

ROBUST TANGENT SPACE ESTIMATION VIA LAPLACIAN EIGENVECTOR GRADIENT ORTHOGONALIZATION

DHRUV KOHLI^{*†}, SAWYER J. ROBERTSON^{*‡}, GAL MISHNE[§], ALEXANDER CLONINGER^{‡,§}

ABSTRACT. Estimating the tangent spaces of a data manifold is a fundamental problem in data analysis. The standard approach, Local Principal Component Analysis (LPCA), struggles in high-noise settings due to a critical trade-off in choosing the neighborhood size. Selecting an optimal size requires prior knowledge of the geometric and noise characteristics of the data that are often unavailable. In this paper, we propose a spectral method, Laplacian Eigenvector Gradient Orthogonalization (LEGO), that utilizes the global structure of the data to guide local tangent space estimation. Instead of relying solely on local neighborhoods, LEGO estimates the tangent space at each data point by orthogonalizing the gradients of low-frequency eigenvectors of the graph Laplacian. We provide two theoretical justifications of our method. First, a differential geometric analysis on a tubular neighborhood of a manifold shows that gradients of the low-frequency Laplacian eigenfunctions of the tube align closely with the manifold's tangent bundle, while an eigenfunction with high gradient in directions orthogonal to the manifold lie deeper in the spectrum. Second, a random matrix theoretic analysis also demonstrates that low-frequency eigenvectors are robust to sub-Gaussian noise. Through comprehensive experiments, we demonstrate that LEGO yields tangent space estimates that are significantly more robust to noise than those from LPCA, resulting in marked improvements in downstream tasks such as manifold learning, boundary detection, and local intrinsic dimension estimation.

Keywords. Tangent space estimation | graph Laplacian | tubular neighborhood | manifold learning | dimensionality reduction | dimensionality estimation.

1. INTRODUCTION

Tangent space estimation is a fundamental geometric task with broad applications across numerous domains, including manifold learning [1, 2, 3, 4, 5, 6, 7], data denoising [8], multi-manifold structure learning [9, 10, 11, 12], local intrinsic dimension estimation [13], connection Laplacian approximation [14, 15, 16], and regression on manifolds [17], among others. The standard procedure for estimating the tangent space at a given point x of a data manifold comprises of two steps: (i) determining the local intrinsic dimension d if not known *a priori*, and (ii) identifying d orthogonal directions in the ambient space that estimate a basis for the tangent space at x .

The most commonly adopted approach for tangent space estimation is Local Principal Component Analysis (LPCA) [1, 2, 3, 5, 8, 15, 17], which constructs the local covariance matrix using the k_{nn} -nearest neighbors of a data point and extracts the leading d eigenvectors as an estimate of the tangent basis at that point. When the local intrinsic dimension d is unknown, it is often inferred by counting the smallest number of top eigenvalues whose normalized cumulative sum (explained variance ratio) exceeds a user-defined threshold.

Due to its local formulation, LPCA offers a straightforward and computationally efficient approach for tangent space estimation. However, a limitation of this local nature is that the presence of noise can significantly degrade the quality of the estimated tangent spaces as demonstrated in Figure 1a and 1b. Specifically, there is a well-known trade-off in the choice of neighborhood size: small neighborhoods are prone to noise corruption, while larger neighborhoods introduce bias due to the underlying curvature and reach of the manifold [18, 19, 20]. One potential strategy to address this limitation involves selecting

^{*}equal contribution

[†]Program in Applied and Computational Mathematics, Princeton University (dhkohli@ucsd.edu)

[‡]Department of Mathematics, UC San Diego (s5robert@ucsd.edu, acloninger@ucsd.edu)

[§]Halicioğlu Data Science Institute, UC San Diego (gmishne@ucsd.edu)

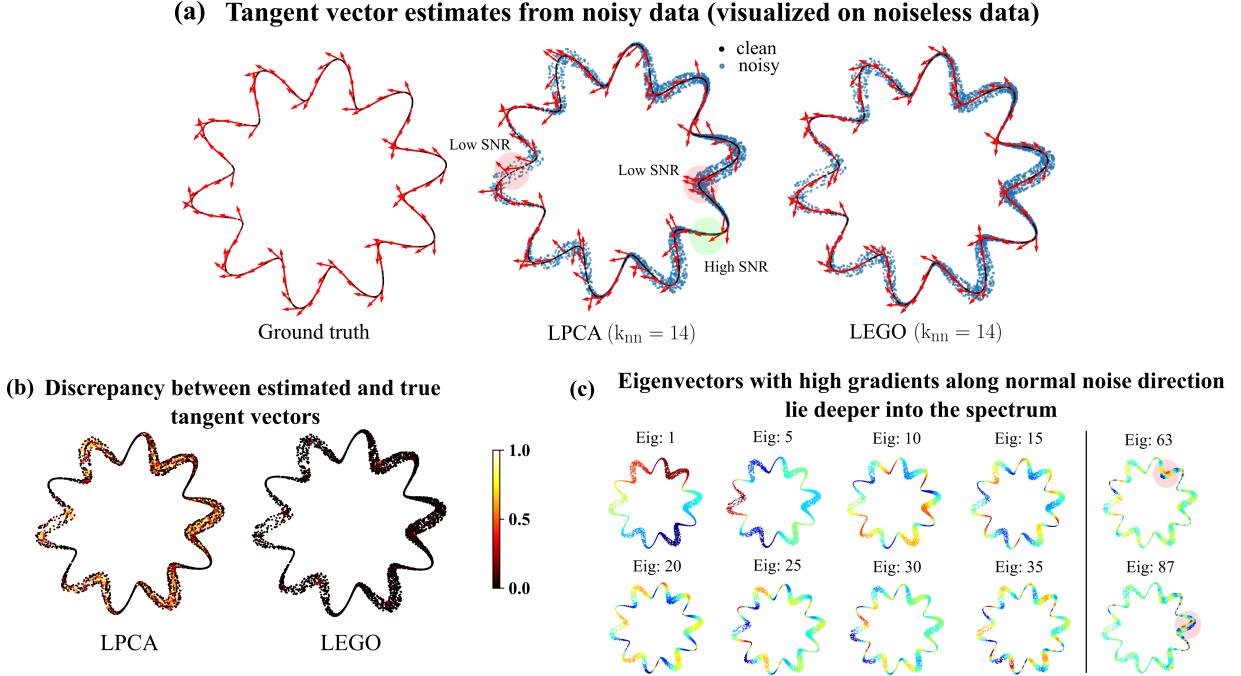


FIGURE 1. Illustration of tangent space estimation using LPCA and LEGO on a noisy point cloud generated by non-uniform sampling of a closed curve—wave on a circle—with heteroskedastic noise added in the normal direction. (a) Clean data points with ground truth tangent vectors, along with tangent vectors estimated from the noisy data using LPCA ($k_{nn} = 14$ and $d = 1$) and LEGO ($k_{nn} = 14$, $m_0 = 20$, $m = 100$ and $d = 1$). (b) Cosine dissimilarity between the true and the estimated tangent vectors. (c) Eigenvectors of the graph Laplacian constructed from noisy data [22], highlighting that those exhibiting high gradient in the noise direction lie deeper into the spectrum.

an adaptive neighborhood size [19] that balances these competing effects. Nonetheless, the practical implementation of such adaptive schemes is hindered by the fact that the geometric quantities—curvature, reach, and the noise level, are typically unknown. As a result, selecting an appropriate neighborhood size becomes a challenging and often ill-posed problem.

In contrast, taking cues from the global structure of the data may offer an alternative route to robust tangent space estimation, avoiding the complexities of adaptive neighborhood sizes while allowing them to remain small. This perspective is widely adopted in literature [21, 22, 23, 24] where the eigenmodes of the graph Laplacian are frequently used to encode the global geometry of data. This naturally raises the question of whether such global eigenmodes can also be leveraged to inform local geometric structure, and in particular, to improve tangent space estimation.

In the continuous setting, such a connection is well established by Jones et al. in [25], showing that under suitable regularity assumptions, for a given point x on a d -dimensional Riemannian manifold [26], there exist d eigenfunctions of the manifold Laplacian which yield a bilipschitz local parameterization of a sufficiently small neighborhood of x into \mathbb{R}^d . Building on this, in [4], we introduced an algorithm, Low Distortion Local Eigenmaps (LDLE), which realizes their result in practice by constructing local parameterizations using global eigenvectors of the graph Laplacian. Unlike traditional approaches that rely on the first d non-trivial eigenvectors [21, 22], LDLE selects customized subsets of d -eigenvectors for each neighborhood to construct their parameterizations into \mathbb{R}^d . These parameterizations typically have low distortion, ensuring their Jacobians are full rank and span the d -dimensional tangent spaces.

This provides empirical and theoretical support for using gradients of graph Laplacian eigenvectors to estimate local tangent spaces on data manifolds. However, in the presence of noise, these eigenvectors may still exhibit non-zero gradients in directions orthogonal to the manifold, causing them to acquire components in the noise directions [27, 28, 29, 30, 31] and consequently distorting the tangent space estimates.

Fortunately, a principle analogous to classical Fourier analysis applies: just as the low-frequency Fourier modes capture the underlying signal while high-frequency modes tend to encode noise [32, 33, 34], it is commonly observed that the eigenvectors corresponding to small eigenvalues of the graph Laplacian are robust to noise while the ones lying deeper into the spectrum may have nontrivial gradient in the noise directions [27, 28, 29, 30, 31], as demonstrated in Figure 1c. Building upon this insight, in this work, we propose an algorithm that estimates the tangent spaces at data points using the gradients of *low-frequency* global eigenvectors of the graph Laplacian. Moreover, we provide differential geometric and random matrix theoretic arguments to support our approach.

Our contributions are as follows. We present a spectral algorithm, LEGO (Laplacian Eigenvector Gradient Orthogonalization), for estimating tangent spaces at each data point by orthogonalizing the gradients of low-frequency *global* eigenvectors of the graph Laplacian derived from a noisy point cloud. Through comprehensive experiments we show that LEGO yields tangent space estimates that are significantly more robust to noise than those obtained via LPCA. We also demonstrate that this increased robustness results in significant improvements across multiple downstream tasks, including manifold learning [1, 2, 35, 3, 5, 36], boundary detection [37, 38, 39], and local intrinsic dimension estimation [13].

On the theoretical front, we offer two complementary justifications for our approach. First, we provide a differential geometric argument where we adopt a noise model in which clean data points lie on a d -dimensional smooth submanifold B in \mathbb{R}^{d+k} , while the noisy observations lie in a tubular neighborhood $\mathcal{T}^{\varepsilon r}$ around B , where r is any number bounded by the global reach of B [40, 41] and $\varepsilon \in (0, 1)$ controls the tube width. In practice, r represents the maximum noise and ε is a parameter that controls the noise level. In particular, we assume that noise perturbs data in directions normal to B . In this setting, the “horizontal space” at a noisy point—the subspace orthogonal to the noise direction—approximates the true tangent space of the corresponding clean point. Thus, estimating the horizontal space provides a principled approach to robust tangent space estimation. To formalize this approach, we study the eigenfunctions of the Laplacian on the tubular neighborhood $\mathcal{T}^{\varepsilon r}$ of B .

Specifically, we establish upper and lower bounds on the eigenvalue λ associated with an eigenfunction ϕ in terms of its horizontal and vertical energies, $\mathcal{E}_B(\phi)$ and $\mathcal{E}_B^\perp(\phi)$, which quantify the net gradient of ϕ across the horizontal spaces and the noise directions, respectively. These bounds indicate that for small ε , λ approximately scales as $\Omega(\varepsilon^{-2}\mathcal{E}_B^\perp(\phi))$ with respect to the vertical energy of ϕ , and as $\mathcal{O}(\mathcal{E}_B(\phi))$ with respect to the horizontal energy. As a result, eigenfunctions with significant gradient across tube cross-sections necessarily correspond to large eigenvalues. Conversely, gradients of eigenfunctions associated with small eigenvalues are largely confined to the horizontal spaces. The practical implication is that the gradients of the low-frequency eigenvectors of the graph Laplacian, constructed from noisy data, exhibit relatively small components along noise directions, and thus their orthogonalization may provide a robust estimate of the tangent spaces.

Second, we provide a random matrix theoretic argument where we adopt an information plus noise-type model in which the clean data is injected with sub-Gaussian noise with variance proxy ε . Our data model is reminiscent of that presented in [42], with a key distinction being that our data dimension is held constant while $\varepsilon = o(1)$ in the limit as the number of data points tends to infinity. This model generalizes from the case where the clean data lies on an embedded submanifold and the noise is confined to directions orthogonal to the tangent spaces. Using a Gaussian kernel with fixed bandwidth s , we construct the adjacency matrices and derive the corresponding random walk graph Laplacians, $\bar{\mathcal{L}}$ for the clean data and \mathcal{L} for the noisy data [43, 44].

By adapting techniques from the random graph literature (see, e.g., [45]), assuming that the variance proxy ε scales as $\mathcal{O}(1/\sqrt{n \log n})$, we prove that the noisy Laplacian \mathcal{L} converges to its clean counterpart $\bar{\mathcal{L}}$ in operator norm at a rate of $n^{-1/2}$ i.e., $\|\mathcal{L} - \bar{\mathcal{L}}\|_2 = \mathcal{O}(n^{-1/2})$ with high probability. By the Davis-Kahan theorem [46], it follows that the eigenvectors of \mathcal{L} remain close to those of $\bar{\mathcal{L}}$, provided the eigengaps of

the clean Laplacian do not decay too rapidly with n . In the submanifold setting, classical spectral results such as Weyl's law (e.g., see [47]) imply that eigengaps tend to shrink deeper in the spectrum. This, when combined with our result, suggest that low-frequency eigenvectors are more stable to noise and therefore their gradients can be reliably used for tangent space estimation.

The remainder of the paper is organized as follows. In Section 2, we introduce our proposed algorithm, LEGO, for tangent space estimation. Theoretical justifications for our method are provided in two parts: a differential geometric perspective in Section 3, and a random matrix theoretic analysis in Section 4. In Section 5, we demonstrate the effectiveness of LEGO through experiments on multiple datasets, highlighting its improved accuracy over LPCA and its benefits for downstream tasks.

2. TANGENT SPACE ESTIMATION VIA GRADIENTS OF LOW-FREQUENCY GLOBAL EIGENVECTORS OF GRAPH LAPLACIAN

Here, we introduce our algorithm, LEGO, for estimating tangent spaces at noise-perturbed data points that are sampled from a tubular neighborhood of a smooth embedded submanifold. Specifically, we assume that the clean data points lie on the submanifold while the noise is constrained to the subspaces orthogonal to their tangent spaces. Our approach estimates orthonormal bases of the tangent spaces at the clean points by orthogonalizing the gradients of low-frequency global eigenvectors of the graph Laplacian constructed from the noisy data.

Let $Y = [Y_1, \dots, Y_n] \in \mathbb{R}^{p \times n}$ be a point cloud sampled from a smooth compact d -dimensional submanifold B embedded in \mathbb{R}^p . Let $X = [X_1, \dots, X_n] \in \mathbb{R}^{p \times n}$ be the noisy point cloud such that X_i is obtained by adding noise to Y_i in the directions orthogonal to the tangent space $T_{Y_i}B$. Let $\mathcal{N}_j = \{j_1, \dots, j_{k_{\text{nn}}}\}$ be a set containing the indices of the k_{nn} -nearest neighbors of X_j obtained using the Euclidean metric in \mathbb{R}^p . Let $\mathcal{L} \in \mathbb{R}^{n \times n}$ be the graph Laplacian constructed from X using one of the following kernel-based methods: the random walk kernel [22, 43], the self-tuned kernel [48, 49] or the doubly stochastic kernel [50, 51]. These construction strategies ensure that, under appropriate scaling of the kernel bandwidth and sampling density, the discrete operator \mathcal{L} converges with high probability to the Laplace–Beltrami operator Δ_{δ_p} on a tubular neighborhood of the submanifold B [52, 43, 53, 54, 49, 55]. Moreover, recent results [49, 54] establish the convergence of the spectrum of \mathcal{L} to that of Δ_{δ_p} under technical conditions on the sampling density, manifold geometry and the kernel bandwidth.

