

When Locally Linear Embedding Hits Boundary

Hau-Tieng Wu

HAUWU@MATH.DUKE.EDU

*Department of Mathematics and Department of Statistical Science, Duke University,
Durham, NC 27708, United States;*

Nan Wu

NAN.WU@UTDALLAS.EDU

*Department of Mathematical Sciences, The University of Texas at Dallas,
Richardson, TX 75080, United States*

Editor: Aapo Hyvärinen

Abstract

Based on the Riemannian manifold model, we study the asymptotic behavior of a widely applied unsupervised learning algorithm, *locally linear embedding (LLE)*, when the point cloud is sampled from a compact, smooth manifold *with boundary*. We show several peculiar behaviors of LLE near the boundary that are different from those diffusion-based algorithms. In particular, we show that LLE pointwisely converges to a *mixed-type differential operator with degeneracy* and we calculate the convergence rate. The impact of the hyperbolic part of the operator is discussed and we propose a clipped LLE algorithm which is a potential approach to recover the Dirichlet Laplace-Beltrami operator.

Keywords: Locally linear embedding; manifold learning; manifold with boundary; mixed-type differential operator; Dirichlet Laplace-Beltrami operator.

1. Introduction

Arguably, unsupervised learning is the holy grail of artificial intelligence. While a lot of challenges are on different fronts, many attempts have been explored, including ISOMAP (Tenenbaum et al., 2000), locally linear embedding (LLE) (Roweis and Saul, 2000), Hessian LLE (Donoho and Grimes, 2003), eigenmap (Belkin and Niyogi, 2003), diffusion map (DM) (Coifman and Lafon, 2006), vector diffusion map (VDM) (Singer and Wu, 2012), t-distributed stochastic neighboring embedding (van der Maaten and Hinton, 2008), maximal variation unfolding (Weinberger and Saul, 2006), to name but a few. In this paper, based on the Riemannian manifold model, we study the asymptotic behavior of LLE when the point cloud is sampled from a compact, smooth manifold with boundary.

LLE is an algorithm based on a rudimentary idea – by well parametrizing the dataset locally, we can patch all local information to recover the global one. It has been widely applied in different fields and has been cited more than 15,800 times according to Google Scholar. However, its theoretical justification for data points sampled on compact manifolds without boundary was only made available at the end of 2017 (Wu and Wu, 2018; Malik et al., 2019). Essentially, the established theory says that under the manifold without boundary setup, LLE has several peculiar behaviors that are different from those of diffusion-based algorithms, including eigenmap and DM. First, unlike DM, LLE may not behave like a diffusion process since the associated kernel function is not always positive. Second, it is sensitive to the regularization, and different regularizations lead to different

differential operators. If the regularization is chosen properly, LLE asymptotically converges to the Laplace-Beltrami operator without extra probability density function (p.d.f.) estimation, even if the p.d.f. is not uniform. However, when the regularization is not chosen properly, LLE converges to a fourth order differential operator in the cases like the spheres. Third, when the regularization is chosen properly, the convergence of LLE to the Laplace-Beltrami operator is comparable to that of DM with a proper normalization (Coifman and Lafon, 2006; Singer and Wu, 2017; Cheng and Wu, 2022). Fourth, the kernel associated with LLE is in general not symmetric, and this asymmetric kernel depends on the curvature and p.d.f. information. Fifth, the kernel depends on the local covariance matrix analysis and the Mahalanobis distance, since it is the mix-up of the ordinary kernel and a special kernel depending on the Mahalanobis distance (Malik et al., 2019).

While several theoretical properties have been discussed in Wu and Wu (2018) and (Malik et al., 2019), there are more open problems about LLE left. In this paper, we are interested in exploring the asymptotic behavior of LLE when the manifold has a boundary. First, we show that asymptotically LLE pointwisely converges to a mixed-type differential operator with degeneracy and we calculate the convergence rate. Second, after showing that the asymptotic operator near the boundary involves singular coefficients, we study the 1-dim manifold case and relate the eigenvalue problem of LLE to a Sturm-Liouville equation. Third, through a series of numerical simulations, we explore the impact of the hyperbolic part of the operator. In those simulations, we modify the LLE by clipping certain points that are close to the boundary, which asymptotically is equivalent to eliminating the hyperbolic part of the operator, then we obtain an algorithm that is potential to recover the Laplace-Beltrami operator with the Dirichlet boundary condition. This enlightens a new approach to recovering the Dirichlet Laplace-Beltrami operator. Fourth, we compare LLE with DM to explain the differences between their behaviors on the boundary.

The paper is organized in the following way. In Section 2, we review the LLE algorithm and provide some spectral properties of LLE on the linear algebra level. In Section 3, we provide the manifold model when the boundary is not empty, and develop the asymptotic theory for the LLE matrix, particularly the associated kernel behavior and its relationship with the geometrical structure of the manifold. In Section 7, we discuss the clipped LLE which potentially leads to the Laplace-Beltrami operator with the Dirichlet boundary condition. Numerical simulations of the clipped LLE are provided. The paper is closed with the discussion in Section 8. Technical proofs are postponed to the Appendices. For reproducibility purposes, the Matlab code to reproduce figures in this paper can be downloaded from <http://hautiengwu.wordpress.com/code/>.

2. Review locally linear embedding

We start with some matrix notations. For $p, r \in \mathbb{N}$ so that $r \leq p$, let $I_r \in \mathbb{R}^{r \times r}$ be the identity matrix.

Denote $J_{p,r} = \begin{bmatrix} I_r \\ 0 \end{bmatrix} \in \mathbb{R}^{p \times r}$, i.e. the (i,i) -th entry of $J_{p,r}$ is 1 for $i = 1, \dots, r$, and the other entries are zero. Denote $\bar{J}_{p,r} = \begin{bmatrix} 0 \\ I_r \end{bmatrix} \in \mathbb{R}^{p \times r}$, i.e. the $(p-r+i, i)$ entry of $J_{p,r}$ is 1 for $i = 1, \dots, r$, and the other entries are zero. Denote $I_{p,r} := J_{p,r}J_{p,r}^\top = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{p \times p}$ and $\bar{I}_{p,r} := \bar{J}_{p,r}\bar{J}_{p,r}^\top = \begin{bmatrix} 0 & 0 \\ 0 & I_r \end{bmatrix} \in \mathbb{R}^{p \times p}$.

Finally, for $d \leq r \leq p$, define $\mathfrak{J}_{p,r-d} := \bar{J}_{p,p-d}J_{p-d,r-d} \in \mathbb{R}^{p \times (r-d)}$.

We quickly recall necessary information about LLE and refer readers with interest in more discussion to (Roweis and Saul, 2000; Wu and Wu, 2018). The key ingredient of LLE is the *barycentric*

coordinate, which is a quantity shown in Wu and Wu (2018) to be parallel to the kernel chosen in the graph Laplacian. Suppose we have the point cloud $\mathcal{X} = \{z_i\}_{i=1}^n \subset \mathbb{R}^p$. There are two nearest neighbor search schemes to proceed. The first one is the *ε -radius ball scheme*. Fix $\varepsilon > 0$. For $z_k \in \mathcal{X}$, assume there are N_k data points, excluding z_k , in the ε -radius ball centered at z_k . The second one is the *K-nearest neighbor (KNN) scheme* used in the original LLE algorithm (Roweis and Saul, 2000); that is, for a fixed $K \in \mathbb{N}$, find the K neighboring points.

Fix one nearest neighbor search scheme, and denote the nearest neighbors of $z_k \in \mathcal{X}$ as $\mathcal{N}_k = \{z_{k,i}\}_{i=1}^{N_k}$. Then the barycentric coordinate of z_k associated with \mathcal{N}_k , denoted as w_k , is defined as the solution of the following optimization problem:

$$w_k = \arg \min_{w \in \mathbb{R}^{N_k}, w^\top \mathbf{1}_{N_k} = 1} \left\| z_k - \sum_{j=1}^{N_k} w(j) z_{k,j} \right\|^2 = \arg \min_{w \in \mathbb{R}^{N_k}, w^\top \mathbf{1}_{N_k} = 1} w^\top G_{n,k}^\top G_{n,k} w \in \mathbb{R}^{N_k}, \quad (2.1)$$

where $\mathbf{1}_{N_k}$ is a vector in \mathbb{R}^{N_k} with all entries 1 and

$$G_{n,k} := \begin{bmatrix} & & & \\ & | & & | \\ z_{k,1} - z_k & \dots & z_{k,N_k} - z_k \\ & | & & | \end{bmatrix} \in \mathbb{R}^{p \times N_k} \quad (2.2)$$

is called the *local data matrix*. In general, $G_{n,k}^\top G_{n,k}$ might be singular, and it is suggested in Roweis and Saul (2000) to stabilize the algorithm by regularizing the equation and solve

$$(G_{n,k}^\top G_{n,k} + cI_{N_k \times N_k})y_k = \mathbf{1}_{N_k}, \quad w_k = \frac{y_k}{y_k^\top \mathbf{1}_{N_k}}, \quad (2.3)$$

where $c > 0$ is the *regularizer* chosen by the user. As is shown in Wu and Wu (2018), the regularizer plays a critical role in LLE. With the barycentric coordinate of x_k for $k = 1, \dots, n$, the *LLE matrix*, which is a $n \times n$ matrix denoted as W , is defined as

$$W_{ki} = \begin{cases} w_k(j) & \text{if } z_i = z_{k,j} \in \mathcal{N}_k; \\ 0 & \text{otherwise.} \end{cases} \quad (2.4)$$

The barycentric coordinates are invariant under rotation and translation, because the matrix $G_{n,k}$ is invariant under translation, and $G_{n,k}^\top G_{n,k}$ is invariant under rotation. As discussed in Wu and Wu (2018), the barycentric coordinates can be understood as the projection of $\mathbf{1}_{N_k}$ onto the null space of $G_{n,k}^\top G_{n,k}$.

Suppose $r_n = \text{rank}(G_{n,k}^\top G_{n,k})$. Note that $r_n = \text{rank}(G_{n,k}) = \text{rank}(G_{n,k}^\top G_{n,k}) = \text{rank}(G_{n,k} G_{n,k}^\top) \leq \min(N_k, p) \leq p$ and $G_{n,k} G_{n,k}^\top$ is positive (semi-)definite. Denote the eigen-decomposition of the matrix $G_{n,k} G_{n,k}^\top$ as $U_n \Lambda_n U_n^\top$, where $\Lambda_n = \text{diag}(\lambda_{n,1}, \lambda_{n,2}, \dots, \lambda_{n,p})$, $\lambda_{n,1} \geq \lambda_{n,2} \geq \dots \geq \lambda_{n,r_n} > \lambda_{n,r_n+1} = \dots = \lambda_{n,p} = 0$, and $U_n \in O(p)$. Denote

$$\mathcal{J}_c(G_{n,k} G_{n,k}^\top) := U_n I_{p,r_n} (\Lambda_n + cI_{p \times p})^{-1} U_n^\top, \quad (2.5)$$

and

$$\mathbf{T}_{n,x_k} := \mathcal{J}_c(G_{n,k} G_{n,k}^\top) G_{n,k} \mathbf{1}_{N_k}. \quad (2.6)$$

Then, it is shown in (Wu and Wu, 2018, Section 2) that the solution to (2.3) is

$$y_k^\top = c^{-1} \mathbf{1}_{N_k}^\top - c^{-1} \mathbf{T}_{n,x_k}^\top G_{n,k}, \quad (2.7)$$

and hence

$$w_k^\top = \frac{\mathbf{1}_{N_k}^\top - \mathbf{T}_{n,x_k}^\top G_{n,k}}{N_k - \mathbf{T}_{n,x_k}^\top G_{n,k} \mathbf{1}_{N_k}}. \quad (2.8)$$

Note that $N_k - \mathbf{T}_{n,x_k}^\top G_{n,k} \mathbf{1}_{N_k}$ in the denominator of (2.8) is the sum of entries of $\mathbf{1}_{N_k}^\top - \mathbf{T}_{n,x_k}^\top G_{n,k}$ in the numerator, so we could view y_k^\top as the “kernel function” associated with LLE, and w_k^\top as the normalized kernel.

To reduce the dimension of \mathcal{X} , it is suggested in Roweis and Saul (2000) to embed \mathcal{X} into a low dimension Euclidean space via

$$z_k \mapsto Y_k = [v_1(k), \dots, v_\ell(k)]^\top \in \mathbb{R}^\ell \quad (2.9)$$

for each $z_k \in \mathcal{X}$, where $\ell \in \mathbb{N}$ is the dimension of the embedded points chosen by the user and $v_1, \dots, v_\ell \in \mathbb{R}^n$ are eigenvectors of $(I - W)^\top (I - W)$ corresponding to the ℓ smallest eigenvalues.

2.1 Spectral properties of the LLE matrix

We provide some spectral properties of the LLE matrix. Unlike the graph Laplacian (GL), in general, W is not a symmetric matrix or a Markov transition matrix, according to the analysis shown in Wu and Wu (2018). For $A \in \mathbb{R}^{n \times n}$, let $\sigma(A) \subset \mathbb{C}$ be the spectrum of A and define $\rho(A)$ to be the spectral radius of A .

Proposition 1 *The LLE matrix $W \in \mathbb{R}^{n \times n}$ satisfies $\rho(W) \geq 1$.*

The proof of the proposition is straightforward. Since $W\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is an n -dim vector with all entries 1, $1 \in \sigma(W)$. Thus we have that $\rho(W) \geq 1$. In Appendix A, we construct an example to show that it is possible for $\rho(W) > 1$.

Since in general, the LLE matrix W may not be symmetric, the eigenvalues might be complex and can be complicated. For example, in the null case that 400 points are sampled independently and identically from a 200-dim Gaussian random vector with 0 mean and identity covariance, the eigenvalue distribution of W spreads on the complex plane. See Figure 1 for the distribution of such a dataset.

However, in some special cases, we can well control the imaginary part of the distribution. Consider the symmetric and anti-symmetric parts of W , $W^+ = (W + W^\top)/2$ and $W^- = (W - W^\top)/2$, so that $W = W^+ + W^-$. By applying the Bauer-Fike theorem with the L^2 norm and Holder's inequality, for any eigenvalue λ of W , there is a real eigenvalue μ of W^+ such that $|\lambda - \mu| \leq \|W^-\|_2 \leq \sqrt{\|W^-\|_1 \|W^-\|_\infty}$. Below we show that the imaginary part of eigenvalues of the LLE matrix W is well controlled under some conditions.

Proposition 2 *Denote $N = \max_k N_k$, where $N_k = |\mathcal{N}_k|$. If $\max_{i,j} |W_{ij} - W_{ji}| \leq \frac{C\varepsilon}{N}$ for some $C \geq 0$, the imaginary part of eigenvalues of the LLE matrix W is of order ε .*

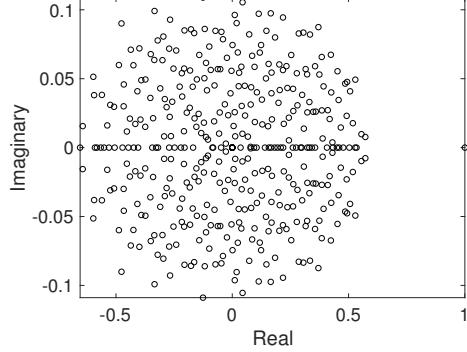


Figure 1: The distribution of eigenvalues of the LLE matrix, where W is constructed with 50 nearest neighbors. In this example, the top eigenvalue is 1.

Proof Note that $W_{ij}^- = 0$ if $\|z_i - z_j\|_2 \geq \varepsilon$ and W_{ij}^- might be nonzero if $\|z_i - z_j\|_2 < \varepsilon$. Since $\sqrt{\|W^-\|_1 \|W^-\|_\infty} \leq N \max_{i,j} |W_{ij} - W_{ji}|$, based on the assumption, the imaginary part of eigenvalues of W is bounded by $O(\varepsilon)$. \blacksquare

Note that $\max_{i,j} |W_{ij} - W_{ji}|$ measures the similarity of different ε -neighborhood \mathcal{N}_k . Thus, the assumption that $\max_{i,j} |W_{ij} - W_{ji}| \leq \frac{C\varepsilon}{N}$ for some $C \geq 0$ means that the affinity graph is “not too imbalanced”. This assumption holds asymptotically under the manifold setup.

Since the KNN scheme and the ε -radius ball scheme are directly related under a suitable manipulation as is shown in Wu and Wu (2018, Section 5), from now on we fix to the ε -radius ball scheme in the rest of the paper for the sake of theoretical analysis.

3. Preliminaries for LLE under the manifold with boundary setup

3.1 Main assumptions

In this subsection, we summarize the major assumptions that we need in this paper. First, we have the following assumption about the manifold M .

Assumption 3.1 Let (M, g) be a d -dimensional compact, smooth Riemannian manifold with boundary isometrically embedded in \mathbb{R}^p via $\iota : M \hookrightarrow \mathbb{R}^p$. We assume the boundary of M is smooth.

Next, we make the following assumption about the sample points on the manifold M .

Assumption 3.2 Suppose (Ω, \mathcal{F}, P) is a probability space, where P is a probability measure defined on the Borel sigma algebra \mathcal{F} on Ω . Let X be a random variable on (Ω, \mathcal{F}, P) with the range on (M, g) . We assume $P := X_*P$ is absolutely continuous with respect to the volume measure on M associated with g so that $dP = PdV$ by the Radon-Nikodym theorem, where dV is the volume form of M and P is a non-negative function defined on M . We call P the probability density function (p.d.f.) associated with X . We further assume $P \in C^2(M)$ and $0 < P_m \leq P(x) \leq P_M$ for all $x \in M$. We assume $\{x_1, \dots, x_n\} \subset M$ are i.i.d. sampled from P .

Remark 3 Under the regularity assumption of the boundary and the density function in this model, in general, the chance to sample a point on the boundary is zero, unless we further assume the knowledge of the boundary and sample on the boundary. Without the knowledge of the boundary, an estimate of the boundary is therefore needed. Such estimate has wide applications including the distance to the boundary estimation and kernel density estimation on a manifold with boundary. We refer the readers to Berry and Sauer (2017) for a discussion.

Remark 4 We refer the readers to Lee (2012) for a discussion of the differentiability of a function on the boundary of the manifold. Compared with the $P \in C^5(M)$ requirement imposed in Wu and Wu (2018), in this work we only assume $P \in C^2(M)$. In Wu and Wu (2018), we need $P \in C^5(M)$ to explore the regularization effect on the whole algorithm. In this work, since we will fix the regularization and focus on the boundary, $P \in C^2(M)$ is sufficient.

We adopt the notations in Section 2. Let $\mathcal{X} = \{z_i = \iota(x_i)\}_{i=1}^n$. Fix $\varepsilon > 0$, we propose the choice of the parameter for regularization:

$$c = n\varepsilon^{d+3}.$$

Then, we construct the LLE matrix W by using \mathcal{X} and $c = n\varepsilon^{d+3}$ as shown in (2.3) and (2.4).

3.2 Manifold with boundary setup

The manifold setup is nowadays standard and has been considered to study several algorithms, including Eigenmap (Belkin and Niyogi, 2007), DM (Coifman and Lafon, 2006; Trillos et al., 2020), VDM (Singer and Wu, 2012, 2017), LLE (Wu and Wu, 2018) and several others, like the gradient estimation (Mukherjee et al., 2010), diffusion on the fiber structure (Lin et al., 2018; Gao, 2021), Bayesian regression (Yang and Dunson, 2016), extrinsic local regression (Lin et al., 2017), image processing model (Osher et al., 2017), sensor fusion algorithm (Shnitzer et al., 2018), to name but a few. Although the manifold model is standard, when the boundary is non-empty, it is less discussed in the literature. We introduce the following setup for manifold with boundary. See Vaughn et al. (2019) for a different treatment.

Denote $d_g(\cdot, \cdot)$ to be the geodesic distance associated with g . For $\varepsilon > 0$, define the ε -neighborhood of ∂M as

$$M_\varepsilon = \{x \in M | d_g(x, \partial M) < \varepsilon\}. \quad (3.1)$$

For the tangent space $T_x M$ on $x \in M$, denote $\iota_* T_x M$ to be the embedded tangent space in \mathbb{R}^p and $(\iota_* T_x M)^\perp$ be the normal space at $\iota(x)$. Let \mathbb{I}_x be the second fundamental form of $\iota(M)$ at $\iota(x)$. Denote S^{d-1} to be the $(d-1)$ -dim unit sphere embedded in \mathbb{R}^p , and $|S^{d-1}|$ be its volume. Denote $\{e_i\}_{i=1}^p$ to be the canonical basis of \mathbb{R}^p , where e_i is a unit vector with 1 in the i -th entry. Since the barycentric coordinate is rotational and translational invariant, without loss of generality, when we analyze local behaviors around $x \in M$ in this paper, we implicitly assume that the manifold has been properly translated and rotated so that $\iota_* T_x M$ is spanned by e_1, \dots, e_d .

We consider the following extension of M and $\iota(M)$ (Wong, 2008). By the Whitney extension Theorem (Whitney, 1992), there is a compact manifold with boundary \tilde{M} and $\delta > 0$ satisfying the following properties:

1. \tilde{M} is an isometric extension of M .
2. $d_{\tilde{M}}(\partial M, \partial \tilde{M}) \geq \delta$, where $d_{\tilde{M}}$ is the geodesic distance measured in \tilde{M} .

3. \tilde{M} is isometrically embedded in \mathbb{R}^p via $\tilde{\iota}$ such that $\tilde{\iota}|_M = \iota$. Thus, $\iota(M) \subset \tilde{\iota}(\tilde{M})$.

Due to the above extension, we abuse the notation and use \exp_x to denote the exponential map of \tilde{M} at $x \in \tilde{M}$ and the exponential map of M at $x \in M$. Note that when $x \in M_\varepsilon$, the exponential map of M at x may not be well defined on $\iota^{-1}(B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \iota(M))$. For example, consider an annulus in the plane and $\iota(x)$ is a point on the inner circle. However, when $\varepsilon < \frac{\delta}{2}$ and ε is small enough, we can make sure that $B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \tilde{\iota}(\tilde{M})$ is contained in the interior of $\tilde{\iota}(\tilde{M})$ and \exp_x is well defined over $\tilde{\iota}^{-1}(B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \tilde{\iota}(\tilde{M}))$. Since $B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \iota(M) \subset B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \tilde{\iota}(\tilde{M})$, \exp_x is well defined over $\tilde{\iota}^{-1}(B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \iota(M))$. Hence, we conclude that \exp_x is well defined over $\tilde{\iota}^{-1}(B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \iota(M))$ for any x in both M_ε and $M \setminus M_\varepsilon$.

Now, we can handle the ε -ball near the boundary. For $x \in M_\varepsilon$, define

$$D_\varepsilon(x) = (\tilde{\iota} \circ \exp_x)^{-1}(B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \iota(M)) \subset T_x \tilde{M},$$

where $T_x \tilde{M}$ is identified with \mathbb{R}^d . Denote $x_\partial := \arg \min_{y \in \partial M} d(y, x)$ and

$$\tilde{\varepsilon}_x = \min_{y \in \partial M} d(y, x). \quad (3.2)$$

Due to the smoothness assumption of the boundary, if ε is sufficiently small, such x_∂ is unique. Clearly, we have $0 \leq \tilde{\varepsilon}_x \leq \varepsilon$ when $x \in M_\varepsilon$. Choose the normal coordinates $\{\partial_i\}_{i=1}^d$ around x so that $x_\partial = \tilde{\iota} \circ \exp_x(\tilde{\varepsilon}_x \partial_d)$. Denote $\gamma_x(t)$ to be the unique geodesic with $\gamma_x(0) = x_\partial$ and $\gamma_x(\tilde{\varepsilon}_x) = x$. We further rotate $\iota(M)$ so that

$$e_d = \iota_* \frac{d}{dt} \gamma_x|_{\tilde{\varepsilon}_x}.$$

Hence, when $x = x_\partial \in \partial M$, e_d is the inward normal direction of $\iota(\partial M)$ at $\iota(x)$.

Since \tilde{M} is an isometric extension of M and $\tilde{\iota}$ is an extension of ι , for $x \in M$, we use Π_x to denote both the second fundamental form of $\iota(M)$ at $\iota(x)$ and the second fundamental form of $\tilde{\iota}(M)$ at $\tilde{\iota}(x)$. Recall that the second fundamental form at x is a symmetric bilinear map from $T_x M \times T_x M$ to $(\iota_* T_x M)^\perp$. We define $\Pi_{ij}(x) = \Pi_x(\partial_i, \partial_j)$ for $i, j = 1, \dots, d$.

When x is close to the boundary, $(\tilde{\iota} \circ \exp_x)^{-1}(B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \iota(\partial M))$ is not empty and can be regarded as the graph of a function. Denote $a_{ij}(x_\partial)$, $i, j = 1, \dots, d-1$, to be the second fundamental form of the embedding of ∂M into M at x_∂ . Then there is a domain $K \subset \mathbb{R}^{d-1}$ and a smooth function q defined on K , such that

$$\begin{aligned} & (\tilde{\iota} \circ \exp_x)^{-1}(B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \iota(\partial M)) \\ &= \left\{ \sum_{l=1}^d u^l \partial_l \in T_x M \mid (u^1, \dots, u^{d-1}) \in K, u^d = q(u^1, \dots, u^{d-1}) \right\}, \end{aligned}$$

where $q(u^1, \dots, u^{d-1})$ can be approximated by

$$\tilde{\varepsilon}_x + \sum_{i,j=1}^{d-1} a_{ij}(x_\partial) u^i u^j$$

up to an error depending on a cubic function of u^1, \dots, u^{d-1} . For the sake of self containedness, we provide a proof of this fact in Lemma 24.

Note that in general the region $D_\varepsilon(x)$ may not be symmetric with respect to x . We define the symmetrized region associated with $D_\varepsilon(x)$.

