies on the opposite side of the boundary. See θ_1 in Figure 7a. One can verify that $\theta_{\max}$, in this case, equals $\arcsin \frac{2\tau}{e}$ and so the optimal value of case (i), for $\theta \in [0, \frac{\pi}{2}]$, is $\arcsin \frac{2\tau}{e}$.

For (ii), the problem reduces to

$$\max_{\theta \in [\frac{\pi}{2}, \pi]} \quad \pi - \theta$$

$$\text{s.t. } |x_1^{(2)}| \leq \tau, \quad \theta_- \in \arcsin[0, \frac{\tau - x_1^{(2)}}{e}], \quad \theta_+ \in \pi - \arcsin[0, \frac{\tau - x_1^{(2)}}{e}]$$

$$\text{or } |x_1^{(2)}| \leq \tau, \quad \theta_- \in \arcsin[\frac{-\tau - x_1^{(2)}}{e}, 0], \quad \theta_+ \in -\pi - \arcsin[\frac{-\tau - x_1^{(2)}}{e}, 0]$$

The two constraint regions simplify the objective function individually to the following two cases:

$$\max_{\theta \in [\frac{\pi}{2}, \pi]} \pi - \theta = \begin{cases} 2 \arcsin \frac{\tau - x_1^{(2)}}{e}, & \text{if } \theta_- = \arcsin \frac{\tau - x_1^{(2)}}{e}, \theta_+ = \pi - \arcsin \frac{\tau - x_1^{(2)}}{e}. \\ 2 \arcsin \frac{\tau + x_1^{(2)}}{e} & \text{if } \theta_- = \arcsin \frac{-\tau - x_1^{(2)}}{e}, \theta_+ = -\pi - \arcsin \frac{-\tau - x_1^{(2)}}{e}. \end{cases}$$

It is easy to verify that in the former case, we need $x_1^{(2)} = -\tau$ and that leads to $\theta = \pi - 2 \arcsin \frac{2\tau}{e}$ and $x_\pm^{(2)} = \tau$. In the later case, we need $x_1^{(2)} = \tau$ and that leads to $\theta = \pi - 2 \arcsin \frac{2\tau}{e}$ again and $x_\pm^{(2)} = -\tau$. We list all the possible configurations that achieve the optimum

24

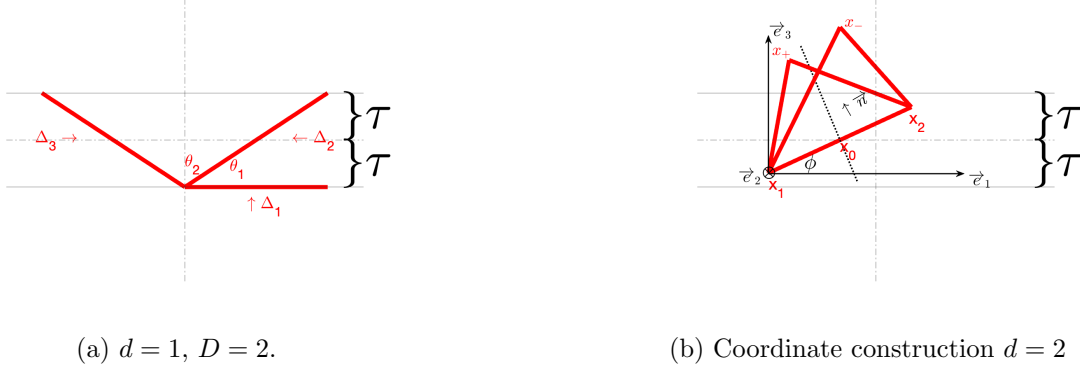(a) $d = 1$, $D = 2$.   (b) Coordinate construction $d = 2$

Figure 7: (a) Optimal configuration for $d = 1$ that attains the largest wLAPD with constraints of noise τ and simplex size e. θ_1 corresponds to the (local) optimal configuration for $\theta = |\theta_+ - \theta_-| < \frac{\pi}{2}$; θ_2 corresponds to the (global) optimal configuration for $\theta = |\theta_+ - \theta_-| > \frac{\pi}{2}$. (b) Construction of coordinates for $d = 2$.

- $x_1^{(2)} = \tau, x_\pm^{(2)} = -\tau$.

- $x_1^{(2)} = -\tau, x_\pm^{(2)} = \tau$.

In words, the optimal configuration for the $\theta \in [\frac{\pi}{2}, \pi]$ case occurs when the shared face lies on one side of the boundary of $T_\tau(\mathcal{M}_1)$ and the two apex vertices lie on the other side of the boundary. See θ_2 in Figure 7a. The maximal value for case (ii) is $2 \arcsin \frac{2\tau}{e}$.

Combining (i) and (ii), we see that (ii) has the configuration that achieves the optimal value for problem (18), which is $2 \arcsin \frac{2\tau}{e} \approx \frac{4\tau}{e}$ when $\tau \ll e$ by Assumption 3.2. The corresponding optimal configuration is

$$x_1^{(2)} = -\tau, \text{ and } x_\pm^{(2)} = \tau,$$

or

$$x_1^{(2)} = \tau, \text{ and } x_\pm^{(2)} = -\tau,$$

and we conclude the proof for $d = 1$.

Now consider $d = 2$ for the same optimization problem in (18). Denote by $\Delta_- = [x_1, x_2, x_-]$ and $\Delta_+ = [x_1, x_2, x_+]$ the two adjacent regular triangles. Following the same logic for $d = 1$, we shift the coordinate system such that $x_1 = (0, 0, x_1^{(3)})^\mathsf{T}$ is the origin of $\mathrm{span}\{e_1, e_2\}$ and x_2 lives in the subspace of $\mathrm{span}\{e_1, e_3\}$. $\overrightarrow{x_1 x_2}$ satisfies the edge length constraint and so we must have $x_2 = (e \cos \phi, 0, x_1^{(3)} + e \sin \phi)^\mathsf{T}$ with $\phi \in [0, \pi)$ denoting the angle of the shared face $\overrightarrow{x_1 x_2}$ flipped from $\mathrm{span}\{e_1, e_2\}$. See Figure 7b for illustration. The case when $\phi \in [\pi, 2\pi)$ can be analogously analyzed with a symmetry argument. Fixing the position of the shared face $\overrightarrow{x_1 x_2}$, the two triangles Δ_-, Δ_+ can rotate about $\overrightarrow{x_1 x_2}$. We introduce $\theta_\pm$ to characterize the angle between $\Delta_\pm$ from $\mathrm{span}\{e_2, \overrightarrow{x_1 x_2}\}$. Again we enforce $\theta_- \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $\theta = |\theta_- - \theta_+| \in [0, \pi]$ w.l.o.g. Due to the edge length constraint, the coordinates of the apex points can be explicitly calculated as

$$x_- = (-\frac{\sqrt{3}}{2} e \sin \theta_- \sin \phi + \frac{e}{2} \cos \phi, \frac{\sqrt{3}}{2} e \cos \theta_-, x_1^{(3)} + \frac{\sqrt{3}}{2} e \sin \theta_- \cos \phi + \frac{e}{2} \sin \phi)^\mathsf{T}$$

$$x_+ = (-\frac{\sqrt{3}}{2} e \sin \theta_+ \sin \phi + \frac{e}{2} \cos \phi, \frac{\sqrt{3}}{2} e \cos \theta_+, x_1^{(3)} + \frac{\sqrt{3}}{2} e \sin \theta_+ \cos \phi + \frac{e}{2} \sin \phi)^\mathsf{T}$$

Note that the middle point x_0 of $\overrightarrow{x_1 x_2}$ is equal to $\frac{x_1 + x_2}{2} = (\frac{e}{2}\cos\phi, 0, x_1^{(3)} + \frac{e}{2}\sin\phi)^\intercal$ and so the two residual vectors $v_- := \overrightarrow{x_0 x_-}$ and $v_+ := \overrightarrow{x_0 x_+}$ can be calculated. The optimization problem in (18) then boils down to

$$\max_{\theta_-, \theta_+, x_1^{(3)}} \min\{\pi - \arccos\frac{v_- \cdot v_+}{\|v_-\|\|v_+\|}, \arccos\frac{v_- \cdot v_+}{\|v_-\|\|v_+\|}\}$$

$$\text{s.t. } |x_1^{(3)}| \leq \tau, \ |x_1^{(3)} + e\sin\phi| \leq \tau,$$

$$|x_1^{(3)} + \frac{\sqrt{3}}{2}e\sin\theta_- \cos\phi + \frac{e}{2}\sin\phi| \leq \tau, \ |x_1^{(3)} + \frac{\sqrt{3}}{2}e\sin\theta_+ \cos\phi + \frac{e}{2}\sin\phi| \leq \tau.$$

Straightforward calculation, using trigonometric properties, reveals that $v_- \cdot v_+ = \frac{3}{4}e^2 \cos\theta$ with $\|v_-\| = \|v_+\| = \frac{\sqrt{3}}{2}e$ and so, the above objective can be further reduced to

$$\max_{\theta \in [0, \pi]} \min\{\pi - \theta, \theta\}$$

Once again we are facing two cases: (i) $\theta \in [0, \frac{\pi}{2}]$ and (ii) $\theta \in [\frac{\pi}{2}, \pi]$.