Let $\phi_i \in \mathbb{R}^n$ be the i -th eigenvector of \mathcal{L} corresponding to the i -th smallest eigenvalue, and $\nabla\phi_i \in \mathbb{R}^{p \times n}$ be a matrix whose j -th column, $\nabla\phi_i(X_j) \in \mathbb{R}^p$, represents the gradient of ϕ_i at X_j . Each of the p components of the gradient $\nabla\phi_i$ is treated a smooth function on X , and thus modeled as a vector in the span of the eigenvectors of \mathcal{L} . Given that eigenvectors corresponding to higher eigenvalues are more susceptible to noise (see, e.g., [24, 31]), we estimate $\nabla\phi_i$ using only the first $m_0 \ll n$ eigenvectors $\{\phi_1, \dots, \phi_{m_0}\}$ of \mathcal{L} . To ensure local fidelity, we require the estimated gradient to approximate ϕ_i , up to first order, on the neighborhood $\{X_{j_s} : j_s \in \mathcal{N}_j\}$ of each point X_j . Precisely, define centered data points and eigenvectors as,

$$(1) \quad \bar{X}_j = \begin{bmatrix} X_{j_1}^T - X_j^T \\ \vdots \\ X_{j_{k_{\text{nn}}}}^T - X_j^T \end{bmatrix} \quad \text{and} \quad \bar{\phi}_i(X_j) = \begin{bmatrix} \phi_i(X_{j_1}) - \phi_i(X_j) \\ \vdots \\ \phi_i(X_{j_{k_{\text{nn}}}}) - \phi_i(X_j) \end{bmatrix},$$

respectively, where $\{X_{j_s}\}_{j_s \in \mathcal{N}_j}$ are the k_{nn} -nearest neighbors of X_j . Then, the estimate $\hat{\nabla}\phi_i \in \mathbb{R}^{p \times n}$ of the gradients $\nabla\phi_i$ is given by,

$$(2) \quad \hat{\nabla}\phi_i = \hat{C}_i U_\Phi^T$$

where $U_\Phi \in \mathbb{R}^{n \times m_0}$ constitutes an orthonormal basis of the range of $\Phi = [\phi_1 \ \phi_2 \ \dots \ \phi_{m_0}] \in \mathbb{R}^{n \times m_0}$ and \hat{C}_i is the solution of the following optimization problem,

$$(3) \quad \hat{C}_i = \underset{C_i \in \mathbb{R}^{p \times m_0}}{\operatorname{argmin}} \frac{1}{n} \sum_{j=1}^n \left\| \bar{X}_j \hat{\nabla}\phi_i(X_j) - \bar{\phi}_i(X_j) \right\|_2^2$$

$$(4) \quad \text{such that } \hat{\nabla}\phi_i = C_i U_\Phi^T.$$

Using the fact that U_Φ has orthonormal columns i.e. $U_\Phi^T U_\Phi = I_{m_0}$, the least squares solution of the above optimization problem is given by,

$$(5) \quad \hat{C}_i = \left[\bar{X}_1^\dagger \bar{\phi}_i(X_1) \quad \dots \quad \bar{X}_n^\dagger \bar{\phi}_i(X_n) \right] U_\Phi$$

where \bar{X}_i^\dagger is the pseudoinverse of \bar{X}_i . Having obtained the gradient estimates of the eigenvectors $\{\phi_1, \dots, \phi_m\}$ at X_j given by,

$$(6) \quad \hat{\nabla}\phi(X_j) = \left[\hat{\nabla}\phi_1(X_j) \quad \dots \quad \hat{\nabla}\phi_m(X_j) \right] \in \mathbb{R}^{p \times m},$$

we obtain an estimate of the orthonormal basis $Q_j \in \mathbb{R}^{p \times d}$ of the d -dimensional tangent space at the j th point by orthogonalizing $\hat{\nabla}\phi(X_j)$ and equivalently, using the top d left singular vectors of the matrix $\hat{\nabla}\phi(X_j)$. If the intrinsic dimension d is not known a priori then one can estimate it by selecting the smallest number of top eigenvalues of the matrix whose normalized cumulative sum exceeds a user-defined threshold f_{var} , as demonstrated in the pseudocode below.

Algorithm 1 Tangent space estimation via LEGO.

Require: $X \in \mathbb{R}^{n \times p}$, \mathcal{L} , k_{nn} , m and m_0 where $m \leq m_0$, either d or $f_{\text{var}} \in (0, 1)$.

```

1:  $\phi_1, \dots, \phi_{m_0} \leftarrow$  eigenvectors of  $\mathcal{L}$  corresponding to  $m_0$  smallest eigenvalues
2: Estimate  $\hat{\nabla}\phi_i$  using Eq. 2 and 5 for  $i \in [1, m]$ 
3: for  $j \in [1, n]$  do
4:   Set  $\hat{\nabla}\phi(X_j)$  as in Eq. 6.
5:    $U \in \mathbb{R}^{p \times p}$ ,  $\sigma_1 \geq \dots \geq \sigma_p \leftarrow$  Left singular vectors and values of  $\hat{\nabla}\phi(X_j)$ 
6:   if  $d$  is provided then
7:      $Q_j \leftarrow$  first  $d$  columns of  $U$ 
8:   else
9:      $d_j \leftarrow \min\{s \in [1, p] : \sum_{i=1}^s \sigma_i^2 / \sum_{i=1}^p \sigma_i^2 \geq f_{\text{var}}\}$ 
10:     $Q_j \leftarrow$  first  $d_j$  columns of  $U$ 
return  $\{Q_j\}_1^n$ 

```

2.1. Time complexity. Assuming that the local intrinsic dimension d is known a priori, the cost of estimating tangent space at each point i.e., computing the top d principal directions from the local neighborhood using LPCA is $\mathcal{O}(k_{\text{nn}}pd)$. Therefore, the total time complexity of applying LPCA to all n points is $\mathcal{O}(nk_{\text{nn}}pd)$. In contrast, the time complexity of each stage of LEGO is as follows: (i) computing eigenvectors $\phi_1, \dots, \phi_{m_0}$ of the graph Laplacian \mathcal{L} (Line 1 in Algorithm 1) using an iterative eigensolver requires $\mathcal{O}(nk_{\text{nn}}m_0T)$ time where T is the number of iterations required for convergence [56]. (ii) Computing orthonormal basis U_Φ of the eigenvectors Φ requires $\mathcal{O}(nm_0^2)$ time. (iii) Estimating the gradients $\hat{\nabla}\phi_i$ for all $i \in [1, m]$ (Line 2 in Algorithm 1) requires $\mathcal{O}(nk_{\text{nn}}p(\min\{k_{\text{nn}}, p\} + m)) + \mathcal{O}(nm_0mp)$ where the first term exclusively corresponds to the estimation of \hat{C}_i in Eq. 5 and the second term corresponds to multiplication by Φ on the right side in Eq. 5 and by Φ^T in Eq. 2. (iv) Computing the orthonormal basis Q_j for all $j \in [1, n]$ using the gradients of eigenvectors takes $\mathcal{O}(nmpd)$. Overall, the total time complexity of LPCA is linear in all the parameters while LEGO is quadratic in $\min\{k_{\text{nn}}, p\}$. In practice, the computation of eigenvectors $\phi_1, \dots, \phi_{m_0}$ seem to dominate the computational cost of LEGO.

2.2. Selection of hyperparameters for LEGO. A typical choice for k_{nn} is on the order $\mathcal{O}(\log(n))$ or $\mathcal{O}(n^\alpha)$ where $\alpha \in (1/d, 1)$. In practice, k_{nn} is selected to be sufficiently small to avoid spurious edges or “shortcuts” in the nearest neighbor graph that could distort the underlying local geometry. Since the appropriate value of k_{nn} is usually unknown, a common approach is to begin with a small value and gradually increase it until the output reaches the desired quality.

As we argue in Section 3, the number of eigenvectors m whose gradients are orthogonalized should be kept relatively small, so as to avoid eigenvectors deeper in the spectrum that may exhibit large gradients in the noise directions [27, 28, 29]. We demonstrate in our experiments that this does not require fine-tuning

as the tangent space estimates remain stable across a wide range of values of m . Lastly, m_0 , the number of eigenvectors used to model the gradients must be sufficiently large to ensure that each component of the gradient—viewed as a smooth function on the manifold—is well-approximated within the span of the first m_0 eigenvectors [57, 44].

3. EIGENFUNCTIONS WITH HIGH GRADIENT ALONG THE CROSS SECTIONS OF A TUBE LIE DEEPER INTO THE SPECTRUM

Noisy data is often modeled as a sample drawn from a tubular neighborhood surrounding an underlying smooth submanifold [41, 58, 59, 60, 61, 62]. Under this noise model, the graph Laplacian constructed from such data [22, 49, 55] converges to the continuous Laplacian of the tubular neighborhood. This motivates the study of the eigenfunctions of the Laplacian on the tube to better understand the behavior of the graph Laplacian eigenvectors derived from noisy data. Here, building on [63], we show that eigenfunctions exhibiting high gradient across the cross sections of the tubular neighborhood necessarily correspond to higher eigenvalues. Consequently, eigenfunctions associated with low eigenvalues exhibit minimal gradient in directions normal to the submanifold. The practical implication of our result is that the gradients of the low-frequency eigenvectors of the graph Laplacian tend to have small components in the noise directions, making them suitable for tangent space estimation. The proofs of our results are provided in the appendix.

3.1. Preliminaries. In the following we describe the necessary constructs from [63] that are needed for our results. Let $B \subset \mathbb{R}^{d+k}$ be a smooth embedded compact d -dimensional submanifold with or without boundary, equipped with the metric g_B induced by the Euclidean metric δ_{d+k} . Let NB be the normal bundle of B equipped with the metric $g_B^\perp = \delta_{d+k}|_{NB}$. Assume that there exist a tubular neighborhood \mathcal{T}^r of B such that $B \subset \mathcal{T}^r \subset \mathbb{R}^{d+k}$, where r is any finite number bounded by the global reach, meaning, it satisfies the property that that normals to B with length less than r do not intersect [41]. Define a map

$$\Psi : NB \rightarrow \mathbb{R}^{d+k}, (x, \nu) = x + \nu$$

which, when restricted to

$$NB^r = \{(x, \nu) \in NB : \|\nu\|_{\delta_{d+k}} < r\} \subset NB,$$

is diffeomorphic to its image \mathcal{T}^r . Let $\pi : NB^r \rightarrow B$ be the canonical projection $\pi(x, \nu) = x$ onto B . By equipping NB^r with the pullback metric $g = \Psi^* \delta_{d+k}$, the tubular neighborhood \mathcal{T}^r is isometric to NB^r . This also holds for ε -tubular neighborhood $\mathcal{T}^{\varepsilon r}$ of B and the normal bundle $NB^{\varepsilon r}$ for $\varepsilon < 1$. To keep the dependence on ε explicit, it is convenient to work with NB^r with the pullback metric $g^\varepsilon = \mathcal{D}_\varepsilon^* g$ where the map $\mathcal{D}_\varepsilon : NB^r \rightarrow NB^{\varepsilon r}$ is given by $\mathcal{D}_\varepsilon(x, \nu) = (x, \varepsilon\nu)$. In fact, NB^r equipped with g^ε is isometric to $\mathcal{T}^{\varepsilon r}$ equipped with Euclidean metric δ_{d+k} . Due to this construction, the Laplacian $-\Delta_{g^\varepsilon}$ on NB^r is unitarily equivalent to the Euclidean Laplacian $\Delta_{\delta_{d+k}}$ on $\mathcal{T}^{\varepsilon r}$ i.e. for functions $\phi \in C_0^\infty(NB^r)$ it holds that

$$\Delta_{g^\varepsilon} \phi = \widehat{\mathcal{D}}_\varepsilon^{-1} \widehat{\Psi} \Delta_{\delta_{d+k}} \widehat{\Psi}^{-1} \widehat{\mathcal{D}}_\varepsilon \phi$$

where $\widehat{\mathcal{D}}_\varepsilon$ and $\widehat{\Psi}$ are the unitary lifts associated with \mathcal{D}_ε and Ψ , respectively. Specifically, $\widehat{\mathcal{D}}_\varepsilon : L^2(NB^r, dV_{g^\varepsilon}) \rightarrow L^2(NB^{\varepsilon r}, dV_g)$ and $\widehat{\Psi} : L^2(\mathcal{T}^{\varepsilon r}, dV_{\delta_{d+k}}) \rightarrow L^2(NB^{\varepsilon r}, dV_g)$ are given by,

$$(\widehat{\mathcal{D}}_\varepsilon \phi)(x, \nu) = \phi(x, \nu/\varepsilon)$$

and

$$\widehat{\Psi} \phi = \phi \circ \Psi.$$

It follows that if ϕ is an eigenfunction of the Laplacian $-\Delta_{\delta_{d+k}}$ on $\mathcal{T}^{\varepsilon r}$ with eigenvalue λ then $\widehat{\mathcal{D}}_\varepsilon^{-1} \widehat{\Psi} \phi$ is an eigenfunction of the Laplacian Δ_{g^ε} on NB^r with the same eigenvalue.

In addition to the metrics introduced above, the metric

$$g_s^\varepsilon = \pi^* g_B + \varepsilon^2 g_B^\perp$$

turns π^* into a Riemannian submersion i.e. an isometry from the horizontal subbundle $\pi^*(TB)$ to the tangent bundle TB . Let $g_s = g_s^{\varepsilon=1}$ be the ε -independent unscaled version of the submersion metric.

Using the unscaled submersion metric and a local coordinate system we define the horizontal energy of an eigenfunction ϕ on the tube $\mathcal{T}^{\varepsilon r}$ that captures the net gradient of ϕ along the submanifold B ,

and the vertical energy of ϕ that measures its net gradient normal to B i.e. across the cross sections of the tubular neighborhood \mathcal{T}^r . To this end, let x^1, \dots, x^d be the local coordinates on B and $\{e_\alpha\}_1^k$ be a locally orthonormal frame of NB^r with respect to g_B^\perp such that every normal vector $\nu(x) \in N_x B$ can be written as $\nu(x) = n^\alpha e_\alpha(x)$. Consequently, $(x^1, \dots, x^d, n^1, \dots, n^k)$ form local coordinates of NB^r and the associated local coordinate vector fields are,

$$(7) \quad \partial_i|_{(x,n)} = \partial_{x^i}, \quad \partial_{d+\alpha}|_{(x,n)} = \partial_{n^\alpha}, \quad i \in [1, d], \alpha \in [1, k].$$

For $f \in C_0^\infty(NB^r)$, define its canonical gradients as

$$(8) \quad \nabla_x f = [\partial_{x^1} f, \dots, \partial_{x^d} f]^T \in \mathbb{R}^d$$

$$(9) \quad \nabla_n f = [\partial_{n^1} f, \dots, \partial_{n^k} f]^T \in \mathbb{R}^k$$

and $\nabla f \in \mathbb{R}^{d+k}$ is the concatenation of the two vectors. Then, for $\phi \in C_0^\infty(\mathcal{T}^{\varepsilon r})$, the horizontal energy $E_B(\phi)$ and the vertical energy $E_B^\perp(\phi)$ of ϕ are given by (here $\widehat{\phi} = \widehat{\mathcal{D}}_\varepsilon^{-1}\widehat{\Psi}\phi$ is the unitary lift of ϕ onto NB^r)

$$(10) \quad E_B(\phi) = \frac{\int_{NB^r} \nabla_x \widehat{\phi}^T g_B^{-1} \nabla_x \widehat{\phi} dV_{g_s}}{\int_{NB^r} \widehat{\phi}^2 dV_{g_s}}$$

$$(11) \quad E_B^\perp(\phi) = \frac{\int_{NB^r} \nabla_n \widehat{\phi}^T \nabla_n \widehat{\phi} dV_{g_s}}{\int_{NB^r} \widehat{\phi}^2 dV_{g_s}}.$$

We further define a normalized version of the horizontal and vertical energy of $\phi \in C_0^\infty(\mathcal{T}^{\varepsilon r})$ as

$$(12) \quad \mathcal{E}_B(\phi) = \frac{1}{\lambda_{B_2}} E_B(\phi) \quad \text{and} \quad \mathcal{E}_B^\perp(\phi) = \frac{r^2}{C_k} E_B^\perp(\phi),$$

respectively, where λ_{B_2} and C_k/r^2 are the first non-zero eigenvalues of the Laplacian $-\Delta_{g_B}$ on B and $-\Delta_{\delta_k}$ on a ball of radius r in \mathbb{R}^k , respectively. Here, C_k is a constant that depends on the dimension k and the choice of the boundary conditions—either Neumann or Dirichlet. Overall, the above normalizations ensure that $\mathcal{E}_B(\phi)$ and $\mathcal{E}_B^\perp(\phi)$ are on similar scale with respect to r and ε .

To further motivate the above definitions, consider the example where $B = [0, l] \times \{0\} \subset \mathbb{R}^2$ is an interval of length l , and its tubular neighborhood of radius r is given by a rectangular domain $\mathcal{T}^r = NB^r = [0, l] \times [-r, r]$. The Neumann eigenfunctions of the Laplacian $\Delta_{\delta_2} = \Delta_{g^\varepsilon} = \Delta_{g_s^\varepsilon}$ on the scaled tube $\mathcal{T}^{\varepsilon r} = NB^{\varepsilon r}$ are of the form $\phi_{i,j}(x, n) = \cos(i\pi x/l) \sin(j\pi n/2\varepsilon r)$ with corresponding eigenvalues $\lambda_{i,j} = (i\pi/l)^2 + (j\pi/2\varepsilon r)^2$ where $i, j \in \mathbb{Z}_{\geq 0}$. Lifting $\phi_{i,j}$ back to the unscaled tube NB^r gives $\widehat{\phi}_{i,j}(x, n) = \cos(i\pi x/l) \sin(j\pi n/2r)$ from which we compute the unnormalized horizontal energy as $E_B(\phi_{i,j}) = (i\pi/l)^2$ and the vertical energy as $E_B^\perp(\phi_{i,j}) = (j\pi/2r)^2$. Consequently, the normalized horizontal and vertical energies are given by $\mathcal{E}_B(\phi_{i,j}) = i^2$ and $\mathcal{E}_B^\perp(\phi_{i,j}) = j^2$, respectively. In particular, the eigenvalue can be expressed as $\lambda_{i,j} = \lambda_{B_2} \mathcal{E}_B(\phi_{i,j}) + C_k(\varepsilon r)^{-2} \mathcal{E}_B^\perp(\phi_{i,j})$ highlighting that, if r is small enough such that $C_k r^{-2} \geq \lambda_{B_2}$ then, the contribution of the vertical energy to $\lambda_{i,j}$ scales as ε^{-2} relative to the horizontal energy. This means that a unit increase in the vertical energy of an eigenfunction results in a much larger increase in the eigenvalue for small ε .