Definition 5 For $x \in M_\varepsilon$ and $\varepsilon > 0$ sufficiently small, the symmetrized region associated with $D_\varepsilon(x)$ is defined as

$$\tilde{D}_\varepsilon(x) = \left\{ (u_1, \dots, u_d) \in T_x \tilde{M} \mid \sum_{i=1}^d u_i^2 \leq \varepsilon^2 \text{ and } u_d \leq \tilde{\varepsilon}_x + \sum_{i,j=1}^{d-1} a_{ij}(x_\partial) u_i u_j \right\}.$$

When $x \in M_\varepsilon$, $\tilde{D}_\varepsilon(x)$ is symmetric across $\partial_1, \dots, \partial_{d-1}$ since if $(u_1, \dots, u_d) \in \tilde{D}_\varepsilon(x)$, then $(u_1, \dots, -u_i, \dots, u_d) \in \tilde{D}_\varepsilon(x)$ for $i = 1, \dots, d-1$ by definition. Clearly, the volume of $\tilde{D}_\varepsilon(x)$ is an approximation of that of $D_\varepsilon(x)$ up to the third order error term. See Corollary 25 for details. For $x \notin M_\varepsilon$ and ε sufficiently small, define the symmetric region associated with $D_\varepsilon(x)$ as

$$\tilde{D}_\varepsilon(x) = \left\{ (u_1, \dots, u_d) \in T_x \tilde{M} \mid \sum_{i=1}^d u_i^2 \leq \varepsilon^2 \right\} \subset T_x \tilde{M}. \quad (3.3)$$

3.3 The augmented vectors and the kernels associated with LLE

Before we define the augmented vectors and the kernels associated with LLE, we recall the definition of the *local covariance matrix*. For $x \in M$, we call

$$C_x := \mathbb{E}[(\iota(X) - \iota(x))(\iota(X) - \iota(x))^\top \chi_{B_\varepsilon^p(\iota(x))}(\iota(X))] \in \mathbb{R}^{p \times p} \quad (3.4)$$

the local covariance matrix at $\iota(x) \in \iota(M)$, which is the covariance matrix considered for the *local principal component analysis (PCA)* (Singer and Wu, 2012; Cheng and Wu, 2013). In this paper, we use the following symbols for the local covariance matrix. For $x \in M$, suppose $\text{rank}(C_x) = r \leq p$. Clearly, r depends on x , but we ignore x for simplicity. Denote the eigen-decomposition of C_x as $C_x = U_x \Lambda_x U_x^\top$, where $U_x \in O(p)$ is composed of eigenvectors and Λ_x is a diagonal matrix with the associated eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_p = 0$. The theoretical property of local PCA on the manifold without boundary has been studied in a sequence of works, like Nadler (2008); Singer and Wu (2012); Tyagi et al. (2013); Cheng and Wu (2013); Kaslovsky and Meyer (2014); Little et al. (2017); Wu and Wu (2018); Dunson and Wu (2021), and the companion property near the boundary will be discussed in Section D.

In this work, the regularizer in (2.3) we have interest in is $c = n\varepsilon^{d+3}$. For the sake of self-containedness, we provide an intuitive explanation for the choice of the regularizer when M has no boundary based on the results in Wu and Wu (2018). Under Assumptions 3.1 and 3.2, let $\mathcal{X} = \{z_i = \iota(x_i)\}_{i=1}^n$. Let $G_{n,k}$ be the local data matrix constructed from \mathcal{X} as defined in (2.2). Then, as $n \rightarrow \infty$, we expect

$$\frac{1}{n} G_{n,k} G_{n,k}^\top \rightarrow C_{x_k},$$

where C_{x_k} is the local covariance matrix at $\iota(x_k)$. In Wu and Wu (2018), the authors show that the first d eigenvalues of C_{x_k} are of order ε^{d+2} , while the rest $p-d$ eigenvalues are bounded by ε^{d+4} when ε is sufficiently small. Those $p-d$ eigenvalues include the extrinsic geometric information, e.g. the second fundamental form of $\iota(M)$. Hence, if ε is chosen properly based on n , we expect that the first d eigenvalues of $G_{n,k} G_{n,k}^\top$ are of order $n\varepsilon^{d+2}$, while the rest $p-d$ eigenvalues containing the extrinsic geometric information are bounded by $n\varepsilon^{d+4}$. Suppose we choose $c = n\varepsilon^{d+3}$ in (2.5). Then last $p-d$ diagonal terms of $\Lambda_n + cI_{p \times p}$ are dominated by c but less than the first d diagonal terms. Hence, we can eliminate the impact of the extrinsic geometry of $\iota(M)$ in (2.5). Since the

Laplace Beltrami operator only depends on the intrinsic geometry of M , we choose $c = n\varepsilon^{d+3}$ to construct the LLE matrix in (2.4) and (2.8) in order to recover the operator. When the boundary is non-empty, since our focus is the boundary effect, we fix this regularizer so that points away from the boundary have a good control.

Definition 6 Define the augmented vector at $x \in M$ as

$$\mathbf{T}(x)^\top = \mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top U_x I_{p,r}(\Lambda_x + \varepsilon^{d+3} I_{p \times p})^{-1} U_x^\top \in \mathbb{R}^p, \quad (3.5)$$

which is a \mathbb{R}^p -valued vector field on M .

The nomination of $\mathbf{T}(x)$ comes from analyzing the kernel associated with LLE. It has been shown in Wu and Wu (2018, Corollary 3.1) that the kernel associated with LLE is not symmetric and is defined as

$$K_\varepsilon(x, y) := \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(y)) - [(\iota(y) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(y))]^\top \mathbf{T}(x). \quad (3.6)$$

We call $\mathbf{T}(x)$ the *augmented vector* since it augments the symmetric 0–1 kernel $K(x, y) = \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(y))$ by the inner product of $\mathbf{T}(x)$ and $[(\iota(y) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(y))]$. Notice that the vector \mathbf{T}_{n,x_k} defined in (2.6) is a discretization of $\mathbf{T}(x)$ and the theoretical justification is provided in Appendix G.

3.4 Empirical estimation of the regularizer

In Wu and Wu (2018), the authors propose to choose the regularizer in LLE as $c = n\varepsilon^{d+3}$ based on the asymptotic analysis. However, in practice, the dimension of the manifold d is unknown. While it is possible to estimate the dimension, it is usually a challenging mission. We thus need a practical way to determine the regularizer without estimating the dimension. In this subsection, we provide empirical estimators of c without estimating the dimension of the underlying manifold given a finite sample $\mathcal{X} = \{z_i = \iota(x_i)\}_{i=1}^n$ satisfying Assumptions 3.1 and 3.2, and the estimator is asymptotically equal to $n\varepsilon^{d+3}$ up to a constant. Suppose $z, z' \in \mathbb{R}^p$. Define $\mathcal{K}_\varepsilon(z, z') = \exp(-\frac{\|z-z'\|_{\mathbb{R}^p}}{\varepsilon})$, where ε is the same bandwidth in the ε radius ball scheme of LLE. We suggest considering the following empirical regularizer

$$\tilde{c} = \varepsilon^3 \text{median} \left(\sum_{j=1}^n \mathcal{K}_\varepsilon(z_i, z_j) \right), \quad (3.7)$$

where the median is evaluated over all $z_i \in \mathcal{X}$. The construction of the above estimator is motivated by the kernel density estimation. Suppose $z = \iota(x)$ for $x \in M$. It is shown in Berry and Sauer (2017) that if $n\varepsilon^d \rightarrow \infty$ and $\varepsilon \rightarrow 0$ as $n \rightarrow \infty$, then $\tilde{c}(z) \rightarrow C(x)P(x)n\varepsilon^{d+3}$, where $C(x)$ depends on d and $d_g(x, \partial M)$ if x is sufficient close to the boundary, and $C(x)$ only depends on d if x is away from the boundary. This estimator is easy to implement and does not require an estimation of d . We shall mention that although we choose the squared exponential kernel for the density estimation, more general kernels can be applied to construct the estimator. We refer the readers to Wu and Wu (2022) for a discussion.

4. Asymptotic analysis of LLE under the manifold with boundary setup

In this section, we provide the asymptotic analysis of LLE under the manifold with boundary setup. With the ε -radius ball scheme, the asymptotic analysis is achieved in the following 4 steps. We also summarize the main results of each step as follows.

Step 1: In subsection 4.1, we study the augmented vector $\mathbf{T}(x)$ for $x \in M$. We show that the vectors $\mathbf{T}(x)$ form a smooth vector field on $\iota(M)$. The geometry of the vector field can be intuitively described as follows. The vector $\mathbf{T}(x)$ almost points toward the normal direction of $\iota(M)$ in the interior region $x \in M \setminus M_\varepsilon$. However, in M_ε , $\mathbf{T}(x)$ leans towards the tangent direction of $\iota(M)$ gradually. It is worth noting that the restriction of the tangent components of $\mathbf{T}(x)$ on $\iota(\partial M)$ forms an inward normal vector field of $\iota(\partial M)$.

Step 2: Since $\mathbf{T}(x)$ is the major ingredient in the definition of the kernel function $K_\varepsilon(x, y)$, where $x, y \in M$, in subsection 4.2, we explore the properties of $K_\varepsilon(x, y)$ by using the properties of $\mathbf{T}(x)$ that we derive in the previous subsection. Obviously, based on the definition, $K_\varepsilon(x, y) = 0$ when $\iota(y) \notin B_\varepsilon^{\mathbb{R}^p}(\iota(x))$. We show that $K_\varepsilon(x, y)$ is approximately equal to 1 when $x \in M \setminus M_\varepsilon$ and $\iota(y) \in B_\varepsilon^{\mathbb{R}^p}(\iota(x))$. However, when $x \in M_\varepsilon$, $K_\varepsilon(x, y)$ is not symmetric and may be negative.

Step 3: In subsection 4.3, we provide the variance analysis. First, we use $K_\varepsilon(x, y)$ to define an integral operator Q_ε on $C(M)$. For $f \in C^2(M)$, let \vec{f} be the discretization of f over $\{x_1, \dots, x_n\} \subset M$. Let W be the LLE matrix. Then, we show that $[(W - I)\vec{f}](k)$ converges to $Q_\varepsilon f(x_k)$ at the rate $O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-1}}\right)$ regardless the x_k in $M \setminus M_\varepsilon$ or M_ε . Hence, the result implies that the pointwise convergence rate of LLE is the same for manifolds with or without boundary.

Step 4: In subsection 4.4, we provide the bias analysis. We define a second order mixed-type differential operator \mathcal{D}_ε . We show that for $f \in C^3(M)$, $Q_\varepsilon f(x)/\varepsilon^2$ can be approximated by $\mathcal{D}_\varepsilon f(x)$ for all $x \in M$. The final result connecting $[(W - I)\vec{f}](k)$ and $\mathcal{D}_\varepsilon f(x_k)$ is achieved by combining the variance and the bias analysis.

4.1 Properties of the augmented vector on manifold with boundary

The main challenge to analyze LLE is dealing with the augmented vector. It involves three main players in the data structure, the p.d.f., the curvature, and the boundary if the boundary is not empty. Clearly, when x is close to the boundary, the term $\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$ in $\mathbf{T}(x)$ includes the geometry of the boundary, and the integration will depend on the p.d.f.. On the other hand, while a manifold can be locally well approximated by an affine space, the curvature appears in the eigenvalues of the local covariance matrix. Hence, the term $(\Lambda_x + \varepsilon^{d+3}I_{p \times p})^{-1}$ in $\mathbf{T}(x)$ involves the curvature. Dealing with these terms requires a careful asymptotic analysis. To alleviate the heavy notation toward this goal, we consider the following functions, and their role will become clear along the theory development.

Definition 7 Suppose ε is sufficiently small. We define the following functions on $[0, \infty)$, where $\frac{|S^{d-2}|}{d-1}$ is defined to be 1 when $d = 1$.

$$\begin{aligned}\sigma_0(t) &:= \begin{cases} \frac{|S^{d-1}|}{2d} + \frac{|S^{d-2}|}{d-1} \int_0^{\frac{t}{\varepsilon}} (1-x^2)^{\frac{d-1}{2}} dx & \text{for } 0 \leq t \leq \varepsilon \\ \frac{|S^{d-1}|}{d} & \text{for } t > \varepsilon \end{cases} \\ \sigma_{1,d}(t) &:= \begin{cases} -\frac{|S^{d-2}|}{d^2-1} (1-(\frac{t}{\varepsilon})^2)^{\frac{d+1}{2}} & \text{for } 0 \leq t \leq \varepsilon \\ 0 & \text{otherwise} \end{cases} \\ \sigma_2(t) &:= \begin{cases} \frac{|S^{d-1}|}{2d(d+2)} + \frac{|S^{d-2}|}{d^2-1} \int_0^{\frac{t}{\varepsilon}} (1-x^2)^{\frac{d+1}{2}} dx & \text{for } 0 \leq t \leq \varepsilon \\ \frac{|S^{d-1}|}{d(d+2)} & \text{otherwise} \end{cases} \\ \sigma_{2,d}(t) &:= \begin{cases} \frac{|S^{d-1}|}{2d(d+2)} + \frac{|S^{d-2}|}{d-1} \int_0^{\frac{t}{\varepsilon}} (1-x^2)^{\frac{d-1}{2}} x^2 dx & \text{for } 0 \leq t \leq \varepsilon \\ \frac{|S^{d-1}|}{d(d+2)} & \text{otherwise} \end{cases} \\ \sigma_3(t) &:= \begin{cases} -\frac{|S^{d-2}|}{(d^2-1)(d+3)} (1-(\frac{t}{\varepsilon})^2)^{\frac{d+3}{2}} & \text{for } 0 \leq t \leq \varepsilon \\ 0 & \text{otherwise} \end{cases} \\ \sigma_{3,d}(t) &:= \begin{cases} -\frac{|S^{d-2}|}{(d^2-1)(d+3)} (2+(d+1)(\frac{t}{\varepsilon})^2)(1-(\frac{t}{\varepsilon})^2)^{\frac{d+1}{2}} & \text{for } 0 \leq t \leq \varepsilon \\ 0 & \text{otherwise} \end{cases}\end{aligned}\tag{4.1}$$

Note that these functions are of order 1 when $t \leq \varepsilon$. These seemingly complicated formulas share a simple geometric picture. If \mathcal{R} is the region between the unit sphere and the hyperspace $x_d = \frac{t}{\varepsilon}$ in \mathbb{R}^d with coordinates $\{x_1, \dots, x_d\}$, where $0 \leq t \leq \varepsilon$, then $\sigma_0(t)$, $\sigma_{1,d}(t)$, $\sigma_2(t)$, $\sigma_{2,d}(t)$, $\sigma_3(t)$ and $\sigma_{3,d}(t)$ are expansions of the integrals of 1, x_d , x_1^2 , x_d^2 , $x_1^2 x_d$ and x_d^3 over \mathcal{R} respectively. All the above functions are differentiable of all orders except when $t = \varepsilon$. The regularity of the functions at $t = \varepsilon$ depends on d . For example $\sigma_0(t)$ is at least C^0 at $t = \varepsilon$ and the other functions are at least C^1 at $t = \varepsilon$.

With these notations, the behavior of $\mathbf{T}(x)$, particularly when x is near the boundary, can be fully described.

Proposition 8 Decompose $\mathbf{T}(x) = \mathbf{T}^{tan}(x) + \mathbf{T}^{per}(x)$, where $\mathbf{T}^{tan}(x)$ is the tangential component of $\mathbf{T}(x)$ and $\mathbf{T}^{per}(x)$ is the normal component of $\mathbf{T}(x)$; that is, $\mathbf{T}^{tan}(x) \in \iota_* T_x M$ and $\mathbf{T}^{per}(x) \in (\iota_* T_x M)^\perp$. If $x \in M_\varepsilon$, then

$$\begin{aligned}\mathbf{T}^{tan}(x) &= \frac{\sigma_{1,d}(\tilde{\varepsilon}_x)}{\sigma_{2,d}(\tilde{\varepsilon}_x)} \frac{1}{\varepsilon} e_d + O(1) \\ \mathbf{T}^{per}(x) &= \frac{P(x)}{2} \left[\left(\sigma_2(\tilde{\varepsilon}_x) - \frac{\sigma_{1,d}(\tilde{\varepsilon}_x)}{\sigma_{2,d}(\tilde{\varepsilon}_x)} \sigma_3(\tilde{\varepsilon}_x) \right) \sum_{j=1}^{d-1} \mathbb{I}_{jj}(x) \right. \\ &\quad \left. + \left(\sigma_{2,d}(\tilde{\varepsilon}_x) - \frac{\sigma_{1,d}(\tilde{\varepsilon}_x)}{\sigma_{2,d}(\tilde{\varepsilon}_x)} \sigma_{3,d}(\tilde{\varepsilon}_x) \right) \mathbb{I}_{dd}(x) \right] \frac{1}{\varepsilon} + O(1).\end{aligned}$$

If $x \in M \setminus M_\varepsilon$, then

$$\begin{aligned}\mathbf{T}^{tan}(x) &= J_{p,d} \frac{\nabla P(x)}{P(x)} + O(\varepsilon) \\ \mathbf{T}^{per}(x) &= \frac{P(x)}{2} \left[\frac{|S^{d-1}|}{d(d+2)} \sum_{j=1}^d \mathbb{I}_{jj}(x) \right] \frac{1}{\varepsilon} + O(1).\end{aligned}$$

The proof is postponed to Appendix E. The above proposition says that when $x \in M_\varepsilon$, both the tangent and normal components of $\mathbf{T}(x)$ are of order $\frac{1}{\varepsilon}$, and the normal component depends on the extrinsic curvature of the manifold at $\iota(x)$. An interesting observation is that a construction of the *inward* normal vector field on $\iota(\partial M)$ is naturally encoded in the LLE algorithm. In particular, $\mathbf{T}^{tan}(x)$ on $\iota(\partial M)$ forms an inward normal vector field on $\iota(\partial M)$ with an order $O(1)$ perturbation. Moreover, the magnitude of $\mathbf{T}^{tan}(x)$ on M_ε only depends on the distance from x to the boundary and it is independent of the p.d.f.. On the other hand, when $x \in M \setminus M_\varepsilon$, $\mathbf{T}(x)$ is of order $\frac{1}{\varepsilon}$ in the normal direction of $M \setminus M_\varepsilon$ with an order $O(1)$ perturbation in the tangential direction. With the theorem developed in Wu and Wu (2018) for the augmented vector field away from the boundary, we have the full knowledge of the augmented vector field.

At last, in Figure 2, we provide a visualization of the augmented vector field in a 2-dim manifold parametrized by $(x, y, x^2 - y^3)$, where $x^2 + y^2 \leq 1$. We sample the manifold in the following way. First, uniformly sample 20,000 points independently on $[-1, 1] \times [-1, 1]$, and keep points with norm less and equal to 1. The i -th point is then constructed by the parametrization. Clearly, the sampling is not uniform. The LLE matrix is constructed with the ε -radius ball nearest neighbor search scheme with $\varepsilon = 0.2$.

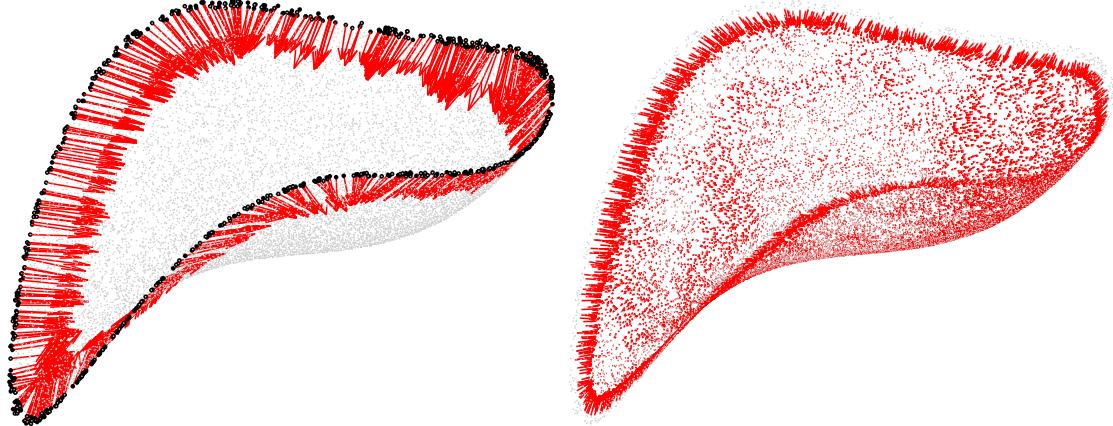


Figure 2: The \mathbf{T} vector field. The sampled point cloud is plotted in gray. Left: the black points indicates points satisfies $0.98 \leq x^2 + y^2 \leq 1$, and the \mathbf{T} on those points are marked in red. Right: the \mathbf{T} on points with $x^2 + y^2 < 0.98$ are marked in red.

Remark 9 Another work that constructs a normal vector field on the boundary of a manifold is Berry and Sauer (2017). Inspired by the kernel density estimation, the authors propose an inward normal vector field by using $\vec{F}(\iota(x)) = \mathbb{E}[(\iota(X) - \iota(x))K(\frac{\|\iota(X) - \iota(x)\|_{\mathbb{R}^p}}{h})] \in \mathbb{R}^p$, where $K : [0, \infty) \rightarrow$

$[0, \infty)$ has an exponential decay and h is the bandwidth. Since the construction originates from the kernel density estimation, when $x \in \partial M$, the magnitude of the vector field depends on the p.d.f.

4.2 Properties of the kernel on manifold with boundary

With the above knowledge of the augmented vector field near the boundary, the behavior of the kernel near the boundary can be well quantified.

Proposition 10 *Let $K_\varepsilon(x, y)$ be the kernel function defined in (3.6). Fix $x \in M$, we summarized the properties of $K_\varepsilon(x, y)$ as follows.*

1. Suppose $x \notin M_\varepsilon$. When $\iota(y) \in B_\varepsilon^{\mathbb{R}^P}(\iota(x))$, $K_\varepsilon(x, y) = 1 - O(\varepsilon)$. Otherwise $K_\varepsilon(x, y) = 0$. Hence, $K_\varepsilon(x, y) \geq 0$, when ε is sufficiently small. The implied constant in $O(\varepsilon)$ depends on the minimum and C^1 norm of P and the maximum of second fundamental form of the manifold.
2. If $x \in M_\varepsilon$ and $\iota(y) \in B_\varepsilon^{\mathbb{R}^P}(\iota(x))$, when ε is sufficiently small,

$$K_\varepsilon(x, y) = 1 - \frac{\sigma_{1,d}(\tilde{\varepsilon}_x)u_d}{\sigma_{2,d}(\tilde{\varepsilon}_x)\varepsilon} + O(\varepsilon), \quad (4.2)$$

where the coordinate u_d of y is defined in Definition 5. The implied constant in $O(\varepsilon)$ depends on the minimum and C^1 norm of P and the maximum of second fundamental form of the manifold. Otherwise $K_\varepsilon(x, y) = 0$. Hence, we have

$$\inf_{x,y} K_\varepsilon(x, y) = 1 - \frac{|S^{d-2}|}{d-1} \frac{2d(d+2)}{(d+1)|S^{d-1}|} + O(\varepsilon) < 0$$

when ε is sufficiently small, where $\frac{|S^{d-2}|}{d-1}$ is defined to be 1 when $d = 1$.

3. For any $x \in M$, we have

$$\mathbb{E}K_\varepsilon(x, X) = C(x)\varepsilon^d + O(\varepsilon^{d+1}), \quad (4.3)$$

where $C(x) > C > 0$, and C is a constant depending only on d and P . Hence, $\mathbb{E}K_\varepsilon(x, X) > 0$ for all $x \in M$ when ε is sufficiently small. The implied constant in $O(\varepsilon^{d+1})$ depends on the C^1 norm of P and the maximum of second fundamental form of the manifold.

This proposition provides several facts about LLE. First, the assumption of Proposition 2 is satisfied when the manifold is boundary free, since the higher order error terms depend on various curvatures of M and M is smooth and compact. So, the eigenvalues of the LLE matrix in the *boundary-free* manifold setup have well controlled imaginary parts. However, when the boundary is not empty, we may lose this control. Second, the kernel function behaves differently when x is near the boundary and away from the boundary. When x is away from the boundary, the kernel is non-negative. However, when x is close to the boundary, then it is possible that $K_\varepsilon(x, y)$ is negative. In particular, when $x \in \partial M$, $\iota(y) \in B_\varepsilon^{\mathbb{R}^P}(\iota(x))$, the geodesic distance between x and y is $\varepsilon + O(\varepsilon^2)$ and the minimizing geodesic between x and y is perpendicular to ∂M , then $K_\varepsilon(x, y) = 1 - \frac{|S^{d-2}|}{d-1} \frac{2d(d+2)}{(d+1)|S^{d-1}|} + O(\varepsilon) < 0$. Although it is possible that $K_\varepsilon(x, y)$ is negative, $\mathbb{E}K_\varepsilon(x, X)$ is always positive if ε is small enough. See Figure 3 for an illustration of the kernel associated with LLE, where the manifold, the sampling scheme, and the LLE matrix are the same as that in Figure 2, expect $\varepsilon = 0.1$.

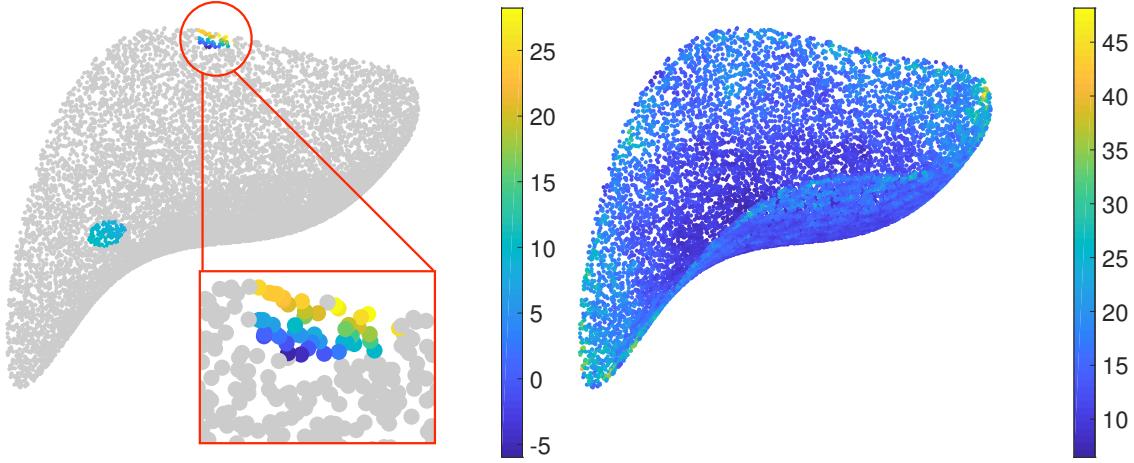


Figure 3: The kernel function associated with LLE. The sampled point cloud is plotted in gray. Left: the kernel function K_ϵ , where $\epsilon = 0.1$, on two points, one is close to the boundary (indicated by the red circle, with the zoomed in enhanced visualization), and one is away from the boundary. It is clear that the kernel close to the boundary changes sign, while the kernel away from the boundary is positive. Right: the $EK_\epsilon(x, X)$. It is clear that the expectations of the kernel at all points are positive.