For (i), the optimization problem becomes

$$\max_{\theta \in [0, \frac{\pi}{2}]} \theta$$

$$\text{s.t. } |x_1^{(3)}| \leq \tau, \ |x_1^{(3)} + e\sin\phi| \leq \tau$$

$$\theta_-, \theta_+ \in \arcsin[\frac{-2\tau - 2x_1^{(3)} - e\sin\phi}{\sqrt{3}e\cos\phi}, \frac{2\tau - 2x_1^{(3)} - e\sin\phi}{\sqrt{3}e\cos\phi}]$$

Note that the maximum is obtained when θ_- and θ_+ take the two endpoints of the domain interval. That is,

$$\max\theta = \arcsin\left(\frac{2\tau - 2x_1^{(3)} - e\sin\phi}{\sqrt{3}e\cos\phi}\right) - \arcsin\left(\frac{-2\tau - 2x_1^{(3)} - e\sin\phi}{\sqrt{3}e\cos\phi}\right)$$

$$= \arcsin\left(\frac{2\tau - (2x_1^{(3)} + e\sin\phi)}{\sqrt{3}e\cos\phi}\right) + \arcsin\left(\frac{2\tau + 2x_1^{(3)} + e\sin\phi}{\sqrt{3}e\cos\phi}\right)$$

$$= \arcsin(a - b) + \arcsin(a + b) =: f(b)$$

Note that for a fixed ϕ, $a = \frac{2\tau}{\sqrt{3}e\cos\phi}$ is a constant whereas $b = \frac{2x_1^{(3)} + e\sin\phi}{\sqrt{3}e\cos\phi} \in [-a, a]$ is the variable. Analogous to $d = 1$, the maximum of $f(b)$ is attained at $b = \pm a$. $b = a$ implies $2x_1^{(3)} + e\sin\phi = 2\tau$. Due to the constraints $|x_1^{(3)}| \leq \tau$, $|x_1^{(3)} + e\sin\phi| \leq \tau$, and $\phi \in [0, \pi)$, this can only happen when $x_1^{(3)} = \tau$ and $x_1^{(3)} + e\sin\phi = \tau$, which uncovers the configuration $x_1^{(3)} = \tau$ and $\phi = 0$. Analogously, $b = -a$ implies $x_1^{(3)} = -\tau, \phi = 0$. In the former case, when $x_1^{(3)} = \tau$ and $\phi = 0$, one of θ_-, θ_+ equals 0 and the other equals $\arcsin\left(\frac{-4\tau}{\sqrt{3}e}\right)$, which suggests that one of $x_-^{(3)}, x_+^{(3)}$ equals τ and the other equals $-\tau$; In the later case, when $x_1^{(3)} = -\tau$ and $\phi = 0$, one of θ_-, θ_+ equals 0 and the other equals $\arcsin\left(\frac{4\tau}{\sqrt{3}e}\right)$, which suggests that one of $x_-^{(3)}, x_+^{(3)}$ equals $-\tau$ and the other one equals τ. We list all the possible configurations that achieve the maximum.

- $x_1^{(3)} = x_2^{(3)} = \tau, x_-^{(3)} = \tau, x_+^{(3)} = -\tau$.
- $x_1^{(3)} = x_2^{(3)} = \tau, x_-^{(3)} = -\tau, x_+^{(3)} = \tau$.

26

- $x_1^{(3)} = x_2^{(3)} = -\tau, x_-^{(3)} = \tau, x_+^{(3)} = -\tau.$

- $x_1^{(3)} = x_2^{(3)} = -\tau, x_-^{(3)} = -\tau, x_+^{(3)} = \tau.$

In other words, the optimal configuration for $\theta \in [0, \frac{\pi}{2}]$ occurs when the shared face $\overrightarrow{x_1 x_2}$ lies on one side of the boundary of $T_\tau(\mathcal{M}_1)$, with one of the two apex vertices $\{x_-, x_+\}$ lying on the same side of the boundary and the other apex vertex lying on the opposite side. One can verify that $\theta_{\max}$, in this case, equals $\arcsin \frac{4\tau}{\sqrt{3}e}$ and so the optimal value of case (i), for $\theta \in [0, \frac{\pi}{2}]$, is $\arcsin \frac{4\tau}{\sqrt{3}e}$.

For (ii), without loss of generality, the optimization problem becomes

$$\max_{\theta \in [\frac{\pi}{2}, \pi]} \pi - \theta$$

$$\text{s.t. } |x_1^{(3)}| \leq \tau, \ |x_1^{(3)} + e\sin\phi| \leq \tau$$

$$\theta_- \in [0, \frac{2\tau - 2x_1^{(3)} - e\sin\phi}{\sqrt{3}e\cos\phi}], \ \theta_+ = \pi - [0, \frac{2\tau - 2x_1^{(3)} - e\sin\phi}{\sqrt{3}e\cos\phi}]$$

$$\text{or } \theta_- \in [\frac{-2\tau - 2x_1^{(3)} - e\sin\phi}{\sqrt{3}e\cos\phi}, 0], \ \theta_+ = -\pi - [\frac{-2\tau - 2x_1^{(3)} - e\sin\phi}{\sqrt{3}e\cos\phi}, 0]$$

where the two constraints on $\theta_\pm$ depend on the quadrants where the two simplices live. The above can be further simplified to

$$\max_{\theta \in [\frac{\pi}{2}, \pi]} \pi - \theta = \begin{cases} 2\arcsin \frac{2\tau - 2x_1^{(3)} - e\sin\phi}{\sqrt{3}e\cos\phi}, & \theta_- \in [0, \frac{\pi}{2}], \theta_+ \in [\frac{\pi}{2}, \pi] \\ 2\arcsin \frac{2\tau + 2x_1^{(3)} + e\sin\phi}{\sqrt{3}e\cos\phi}, & \theta_- \in [-\frac{\pi}{2}, 0], \theta_+ \in [-\pi, -\frac{\pi}{2}] \end{cases}$$

It is easy to verify that in the former case, we need $2x_1^{(3)} + e\sin\phi = -2\tau$ and in the later case, we need $2x_1^{(3)} + e\sin\phi = 2\tau$. Due to the constraints, it has to be that in the former case, $x_1^{(3)} = -\tau, x_1^{(3)} + e\sin\phi = -\tau$, which indicates $\phi = 0$ and so $x_- = x_+ = \tau$. Similarly, in the later case, $x_1^{(3)} = \tau$, which indicates $\phi = 0$ and so $x_- = x_+ = -\tau$. We list all the possible configurations that achieve the maximum.

- $x_1^{(3)} = x_2^{(3)} = \tau, x_\pm^{(3)} = -\tau$

- $x_1^{(3)} = x_2^{(3)} = -\tau, x_\pm^{(3)} = \tau$

In words, the optimal configuration for $\theta \in [\frac{\pi}{2}, \pi]$ occurs when the shared face lies on one side of the boundary of $T_\tau(\mathcal{M}_1)$ and the two apex vertices lie on the other side of the boundary. The maximal value for case (ii) is $2\arcsin \frac{4\tau}{\sqrt{3}e}$.

Combining (i) and (ii), we see that (ii) has the configuration that achieves the (global) optimal value for the objective function (18) for $d = 2$. See Figure 8. In particular, the optimal value equals $2\arcsin \frac{4\tau}{\sqrt{3}e} \approx \frac{8\tau}{\sqrt{3}e}$ when $\tau \ll e$ by Assumption 3.2. The optimal configuration is
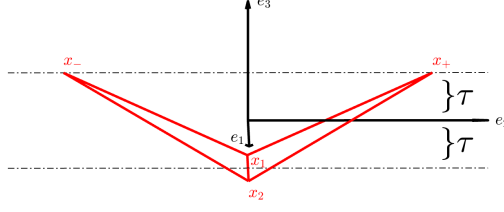
$$x_1^{(3)} = x_2^{(3)} = -\tau, \text{ and } x_\pm^{(3)} = \tau,$$

or

$$x_1^{(3)} = x_2^{(3)} = \tau, \text{ and } x_\pm^{(3)} = -\tau.$$

We therefore finish the proof for $d = 2$.

To draw the same conclusion for $d > 2$, one needs $x_1^{(d+1)}$ and $d + 1$ additional projection parameters (the angles) to capture the constraints, i.e. one must solve an (nonconvex) optimization

27

Figure 8: Optimal Configuration for $d = 2$.

problem with $d + 2$ unknowns. The analytical derivation of the solution is straightforward but cumbersome, and the optimal configuration for any dimension d is always when the apex vertices and the shared face appear on the opposite side of $T_\tau(\mathcal{M}_1)$. We skip the proof for the sake of brevity. ∎

A.0.2 wLAPD UPPER BOUND

In the last section, we showed that when $q = 1$, the global optimum for (18) occurs when $x_i^{(d+1)}$ for $i \in [d]$ and $x_\pm^{(d+1)}$ have the same scale τ but opposite sign. Following this optimal configuration, we are going to find a general formula for the upper bound on (18) for arbitrary d's. Note that as a byproduct of the calculation in the last section, this upper bound for $d = 1$ and $d = 2$ were found to be $\frac{4\tau}{e}$ and $\frac{8\tau}{\sqrt{3}e}$, respectively. Our general formula to be found should verify these results.

Since we can translate the optimal configuration along the manifold, without loss of generality we can make the centroid of the shared face be at zero in the first d coordinates, i.e.

$$x_0 = (0, \ldots, 0, -\tau).$$

Also, since we can always rotate the optimal configuration in the manifold dimensions, we can assume without loss of generality that the first $d - 1$ coordinates of x_-, x_+ are zero (they span a 2-dimensional space, but only 1 manifold dimension, which we can rotate to the dth coordinate), and symmetry will give:

$$x_- = (0, \ldots, -x, \tau), \quad x_+ = (0, \ldots, x, \tau)$$

for some positive constant x. However, from the Pythagorean theorem we know $\|\overrightarrow{x_0 x_-}\| = \|\overrightarrow{x_0 x_+}\| = e\sqrt{\frac{d+1}{2d}}$. Solving for x gives $x = \sqrt{e^2 \frac{(d+1)}{2d} - (2\tau)^2}$, so that the angle θ between $\overrightarrow{x_0 x_-}, \overrightarrow{x_0 x_+}$ satisfies $\cos(\theta) = \frac{16d}{d+1} \frac{\tau^2}{e^2} - 1$. Converting this to $\sin(\theta)$ and by the assumption $\tau \ll e$ we have

$$\theta \approx \sin(\theta) = \sqrt{\frac{32d}{d+1} \frac{\tau^2}{e^2} - \frac{(16d)^2}{(d+1)^2} \frac{\tau^4}{e^4}} \approx \sqrt{\frac{32d}{d+1}} \frac{\tau}{e}.$$

Hence for any d, wLAPD has an upper bound $\sqrt{\frac{32d}{d+1}} \frac{\tau}{e} \sim O\left(\frac{\tau}{e}\right)$. This upper bound verifies our finding from explicit construction and calculation in the last section for $d = 1, 2$. Note that the empirical result from optimization ($q = 1$ curve in Figure 9b) reflects this theoretical bound.