We end this subsection by defining a few constructs that capture the geometry of B and are utilized in our results. Let Π be the second fundamental form on B then the coefficients of the scalar second fundamental form on B are given by,

$$(13) \quad h_{\alpha i}^j = g_B^\perp(e_\alpha, \Pi(\partial_{x^i}, \partial_{x^j})) = h_{\alpha j}^i.$$

Let ∇^\perp be the normal connection with respect to $\{e_\alpha\}_1^k$ then the Christoffel symbols of the normal connection are given by,

$$(14) \quad \gamma_{i\alpha}^\beta = g_B^\perp(\nabla_{\partial_{x^i}}^\perp e_\alpha, e_\beta).$$

Define a symmetric matrix $H_\alpha(x) \in \mathbb{R}^{d \times d}$ such that for $i, j \in [1, d]$,

$$(15) \quad (H_\alpha)_{i,j} = h_{\alpha i}^j.$$

Also define a rectangular matrix $\Gamma_\beta(x) \in \mathbb{R}^{d \times k}$ such that for $i, j \in [1, d]$ and $\alpha \in [1, k]$,

$$(16) \quad (\Gamma_\beta)_{j,\alpha} = \gamma_{j\alpha}^\beta$$

3.2. Our results. First, we note that the span of $\{\partial_i|_{(x,n)}\}_1^d$ is not necessarily orthogonal to $\{\partial_{d+\alpha}|_{(x,n)}\}_1^k$. To aid the analysis, a new basis $\{\partial_i^H|_{(x,n)}\}_1^d$ is obtained by projecting each $\partial_i|_{(x,n)}$ orthogonal to the span of $\{\partial_{d+\alpha}|_{(x,n)}\}_1^k$ via Gram-Schmidt orthogonalization. Through basic manipulations we then obtain the following lemma.

Lemma 1. *The pullback metric $g^\varepsilon = \mathcal{D}_\varepsilon^* g = \mathcal{D}_\varepsilon^* \Psi^* \delta_{d+k}$ with respect to the coordinate vector fields $\{\partial_1^H|_{(x,n)}, \dots, \partial_d^H|_{(x,n)}, \partial_{d+1}|_{(x,n)}, \dots, \partial_{d+k}|_{(x,n)}\}$ on NB^r , is given by*

$$(17) \quad g^\varepsilon(x, n) = \left[g_B^{1/2} (I_d - \varepsilon n^\alpha g_B^{-1/2} H_\alpha g_B^{-1/2})^2 g_B^{1/2} \right]_{\varepsilon^2 I_k}.$$

Consequently, the Riemannian gradient of $\hat{\phi} \in C_0^\infty(NB^r)$ with respect to g^ε is given by,

$$(18) \quad \text{grad}\hat{\phi}(x, n) = \left[g_B^{-1/2} (I_d - \varepsilon n^\alpha g_B^{-1/2} H_\alpha g_B^{-1/2})^{-2} g_B^{-1/2} \left(\nabla_x \hat{\phi}(x, n) - n^\beta \Gamma_\beta \nabla_n \hat{\phi}(x, n) \right) \right]_{\varepsilon^{-2} \nabla_n \hat{\phi}(x, n)}.$$

Note that g^ε is guaranteed to be positive semidefinite. However, for large ε it can become singular for certain values of (n^1, \dots, n^k) . The following lemma provides a sufficient and necessary condition on ε that ensures the positivity of g^ε throughout NB^r .

Lemma 2. *Let $\kappa(x) \in \mathbb{R}_{\geq 0}$ and $\kappa^* \in \mathbb{R}_{\geq 0}$ be the absolute maximum principal curvature at $x \in B$ and across B , respectively, given by $\kappa^* = \max_{x \in B} \kappa(x)$ where*

$$\kappa(x) = \max_{\|v\|_2=1} \left(\sum_{\alpha=1}^k \left(v^T g_B(x)^{-1/2} H_\alpha(x) g_B(x)^{-1/2} v \right)^2 \right)^{1/2}.$$

Then, g^ε is positive definite on NB^r if and only if $\varepsilon r \kappa^* < 1$. Moreover,

$$(19) \quad \varepsilon^{-2k} \det(g_B) (1 - \varepsilon r \kappa^*)^{2d} \leq \det(g^\varepsilon) \leq \varepsilon^{-2k} \det(g_B) (1 + \varepsilon r \kappa^*)^{2d}.$$

By definition, we have $r \kappa^* < 1$ and $\varepsilon < 1$. Consequently, the condition $\varepsilon r \kappa^* < 1$ is satisfied in our setting, which ensures that g^ε is positive definite on NB^r .

Now we state our main result which shows that the eigenvalue λ corresponding to an eigenfunction ϕ scales as $\Omega\left(\frac{C_k}{(\varepsilon r)^2} \frac{(1 - \varepsilon r \kappa^*)^d}{(1 + \varepsilon r \kappa^*)^d}\right)$ with respect to a unit increase in $\mathcal{E}_B^\perp(\phi)$ versus $\mathcal{O}\left(\lambda_{B_2} \frac{(1 + \varepsilon r \kappa^*)^d}{(1 - \varepsilon r \kappa^*)^{d+2}}\right)$ with respect to a unit increase in $\mathcal{E}_B(\phi)$.

Theorem 3. *If ϕ is a Neumann or Dirichlet eigenfunction of the Laplacian $\Delta_{\delta_{d+k}}$ on $\mathcal{T}^{\varepsilon r}$ then the corresponding eigenvalue λ satisfies*

$$(20) \quad \lambda \geq \frac{C_k}{(\varepsilon r)^2} \frac{(1 - \varepsilon r \kappa^*)^d}{(1 + \varepsilon r \kappa^*)^d} \mathcal{E}_B^\perp(\phi),$$

$$(21) \quad \lambda \geq \frac{(1 - \varepsilon r \kappa^*)^d}{(1 + \varepsilon r \kappa^*)^d} \left(\frac{\lambda_{B_2} \mathcal{E}_B(\phi)}{(1 + \varepsilon r \kappa^*)^2} + \frac{C_k \mathcal{E}_B^\perp(\phi)}{(\varepsilon r)^2} - \frac{2\kappa^{\perp*} \sqrt{\lambda_{B_2} C_k \mathcal{E}_B(\phi) \mathcal{E}_B^\perp(\phi)}}{(1 + \varepsilon r \kappa^*)^2} \right)$$

and

$$(22) \quad \lambda \leq \frac{(1 + \varepsilon r \kappa^*)^d}{(1 - \varepsilon r \kappa^*)^d} \left(\frac{\lambda_{B_2} \mathcal{E}_B(\phi)}{(1 - \varepsilon r \kappa^*)^2} + \left(\left(\frac{\kappa^{\perp*}}{1 - \varepsilon r \kappa^*} \right)^2 + \frac{1}{(\varepsilon r)^2} \right) C_k \mathcal{E}_B^\perp(\phi) \right)$$

where $\kappa^\perp(x) \in \mathbb{R}_{\geq 0}^k$ quantifies the maximum rate of twisting of e_β , $\beta \in [1, k]$, in any direction in the tangent space $T_x B$, and $\kappa^{\perp*} \in \mathbb{R}_{\geq 0}$ quantifies maximum twisting among all normal directions. Specifically, $\kappa^{\perp*} = \max_{x \in B} \|\kappa^\perp(x)\|_2$ where

$$\kappa_\beta^\perp(x) = \left\| g_B(x)^{-1/2} \Gamma_\beta(x) \right\|_2, \quad \beta \in [1, k].$$

Due to the fact that $r\kappa^* < 1$, the above simplifies to,

Corollary 4. *If $r \leq \sqrt{C_k/\lambda_B}$ and $(1 - \varepsilon)^{d+1} \geq (1 + \varepsilon)^d \varepsilon^{1-t}$ for some $t \in (0, 1)$ then*

$$(23) \quad \lambda = \Omega(\varepsilon^{-2t} \mathcal{E}_B^\perp(\phi)) \text{ and } \lambda = \mathcal{O}(\mathcal{E}_B(\phi)).$$

Remark 5. *It easily follows that the above inequalities remain valid for a tubular neighborhood of B with varying width i.e. when ε depends on x , by replacing ε in the inequalities with $\max_{x \in B} \varepsilon(x)$.*

An immediate consequence of the above result is that when $\varepsilon \ll 1$ so that t is close to 1, any eigenvalue λ of $\Delta_{\delta_{d+k}}$ that is not too large—say of order $\mathcal{O}(\varepsilon^{-2\eta})$ for some $\eta \in (0, 1/2)$ —has a corresponding eigenfunction ϕ whose vertical energy $\mathcal{E}_B^\perp(\phi)$ is small, of order $\mathcal{O}(\varepsilon^{2t-2\eta})$. Consequently, the gradient of such an eigenfunction has a small component in the normal directions to B , making it a suitable candidate for tangent space estimation.

In the following results, we show that such eigenfunctions do exist if the Laplacian Δ_{g_B} on B has sufficiently small eigenvalues. First, using an eigenfunction of the Laplacian Δ_{g_B} on B with eigenvalue λ_B , we construct a function on $\mathcal{T}^{\varepsilon r}$ whose quadratic form is bounded by $\frac{(1+\varepsilon r\kappa^*)^d}{(1-\varepsilon r\kappa^*)^{d+2}} \lambda_B$.

Theorem 6. *Let ϕ_B be a Neumann or Dirichlet eigenfunction of the Laplacian Δ_{g_B} on B with the corresponding eigenvalue,*

$$\lambda_B = \frac{\int \langle \operatorname{grad}\phi_B, \operatorname{grad}\phi_B \rangle_{g_B} dV_{g_B}}{\int_B \phi_B^2 dV_{g_B}}.$$

Define an extension $\widehat{\phi} : NB^{\varepsilon r} \rightarrow \mathbb{R}$ of ϕ_B onto $NB^{\varepsilon r}$ which has constant value along the cross sections,

$$\widehat{\phi}(x, \nu) = \frac{\phi_B(x)}{\left(\int_{NB^r} (\phi_B \circ \pi)^2 dV_{g^r}\right)^{1/2}}.$$

Then, the Dirichlet energy of $\phi = \widehat{\Psi}^{-1} \widehat{\phi}$ defined on $\mathcal{T}^{\varepsilon r}$ satisfies,

$$(24) \quad \frac{(1 - \varepsilon r \kappa^*)^d \lambda_B}{(1 + \varepsilon r \kappa^*)^{d+2}} \leq - \int_{\mathcal{T}^{\varepsilon r}} \phi \Delta_{\delta_{d+k}} \phi dV_{\delta_{d+k}} \leq \frac{(1 + \varepsilon r \kappa^*)^d \lambda_B}{(1 - \varepsilon r \kappa^*)^{d+2}}.$$

By combining the above result with Sturm-Liouville decomposition, we conclude that there exist eigenfunctions of $\Delta_{\delta_{d+k}}$ on $\mathcal{T}^{\varepsilon r}$ whose eigenvalues are also bounded by $\frac{(1+\varepsilon r\kappa^*)^d}{(1-\varepsilon r\kappa^*)^{d+2}} \lambda_B$. Combining this with Theorem 3, we obtain the following corollary which shows that if there is an eigenvalue λ_B of order $\mathcal{O}\left(\frac{1}{(\varepsilon^{\eta} r)^2} \frac{(1-\varepsilon r\kappa^*)^{2d+2}}{(1+\varepsilon r\kappa^*)^{2d}}\right)$, $\eta \in (0, 1)$, then there exist an eigenfunction ϕ of $\Delta_{\delta_{d+k}}$ on $\mathcal{T}^{\varepsilon r}$ whose vertical energy $\mathcal{E}_B^\perp(\phi)$ is small and is of order $\mathcal{O}(\varepsilon^{2-2\eta})$.

Corollary 7. *Let λ_B be a Neumann (Dirichlet) eigenvalue of Δ_{g_B} on B . Then there exist a Neumann (Dirichlet) eigenfunction ϕ of $\Delta_{\delta_{d+k}}$ on $\mathcal{T}^{\varepsilon r}$ whose vertical energy satisfies,*

$$(25) \quad \mathcal{E}_B^\perp(\phi) \leq \varepsilon^2 \frac{\lambda_B r^2}{C_k} \frac{(1 + \varepsilon r \kappa^*)^{2d}}{(1 - \varepsilon r \kappa^*)^{2d+2}}.$$

4. ROBUSTNESS OF LAPLACIAN EIGENVECTORS UNDER NOISE

A small gradient component of the low-frequency eigenvectors of the Laplacian in the noise direction is a direct consequence of the noise stability of the eigenvectors as illustrated in Figure 2. In this section, we analyze the stability of the Laplacian eigenvectors from the angle of robustness of random kernel Laplacians to sub-Gaussian noise. In turn, by the Davis-Kahan theorem [46], if the low-frequency eigengaps of the random kernel Laplacians do not vanish too quickly, the corresponding Laplacian eigenvectors will be robust to noise perturbations as well. The proofs of our results are provided in the appendix.

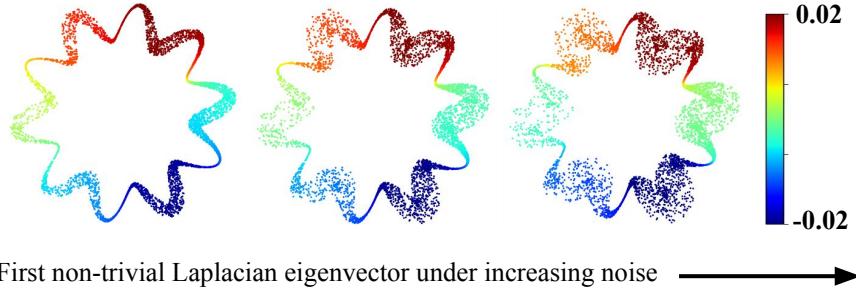


FIGURE 2. The first non-trivial eigenvector of the Laplacian \mathcal{L} is plotted against increasing noise level ε . Here, the noise is independent but non-identically distributed as in Figure 1a (also see Remark 9). The common colorbar represents the range of values, from minimum to maximum, across the eigenvectors.

As before, let $Y = \{Y_1, \dots, Y_n\}$ be the clean data points sampled from a bounded region in \mathbb{R}^p . Fixing a bandwidth $s > 0$, we define the clean or ground-truth kernel adjacency matrix $\bar{A} \in \mathbb{R}^{n \times n}$ entrywise by the formula

$$(26) \quad \bar{A}_{ij} = \sigma_s(Y_i - Y_j).$$

where $\sigma_s : \mathbb{R}^p \rightarrow (0, \infty)$ is the Gaussian kernel defined as,

$$(27) \quad \sigma_s(z) = e^{-\|z\|_2^2/s^2}.$$

Later, we will make use of the following estimate on the Lipschitz constant of σ_s .

Lemma 8. *For any $s > 0$ and $z_1, z_2 \in \mathbb{R}^p$, it follows that*

$$(28) \quad |\sigma_s(z_1) - \sigma_s(z_2)| \leq \frac{\sqrt{2/e}}{s} \|z_1 - z_2\|_2.$$

Noise is injected into each datapoint through a random variable Z_i , leading to a corrupted dataset $\{X_1, \dots, X_n\}$ given by

$$(29) \quad X_i = Y_i + Z_i, \quad 1 \leq i \leq n.$$

We assume the Z_i are mean-zero, independent and identically distributed, and are sub-Gaussian with variance proxy $\varepsilon \geq 0$. We define sub-Gaussian noise precisely below.

Remark 9. *Two remarks are in order:*

- (1) *The above setup generalizes the case where the clean data points lie on a d -dimensional submanifold embedded in $p = d + k$ -dimensional Euclidean space while the noisy data points are a sample from a tubular neighborhood of the manifold i.e. $Z_i = Q_i^\perp z_i$ where Q_i^\perp is an orthonormal basis of the normal space at Y_i and $z_i \in \mathbb{R}^k$ is uniformly distributed in a ball centered at origin and of radius ε .*
- (2) *If Z_i are non-identically distributed sub-Gaussian random variables with variance proxy ε_i then the following results hold by replacing ε with $\max_1^n \varepsilon_i$.*

Let $Z \in \mathbb{R}^p$ be a random vector. We say that Z is a sub-Gaussian random vector with variance proxy $\varepsilon \geq 0$ and center $\mu \in \mathbb{R}^p$ if, for all $\alpha \in \mathbb{R}^p$, it holds

$$(30) \quad \mathbb{E}(\exp(\alpha^\top(Z - \mu))) \leq \exp(\|\alpha\|_2^2 \varepsilon^2 / 2).$$

We say that Z is centered if Eq. 30 holds when $\mu = 0_p$.

Eq. 29 leads to a noisy kernel adjacency matrix, given entrywise by the expression

$$(31) \quad A_{ij} = s(X_i - X_j) = \exp(-\|X_i - X_j\|_2^2/s^2).$$

We investigate the concentration of $\|A - \bar{A}\|$ in the regime where:

- (i) The variance proxy ε of the noise shrinks as n gets large.

(ii) The feature dimension p of the data remains fixed.

This regime is distinguished from the approach which has been considered in, e.g., [28] and where the authors instead allow the variance proxy to remain bounded from below while the feature dimension is taken to be arbitrarily large. On the other hand, we make assumptions on the decay of the variance proxy ε to obtain guarantees on the concentration of $\|A - \bar{A}\|$ at a distance of roughly $O(n^{1/2})$ w.h.p. as n gets large. Note that we do not need to assume the clean data are uniformly bounded to ensure concentration of $\|A - \bar{A}\|$; however, this will be required later when we need to ensure that degrees are bounded from below.