4.3 Variance analysis of LLE on manifold with boundary

Define the integral operator from $C(M)$ to $C(M)$:

$$Q_\epsilon f(x) := \frac{\mathbb{E}[K_\epsilon(x, X)f(X)]}{\mathbb{E}K_\epsilon(x, X)} - f(x), \quad (4.4)$$

where $f \in C(M)$. We now show that when the boundary is not empty, the LLE matrix W converges to the integral operator Q_ϵ when $n \rightarrow \infty$. The proof of the theorem is postponed to Appendix G.

Theorem 11 (Variance analysis) Suppose $f \in C^2(M)$. Suppose $\epsilon = \epsilon(n)$ so that $\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2+1}} \rightarrow 0$ and $\epsilon \rightarrow 0$ as $n \rightarrow \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \dots, n$,

$$\sum_{j=1}^{N_k} y_k(j) = \frac{\mathbb{E}K_\epsilon(x_k, X)}{\epsilon^{d+3}} + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2+3}}\right) \quad (4.5)$$

$$\sum_{j=1}^n [W - I_{n \times n}]_{kj} f(x_j) = Q_\epsilon f(x_k) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-1}}\right), \quad (4.6)$$

where y_k is defined in (2.3). The implied constants in the error terms depend on C^2 norm of f , C^1 norm of P and the L^∞ norm of $\max_{i,j=1,\dots,d} \|\Pi_{ij}(x)\|$.

Note that the order of the variance does not depend on the location of x_j . By combining (4.5) and (4.3) in Proposition 10, we know that if n is sufficiently large, the sum of all components of y_k is positive. This result restates the fact that w_k defined in (2.8) does not blow up.

4.4 Bias analysis of LLE on manifold with boundary and the main result

In this subsection, we study the integral operator Q_ε by relating it to the following differential operator.

Definition 12 Fix $\varepsilon > 0$. Define a differential operator on $C^2(M)$ as

$$\mathcal{D}_\varepsilon f(x) = \phi_1(\tilde{\varepsilon}_x) \sum_{i=1}^{d-1} \partial_{ii}^2 f(x) + \phi_2(\tilde{\varepsilon}_x) \partial_{dd}^2 f(x) + V(x) \partial_d f(x), \quad (4.7)$$

where ϕ_1 and ϕ_2 are functions defined on $[0, \infty)$ by

$$\phi_1(t) = \frac{1}{2} \frac{\sigma_{2,d}(t)\sigma_2(t) - \sigma_3(t)\sigma_{1,d}(t)}{\sigma_{2,d}(t)\sigma_0(t) - \sigma_{1,d}^2(t)}, \quad (4.8)$$

$$\phi_2(t) = \frac{1}{2} \frac{\sigma_{2,d}^2(t) - \sigma_{3,d}(t)\sigma_{1,d}(t)}{\sigma_{2,d}(t)\sigma_0(t) - \sigma_{1,d}^2(t)}, \quad (4.9)$$

and V is a function on M defined by

$$V(x) = \frac{\sigma_{1,d}(\tilde{\varepsilon}_x)}{P(x)(\sigma_{2,d}(\tilde{\varepsilon}_x)\sigma_0(\tilde{\varepsilon}_x) - \sigma_{1,d}^2(\tilde{\varepsilon}_x))}. \quad (4.10)$$

Next, we take a closer look at coefficients of \mathcal{D}_ε .

Proposition 13 Fix $\varepsilon > 0$. We have the following properties of coefficients of the differential operator \mathcal{D}_ε .

1. $\phi_1(t) > 0$. When $t \geq \varepsilon$,

$$\phi_1(t) = \frac{1}{2(d+2)}. \quad (4.11)$$

Moreover, $\phi_1(t)$ is differentiable of all orders at all $t > 0$ except at $t = \varepsilon$, where it is at least first order differentiable.

2. $\phi_2(0) < 0$. If $t \geq \varepsilon$, then

$$\phi_2(t) = \frac{1}{2(d+2)}. \quad (4.12)$$

Moreover, $\phi_2(t)$ is differentiable of all orders at all $t > 0$ except at $t = \varepsilon$, where it is at least first order differentiable. Hence, there is a set $\mathcal{S} \subset M_\varepsilon$ diffeomorphic to ∂M and $\phi_2(\tilde{\varepsilon}_x)$ vanishes on \mathcal{S} . Denote the geodesic distance from $x \in \mathcal{S}$ to ∂M as $t^*(x)$. $t^*(x)$ depends only on ε and d . In fact, $\delta_1\varepsilon < t^*(x) < \delta_2\varepsilon$, where

$$\delta_1 = \left(1 - \left[\frac{1 + \frac{(d^2-1)|S^{d-1}|}{2d(d+2)|S^{d-2}|}}{1 + \sqrt{\frac{2}{d+3}}} \right]^{\frac{2}{d+1}} \right)^{\frac{1}{2}}, \quad (4.13)$$

$$\delta_2 = \left(1 - \left[\frac{(d^2-1)|S^{d-1}|}{4d(d+2)|S^{d-2}|} + \frac{1}{d+3} \right]^{\frac{2}{d+1}} \right)^{\frac{1}{2}} < 1 \quad (4.14)$$

and $\delta_2 \rightarrow 0$ as $d \rightarrow \infty$.

3. $V(x) \leq 0$. Moreover, $V(x) = O(\varepsilon^2)$ is differentiable of all orders at all x except when $\tilde{\varepsilon}_x = \varepsilon$, where it is at least differentiable of the first order. If $x \in M_\varepsilon$ satisfies $\tilde{\varepsilon}_x \geq \varepsilon$, in other words, $x \in M \setminus M_\varepsilon$, then $V(x) = 0$. In particular, if $P(x)$ is constant, then $V(x)$ is an increasing function of $\tilde{\varepsilon}_x$.

Denote M_w to be the interior subset of the region between \mathcal{S} and ∂M on M . Denote M_e to be the interior subset of $M \setminus M_w$ on M . Clearly, M_w is a strict subset of M_ε . According to Proposition 13, \mathcal{D}_ε is of hyperbolic type over M_w , of elliptic type over M_e , and degenerate over \mathcal{S} . We thus call M_w the *wave region*, M_e the *elliptic region*, and \mathcal{S} the degenerate region. We conclude that the operator \mathcal{D}_ε is a mixed-type differential operator with degeneracy.

Remark 14 In fact, t^* is the solution of the following nonlinear equation of t :

$$\begin{aligned} & \left(\frac{|S^{d-1}|}{2d(d+2)} + \frac{|S^{d-2}|}{d-1} \int_0^{\frac{t}{\varepsilon}} (1-x^2)^{\frac{d-1}{2}} x^2 dx \right)^2 \\ &= \frac{|S^{d-2}|^2}{(d^2-1)^2(d+3)} \left[2 + (d+1) \left(\frac{t}{\varepsilon} \right)^2 \right] \left[1 - \left(\frac{t}{\varepsilon} \right)^2 \right]^{d+1}, \end{aligned} \quad (4.15)$$

where $t > 0$

We have the following theorem describing how Q_ε is related to \mathcal{D}_ε when ε is sufficiently small. The proof is postponed to Appendix F.

Theorem 15 (Bias analysis) Suppose $f \in C^3(M)$ and $P \in C^2(M)$. We have

$$Q_\varepsilon f(x) = \mathcal{D}_\varepsilon f(x) \varepsilon^2 + O(\varepsilon^3). \quad (4.16)$$

By combining the bias and variance analyses, we have the following pointwise convergence result.

Theorem 16 Suppose $f \in C^3(M)$ and $P \in C^2(M)$. Suppose $\varepsilon = \varepsilon(n)$ so that $\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2+1}} \rightarrow 0$ and $\varepsilon \rightarrow 0$ as $n \rightarrow \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \dots, n$,

$$\sum_{j=1}^n [W - I_{n \times n}]_{kj} f(x_j) = \mathcal{D}_\varepsilon f(x) \varepsilon^2 + O(\varepsilon^3) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-1}}\right), \quad (4.17)$$

where the implied constants in the error terms depend on the C^2 norm of f , the C^1 norm of P , and the L^∞ norm of $\max_{i,j=1,\dots,d} \|\mathbb{I}_{ij}(x)\|$.

Note that when M is a manifold without boundary, then $\mathcal{D}_\varepsilon = \frac{1}{2(d+2)}\Delta$. Hence, the above theorem is consistent with the result in Wu and Wu (2018) when M has no boundary. However, the regularizer in Wu and Wu (2018) is chosen as $c = n\varepsilon^{d+\rho}$, where ρ may be different than 3. The main result there is presented in different cases capturing the interaction between ρ and different local covariance matrix structures. In contrast, the above theorem is much simpler as we only focus on the case when $\rho = 3$.

4.5 Relationship between $W - I$ and $(W - I)^\top(W - I)$

In this paper and Wu and Wu (2018), we focus on studying the asymptotic behavior of $W - I$. However, in the original LLE algorithm, it is the eigenvector of the matrix $(W - I)^\top(W - I)$ that is used to reduce the dimension of \mathcal{X} . We shall clarify the relationship between $W - I$ and $(W - I)^\top(W - I)$ from two aspects – spectral geometry and linear algebra.

Recall that when a manifold is compact without boundary, based on the spectral embedding results (Bérard et al., 1994; Bates, 2014; Portegies, 2016), the eigenfunctions of $-\Delta$ can be applied to construct an embedding of the manifold in a Euclidean space, and the embedding is almost isometric if both eigenfunctions and eigenvalues are properly used. Therefore, if we could obtain eigenpairs of Δ from the database \mathcal{X} sampled from a manifold, we could recover the manifold via this spectral embedding. However, it is not clear what the eigenvectors of $(W - I)^\top(W - I)$ or $W - I$ mean directly from the algorithm. An immediate approach to answering this question is via the pointwise convergence. In Wu and Wu (2018), if M is a compact manifold without boundary, $W - I$ pointwisely converges to the operator $\frac{\varepsilon^2}{2(d+2)}\Delta$. In Theorem 16 of the current work, we show that the same result holds over the region $M \setminus M_\varepsilon$ when $\partial M \neq \emptyset$, and near the boundary the asymptotic behavior is complicated with degeneracy. Hence, one may guess that $(W - I)^\top(W - I)$ will pointwisely converge to $\frac{\varepsilon^4}{4(d+2)^2}\Delta^2$ over the data points in the region $M \setminus M_\varepsilon$. However, it is in general *not* true, particularly when the sampling is nonuniform. It is because the nonsymmetry of $W - I$ plays an important role in eliminating the impact of nonuniform distribution of the point cloud on M (Wu and Wu, 2018), and we lose such property if we consider $(W - I)^\top(W - I)$. Consider the following analysis of $(W - I)^\top(W - I)$ in a simple manifold for an illustration. In contrast to the operator Q_ε , we define a new integral operator from $C(M)$ to $C(M)$:

$$\bar{Q}_\varepsilon f(x) := \frac{\mathbb{E}[K_\varepsilon(X, x)f(X)]}{\mathbb{E}K_\varepsilon(X, x)} - f(x), \quad (4.18)$$

where $f \in C(M)$. By the law of large number, we would have

$$\sum_{j=1}^n [W - I_{n \times n}]_{kj}^\top f(x_j) \rightarrow \bar{Q}_\varepsilon f(x_k)$$

and

$$\sum_{i=1}^n [W - I_{n \times n}]_{ki}^\top \sum_{j=1}^n [W - I_{n \times n}]_{ij} f(x_j) \rightarrow (\bar{Q}_\varepsilon(Q_\varepsilon f))(x_k)$$

as $n \rightarrow \infty$. Now, suppose $\iota(M) = [-1, 1] \subset \mathbb{R}$ and let $\mathcal{X} = \{\iota(x_i)\}_{i=1}^n$, where $\{x_1, x_2, \dots, x_n\}$ are i.i.d. sampled following a p.d.f. $P \in C^2(M)$. Then, for any ε small enough, suppose $\iota(x_k) \in [-1 + \varepsilon, 1 - \varepsilon]$, for any $f \in C^5(M)$, we have

$$(\bar{Q}_\varepsilon(Q_\varepsilon f))(x_k) = \frac{1}{9} \left(\frac{1}{4} f'''(x_k) P'(x_k) + O(\varepsilon) \right) + O(\varepsilon), \quad (4.19)$$

where $\vec{f} = [f(x_1), \dots, f(x_n)]^\top$. The detailed calculation of (4.19) is postponed to Appendix H. In this result, although the manifold is flat and there is no extrinsic geometric information involved, asymptotically $(W - I)^\top(W - I)$ involves not only the desired Δ^2 but also the sampling distribution. This says that even if we do not consider the boundary, the asymptotic differential operator is more

complicated than the bi-Laplacian Δ^2 . A similar result can be derived when M is a general manifold without boundary, but we omit details here.

Based on the above discussion, we could reasonably conjecture that the eigenvectors of $W - I$ approximate the eigenfunctions of Δ when $\partial M = \emptyset$ and the eigenfunctions of more complicated second order differential operator with degeneracy when $\partial M \neq \emptyset$, and the eigenvectors of $(W - I)^\top(W - I)$ approximate the eigenfunctions of more complicated fourth order differential operator. To prove these conjectures, we need to establish the spectral convergence results, which is out of the scope of this paper. Note that previous work on spectral convergence of graph Laplacian (Dunson et al., 2021; Calder et al., 2022; Wormell and Reich, 2021) mainly focuses on symmetric kernel matrices, except the work discussing the kNN kernel construction Calder et al. (2022). However, these kernels are a priori assigned, so their approaches cannot be directly applied to study $W - I$. Specifically, the LLE matrix is not only nonsymmetric but also determined by the dataset. We need different analysis tools to establish the spectral convergence. On the other hand, a complete understanding of the original LLE is certainly via understanding $(W - I)^\top(W - I)$. We shall mention that even for bi-Laplacian to which $(W - I)^\top(W - I)$ pointwisely converges under special conditions, it is still challenging to derive the spectral convergence. Note that it is still an active research field to study bi-Laplacian and its spectral behavior (Chang et al., 1999; Cuccu and Porru, 2009). To sum up, with the help of pointwise convergence results, we could conjecture the spectral behavior of $W - I$ and $(W - I)^\top(W - I)$, and their behaviors are different in general.

From the linear algebra perspective, there are several interesting relationships between $(W - I)^\top(W - I)$ and $W - I$. First, $W - I$ is a sparse matrix and in general sparser than $(W - I)^\top(W - I)$. Since $(W - I)^\top(W - I)$ is symmetric, its eigendecomposition always exists, while $W - I$ is not always diagonalizable. Second, eigenvalues of $(W - I)^\top(W - I)$ are the square of the singular values of $W - I$ and the eigenvectors of $(W - I)^\top(W - I)$ are the same as the right singular vectors of the $W - I$. While $W - I$ is in general not diagonalizable, based on Proposition 10 under the manifold setup, the conditions in Proposition 2 are satisfied, which says that when the sample size of the data is large enough, $W - I$ is close to the symmetric matrix $\frac{W + W^\top}{2} - I$ and the imaginary parts of the eigenvalues of $W - I$ are small. Note that even if $W - I$ is diagonalizable, the right singular vectors of $W - I$ are different from the right eigenvectors of $W - I$. It echos what we discuss above — under the boundary-free manifold setup, we conjecture that the eigenvectors of $W - I$ approximate the eigenfunctions of Δ , while the eigenvectors of $(W - I)^\top(W - I)$, and hence the singular vectors of $W - I$, approximate the eigenfunctions of a fourth order differential operator that involves the nonuniform sampling information. Third, we shall mention that numerically we consistently found that under the manifold setup, when n is sufficiently large, the leading eigenvectors of $W - I$ recover the corresponding eigenfunctions of Δ . Note that a theoretical justification of this numerical finding is part of the spectral convergence conjecture listed above. Therefore, if we consider $W - I$ for the dimension reduction purpose, we propose to use the real parts of the top eigenvectors of $W - I$ corresponding to the leading eigenvalues listed in the decreasing order of their real parts to define the embedding. We provide a numerical comparison of the embedding by $W - I$ and $(W - I)^\top(W - I)$ in Figure 4, where we nonuniformly sample 3513 points from the unit disk on \mathbb{R}^2 (shown on the top left subfigure superimposed with the radius as the color), construct W with the $\varepsilon = 0.04$ radius ball and the regularizer $n\varepsilon^5$, and embed the data using the top 2 non-trivial eigenvectors of $W - I$ (shown on the top middle subfigure) and $(W - I)^\top(W - I)$ (shown on the top right subfigure). Note that in the top middle and right subfigures, the original radius of each point is superimposed as the color, which indicates how the embedding behaves. The top three nontrivial eigenvectors of $W - I$ are

shown in the middle panel, and the top three nontrivial eigenvectors of $(W - I)^\top(W - I)$ are shown in the bottom panel. In this example, we see that the embeddings by $W - I$ and $(W - I)^\top(W - I)$ are different. Based on the analysis in the current work, it is not surprising that the embedding by $W - I$ is impacted by the boundary, but it is interesting to see that $(W - I)^\top(W - I)$ is less impacted by the boundary. The nonuniform sampling also plays a role here. While we do not show it here, we found that when the sampling from the unit disk is uniform, the embeddings by $W - I$ and $(W - I)^\top(W - I)$ are similar, which suggests the interaction between the sampling and boundary. The above interesting findings warrant further study of the behavior of $(W - I)^\top(W - I)$ from various aspects to fully understand how LLE functions.

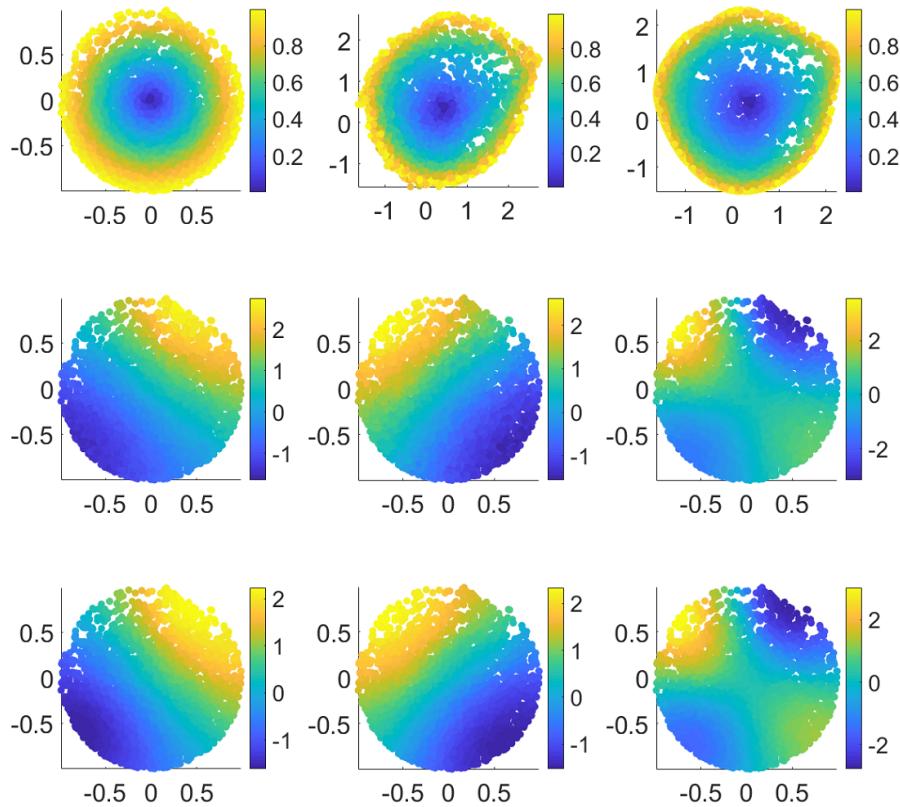


Figure 4: Top left subfigure: the original dataset uniformly sampled from the unit disk superimposed with the radius as the color. Top middle (right respectively) subfigure: the embedding by the top two nontrivial eigenvectors of $W - I$ ($(W - I)^\top(W - I)$ respectively), where the radius of the original radius of each point is superimposed as the color. The top three nontrivial eigenvectors of $W - I$ are shown in the middle panel, and the top three nontrivial eigenvectors of $(W - I)^\top(W - I)$ are shown in the bottom panel.

5. Exploration of LLE on 1-dim manifold with boundary

In this section, we further explore the LLE matrix W . In Theorem 16, we show that the matrix $W - I$ converges pointwisely to the differential operator \mathcal{D}_ε . Then, we illustrate how the differential operator \mathcal{D}_ε looks like in the 1 dimensional case.

Corollary 17 *Let M be a regular smooth curve in \mathbb{R}^p . Let $\gamma(t) : [0, a] \rightarrow \mathbb{R}^p$ be the arclength parametrization. Let $P(t)$ be the probability density function. Then, we have, for $f \in C^2(M)$,*

$$\mathcal{D}_\varepsilon f(t) = \begin{cases} -\frac{1}{12}(1 - 4(\frac{t}{\varepsilon}) + (\frac{t}{\varepsilon})^2)f''(t) + \frac{6\varepsilon^2(\varepsilon-t)}{P(t)(\varepsilon+t)^3}f'(t) & \text{if } t \in [0, \varepsilon]; \\ \frac{1}{6}f''(t) & \text{if } t \in [\varepsilon, a-\varepsilon]; \\ -\frac{1}{12}(1 - 4(\frac{a-t}{\varepsilon}) + (\frac{a-t}{\varepsilon})^2)f''(t) - \frac{6\varepsilon^2(\varepsilon+t-a)}{P(t)(\varepsilon+a-t)^3}f'(t) & \text{if } t \in [a-\varepsilon, a]. \end{cases}$$

Specifically, $\mathcal{D}_\varepsilon f(t)$ degenerates to $\frac{2\sqrt{3}}{P(t)}f'(t)$ at $t = (2 - \sqrt{3})\varepsilon$ and $t = a - (2 - \sqrt{3})\varepsilon$.

This corollary comes from a direct expansion of the formula in Definition 12. Note that e_d is in the outward normal direction by definition. Therefore, $\partial_d f(t) = -f'(t)$, when $t \in [0, \varepsilon]$. And $\partial_d f(t) = f'(t)$, when $t \in [a - \varepsilon, a]$. To study the spectral property of \mathcal{D}_ε , it is natural to consider converting \mathcal{D}_ε into the Sturm-Liouville (SL) form by the integrating factor. However, due to the degeneracy of a_ε , several technical details need to be taken care of. Here we provide a summary of known facts about the SL theory (Naïmark, 1967, Chapter V).

Fix $a > 0$. The SL problem on $(0, a)$ is finding a complex function $f(x)$ defined on $(0, a)$ that solves

$$-(p(x)f(x))' + q(x)f(x) = \lambda w(x)f(x), \quad (5.1)$$

where $p(x)$, $q(x)$ and $w(x)$ are measurable real functions on (a, b) and λ is a complex number. The SL problem is called *regular* if $1/p$, q , and w are all functions in $L^1(a, b)$; otherwise, it is called *singular*. A complex function $f(x)$ is a *solution* of the SL problem (5.1), if $f^{[0]}(x)$ and $f^{[1]}(x)$ exist, where $f^{[0]}(x) := f(x)$ (respectively $f^{[1]}(x) := p(x)f'(x)$) is the zero (respectively first) order quasi-derivative of f , and are absolutely continuous on any compact subinterval of (a, b) . It is worth noting that in general $f'(x)$ may not be absolutely continuous.

It is stated in (Naïmark, 1967, Chapter V) that for $x_0 \in (0, a)$ and complex numbers c_0 and c_1 , there exists a unique solution to the regular SL problem with $f^{[0]}(x_0) = c_0$ and $f^{[1]}(x_0) = c_1$. As an eigenvalue problem, by (Atkinson, 1964), given boundary conditions $A_1 f^{[0]}(0) + A_2 f^{[1]}(0) = 0$ and $B_1 f^{[0]}(a) + B_2 f^{[1]}(a) = 0$ with $A_1^2 + A_2^2 > 0$ and $B_1^2 + B_2^2 > 0$, if the SL problem is regular and $p > 0$ and $w > 0$ on $(0, a)$, then the eigenvalues are discrete and bounded from below; that is, we have eigenvalues $-\infty < \lambda_0 < \lambda_1 < \lambda_2 < \dots$ so that $\lambda_n \rightarrow \infty$ as $n \rightarrow \infty$. Moreover, if f_n is the corresponding eigenvalues of λ_n , then f_n has exactly n zeros in $(0, a)$.

Now we come back to the challenge. Suppose $P(t) = 1/a$; that is, the sampling is uniform. By Corollary 17, when ε is sufficiently small, the second order ordinary differential equation

$$\mathcal{D}_\varepsilon f(t) = A_\varepsilon(t)f''(t) + B_\varepsilon(t)f'(t) \quad (5.2)$$

dominates. Note that $A_\varepsilon(t) > 0$ on the elliptic region $((2 - \sqrt{3})\varepsilon, a - (2 - \sqrt{3})\varepsilon)$, and $A_\varepsilon(t) < 0$ on the wave region $[0, (2 - \sqrt{3})\varepsilon) \cup (a - (2 - \sqrt{3})\varepsilon, a]$, while $B_\varepsilon(t) \geq 0$ on $[0, a]$.