A.0.3 GENERALIZATION TO NONREGULAR SIMPLICES

In the previous section, we found the theoretical upper bound on wLAPD for $q = 1$, i.e. when all the simplices are regular. It is hard to navigate the situation when q need be relaxed, which is always

the case in practice. We provide a numerical result from optimization in Figure 9b. It is clearly seen that wLAPD remains upper bounded when q is moderately relaxed.
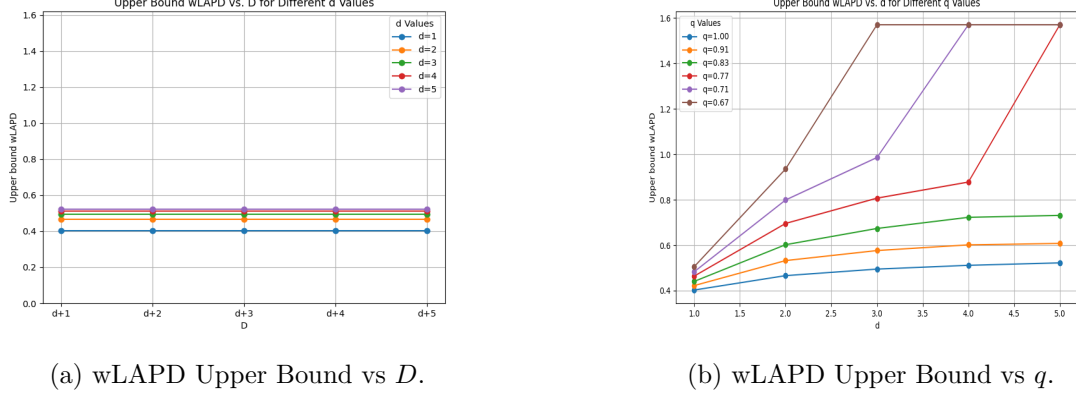

(a) wLAPD Upper Bound vs D.

(b) wLAPD Upper Bound vs q.

Figure 9: wLAPD upper bound simulation results. In (a), $q = 1$ is fixed. It verifies the independence of the optimial configuration and wLAPD upper bound from the ambient dimension D. In (b), $D = d+1$ is fixed but q is relaxed. It reveals that moderately relaxing q maintains a reasonable upper bound on wLAPD.

Appendix B. Proof of Theorem 4.4

We now prove the lower bound on bLAPD given in Theorem 4.4. Due to the relationship between the two weight constructions $\text{LAPD}^1 \geq \text{LAPD}^2$ by (17), it is sufficient to obtain a lower bound for bLAPD under the two-way weight construction:

$$W_S^2(\Delta_i, \Delta_j) = \min\{\pi - \theta_{ij}, \theta_{ij}\}.$$

Recall we have n points i.i.d sampled from the uniform volume measure vol_D on $T_\tau(\mathcal{M}) = T_\tau(\mathcal{M}_1) \cup T_\tau(\mathcal{M}_2)$, where vol_D refers to the D-dimensional Hausdorff measure. For adjacent simplices Δ_i, Δ_j, we say Δ_i ϵ-links to Δ_j (denoted $\Delta_i \to_\epsilon \Delta_j$) if $W_S^2(\Delta_i, \Delta_j) \leq \epsilon$. We also define a simplex to be *straddle* if it forms an angle of at least $\frac{\Theta}{4}$ with both $\mathcal{M}_1$ and $\mathcal{M}_2$, and let S^{str} denote the set of valid straddle simplices. To lower bound bLAPD by a path enumeration argument, we will need the following supporting results.

Lemma B.1 *Suppose $R \geq \frac{5e}{q \sin \Theta}$. Then any mixed simplex is formed by points contained in $T_R(\mathcal{M}_1 \cap \mathcal{M}_2)$.*

Proof Let P denote projection onto the intersection. We need to show that if Δ is a mixed simplex, then $\|x - Px\| \leq \frac{5e}{q \sin \Theta}$ for all $x \in \Delta$. Pick an arbitrary $x_1 \in \Delta$; without loss of generality assume $z_1 \in \mathcal{M}_1$, i.e. the noiseless sample corresponding to x_1 came from $\mathcal{M}_1$; since Δ is mixed, there exists $x_2 \in \Delta$ such that $z_2 \in \mathcal{M}_2$. Since $\|x_1 - x_2\| \leq \frac{e}{q}$ by construction of valid simplex set S in (2), $\|z_1 - z_2\| \leq \frac{e}{q} + 2\tau$ by the triangle inequality. Note:

$$z_2 - z_1 = Pz_1 - z_1 + Pz_2 - Pz_1 + z_2 - Pz_2 := p_1 + p_{\text{int}} + p_2$$

29

where $p_1 = Pz_1 - z_1$, $p_{\text{int}} = Pz_2 - Pz_1$, $p_2 = z_2 - Pz_2$. Note that p_{int} is orthogonal to $p_1 + p_2$, so

$$\|z_2 - z_1\| = \|p_{\text{int}}\| + \|p_1 + p_2\|$$
$$\implies \|p_1 + p_2\| \leq \|z_2 - z_1\| \leq \frac{e}{q} + 2\tau\,.$$

Furthermore, by the Law of Sines $\|p_1\| \leq \frac{\|p_1 + p_2\|}{\sin \Theta} \leq \frac{\frac{e}{q} + 2\tau}{\sin \Theta}$ and similarly for $\|p_2\|$. Thus we conclude $\|Pz_1 - z_1\| \leq \frac{\frac{e}{q} + 2\tau}{\sin \Theta}$. Finally we relate to the noisy distances:

$$\begin{aligned}
\|x_1 - Px_1\| &= \|x_1 - z_1 + z_1 - Pz_1 + Pz_1 - Px_1\| \\
&\leq \|x_1 - z_1\| + \|z_1 - Pz_1\| + \|Pz_1 - Px_1\| \\
&\leq \|x_1 - z_1\| + \|z_1 - Pz_1\| + \|z_1 - x_1\| \\
&\leq \frac{\frac{e}{q} + 2\tau}{\sin \Theta} + 2\tau \leq \frac{\frac{e}{q} + 4\tau}{\sin \Theta} \leq \frac{5e}{q \sin \Theta}
\end{aligned}$$

where the last inequality is due to Assumption 3.2. ∎

Lemma B.2 *Let n_T denote the number of points which land in $\mathcal{T}_R = T_R(\mathcal{M}_1 \cap \mathcal{M}_2) \cap T_\tau(\mathcal{M})$ for some $R \in [\tau, R_{\max}]$ as defined in Assumption 4.2. Then $n_T \leq 2C_2 nR$ with probability at least $1 - \exp\left(-\frac{C_1 nR}{3}\right)$, with C_1, C_2 defined in (14).*

Proof The number of points landing in $\mathcal{T}_R$ is a Binomial random variable with parameters (n, p) where p is the volume ratio between the tubular region of radius R around the $d-1$ dimensional intersection $\text{vol}_D(\mathcal{T}_R)$ and the tubular region of radius τ around the collection of d dimensional manifolds $\text{vol}_D(T_\tau(\mathcal{M}))$, i.e. by (14)

$$p = \frac{\text{vol}_D(\mathcal{T}_R)}{\text{vol}_D(T_\tau(\mathcal{M}))} \in \frac{[C_1, C_2] R \tau^{D-d}}{\text{vol}_d(\mathcal{M}) \tau^{D-d}} \sim R.$$

Hence, the Binomial distribution of n_T has a mean that scales like $\mu \sim nR$. A multiplicative Chernoff bound gives

$$\mathbb{P}(n_T \geq (1+t)\mu) \leq \exp\left(-\frac{t^2 \mu}{2+t}\right)$$

for $t \geq 0$. For convenience, we choose $t = 1$, and using (14), we have $C_1 nR \leq \mu \leq C_2 nR$, so that

$$\mathbb{P}(n_T \geq 2C_2 nR) \leq \mathbb{P}(n_T \geq 2\mu) \leq \exp\left(-\frac{\mu}{3}\right) \leq \exp\left(-\frac{C_1 nR}{3}\right)\,.$$

∎

Lemma B.3 *If $bLAPD \leq \epsilon$, then there exists a path of $\frac{\Theta}{2\epsilon}$ consecutive straddle simplices which link with a weight bounded by ϵ.*

Proof Since $bLAPD \leq \epsilon$, there exists a path of all weights $W_S^2(\Delta_i, \Delta_j)$ less than ϵ between the manifolds; that is, we have a path connecting the manifolds where the sequence $a_i = W_S^2(\Delta_i, \Delta_{i+1})$

30

is small. Consider also the dihedral angle made by Δ_i with the two manifolds $b_i = \angle(\Delta_i, \mathcal{M}_1)$ and $c_i = \angle(\Delta_i, \mathcal{M}_2)$. Note that we have the following triangle inequality

$$\angle(\Delta_i, \mathcal{M}_1) \leq W_S^2(\Delta_i, \Delta_{i+1}) + \angle(\Delta_{i+1}, \mathcal{M}_1) \leq \epsilon + \angle(\Delta_{i+1}, \mathcal{M}_1)$$

$$\angle(\Delta_{i+1}, \mathcal{M}_1) \leq W_S^2(\Delta_{i+1}, \Delta_i) + \angle(\Delta_i, \mathcal{M}_1) \leq \epsilon + \angle(\Delta_i, \mathcal{M}_1)$$

$$\implies |b_i - b_{i+1}| = |\angle(\Delta_i, \mathcal{M}_1) - \angle(\Delta_{i+1}, \mathcal{M}_1)| \leq \epsilon,$$

and similarly $|c_i - c_{i+1}| \leq \epsilon$. Now the sequence b_i increases from $O(\frac{\tau}{e})$ to $\Theta - O(\frac{\tau}{e})$ and c_i decreases from $\Theta - O(\frac{\tau}{e})$ to $O(\frac{\tau}{e})$. At some point b^* it is very close to $\frac{\Theta}{2}$. Note that since the sequence can move by at most ϵ in one step, if we take $\lfloor \frac{\Theta}{4\epsilon} \rfloor$ steps before and after b^*, we will not exit the interval $[\frac{1}{4}\Theta, \frac{3}{4}\Theta]$. That is, there has to be a sequence of $\lfloor \frac{\Theta}{2\epsilon} \rfloor$ consecutive links $\Delta_i \to \Delta_{i+1}$ where $b_i \in [\frac{1}{4}\Theta, \frac{3}{4}\Theta]$. Application of the triangle inequality ($b_i + c_i \geq \Theta$) gives $c_i \geq \Theta - b_i \geq \frac{\Theta}{4}$. Hence we conclude that there is a sequence of $\lfloor \frac{\Theta}{2\epsilon} \rfloor$ consecutive simplices whose angles with both $\mathcal{M}_1$ and $\mathcal{M}_2$ are at least $\frac{\Theta}{4}$, which is our definition of the straddle simplices. ∎