Theorem 10. *Assume that there exists $c > 0$ for which $\varepsilon < \frac{\sqrt{c}}{\sqrt{n \log n}}$, and let $r > 2$ be fixed. Then there exists a positive constant $C_1 \equiv C_1(s, r, c)$ such that for $n \geq N_1(p, r)$ sufficiently large, it holds*

$$(32) \quad \|A - \bar{A}\|_F \leq C_1 n^{1/2} \text{ and } \|A - \bar{A}\|_\infty \leq C_1 n^{1/2},$$

each with probability at least $1 - n^{-r+2}$.

The proof follows from the Lipschitz continuity of σ_s and an application of the well-known tail bound for quadratic forms of sub-Gaussian random vectors (see [64]).

Next, we define the degree of a node i with clean and noisy adjacency matrices as,

$$(33) \quad d_i = \sum_{j=1}^n A_{ij}, \quad \bar{d}_i = \sum_{j=1}^n \bar{A}_{ij}.$$

Let D, \bar{D} to be the diagonal matrices consisting of d_i and \bar{d}_i , respectively. The following Lemma utilizes the uniform boundedness of the clean data to bound the degrees from below. We will use this result to de-randomize bounds in the subsequent results on the stability of the normalized adjacency matrix and the resulting graph Laplacian.

Lemma 11. *Assume that there exists $c > 0$ for which $\varepsilon < \frac{\sqrt{c}}{\sqrt{n \log n}}$, and let $r > 2$ be fixed. Also assume that for some $R > 0$, $\|Y_i\| \leq R$ for all $i \in [1, n]$. Then there exists a positive constant $C_2 \equiv C_2(p, r, s, c)$ for which*

$$(34) \quad \min \{d_{\min}, \bar{d}_{\min}\} \geq C_2 n$$

with probability at least $1 - n^{-r+2}$, where $d_{\min} = \min_{i=1}^n d_i$ and $\bar{d}_{\min} = \min_{i=1}^n \bar{d}_i$.

We define the clean and noisy normalized adjacency matrices to be

$$(35) \quad \bar{\mathcal{K}} = \bar{D}^{-1} \bar{A} \bar{D}^{-1}, \quad \mathcal{K} = D^{-1} A D^{-1}.$$

Using Theorem 10 and Lemma 11, we obtain concentration of $\|\mathcal{K} - \bar{\mathcal{K}}\|$ as follows.

Theorem 12. *Assume that there exists $c > 0$ for which $\varepsilon < \frac{\sqrt{c}}{\sqrt{n \log n}}$, and let $r > 2$ be fixed. Then there exists a positive constant $C_3 \equiv C_3(p, s, c, r)$ such that for n sufficiently large,*

$$(36) \quad \|\mathcal{K} - \bar{\mathcal{K}}\|_F \leq C_3 n^{-3/2} \text{ and } \|\mathcal{K} - \bar{\mathcal{K}}\|_\infty \leq C_3 n^{-3/2},$$

each with probability at least $1 - 2n^{-r+2}$.

The proof is an adaptation of the proof of Theorem 4 of Deng, Ling, and Strohmer [45]. We first show that $\|\mathcal{K} - \bar{\mathcal{K}}\| \leq C_1 n^{5/2} / \min \{d_{\min}, \bar{d}_{\min}\}^4$ with high probability. This is a randomized bound that depends on d_{\min} , which is de-randomized with high probability using Lemma 11, to obtain the final result.

Next, we define the normalized degrees to be

$$(37) \quad \delta_i = \sum_{j=1}^n \mathcal{K}_{ij}, \quad \bar{\delta}_i = \sum_{j=1}^n \bar{\mathcal{K}}_{ij}.$$

Let $\mathcal{D}, \bar{\mathcal{D}}$ be the diagonal matrices consisting of δ_i and $\bar{\delta}_i$, respectively. Finally, we define the random walk graph Laplacians [43, 44] on the clean and the noisy data as

$$(38) \quad \bar{\mathcal{L}} = I_n - \bar{\mathcal{D}}^{-1} \bar{\mathcal{K}}, \quad \mathcal{L} = I_n - \mathcal{D}^{-1} \mathcal{K}.$$

The above construction of the graph Laplacian is used in all of our experiments. Using Theorem 12, we derive the stability of the Laplacian as follows.

Theorem 13. *Assume that there exists $c > 0$ for which $\varepsilon < \frac{\sqrt{c}}{\sqrt{n} \log n}$, and let $r > 2$ be fixed. Then there exists a positive constant $C_3 \equiv C_3(p, s, c, r)$ such that for n sufficiently large,*

$$(39) \quad \|\mathcal{L} - \bar{\mathcal{L}}\|_F \leq C_3 n^{-1/2}$$

with probability at least $1 - 2n^{-r+2}$.

It follows from the Davis-Kahan theorem (see [46]) that as long as the eigengaps of the clean Laplacian $\bar{\mathcal{L}}$ do not vanish too quickly as n gets large, the eigenvectors of \mathcal{L} and $\bar{\mathcal{L}}$ will remain close as well. We state a somewhat simplified version of this below.

Corollary 14. *Instate the assumptions of Theorem 10 and Theorem 13. Enumerate the eigenvalues of \mathcal{L} and $\bar{\mathcal{L}}$, respectively, as follows:*

$$(40) \quad \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n,$$

$$(41) \quad \xi_1 \leq \xi_2 \leq \dots \leq \xi_n.$$

Let $m \ll n$ be fixed and assume that for each $1 \leq i \leq m-1$, it holds

$$(42) \quad |\xi_i - \xi_{i+1}| = \omega(n^{-1/2}).$$

Then for each $1 \leq i \leq m-1$ fixed, there exists a choice of sign $\tau_i \in \{\pm 1\}$ so that if ϕ_i and ψ_i , respectively, denote the eigenvectors of \mathcal{L} and $\bar{\mathcal{L}}$ with eigenvalues λ_i and ξ_i , then it holds

$$(43) \quad \|\phi_i - \tau_i \psi_i\|_2 = o(1)$$

with probability at least $1 - o(1)$.

Note that there are natural improvements of Corollary 14 to settings where as n gets large $\bar{\mathcal{L}}$ picks up eigenvalues with vanishingly small gaps. We would ask instead that first m eigenvalues of $\bar{\mathcal{L}}$ break into finitely many small groups and the distances between such groups decay at a rate no worse than $\omega(n^{-1/2})$. In this scenario, the distance $\|\phi_i - \tau_i \psi_i\|_2$ would be replaced with distance between the corresponding eigenspaces; i.e., convergence of the eigenvectors up to some orthogonal alignment matrix. For simplicity we do not state such improvements here.

5. EXPERIMENTS

In this section, we estimate tangent spaces on several noisy synthetic and real-world datasets using LPCA and LEGO, compare the estimated tangent spaces against the ground truth, and assess their utility in the following downstream tasks: (a) manifold learning, where we compute an intrinsic-dimensional parametrization of the underlying data manifold; (b) boundary detection, where we identify points that lie on or near the boundary of the data manifold; and (c) local intrinsic dimension estimation, where we determine the dimension of the tangent space at each data point. For completeness, we briefly describe how the estimated tangent spaces are utilized in these tasks in Section C.

To quantify the accuracy of a tangent space estimate $Q_j \in \mathbb{R}^{p \times d}$ at the j -th data point, we compute its deviation from the ground-truth tangent space Q_j^* , obtained from clean data. Specifically, we calculate the principal angles $\theta_{j,1}, \dots, \theta_{j,d}$ between the subspaces spanned by Q_j and Q_j^* [65]. The quality of the estimate is measured by the discrepancy score:

$$(44) \quad D_j = \sum_{i=1}^d (1 - \cos(\theta_{j,i})).$$

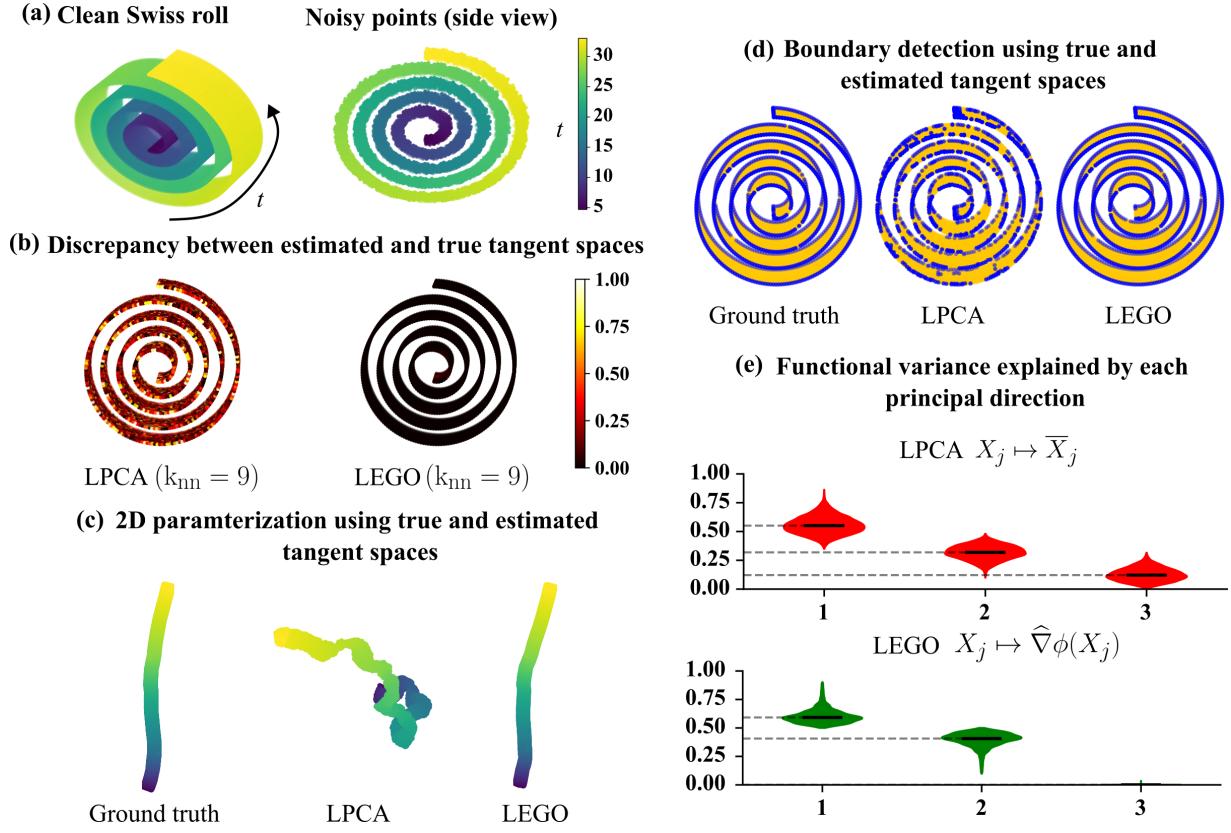


FIGURE 3. (a) Clean and noisy Swiss roll with high-aspect ratio in \mathbb{R}^3 colored by the ‘‘roll’’ parameter. (b) Discrepancy between the true and the estimated tangent spaces due to LPCA ($k_{nn} = 9$) and LEGO ($k_{nn} = 9, m_0 = 100, m = 40$), as computed using Eq. 44. (c, d) 2-dimensional parameterization of the noisy data, and the boundary points detected from the noisy data using the estimated and the true tangent spaces (see Section C.1 and C.2 for details) (e) The functional variance explained by each of the three principal directions in LPCA and LEGO (see Section C.3).

5.1. High-aspect ratio Swiss roll and a truncated torus. We begin with two synthetic datasets: a high-aspect-ratio Swiss roll and a truncated torus. For the Swiss roll, we generate $n = 10700$ uniformly distributed points in \mathbb{R}^3 , forming the clean dataset Y (Figure 3a). Each point is perturbed by adding uniform noise in the direction normal to the underlying tangent space. Specifically, the noisy data points are given by $X_j = Y_j + \eta_j \nu_j$, where ν_j is outward normal to the tangent space at Y_j and the coefficient η_j is uniformly distributed in $(-\varepsilon, \varepsilon)$ where $\varepsilon = 0.0175$. The resulting noisy dataset X is shown in Figure 3a.

For the truncated torus, we sample $n = 3617$ uniformly distributed points on a subset of the torus in \mathbb{R}^3 as shown in Figure 4a. Here, each data point Y_j is parameterized by $(u, v) \in [0, 2\pi]^2$ i.e. $Y_j \equiv Y_j(u_j, v_j)$. We corrupt the clean data with heteroskedastic noise added in the normal direction to the tangent space at each point. The noisy data points are given by $X_j = Y_j + \eta_j \nu_j$, where ν_j is the outward normal direction to the tangent space at Y_j and the coefficient η_j is uniformly distributed in $(-\varepsilon(u_j), \varepsilon(u_j))$ where $\varepsilon(u) = 10^{-2} + 2.5 \times 10^{-3}(1 + \cos(2u))$. The noisy dataset X is shown in Figure 4a.

For both datasets, we estimate an orthonormal basis Q_j of the 2-dimensional tangent space at each X_j using LPCA and LEGO, and then compute the discrepancy D_j (Eq. 44) between the estimates Q_j and the ground-truth Q_j^* , as shown in Figure 3b and 4b. These results show that LEGO produces significantly more accurate estimates while LPCA estimates are highly sensitive to noise. Noise ablation (Figure 6) confirms LPCA estimates degrade rapidly with noise, whereas LEGO consistently yields reliable estimates.

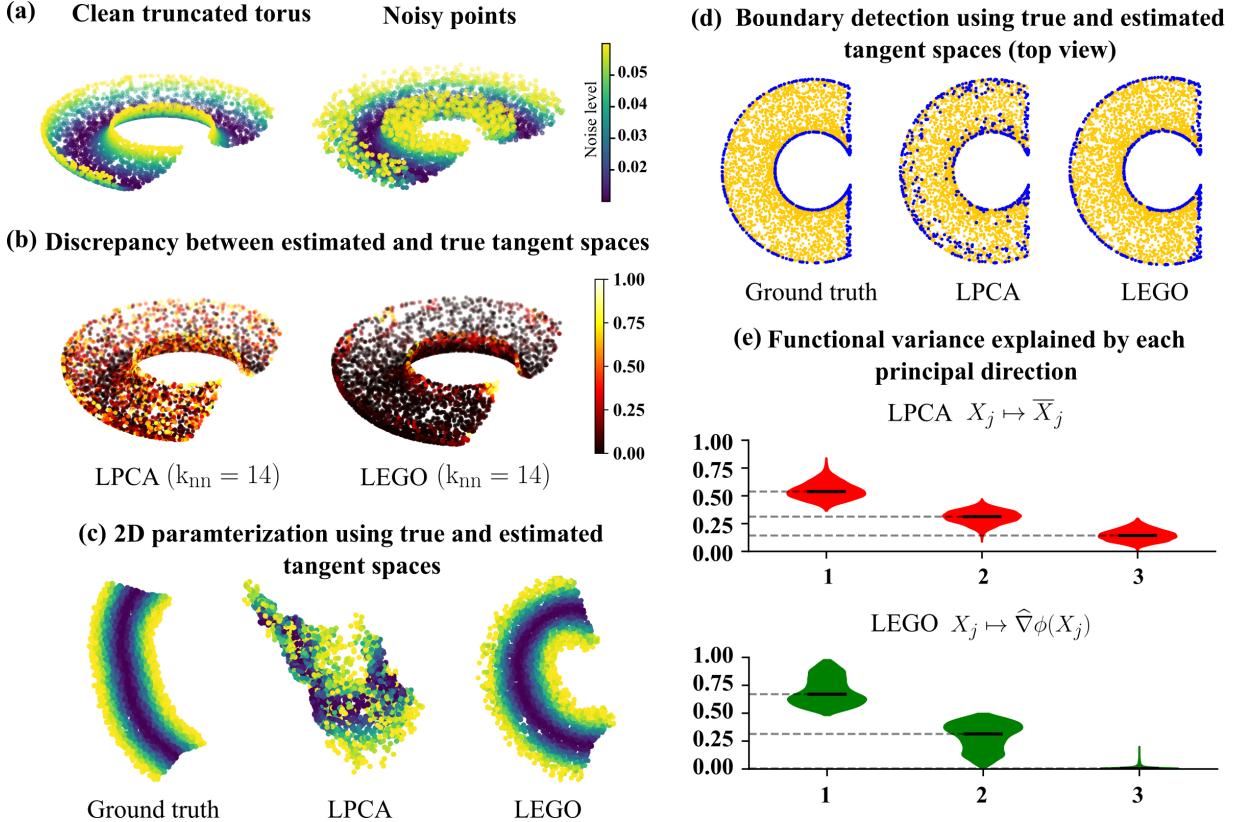


FIGURE 4. (a) Clean and noisy truncated torus in \mathbb{R}^3 colored by the noise level. (b) Discrepancy between the true and the estimated tangent spaces due to LPCA ($k_{nn} = 14$) and LEGO ($k_{nn} = 14, m_0 = 100, m = 20$), as computed using Eq. 44. (c, d) 2-dimensional parameterization of the noisy data, and the boundary points detected from the noisy data using the estimated and the true tangent spaces (see Section C.1 and C.2 for details) (e) The functional variance explained by each of the three principal directions in LPCA and LEGO (see Section C.3).

Hyperparameter analysis (Figure 7) also shows that LEGO estimates remain stable across a broad range of values for m and m_0 .