To convert \mathcal{D}_ε into the SL form, we define two more functions over $t \in [0, \varepsilon]$. First,

$$g(t) := |t - (2 - \sqrt{3})\varepsilon|^{(4+2\sqrt{3})a\varepsilon} |t - (2 + \sqrt{3})\varepsilon|^{(4-2\sqrt{3})a\varepsilon} (t + \varepsilon)^{8a\varepsilon} e^{[\frac{12ae^3}{(\varepsilon+t)^2} + \frac{12ae^2}{\varepsilon+t}]},$$

over $t \in [0, \varepsilon]$. Clearly, $g(t) > 0$. Moreover, g is continuous and is smooth except at $t = (2 - \sqrt{3})\varepsilon$. Second,

$$h(t) := |t - (2 - \sqrt{3})\varepsilon|^{(4+2\sqrt{3})a\varepsilon-1} |t - (2 + \sqrt{3})\varepsilon|^{(4-2\sqrt{3})a\varepsilon-1} (t + \varepsilon)^{8a\varepsilon} e^{[\frac{12a\varepsilon^3}{(\varepsilon+t)^2} + \frac{12a\varepsilon^2}{\varepsilon+t}]},$$

over $t \in [0, \varepsilon]$. By a direct check, we know that $h(t) > 0$ on $[0, (2 - \sqrt{3})\varepsilon] \cup ((2 - \sqrt{3})\varepsilon, \varepsilon]$ and $h(t) \rightarrow \infty$ when $t \rightarrow (2 - \sqrt{3})\varepsilon$ since $(4 + 2\sqrt{3})a\varepsilon - 1 < 0$. With g and h , define

$$p(t) := \begin{cases} -g(t) & \text{if } t \in [0, (2 - \sqrt{3})\varepsilon]; \\ g(t) & \text{if } t \in [(2 - \sqrt{3})\varepsilon, \varepsilon]; \\ g(\varepsilon) & \text{if } t \in [\varepsilon, a - \varepsilon]; \\ g(a - t) & \text{if } t \in [a - \varepsilon, a - (2 - \sqrt{3})\varepsilon]; \\ -g(a - t) & \text{if } t \in [a - (2 - \sqrt{3})\varepsilon, a] \end{cases} \quad (5.3)$$

on $[0, a]$ and

$$w(t) := \begin{cases} h(t) & \text{if } t \in [0, \varepsilon]; \\ h(\varepsilon) & \text{if } t \in [\varepsilon, a - \varepsilon]; \\ h(a - t) & \text{if } t \in [a - \varepsilon, a]. \end{cases} \quad (5.4)$$

on $[0, a]$. We have the following proposition summarizing the behavior of p and w .

Proposition 18 Suppose ε is sufficiently small. The defined function p satisfies the following properties.

1. $p(t) > 0$ on the elliptic region $((2 - \sqrt{3})\varepsilon, a - (2 - \sqrt{3})\varepsilon)$, $p(t) < 0$ on the wave region $[0, (2 - \sqrt{3})\varepsilon] \cup (a - (2 - \sqrt{3})\varepsilon, a]$, and $p(t) = 0$ when $t = (2 - \sqrt{3})\varepsilon$ or $t = a - (2 - \sqrt{3})\varepsilon$.
2. $p(t)$ is C^1 on $[0, a]$ except at $t = (2 - \sqrt{3})\varepsilon$ and $t = a - (2 - \sqrt{3})\varepsilon$. In particular, $p'(t) \rightarrow \infty$ as $t \rightarrow (2 - \sqrt{3})\varepsilon$ from right or $t \rightarrow a - (2 - \sqrt{3})\varepsilon$ from left; $p'(t) \rightarrow -\infty$ as $t \rightarrow (2 - \sqrt{3})\varepsilon$ from left or $t \rightarrow a - (2 - \sqrt{3})\varepsilon$ from right.
3. $p(t)$ is absolutely continuous.
4. $1/p \in L^1$ on $[0, a]$.

The defined function w satisfies the following properties.

1. $w(t)$ is C^1 on $[0, a]$ except at $t = (2 - \sqrt{3})\varepsilon$ and $t = a - (2 - \sqrt{3})\varepsilon$.
2. $w(t) \rightarrow \infty$ as $t \rightarrow (2 - \sqrt{3})\varepsilon$ or $t \rightarrow a - (2 - \sqrt{3})\varepsilon$.
3. $w \in L^1$ on $[0, a]$.

The defined functions p and w are related to A_ε and B_ε and satisfy the following properties.

1. $\frac{p(t)}{w(t)} = A_\varepsilon(t)$ and $\frac{p'(t)}{w(t)} = B_\varepsilon(t)$, when $t \neq (2 - \sqrt{3})\varepsilon$ and $t \neq a - (2 - \sqrt{3})\varepsilon$.
2. $\frac{p(t)}{w(t)} \rightarrow A_\varepsilon(t)$ and $\frac{p'(t)}{w(t)} \rightarrow B_\varepsilon(t)$ when $t \rightarrow (2 - \sqrt{3})\varepsilon$ or $t \rightarrow a - (2 - \sqrt{3})\varepsilon$.

With this proposition, we conclude that except at $t = (2 - \sqrt{3})\varepsilon$ and $t = a - (2 - \sqrt{3})\varepsilon$, the following relationship holds:

$$A_\varepsilon(t)f''(t) + B_\varepsilon(t)f'(t) = \frac{p(t)}{w(t)}f''(t) + \frac{p'(t)}{w(t)}f'(t) = \frac{(p(t)f'(t))'}{w(t)}. \quad (5.5)$$

Also, the corresponding eigenvalue problem

$$-(p(t)f'(t))' = \lambda w(t)f(t), \quad (5.6)$$

is regular when ε is small enough. If f is a solution to the above problem, then f is absolutely continuous, and hence it is differentiable almost everywhere. Moreover, $p(t)f'(t)$ is absolutely continuous and $p(t)$ is differentiable and nonzero except at $t = (2 - \sqrt{3})\varepsilon$ and $t = a - (2 - \sqrt{3})\varepsilon$. By the quotient rule f is twice differentiable almost everywhere.

With $p(t)$ defined in (5.3), define

$$H_p := \{f^{[0]}(t) \text{ and } f^{[1]}(t) \text{ are absolutely continuous on } [0, a]\}.$$

With the above discussion, we have the following corollary based on Atkinson (1964) that are related to understanding the spectrum of the LLE matrix.

Corollary 19 Suppose we impose the Dirichlet boundary condition for the eigenvalue problem, $D_\varepsilon f(t) = \lambda f(t)$, over the elliptic region $[(2 - \sqrt{3})\varepsilon, a - (2 - \sqrt{3})\varepsilon]$; that is,

$$f((2 - \sqrt{3})\varepsilon) = f(a - (2 - \sqrt{3})\varepsilon) = 0.$$

Then the eigenvalues are discrete and bounded from above; that is, the eigenvalues are $\infty > \lambda_0 > \lambda_1 > \lambda_2 > \dots$ so that $\lambda_n \rightarrow -\infty$ as $n \rightarrow \infty$. If $f_n \in H_p$ is the corresponding eigenfunctions of λ_n , then f_n has exactly n zeros in $((2 - \sqrt{3})\varepsilon, a - (2 - \sqrt{3})\varepsilon)$.

We mention that this corollary is only for theoretical interest but not for practical interest since we need extra steps to “clip” the wave region and impose the Dirichlet boundary condition when we only have a point cloud. Also, the above conclusion may not hold when $P(t)$ is not uniform. In fact, it is not hard to show that if $P(t)$ behaves like $t + \varepsilon$ in $[0, \varepsilon]$, then the corresponding SL problem is singular.

6. A comparison of LLE and DM

We provide a comparison of LLE and DM (Coifman and Lafon, 2006) on a manifold with smooth boundary. Recall that, unlike LLE, when we run DM, the affinity matrix is defined by composing a fixed kernel function chosen by the user with the distance between pairs of sampled points. Below we summarize the bias analysis result of DM using our notations for a further comparison when the manifold has a non-empty boundary. A full calculation can be found in Coifman and Lafon (2006); Singer and Wu (2017). To simplify the comparison, we consider the Gaussian kernel $H(t) = e^{-t^2}$. More general kernels can be considered, and we refer the reader with interest to, e.g., (Coifman and Lafon, 2006; Singer and Wu, 2017). Also, see (Vaughn, 2020) for more relevant results. For

$x, y \in M$, we define $H_\varepsilon(x, y) = \exp\left(-\frac{\|\iota(x) - \iota(y)\|^2}{\varepsilon^2}\right)$, where $\varepsilon > 0$ is the bandwidth. For $0 \leq \alpha \leq 1$, we define the α -normalized kernel as

$$H_{\varepsilon, \alpha}(x, y) := \frac{H_\varepsilon(x, y)}{p_\varepsilon^\alpha(x)p_\varepsilon^\alpha(y)},$$

where $p_\varepsilon(x) := \mathbb{E}[H_\varepsilon(x, X)]$. With the α -normalized kernel $H_{\varepsilon, \alpha}$, for $f \in C^3(M)$, the diffusion operator associated with the α -normalized DM is

$$\mathcal{H}_{\varepsilon, \alpha} f(x) := \frac{\mathbb{E}[H_{\varepsilon, \alpha}(x, X)f(X)]}{\mathbb{E}[H_{\varepsilon, \alpha}(x, X)]}. \quad (6.1)$$

The behavior of the operator $\mathcal{H}_{\varepsilon, \alpha}$ is summarized below.

Theorem 20 (Bias analysis of Diffusion map) *Let (M, g) be a d -dimensional compact, smooth Riemannian manifold isometrically embedded in \mathbb{R}^p , with a non-empty smooth boundary. Suppose $f \in C^3(M)$ and $P \in C^2(M)$. If $\alpha = 1$, we have*

$$\begin{aligned} \mathcal{H}_{\varepsilon, \alpha} f(x) &= \frac{\sigma_{1,d}(\tilde{\varepsilon}_x)}{\sigma_0(\tilde{\varepsilon}_x)} \partial_d f(x) \varepsilon + \left[\psi_1(\tilde{\varepsilon}_x) \sum_{i=1}^{d-1} \partial_{ii}^2 f(x) + \psi_2(\tilde{\varepsilon}_x) \partial_{dd}^2 f(x) \right] \varepsilon^2 \\ &\quad + U(\tilde{\varepsilon}_x) \partial_d f(x) \varepsilon^2 + O(\varepsilon^3), \end{aligned} \quad (6.2)$$

where ψ_1, ψ_2 and U are scalar value functions defined on $[0, \infty)$ so that

$$\psi_1(t) := \frac{1}{2} \frac{\sigma_2(t)}{\sigma_0(t)}, \quad \psi_2(t) := \frac{1}{2} \frac{\sigma_{2,d}(t)}{\sigma_0(t)},$$

$U(t) = 0$ if $t \geq \varepsilon$, $U(\tilde{\varepsilon}_x)$ depends on the second fundamental form of ∂M in M at x , and $U(\tilde{\varepsilon}_x)$ is independent of P . In fact,

$$U(\tilde{\varepsilon}_x) = \frac{\int_{\tilde{D}_\varepsilon(x)} u_d du}{\int_{\tilde{D}_\varepsilon(x)} 1 du} - \frac{\sigma_{1,d}(\tilde{\varepsilon}_x)}{\sigma_0(\tilde{\varepsilon}_x)} \varepsilon, \quad (6.3)$$

where $\frac{\int_{\tilde{D}_\varepsilon(x)} u_d du}{\int_{\tilde{D}_\varepsilon(x)} 1 du}$ is a function depending on $\tilde{\varepsilon}_x$ and the second fundamental form of ∂M as a codimension 1 submanifold embedded in M at x by Definition 5.

Compared with the differential operator \mathcal{D}_ε associated with LLE, the differential operator associated with DM has a very different behavior. First, in DM, when $\varepsilon > 0$ is finite, asymptotically the first order differential operator exists in the ε order (Coifman and Lafon, 2006), which suggests that the boundary condition is *Neumann*. In Vaughn (2020), it is shown that the graph Laplacian converges in the weak sense to the Laplace-Beltrami operator with the Neumann boundary condition. With this boundary condition, the spectral convergence of DM when the boundary is non-empty without a convergence rate was provided in Singer and Wu (2017) as a special case when the considered group action is $SO(1)$. Another spectral convergence result when the boundary exists can be found in Peoples and Harlim (2021). Note that when the boundary is empty, more spectral convergence results with a convergence rate can be found in Von Luxburg et al. (2008); Trillos et al. (2020); Dunson et al. (2021), while none provide the rate is optimal to our knowledge. Second, the coefficients of the second order differential operator, ψ_1 and ψ_2 , do not change sign and do not

degenerate over the whole manifold. Third, in DM, there is an extra first order differential operator in the ε^2 order. As a result, in addition to what has been explored in Wu and Wu (2018), we see more differences between LLE and DM.

While LLE and DM are different, they are intimately related. Here we provide a brief exploration for this relationship from the kernel perspective. Observe that we can rewrite the kernel as

$$\begin{aligned} \frac{\mathbb{E}[K(x, X)f(X)]}{\mathbb{E}K(x, X)} &= \frac{\mathbb{E}[\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X)) - (\iota(X) - \iota(x))^\top \mathbf{T}(x)\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]f(X)}{\mathbb{E}[\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X)) - (\iota(X) - \iota(x))^\top \mathbf{T}(x)\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]} \\ &= \frac{\mathbb{E}[\frac{1}{2}\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X)) - \frac{1}{2}(\iota(X) - \iota(x))^\top \mathbf{T}(x)\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]f(X)}{\mathbb{E}[\frac{1}{2}\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X)) - \frac{1}{2}(\iota(X) - \iota(x))^\top \mathbf{T}(x)\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]}, \end{aligned}$$

which means that the kernel function is an “average” of two functions,

$$K_1(x, y) := \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(y))$$

and

$$K_2(x, y) := -(\iota(y) - \iota(x))^\top \mathbf{T}(x)\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(y)).$$

We can thus consider the following kernel generalizing the LLE kernel:

$$K^{(\alpha)}(x, y) = \alpha K_1(x, y) + (1 - \alpha)K_2(x, y),$$

where $\alpha \in [0, 1]$. Note that K_1 can be viewed as a 0–1 kernel that is commonly used in DM, so when $\alpha = 1$, we recover the DM. When $\alpha = 1/2$, it is clear that $K^{(1/2)}$ is the LLE kernel. When $\alpha = 0$, we get a different kernel with different behavior. Recall Definition 6. We have

$$\begin{aligned} K_2(x, y) &= -\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top I_{p,r}(C_x + cI_{p \times p})^{-1}I_{p,r}(\iota(y) - \iota(x)) \\ &= \int_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))} (\iota(z) - \iota(x))^\top I_{p,r}(C_x + cI_{p \times p})^{-1}I_{p,r}(\iota(y) - \iota(x))dz. \end{aligned} \quad (6.4)$$

As discussed in Malik et al. (2019), since $I_{p,r}(C_x + cI_{p \times p})^{-1}I_{p,r}$ can be viewed as the “regularized precision matrix”, we can view $(\iota(z) - \iota(x))^\top I_{p,r}(C_x + cI_{p \times p})^{-1}I_{p,r}(\iota(y) - \iota(x))$ as the local Mahalanobis distance between z and y , or the distance between the latent variables related to z and y . Thus, when $\alpha = 0$, the kernel comes from averaging out the pairwise local Mahalanobis distance, and hence depends on the local geometric structure. The relationship between the second fundamental form of M at x and the latent space will be explored in the future work.

It is natural to ask if we can “alleviate” the impact of the wave region by choosing different α . To answer this question, we briefly discuss the behavior of $\mathbb{E}[K^\alpha(x, X)f(X)]$ with different choices of α , particularly when $\alpha < 1$. Let us take a more careful look at the case when $\alpha = 1/2$; that is, the kernel for LLE; particularly, we look into the reason why LLE does not have the Neurman boundary condition, and why LLE is independent of the nonuniform density function from the kernel perspective. We have

$$\begin{aligned} \mathbb{E}[K^{(1/2)}(x, X)f(X)] &= \frac{1}{2}\mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] \\ &\quad - \frac{1}{2}\mathbb{E}[(\iota(X) - \iota(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \mathbf{T}(x). \end{aligned} \quad (6.5)$$

Recall the behavior of the two terms on the right hand side. By a direct calculation, we know $\mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(t(x))}(t(X))]$ becomes $P(x)\sigma_{1,d}(\tilde{\varepsilon}_x)\partial_d f(x)\varepsilon^{d+1} + O(\varepsilon^{d+2})$, when $x \in M_\varepsilon$, and $[\frac{1}{2}P(x)\Delta f(x) + \nabla f(x) \cdot \nabla P(x)]\varepsilon^{d+2} + O(\varepsilon^{d+3})$ when $x \notin M_\varepsilon$. Note that this is the behavior of the kernel $K^{(1)}$. On the other hand, $\mathbb{E}[(t(X) - t(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(t(x))}(t(X))]\top \mathbf{T}(x)$ becomes $P(x)\sigma_{1,d}(\tilde{\varepsilon}_x)\partial_d f(x)\varepsilon^{d+1} + O(\varepsilon^{d+2})$ when $x \in M_\varepsilon$, and $\nabla f(x) \cdot \nabla P(x)\varepsilon^{d+2} + O(\varepsilon^{d+3})$ when $x \notin M_\varepsilon$. Note that this is the behavior of the kernel $K^{(0)}$. As a result, when $x \in M_\varepsilon$, since the common term $P(x)\sigma_{1,d}(\tilde{\varepsilon}_x)\partial_d f(x)\varepsilon^{d+1}$ cancels, there is no such Neumann boundary behavior when $\alpha = 1/2$ as that in DM. When $x \notin M_\varepsilon$, then the common term $\nabla f(x) \cdot \nabla P(x)\varepsilon^{d+2}$ cancels. Hence, the behavior of LLE in the interior of the manifold is independent of the density function.

However, it is worth noting that one cannot remove the wave region M_w through adjusting α after the above analysis – since both $\mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(t(x))}(t(X))]$ and $\mathbb{E}[(t(X) - t(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(t(x))}(t(X))]\top \mathbf{T}(x)$ are dominated by $P(x)\sigma_{1,d}(\tilde{\varepsilon}_x)\partial_d f(x)\varepsilon^{d+1}$ when x is near the boundary, if $\alpha \neq 1/2$, the first order term remains.

7. Clipped LLE matrix

Based on the above theoretical results, we provide an immediate application. The Laplace-Beltrami operator with the Dirichlet boundary condition is widely used in various fields, like in the analysis of stochastic dynamics. For example, in Georgiou et al. (2017) the eigenfunctions of the Laplace-Beltrami operators with the Neumann and Dirichlet boundary conditions are used together to reconstruct the conformational space of a stochastic gradient system. According to the developed theory in Theorem 16, the asymptotic operator in general behaves well away from the boundary. We thus consider the following modification of LLE that echoes the theoretical development in Section 3 and this algorithm is a potential candidate to recover the Laplace-Beltrami operator with the Dirichlet boundary condition on manifold with boundary.

For a given sampling set $\mathcal{X} = \{x_i\}_{i=1}^n$, due to the ε -radius nearest neighbor scheme, assume we can divide the LLE matrix $W \in \mathbb{R}^{n \times n}$ into blocks according to four portions, the interior, transition, wave-boundary and non-wave-boundary portions. The interior portion $\mathcal{X}_I := \{x_i \in \mathcal{X} : d(x_i, \partial M) > 2\varepsilon\}$ that includes points far away from the boundary, the wave-boundary portion $\mathcal{X}_w := \{x_i \in \mathcal{X} : x_i \in M_w \cup \partial M\}$ that includes points in the wave region, $\mathcal{X}_B := \{x_i \in \mathcal{X} : x_i \in M_\varepsilon \setminus (M_w \cup \partial M)\}$ and the transition portion $\mathcal{X}_T := \{x_i \in \mathcal{X} : \varepsilon \leq d(x_i, \partial M) \leq 2\varepsilon\}$ that includes the remaining points touching the other three portions. The W matrix is thus divided into

$$W = \begin{bmatrix} W_{ww} & W_{wB} & W_{wT} & 0 \\ W_{Bw} & W_{BB} & W_{BT} & 0 \\ W_{Tw} & W_{TB} & W_{TT} & W_{TI} \\ 0 & 0 & W_{IT} & W_{II} \end{bmatrix},$$

where $W_{ww} \in \mathbb{R}^{|\mathcal{X}_w| \times |\mathcal{X}_w|}$ represents the wave-boundary portion of the LLE matrix, $W_{BB} \in \mathbb{R}^{|\mathcal{X}_B| \times |\mathcal{X}_B|}$ represents the non-wave-boundary portion of the LLE matrix, $W_{TT} \in \mathbb{R}^{|\mathcal{X}_T| \times |\mathcal{X}_T|}$ represents the transition portion of the LLE matrix, $W_{II} \in \mathbb{R}^{|\mathcal{X}_I| \times |\mathcal{X}_I|}$ represents the interior portion of the LLE matrix, and the other submatrices represent the interaction of the four portions of the LLE matrix.

Construct a new matrix $W_r \in \mathbb{R}^{n' \times n'}$, where $n' = n - |\mathcal{X}_w|$, by restricting W to \mathcal{Y} ; that is,

$$W_r = \begin{bmatrix} W_{BB} & W_{BT} & 0 \\ W_{TB} & W_{TT} & W_{TI} \\ 0 & W_{IT} & W_{II} \end{bmatrix}.$$

We call W_r the *clipped LLE matrix* for simplicity. Note that in general the matrix W_r is not symmetric, not a transition matrix, and the rows may not sum to 1. We shall emphasize that this algorithm depends on the knowledge of the wave region. In general, we need an algorithm to detect the wave region from a point cloud.

Below we see some numerical results. We uniformly and independently sample points from $M_1 := [0, 1] \subset \mathbb{R}^1$. The LLE matrix is constructed with the ε -radius scheme, where $\varepsilon = 0.01$. The first 5 eigenfunctions are shown in the top panel of Figure 5. Note that the first eigenfunction is constant, and the second eigenfunction is linear, and both are with eigenvalue 1; that is, these two eigenfunctions form the null space of $I - W$. The other eigenfunctions “look like” the eigenfunctions of the Laplace-Beltrami operator with the Dirichlet boundary condition, but higher eigenfunctions become “irregular” when getting closer to the boundary. Next, consider another 1-dim curve M_2 embedded in \mathbb{R}^3 that is parametrized by $t \rightarrow (t, \log(0.5 + t), \cos(\pi t))^\top \in \mathbb{R}^3$, where $t \in [0, 1]$. We uniformly and independently sample 8,000 points from $[0, 1]$ and mapped them to M_2 . Denote the sampled points $\mathcal{X} := \{x_i\}_{i=1}^{8,000} \subset \mathbb{R}^3$. Note that the sample is not uniform. The LLE matrix $W \in \mathbb{R}^{8,000 \times 8,000}$ is constructed with the ε -radius scheme, where $\varepsilon = 0.01$, and hence the clipped LLE matrix W_r . The first 10 eigenfunctions of W_r constructed from M_1 and M_2 are shown in the middle and bottom panels in Figure 5. According to Corollary 17 and the discussion in Subsection 5, the asymptotic operator is well behaved in $[(2 - \sqrt{3})\varepsilon, 1 - (2 - \sqrt{3})\varepsilon]$. This theoretical finding fits the numerical results – the eigenfunctions are all 0 at the “boundary points” $(2 - \sqrt{3})\varepsilon$ and $1 - (2 - \sqrt{3})\varepsilon$.

Second, we uniformly sample points from a unit disk, $M_3 \subset \mathbb{R}^2$, by keeping points with norm less than or equal to 1 from 20,000 points sampled uniformly and independently from $[-1, 1] \times [-1, 1]$. The LLE matrix is constructed with the ε -radius ball nearest neighbor search scheme, where $\varepsilon = 0.1$. The first 20 eigenfunctions are shown in Figure 6. The first eigenfunction is constant, and the second and third eigenfunctions are linear, and these three eigenfunctions are associated with eigenvalue 1. These three eigenfunctions form the null space of $I - W$. We can see three types of eigenfunctions – those of the first type “look like” the eigenfunctions of the Laplace-Beltrami operator with the Dirichlet boundary condition, those of the second type “look like” the eigenfunctions of the Laplace-Beltrami operator restricted to the “rim” near the boundary, which is topologically a closed manifold S^1 , and those of the third type “look like” the mix-up of the first two types. Next, we explore the clipped LLE matrix on the unit disk $M_3 \subset \mathbb{R}^2$ with the same uniform sampling scheme. For M_3 , we remove rows and columns associated with points with norm greater than $1 - (2 - \sqrt{3})\varepsilon$, where $\varepsilon = 0.1$. The first 20 eigenfunctions are shown in Figure 7. It is clear that compared with those shown in Figure 6, all eigenfunctions are 0 at the “boundary”.

Next, the first 20 eigenfunctions of the LLE matrix and the clipped LLE matrix of the surface shown in Figures 2 and 3 with the same sampling scheme are shown in Figures 8 and 9. It is clear that while the first 3 eigenfunctions of the LLE matrix behave like constant or linear functions, the other eigenfunctions are not easy to describe. However, all eigenfunctions of the clipped LLE matrix are zero on the “boundary”, which behaves like the eigenfunctions of the Laplace-Beltrami operator with the Dirichlet boundary condition.

The above examples are all manifolds without interesting topological structure since they can all be parametrized by one chart. In the final example we show a two dimensional manifold with non-trivial topology. Consider a torus embedded in \mathbb{R}^3 , which is parametrized by

$$\Phi : (\theta, \phi) \mapsto ((3 + 1.2 \cos(\theta)) \cos(\phi), (3 + 1.2 \cos(\theta)) \sin(\phi), 1.2 \sin(\phi))^{\top} \in \mathbb{R}^3,$$

where $\theta, \phi \in [0, 2\pi)$. The manifold M_4 is defined as

$$M_4 = \{\Phi(\theta, \phi) : \theta, \phi \in [0, 2\pi) \text{ and } (3 + 1.2 \cos(\theta)) \cos(\phi) > -3.4\};$$

that is, M_4 is a truncated torus with the boundary diffeomorphic to S^1 . We sample uniformly 25,000 points on $[0, 2\pi] \times [0, 2\pi]$, and remove points associated with $(3 + 1.2 \cos(\theta)) \cos(\phi) > -3.4$. Note that this is a nonuniform sampling scheme from M_4 . Then, establish the LLE matrix and the clipped LLE matrix with $\varepsilon = 0.3$. The results are shown in Figures 10 and 11. Again, it is clear that the eigenfunctions of the clipped LLE matrix are zero on the “boundary”.