Proposition B.4 (Single Link Estimate) *Given a fixed simplex $[x_1, \ldots, x_{d+1}] = \Delta \in S^{str}$, the volume of the region where a new point y can be placed so that $\Gamma = [x_2, \ldots, x_{d+1}, y] \to_\epsilon \Delta$ is bounded by $C_{\Theta,q,d,r_0} e^d \epsilon \tau^{D-d}$ for ϵ small enough.*

Proof See Appendix D. ∎

Utilizing the above, we are ready to prove Theorem 4.4, which we restate here.

Theorem 4.4 (bLAPD lower bound) *Let Assumptions 2.1, 2.2, 3.1, 3.2, 4.1, and 4.2 hold and assume the construction outlined in Section 3 with either weight construction (4) or (16). If e satisfies $ne^d \geq \frac{2(d+1)}{C_5 \Theta \log 2} \log n$ and $e \leq 1$, then for n large enough*

$$bLAPD \geq (2C_5 n e^d)^{-1}$$

with probability at least $1 - 3C_4 n^{-(d+1)}$, where $C_4 = \left(\frac{2C_2}{q \sin \Theta}\right)^{d+1}$, $C_5 = \frac{2C_2 C_{\Theta,q,d}}{C_1}$, C_1, C_2 are the constants from (14), and C_{Θ,q,d,r_0} is the constant from Proposition B.4.

Proof We first control the number of points landing near the intersection. Defining the event $\Omega := \{n_T \leq 2C_2 nR\}$, we have $P(\Omega^C) \leq \exp\left(-\frac{C_1 nR}{3}\right)$ by Lemma B.2. Thus:

$$\mathbb{P}(bLAPD < \epsilon) = \mathbb{P}(bLAPD < \epsilon | \Omega) \mathbb{P}(\Omega) + \mathbb{P}(bLAPD < \epsilon | \Omega^C) \mathbb{P}(\Omega^C)$$

$$\leq \mathbb{P}(bLAPD < \epsilon | \Omega) + \mathbb{P}(\Omega^C)$$

$$\leq \mathbb{P}(bLAPD < \epsilon | n_T = w) + \mathbb{P}(\Omega^C)$$

where $w := \lfloor 2C_2 nR \rfloor$. Note that the conditional probability above is simply the probability that bLAPD $< \epsilon$ if we sample w i.i.d. points from $\mathcal{T}_R$. If bLAPD $< \epsilon$, then by Lemma B.3 there exists a path of $s = \frac{\Theta}{2\epsilon}$ consecutive straddle simplices which link with weights bounded by ϵ, i.e. there exists a path of straddle simplices such that $\Delta_1 \to_\epsilon \ldots \to_\epsilon \Delta_s$. We will denote $\Delta_i \to_{\epsilon,str} \Delta_j$ if $\Delta_i \to_\epsilon \Delta_j$ and Δ_i is straddle. By Lemma B.1, if we choose $R = \frac{5e}{q \sin \Theta}$, then all of these simplices are in the tube $\mathcal{T}_R$. We note there is a one-to-one correspondence between a simplex path $\Delta_1 \to \ldots \to \Delta_s$ of s simplices and a point path $x_1 \to \ldots \to x_{s+d}$ of $d + s$ points. Consider now a fixed simplex path $\Delta_1 \to \ldots \to \Delta_s$, i.e. the simplex path defined by a fixed $x_1 \to \ldots \to x_{s+d}$. Thus

$$\mathbb{P}(\Delta_1 \to_{\epsilon,str} \ldots \to_{\epsilon,str} \Delta_s) = \mathbb{P}(\Delta_1 \to_{\epsilon,str} \Delta_2) \mathbb{P}(\Delta_2 \to_{\epsilon,str} \Delta_3 | \Delta_1 \to_{\epsilon,str} \Delta_2)$$

$$\ldots \mathbb{P}(\Delta_{s-1} \to_{\epsilon,str} \Delta_s | \Delta_1 \to_{\epsilon,str} \ldots \to_{\epsilon,str} \Delta_{s-1})$$

Now by Proposition B.4, each term in the above product is bounded by

$$\frac{C_{\Theta,q,d}e^{d}\epsilon\tau^{D-d}}{\mathrm{vol}_D(\mathcal{T}_R)} \leq \frac{C_{\Theta,q,d}e^{d}\tau^{D-d}\epsilon}{C_1 R\tau^{D-d}} \leq \frac{C_{\Theta,q,d}e^{d}\epsilon q \sin\Theta}{5eC_1} = C_3 e^{d-1}\epsilon, \quad C_3 = \frac{C_{\Theta,q,d}q \sin\Theta}{5C_1},$$

so that for a fixed path,

$$\mathbb{P}(\Delta_1 \to_{\epsilon,\mathrm{str}} \ldots \to_{\epsilon,\mathrm{str}} \Delta_s) \leq \left(C_3 e^{d-1}\epsilon\right)^{s-1}.$$

Note here we have used the fact that

$$\begin{aligned}
\mathbb{P}(\Delta_i \to_{\epsilon,\mathrm{str}} \Delta_{i+1}) &= \mathbb{P}(\Delta_i \to_\epsilon \Delta_{i+1} \cap \Delta_i \in S^{\mathrm{str}}) \\
&= \mathbb{P}(\Delta_i \to_\epsilon \Delta_{i+1} | \Delta_i \in S^{\mathrm{str}})\mathbb{P}(\Delta_i \in S^{\mathrm{str}}) \\
&\leq \mathbb{P}(\Delta_i \to_\epsilon \Delta_{i+1} | \Delta_i \in S^{\mathrm{str}})
\end{aligned}$$

(similarly under conditioning on the previous links). Since we have w possible points from which to form paths, there are a total of $\frac{w!}{(w-(s+d))!} \leq w^{s+d}$ ordered sets of points of size $s + d$, so that

$$\begin{aligned}
\mathbb{P}(\mathrm{bLAPD} < \epsilon | n_T = w) &\leq w^{s+d}(C_3 e^{d-1}\epsilon)^{s-1} \\
&= w^{d+1}(C_3 w e^{d-1}\epsilon)^{s-1} \\
&= C_4(ne)^{d+1}(C_5 ne^d\epsilon)^{\frac{\Theta}{2\epsilon}-1},
\end{aligned}$$

where $C_4 = \left(\frac{10C_2}{q\sin\Theta}\right)^{d+1}$, $C_5 = \frac{10C_2 C_3}{q\sin\Theta} = \frac{2C_{\Theta,q,d}C_2}{C_1}$. We define $\epsilon := (2C_5 ne^d)^{-1}$ and note that

$$ne^d \geq \frac{2(d+1)}{C_5\Theta\log 2}\log n \quad \implies \quad \frac{\Theta}{4\epsilon} = \frac{\Theta C_5 ne^d}{2} \geq \frac{(d+1)\log n}{\log 2}$$

One can check that the above inequality implies $(ne)^{d+1} \leq (C_5 ne^d\epsilon)^{-\frac{\Theta}{4\epsilon}} = \left(\frac{1}{2}\right)^{-\frac{\Theta}{4\epsilon}} = 2^{\frac{\Theta}{4\epsilon}}$. To see this, note that since $e \leq 1$ by assumption, we have $(ne)^{d+1} \leq n^{d+1} = 2^{(d+1)\frac{\log n}{\log 2}} \leq 2^{\frac{\Theta}{4\epsilon}}$ and so

$$\mathbb{P}(\mathrm{bLAPD} < \epsilon | n_T = w) \leq C_4 2^{1-\frac{\Theta}{4\epsilon}} \leq 2C_4 2^{-(d+1)\frac{\log n}{\log 2}} = 2C_4 n^{-(d+1)}.$$

Note this probability dominates $\mathbb{P}(\Omega^C)$. In summary, we have shown that provided that

$$ne^d \geq \frac{2(d+1)}{C_5\Theta\log 2}\log n,$$

one has

$$\mathrm{bLAPD} > (2C_5 ne^d)^{-1}$$

with probability at least $1 - 2C_4 n^{-(d+1)} - \mathbb{P}(\Omega^C) \geq 1 - 3C_4 n^{-(d+1)}$ where $C_4 = \left(\frac{10C_2}{q\sin\Theta}\right)^{d+1}$ and $C_5 = \frac{2C_{\Theta,q,d}C_2}{C_1}$. ∎

Appendix C. Proof of Theorem 4.5

We now show how a very similar argument can be used to guarantee that if we incorporate a denoising procedure, we can achieve a bLAPD which is $O(\Theta)$. To make the argument precise, we define super straddle simplices to be simplices which have an angle of at least $\frac{3\Theta}{8}$ with both manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$.