To assess how these tangent space estimates affect downstream tasks, we use them to compute a 2-dimensional embedding of the noisy data and to detect boundary points (see Section C). As shown in Figure 3c and 4c, and Figure 3d and 4d, the embeddings and the detected boundary points based on LPCA estimates are severely degraded by noise, while those based on LEGO closely match the results obtained using the true tangent spaces. This is not surprising as the accuracy of the tangent space estimation is critical to the performance of several algorithms [1, 35, 3, 5, 38, 37] designed for these downstream tasks.

Finally, by setting $d = p = 3$, we compute the functional variance explained by each principal direction (Section C.3). As shown in Figure 3e and 4e, LEGO concentrates functional variance in the first two directions, aligning with the true intrinsic structure, while LPCA spuriously allocates variance to the third direction, reflecting noise-sensitivity of the local intrinsic dimension estimates due to LPCA.

5.2. Puppets data. In this real-world experiment, we use an image dataset from [66], consisting of $n = 8100$ camera snapshots of a platform with two rotating objects—Yoda and a bulldog—each rotating about its vertical axis at distinct frequencies. As a result, the intrinsic geometry of the dataset corresponds to a 2-dimensional flat torus. The original images of size $320 \times 240 \times 3$ are first normalized to the range

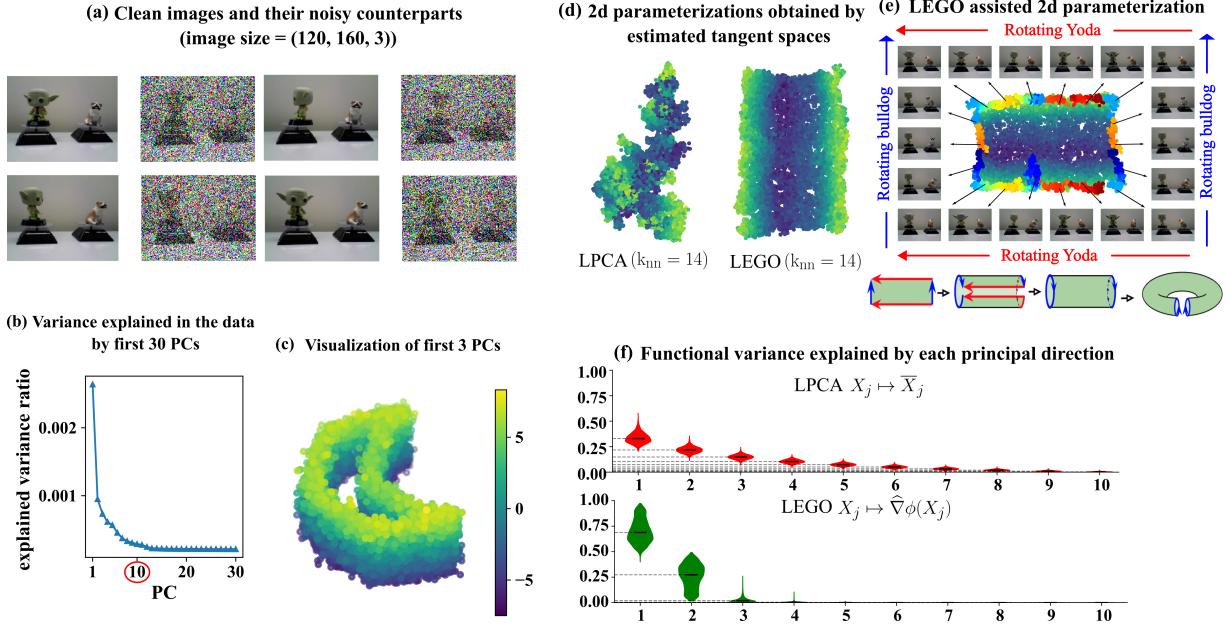


FIGURE 5. (a) Sample clean images from the Yoda and Bulldog dataset [66] (first and third columns), along with their noise-perturbed versions (second and fourth columns). (b) Explained variance ratio for the first 30 principal directions obtained via PCA. As the variance saturates after 10 dimensions, we project the noisy images into \mathbb{R}^{10} using PCA. (c) Visualization of the noisy data using its first three principal components. The colorbar corresponds to the third component. (d) Two-dimensional torn embeddings of the noisy data using the estimated tangent spaces (see Section C.1 and [5] for details). (e) The torn 2d embedding obtained using LEGO estimates, equipped with the gluing instructions that identify the same colored points along the tear, reveals a toroidal topology. The corresponding clean images along the opposite edges further confirm this structure. (f) Functional variance explained by each of the 10 principal directions obtained from LPCA and LEGO (see Section C.3).

[0, 1], followed by addition of uniformly distributed noise in $(-1, 1)$ to each pixel channel. Examples of both clean and noisy images are shown in Figure 5a (the pixel values are clipped between [0, 1] for visualization). Due to computational constraints, we first reduce the dimensionality of the noisy dataset. Based on the explained variance ratio shown in Figure 5b, we project the data to $p = 10$ dimensions, resulting in the final dataset X which is utilized for tangent space estimation.

We then estimate the 2-dimensional tangent spaces using both LPCA ($k_{nn} = 14$) and LEGO ($k_{nn} = 14$, $m_0 = 100$, $m = 20$). These estimates are used to compute a 2-dimensional embedding of the noisy data. Because the data lies on a closed manifold, directly aligning the local intrinsic-dimensional embeddings derived from the tangent space estimates using standard methods leads to a collapse—specifically, the resulting intrinsic parameterization is non-injective. To obtain an injective embedding, we adopt the *tear-enabled* alignment framework introduced in [5], which produces a torn 2d embedding of the data. As shown in Figure 5d, the embedding based on LPCA estimates is non-interpretable, whereas LEGO produces a clear rectangular embedding. When visualized with gluing instructions (Figure 5e)—which identifies the same-colored points along the tear—it becomes evident that opposite edges of the rectangle should be glued, revealing the underlying toroidal topology. Moreover, examining the clean images corresponding to the points on opposite edges shows that only one of the two puppets undergoes rotation, further supporting the toroidal structure.

Finally, by setting $d = p = 10$, we compute the functional variance explained by each of the 10 principal directions obtained by applying LPCA and LEGO to the noisy data (see Section C.3). As shown in Figure 5f, LEGO concentrates the functional variance in the first two directions, faithfully capturing the underlying 2d structure. In contrast, LPCA distributes the variance across multiple dimensions, highlighting its sensitivity to noise and its inability to accurately recover the local intrinsic geometry in the noisy setting.

APPENDIX A. PROOFS FROM 3

Proof of Lemma 1. The following form of the pullback metric $g = \Psi^* \delta_{d+k}$ with respect to the coordinate vector fields defined in Eq. 7, and as derived in [63, Lemma 4.1], is given by

$$(45) \quad g_{i,j}(x, n) = g_B(\partial_{x^i}, \partial_{x^j}) - 2\Pi_\nu(\partial_{x^i}, \partial_{x^j}) + g_B(W_\nu(\partial_{x^i}), W_\nu(\partial_{x^j})) + g_B^\perp(\nabla_{\partial_{x^i}}^\perp \nu, \nabla_{\partial_{x^j}}^\perp \nu)$$

$$(46) \quad g_{i,d+\alpha}(x, n) = g_B^\perp(\nabla_{\partial_{x^i}}^\perp \nu, e_\alpha)$$

$$(47) \quad g_{d+\alpha,d+\beta}(x, n) = g_B^\perp(e_\alpha, e_\beta) = \delta_{\alpha\beta},$$

for $i, j \in [1, d]$, $\alpha, \beta \in [1, k]$ and where (i) $W_{e_\alpha} : \mathfrak{X}(B) \rightarrow \mathfrak{X}(B)$ is the Weingarten map that captures the projection of $\nabla_{\partial_{x^i}} e_\alpha(x)$ on TB i.e. $W_{e_\alpha}(\partial_{x^i}) = -(\nabla_{\partial_{x^i}}^{\mathbb{R}^{d+k}} e_\alpha(x))^\top$, and (ii) $\nabla^\perp : \mathfrak{X}(B) \times \mathfrak{X}(NB) \rightarrow \mathfrak{X}(NB)$ is the normal connection that captures the projection of $\nabla_{\partial_{x^i}}^{\mathbb{R}^{d+k}} e_\alpha(x)$ on NB , and is given by $\nabla_{\partial_{x^i}}^\perp e_\alpha(x) = (\nabla_{\partial_{x^i}}^{\mathbb{R}^{d+k}} e_\alpha(x))^\perp$. Since $\Pi_\nu(\partial_{x^i}, \partial_{x^j}) = g_B^\perp(\nu, \Pi(\partial_{x^i}, \partial_{x^j})) = g_B(\partial_{x^i}, W_\nu(\partial_{x^j})) = g_B(\partial_{x^j}, W_\nu(\partial_{x^i}))$, therefore, using the definitions of $h_{\alpha i}^j$ and $\gamma_{i\alpha}^\beta$ in Eq. 13 and 14,

$$(48) \quad \Pi_\nu(\partial_{x^i}, \partial_{x^j}) = n^\alpha h_{\alpha i}^j = n^\alpha h_{\alpha j}^i.$$

$$(49) \quad W_\nu(\partial_{x^i}) = n^\alpha g_B^{kk'} h_{\alpha k'}^i \partial_{x^k}$$

$$(50) \quad \nabla_{\partial_{x^i}}^\perp \nu = n^\alpha \gamma_{i\alpha}^\beta e_\beta.$$

Therefore,

$$(51) \quad g_{i,j}(x, n) = (g_B)_{ij} - 2n^\alpha h_{\alpha i}^j + n^\alpha n^\beta (g_B)_{kl} g_B^{kk'} h_{\alpha k'}^i g_B^{ll'} h_{\beta l'}^j + n^\alpha n^\beta \gamma_{i\alpha}^\omega \gamma_{j\beta}^{\omega'} \delta_{\omega\omega'}$$

$$(52) \quad = (g_B)_{ij} - 2n^\alpha h_{\alpha i}^j + n^\alpha n^\beta g_B^{k'l'} h_{\alpha k'}^i h_{\beta l'}^j + n^\alpha n^\beta \gamma_{i\alpha}^\omega \gamma_{j\beta}^{\omega'} \delta_{\omega\omega'}$$

$$(53) \quad g_{i,d+\alpha}(x, n) = n^\beta \gamma_{i\beta}^\alpha$$

$$(54) \quad g_{d+\alpha,d+\beta}(x, n) = \delta_{\alpha\beta}.$$

Consequently, the scaled pullback metric $g^\varepsilon = \mathcal{D}_\varepsilon^* \Psi^* \delta_{d+k}$ with respect to the coordinate vector fields in Eq. 7 is given by,

$$(55) \quad g_{i,j}^\varepsilon(x, n) = (g_B)_{ij} - 2\varepsilon n^\alpha h_{\alpha i}^j + \varepsilon^2 n^\alpha n^\beta g_B^{k'l'} h_{\alpha k'}^i h_{\beta l'}^j + \varepsilon^2 n^\alpha n^\beta \gamma_{i\alpha}^\omega \gamma_{j\beta}^{\omega'} \delta_{\omega\omega'}$$

$$(56) \quad g_{i,d+\alpha}^\varepsilon(x, n) = \varepsilon^2 n^\beta \gamma_{i\beta}^\alpha$$

$$(57) \quad g_{d+\alpha,d+\beta}^\varepsilon(x, n) = \varepsilon^2 \delta_{\alpha\beta}$$

Now, the new basis vector $\partial_i^H|_{(x,n)}$ obtained by projecting $\partial_i|_{(x,n)}$ orthogonal to the span of $\{\partial_{d+\alpha}|_{(x,n)}\}_1^k$ is given by,

$$(58) \quad \partial_i^H|_{(x,n)} = \partial_i|_{(x,n)} - \sum_{\alpha=1}^k g_B^\perp(\nabla_{\partial_{x^i}}^\perp \nu, e_\alpha) \partial_{d+\alpha}|_{(x,n)} = \partial_i|_{(x,n)} - n^\beta \gamma_{i\beta}^\alpha \partial_{d+\alpha}|_{(x,n)}.$$

Consequently, the pullback metric g^ε in the new local coordinate fields $\{\partial_i^H|_{(x,n)}\}_1^d$ and $\{\partial_{d+\alpha}|_{(x,n)}\}_1^k$, is given by

$$(59) \quad g_{i,j}^\varepsilon(x, n) = g_B(\partial_{x^i}, \partial_{x^j}) - 2\varepsilon \Pi_\nu(\partial_{x^i}, \partial_{x^j}) + \varepsilon^2 g_B(W_\nu(\partial_{x^i}), W_\nu(\partial_{x^j}))$$

$$(60) \quad = (g_B)_{ij} - 2\varepsilon n^\alpha h_{\alpha i}^j + \varepsilon^2 n^\alpha n^\beta g_B^{k'l'} h_{\alpha k'}^i h_{\beta l'}^j$$

$$(61) \quad g_{i,d+\alpha}^\varepsilon(x, n) = 0$$

$$(62) \quad g^\varepsilon_{d+\alpha, d+\beta}(x, n) = g_B^\perp(e_\alpha, e_\beta) = \varepsilon^2 \delta_{\alpha\beta}.$$

Using the definition of H_α in Eq. 15,

$$(63) \quad g^\varepsilon(x, n) = \begin{bmatrix} g_B - 2\varepsilon n^\alpha H_\alpha + \varepsilon^2 n^\alpha n^\beta H_\alpha g_B^{-1} H_\beta & \\ & \varepsilon^2 I_k \end{bmatrix}$$

$$(64) \quad = \begin{bmatrix} (g_B^{1/2} - \varepsilon n^\alpha H_\alpha g_B^{-1/2})(g_B^{1/2} - \varepsilon g_B^{-1/2} n^\beta H_\beta) & \\ & \varepsilon^2 I_k \end{bmatrix}$$

$$(65) \quad = \begin{bmatrix} g_B^{1/2} (I_d - \varepsilon n^\alpha g_B^{-1/2} H_\alpha g_B^{-1/2})^2 g_B^{1/2} & \\ & \varepsilon^2 I_k \end{bmatrix}$$

Finally, we decompose $\text{grad}\hat{\phi}$ into a component $(\text{grad}\hat{\phi})^H$ on $\pi^*(TB)$ and a component $(\text{grad}\hat{\phi})^V$ on $\ker(\pi_*)$. Specifically, $\text{grad}\hat{\phi} = (\text{grad}\hat{\phi})^H + (\text{grad}\hat{\phi})^V$ where

$$(66) \quad (\text{grad}\hat{\phi})^H = g^{\varepsilon^{ij}} \partial_j^H \hat{\phi} \partial_i^H = g^{\varepsilon^{ij}} \left(\frac{\partial \hat{\phi}}{\partial x^j} - n^\beta \gamma_{j\beta}^\alpha \frac{\partial \hat{\phi}}{\partial n^\alpha} \right) \partial_i^H \text{ and}$$

$$(67) \quad (\text{grad}\hat{\phi})^V = g^{\varepsilon^{d+\alpha, d+\beta}} \frac{\partial \hat{\phi}}{\partial n^\beta} \partial_{d+\alpha} = \varepsilon^{-2} \frac{\partial \hat{\phi}}{\partial n^\alpha} \partial_{d+\alpha}.$$

Using the definition of Γ_β in Eq. 16,

$$(68) \quad \text{grad}\hat{\phi}|_{(x, n)} = \begin{bmatrix} g_B^{-1/2} (I_d - \varepsilon n^\alpha g_B^{-1/2} H_\alpha g_B^{-1/2})^{-2} g_B^{-1/2} (\nabla_x \hat{\phi}(x, n) - n^\beta \Gamma_\beta \nabla_n \hat{\phi}(x, n)) \\ \varepsilon^{-2} \nabla_n \hat{\phi}(x, n) \end{bmatrix}.$$

□

Proof of Lemma 2. Using the expression of g^ε we obtain,

$$(69) \quad \det(g^\varepsilon) = \varepsilon^{2k} \det(g_B) \det \left(I_d - \varepsilon n^\alpha g_B^{-1/2} H_\alpha g_B^{-1/2} \right)^2.$$

Using Cauchy-Schwarz inequality, we obtain

$$(70) \quad \left\| n^\alpha g_B^{-1/2} H_\alpha g_B^{-1/2} \right\|_{\delta_k} = \sup_{\|v\|_2=1} v^T n^\alpha g_B^{-1/2} H_\alpha g_B^{-1/2} v \leq r\kappa(x) \leq r\kappa^*.$$

Since, for each $x \in B$ the maximum value of $r\kappa(x)$ can be realized for some v dependent on x , therefore $\det(g^\varepsilon) > 0$ if and only if $\varepsilon r\kappa(x) < 1$. Under this constraint, it follows that

$$(71) \quad \varepsilon^{2k} \det(g_B) (1 - \varepsilon r\kappa^*)^{2d} \leq \det(g^\varepsilon) \leq \varepsilon^{2k} \det(g_B) (1 + \varepsilon r\kappa^*)^{2d}.$$

□

Proof of Theorem 3. First note that for $f \in C_0^\infty(\mathcal{T}^{\varepsilon r})$,

$$(72) \quad \int_{\mathcal{T}^{\varepsilon r}} f dV_{\delta_{d+k}} = \int_{NB^{\varepsilon r}} (\widehat{\Psi} f) dV_{\Psi^* \delta_{d+k}} = \int_{NB^r} (\widehat{\mathcal{D}}_\varepsilon^{-1} \widehat{\Psi} f) dV_{\mathcal{D}_\varepsilon^* \Psi^* \delta_{d+k}} = \int_{NB^r} (\widehat{\mathcal{D}}_\varepsilon^{-1} \widehat{\Psi} f) dV_{g^\varepsilon}.$$