With these numerical results, we conjecture that if we clip the wave region, the operator $\mathcal{D}_{\varepsilon}$ over $M \setminus (M_w \cup \partial M)$ asymptotically converges to the Laplace-Beltrami operator with the Dirichlet boundary condition over $M \setminus (M_w \cup \partial M)$ in the spectral sense when $\varepsilon \rightarrow 0$.¹ We will explore this problem in our future work.

8. Discussion and Conclusion

In this paper, we provide an exploration of LLE when the manifold has boundary. We mention several interesting problems that we will explore in our future work.

First, the distribution of the LLE matrix eigenvalues under the null case has an interesting distribution behavior, which rings the bell of the interaction between kernel random matrix and random matrix theory. Recently, the spectral behavior of graph Laplacian has been studied from the random matrix perspective in Ding and Wu (2022). However, due to the non-symmetric nature of the LLE matrix, such an approach cannot be directly applied. Understanding the behavior of LLE will pave the road toward statistical inference of unsupervised manifold learning.

Second, the potential Dirichlet boundary condition associated with the clipped LLE matrix and its relationship with the Neuman boundary condition associated with the GL suggest exploring the Dirichlet-to-Neumann map and Schur’s complement as a future direction. Moreover, from the practical perspective, when we do not know where is the wave region, we shall design an effective algorithm to determine all points from the wave region, so that we can clip the LLE matrix.

Third, as is shown in Theorem 16, LLE converges pointwisely to a mixed-type differential operator with degeneracy, which is an SL equation with a peculiar structure in the one-dimensional case. In other words, we have a degenerate mixed-type differential equation and the boundary condition is not known a priori. Understanding the spectral behavior of this operator is of interest on its own from the theoretical perspective and it might be necessary in order to explore the spectral behavior of LLE when there is a boundary.

Fourth, the spectral convergence of LLE is so far an open problem to our knowledge. To attack this problem, we shall again compare the original proposed $(W - I)^{\top}(W - I)$ and $W - I$ considered

1. In the 1-dim case, this is related to a different differential equation, the Kimura equation Epstein and Mazzeo (2013), that shares the same degeneracy on the boundary. In the Kimura equation, the boundary condition is adaptively encoded in the functional space that we search for the eigenfunctions.

in this paper and (Wu and Wu, 2018). Assume first that $\partial M = \emptyset$. We may learn from what has been done in the literature. To study the spectral convergence of DM (Trillos et al., 2020; Dunson et al., 2021), we need at least two pieces of information. The first one is the knowledge of the spectral behavior of the asymptotic differential operator, and the second one is the pointwise convergence result. The pointwise convergence result of $W - I$ has been studied in Wu and Wu (2018) when $\partial M = \emptyset$, which is generalized to the case $\partial M \neq \emptyset$ in this paper. It might be intuitive to conclude that since the spectral behavior of the Laplace-Beltrami operator has been well known, when $\partial M = \emptyset$, we may easily obtain the spectral convergence of $W - I$. However, since the matrix $W - I$ is not symmetric, the associated integral operator of LLE is not self-adjoint. Thus, we cannot directly apply the same method in Trillos et al. (2020); Dunson et al. (2021) to prove the spectral convergence of $W - I$, and new tools are needed. Moreover, since the fourth order differential operator is involved in $(W - I)^\top(W - I)$ via a pointwise convergence analysis, and its spectral behavior is less well known, it is more challenging to study the spectral convergence of $(W - I)^\top(W - I)$. The situation is certainly more complicated when $\partial M \neq \emptyset$ since the associated kernel, the integral operator behavior, and the asymptotic differential operator are all different. Even for $W - I$, it is still an open problem as mentioned above as the third point, not to mention $(W - I)^\top(W - I)$ and if the nearest neighbor scheme is taken into consideration. A further systematic pointwise and spectral convergence study of $(W - I)^\top(W - I)$ is thus needed to advance the field.

Acknowledgments

The authors acknowledge the fruitful discussion with Professor Jun Kitagawa about the boundary condition. They also thank the anonymous reviewers for their valuable comments that help strengthen this paper.

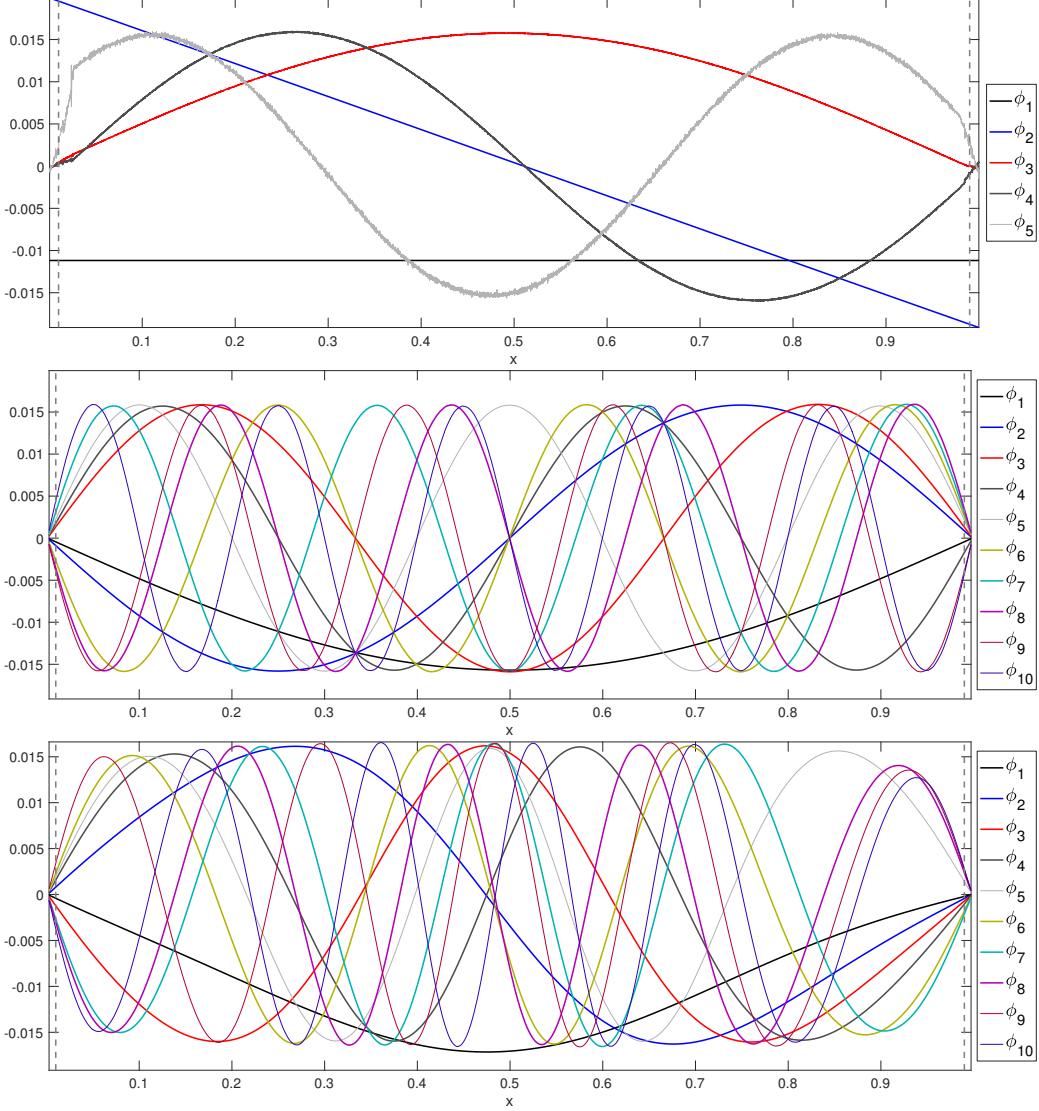


Figure 5: Top: The first 5 eigenfunctions of the LLE matrix for a point cloud sampled from the $[0, 1]$ interval are plotted with different colors. The dashed vertical gray lines indicate ϵ and $1 - \epsilon$. It is clear that the third, fourth, and fifth eigenfunctions “look like” the eigenfunctions of the Laplace-Beltrami operator with the Dirichlet boundary condition, but higher eigenfunctions become “irregular” when getting closer to the boundary. Middle: the first 10 eigenfunctions of the clipped LLE matrix W_r for a point cloud sampled from $M_1 = [0, 1]$ are plotted with different colors. The dashed vertical gray lines indicate ϵ and $1 - \epsilon$. Bottom: the first 10 eigenfunctions of the clipped LLE matrix W_r for a point cloud sampled from the curve $M_3 \subset \mathbb{R}^3$ are plotted with different colors. The dashed vertical gray lines indicate ϵ and $1 - \epsilon$. Compared with those eigenfunctions of M_1 , the amplitude of higher eigenfunctions of M_3 becomes less constant, which is expected due to the nonuniform sampling effect. It is clear that these eigenfunctions “look like” the eigenfunctions of the Laplace-Beltrami operator with the Dirichlet boundary condition without the “irregularity” behavior close to the boundary observed in the top panel.

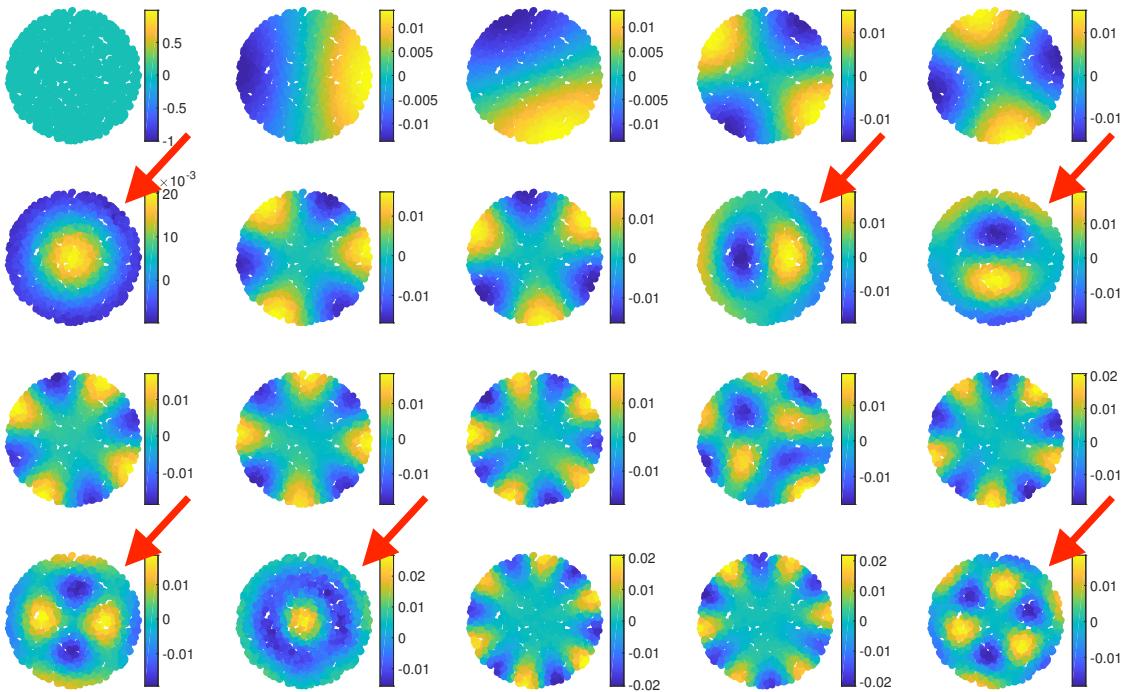


Figure 6: The first 20 eigenfunctions of the LLE matrix for a point cloud sampled from the unit disk are plotted from top left to bottom right. It is clear that some eigenfunctions (indicated by red arrows) “look like” the eigenfunctions of the Laplace-Beltrami operator with the Dirichlet boundary condition combined with the eigenfunctions of Laplace-Beltrami operator of the “rim” near the boundary. Note that the “rim” near the boundary is close to S^1 in the Gromov-Hausdorff sense.

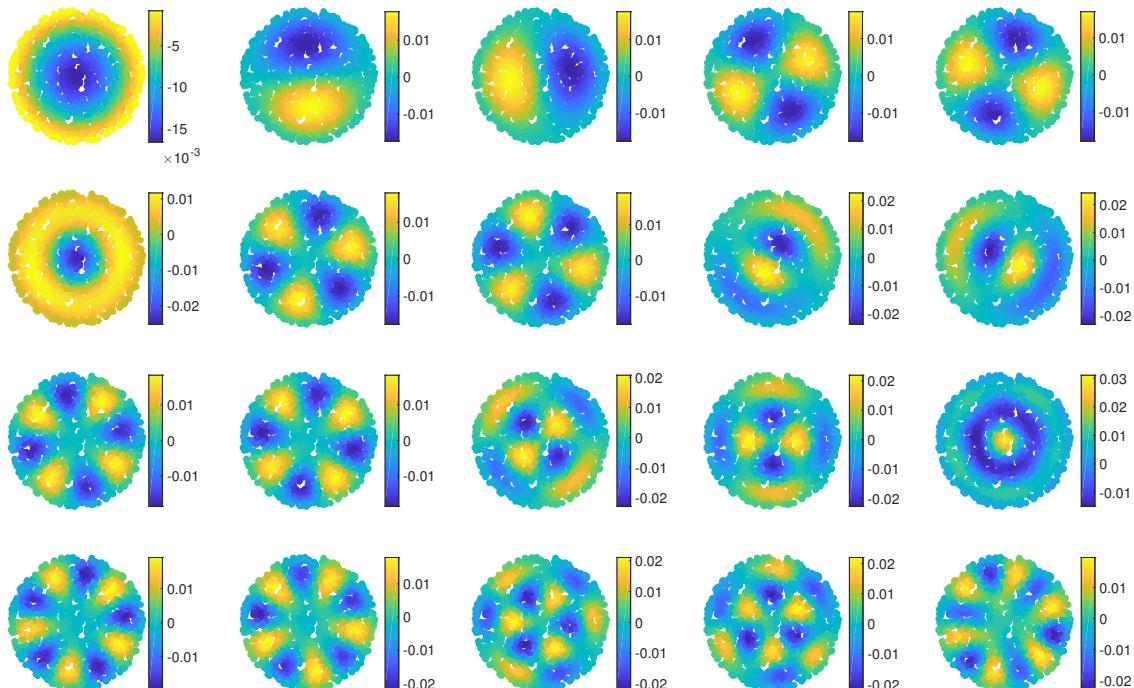


Figure 7: The first 20 eigenfunctions of the clipped LLE matrix for a point cloud sampled from the unit disk are plotted from top left to bottom right. It is clear that all eigenfunctions “look like” the eigenfunctions of the Laplace-Beltrami operator with the Dirichlet boundary condition.

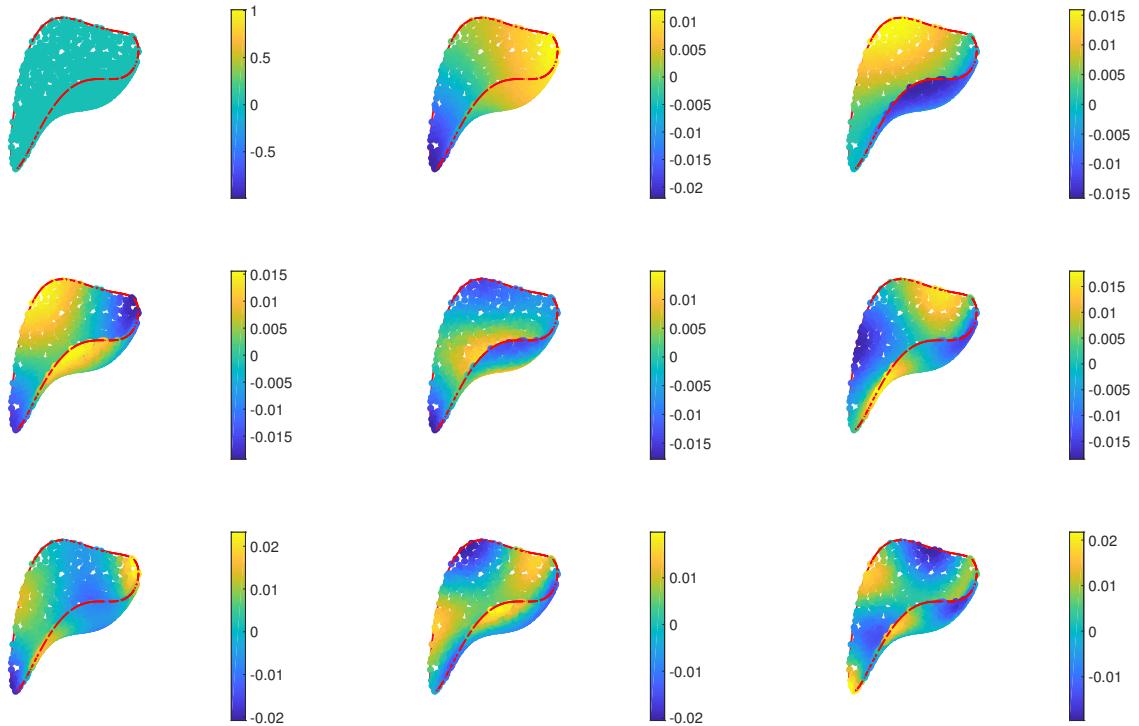


Figure 8: The first 9 eigenfunctions of the LLE matrix for a point cloud sampled from the surface in Figures 2 and 3 are plotted from top left to bottom right. To enhance the visualization, the boundary of the surface is colored by red. It is clear that the first three eigenfunctions are either constant or linear, while the behavior of other eigenfunctions is not easy to describe, while compared with those shown in Figure 6.

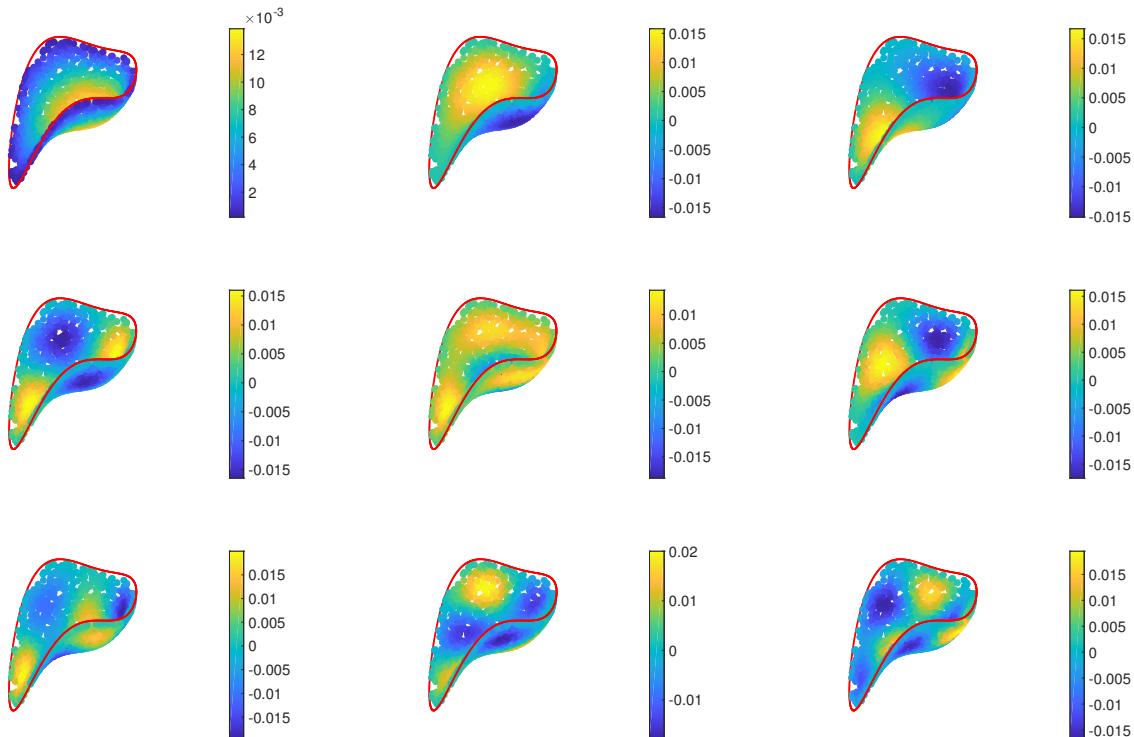


Figure 9: The first 9 eigenfunctions of the clipped LLE matrix for a point cloud sampled from the surface in Figures 2 and 3 are plotted from top left to bottom right. To enhance the visualization, the boundary of the surface is colored by red. It is clear that all eigenfunctions are zero on the boundary, and the behavior “looks like” the eigenfunctions of the Laplace-Beltrami operator with the Dirichlet boundary condition.

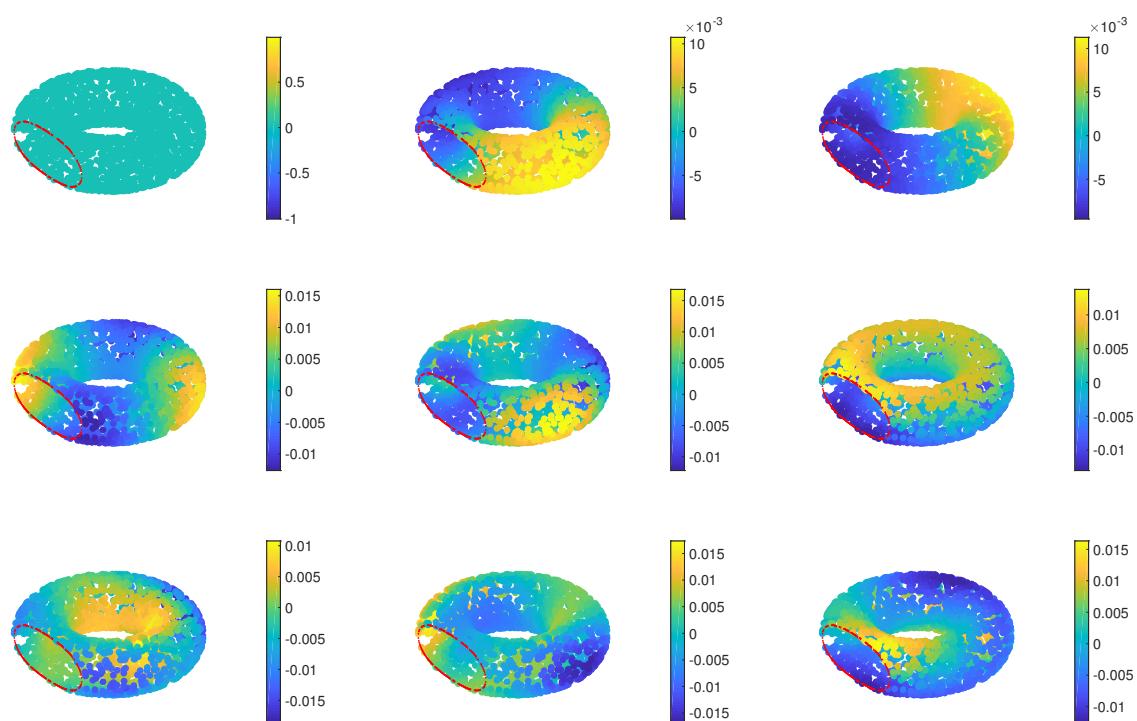


Figure 10: The first 9 eigenfunctions of the LLE matrix for a point cloud sampled from the truncated torus M_4 are plotted from top left to bottom right. To enhance the visualization, the boundary of the truncated torus is colored by red.

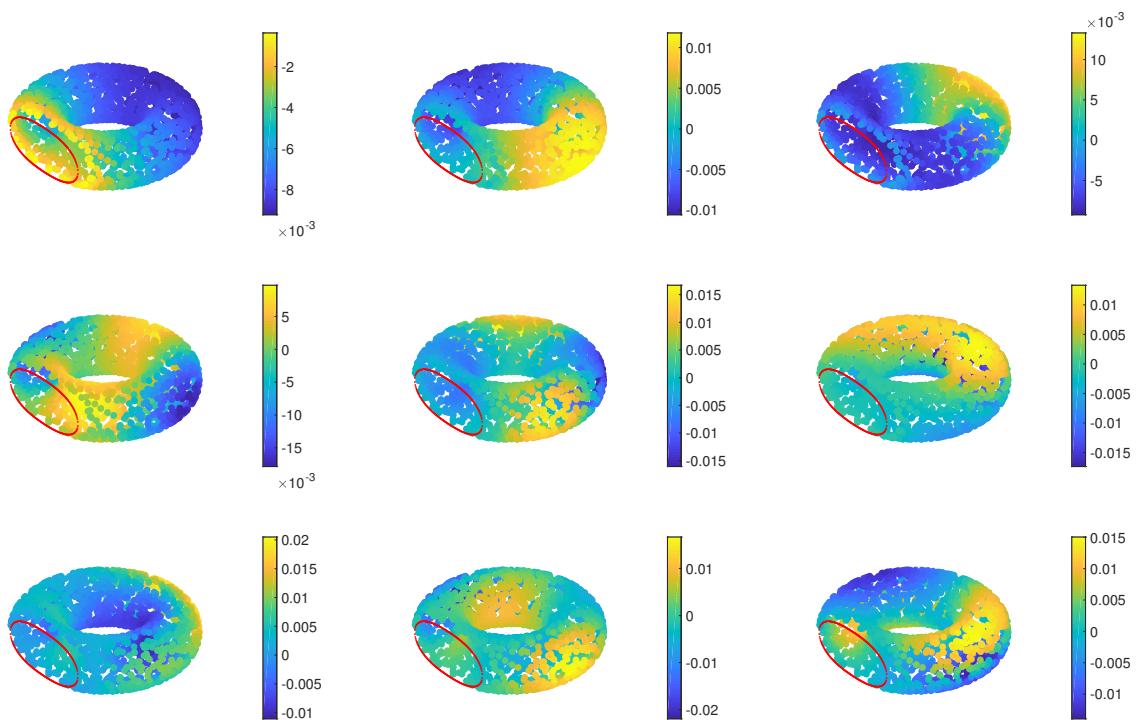


Figure 11: The first 9 eigenfunctions of the clipped LLE matrix for a point cloud sampled from the truncated torus M_4 are plotted from top left to bottom right. To enhance the visualization, the boundary of the truncated torus is colored by red. It is clear that all eigenfunctions are zero on the “boundary”.