Theorem 4.5 (bLAPD lower bound with denoising) *Let Assumptions 2.1, 2.2, 3.1, 3.2, 4.1, and 4.2 hold and assume the construction outlined in Section 3 with either weight construction (4) or (16). Assume also that e satisfies $ne^d \geq \frac{8(d+1)}{C_5 \Theta \log 2} \log n$, $e \leq 1$ and that we remove all simplices whose LAPD to their $\kappa = \frac{\Theta}{4} C_5 ne^d$-th nearest neighbor exceeds $\eta = (2C_5 ne^d)^{-1}$. Then for n large enough, if LAPD is recomputed after denoising,*

$$bLAPD_{dns} \geq \frac{\Theta}{4}$$

with probability at least $1 - 3C_4 n^{-(d+1)}$, where constants C_4, C_5 are as in Theorem 4.4.

Proof We argue that with high probability all super straddle simplices are removed during denoising, so that the between distance after denoising must be at least $\frac{\Theta}{4}$. Note that if a path starts at a super straddle simplex and takes ϵ-links, the path will contain only straddle simplices as long as the number of links does not exceed $s = \frac{\Theta}{8\epsilon}$. Now suppose there exists a super straddle simplex which survives denoising; then there exists a super straddle simplex that ϵ links to $s = \frac{\Theta}{8\epsilon}$ neighbors (all of which must be straddle). That is, there exists a path of simplices such that $\Delta_1 \to_{\epsilon,\text{str}} \cdots \to_{\epsilon,\text{str}} \Delta_s$. We simply repeat the proof of Theorem 4.4 to show this is unlikely, with $s = \frac{\Theta}{8\epsilon}$ replacing $s = \frac{\Theta}{2\epsilon}$. The only change is the constant. ∎

Appendix D. Single-link estimate: proof of Proposition B.4

We provide an estimate for linking straddle simplices, summarized in Proposition B.4. We first prove the result for the noiseless case, $\tau = 0$. In this case, the manifolds are strictly in a d-dimensional space, so that our estimates are with respect to d-dimensional volume.

Lemma D.1 (Noiseless single link estimate) *Let $\mathcal{M}_1$, $\mathcal{M}_2$, S, d, q, e, r_0 be fixed that satisfy Assumptions 2.1, 3.1 and 4.1. Then there are positive constants ϵ_0 and C such that the following is true: Consider any d-simplex $\Delta = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{d+1}] \in S^{str}$. With $\boldsymbol{v}$ any point in $\mathcal{M}_2$, let $\Delta_j(\boldsymbol{v})$ denote a simplex formed by the $(d+1)$ points $\boldsymbol{v}$ and $\{\boldsymbol{x}_\ell\}_{\ell \neq j}$. Consider the set of points in $\mathcal{M}_2$ such that a new valid straddle simplex $\Delta_j(\boldsymbol{v})$ can be formed adjacent to Δ for any j:*

$$A := \left\{ \boldsymbol{v} \in \mathcal{M}_2 \ \big| \ \Delta_j(\boldsymbol{v}) \in S \text{ and } W_S^2(\Delta, \Delta_j(\boldsymbol{v})) \leq \epsilon \text{ for any } j \in [d+1] \right\}. \qquad (19)$$

Then whenever $\epsilon \leq \epsilon_0$,

$$\text{vol}_d(A) \leq Ce^d \epsilon$$

Proof First note that by Assumption 2.1, $\mathcal{M}$ is contained in a $(d+1)$-dimensional Euclidean space, and so without loss of generality we assume $D = d + 1$ throughout the proof. Within this proof, we use boldface letters, e.g. $\boldsymbol{x}$ to denote vectors or points in multi-dimensional space, and normally typeset letters, e.g., x to denote scalars. Our overall strategy will be to use subadditivity of the volume of A. With

$$A_j = \left\{ \boldsymbol{v} \in \mathcal{M}_2 \ \big| \ \Delta_j(\boldsymbol{v}) \in S \text{ and } W_S^2(\Delta, \Delta_j(\boldsymbol{v})) \leq \epsilon \right\},$$

then

$$A = \bigcup_{j \in [d+1]} A_j \quad \implies \quad \text{vol}_d(A) \leq \sum_{j \in [d+1]} \text{vol}_d(A_j),$$

33

and we will proceed to compute $\mathrm{vol}_d(A_j)$. Fixing $j \in [d+1]$, and taking some $\boldsymbol{v} \in A_j$, we will study the cosine of the angle between Δ and $\Delta_j(\boldsymbol{v})$:

$$\alpha_j(\boldsymbol{v}) := \boldsymbol{n}(\Delta) \cdot \boldsymbol{n}(\Delta_j(\boldsymbol{v})) = \cos \angle (\Delta, \Delta_j(\boldsymbol{v})).$$

where $\boldsymbol{n}(\cdot)$ denotes a(ny) unit normal vector to the input d-simplex. Then by definition we have,

$$\boldsymbol{v} \in A_j \implies W_S^2(\Delta, \Delta_j(\boldsymbol{v})) \leq \epsilon \overset{(16)}{\implies} |\alpha_j(\boldsymbol{v})| \geq 1 - \frac{\epsilon^2}{2},$$

where we have used $|\cos x| \geq 1 - x^2/2$. Hence we have,

$$A_j \subset B_j, \qquad B_j := \left\{ \boldsymbol{v} \in \mathcal{M}_2 \mid \Delta_j(\boldsymbol{v}) \in S \text{ and } |\alpha_j(\boldsymbol{v})| \geq 1 - \epsilon^2/2 \right\}.$$

We proceed to bound the volume of B_j, which we accomplish by deriving necessary conditions on the coordinates of $\boldsymbol{v}$ for B_j membership. Without loss we assume $\boldsymbol{x}_{d+1} \in \mathcal{M}_2$. (If not, re-index the vertices of Δ. Because $\Delta \in S^{\mathrm{str}}$, at least one of its vertices lies in $\mathcal{M}_2$.) We can assume a certain geometric configuration for the problem: Volumes, angles, simplex qualities, and edge lengths are all invariant under affine unitary transformations, i.e., rigid rotations, reflections, and translations. In addition, we subsequently perform an isotropic scaling to extract dependence on e. In D-dimensional space, we choose such a transformation corresponding to, in order, a rigid translation, followed by $d = D - 1$ rotations, followed by isotropic scaling, so this mapping corresponds to the map $\boldsymbol{x} \mapsto \boldsymbol{z}_* \mapsto \boldsymbol{z}_{**} \mapsto \boldsymbol{z}_{***} \mapsto \boldsymbol{z}$:

$$\boldsymbol{z}_*(\boldsymbol{x}_{d+1}) = \boldsymbol{0}, \tag{20a}$$

$$\boldsymbol{z}_{**}(\mathcal{M}_2) \subset \mathrm{span}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\} \tag{20b}$$

$$\boldsymbol{z}_{***}\left(\mathrm{span}\{\mathrm{span}\{\Delta\} \cap \mathcal{M}_2\}\right) = \mathrm{span}\{\boldsymbol{e}_2, \ldots, \boldsymbol{e}_d\} \tag{20c}$$

$$\boldsymbol{z} = \frac{\boldsymbol{z}_{***}}{e} \tag{20d}$$

Condition (20a) is a rigid translation. Condition (20b) corresponds to rotation of a normal vector to $\mathcal{M}_2$ passing through $\boldsymbol{x}_{d+1} = \boldsymbol{0}$. Condition (20c) corresponds to a $(d-1)$-dimensional rotation around a normal vector to $\mathcal{M}_2$ passing through $\boldsymbol{x}_{d+1} = \boldsymbol{0}$. In summary, the map $\boldsymbol{x} \mapsto \boldsymbol{z}$ has the form,

$$\boldsymbol{z} = \frac{1}{e} \boldsymbol{U} \left(\boldsymbol{x} - \boldsymbol{x}_{d+1}\right), \qquad\qquad \boldsymbol{U} \text{ is unitary}$$

We also define $\boldsymbol{z}$-images of the given vertices of $\Delta = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{d+1}]$:

$$\boldsymbol{z}_j := \boldsymbol{z}(\boldsymbol{x}_j), \qquad\qquad j \in [d+1], \qquad\qquad \boldsymbol{z}_{d+1} = \boldsymbol{0}$$

While distances change by a factor of e under the $\boldsymbol{x} \mapsto \boldsymbol{z}$ mapping, angles do not because the terminal scaling factor is isotropic. The $\boldsymbol{z}$-image of the set B_j is a rigid translation, rotation, and isotropic scaling of B_j, and hence satisfies,

$$C_j := \boldsymbol{z}(B_j), \qquad\qquad \mathrm{vol}_d(B_j) = e^d \mathrm{vol}_d(C_j),$$

We define corresponding quantities and notation in $\boldsymbol{z}$-space:

- With $\boldsymbol{y} = \boldsymbol{z}(\boldsymbol{v})$, then $\boldsymbol{v} \in B_j \iff \boldsymbol{y} \in C_j$.

- $\widetilde{\Delta} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{d+1}]$ is the $\boldsymbol{z}$-image of Δ

- $\widetilde{\Delta}_j(\boldsymbol{y}) = [\boldsymbol{y}, \{\boldsymbol{z}_\ell\}_{\ell \neq j}]$ is the simplex adjacent to $\widetilde{\Delta}$ sharing all vertices except $\boldsymbol{z}_j$.