Therefore, if (λ, ϕ) is an eigenpair of $\Delta_{\delta_{d+k}}$ on $\mathcal{T}^{\varepsilon r}$ with Neumann or Dirichlet boundary conditions then it follows that $\widehat{\phi} = \widehat{\mathcal{D}}_\varepsilon^{-1} \widehat{\Psi} \phi$ is an eigenfunction of Δ_{g^ε} with the same eigenvalue. Specifically,

$$(73) \quad \lambda = \frac{- \int_{\mathcal{T}^{\varepsilon r}} \phi \Delta_{\delta_{d+k}} \phi dV_{\delta_{d+k}}}{\int_{\mathcal{T}^{\varepsilon r}} \phi^2 dV_{\delta_{d+k}}} = \frac{- \int_{NB^r} \widehat{\phi} \Delta_{g^\varepsilon} \widehat{\phi} dV_{g^\varepsilon}}{\int_{NB^r} \widehat{\phi}^2 dV_{g^\varepsilon}} = \frac{\int_{NB^r} \langle \text{grad}\widehat{\phi}, \text{grad}\widehat{\phi} \rangle_{g^\varepsilon} dV_{g^\varepsilon}}{\int_{NB^r} \widehat{\phi}^2 dV_{g^\varepsilon}}.$$

Using Lemma 1, the definition of κ^* in Lemma 2 and Cauchy-Schwarz inequality, we obtain

$$\langle \text{grad}\widehat{\phi}, \text{grad}\widehat{\phi} \rangle_{g^\varepsilon} = \text{grad}\widehat{\phi}^T g^\varepsilon \text{grad}\widehat{\phi} \geq \frac{\nabla_n \widehat{\phi}^T \nabla_n \widehat{\phi}}{\varepsilon^2}$$

and

$$\langle \text{grad}\widehat{\phi}, \text{grad}\widehat{\phi} \rangle_{g^\varepsilon} = \text{grad}\widehat{\phi}^T g^\varepsilon \text{grad}\widehat{\phi}$$

$$\begin{aligned}
&= \left(\nabla_x \hat{\phi} - n^\beta \Gamma_\beta \nabla_n \hat{\phi} \right)^T g_B^{-1/2} (I_d - \varepsilon n^\alpha g_B^{-1/2} H_\alpha g_B^{-1/2})^{-2} g_B^{-1/2} \left(\nabla_x \hat{\phi} - n^\beta \Gamma_\beta \nabla_n \hat{\phi} \right) + \frac{\nabla_n \hat{\phi}^T \nabla_n \hat{\phi}}{\varepsilon^2} \\
&\geq \frac{1}{(1 + \varepsilon r \kappa^*)^2} \left(\nabla_x \hat{\phi} - n^\beta \Gamma_\beta \nabla_n \hat{\phi} \right)^T g_B^{-1} \left(\nabla_x \hat{\phi} - n^\beta \Gamma_\beta \nabla_n \hat{\phi} \right) + \frac{\nabla_n \hat{\phi}^T \nabla_n \hat{\phi}}{\varepsilon^2} \\
&= \frac{1}{(1 + \varepsilon r \kappa^*)^2} \left(\nabla_x \hat{\phi}^T g_B^{-1} \nabla_x \hat{\phi} + \left\| n^\beta g_B^{-1/2} \Gamma_\beta \nabla_n \hat{\phi} \right\|_{\delta_k}^2 - 2n^\beta \nabla_x \hat{\phi}^T g_B^{-1} \Gamma_\beta \nabla_n \hat{\phi} \right) + \frac{\nabla_n \hat{\phi}^T \nabla_n \hat{\phi}}{\varepsilon^2} \\
&\geq \frac{1}{(1 + \varepsilon r \kappa^*)^2} \left(\nabla_x \hat{\phi}^T g_B^{-1} \nabla_x \hat{\phi} - 2n^\beta \nabla_x \hat{\phi}^T g_B^{-1} \Gamma_\beta \nabla_n \hat{\phi} \right) + \frac{\nabla_n \hat{\phi}^T \nabla_n \hat{\phi}}{\varepsilon^2} \\
&\geq \frac{1}{(1 + \varepsilon r \kappa^*)^2} \left(\nabla_x \hat{\phi}^T g_B^{-1} \nabla_x \hat{\phi} - 2 \left(\nabla_x \hat{\phi}^T g_B^{-1} \nabla_x \hat{\phi} \right)^{1/2} \left\| n^\beta g_B^{-1/2} \Gamma_\beta \nabla_n \hat{\phi} \right\|_{\delta_k} \right) + \frac{\nabla_n \hat{\phi}^T \nabla_n \hat{\phi}}{\varepsilon^2} \\
&\geq \frac{1}{(1 + \varepsilon r \kappa^*)^2} \left(\nabla_x \hat{\phi}^T g_B^{-1} \nabla_x \hat{\phi} - 2 \left(\nabla_x \hat{\phi}^T g_B^{-1} \nabla_x \hat{\phi} \right)^{1/2} |n^\beta \kappa_\beta^\perp| \left\| \nabla_n \hat{\phi} \right\|_{\delta_k} \right) + \frac{\nabla_n \hat{\phi}^T \nabla_n \hat{\phi}}{\varepsilon^2} \\
&\geq \frac{1}{(1 + \varepsilon r \kappa^*)^2} \left(\nabla_x \hat{\phi}^T g_B^{-1} \nabla_x \hat{\phi} - 2r\kappa^{\perp*} \left(\nabla_x \hat{\phi}^T g_B^{-1} \nabla_x \hat{\phi} \right)^{1/2} \left\| \nabla_n \hat{\phi} \right\|_{\delta_k} \right) + \frac{\nabla_n \hat{\phi}^T \nabla_n \hat{\phi}}{\varepsilon^2}.
\end{aligned}$$

In the last two equations we used $\|n\|_{\delta_k} \leq r$ and the definitions of κ_β^\perp and $\kappa^{\perp*}$ provided in the statement of the theorem. Combining the above with the bounds on $\det(g^\varepsilon)$ in Lemma 2, we obtain

$$\begin{aligned}
\lambda &= \frac{\int_{NB^r} \langle \text{grad} \hat{\phi}, \text{grad} \hat{\phi} \rangle_{g^\varepsilon} dV_{g^\varepsilon}}{\int_{NB^r} \hat{\phi}^2 dV_{g^\varepsilon}} = \frac{\int_{NB^r} \langle \text{grad} \hat{\phi}, \text{grad} \hat{\phi} \rangle_{g^\varepsilon} \sqrt{\det(g^\varepsilon)} dx^1 \dots dx^d dn^1 \dots dn^k}{\int_{NB^r} \hat{\phi}^2 \sqrt{\det(g^\varepsilon)} dx^1 \dots dx^d dn^1 \dots dn^k} \\
&\geq \frac{(1 - \varepsilon r \kappa^*)^d}{(1 + \varepsilon r \kappa^*)^d} \frac{\int_{NB^r} \langle \text{grad} \hat{\phi}, \text{grad} \hat{\phi} \rangle_{g^\varepsilon} \sqrt{\det(g_B)} dx^1 \dots dx^d dn^1 \dots dn^k}{\int_{NB^r} \hat{\phi}^2 \sqrt{\det(g_B)} dx^1 \dots dx^d dn^1 \dots dn^k} \\
&= \frac{(1 - \varepsilon r \kappa^*)^d}{(1 + \varepsilon r \kappa^*)^d} \frac{\int_{NB^r} \langle \text{grad} \hat{\phi}, \text{grad} \hat{\phi} \rangle_{g^\varepsilon} dV_{g_s}}{\int_{NB^r} \hat{\phi}^2 dV_{g_s}} \\
&\geq \frac{(1 - \varepsilon r \kappa^*)^d}{(1 + \varepsilon r \kappa^*)^d} \left(\frac{E_B(\phi) - 2r\kappa^{\perp*} \frac{\int_{NB^r} (\nabla_x \hat{\phi}^T g_B^{-1} \nabla_x \hat{\phi})^{1/2} \left\| \nabla_n \hat{\phi} \right\|_{\delta_k} dV_{g_s}}{\int_{NB^r} \hat{\phi}^2 dV_{g_s}}}{(1 + \varepsilon r \kappa^*)^2} + \frac{E_B^\perp(\phi)}{\varepsilon^2} \right) \\
&\geq \frac{(1 - \varepsilon r \kappa^*)^d}{(1 + \varepsilon r \kappa^*)^d} \left(\frac{E_B(\phi) - 2r\kappa^{\perp*} \sqrt{E_B(\phi) E_B^\perp(\phi)}}{(1 + \varepsilon r \kappa^*)^2} + \frac{E_B^\perp(\phi)}{\varepsilon^2} \right).
\end{aligned}$$

The result follows from the definitions of normalized horizontal and vertical energies in Eq. 12. Similarly,

$$\begin{aligned}
&\langle \text{grad} \hat{\phi}, \text{grad} \hat{\phi} \rangle_{g^\varepsilon} \\
&= \text{grad} \hat{\phi}^T g^\varepsilon \text{grad} \hat{\phi} \\
&= \left(\nabla_x \hat{\phi} - n^\beta \Gamma_\beta \nabla_n \hat{\phi} \right)^T g_B^{-1/2} (I_d - \varepsilon n^\alpha g_B^{-1/2} H_\alpha g_B^{-1/2})^{-2} g_B^{-1/2} \left(\nabla_x \hat{\phi} - n^\beta \Gamma_\beta \nabla_n \hat{\phi} \right) + \frac{\nabla_n \hat{\phi}^T \nabla_n \hat{\phi}}{\varepsilon^2} \\
&\leq \frac{\left(\nabla_x \hat{\phi} - n^\beta \Gamma_\beta \nabla_n \hat{\phi} \right)^T g_B^{-1} \left(\nabla_x \hat{\phi} - n^\beta \Gamma_\beta \nabla_n \hat{\phi} \right)}{(1 - \varepsilon r \kappa^*)^2} + \frac{\nabla_n \hat{\phi}^T \nabla_n \hat{\phi}}{\varepsilon^2} \\
&= \frac{\left(g_B^{-1/2} \nabla_x \hat{\phi} - n^\beta g_B^{-1/2} \Gamma_\beta \nabla_n \hat{\phi} \right)^T \left(g_B^{-1/2} \nabla_x \hat{\phi} - n^\beta g_B^{-1/2} \Gamma_\beta \nabla_n \hat{\phi} \right)}{(1 - \varepsilon r \kappa^*)^2} + \frac{\nabla_n \hat{\phi}^T \nabla_n \hat{\phi}}{\varepsilon^2} \\
&\leq \frac{\left(\nabla_x \hat{\phi}^T g_B^{-1} \nabla_x \hat{\phi} + \left\| n^\beta g_B^{-1/2} \Gamma_\beta \nabla_n \hat{\phi} \right\|_{\delta_k}^2 \right)}{(1 - \varepsilon r \kappa^*)^2} + \frac{\nabla_n \hat{\phi}^T \nabla_n \hat{\phi}}{\varepsilon^2}
\end{aligned}$$

$$\begin{aligned} &\leq \frac{\left(\nabla_x \widehat{\phi}^T g_B^{-1} \nabla_x \widehat{\phi} + |n^\beta \kappa_\beta^\perp|^2 \nabla_n \widehat{\phi}^T \nabla_n \widehat{\phi}\right)}{(1 - \varepsilon r \kappa^*)^2} + \frac{\nabla_n \widehat{\phi}^T \nabla_n \widehat{\phi}}{\varepsilon^2} \\ &\leq \frac{\left(\nabla_x \widehat{\phi}^T g_B^{-1} \nabla_x \widehat{\phi} + (r \kappa^{\perp*})^2 \nabla_n \widehat{\phi}^T \nabla_n \widehat{\phi}\right)}{(1 - \varepsilon r \kappa^*)^2} + \frac{\nabla_n \widehat{\phi}^T \nabla_n \widehat{\phi}}{\varepsilon^2} \end{aligned}$$

Combining the above with the bounds on $\det(g^\varepsilon)$ in Lemma 2, we obtain

$$\begin{aligned} \lambda &= \frac{\int_{NB^r} \langle \text{grad} \widehat{\phi}, \text{grad} \widehat{\phi} \rangle_{g^\varepsilon} dV_{g^\varepsilon}}{\int_{NB^r} \widehat{\phi}^2 dV_{g^\varepsilon}} = \frac{\int_{NB^r} \langle \text{grad} \widehat{\phi}, \text{grad} \widehat{\phi} \rangle_{g^\varepsilon} \sqrt{\det(g^\varepsilon)} dx^1 \dots dx^d dn^1 \dots dn^k}{\int_{NB^r} \widehat{\phi}^2 \sqrt{\det(g^\varepsilon)} dx^1 \dots dx^d dn^1 \dots dn^k} \\ &\leq \frac{(1 + \varepsilon r \kappa^*)^d}{(1 - \varepsilon r \kappa^*)^d} \frac{\int_{NB^r} \langle \text{grad} \widehat{\phi}, \text{grad} \widehat{\phi} \rangle_{g^\varepsilon} \sqrt{\det g_B} dx^1 \dots dx^d dn^1 \dots dn^k}{\int_{NB^r} \widehat{\phi}^2 \sqrt{\det g_B} dx^1 \dots dx^d dn^1 \dots dn^k} \\ &= \frac{(1 + \varepsilon r \kappa^*)^d}{(1 - \varepsilon r \kappa^*)^d} \frac{\int_{NB^r} \langle \text{grad} \widehat{\phi}, \text{grad} \widehat{\phi} \rangle_{g_s} dV_{g_s}}{\int_{NB^r} \widehat{\phi}^2 dV_{g_s}} \\ &\leq \frac{(1 + \varepsilon r \kappa^*)^d}{(1 - \varepsilon r \kappa^*)^d} \left(\frac{E_B(\phi)}{(1 - \varepsilon r \kappa^*)^2} + \left(\left(\frac{r \kappa^{\perp*}}{1 - \varepsilon r \kappa^*} \right)^2 + \varepsilon^{-2} \right) E_B^\perp(\phi) \right). \end{aligned}$$

The result follows from the definitions of horizontal and vertical energies in Eq. 12. \square

Proof of Theorem 6. Using the fact that $\nabla_n \widehat{\phi} = 0$, the proof is similar to the proof of Theorem 3. \square

APPENDIX B. PROOFS FROM 4

Proof of Lemma 8. Note that $\nabla \sigma_s(z) = -\frac{2}{s^2} z \sigma_s(z)$. Thus, $\|\nabla \sigma_s(z)\|_2 \leq \frac{2}{s^2} \sigma_s(z) \|z\|_2 \leq \frac{\sqrt{2/e}}{s}$. The claim follows from the properties of Lipschitz functions. \square

Lemma 15 (Specialized version of Theorem 2.1 of [64]). *Suppose Z is a centered sub-Gaussian random vector with parameter $\varepsilon \geq 0$. Then for all $t > 0$, it holds*

$$(74) \quad \mathbb{P}(\|Z\|_2^2 > \varepsilon^2(p + 2\sqrt{pt} + 2t)) \leq e^{-t}.$$

Proof of Theorem 10. We begin by writing

$$(75) \quad \|A - \bar{A}\|_F^2 = \sum_{i,j=1}^n (A_{ij} - \bar{A}_{ij})^2 = \sum_{i,j=1}^n |\sigma_s(X_i - X_j) - \sigma_s(Y_i - Y_j)|^2.$$

Using Lemma 8, we have

$$(76) \quad |\sigma_s(X_i - X_j) - \sigma_s(Y_i - Y_j)| \leq \frac{\sqrt{2/e}}{s} \|(X_i - X_j) - (Y_i - Y_j)\|_2 = \|Z_i - Z_j\|_2.$$

Writing $Z_{ij} = Z_i - Z_j$, we have

$$(77) \quad \|A - \bar{A}\|_F^2 \leq \frac{2/e}{s^2} \sum_{i,j=1}^n \|Z_{ij}\|_2^2, \text{ and } \|A - \bar{A}\|_\infty \leq \frac{\sqrt{2/e}}{s} \max_{i=1}^n \sum_{j=1}^n \|Z_{ij}\|_2.$$

We now bound $\sum_{i,j} \|Z_{ij}\|_2^2$ and $\max_i \sum_j \|Z_{ij}\|_2$ using the tail bound given in Lemma 15. Since Z_i and Z_j are independent, each Z_{ij} is a centered sub-Gaussian vector with parameter 2ε so that we have, for $t > 0$,

$$(78) \quad \mathbb{P}(\|Z_{ij}\|_2^2 > 4\varepsilon^2(p + 2\sqrt{pt} + 2t)) \leq e^{-t}.$$