Appendix A. Examples for Proposition 1

To show that it is possible $\rho(W) = 1$, consider the following example. Let $n = 2m$, where $m \geq 2$ is an integer. Suppose $\mathcal{X} = \{z_1, z_2, \dots, z_n\}$ is a uniform grid of $S^1 \subset \mathbb{R}^2$ so that $z_i = (\cos(\frac{2\pi(i-1)}{n}), \sin(\frac{2\pi(i-1)}{n}))$, where $i = 1, \dots, n$. We choose ε so that \mathcal{N}_k only contains two data points $(\cos(\frac{2\pi(i-2)}{n}), \sin(\frac{2\pi(i-2)}{n}))$ and $(\cos(\frac{2\pi i}{n}), \sin(\frac{2\pi i}{n}))$. Fix z_k , and $z_{k,1}$ and $z_{k,2}$ are the two data points in \mathcal{N}_k . Without loss of generality, we assume that $z_k = (0, 0)$, $z_{k,1} = (a, b)$ and $z_{k,2} = (-a, b)$. Hence, $G_{n,k}$ at z_k is

$$G_{n,k} = \begin{bmatrix} a & -a \\ b & b \end{bmatrix}, \quad (\text{A.1})$$

and the solution $y_k^\top = [y_{k,1}, y_{k,2}]$ to the regularized equation (2.3) with the regularizer $c > 0$ satisfies

$$\begin{bmatrix} a^2 + b^2 + c & -a^2 + b^2 \\ -a^2 + b^2 & a^2 + b^2 + c \end{bmatrix} \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \quad (\text{A.2})$$

Therefore, we have $y_{k,1} = y_{k,2}$, $w_k^\top = [1/2, 1/2]$, and

$$W_{ki} = \begin{cases} 1/2 & \text{if } z_i = z_{k,j} \in \mathcal{N}_k; \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.3})$$

Suppose $\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$ are the eigenvalues of W . Then $\lambda_0 = -1$, $\lambda_{n-1} = 1$ and $\lambda_{2i-1} = \lambda_{2i} = \cos(\frac{\pi(m-i)}{m})$ for $i = 1, \dots, m-1$.

We provide another example to show that in general it is possible that $\rho(W) > 1$. Consider a point cloud with ten points in \mathbb{R}^3 , $(-0.56, -0.34, 1.03)$, $(-0.51, 0.32, -0.02)$, $(-0.53, -1.47, -0.57)$, $(1.34, 0.47, -0.15)$, $(1.01, -1.56, 1.22)$, $(-0.55, -1, -0.07)$, $(0.09, -1.04, -0.2)$, $(-1.27, 2.07, -0.9)$, $(1.26, -0.71, -1.2)$, and $(1.46, 0, 0.61)$. The LLE matrix of this point cloud with 5 nearest neighbors and the regularizer $c = 10^{-3}$ has an eigenvalue -2.4233 .

Appendix B. Technical lemmas for some geometric quantities

In this section we collect several technical lemmas for some geometric quantities we will encounter in the proof. They might be also useful for other works when the manifold with boundary setup is considered.

For the manifold with boundary, denoted as M , with the isometric embedding ι into \mathbb{R}^p , we consider the extensions \tilde{M} and $\tilde{\iota}$ introduced in Section 3.2. Let \exp_x be the exponential map of \tilde{M} at $x \in \tilde{M}$. Recall that \exp_x is well defined over $\tilde{\iota}^{-1}(B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \iota(M))$ for any $x \in M$. For $x \in M$, we use Π_x to denote both the second fundamental form of $\iota(M)$ at $\iota(x)$ and the second fundamental form of $\tilde{\iota}(M)$ at $\tilde{\iota}(x)$. The first three lemmas are basic facts about \exp_x , the normal coordinate, and the volume form. The proof of these three lemmas can be found in Singer and Wu (2012).

Lemma 21 *Fix $x \in M$. Consider the extension \tilde{M} . If we use the Cartesian coordinate to parametrize $T_x \tilde{M}$, the volume form has the following expansion*

$$dV = \left(1 - \frac{1}{6} \sum_{i,j=1}^d Ric_x(i, j) u_i u_j + O(u^3) \right) du, \quad (\text{B.1})$$

where $u = \sum_{i=1}^d u_i e_i \in T_x \tilde{M}$, $Ric_x(i, j) = Ric_x(e_i, e_j)$.

Lemma 22 Fix $x \in M$. Consider the extensions \tilde{M} and $\tilde{\iota}$. For $u \in T_x\tilde{M}$ with $\|u\|$ sufficiently small, we have the following Taylor expansion:

$$\tilde{\iota} \circ \exp_x(u) - \tilde{\iota}(x) = \tilde{\iota}_*(u) + \frac{1}{2}\mathbb{II}_x(u, u) + O(\|u\|^3). \quad (\text{B.2})$$

In the next Lemma, we compare the geodesic distance and the Euclidean distance.

Lemma 23 Fix $x \in M$. Consider the extensions \tilde{M} and $\tilde{\iota}$. If we use the polar coordinate $(t, \theta) \in [0, \infty) \times S^{d-1}$ to parametrize $T_x\tilde{M}$, when $t > 0$ is sufficiently small and $\tilde{t} = \|\tilde{\iota} \circ \exp_x(\theta t) - \tilde{\iota}(x)\|_{\mathbb{R}^p}$, then

$$\begin{aligned} \tilde{t} &= t - \frac{1}{24}\|\mathbb{II}_x(\theta, \theta)\|^2 t^3 + O(t^4) \\ t &= \tilde{t} + \frac{1}{24}\|\mathbb{II}_x(\theta, \theta)\|^2 \tilde{t}^3 + O(\tilde{t}^4), \end{aligned} \quad (\text{B.3})$$

where $\theta \in S^{d-1} \subset T_x\tilde{M}$.

The following lemma describes a parametrization of the boundary set. This parametrization is needed when we analyze the LLE matrix near the boundary.

Lemma 24 Fix $x \in M_\varepsilon$. Consider the extensions \tilde{M} and $\tilde{\iota}$.

$$\begin{aligned} &(\tilde{\iota} \circ \exp_x)^{-1}(B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \iota(\partial M)) \\ &= \left\{ \sum_{l=1}^d u^l \partial_l \in T_x\tilde{M} \mid (u^1, \dots, u^{d-1}) \in K, u^d = q(u^1, \dots, u^{d-1}) \right\}, \end{aligned} \quad (\text{B.4})$$

where

$$q(u^1, \dots, u^{d-1}) = \tilde{\varepsilon}_x + \sum_{i,j=1}^{d-1} a_{ij}(x_\partial) u^i u^j + O(\|u\|^3), \quad (\text{B.5})$$

and $a_{ij}(x_\partial)$ is the second fundamental form of the embedding of ∂M in M at x_∂ .

Proof Note that $(\tilde{\iota} \circ \exp_x)^{-1}(B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \iota(\partial M))$ is a hypersurface with boundary in $T_x M$. Since ∂M is smooth, by the implicit function theorem, if ε is small enough,

$$\begin{aligned} &(\tilde{\iota} \circ \exp_x)^{-1}(B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \iota(\partial M)) \\ &= \left\{ \sum_{l=1}^d u^l \partial_l \in T_x\tilde{M} \mid (u^1, \dots, u^{d-1}) \in K, u^d = q(u^1, \dots, u^{d-1}) \right\} \end{aligned} \quad (\text{B.6})$$

for a smooth function q of u^1, \dots, u^{d-1} . By Taylor's expansion, we have

$$q(u^1, \dots, u^{d-1}) = \tilde{\varepsilon}_x + \sum_{i,j=1}^{d-1} a_{ij}(x) u^i u^j + O(\|u\|^3), \quad (\text{B.7})$$

where the first order disappears since the tangent space of $(\tilde{\iota} \circ \exp_x)^{-1}(B_\varepsilon^{\mathbb{R}^p}(\iota(x)) \cap \iota(\partial M))$ in $T_x M$ at $\exp_x^{-1}(x_\partial)$ is perpendicular to u_d direction by Gauss's lemma, and $a_{ij}(x)$ is the coefficient of the

second order expansion. Due to the smoothness of the manifold, $a_{ij}(x)$ is smooth along the minimizing geodesic from x_∂ to x . Also, when $x = x_\partial$, $a_{ij}(x_\partial)$ is the second fundamental form of the embedding of ∂M in M at x_∂ . Therefore, by another Taylor's expansion, $a_{ij}(x) = a_{ij}(x_\partial) + O(u_d)$, the conclusion follows. \blacksquare

Next Lemma describes the discrepancy between $\int_{D_\varepsilon(x)} f(u)du$ and $\int_{\tilde{D}_\varepsilon(x)} f(u)du$. Note that the order of the discrepancy does not depend on the location of x .

Corollary 25 *Fix $x \in M$. When $\varepsilon > 0$ is sufficiently small, we have*

$$\left| \int_{D_\varepsilon(x)} du - \int_{\tilde{D}_\varepsilon(x)} du \right| = O(\varepsilon^{d+2}). \quad (\text{B.8})$$

Proof Based on Lemma 23 and the definition of $\tilde{D}_\varepsilon(x)$, the distance between the boundary of $D_\varepsilon(x)$ and the boundary of $\tilde{D}_\varepsilon(x)$ is of order ε^3 . The volume of the boundary $\tilde{D}_\varepsilon(x)$ is of order ε^{d-1} . Hence the volume difference between $\tilde{D}_\varepsilon(x)$ and $D_\varepsilon(x)$ is of order $\varepsilon^{d-1} \cdot \varepsilon^3 = \varepsilon^{d+2}$. The conclusion follows. \blacksquare

Appendix C. Technical lemmas for the kernel analysis

To have a closer look at the kernel, we need the following quantities. First, we introduce some notations. For $v \in \mathbb{R}^p$, denote

$$v = [[v_1, v_2]] \in \mathbb{R}^p, \quad (\text{C.1})$$

where $v_1 \in \mathbb{R}^d$ forms the first d coordinates of v and $v_2 \in \mathbb{R}^{p-d}$ forms the last $p-d$ coordinates of v . Thus, for $v = [[v_1, v_2]] \in T_{t(x)} \mathbb{R}^p$, $v_1 = J_{p,d}^\top v$ is the coordinate of the tangential component of v on $t_* T_x M$ and $v_2 = \bar{J}_{p,p-d}^\top v$ is the coordinate of the normal component of v associated with a chosen basis of the normal bundle. Define

$$\mathfrak{N}_{ij}(x) := \bar{J}_{p,p-d}^\top \mathbb{I}_{ij}(x).$$

Note that $\mathfrak{N}_{ij}(x) = \mathfrak{N}_{ji}(x)$.

Definition 26 (Moments) *For $x \in M$, consider the following moments that capture the geometric asymmetry:*

$$\mu_v(x, \varepsilon) := \int_{\tilde{D}_\varepsilon(x)} \prod_{i=1}^d u_i^{v_i} du,$$

where $v = [v_1, \dots, v_d]^\top$ describes the moment order.

In next lemma, we quantitatively describe all the moments up to the third order. This Lemma tells us that when $x \notin M_\varepsilon$ (when x is far away from the boundary), all odd order moments disappear due to the symmetry of the integration domain. However, when $x \in M_\varepsilon$, it no longer holds – the integration domain becomes asymmetric, and the odd moments no longer disappear. Therefore, we can show that $\mu_0(x, \varepsilon)$, $\mu_{e_d}(x, \varepsilon)$, $\mu_{2e_i}(x, \varepsilon)$ and $\mu_{2e_i+e_d}(x, \varepsilon)$ are all the non-trivial moments needed in analyzing LLE. The proof follows from the symmetry argument and a straightforward integration, so we omit it here.

Lemma 27 [Symmetry] Suppose ε is sufficiently small. Then, the moments up to order three can be quantitatively described as follows. In fact, $\mu_0(x, \varepsilon)$, $\mu_{e_d}(x, \varepsilon)$, μ_{2e_i} and $\mu_{2e_i+e_d}(x, \varepsilon)$ for all $i = 1, \dots, d$ are the only non-trivial moments. And they are continuous functions of x on M . Define $\frac{|S^{d-2}|}{d-1} = 1$ when $d = 1$, we have

1. **Zero order moment**, μ_0

If $x \in M_\varepsilon$, μ_0 is an increasing function of $\tilde{\varepsilon}_x$ and

$$\mu_0(x, \varepsilon) = \frac{|S^{d-1}|}{2d} \varepsilon^d + \int_0^{\tilde{\varepsilon}_x} \frac{|S^{d-2}|}{d-1} (\varepsilon^2 - h^2)^{\frac{d-1}{2}} dh + O(\varepsilon^{d+1}).$$

If $x \notin M_\varepsilon$, then

$$\mu_0(x, \varepsilon) = \frac{|S^{d-1}|}{d} \varepsilon^d.$$

In general, the following bound holds for $\mu_0(x, \varepsilon)$:

$$\frac{|S^{d-1}|}{2d} \varepsilon^d + O(\varepsilon^{d+1}) \leq \mu_0(x, \varepsilon) \leq \frac{|S^{d-1}|}{d} \varepsilon^d.$$

2. **First order moment**, μ_{e_d}

If $x \in M_\varepsilon$, μ_{e_d} is an increasing function of $\tilde{\varepsilon}_x$ and

$$\mu_{e_d}(x, \varepsilon) = -\frac{|S^{d-2}|}{d^2-1} (\varepsilon^2 - \tilde{\varepsilon}_x^2)^{\frac{d+1}{2}} + O(\varepsilon^{d+2}).$$

If $x \notin M_\varepsilon$, then

$$\mu_{e_d}(x, \varepsilon) = 0.$$

In general, $\mu_{e_d}(x, \varepsilon)$ is of order ε^{d+1} . For the rest of the moments, $\mu_{e_i} = 0$, for $i = 1, \dots, d-1$.

3. **Second order moment**, μ_{2e_i}

If $x \in M_\varepsilon$, μ_{2e_i} is an increasing function of $\tilde{\varepsilon}_x$ for $i = 1, \dots, d$. We have

$$\mu_{2e_i}(x, \varepsilon) = \frac{|S^{d-1}|}{2d(d+2)} \varepsilon^{d+2} + \int_0^{\tilde{\varepsilon}_x} \frac{|S^{d-2}|}{d^2-1} (\varepsilon^2 - h^2)^{\frac{d+1}{2}} dh + O(\varepsilon^{d+3}),$$

for $i = 1, \dots, d-1$, and

$$\mu_{2e_d}(x, \varepsilon) = \frac{|S^{d-1}|}{2d(d+2)} \varepsilon^{d+2} + \int_0^{\tilde{\varepsilon}_x} \frac{|S^{d-2}|}{d-1} (\varepsilon^2 - h^2)^{\frac{d-1}{2}} h^2 dh + O(\varepsilon^{d+3}).$$

If $x \notin M_\varepsilon$, then

$$\mu_{2e_i}(x, \varepsilon) = \frac{|S^{d-1}|}{d(d+2)} \varepsilon^{d+2}.$$

In general, the following bounds hold for $\mu_{2e_i}(x, \varepsilon)$, where $i = 1, \dots, d$:

$$\frac{|S^{d-1}|}{2d(d+2)} \varepsilon^{d+2} + O(\varepsilon^{d+3}) \leq \mu_{2e_i}(x, \varepsilon) \leq \frac{|S^{d-1}|}{d(d+2)} \varepsilon^{d+2},$$

For the rest of the moments, $\mu_{e_i+e_j} = 0$, whenever $i \neq j$.

4. Third order moment, $\mu_{e_i+e_j+e_k}$

If $x \in M_\varepsilon$, $\mu_{2e_i+e_d}$ is an increasing function of $\tilde{\varepsilon}_x$ and

$$\mu_{2e_i+e_d}(x, \varepsilon) = -\frac{|S^{d-2}|}{(d^2-1)(d+3)}(\varepsilon^2 - \tilde{\varepsilon}_x^2)^{\frac{d+3}{2}} + O(\varepsilon^{d+4}),$$

for $i = 1, \dots, d-1$, and

$$\mu_{3e_d}(x, \varepsilon) = -\frac{|S^{d-2}|}{(d^2-1)(d+3)}(\varepsilon^2 - \tilde{\varepsilon}_x^2)^{\frac{d+1}{2}}(2\varepsilon^2 + (d+1)\tilde{\varepsilon}_x^2) + O(\varepsilon^{d+4}).$$

If $x \notin M_\varepsilon$, then

$$\mu_{3e_d}(x, \varepsilon) = 0.$$

In general, $\mu_{2e_i+e_d}(x, \varepsilon)$ is of order ε^{d+3} . And $\mu_{e_i+e_j+e_k} = 0$, for the rest of the cases.

Below, we relate those non-trivial moments described in the previous lemma to those σ functions in Definition 7. The proof follows from a straightforward change of variable, so we omit the details.

Corollary 28 *The relationship between the non-trivial moments in Lemma 27 and the functions σ defined in Definition 7 satisfies:*

$$\begin{aligned}\mu_0(x, \varepsilon) &= \sigma_0(\tilde{\varepsilon}_x)\varepsilon^d + O(\varepsilon^{d+1}) \\ \mu_{e_d}(x, \varepsilon) &= \sigma_{1,d}(\tilde{\varepsilon}_x)\varepsilon^{d+1} + O(\varepsilon^{d+2}). \\ \mu_{2e_i}(x, \varepsilon) &= \sigma_2(\tilde{\varepsilon}_x)\varepsilon^{d+2} + O(\varepsilon^{d+3}),\end{aligned}$$

for $i = 1, \dots, d-1$, and

$$\mu_{2e_d}(x, \varepsilon) = \sigma_{2,d}(\tilde{\varepsilon}_x)\varepsilon^{d+2} + O(\varepsilon^{d+3}).$$

Moreover,

$$\mu_{2e_i+e_d}(x, \varepsilon) = \sigma_3(\tilde{\varepsilon}_x)\varepsilon^{d+3} + O(\varepsilon^{d+4}),$$

for $i = 1, \dots, d-1$, and

$$\mu_{3e_d}(x, \varepsilon) = \sigma_{3,d}(\tilde{\varepsilon}_x)\varepsilon^{d+3} + O(\varepsilon^{d+4}).$$

Next, we prove the following lemma about the ratio between the volumes of $d-1$ sphere and $d-2$ sphere. We need it to study the relation between different σ functions later.

Lemma 29 *For $d \in \mathbb{N}$, we have*

$$\frac{(d+1)^2(d+3)}{8d^2(d+2)^2} < \frac{|S^{d-2}|^2}{(d-1)^2|S^{d-1}|^2} < \frac{(d+1)^2}{4d^2(d+2)}, \quad (\text{C.2})$$

where $\frac{|S^{d-2}|}{d-1}$ is defined as 1 when $d = 1$.

Proof The inequality can be verified by a straightforward calculation for $d \leq 6$.

Next, we show that $\frac{|S^{d-2}|^2}{|S^{d-1}|^2} < \frac{(d^2-1)^2}{4d^2(d+2)}$ for $d > 6$. Note that $\frac{|S^{d-2}|^2}{|S^{d-1}|^2} = \frac{\Gamma(\frac{d}{2})^2}{\pi\Gamma(\frac{d-1}{2})^2}$. Hence, it suffices to prove $\frac{\Gamma(\frac{d}{2})^2}{\Gamma(\frac{d-1}{2})^2} < \frac{\pi(d^2-1)^2}{4d^2(d+2)}$. In Kershaw (1983), it is proved that for all $x > 0$ and $0 < s < 1$,

$$(x + \frac{s}{2})^{1-s} < \frac{\Gamma(x+1)}{\Gamma(x+s)} < e^{(1-s)\psi(x+\frac{1+s}{2})}, \quad (\text{C.3})$$

where $\psi(y) = \frac{\Gamma'(y)}{\Gamma(y)}$. Choose $x = \frac{d}{2} - 1$ and $s = \frac{1}{2}$, then $\frac{\Gamma(\frac{d}{2})^2}{\Gamma(\frac{d-1}{2})^2} < e^{\psi(\frac{d}{2}-\frac{1}{4})}$. Hence, it suffice to show that

$$e^{\psi(\frac{d}{2}+\frac{1}{4})} < \frac{\pi(d^2-1)^2}{4d^2(d+2)}. \quad (\text{C.4})$$

Actually, we have $e^{\psi(y)} < y$ for any postive y . The conclusion follows by verifying $\frac{d}{2} + \frac{1}{4} < \frac{\pi(d^2-1)^2}{4d^2(d+2)}$ for $d > 6$.

At last, we show that $\frac{(d^2-1)^2(d+3)}{8d^2(d+2)^2} < \frac{|S^{d-2}|^2}{|S^{d-1}|^2}$, which is equivalent to $\frac{\pi(d^2-1)^2(d+3)}{8d^2(d+2)^2} < \frac{\Gamma(\frac{d}{2})^2}{\Gamma(\frac{d-1}{2})^2}$. By (C.3), with $x = \frac{d}{2} - 1$ and $s = \frac{1}{2}$, we have $\frac{d}{2} - \frac{3}{4} < \frac{\Gamma(\frac{d}{2})^2}{\Gamma(\frac{d-1}{2})^2}$. The conclusion follows by verifying $\frac{\pi(d^2-1)^2(d+3)}{8d^2(d+2)^2} < \frac{d}{2} - \frac{3}{4}$ for d large. ■

We calculate some major ingredients that we are going to use in the proof of the main theorem. Specifically, we calculate the first two order terms in $\mathbb{E}[\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$, $\mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$, and the first two order terms in the tangent component of $\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$ and $\mathbb{E}[(\iota(X) - \iota(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$. This long Lemma is the generalization of (Wu and Wu, 2018, Lemma B.5) to the boundary. In particularly, when $x \notin M_\varepsilon$, we recover (Wu and Wu, 2018, Lemma B.5).

Lemma 30 Fix $x \in M$ and $f \in C^3(M)$. When $\varepsilon > 0$ is sufficiently small, the following expansions hold.

1. $\mathbb{E}[\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$ satisfies

$$\mathbb{E}[\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] = P(x)\mu_0(x, \varepsilon) + \partial_d P(x)\mu_{e_d}(x, \varepsilon) + O(\varepsilon^{d+2}).$$

2. $\mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$ satisfies

$$\begin{aligned} \mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] &= P(x)\partial_d f(x)\mu_{e_d}(x, \varepsilon) \\ &\quad + \sum_{i=1}^d \left(\frac{P(x)}{2}\partial_{ii}^2 f(x) + \partial_i f(x)\partial_i P(x) \right) \mu_{2e_i}(x, \varepsilon) + O(\varepsilon^{d+3}). \end{aligned}$$

3. The vector $\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$ satisfies

$$\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] = [v_1, v_2],$$

where

$$\begin{aligned} v_1 &= P(x)\mu_{e_d}(x, \varepsilon)J_{p,d}^\top e_d + \sum_{i=1}^d (\partial_i P(x)\mu_{2e_i}(x, \varepsilon))J_{p,d}^\top e_i + O(\varepsilon^{d+3}) \\ v_2 &= \frac{P(x)}{2} \sum_{i=1}^d \mathfrak{N}_{ii}(x)\mu_{2e_i}(x, \varepsilon) + O(\varepsilon^{d+3}). \end{aligned}$$

4. The vector $\mathbb{E}[(\iota(X) - \iota(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$ satisfies

$$\mathbb{E}[(\iota(X) - \iota(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] = [\![v_1, v_2]\!],$$

where

$$\begin{aligned} v_1 &= P(x) \sum_{i=1}^d (\partial_i f(x)\mu_{2e_i}(x, \varepsilon)) J_{p,d}^\top e_i \\ &\quad + \sum_{i=1}^{d-1} [\partial_i f(x)\partial_d P(x) + \partial_d f(x)\partial_i P(x) + P(x)\partial_{id}^2 f(x)]\mu_{2e_i+e_d}(x, \varepsilon) J_{p,d}^\top e_i \\ &\quad + \sum_{i=1}^d \left([\partial_i f(x)\partial_i P(x) + \frac{P(x)}{2}\partial_{ii}^2 f(x)]\mu_{2e_i+e_d}(x, \varepsilon) \right) J_{p,d}^\top e_d + O(\varepsilon^{d+4}), \\ v_2 &= P(x) \sum_{i=1}^{d-1} \partial_i f(x)\mathfrak{N}_{id}(x)\mu_{2e_i+e_d}(x, \varepsilon) \\ &\quad + \frac{P(x)}{2} \partial_d f(x) \sum_{i=1}^d \mathfrak{N}_{ii}(x)\mu_{2e_i+e_d}(x, \varepsilon) + O(\varepsilon^{d+4}). \end{aligned}$$

Proof We use the extensions \tilde{M} and $\tilde{\iota}$ introduced in Section 3.2. For any $x \in M$, let \exp_x be the exponential map of \tilde{M} . First, we calculate $\mathbb{E}[\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$.

$$\begin{aligned} &\mathbb{E}[\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] \tag{C.5} \\ &= \int_{D_\varepsilon(x)} \left(P(x) + \sum_{i=1}^d \partial_i P(x)u_i + O(u^2) \right) \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i, j)u_i u_j + O(u^3) \right) du \\ &= P(x) \int_{\tilde{D}_\varepsilon(x)} du + \int_{\tilde{D}_\varepsilon(x)} \sum_{i=1}^d \partial_i P(x)u_i du + O(\varepsilon^{d+2}) \\ &= P(x)\mu_0(x, \varepsilon) + \partial_d P(x)\mu_{e_d}(x, \varepsilon) + O(\varepsilon^{d+2}), \end{aligned}$$

where the second equality holds by applying Corollary 25 that the error of changing domain from $D_\varepsilon(x)$ to $\tilde{D}_\varepsilon(x)$ is of order ε^{d+2} . We use Lemma 27 in the last step. Note that $P(x)$ is bounded away from 0.