Two particularly useful consequences of this $\boldsymbol{z} \mapsto \boldsymbol{x}$ map are,

$$\alpha_j(\boldsymbol{v}) = \boldsymbol{n}(\Delta) \cdot \boldsymbol{n}(\Delta_j(\boldsymbol{v})) = \boldsymbol{n}(\widetilde{\Delta}) \cdot \boldsymbol{n}(\widetilde{\Delta}_j(\boldsymbol{y})) =: \widetilde{\alpha}_j(\boldsymbol{y}),$$
$$\Delta_j(\boldsymbol{v}) \in S \iff \widetilde{\Delta}_j(\boldsymbol{y}) \in \widetilde{S}$$

where $\widetilde{S}$ is the set of valid simplices (cf. Assumption 3.1) whose minimal edge length is unity. I.e., $\Gamma = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{d+1}] \in \widetilde{S}$ means,

$$r_0 V_0 \le \text{vol}_d(\Gamma) \le \frac{V_0}{q^d}, \qquad\qquad 1 \le |\boldsymbol{u}_i - \boldsymbol{u}_j| \le \frac{1}{q} \ \ \forall \, i, j \in [d+1], \ i < j.$$

In summary, we have the following equivalent definition for C_j:

$$C_j = \left\{ \boldsymbol{y} \in \text{span}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\} \ \middle| \ \widetilde{\Delta}_j(\boldsymbol{y}) \in \widetilde{S} \text{ and } |\widetilde{\alpha}_j(\boldsymbol{y})| \ge 1 - \epsilon^2/2 \right\}.$$

The conditions defining C_j allow us to derive constraints on the coordinates of a fixed, arbitrary element $\boldsymbol{y} \in C_j$. By the property that $\boldsymbol{z}(\mathcal{M}_2)$ lies in $\text{span}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$, then,

$$\boldsymbol{y} = (y_1, \ldots, y_d, 0)^T \in \mathbb{R}^D$$

The fact that $\widetilde{\Delta}_j(\boldsymbol{y})$ is a valid simplex provides a bound for $|\boldsymbol{y}|$:

$$|\boldsymbol{y}| - |\boldsymbol{z}_j| \le |\boldsymbol{y} - \boldsymbol{z}_j| \overset{(2)}{\le} \frac{1}{q}$$

This, along with the fact that $\boldsymbol{z}_j = \boldsymbol{z}_j - \boldsymbol{z}_{d+1}$ is an edge for the valid simplex $\widetilde{\Delta}$ results in the bound,

$$\boldsymbol{y} \in C_j \quad \implies \quad |\boldsymbol{y}| \le \frac{1}{q} + |\boldsymbol{z}_j| \overset{(2)}{\le} \frac{2}{q}. \tag{21}$$

This provides a bound on each coordinate of $\boldsymbol{y}$:

$$|y_\ell| \le |\boldsymbol{y}| \le \frac{2}{q}, \qquad\qquad\qquad \ell \in [d]. \tag{22}$$

The remainder of our arguments will craft an additional, more stringent constraint for y_1 that is necessary to achieve $|\widetilde{\alpha}_j(\boldsymbol{y})| \ge 1 - \epsilon^2/2$. To this end, an important observation is that by definition $|\widetilde{\alpha}_j(\boldsymbol{y})| = 1$ iff $\widetilde{\Delta}$ and $\widetilde{\Delta}_j(\boldsymbol{y})$ are coplanar, which occurs iff $\boldsymbol{y} \in \text{span}\{\widetilde{\Delta}\}$. I.e.,

$$\boldsymbol{y} \in C_j \text{ and } |\widetilde{\alpha}_j(\boldsymbol{y})| = 1 \iff \boldsymbol{y} \perp \boldsymbol{e}_{d+1} \text{ and } \boldsymbol{y} \in \text{span}\{\widetilde{\Delta}\} \overset{(20c)}{\iff} \boldsymbol{y} = (0, y_2, \ldots, y_d, 0)^T \tag{23}$$

An informal continuity argument then suggests that membership in C_j, i.e., "large" values of $\widetilde{\alpha}_j(\boldsymbol{y})$, requires "small" values of y_1. We make this precise below. To derive a formula for $\widetilde{\alpha}_j(\boldsymbol{y})$, we compute expressions for the normal vectors $\boldsymbol{n}(\widetilde{\Delta})$ and $\boldsymbol{n}(\widetilde{\Delta}_j(\boldsymbol{y}))$. We define ψ as the angle between $\widetilde{\Delta}$ and $\boldsymbol{z}(\mathcal{M}_2)$, i.e.,

$$\psi = W^2(\widetilde{\Delta}, \boldsymbol{z}(\mathcal{M}_2)) = \angle(\widetilde{\Delta}, \boldsymbol{z}(\mathcal{M}_2)) = \angle(\Delta, \mathcal{M}_2) \in \left[\frac{\Theta}{4}, \frac{\pi}{2}\right],$$

where $\psi \ge \Theta/4$ because $\Delta \in S^{\text{str}}$. Then $\boldsymbol{n}(\widetilde{\Delta})$ can be determined by noting that it's a unit vector, that $\boldsymbol{e}_2, \ldots \boldsymbol{e}_d$ are all vectors in the span of $\widetilde{\Delta}$, and also that $\left|\boldsymbol{n}(\widetilde{\Delta}) \cdot \boldsymbol{n}(\boldsymbol{z}(\mathcal{M}_2))\right| = \left|\boldsymbol{n}(\widetilde{\Delta}) \cdot \boldsymbol{e}_{d+1}\right| = \cos\psi$:

$$\boldsymbol{n}(\widetilde{\Delta}) \cdot \boldsymbol{e}_\ell = 0, \ \ell \in \{2, \ldots, d\} \quad \implies \quad \boldsymbol{n}(\widetilde{\Delta}) = \sigma_1 \boldsymbol{e}_1 \sin\psi + \sigma_2 \boldsymbol{e}_{d+1} \cos\psi,$$

where $\sigma_1, \sigma_2 \in \{-1, 1\}$ are some fixed signs (determined uniquely by any particular choice of $\boldsymbol{n}(\widetilde{\Delta})$). Let $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_d$ be the set of d vertices of $\widetilde{\Delta}_j(\boldsymbol{y})$ that are shared by $\widetilde{\Delta}$. I.e., they are any identification of the set of d vectors $\{\boldsymbol{z}_\ell\}_{\ell \in [d+1] \setminus \{j\}}$. The remaining vertex is $\boldsymbol{y}$. Then a normal vector to $\widetilde{\Delta}_j(\boldsymbol{y})$ is given by,

$$\boldsymbol{n}(\widetilde{\Delta}_j(\boldsymbol{y})) = \frac{\boldsymbol{m}}{|\boldsymbol{m}|}, \qquad \boldsymbol{m} = \det \begin{pmatrix} \boldsymbol{e}_1 & \cdots & \boldsymbol{e}_{d+1} \\ - & (\boldsymbol{y} - \boldsymbol{w}_1)^T & - \\ - & (\boldsymbol{w}_2 - \boldsymbol{w}_1)^T & - \\ & \vdots & \\ - & (\boldsymbol{w}_d - \boldsymbol{w}_1)^T & - \end{pmatrix}$$

We make the observations that $\boldsymbol{m}$ is a multilinear function of the entries in the second row, and hence the entries of $\boldsymbol{m}$ are affine functions of y_1. Then since $\boldsymbol{n}(\widetilde{\Delta})$ is independent of $\boldsymbol{y}$, we have,

$$\widetilde{\alpha}_j(\boldsymbol{y}) = \boldsymbol{n}(\widetilde{\Delta}_j(\boldsymbol{y})) \cdot \boldsymbol{n}(\widetilde{\Delta}) = \frac{\boldsymbol{m}}{|\boldsymbol{m}|} \cdot \boldsymbol{n}(\widetilde{\Delta}) = \frac{A + By_1}{\sqrt{A_0^2 + \sum_{\ell \in [d]} (B_\ell y_1 + C_\ell)^2}},$$

where $A, B, A_0, B_\ell, C_\ell$ all depend *smoothly* (are infinitely differentiable) with respect to $\psi, y_2, \ldots, y_d$ because the entries of $\boldsymbol{m}$ are smooth functions of these parameters. Note that by the Cauchy-Schwarz inequality,

$$|\widetilde{\alpha}_j(\boldsymbol{y})| \le 1, \tag{24}$$

and recall from (23) that $|\widetilde{\alpha}_j(\boldsymbol{y})| = 1$ iff $y_1 = 0$. When $y_1 = 0$, let $\widetilde{\alpha}_j(0, y_2, \ldots, y_d, 0) = \sigma \in \{+1, -1\}$. Note that it cannot switch signs as a function of $\psi, y_2, \ldots, y_d$ due to continuity in those variables, and hence σ is a constant in those variables. To explicitly separate dependence of $\widetilde{\alpha}_j$ on y_1 versus other quantities, we'll write,

$$\widetilde{\alpha}_j(\boldsymbol{y}) = \alpha_j(y_1, \boldsymbol{h}), \qquad \boldsymbol{h} = (y_2, ,\ldots, y_d, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_d) \in H(y_1) \subset \mathbb{R}^{d-1+d(d+1)},$$

The set $H(y_1)$ in $\mathbb{R}^{d-1+d(d+1)}$ corresponds to the set of values that conform to our assumptions; as suggested by the notation, this set depends on y_1. We are concerned only with $y_1 = 0$. We claim the set $H(0)$ is compact, for which it suffices to show it's closed and bounded. The bound (22) on the entries $y_2, \ldots, y_d$ of $\boldsymbol{y}$, implies that the first $d - 1$ components of $\boldsymbol{h}$ are bounded. The values of each element of $\boldsymbol{w}_j$, $j \in [d]$ is bounded because $\boldsymbol{w}_j$ is a bounded vector:

$$\widetilde{S} \ni \Delta = [\boldsymbol{0}, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_d] \implies \|\boldsymbol{w}_j\| \le 1/q,$$

Hence, $H(0)$ is a bounded set. An arbitrary $\boldsymbol{h}$ has membership in $H(0)$ iff with $y_1 = 0$,

$$f_1(\boldsymbol{h}) = \mathrm{vol}_d(\widetilde{\Delta}) \in \left[r_0 V_0, \frac{V_0}{q^d} \right], \qquad \text{(Volume distortion of } \widetilde{\Delta})$$

$$f_2(\boldsymbol{h}) = \mathrm{vol}_d(\widetilde{\Delta}_j(\boldsymbol{y})) \in \left[r_0 V_0, \frac{V_0}{q^d} \right], \qquad \text{(Volume distortion of } \widetilde{\Delta}_j(\boldsymbol{y}))$$

$$f_3(\boldsymbol{h}) = \min_{\substack{\boldsymbol{a}, \boldsymbol{b} \in \mathrm{ex}(\widetilde{\Delta}) \\ \boldsymbol{a} \ne \boldsymbol{b}}} \|\boldsymbol{a} - \boldsymbol{b}\| \in \left[1, \frac{1}{q} \right], \qquad \text{(Edgelength distortion of } \widetilde{\Delta})$$

$$f_4(\boldsymbol{h}) = \min_{\substack{\boldsymbol{a}, \boldsymbol{b} \in \mathrm{ex}(\widetilde{\Delta}_j(\boldsymbol{y})) \\ \boldsymbol{a} \ne \boldsymbol{b}}} \|\boldsymbol{a} - \boldsymbol{b}\| \in \left[1, \frac{1}{q} \right], \qquad \text{(Edgelength distortion of } \widetilde{\Delta}_j(\boldsymbol{y}))$$