Therefore, by taking $t = r \log n$ in the above inequality and using the union bound, we have

$$(79) \quad \sum_{i,j=1}^n \|Z_{ij}\|_2^2 \leq 4\varepsilon^2 n^2 (p + 2\sqrt{rp \log n} + 2r \log n),$$

with probability at least $1 - n^2 e^{-r \log n}$, and

$$(80) \quad \max_{i=1}^n \sum_{j=1}^n \|Z_{ij}\|_2 \leq 2\varepsilon n(p + 2\sqrt{rp \log n} + 2r \log n)^{1/2},$$

also with probability at least $1 - n^2 e^{-r \log n}$. Using $\varepsilon \leq \sqrt{c}/\sqrt{n \log n}$, we have that for n large enough so that $r \log n \geq \max\{p, 2\sqrt{rp \log n}\}$, it holds

$$(81) \quad 4\varepsilon^2 n^2(p + 2\sqrt{rp \log n} + 2r \log n) \leq 16crn.$$

Substituting in Eq 79 and 80, and using Eq 77, the result follows. \square

Proof of Lemma 11. We proceed with a two step argument. First, we show that with high probability, $d_{\min} \geq C\bar{d}_{\min}$ for some constant $C \in (0, 1)$, and then we show that \bar{d}_{\min} is bounded from below in general. To this end, note that since

$$(82) \quad \|Y_i - Y_j + Z_i - Z_j\|_2^2 \leq 2(\|Y_i - Y_j\|_2^2 + \|Z_i - Z_j\|_2^2),$$

we have, for i fixed,

$$\begin{aligned} d_i &= \sum_{j=1}^n \sigma_s(X_i - X_j) = \sum_{j=1}^n \exp(-\|Y_i - Y_j + Z_i - Z_j\|_2^2/s^2) \\ &\geq \sum_{j=1}^n \exp(-2\|Y_i - Y_j\|_2^2/s^2) \exp(-2\|Z_i - Z_j\|_2^2/s^2) \\ &\geq \sum_{j=1}^n \exp(-\|Y_i - Y_j\|_2^2/s^2) \times \min_{1 \leq j \leq n} \exp(-\|Y_i - Y_j\|_2^2/s^2) \times \min_{1 \leq j \leq n} \exp(-2\|Z_i - Z_j\|_2^2/s^2). \end{aligned}$$

Taking each term one by one, we first have $\sum_{j=1}^n \exp(-\|Y_i - Y_j\|_2^2/s^2) = \bar{d}_i$. Then, we note that since $Y_i, Y_j \in B(0, R)$ where $B(0, R)$ is the ball of radius $R > 0$ in \mathbb{R}^p , it holds

$$(83) \quad \exp(-\|Y_i - Y_j\|_2^2/s^2) \geq \exp(-4R^2/s^2),$$

and therefore

$$(84) \quad \min_{1 \leq j \leq n} \exp(-\|Y_i - Y_j\|_2^2/s^2) \geq \exp(-4R^2/s^2).$$

Finally, we focus on the third term. Letting i, j be fixed, note that $Z_i - Z_j$ is a mean zero centered sub-Gaussian with variance proxy 2ε . Therefore, using Lemma 15, for $t > 0$, we have

$$(85) \quad \mathbb{P}(\|Z_i - Z_j\|_2^2 > 4\varepsilon^2(p + 2\sqrt{pt} + 2t)) \leq e^{-t}.$$

Picking $t = r \log n$, we have that by the union bound,

$$(86) \quad \|Z_i - Z_j\|_2^2 \leq 4\varepsilon^2(p + 2\sqrt{pr \log n} + 2r \log n)$$

for all $j \in [1, n]$ with probability at least $1 - n^{-r+1}$. Picking n large enough so that $2r \log n \geq \max\{p, 2\sqrt{pr \log n}\}$ (as in the proof of Theorem 10) and $\varepsilon \leq \sqrt{c}/\sqrt{n \log n}$, it holds

$$(87) \quad 4\varepsilon^2(p + 2\sqrt{pr \log n} + 2r \log n) \leq \frac{16cr}{n}.$$

Thus, with probability at least $1 - n^{-r+1}$, and since $n \geq e^{p/2r}$ as before,

$$(88) \quad \min_{1 \leq j \leq n} \exp(-2\|Z_i - Z_j\|_2^2/s^2) \geq \exp\left(-\frac{32cr}{s^2 n}\right) \geq \exp\left(-\frac{32cr}{s^2 e^{p/2r}}\right).$$

Thus, if we put

$$(89) \quad c_1 = \exp\left(-\frac{32cr}{s^2 e^{p/2r}} - \frac{4R^2}{s^2}\right),$$

then $0 < c_1 < 1$ and with probability at least $1 - n^{-r+1}$, $d_i \geq c_1 \bar{d}_i$. Therefore, by the union bound, with probability at least $1 - n^{-r+2}$, it holds $d_{\min} \geq c_1 \bar{d}_{\min}$. Note that by Eq 84, we have that

$$(90) \quad \bar{d}_{\min} \geq n \min_{1 \leq i, j \leq n} \exp(-\|Y_i - Y_j\|_2^2/s^2) \geq n \exp(-4R^2/s^2),$$

so that $\bar{d}_{\min} \geq c_2 n$ in turn. Therefore, with high probability, $\min\{d_{\min}, \bar{d}_{\min}\} \geq c_1 c_2 n$. \square

Proof of Theorem 12. In the following proof, we use $\|\cdot\|$ to denote both $\|\cdot\|_F$ and $\|\cdot\|_\infty$.

Assume, as in the proof of Theorem 10, that n is sufficiently large. Begin by writing

$$\|\mathcal{K} - \bar{\mathcal{K}}\|_2 = \|D^{-1}AD^{-1} - \bar{D}^{-1}\bar{A}\bar{D}^{-1}\| \leq \|D^{-1}(A - \bar{A})D^{-1}\| + \|D^{-1}\bar{A}D^{-1} - \bar{D}^{-1}\bar{A}\bar{D}^{-1}\|.$$

The first term can be bounded easily using Theorem 10, namely,

$$(91) \quad \|D^{-1}(A - \bar{A})D^{-1}\| \leq \|D^{-1}\|_2 \|A - \bar{A}\| \|D^{-1}\|_2 \leq \frac{\|A - \bar{A}\|}{d_{\min}^2} \leq \frac{C_1 n^{1/2}}{d_{\min}^2}.$$

with probability at least $1 - n^{-r+2}$ and where $C_1 \equiv C_1(s, r, c) > 0$.

For the second half, let $v = A\mathbf{1}_n$, $\bar{v} = \bar{A}\mathbf{1}_n$ and $d_0 = \min\{d_{\min}, \bar{d}_{\min}\}$. Then, in the case of Frobenius norm,

$$\begin{aligned} \|D^{-1}\bar{A}D^{-1} - \bar{D}^{-1}\bar{A}\bar{D}^{-1}\|_F &= \sqrt{\sum_{i,j=1}^n |\sigma_s(Y_i - Y_j)|^2 \left(\frac{1}{d_i d_j} - \frac{1}{\bar{d}_i \bar{d}_j}\right)^2} \leq \sqrt{\sum_{i,j=1}^n \left(\frac{\bar{d}_i \bar{d}_j - d_i d_j}{d_i d_j \bar{d}_i \bar{d}_j}\right)^2} \\ &\leq \frac{1}{d_0^4} \sqrt{\sum_{i,j=1}^n (\bar{d}_i \bar{d}_j - d_i d_j)^2} \leq \frac{1}{d_0^4} \|vv^T - \bar{v}\bar{v}^T\|_F \leq \frac{1}{d_0^4} (\|v(v - \bar{v})^T\|_F + \|\bar{v}(v - \bar{v})^T\|_F) \\ &\leq \frac{1}{d_0^4} (\|v\|_2 \|v - \bar{v}\|_2 + \|\bar{v}\|_2 \|v - \bar{v}\|_2) \leq \frac{1}{d_0^4} (\|A\|_2 + \|\bar{A}\|_2) \|A - \bar{A}\|_2 \|\mathbf{1}_n\|_2^2. \end{aligned}$$

Applying Theorem 10 once again and using $\|\bar{A}\|_2 \leq \|\bar{A}\|_F = \sqrt{\sum_{i,j} \bar{A}_{ij}^2} \leq n \max_{i,j} \bar{A}_{ij} \leq n$, $\|A\|_2 \leq n$ and $\|A - \bar{A}\|_2 \leq \|A - \bar{A}\|_F$, we have

$$(92) \quad \|D^{-1}\bar{A}D^{-1} - \bar{D}^{-1}\bar{A}\bar{D}^{-1}\|_F \leq \frac{C_1 n^{5/2}}{d_0^4}$$

with probability at least $1 - n^{-r+2}$ and where $C_1 = C_1(s, r, c) > 0$.

Similarly, in the case of infinity norm, using $d_i \leq n$, $\bar{d}_i \leq n$ and $|d_i - \bar{d}_i| \leq \|A - \bar{A}\|_\infty$,

$$\begin{aligned} \|D^{-1}\bar{A}D^{-1} - \bar{D}^{-1}\bar{A}\bar{D}^{-1}\|_\infty &= \max_{i=1}^n \sum_{j=1}^n |\sigma_s(Y_i - Y_j)| \left| \frac{1}{d_i d_j} - \frac{1}{\bar{d}_i \bar{d}_j} \right| \leq \frac{1}{d_0^4} \sum_{j=1}^n |\bar{d}_i \bar{d}_j - d_i d_j| \\ &\leq \frac{1}{d_0^4} \|\bar{d}_i \bar{v} - d_i v\|_1 \leq \frac{\sqrt{n}}{d_0^4} \|\bar{d}_i \bar{v} - d_i v\|_2 \leq \frac{\sqrt{n}}{d_0^4} (d_i \|\bar{v} - v\|_2 + |d_i - \bar{d}_i| \|\bar{v}\|_2) \\ &\leq \frac{n^2}{d_0^4} (\|A - \bar{A}\|_2 + |d_i - \bar{d}_i|) \leq \frac{n^2}{d_0^4} (\|A - \bar{A}\|_2 + \|A - \bar{A}\|_\infty) \leq \frac{C_1 n^{5/2}}{d_0^4} \end{aligned}$$

with probability at least $1 - n^{-r+2}$ and where $C_1 = C_1(s, r, c) > 0$. Therefore,

$$\|D^{-1}\bar{A}D^{-1}\mathbf{1}_n - \bar{D}^{-1}\bar{A}\bar{D}^{-1}\mathbf{1}_n\|_\infty \leq \frac{C_1 n^{5/2}}{d_0^4}.$$

with probability at least $1 - n^{-r+2}$ and where $C_1 = C_1(s, r, c) > 0$. Combining the two estimates, we then have (for both Frobenius and infinity norms)

$$\|\mathcal{K} - \bar{\mathcal{K}}\| \leq \frac{C_1 n^{1/2}}{d_{\min}^2} + \frac{C_1 n^{5/2}}{d_0^4} \leq \frac{C_1 \bar{d}_{\min}^2 n^{1/2}}{\bar{d}_{\min}^2 d_{\min}^2} + \frac{C_1 n^{5/2}}{d_0^4} \leq \frac{C_1 n^{5/2}}{d_0^4}$$

with probability at least $1 - n^{-r+2}$.

Finally, using Lemma 11, we de-randomize the above bound,

$$(93) \quad \|\mathcal{K} - \bar{\mathcal{K}}\| \leq \frac{C_1 C_2 n^{5/2}}{d_0^4} \leq C_3 n^{-3/2}.$$

with probability at least $1 - 2n^{-r+2}$ and where $C_3 \equiv C_3(p, r, s, c) > 0$. \square

Proof of Theorem 13. Assume, as in the proof of Theorem 12, that n is sufficiently large. Begin by writing

$$(94) \quad \|\mathcal{L} - \bar{\mathcal{L}}\|_F = \|\mathcal{D}^{-1}\mathcal{K} - \bar{\mathcal{D}}^{-1}\bar{\mathcal{K}}\|_2 \leq \|\mathcal{D}^{-1}(\mathcal{K} - \bar{\mathcal{K}})\|_F + \|\mathcal{D}^{-1}\bar{\mathcal{K}} - \bar{\mathcal{D}}^{-1}\bar{\mathcal{K}}\|_F.$$

The first term can be bounded easily using Theorem 10, namely,

$$(95) \quad \|\mathcal{D}^{-1}(\mathcal{K} - \bar{\mathcal{K}})\|_F \leq \|\mathcal{D}^{-1}\|_2 \|\mathcal{K} - \bar{\mathcal{K}}\|_F \leq \frac{\|\mathcal{K} - \bar{\mathcal{K}}\|_F}{\delta_{\min}} \leq \frac{C_3 n^{-3/2}}{\delta_{\min}}$$

with probability at least $1 - 2n^{-r+2}$ and where $C_3 \equiv C_3(s, r, c) > 0$. Note that

$$(96) \quad \delta_i = \sum_{j=1}^n \mathcal{K}_{ij} = \frac{1}{d_i} \sum_{j=1}^n \frac{A_{ij}}{d_j} \geq \frac{1}{d_{\max}} \frac{1}{d_i} \sum_{j=1}^n A_{ij} = \frac{1}{d_{\max}} \geq \frac{1}{n} \implies \frac{1}{\delta_{\min}} \leq n.$$

Therefore, $\|\mathcal{D}^{-1}(\mathcal{K} - \bar{\mathcal{K}})\|_2 \leq C_3 n^{-1/2}$ with probability at least $1 - n^{-r+2}$ and where $C_3 \equiv C_3(s, r, c) > 0$.

For the second half, let $v = \mathcal{K}\mathbf{1}_n$ and $\bar{v} = \bar{\mathcal{K}}\mathbf{1}_n$. Then, using the fact that spectral radius of row-stochastic matrices equals 1,

$$\begin{aligned} \|\mathcal{D}^{-1}\bar{\mathcal{K}} - \bar{\mathcal{D}}^{-1}\bar{\mathcal{K}}\|_2 &\leq \|\mathcal{D}^{-1}\bar{\mathcal{D}} - I_n\|_2 \|\bar{\mathcal{D}}^{-1}\bar{\mathcal{K}}\|_2 \leq \frac{1}{\delta_{\min}} \|\bar{\mathcal{D}} - \mathcal{D}\|_2 \\ &\leq \frac{1}{\delta_{\min}} \|v - \bar{v}\|_{\infty} \leq \frac{1}{\delta_{\min}} \|\mathcal{K} - \bar{\mathcal{K}}\|_{\infty} \leq C_3 n^{-1/2}. \end{aligned}$$

\square

APPENDIX C. OVERVIEW OF DOWNSTREAM TASKS INVOLVING TANGENT SPACE ESTIMATION

C.1. Bottom-up manifold learning. Given data points $X = [X_1, \dots, X_n] \in \mathbb{R}^{p \times n}$ sampled from a d -dimensional data manifold where $p \geq d$, bottom-up manifold learning algorithms [1, 2, 3, 4, 5] aim to recover a d -dimensional parameterization of the data by first constructing local embeddings and then aligning them on the overlaps to obtain a global embedding. To construct local embeddings (also called local views), an orthonormal basis $Q_j \in \mathbb{R}^{p \times d}$ is estimated for the tangent space at each data point X_j . Points in the neighborhood \mathcal{N}_j of X_j are then projected onto the estimated tangent space of X_j to obtain d -dimensional local coordinates. Precisely, the local coordinate of a neighbor X_{j_s} , where $j_s \in \mathcal{N}_j$, is given by:

$$(97) \quad \theta_{j_s, j} = Q_j^T(X_j - \mu_j) \quad \text{where} \quad \mu_j = \frac{1}{k_{\text{nn}}} \sum_{s=1}^{k_{\text{nn}}} X_{j_s}.$$

In our experiments we estimate Q_j either using LPCA or our proposed method LEGO.

Once the local views are computed, they are aligned on the overlaps using rigid transformations, to obtain a global embedding [36, 67, 68]. This leads to an optimization problem where, for each local view, we estimate an orthogonal matrix $S_j^* \in \mathbb{O}(d)$ and a translation vector $t_j^* \in \mathbb{R}^d$ that minimize the alignment error:

$$(98) \quad \min_{(S_j, t_j)} \sum_{j=1}^n \sum_{k \in \mathcal{N}_i \cap \mathcal{N}_j} \|(S_i^T \theta_{k,i} + t_i) - (S_j^T \theta_{k,j} + t_j)\|_2^2.$$

In our experiments, we follow the procedure in [68, 5] where we first initialize the rigid transformations using Procrustes analysis and then refine them using Riemannian gradient descent (RGD) [69, 68, 5]. Once

the rigid transformations are obtained, the global embedding is computed by averaging the transformed local coordinates:

$$(99) \quad \Theta_k = \frac{\sum_{k \in \mathcal{N}_j} S_j^{*T} \theta_{k,j} + t_j^*}{|\{j : k \in \mathcal{N}_j\}|}.$$

Note that when the data lies on a closed manifold—as in the case of the Yoda-Bulldog dataset—and the goal is to recover an embedding in the intrinsic dimension, the above alignment strategy leads to a collapsed non-injective embedding [5]. To address this, a tear-enabled rigid alignment framework was introduced in [5], which tears the data manifold to avoid collapse. We adopt this framework to compute a torn embedding in 2 dimensions of the Yoda-Bulldog dataset, along with gluing instructions at the tear (see Figure 5e). For further details regarding the tear-enabled alignment framework, we refer the reader to [5].

C.2. Boundary detection. Berry and Sauer in [38, 39] proposed a method to estimate the boundary points using a local statistic that approximates the normal direction near the boundary and yields a zero vector in the interior. We recently developed a robust variant of this approach in [37], where we combined the doubly stochastic kernel [51, 70] along with tangent space estimates to detect the boundary points. The method starts by approximating the normal direction at X_j as,

$$(100) \quad \nu_j := \frac{1}{n-1} \sum_{i=1}^n W_{ij} Q_j^T (X_i - X_j).$$

Here W is the doubly stochastic kernel derived from the Gaussian kernel on the data via Sinkhorn iterations (see [70, 51] for details), and Q_j represents an estimate of the orthonormal basis of the tangent space at X_j . After computing ν_j for each $j \in [1, n]$, boundary points are identified by thresholding the norm $\|\nu_j\|_2$. In our experiments, we select a fixed percentile of these values as the threshold, labeling X_j as a boundary point if $\|\nu_j\|_2$ exceeds this threshold. The same percentile is used consistently across all methods for a given dataset.