Second, we calculate $\mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$. Note that when ε is sufficiently small, we have

$$f \circ \exp_x(u) - f(x) = \sum_{i=1}^d \partial_i f(x)u_i + \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 f(x)u_i u_j + O(u^3), \tag{C.6}$$

which is of order ε for $u \in D_\varepsilon(x)$. By a direct expansion, we have

$$\begin{aligned} & \mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(t(x))}(t(X))] \\ &= \int_{D_\varepsilon(x)} \left(\sum_{i=1}^d \partial_i f(x) u_i + \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 f(x) u_i u_j + O(u^3) \right) (P(x) + \sum_{i=1}^d \partial_i P(x) u_i + O(u^2)) \right. \\ &\quad \times \left. \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i,j) u_i u_j + O(u^3) \right) du, \right) \end{aligned} \tag{C.7}$$

which by Corollary 25 and Lemma 27 becomes

$$\begin{aligned} & \int_{D_\varepsilon(x)} \left[P(x) \sum_{i=1}^d \partial_i f(x) u_i + \frac{P(x)}{2} \sum_{i,j=1}^d \partial_{ij}^2 f(x) u_i u_j + \sum_{i=1}^d \partial_i f(x) u_i \sum_{j=1}^d \partial_j P(x) u_j + O(u^3) \right] du \\ &= P(x) \partial_d f(x) \int_{\tilde{D}_\varepsilon(x)} u_d du + \sum_{i=1}^d \left(\frac{P(x)}{2} \partial_{ii}^2 f(x) + \partial_i f(x) \partial_i P(x) \right) \int_{\tilde{D}_\varepsilon(x)} u_i^2 du + O(\varepsilon^{d+3}) \\ &= P(x) \partial_d f(x) \mu_{e_d}(x, \varepsilon) + \sum_{i=1}^d \left(\frac{P(x)}{2} \partial_{ii}^2 f(x) + \partial_i f(x) \partial_i P(x) \right) \mu_{2e_i}(x, \varepsilon) + O(\varepsilon^{d+3}). \end{aligned}$$

Note that the leading term in the integral is of order ε , so the error of changing the domain from $D_\varepsilon(x)$ to $\tilde{D}_\varepsilon(x)$ is of order ε^{d+3} .

Third, by a direct expansion, we have

$$\begin{aligned} & \mathbb{E}[(t(X) - t(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(t(x))}(t(X))] \\ &= \int_{D_\varepsilon(x)} (\tilde{t}_* u + \frac{1}{2} \text{II}_x(u, u) + O(u^3)) (P(x) + \sum_{i=1}^d \partial_i P(x) u_i + O(u^2)) \\ &\quad \times \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i,j) u_i u_j + O(u^3) \right) du, \end{aligned} \tag{C.8}$$

which is a vector in \mathbb{R}^p . We then find the tangential part and the normal part of $\mathbb{E}[(t(X) - t(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(t(x))}(t(X))]$ respectively. The tangential part is

$$\begin{aligned} & \int_{D_\varepsilon(x)} (\tilde{t}_* u + O(u^3)) (P(x) + \sum_{i=1}^d \partial_i P(x) u_i + O(u^2)) \\ &\quad \times \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i,j) u_i u_j + O(u^3) \right) du \\ &= \int_{\tilde{D}_\varepsilon(x)} (\tilde{t}_* u + O(u^3)) (P(x) + \sum_{i=1}^d \partial_i P(x) u_i + O(u^2)) \\ &\quad \times \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i,j) u_i u_j + O(u^3) \right) du + O(\varepsilon^{d+3}), \end{aligned} \tag{C.9}$$

where the equality holds by Corollary 25. Similarly, by Corollary 25, the normal part is

$$\begin{aligned}
& \int_{D_\varepsilon(x)} \left(\frac{1}{2} \mathbb{I}_x(u, u) + O(u^3) \right) (P(x) + \sum_{i=1}^d \partial_i P(x) u_i + O(u^2)) \\
& \quad \times \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i, j) u_i u_j + O(u^3) \right) du \\
& = \int_{\tilde{D}_\varepsilon(x)} \left(\frac{1}{2} \mathbb{I}_x(u, u) + O(u^3) \right) (P(x) + \sum_{i=1}^d \partial_i P(x) u_i + O(u^2)) \\
& \quad \times \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i, j) u_i u_j + O(u^3) \right) du + O(\varepsilon^{d+4})
\end{aligned} \tag{C.10}$$

since the leading term $P(x)\mathbb{I}_x(u, u)$ is of order ε^2 on $D_\varepsilon(x)$. As a result, by putting the tangent part and normal part together, $\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] = [\![v_1, v_2]\!]$, where

$$\begin{aligned}
v_1 &= J_{p,d}^\top \left[P(x) \int_{\tilde{D}_\varepsilon(x)} \tilde{\iota}_* u du + \int_{\tilde{D}_\varepsilon(x)} \tilde{\iota}_* u \sum_{i=1}^d \partial_i P(x) u_i du + O(\varepsilon^{d+3}) \right] \\
&= \left(P(x) \int_{\tilde{D}_\varepsilon(x)} u_d du \right) J_{p,d}^\top e_d + \sum_{i=1}^d \left(\partial_i P(x) \int_{\tilde{D}_\varepsilon(x)} u_i^2 du \right) J_{p,d}^\top e_i + O(\varepsilon^{d+3}) \\
&= P(x) \mu_{e_d}(x, \varepsilon) J_{p,d}^\top e_d + \sum_{i=1}^d \partial_i P(x) \mu_{2e_i}(x, \varepsilon) J_{p,d}^\top e_i + O(\varepsilon^{d+3})
\end{aligned} \tag{C.11}$$

and

$$v_2 = \frac{P(x)}{2} \bar{J}_{p,p-d}^\top \int_{\tilde{D}_\varepsilon(x)} \mathbb{I}_x(u, u) du + O(\varepsilon^{d+3}) = \frac{P(x)}{2} \sum_{i=1}^d \mathfrak{N}_{ii}(x) \mu_{2e_i}(x, \varepsilon) + O(\varepsilon^{d+3}).$$

Finally, we evaluate $\mathbb{E}[(\iota(X) - \iota(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$ and then find the tangential part and the normal part. By a direct expansion,

$$\begin{aligned}
& \mathbb{E}[(\iota(X) - \iota(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] \\
&= \int_{D_\varepsilon(x)} (\tilde{\iota}_* u + \frac{1}{2} \mathbb{I}_x(u, u) + O(u^3)) \left(\sum_{i=1}^d \partial_i f(x) u_i + \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 f(x) u_i u_j + O(u^3) \right) \\
& \quad \times \left(P(x) + \sum_{i=1}^d \partial_i P(x) u_i + O(u^2) \right) \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i, j) u_i u_j + O(u^3) \right) du.
\end{aligned} \tag{C.12}$$

The tangential part is

$$\begin{aligned}
& \int_{D_\varepsilon(x)} (\tilde{\iota}_* u + O(u^3)) \left(\sum_{i=1}^d \partial_i f(x) u_i + \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 f(x) u_i u_j + O(u^3) \right) \\
& \quad \times \left(P(x) + \sum_{i=1}^d \partial_i P(x) u_i + O(u^2) \right) \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i, j) u_i u_j + O(u^3) \right) du.
\end{aligned} \tag{C.13}$$

The leading term $P(x)\tilde{\iota}_*u\sum_{i=1}^d \partial_i f(x)u_i$ is of order ε^2 on $D_\varepsilon(x)$, therefore the error of changing domain from $D_\varepsilon(x)$ to $\tilde{D}_\varepsilon(x)$ is of order ε^{d+4} . The normal part is

$$\begin{aligned} & \int_{D_\varepsilon(x)} \left(\frac{1}{2}\mathbb{I}_x(u,u) + O(u^3) \right) \left(\sum_{i=1}^d \partial_i f(x)u_i + \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 f(x)u_i u_j + O(u^3) \right) \\ & \quad \times \left(P(x) + \sum_{i=1}^d \partial_i P(x)u_i + O(u^2) \right) \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i,j)u_i u_j + O(u^3) \right) du. \end{aligned} \quad (\text{C.14})$$

The leading term $P(x)\mathbb{I}_x(u,u)\sum_{i=1}^d \partial_i f(x)u_i$ is of order ε^3 on $D_\varepsilon(x)$. Therefore, the error of changing domain from $D_\varepsilon(x)$ to $\tilde{D}_\varepsilon(x)$ is of order ε^{d+5} . Putting the above together, $\mathbb{E}[(\iota(X) - \iota(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] = [\![v_1, v_2]\!]$, where by the symmetry of $\tilde{D}_\varepsilon(x)$ we have

$$\begin{aligned} v_1 &= J_{p,d}^\top \left[P(x) \int_{\tilde{D}_\varepsilon(x)} \tilde{\iota}_*u \sum_{i=1}^d \partial_i f(x)u_i du + \int_{\tilde{D}_\varepsilon(x)} \tilde{\iota}_*u \sum_{i=1}^d \partial_i f(x)u_i \sum_{j=1}^d \partial_j P(x)u_j du \right. \\ & \quad \left. + \frac{P(x)}{2} \int_{\tilde{D}_\varepsilon(x)} \tilde{\iota}_*u \sum_{i,j=1}^d \partial_{ij}^2 f(x)u_i u_j du + O(\varepsilon^{d+4}) \right] \\ &= P(x) \sum_{i=1}^d \left(\partial_i f(x) \int_{\tilde{D}_\varepsilon(x)} u_i^2 du \right) J_{p,d}^\top e_i \\ & \quad + \sum_{i=1}^{d-1} \left[\partial_i f(x) \partial_d P(x) + \partial_d f(x) \partial_i P(x) + P(x) \partial_{id}^2 f(x) \right] \int_{\tilde{D}_\varepsilon(x)} u_i^2 u_d du J_{p,d}^\top e_i \\ & \quad + \sum_{i=1}^d \left(\left[\partial_i f(x) \partial_i P(x) + \frac{P(x)}{2} \partial_{ii}^2 f(x) \right] \int_{\tilde{D}_\varepsilon(x)} u_i^2 u_d du \right) J_{p,d}^\top e_d + O(\varepsilon^{d+4}) \\ &= P(x) \sum_{i=1}^d \partial_i f(x) \mu_{2e_i}(x, \varepsilon) J_{p,d}^\top e_i \\ & \quad + \sum_{i=1}^{d-1} \left[\partial_i f(x) \partial_d P(x) + \partial_d f(x) \partial_i P(x) + P(x) \partial_{id}^2 f(x) \right] \mu_{2e_i+e_d}(x, \varepsilon) J_{p,d}^\top e_i \\ & \quad + \sum_{i=1}^d \left[\partial_i f(x) \partial_i P(x) + \frac{P(x)}{2} \partial_{ii}^2 f(x) \right] \mu_{2e_i+e_d}(x, \varepsilon) J_{p,d}^\top e_d + O(\varepsilon^{d+4}), \end{aligned} \quad (\text{C.15})$$

and

$$\begin{aligned} v_2 &= \frac{P(x)}{2} \bar{J}_{p,p-d}^\top \sum_{i=1}^d \partial_i f(x) \int_{\tilde{D}_\varepsilon(x)} \mathbb{I}_x(u,u)u_i du + O(\varepsilon^{d+4}) \\ &= P(x) \sum_{i=1}^{d-1} \partial_i f(x) \mathfrak{N}_{id}(x) \mu_{2e_i+e_d}(x, \varepsilon) + \frac{P(x)}{2} \partial_d f(x) \sum_{i=1}^d \mathfrak{N}_{ii}(x) \mu_{2e_i+e_d}(x, \varepsilon) + O(\varepsilon^{d+4}). \end{aligned}$$

■

Appendix D. Structure of the local covariance matrix under the manifold setup

In this section we provide detailed analysis for the local covariance matrix $C_x = \mathbb{E}[(\iota(X) - \iota(x))(\iota(X) - \iota(x))^\top \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$. This Lemma could be viewed as the generalization of (Wu and Wu, 2018, Proposition 3.2) in the sense that when $x \notin M_\varepsilon$, the result is reduced to that of (Wu and Wu, 2018, Proposition 3.2). To handle the boundary effect, we only need to calculate the first two order terms in eigenvalues and orthonormal eigenvectors of C_x .

Lemma 31 *Fix $x \in M$. Suppose that $\text{rank}(C_x) = r$, there is a choice of e_{d+1}, \dots, e_p so that we have $e_i^\top \mathbb{I}_x(e_j, e_j) = 0$ for all $i = r+1, \dots, p$ and $j = 1, \dots, d$. We have*

$$C_x = P(x) \begin{bmatrix} M^{(0)}(x, \varepsilon) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} M^{(11)}(x, \varepsilon) & M^{(12)}(x, \varepsilon) & 0 \\ M^{(21)}(x, \varepsilon) & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} O(\varepsilon^{d+4}) & O(\varepsilon^{d+4}) & 0 \\ O(\varepsilon^{d+4}) & M^{(3)}(x, \varepsilon) + O(\varepsilon^{d+5}) & 0 \end{bmatrix}, \quad (\text{D.1})$$

where $M^{(0)}$ is a $d \times d$ diagonal matrix with the m -th diagonal entry $\mu_{2e_m}(x, \varepsilon)$. $M^{(11)}$ is a symmetric $d \times d$ matrix. $M^{(12)} \in \mathbb{R}^{d \times (r-d)}$. $M^{(21)} = M^{(12)^\top}$. In particular, when $x \notin M_\varepsilon$, $\begin{bmatrix} M^{(11)}(x, \varepsilon) & M^{(12)}(x, \varepsilon) & 0 \\ M^{(21)}(x, \varepsilon) & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0$. $M^{(3)}(x, \varepsilon)$ is diagonal $(p-d) \times (p-d)$ matrix and is of order ε^{d+4} . The first d eigenvalues of C_x are

$$\lambda_i = P(x)\mu_{2e_i}(x, \varepsilon) + \lambda_i^{(1)}(x, \varepsilon) + O(\varepsilon^{d+4}), \quad (\text{D.2})$$

where $i = 1, \dots, d$. And $\lambda_i^{(1)}(x, \varepsilon) = O(\varepsilon^{d+3})$. If $x \notin M_\varepsilon$, $\lambda_i^{(1)}(x, \varepsilon) = 0$. The last $p-d$ eigenvalues of C_x are $\lambda_i = O(\varepsilon^{d+4})$, where $i = d+1, \dots, p$.

The corresponding orthonormal eigenvector matrix is

$$X(x, \varepsilon) = X(x, 0) + X(x, 0)S(x, \varepsilon) + O(\varepsilon^2), \quad (\text{D.3})$$

where

$$X(x, 0) = \begin{bmatrix} X_1(x) & 0 & 0 \\ 0 & X_2(x) & 0 \\ 0 & 0 & X_3(x) \end{bmatrix}, \quad S(x, \varepsilon) = \begin{bmatrix} S_{11}(x, \varepsilon) & S_{12}(x, \varepsilon) & S_{13}(x, \varepsilon) \\ S_{21}(x, \varepsilon) & S_{22}(x, \varepsilon) & S_{23}(x, \varepsilon) \\ S_{31}(x, \varepsilon) & S_{32}(x, \varepsilon) & S_{33}(x, \varepsilon) \end{bmatrix}, \quad (\text{D.4})$$

$X_1 \in O(d)$, $X_2 \in O(r-d)$ and $X_3 \in O(p-r)$. The matrix $S(x, \varepsilon)$ is divided into blocks the same as $X(x, 0)$. Moreover, $S(x, \varepsilon)$ is an antisymmetric matrix with 0 on the diagonal entries. In particular, if $x \notin M_\varepsilon$, $S(x, \varepsilon) = 0$.

The proof is essentially the same as that of (Wu and Wu, 2018, Proposition 3.2), except that when x is close to the boundary, the integral domain is no longer symmetric.

Proof By definition, the (m, n) -th entry of C_x is

$$e_m^\top C_x e_n = \int_{D_\varepsilon(x)} (\iota(y) - \iota(x))^\top e_m (\iota(y) - \iota(x))^\top e_n P(y) dV(y). \quad (\text{D.5})$$

By the expression

$$\tilde{\iota} \circ \exp_x(u) - \tilde{\iota}(x) = \tilde{\iota}_* u + \frac{1}{2} \mathbb{I}_x(u, u) + O(u^3), \quad (\text{D.6})$$

we have

$$\begin{aligned} & (\iota(y) - \iota(x))^\top e_m (\iota(y) - \iota(x))^\top e_n \\ &= (e_m^\top \tilde{\iota}_* u)(e_n^\top \tilde{\iota}_* u) + \frac{1}{2} (e_m^\top \tilde{\iota}_* u)(e_n^\top \mathbb{I}_x(u, u)) + \frac{1}{2} (e_m^\top \mathbb{I}_x(u, u))(e_n^\top \tilde{\iota}_* u) + O(u^4). \end{aligned}$$

Thus, (D.5) is reduced to

$$\begin{aligned} e_m^\top C_x e_n &= \int_{D_\varepsilon(x)} ((e_m^\top \tilde{\iota}_* u)(e_n^\top \tilde{\iota}_* u) + \frac{1}{2} (e_m^\top \tilde{\iota}_* u)(e_n^\top \mathbb{I}_x(u, u)) + \frac{1}{2} (e_m^\top \mathbb{I}_x(u, u))(e_n^\top \tilde{\iota}_* u) \\ &\quad + \frac{1}{4} [e_m^\top \mathbb{I}_x(u, u) e_n^\top \mathbb{I}_x(u, u)] + O(u^5)) \\ &\quad \times (P(x) + \nabla_u P(x) + O(u^2)) \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i, j) u_i u_j + O(u^3)\right) du. \end{aligned} \quad (\text{D.7})$$

For $1 \leq m, n \leq d$, $(e_m^\top \tilde{\iota}_* u)(e_n^\top \tilde{\iota}_* u) = u_m u_n$. Moreover, $e_n^\top \mathbb{I}_x(u, u)$ and $e_m^\top \mathbb{I}_x(u, u)$ are zero, so

$$\begin{aligned} & e_m^\top C_x e_n \\ &= \int_{D_\varepsilon(x)} (u_m u_n + O(u^4)) (P(x) + \nabla_u P(x) + O(u^2)) \\ &\quad \times \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i, j) u_i u_j + O(u^3)\right) du \\ &= P(x) \int_{\tilde{D}_\varepsilon(x)} u_m u_n du + \int_{\tilde{D}_\varepsilon(x)} u_m u_n \sum_{k=1}^d u_k \partial_k P(x) du + O(\varepsilon^{d+4}). \end{aligned} \quad (\text{D.8})$$

where we use Lemma 25 to handle the error of changing domain from $D_\varepsilon(x)$ to $\tilde{D}_\varepsilon(x)$, which is $O(\varepsilon^{d+4})$. By the symmetry of domain $\tilde{D}_\varepsilon(x)$, if $1 \leq m = n \leq d$,

$$M_{m,n}^{(0)} = \int_{\tilde{D}_\varepsilon(x)} u_m^2 du = \mu_{2e_m}(x, \varepsilon) \quad (\text{D.9})$$

and $M_{m,n}^{(0)}$ is 0 otherwise.

Next,

$$M_{m,n}^{(11)} = \int_{\tilde{D}_\varepsilon(x)} u_m u_n \sum_{k=1}^d u_k \partial_k P(x) du \quad (\text{D.10})$$

So, by the symmetry of domain $\tilde{D}_\varepsilon(x)$, we have

$$M_{m,n}^{(11)} = \begin{cases} \partial_d P(x) \mu_{2e_m+e_d}(x, \varepsilon) & 1 \leq m = n \leq d, \\ \partial_n P(x) \mu_{2e_n+e_d}(x, \varepsilon) & m = d, 1 \leq n \leq d, \\ \partial_m P(x) \mu_{2e_m+e_d}(x, \varepsilon) & n = d, 1 \leq m \leq d, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{D.11})$$

For $d+1 \leq m \leq p$ and $d+1 \leq n \leq p$, we have

$$\begin{aligned} e_m^\top C_x e_n &= \int_{D_\varepsilon(x)} \left(\frac{1}{4} [e_m^\top \mathbb{I}_x(u, u) e_n^\top \mathbb{I}_x(u, u)] + O(u^5) \right) (P(x) + O(u)) (1 + O(u)) du \\ &= \frac{P(x)}{4} \int_{\tilde{D}_\varepsilon(x)} e_m^\top \mathbb{I}_x(u, u) e_n^\top \mathbb{I}_x(u, u) du + O(\varepsilon^{d+5}). \end{aligned} \quad (\text{D.12})$$

Hence, we have

$$M_{m-d, n-d}^{(3)}(x, \varepsilon) = \frac{P(x)}{4} \int_{\tilde{D}_\varepsilon(x)} e_m^\top \mathbb{I}_x(u, u) e_n^\top \mathbb{I}_x(u, u) du. \quad (\text{D.13})$$

Since $M_{m-d, m-d}^{(3)}(x, \varepsilon)$ is symmetric, we can choose e_{d+1}, \dots, e_p so that it is diagonal. Then $M_{m-d, m-d}^{(3)}(x, \varepsilon) = 0$ implies

$$\int_{\tilde{D}_\varepsilon(x)} (e_m^\top \mathbb{I}_x(u, u))^2 du = 0. \quad (\text{D.14})$$

Note that since $e_m^\top \mathbb{I}_x(u, u)$ is a quadratic form of u , we have $e_m^\top \mathbb{I}_x(u, u) = 0$. Since C_x has rank r ,

$M_{m-d, m-d}^{(3)}(x, \varepsilon) = 0$ for $m = r+1, \dots, p$, and $e_m^\top \mathbb{I}_x(e_i, e_j) = 0$ for $m = r+1, \dots, p$ and $i, j = 1, \dots, d$.

For $1 \leq m \leq d$ and $n \geq d$,

$$\begin{aligned} e_m^\top C_x e_n &= \int_{D_\varepsilon(x)} \left(\frac{1}{2} (e_m^\top \tilde{\mathbf{l}}_* u) (e_n^\top \mathbb{I}_x(u, u)) + O(u^4) \right) (P(x) + \nabla_u P(x) + O(u^2)) \\ &\quad \times \left(1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i, j) u_i u_j + O(u^3) \right) du \\ &= \frac{P(x)}{2} \int_{D_\varepsilon(x)} u_m (e_n^\top \mathbb{I}_x(u, u)) du + O(\varepsilon^{d+4}). \end{aligned} \quad (\text{D.15})$$

We use Lemma 25 to handle the error of changing domain from $D_\varepsilon(x)$ to $\tilde{D}_\varepsilon(x)$, which is $O(\varepsilon^{d+5})$. Hence, for $1 \leq m \leq d$ and $d+1 \leq n \leq r$,

$$M^{(12)}(x)_{m, n-d} = \frac{P(x)}{2} \int_{\tilde{D}_\varepsilon(x)} u_m (e_n^\top \mathbb{I}_x(u, u)) du. \quad (\text{D.16})$$

By symmetry of C_x , we have $M^{(21)} = M^{(12)^\top}$.

For $1 \leq m \leq d$ and $r+1 \leq n \leq p$,

$$e_m^\top C_x e_n = \frac{P(x)}{2} \int_{\tilde{D}_\varepsilon(x)} u_m (e_n^\top \mathbb{I}_x(u, u)) du + O(\varepsilon^{d+4}) = O(\varepsilon^{d+4}). \quad (\text{D.17})$$

For $1 \leq n \leq d$ and $r+1 \leq m \leq p$, $e_m^\top C_x e_n = O(\varepsilon^{d+4})$ by symmetry.

Based on Lemma 27, $\begin{bmatrix} M^{(0)}(x, \varepsilon) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ is of order ε^{d+2} and $\begin{bmatrix} M^{(11)}(x, \varepsilon) & M^{(12)}(x, \varepsilon) & 0 \\ M^{(21)}(x, \varepsilon) & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

is of order ε^{d+3} . Note that the entries of $\begin{bmatrix} M^{(11)}(x, \varepsilon) & M^{(12)}(x, \varepsilon) & 0 \\ M^{(21)}(x, \varepsilon) & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ are integrals of odd-order

polynomials over $\tilde{D}_\varepsilon(x)$. Hence, the matrix is 0 when $x \notin M_\varepsilon$. By applying the perturbation theory (see, for example, (Wu and Wu, 2018, Appendix A)), the first d eigenvalues of C_x are

$$\lambda_i = P(x)\mu_{2e_i}(x, \varepsilon) + \lambda_i^{(1)}(x, \varepsilon) + O(\varepsilon^{d+4}), \quad (\text{D.18})$$

for $i = 1, \dots, d$ and any $x \in M$, where $\{\lambda_i^{(1)}(x, \varepsilon)\}$ are of order ε^{d+3} . The calculation of $\{\lambda_i^{(1)}(x, \varepsilon)\}$ depends on $M^{(11)}(x, \varepsilon)$ and whether $\mu_{2e_i}(x, \varepsilon)$ are the same. Moreover, $\lambda_i = O(\varepsilon^{d+4})$ for $i = d+1, \dots, p$.