36

Because (i) the volume of any simplex is a continuous function of its extreme points, and (ii) the norm $\|\cdot\|$ is a continuous function of its argument, and (iii) the function $\min(\cdot,\cdot,\ldots,\cdot) : \mathbb{R}^{d(d+1)/2} \to \mathbb{R}$ is a continuous function of its inputs, then the map $\boldsymbol{h} \mapsto \boldsymbol{f}(\boldsymbol{h})$ is a continuous function, and,

$$H(0) = \boldsymbol{f}^{-1}(C), \qquad\qquad C = \left[r_0 V_0, \frac{V_0}{q^d}\right]^2 \times \left[1, \frac{1}{q}\right]^2 \subset \mathbb{R}^4$$

where again we take $y_1 = 0$. Since C is a closed set and $\boldsymbol{f}$ is continuous, then H is closed. Therefore, we've shown that $H(0)$ is bounded and continuous, hence compact. For a fixed but arbitrary $\boldsymbol{h} \in H(0)$, we use the shorthand notation $\widetilde{\alpha}'_j(y_1, \boldsymbol{h}) = \frac{\partial}{\partial y_1}\widetilde{\alpha}_j(y_1, \boldsymbol{h})$. Then:

- $\sigma\widetilde{\alpha}_j(0, \boldsymbol{h}) = \sigma^2 = 1$, due to (23).

- $\widetilde{\alpha}'_j(0, \boldsymbol{h}) = 0$. This is true because (24) implies that $\widetilde{\alpha}_j(\cdot, \boldsymbol{h})$ reaches its global maximum at $y_1 = 0$, and since by (23) $|\widetilde{\alpha}_j(y_1, \boldsymbol{h})| < 1$ for $y_1 \neq 0$, then $y_1 = 0$ is a strict global extremum of the smooth function $\widetilde{\alpha}_j(y_1, \boldsymbol{h})$, so necessarily $y_1 = 0$ is a stationary point.

- $\sigma\widetilde{\alpha}''_j(0, \boldsymbol{h}) < 0$. If this were false, then the Mean Value Theorem would imply that $\sigma\widetilde{\alpha}_j(\cdot, \boldsymbol{h})$ would be locally non-decreasing around 0, which violates the global extremum property of $y_1 = 0$.

The observations above imply that $\widetilde{\alpha}_j$ is y_1-stationary and strictly concave at $y_1 = 0$ for fixed $\boldsymbol{h}$. Our ultimate goal is to show $\boldsymbol{h}$-uniform and Δ-uniform *strong* concavity. I.e., we seek to show that $\sigma\widetilde{\alpha}''_j \leq m < 0$, where m is independent both of $\boldsymbol{h} \in H(0)$ and also of the particular choice of simplex $\Delta \in S$.

We define,

$$K = \min_{\boldsymbol{h} \in H(0)} -\sigma\widetilde{\alpha}''_j(0, \boldsymbol{h}) > 0,$$

where the inequality holds because $\widetilde{\alpha}''(\cdot, \boldsymbol{h})$ is a smooth (in particular continuous) function that is strictly negative on the compact set $H(0)$. Note that because we minimize over all elements $\boldsymbol{h} \in H(0)$, i.e., over all valid configurations of points in $\mathcal{M}_1$ and $\mathcal{M}_2$, then K is *independent* of the particular simplex Δ we start with, and depends only on manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$, their dimension and intersection angle d, Θ, and the geometric parameters e, q, r_0 defining our simplices. Given this value of K, we now consider a fixed $\boldsymbol{h}$ and define a quantity $\eta(\boldsymbol{h})$ as,

$$\eta(\boldsymbol{h}) = \max\left\{\eta > 0 \mid \sigma\widetilde{\alpha}''_j(y_1, \boldsymbol{h}) \leq -\frac{K}{2} \text{ for all } |y_1| \leq \eta\right\} > 0.$$

We note the continuity of $\widetilde{\alpha}''(y_1, \boldsymbol{h})$ in y_1 and $\boldsymbol{h}$ imply strict positivity $\eta(\boldsymbol{h}) > 0$ and continuity of $\eta(\boldsymbol{h})$ in $\boldsymbol{h}$, respectively. We will next need the smallest value of η:

$$\eta_0 := \min_{\boldsymbol{h} \in H(0)} \eta(\boldsymbol{h}) > 0,$$

where the strict inequality is again appeals to the fact that $\eta(\cdot)$ is strictly negative over the compact set $H(0)$. In addition, by minimizing over all $\boldsymbol{h} \in H(0)$, then η_0 is also *independent* of the particular simplex Δ, and depends only on $\mathcal{M}_1, \mathcal{M}_2, d, \Theta, e, q, r_0$. By this definition, $(-\eta_0, \eta_0)$ is an interval where $\sigma\widetilde{\alpha}''(\cdot, \boldsymbol{h})$ is no larger than $-K/2$ for all $\boldsymbol{h} \in H$. Hence, applying Taylor's Theorem (with the Mean Value remainder) to the function $y_1 \mapsto \widetilde{\alpha}_j(y_1, \boldsymbol{h})$ yields the $\boldsymbol{h}$-uniform bound,

$$\sigma\widetilde{\alpha}_j(y_1, \boldsymbol{h}) \leq \sigma\widetilde{\alpha}_j(0, \boldsymbol{h}) + \sigma\widetilde{\alpha}'_j(0, \boldsymbol{h})\, y_1 + \frac{-K/2}{2}y_1^2$$

$$= 1 - \frac{K}{4}y_1^2,$$

uniformly in $\boldsymbol{h}$ whenever $|y_1| \leq \eta_0$. Finally, note that if $\boldsymbol{y} \in C_j$, then

$$\frac{\epsilon^2}{2} \geq 1 - |\widetilde{\alpha}_j(y_1; \boldsymbol{h})| = 1 - \sigma \widetilde{\alpha}_j(y_1; \boldsymbol{h}) \geq \frac{K}{4} y_1^2,$$

whenever $y_1 \leq \eta_0$. Hence,

$$\boldsymbol{y} \in A_j \text{ and } \epsilon \leq \eta_0 \sqrt{\frac{K}{2}} \quad \implies \quad |y_1| \leq \epsilon \sqrt{\frac{2}{K}}. \tag{25}$$

Then combining (22) with (25), we conclude that if $\epsilon < \eta_0 \sqrt{\frac{K}{2}}$, then

$$\mathrm{vol}_d(B_j) = e^d \mathrm{vol}_d(C_j) \leq e^d 2^d \epsilon \sqrt{\frac{2}{K}} \left(\frac{2}{q}\right)^{d-1} \leq \epsilon e^d \frac{2^{2d-1}}{q^{d-1}\sqrt{K}},$$

and finally exercising subadditivity of volume,

$$\mathrm{vol}_d(A) \leq \sum_{j=0}^{d} \mathrm{vol}_d(A_j) \leq \sum_{j=0}^{d} \mathrm{vol}_d(B_j) \leq \epsilon e^d \left[\frac{(d+1)2^{2d-1}}{q^{d-1}\sqrt{K}}\right] =: C \epsilon e^d.$$

This bound requires $\epsilon \leq \epsilon_0 := \eta_0 \sqrt{\frac{K}{2}}$, but both η_0 and K are independent of Δ, and depend only on $\mathcal{M}_1, \mathcal{M}_2, d, \Theta, e, q, r_0$. ∎

The argument presented in the proof of Lemma D.1 contains the essential idea for the desired estimate in Proposition B.4. We expect the $\tau > 0$ to have essentially the same character but to be considerably more technical because noise in the nuisance $D - d$ dimensions can lift points $\boldsymbol{y}$ off the manifold $\mathcal{M}_2$. Therefore, in lieu of formally generalizing this to the noisy $\tau > 0$ case, we provide an informal argument for the noisy setup: when $\tau > 0$, we would be interested in bounding the volume of a tubular version of the set A in (19):

$$A_\tau := \left\{\boldsymbol{y} \in T_\tau(\mathcal{M}_2) \mid \Delta_j(\boldsymbol{y}) \in S \text{ and } W_S^2(\Delta, \Delta_j(\boldsymbol{y})) \leq \epsilon \text{ for any } j = 0, \ldots, d\right\}. \tag{26}$$

Since this region is in D-dimensional space, we would bound $\mathrm{vol}_D(A_\tau)$. As in the proof of Lemma D.1, we would use membership in the set A_τ to derive bounds on the coordinates of a candidate point $\boldsymbol{y} \in \mathbb{R}^D$. The treatement of the first d coordinates would be similar, corresponding to points on $\mathcal{M}_2$. The treatment of the remaining $D - d$ coordinates would be subject to perturbations bounded by τ. The bounds for these remaining coordinates would be dictated by ensuring $|\alpha_j(\boldsymbol{y}, q)| \geq 1 - \epsilon^2/2$. Because $\alpha_j(\boldsymbol{y}, q)$ is continuous in the coordinates of $\boldsymbol{y}$, a rough Taylor series argument suggests that,

$$|y_j| \leq \tau \text{ for } j \in [d+1, D] \quad \sim \quad |\alpha_j(\boldsymbol{y}, q)| \gtrsim 1 - \epsilon^2,$$

for $\tau \lesssim \epsilon$. Then the remaining $D - d$ dimensions correspond to the restriction $|y_j| \leq \tau$, and so there is a constant C such that,

$$\mathrm{vol}_D(A_\tau) \sim C \epsilon e^d \tau^{D-d},$$

which is the bound stated in Proposition B.4.