C.3. Local intrinsic dimension. When the local intrinsic dimension d_j at the data point X_j is not known apriori, a typical procedure to estimate it—as used in LPCA—is to compute the eigenvalues $\tilde{\lambda}_1 \geq \dots \geq \tilde{\lambda}_p$ of the local covariance matrix

$$(101) \quad C_j = \sum_{s=1}^{k_{nn}} (X_{js} - \mu_j)(X_{js} - \mu_j)^T \quad \text{where } \mu_j = \frac{1}{k_{nn}} \sum_{s=1}^{k_{nn}} X_{js},$$

and then assess the variance explained by the i th principal direction (the i -th eigenvector of C_j) at X_j via $\tilde{\lambda}_i / \sum_{k=1}^p \tilde{\lambda}_k$. The local intrinsic dimension at X_j is then selected to be the smallest index at which the cumulative variance explained exceeds a user defined threshold.

This strategy can be readily adapted to our method, with the key difference being that the eigenvalues $\tilde{\lambda}_1, \dots, \tilde{\lambda}_p$ are now derived from the surrogate covariance matrix $\widehat{\nabla}\phi(X_j)\widehat{\nabla}\phi(X_j)^T$, constructed from the gradients of the low-frequency eigenvectors in LEGO. As demonstrated in our experiments, this provides a robust estimate of the local intrinsic dimension in the presence of noise and varying sampling density.

ACKNOWLEDGMENTS

DK and SJR were partially supported by the Halicioğlu Data Science Institute PhD Fellowship. DK and GM acknowledge support from NSF EFRI 2223822. AC and GM acknowledge support from NSF CISE 2403452.

REFERENCES

- [1] Zhenyue Zhang and Hongyuan Zha. “Principal manifolds and nonlinear dimensionality reduction via tangent space alignment”. In: *SIAM journal on scientific computing* 26.1 (2004), pp. 313–338.
- [2] Sam T Roweis and Lawrence K Saul. “Nonlinear dimensionality reduction by locally linear embedding”. In: *science* 290.5500 (2000), pp. 2323–2326.

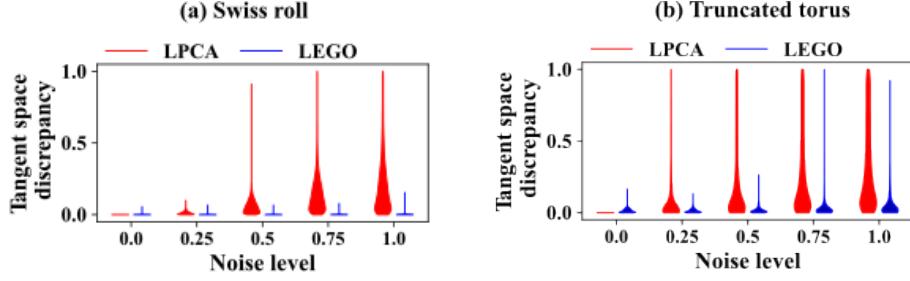


FIGURE 6. The discrepancy between true and the estimated tangent spaces from the noisy data $X_j = Y_j + \sigma \eta_j \nu_j$, $j \in [1, n]$, as the noise level σ varies between 0 and 1. In our experiments in Section 5, we used the maximum noise level i.e. $\sigma = 1$.

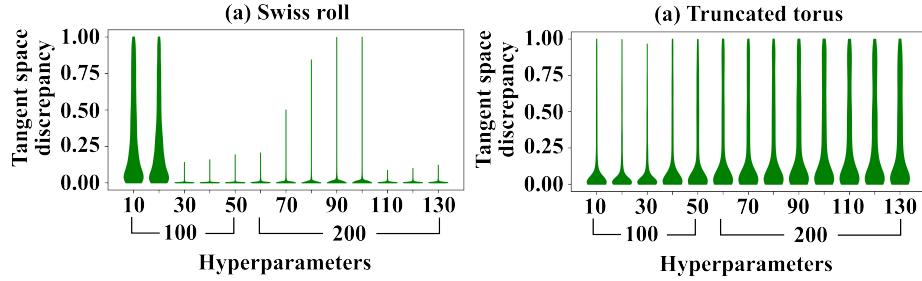


FIGURE 7. The discrepancy between the true and the estimated tangent spaces on the noisy datasets described in Section 5, against several different values of the hyperparameters m and m_0 in LEGO, provided at the top and the bottom of the x -axis, respectively. The noise level corresponds to the maximum noise in Figure 6. Note that $m = 10$ and 20 result in high tangent space discrepancy for the Swiss roll. This is because of its high aspect ratio which makes the gradients of the first 20 eigenvectors to be restricted to a one-dimensional subspace.

- [3] David L Donoho and Carrie Grimes. “Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data”. In: *Proceedings of the National Academy of Sciences* 100.10 (2003), pp. 5591–5596.
- [4] Dhruv Kohli, Alexander Cloninger, and Gal Mishne. “LDLE: Low Distortion Local Eigenmaps”. In: *Journal of Machine Learning Research* 22.282 (2021), pp. 1–64.
- [5] Dhruv Kohli et al. “RATS: Unsupervised manifold learning using low-distortion alignment of tangent spaces”. In: *bioRxiv* (2024).
- [6] Marina Meilă and Hanyu Zhang. “Manifold learning: What, how, and why”. In: *Annual Review of Statistics and Its Application* 11 (2024).
- [7] Cyril de Bodt et al. “Low-dimensional embeddings of high-dimensional data”. In: *arXiv preprint arXiv:2508.15929* (2025).
- [8] Dian Gong, Fei Sha, and Gérard Medioni. “Locally linear denoising on image manifolds”. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings. 2010, pp. 265–272.
- [9] Nicolas Garcia Trillos, Pengfei He, and Chenghui Li. “Large sample spectral analysis of graph-based multi-manifold clustering”. In: *Journal of Machine Learning Research* 24.143 (2023), pp. 1–71.
- [10] Yong Wang et al. “Spectral clustering on multiple manifolds”. In: *IEEE Transactions on Neural Networks* 22.7 (2011), pp. 1149–1161.
- [11] Dian Gong, Xuemei Zhao, and Gérard Medioni. “Robust multiple manifolds structure learning”. In: *arXiv preprint arXiv:1206.4624* (2012).

- [12] Ery Arias-Castro, Guangliang Chen, and Gilad Lerman. “Spectral clustering based on local linear approximations”. In: *Electronic Journal of Statistics* 5.none (2011), pp. 1537–1587.
- [13] Elizaveta Levina and Peter Bickel. “Maximum likelihood estimation of intrinsic dimension”. In: *Advances in neural information processing systems* 17 (2004).
- [14] Amit Singer and Hau-Tieng Wu. “Spectral convergence of the connection Laplacian from random samples”. In: *Information and Inference: A Journal of the IMA* 6.1 (2017), pp. 58–123.
- [15] Amit Singer and H-T Wu. “Vector diffusion maps and the connection Laplacian”. In: *Communications on pure and applied mathematics* 65.8 (2012), pp. 1067–1144.
- [16] Sawyer Robertson et al. “On a generalization of Wasserstein distance and the Beckmann problem to connection graphs”. In: *arXiv:2312.10295* (2023).
- [17] Ming-Yen Cheng and Hau-tieng Wu. “Local linear regression on manifolds and its geometric interpretation”. In: *Journal of the American Statistical Association* 108.504 (2013), pp. 1421–1434.
- [18] Eddie Aamari and Clément Levraud. “Nonasymptotic rates for manifold, tangent space and curvature estimation”. In: *The Annals of Statistics* 47.1 (2019), pp. 177–204.
- [19] Daniel N Kaslovsky and François G Meyer. “Non-asymptotic analysis of tangent space perturbation”. In: *Information and Inference: a Journal of the IMA* 3.2 (2014), pp. 134–187.
- [20] Hemant Tyagi, Elif Vural, and Pascal Frossard. “Tangent space estimation for smooth embeddings of Riemannian manifolds^(R)”. In: *Information and Inference: A Journal of the IMA* 2.1 (2013), pp. 69–114.
- [21] Mikhail Belkin and Partha Niyogi. “Laplacian eigenmaps for dimensionality reduction and data representation”. In: *Neural computation* 15.6 (2003), pp. 1373–1396.
- [22] Ronald R Coifman and Stéphane Lafon. “Diffusion maps”. In: *Applied and computational harmonic analysis* 21.1 (2006), pp. 5–30.
- [23] Yu-Chia Chen and Marina Meila. “Selecting the independent coordinates of manifolds with large aspect ratios”. In: *Advances in Neural Information Processing Systems* 32 (2019).
- [24] Alex Kokot, Octavian-Vlad Murad, and Marina Meila. “The Noisy Laplacian: a Threshold Phenomenon for Non-Linear Dimension Reduction”. In: *Forty-second International Conference on Machine Learning*. 2025.
- [25] Peter W Jones, Mauro Maggioni, and Raanan Schul. “Universal local parametrizations via heat kernels and eigenfunctions of the Laplacian”. In: *arXiv preprint arXiv:0709.1975* (2007).
- [26] John M Lee. *Introduction to Riemannian manifolds*. Vol. 2. Springer, 2018.
- [27] François G Meyer and Xilin Shen. “Perturbation of the eigenvectors of the graph Laplacian: Application to image denoising”. In: *Applied and Computational Harmonic Analysis* 36.2 (2014), pp. 326–334.
- [28] Xiucai Ding and Hau-Tieng Wu. “Impact of signal-to-noise ratio and bandwidth on graph Laplacian spectrum from high-dimensional noisy point cloud”. In: *IEEE Transactions on Information Theory* 69.3 (2022), pp. 1899–1931.
- [29] David I Shuman et al. “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains”. In: *IEEE signal processing magazine* 30.3 (2013), pp. 83–98.
- [30] David K Hammond, Pierre Vandergheynst, and Rémi Gribonval. “Wavelets on graphs via spectral graph theory”. In: *Applied and Computational Harmonic Analysis* 30.2 (2011), pp. 129–150.
- [31] Xiuyuan Cheng and Gal Mishne. “Spectral embedding norm: Looking deep into the spectrum of the graph Laplacian”. In: *SIAM journal on imaging sciences* 13.2 (2020), pp. 1015–1048.
- [32] Elias M Stein and Rami Shakarchi. *Fourier analysis: an introduction*. Vol. 1. Princeton University Press, 2011.
- [33] Yitzhak Katznelson. *An introduction to harmonic analysis*. Cambridge University Press, 2004.
- [34] Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999.
- [35] Zhenyue Zhang and Jing Wang. “MLLE: Modified locally linear embedding using multiple weights”. In: *Advances in neural information processing systems* 19 (2006).
- [36] Kunal N Chaudhury, Yuehaw Khoo, and Amit Singer. “Global registration of multiple point clouds using semidefinite programming”. In: *SIAM Journal on Optimization* 25.1 (2015), pp. 468–501.

- [37] Dhruv Kohli et al. “Robust estimation of boundary using doubly stochastic scaling of Gaussian kernel”. In: *arXiv:2411.18942* (2024).
- [38] Tyrus Berry and Timothy Sauer. “Density estimation on manifolds with boundary”. In: *Computational Statistics & Data Analysis* 107 (2017), pp. 1–17.
- [39] Ryan Vaughn, Tyrus Berry, and Harbir Antil. “Diffusion maps for embedded manifolds with boundary with applications to PDEs”. In: *Applied and Computational Harmonic Analysis* 68 (2024), p. 101593. ISSN: 1063-5203.
- [40] Herbert Federer. “Curvature measures”. In: *Transactions of the American Mathematical Society* 93.3 (1959), pp. 418–491.
- [41] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. “Finding the homology of submanifolds with high confidence from random samples”. In: *Discrete & Computational Geometry* 39 (2008), pp. 419–441.
- [42] Noureddine El Karoui. “On information plus noise kernel random matrices”. In: *Annals of statistics* 38.5 (2010), pp. 3191–3216.
- [43] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. “Graph Laplacians and their convergence on random neighborhood graphs.” In: *Journal of Machine Learning Research* 8.6 (2007).
- [44] Ronald R Coifman and Stéphane Lafon. “Diffusion maps”. In: *Applied and computational harmonic analysis* 21.1 (2006), pp. 5–30.
- [45] Shaofeng Deng, Shuyang Ling, and Thomas Strohmer. “Strong consistency, graph Laplacians, and the stochastic block model”. In: *Journal of Machine Learning Research* 22.117 (2021), pp. 1–44.
- [46] Y. Yu, T. Wang, and R. J. Samworth. “A Useful Variant of the Davis–Kahan Theorem for Statisticians”. In: *Biometrika* 102.2 (June 2015), pp. 315–323.
- [47] Isaac Chavel. *Eigenvalues in Riemannian geometry*. Vol. 115. Academic press, 1984.
- [48] Lihi Zelnik-Manor and Pietro Perona. “Self-tuning spectral clustering”. In: *Advances in Neural Information Processing Systems* (2005), pp. 1601–1608.
- [49] Xiuyuan Cheng and Hau-Tieng Wu. “Convergence of graph Laplacian with kNN self-tuned kernels”. In: *Information and Inference: A Journal of the IMA* 11.3 (2022), pp. 889–957.
- [50] Nicholas F Marshall and Ronald R Coifman. “Manifold learning with bi-stochastic kernels”. In: *IMA Journal of Applied Mathematics* 84.3 (2019), pp. 455–482.
- [51] Boris Landa, Ronald R Coifman, and Yuval Kluger. “Doubly stochastic normalization of the Gaussian kernel is robust to heteroskedastic noise”. In: *SIAM journal on mathematics of data science* 3.1 (2021), pp. 388–413.
- [52] Mikhail Belkin and Partha Niyogi. “Towards a theoretical foundation for Laplacian-based manifold methods”. In: *Journal of Computer and System Sciences* 74.8 (2008), pp. 1289–1308.
- [53] Amit Singer. “From graph to manifold Laplacian: The convergence rate”. In: *Applied and Computational Harmonic Analysis* 21.1 (2006), pp. 128–134.
- [54] Nicolás García Trillo et al. “Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace–Beltrami operator”. In: *Foundations of Computational Mathematics* 20.4 (2020), pp. 827–887.
- [55] Xiuyuan Cheng and Boris Landa. “Bi-stochastically normalized graph Laplacian: convergence to manifold Laplacian and robustness to outlier noise”. In: *Information and Inference: A Journal of the IMA* 13.4 (2024), iaae026.
- [56] Richard B Lehoucq, Danny C Sorensen, and Chao Yang. *ARPACK users’ guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. SIAM, 1998.
- [57] Mikhail Belkin and Partha Niyogi. “Laplacian eigenmaps for dimensionality reduction and data representation”. In: *Neural computation* 15.6 (2003), pp. 1373–1396.
- [58] Yariv Aizenbud and Barak Sober. “Non-parametric estimation of manifolds from noisy data”. In: *arXiv preprint arXiv:2105.04754* (2021).
- [59] Yariv Aizenbud and Barak Sober. “Estimation of Local Geometric Structure on Manifolds from Noisy Data”. In: *arXiv preprint arXiv:2503.07220* (2025).
- [60] Anna V Little. “Estimating the intrinsic dimension of high-dimensional data sets: a multiscale, geometric approach”. PhD thesis. Duke University, 2011.

- [61] Anna V Little, Mauro Maggioni, and Lorenzo Rosasco. “Multiscale geometric methods for data sets I: Multiscale SVD, noise and curvature”. In: *Applied and Computational Harmonic Analysis* 43.3 (2017), pp. 504–567.
- [62] Christopher R Genovese et al. “Minimax manifold estimation”. In: *The Journal of Machine Learning Research* 13.1 (2012), pp. 1263–1291.
- [63] Stefan Haag, Jonas Lampart, and Stefan Teufel. “Generalised quantum waveguides”. In: *Annales Henri Poincaré*. Vol. 16. Springer. 2015, pp. 2535–2568.
- [64] Daniel Hsu, Sham Kakade, and Tong Zhang. “A tail inequality for quadratic forms of subgaussian random vectors”. In: *Electronic Communications in Probability* none (2012), pp. 1–6. DOI: 10.1214/ECP.v17-2079.
- [65] Andrew V Knyazev and Merico E Argentati. “Principal angles between subspaces in an A-based scalar product: algorithms and perturbation estimates”. In: *SIAM Journal on Scientific Computing* 23.6 (2002), pp. 2008–2040.
- [66] Roy R Lederman and Ronen Talmon. “Learning the geometry of common latent variables using alternating-diffusion”. In: *Applied and Computational Harmonic Analysis* 44.3 (2018), pp. 509–536.
- [67] Shuyang Ling. “Generalized power method for generalized orthogonal Procrustes problem: global convergence and optimization landscape analysis”. In: *arXiv preprint arXiv:2106.15493* (2021).
- [68] Dhruv Kohli, Gal Mishne, and Alexander Cloninger. “Non-degenerate rigid alignment in a patch framework”. In: *arXiv:2303.11620* (2023).
- [69] Shankar Krishnan et al. “Global registration of multiple 3D point sets via optimization-on-a-manifold.” In: *Symposium on Geometry Processing*. 2005, pp. 187–196.
- [70] Boris Landa and Xiuyuan Cheng. “Robust inference of manifold density and geometry by doubly stochastic scaling”. In: *SIAM Journal on Mathematics of Data Science* 5.3 (2023), pp. 589–614.