References

Maryam Abdolali and Nicolas Gillis. Beyond linear subspace clustering: A comparative study of nonlinear manifold clustering algorithms. *Computer Science Review*, 42:100435, November 2021. ISSN 1574-0137. doi: 10.1016/j.cosrev.2021.100435. URL https://www.sciencedirect.com/science/article/pii/S1574013721000757.

Ery Arias-Castro, Gilad Lerman, and Teng Zhang. Spectral Clustering Based on Local PCA. *Journal of Machine Learning Research*, 18(9):1–57, 2017. ISSN 1533-7928. URL `http://jmlr.org/papers/v18/14-318.html`.

Amir Babaeian, Alireza Bayestehtashk, and Mojtaba Bandarabadi. Multiple Manifold Clustering Using Curvature Constrained Path. *PLOS ONE*, 10(9):e0137986, September 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0137986. URL `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0137986`. Publisher: Public Library of Science.

R. Basri and D.W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):218–233, 2003. ISSN 0162-8828. doi: 10.1109/TPAMI.2003.1177153.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003. doi: 10.1162/089976603321780317.

Hoss Belyadi, Alireza Haghighat, and Doug Yarranton. *Machine Learning Guide for Oil and Gas Using Python - A Step-by-Step Breakdown with Data, Algorithms, Codes, and Applications*. Elsevier, 2021. ISBN 9780323902321.

P.S. Bradley and O.L. Mangasarian. k-Plane Clustering. *Journal of Global Optimization*, 16(1): 23–32, January 2000. ISSN 1573-2916. doi: 10.1023/A:1008324625522. URL `https://doi.org/10.1023/A:1008324625522`.

Tadeusz Caliński and Harabasz JA. A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3:1–27, 01 1974. doi: 10.1080/03610927408827101.

Paolo M. Camerini. The min-max spanning tree problem and some extensions. *Information Processing Letters*, 7(1):10–14, 1978.

Haoyu Chen, Anna Little, and Akil Narayan. Largest angle path distance for multi-manifold clustering. In *2023 International Conference on Sampling Theory and Applications (SampTA)*, pages 1–7, 2023. doi: 10.1109/SampTA59647.2023.10301401.

J.P. Costeira and T. Kanade. A Multibody Factorization Method for Independently Moving Objects. *International Journal of Computer Vision*, 29(3):159–179, 1998. ISSN 0920-5691. doi: 10.1023/A:1008000628999.

Li Deng. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.

Ehsan Elhamifar and René Vidal. Sparse Manifold Clustering and Embedding. In *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011. URL `https://proceedings.neurips.cc/paper/2011/hash/fc490ca45c00b1249bbe3554a4fdf6fb-Abstract.html`.

Ehsan Elhamifar and René Vidal. Sparse Subspace Clustering: Algorithm, Theory, and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, November 2013. ISSN 1939-3539. doi: 10.1109/TPAMI.2013.57. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

Harold N Gabow and Robert E Tarjan. Algorithms for two bottleneck optimization problems. *Journal of Algorithms*, 9(3):411–417, 1988.

Attila Gere. Recommendations for validating hierarchical clustering in consumer sensory projects. *Current Research in Food Science*, 6:100522, 2023. ISSN 2665-9271. doi: https://doi.org/10.1016/j.crfs.2023.100522. URL `https://www.sciencedirect.com/science/article/pii/S2665927123000904`.

Dian Gong, Xuemei Zhao, and Gérard Medioni. Robust multiple manifolds structure learning. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, pages 25–32, Madison, WI, USA, June 2012. Omnipress. ISBN 978-1-4503-1285-1.

Trevor Hastie and Patrice Simard. Metrics and Models for Handwritten Character Recognition. In Carlo Malaguerra, Stephan Morgenthaler, and Elvezio Ronchetti, editors, *Conference on Statistical Science Honouring the Bicentennial of Stefano Franscini's Birth*, Monte Verità, pages 203–219, Basel, 1997. Birkhäuser. ISBN 978-3-0348-8930-8. doi: 10.1007/978-3-0348-8930-8_16.

J. Ho, Ming-Husang Yang, Jongwoo Lim, Kuang-Chih Lee, and D. Kriegman. Clustering appearances of objects under varying illumination conditions. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 1, pages I–I, June 2003. doi: 10.1109/CVPR.2003.1211332. ISSN: 1063-6919.

TC Hu. The maximum capacity route problem. *Operations Research*, 9(6):898–900, 1961.

J. J. Hull. A database for handwritten text recognition research. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5):550–554, 1994. doi: 10.1109/34.291440.

Pan Ji, T. Zhang, Hongdong Li, Mathieu Salzmann, and Ian D. Reid. Deep subspace clustering networks. In *Neural Information Processing Systems*, 2017. URL `https://api.semanticscholar.org/CorpusID:3470712`.

Jun Li, Hongfu Liu, Zhiqiang Tao, Handong Zhao, and Yun Fu. Learnable subspace clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 33(3):1119–1133, 2022a. doi: 10.1109/TNNLS.2020.3040379.

Zengyi Li, Yubei Chen, Yann LeCun, and Friedrich T. Sommer. Neural manifold clustering and embedding. *ArXiv*, abs/2201.10000, 2022b. URL `https://api.semanticscholar.org/CorpusID:246275780`.

John Lipor, David Hong, Yan Shuo Tan, and Laura Balzano. Subspace clustering using ensembles of K-subspaces. *Information and Inference: A Journal of the IMA*, 10(1):73–107, 11 2020. ISSN 2049-8772. doi: 10.1093/imaiai/iaaa031. URL `https://doi.org/10.1093/imaiai/iaaa031`.

Anna Little, Mauro Maggioni, and James M. Murphy. Path-based spectral clustering: guarantees, robustness to outliers, and fast algorithms. *The Journal of Machine Learning Research*, 21(1): 6:184–6:249, 2020.

Anna V. Little, Mauro Maggioni, and Lorenzo Rosasco. Multiscale geometric methods for data sets i: Multiscale svd, noise and curvature. *Applied and Computational Harmonic Analysis*, 43(3):504–567, 2017. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2015.09.009. URL `https://www.sciencedirect.com/science/article/pii/S1063520315001414`.

G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2012.

Ryan McConville, Raúl Santos-Rodríguez, Robert J Piechocki, and Ian Craddock. N2d: (not too) deep clustering via clustering the local manifold of an autoencoded embedding. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 5145–5152, 2021. doi: 10.1109/ICPR48806.2021.9413131.

Sameer A Nene, Shree K Nayar, Hiroshi Murase, et al. Columbia object image library (coil-20). 1996.

Vishal M. Patel and René Vidal. Kernel sparse subspace clustering. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 2849–2853, 2014. doi: 10.1109/ICIP.2014.7025576.

Xi Peng, Jiashi Feng, Joey Tianyi Zhou, Yingjie Lei, and Shuicheng Yan. Deep subspace clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 31(12):5509–5521, 2020. doi: 10.1109/TNNLS.2020.2968848.

R. Souvenir and R. Pless. Manifold clustering. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 648–653 Vol. 1, 2005. doi: 10.1109/ICCV.2005.149.

Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290 5500:2319–23, 2000. URL `https://api.semanticscholar.org/CorpusID:221338160`.

Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 63(2):411–423, 01 2002. ISSN 1369-7412. doi: 10.1111/1467-9868.00293. URL `https://doi.org/10.1111/1467-9868.00293`.

M.E. Tipping and C.M. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, 1999. ISSN 0899-7667. doi: 10.1162/089976699300016728.

C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992. ISSN 0920-5691. doi: 10.1007/BF00129684.

Nicolas Garcia Trillos, Pengfei He, and Chenghui Li. Large sample spectral analysis of graph-based multi-manifold clustering, 2021. URL `http://arxiv.org/abs/2107.13610`. arXiv:2107.13610 [cs, math, stat].

R. Vidal. Subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52–68, 2011. ISSN 1053-5888. doi: 10.1109/MSP.2010.939739.

René Vidal, Yi Ma, and S. Shankar Sastry. Principal Component Analysis. In René Vidal, Yi Ma, and S.S. Sastry, editors, *Generalized Principal Component Analysis*, Interdisciplinary Applied Mathematics, pages 25–62. Springer, New York, NY, 2016. ISBN 978-0-387-87811-9. doi: 10.1007/978-0-387-87811-9_2. URL `https://doi.org/10.1007/978-0-387-87811-9_2`.

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. PMLR, 2015. URL `https://proceedings.mlr.press/v38/wang15b.html`.

John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, and Yi Ma. Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31 (2):210–227, February 2009. ISSN 1939-3539. doi: 10.1109/TPAMI.2008.79. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

Lirong Wu, Zicheng Liu, Jun Xia, Zelin Zang, Siyuan Li, and Stan Z Li. Generalized clustering and multi-manifold learning with geometric structure preservation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 139–147, 2022a.

Lirong Wu, Lifan Yuan, Guojiang Zhao, Haitao Lin, and Stan Z Li. Deep clustering and visualization for end-to-end high-dimensional data analysis. *IEEE Transactions on Neural Networks and Learning Systems*, 2022b.

Xulun Ye and Jieyu Zhao. Multi-manifold clustering: A graph-constrained deep nonparametric method. 93:215–227. ISSN 0031-3203. doi: 10.1016/j.patcog.2019.04.029. URL `https://www.sciencedirect.com/science/article/pii/S003132031930175X`.

Zheng Zhang, Yong Xu, Jian Yang, Xuelong Li, and David Zhang. A Survey of Sparse Representation: Algorithms and Applications. *IEEE Access*, 3:490–530, 2015. ISSN 2169-3536. doi: 10.1109/ACCESS.2015.2430359. Conference Name: IEEE Access.

Pan Zhou, Yunqing Hou, and Jiashi Feng. Deep adversarial subspace clustering. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1596–1604, 2018. doi: 10.1109/CVPR.2018.00172.

Linlin Zong, Xianchao Zhang, Long Zhao, Hong Yu, and Qianli Zhao. Multi-view clustering via multi-manifold regularized non-negative matrix factorization. 88:74–89. ISSN 0893-6080. doi: 10.1016/j.neunet.2017.02.003. URL `https://www.sciencedirect.com/science/article/pii/S089360801730028X`.