Suppose that $\text{rank}(C_x) = r$, based on the perturbation theory (see, for example, (Wu and Wu, 2018, Appendix A)), the orthonormal eigenvector matrix of C_x is in the form

$$X(x, \varepsilon) = \begin{bmatrix} X_1(x) & 0 & 0 \\ 0 & X_2(x) & 0 \\ 0 & 0 & X_3(x) \end{bmatrix} + \begin{bmatrix} X_1(x) & 0 & 0 \\ 0 & X_2(x) & 0 \\ 0 & 0 & X_3(x) \end{bmatrix} S(x, \varepsilon) + O(\varepsilon^2), \quad (\text{D.19})$$

where $X_1(x) \in O(d)$, $X_2(x) \in O(r-d)$ and $X_3(x) \in O(p-r)$. And

$$S(x, \varepsilon) = \begin{bmatrix} S_{11}(x, \varepsilon) & S_{12}(x, \varepsilon) & S_{13}(x, \varepsilon) \\ S_{21}(x, \varepsilon) & S_{22}(x, \varepsilon) & S_{23}(x, \varepsilon) \\ S_{31}(x, \varepsilon) & S_{32}(x, \varepsilon) & S_{33}(x, \varepsilon) \end{bmatrix}.$$

$S(x, \varepsilon)$ is an antisymmetric matrix with 0 on the diagonal entries. It is of order ε and depends on those terms of C_x of order ε^{d+2} , order ε^{d+3} and higher orders. In particular,

$$X_1(x)S_{12}(x, \varepsilon) = -[P(x)M^{(0)}(x, \varepsilon)]^{-1}M^{(12)}(x, \varepsilon)X_2.$$

And a straightforward calculation shows that

$$e_i^\top J_{p,d}X_1(x)S_{12}(x, \varepsilon) = -\frac{\mu_{2e_i+e_d}(x, \varepsilon)}{\mu_{2e_i}(x, \varepsilon)}\mathfrak{N}_{id}^\top(x)J_{p-d,r-d}X_2(x), \quad (\text{D.20})$$

for $i = 1, \dots, d-1$, and

$$e_d^\top J_{p,d}X_1(x)S_{12}(x, \varepsilon) = -\frac{1}{2}\sum_{j=1}^d \frac{\mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)}\mathfrak{N}_{jj}^\top(x)J_{p-d,r-d}X_2(x). \quad (\text{D.21})$$

If $x \notin M_\varepsilon$, $S(x, \varepsilon) = 0$. Moreover, if among first $\mu_{2e_1}, \dots, \mu_{2e_d}$, there are $1 \leq k \leq d$ distinct ones, then there is a choice of the basis in the tangent space of M so that

$$X_1(x) = \begin{bmatrix} X_1^{(1)}(x) & 0 & \cdots & 0 \\ 0 & X_1^{(2)}(x) & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & X_1^{(k)}(x) \end{bmatrix}, \quad (\text{D.22})$$

where each $X_1^{(i)}(x)$ is an orthogonal matrix corresponding to the same of μ_{2e_i} . The conclusion follows. \blacksquare

Appendix E. Analysis on the augmented vector $\mathbf{T}(x)$

We now calculate $\mathbf{T}(x)$. For our purpose, we need an asymptotic expansion up to the first two orders for the tangent component of $\mathbf{T}(x)$ and to the first order for the normal component when ε is sufficiently small.

Lemma 32 $\mathbf{T}(x) = [\![v_1^{(-1)} + v_{1,1}^{(0)} + v_{1,2}^{(0)} + v_{1,3}^{(0)} + v_{1,4}^{(0)}, v_2^{(-1)}]\!] + [\![O(\varepsilon), O(1)]\!]$, where

$$\begin{aligned} v_1^{(-1)} &= \frac{\mu_{e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} J_{p,d}^\top e_d, \\ v_{1,1}^{(0)} &= \frac{\nabla P(x)}{P(x)}, \\ v_{1,2}^{(0)} &= -\frac{\varepsilon^{d+3} \mu_{e_d}(x, \varepsilon)}{P(x)(\mu_{2e_d}(x, \varepsilon))^2} J_{p,d}^\top e_d, \\ v_{1,3}^{(0)} &= -\sum_{i=1}^d \frac{\partial_i P(x) \mu_{e_d}(x, \varepsilon) \mu_{2e_i+e_d}(x, \varepsilon)}{P(x) \mu_{2e_i}(x, \varepsilon) \mu_{2e_d}(x, \varepsilon)} J_{p,d}^\top e_i, \\ v_{1,4}^{(0)} &= \frac{P(x)}{2\varepsilon^{d+3}} \sum_{i=1}^{d-1} \sum_{j=1}^d \left[\left(\frac{\mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} - \mu_{2e_j}(x, \varepsilon) \right) \frac{\mu_{2e_i+e_d}(x, \varepsilon)}{\mu_{2e_i}(x, \varepsilon)} \mathfrak{N}_{jj}^\top(x) \right] \mathfrak{N}_{id}(x) J_{p,d}^\top e_i \\ &\quad + \frac{P(x)}{4\varepsilon^{d+3}} \sum_{i=1}^d \sum_{j=1}^d \left[\left(\frac{\mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} - \mu_{2e_j}(x, \varepsilon) \right) \frac{\mu_{2e_i+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \mathfrak{N}_{jj}^\top(x) \right] \mathfrak{N}_{ii}(x) J_{p,d}^\top e_d, \end{aligned} \tag{E.1}$$

and

$$v_2^{(-1)} = \frac{P(x)}{2\varepsilon^{d+3}} \sum_{j=1}^d \left(\mu_{2e_j}(x, \varepsilon) - \frac{\mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \right) \mathfrak{N}_{jj}. \tag{E.2}$$

Note that by Lemma 27, $v_1^{(-1)}$ is of order ε^{-1} when $x \in M_\varepsilon$ and 0 when $x \notin M_\varepsilon$; $v_{1,2}^{(0)}$ is of order 1 since $\mu_{e_d}(x, \varepsilon)$ is of order ε^{d+1} and $\mu_{2e_i}(x, \varepsilon)$ is of order ε^{d+2} for $i = 1, \dots, d$. Moreover, when $x \notin M_\varepsilon$, we have $\mu_{e_d}(x, \varepsilon) = 0$ and $\mu_{2e_i+e_d}(x, \varepsilon) = 0$. Hence, $v_{1,2}^{(0)} = 0, v_{1,3}^{(0)} = 0$ and $v_{1,4}^{(0)} = 0$. Similarly, $v_2^{(-1)}$ is of order ε^{-1} .

Proof Recall that

$$\mathbf{T}(x)^\top = \sum_{i=1}^r \frac{\mathbb{E}[(\mathbf{l}(X) - \mathbf{l}(x)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\mathbf{l}(x))}(\mathbf{l}(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}}. \tag{E.3}$$

To show the proof, we evaluate the terms in $\mathbf{T}(x)$ one by one.

Based on Lemma 31, the first d eigenvalues are $\lambda_i = P(x) \mu_{2e_i}(x, \varepsilon) + \lambda_i^{(1)}(x, \varepsilon) + O(\varepsilon^{d+4})$, where $i = 1, \dots, d$, and the corresponding eigenvectors are

$$\beta_i = \begin{bmatrix} X_1(x) J_{p,d}^\top e_i \\ 0_{(p-d) \times 1} \end{bmatrix} + \begin{bmatrix} X_1(x) S_{11}(x, \varepsilon) J_{p,d}^\top e_i + O(\varepsilon^2) \\ O(\varepsilon) \end{bmatrix}, \tag{E.4}$$

where $X_1(x) \in O(d)$. For $i = d+1, \dots, r$, $\lambda_i = O(\varepsilon^{d+4})$, and the corresponding eigenvectors are

$$\beta_i = \begin{bmatrix} 0_{d \times 1} \\ J_{p-d, r-d} X_2(x) \mathfrak{J}_{p,r-d}^\top e_i \end{bmatrix} + \begin{bmatrix} X_1(x) S_{12}(x, \varepsilon) \mathfrak{J}_{p,r-d}^\top e_i + O(\varepsilon^2) \\ O(\varepsilon) \end{bmatrix}, \tag{E.5}$$

where $X_2(x) \in O(r-d)$.

By Lemma 30, we have

$$\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] = [[v_1, v_2]], \quad (\text{E.6})$$

where

$$\begin{aligned} v_1 &= P(x)\mu_{e_d}(x, \varepsilon)J_{p,d}^\top e_d + \sum_{i=1}^d \partial_i P(x)\mu_{2e_i}(x, \varepsilon)J_{p,d}^\top e_i + O(\varepsilon^{d+3}), \\ v_2 &= \frac{P(x)}{2} \sum_{i=1}^d \mathfrak{N}_{ii}(x)\mu_{2e_i}(x, \varepsilon) + O(\varepsilon^{d+3}). \end{aligned} \quad (\text{E.7})$$

Next, we calculate $\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i$, for $i = 1, \dots, d$. Note that the normal component of β_i is of order ε and the normal component of $\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]$ is of order ε^{d+2} , so they will only contribute in the $O(\varepsilon^{d+3})$ term. Therefore, for $i = 1, \dots, d$, the first two order terms of $\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i$ are

$$\begin{aligned} &\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \\ &= (P(x)\mu_{e_d}(x, \varepsilon))(e_d^\top J_{p,d}X_1(x)J_{p,d}^\top e_i) + (P(x)\mu_{e_d}(x, \varepsilon))(e_d^\top J_{p,d}X_1(x)\mathsf{S}_{11}(x, \varepsilon)J_{p,d}^\top e_i) \\ &\quad + \sum_{j=1}^d (\partial P_j(x)\mu_{2e_j}(x, \varepsilon))(e_j^\top J_{p,d}X_1(x)J_{p,d}^\top e_i) + O(\varepsilon^{d+3}). \end{aligned}$$

By putting the above expressions together, a direct calculation shows that the normal component of $\sum_{i=1}^d \frac{\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}}$ is of order 1 and the tangent component of $\sum_{i=1}^d \frac{\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}}$ is of order ε^{-1} :

$$\begin{aligned} &P(x)\mu_{e_d}(x, \varepsilon) \sum_{i=1}^d \frac{(e_d^\top J_{p,d}X_1(x)J_{p,d}^\top e_i)X_1(x)J_{p,d}^\top e_i}{\lambda_i + \varepsilon^{d+3}} \\ &+ \sum_{i=1}^d \frac{\sum_{j=1}^d (\partial P_j(x)\mu_{2e_j}(x, \varepsilon))(e_j^\top J_{p,d}X_1(x)J_{p,d}^\top e_i)X_1(x)J_{p,d}^\top e_i}{\lambda_i + \varepsilon^{d+3}} \\ &+ P(x)\mu_{e_d}(x, \varepsilon) \sum_{i=1}^d \frac{(e_d^\top J_{p,d}X_1(x)\mathsf{S}_{11}(x, \varepsilon)J_{p,d}^\top e_i)X_1(x)J_{p,d}^\top e_i}{\lambda_i + \varepsilon^{d+3}} \\ &+ P(x)\mu_{e_d}(x, \varepsilon) \sum_{i=1}^d \frac{(e_d^\top J_{p,d}X_1(x)J_{p,d}^\top e_i)X_1(x)\mathsf{S}_{11}(x, \varepsilon)J_{p,d}^\top e_i}{\lambda_i + \varepsilon^{d+3}} + O(\varepsilon), \end{aligned} \quad (\text{E.8})$$

where the first term is of order ε^{-1} , the second to the fourth terms are of order 1 since $\mu_{e_d}(x, \varepsilon)$ is of order ε^{d+1} , $\mu_{2e_i}(x, \varepsilon)$ is of order ε^{d+2} for $i = 1, \dots, d$, $\mathsf{S}_{11}(x, \varepsilon)$ is of order ε and λ_i is of order ε^{d+2} for $i = 1, \dots, d$. Note that above formula involves λ_i and $\mathsf{S}_{11}(x, \varepsilon)$. We are going to express those terms by μ_{2e_i} and $\mu_{2e_i+e_d}$. In the following paragraph, we prepare some necessary ingredients to simplify the formula of tangent component of $\sum_{i=1}^d \frac{\mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}}$. Recall that from

(D.22),

$$X_1(x) = \begin{bmatrix} X_1^{(1)}(x) & 0 & \cdots & 0 \\ 0 & X_1^{(2)}(x) & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & X_1^{(k)}(x) \end{bmatrix},$$

$1 \leq k \leq d$. Here different $X_1^{(i)}$ corresponds to different μ_{2e_i} . Each $X_1^{(i)}$ is an orthogonal matrix. By re-ordering the basis $\{e_1, \dots, e_d\}$ of the tangent space $T_x M$, we suppose that among $\mu_{2e_1}(x, \varepsilon), \dots, \mu_{2e_{d-1}}(x, \varepsilon)$, the first t terms of them are different from μ_{2e_d} . Define

$$X_{1,1}(x) := \begin{bmatrix} X_1^{(1)}(x) & 0 & \cdots & 0 \\ 0 & X_1^{(2)}(x) & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & X_1^{(k-1)}(x) \end{bmatrix}. \quad (\text{E.9})$$

Hence, we have

$$X_1(x) = \begin{bmatrix} X_{1,1}(x) & 0 \\ 0 & X_{1,2}(x) \end{bmatrix},$$

where $X_{1,1}(x) \in O(t)$, $X_{1,2} \in O(d-t)$, and $0 \leq t \leq d-1$. Divide $M^{(11)}(x, \varepsilon)$ in (D.1) and $S_{11}(x, \varepsilon)$ in (D.4) corresponding to $\begin{bmatrix} X_{1,1}(x) & 0 \\ 0 & X_{1,2}(x) \end{bmatrix}$:

$$M^{(11)}(x, \varepsilon) = \begin{bmatrix} M_1^{(11)}(x, \varepsilon) & M_2^{(11)}(x, \varepsilon) \\ M_3^{(11)}(x, \varepsilon) & M_4^{(11)}(x, \varepsilon) \end{bmatrix}, \quad S_{11}(x, \varepsilon) = \begin{bmatrix} S_{11,1}(x, \varepsilon) & S_{11,2}(x, \varepsilon) \\ S_{11,3}(x, \varepsilon) & S_{11,4}(x, \varepsilon) \end{bmatrix}.$$

Recall that

$$M_{m,n}^{(11)} = \begin{cases} \partial_d P(x) \mu_{2e_m+e_d}(x, \varepsilon) & 1 \leq m = n \leq d, \\ \partial_n P(x) \mu_{2e_n+e_d}(x, \varepsilon) & m = d, 1 \leq n \leq d, \\ \partial_m P(x) \mu_{2e_m+e_d}(x, \varepsilon) & n = d, 1 \leq m \leq d, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{E.10})$$

By the perturbation theory (see, e.g., (Wu and Wu, 2018, Appendix A)), $\lambda_{t+1}^{(1)}(x, \varepsilon), \dots, \lambda_d^{(1)}(x, \varepsilon)$ are the eigenvalues of $M_4^{(11)}(x, \varepsilon)$ and $X_{1,2}(x)$ is the orthonormal eigenvector matrix of $M_4^{(11)}(x, \varepsilon)$. We have

$$S_{11,2}(x, \varepsilon) = X_{1,1}^\top(x) [P(x) \mu_{2e_d}(x, \varepsilon) I_{t \times t} - \Lambda]^{-1} M_2^{(11)}(x, \varepsilon) X_{1,2}(x), \quad (\text{E.11})$$

where

$$\Lambda = \begin{bmatrix} P(x) \mu_{2e_1}(x, \varepsilon) & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & P(x) \mu_{2e_t}(x, \varepsilon) \end{bmatrix},$$

Next, we simplify the terms in equation (E.8) one by one. We start from the first one. Recall that based on the structure of X_1 , we have $e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i = 0$ for $i = 1, \dots, t$. Hence,

$$\begin{aligned}
 & P(x) \mu_{e_d}(x, \varepsilon) \sum_{i=1}^d \frac{(e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i}{\lambda_i + \varepsilon^{d+3}} \\
 &= P(x) \mu_{e_d}(x, \varepsilon) \sum_{i=t+1}^d \frac{(e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i}{P(x) \mu_{2e_d}(x, \varepsilon) + \lambda_i^{(1)}(x, \varepsilon) + \varepsilon^{d+3} + O(\varepsilon^{d+4})} \\
 &= P(x) \mu_{e_d}(x, \varepsilon) \sum_{i=t+1}^d \left[\frac{1}{P(x) \mu_{2e_d}(x, \varepsilon)} - \frac{\lambda_i^{(1)}(x, \varepsilon) + \varepsilon^{d+3}}{(P(x) \mu_{2e_d}(x, \varepsilon))^2} + O(\varepsilon^{-d}) \right] \\
 &\quad \times (e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i \\
 &= \frac{\mu_{e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \sum_{i=t+1}^d (e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i \\
 &\quad - \frac{\mu_{e_d}(x, \varepsilon)}{P(x) (\mu_{2e_d}(x, \varepsilon))^2} \sum_{i=t+1}^d \lambda_i^{(1)}(x, \varepsilon) (e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i \\
 &\quad - \frac{\varepsilon^{d+3} \mu_{e_d}(x, \varepsilon)}{P(x) (\mu_{2e_d}(x, \varepsilon))^2} \sum_{i=t+1}^d (e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i + O(\varepsilon).
 \end{aligned} \tag{E.12}$$

Note that we use (D.2) in the first step. Moreover, we have

$$\frac{\mu_{e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \sum_{i=t+1}^d (e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i = \frac{\mu_{e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} J_{p,d}^\top e_d \tag{E.13}$$

and

$$\frac{\varepsilon^{d+3} \mu_{e_d}(x, \varepsilon)}{P(x) (\mu_{2e_d}(x, \varepsilon))^2} \sum_{i=t+1}^d (e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i = \frac{\varepsilon^{d+3} \mu_{e_d}(x, \varepsilon)}{P(x) (\mu_{2e_d}(x, \varepsilon))^2} J_{p,d}^\top e_d. \tag{E.14}$$

By using the eigen-decomposition of $M_4^{(11)}(x, \varepsilon)$, we have

$$\begin{aligned}
 & e_k^\top J_{p,d} \sum_{i=t+1}^d \lambda_i^{(1)}(x, \varepsilon) (e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i \\
 &= \sum_{i=t+1}^d \lambda_i^{(1)}(x, \varepsilon) (e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) (e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) \\
 &= \sum_{i=t+1}^d \lambda_i^{(1)}(x, \varepsilon) (e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) (e_i^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_k) \\
 &= \partial_k P(x) \mu_{2e_k+e_d}(x, \varepsilon)
 \end{aligned} \tag{E.15}$$

for $t+1 \leq k \leq d$ and this quantity is 0 if $1 \leq k \leq d$. Thus, if we sum up the above terms, we have

$$\begin{aligned} & P(x)\mu_{e_d}(x, \varepsilon) \sum_{i=1}^d \frac{(e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i}{\lambda_i + \varepsilon^{d+3}} \\ &= \frac{\mu_{e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} J_{p,d}^\top e_d - \frac{\varepsilon^{d+3} \mu_{e_d}(x, \varepsilon)}{P(x)(\mu_{2e_d}(x, \varepsilon))^2} J_{p,d}^\top e_d \\ &\quad - \sum_{j=t+1}^d \frac{\partial_j P(x) \mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{P(x)(\mu_{2e_d}(x, \varepsilon))^2} J_{p,d}^\top e_j + O(\varepsilon). \end{aligned} \quad (\text{E.16})$$

Next, we simplify the second term in (E.8). Recall the description of X_1 (e.g. (D.22)). We have

$$\sum_{i=1}^d \frac{(e_j^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i}{\mu_{2e_j}(x, \varepsilon)} = \frac{1}{\mu_{2e_j}(x, \varepsilon)} J_{p,d}^\top e_j. \quad (\text{E.17})$$

Hence,

$$\begin{aligned} & \sum_{i=1}^d \frac{\sum_{j=1}^d (\partial P_j(x) \mu_{2e_j}(x, \varepsilon)) (e_j^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i}{\lambda_i + \varepsilon^{d+3}} \\ &= \sum_{i=1}^d \frac{\sum_{j=1}^d (\partial P_j(x) \mu_{2e_j}(x, \varepsilon)) (e_j^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i}{P(x) \mu_{2e_i}(x, \varepsilon) + \lambda_i^{(1)}(x, \varepsilon) + \varepsilon^{d+3} + O(\varepsilon^{d+4})} \\ &= \sum_{j=1}^d \frac{\partial P_j(x) \mu_{2e_j}(x, \varepsilon)}{P(x)} \sum_{i=1}^d \frac{(e_j^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i}{\mu_{2e_i}(x, \varepsilon)} + O(\varepsilon) \\ &= \frac{\nabla P(x)}{P(x)} + O(\varepsilon). \end{aligned} \quad (\text{E.18})$$

At last, we simplify the third and the last terms in (E.8) together, because we need to use the antisymmetric property of $S_{11}(x, \varepsilon)$.

$$\begin{aligned} & P(x)\mu_{e_d}(x, \varepsilon) \sum_{i=1}^d \frac{(e_d^\top J_{p,d} X_1(x) S_{11}(x, \varepsilon) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i}{\lambda_i} \\ &\quad + P(x)\mu_{e_d}(x, \varepsilon) \sum_{i=1}^d \frac{(e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) S_{11}(x, \varepsilon) J_{p,d}^\top e_i}{\lambda_i} \\ &= \mu_{e_d}(x, \varepsilon) \sum_{i=1}^d \left[\frac{(e_d^\top J_{p,d} X_1(x) S_{11}(x, \varepsilon) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i}{\mu_{2e_i}(x, \varepsilon) + \lambda_i^{(1)}(x, \varepsilon)/P(x) + \varepsilon^{d+3} + O(\varepsilon^{d+4})} \right. \\ &\quad \left. + \frac{(e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) S_{11}(x, \varepsilon) J_{p,d}^\top e_i}{\mu_{2e_i}(x, \varepsilon) + \lambda_i^{(1)}(x, \varepsilon)/P(x) + \varepsilon^{d+3} + O(\varepsilon^{d+4})} \right] \\ &= \bar{v} + O(\varepsilon), \end{aligned} \quad (\text{E.19})$$

where we denote

$$\begin{aligned} \bar{v} := & \mu_{e_d}(x, \varepsilon) \sum_{i=1}^d \left[\frac{(e_d^\top J_{p,d} X_1(x) S_{11}(x, \varepsilon) J_{p,d}^\top e_i) X_1(x) J_{p,d}^\top e_i}{\mu_{2e_i}(x, \varepsilon)} \right. \\ & \left. + \frac{(e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) X_1(x) S_{11}(x, \varepsilon) J_{p,d}^\top e_i}{\mu_{2e_i}(x, \varepsilon)} \right]. \end{aligned} \quad (\text{E.20})$$

We now simplify \bar{v} . Note that, for $1 \leq k \leq d$,

$$\begin{aligned}
 & e_k^\top J_{p,d} \bar{v} \\
 &= \mu_{e_d}(x, \varepsilon) \sum_{i=1}^d \left(\sum_{j=1}^d e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_j \frac{e_j^\top J_{p,d} S_{11}(x, \varepsilon) J_{p,d}^\top e_i}{\mu_{2e_i}(x, \varepsilon)} \right) e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i \\
 &\quad + \mu_{e_d}(x, \varepsilon) \sum_{i=1}^d e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_i \left(\sum_{j=1}^d e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_j \frac{e_j^\top J_{p,d} S_{11}(x, \varepsilon) J_{p,d}^\top e_i}{\mu_{2e_i}(x, \varepsilon)} \right) \\
 &= \mu_{e_d}(x, \varepsilon) \sum_{j=1}^d e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_j \left(\sum_{i=1}^d e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i \frac{e_j^\top J_{p,d} S_{11}(x, \varepsilon) J_{p,d}^\top e_i}{\mu_{2e_i}(x, \varepsilon)} \right) \\
 &\quad + \mu_{e_d}(x, \varepsilon) \sum_{j=1}^d e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_j \left(\sum_{i=1}^d e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i \frac{e_i^\top J_{p,d} S_{11}(x, \varepsilon) J_{p,d}^\top e_j}{\mu_{2e_j}(x, \varepsilon)} \right) \\
 &= \mu_{e_d}(x, \varepsilon) \sum_{j=1}^d e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_j \\
 &\quad \times \left[\sum_{i=1}^d e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i \left(\frac{1}{\mu_{2e_j}(x, \varepsilon)} - \frac{1}{\mu_{2e_i}(x, \varepsilon)} \right) e_i^\top J_{p,d} S_{11}(x, \varepsilon) J_{p,d}^\top e_j \right].
 \end{aligned}$$

In the last step, we use the fact that $S_{11}(x, \varepsilon)$ is antisymmetric. Based on the structure of $X_1(x)$, $e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_j = 0$ for $j = 1, \dots, t$, $e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i = 0$, for $k = 1, \dots, t$ and $i = t+1, \dots, d$ and $e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i = 0$ for $k = t+1, \dots, d$ and $i = 1, \dots, t$. We can further simplify \bar{v} as

$$\begin{aligned}
 e_k^\top J_{p,d} \bar{v} &= \mu_{e_d}(x, \varepsilon) \sum_{j=t+1}^d e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_j \\
 &\quad \times \left[\sum_{i=1}^t e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i \left(\frac{1}{\mu_{2e_d}(x, \varepsilon)} - \frac{1}{\mu_{2e_i}(x, \varepsilon)} \right) e_i^\top J_{p,d} S_{11}(x, \varepsilon) J_{p,d}^\top e_j \right], \tag{E.21}
 \end{aligned}$$

for $k = 1, \dots, t$, and

$$\begin{aligned}
 & e_k^\top J_{p,d} \bar{v} \\
 &= \mu_{e_d}(x, \varepsilon) \sum_{j=t+1}^d e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_j \\
 &\quad \times \left[\sum_{i=t+1}^d e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i \left(\frac{1}{\mu_{2e_d}(x, \varepsilon)} - \frac{1}{\mu_{2e_i}(x, \varepsilon)} \right) e_i^\top J_{p,d} S_{11}(x, \varepsilon) J_{p,d}^\top e_j \right] \\
 &= 0, \tag{E.22}
 \end{aligned}$$

for $k = t+1, \dots, d$, where we use the fact that $\mu_{2e_{t+1}}(x, \varepsilon) = \dots = \mu_{2e_d}(x, \varepsilon)$.

Next, we focus on the case when $1 \leq k \leq t$. By (E.11), for $1 \leq i \leq t$ and $t+1 \leq j \leq d$, we have

$$\begin{aligned}
 e_i^\top J_{p,d} S_{11}(x, \varepsilon) J_{p,d}^\top e_j &= \sum_{l=1}^t e_i^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_l \\
 &\quad \times \sum_{m=t+1}^d \frac{(e_l^\top J_{p,d} M^{(11)}(x, \varepsilon) J_{p,d}^\top e_m)(e_m^\top J_{p,d} X_1(x) J_{p,d}^\top e_j)}{P(x)\mu_{2e_d}(x, \varepsilon) - P(x)\mu_{2e_l}(x, \varepsilon)}. \tag{E.23}
 \end{aligned}$$

Note by Lemma 31, if $1 \leq l \leq t$, and $t+1 \leq m < d$, then $e_l^\top J_{p,d} M^{(11)}(x, \varepsilon) J_{p,d}^\top e_m = 0$. And

$$e_l^\top J_{p,d} M^{(11)}(x, \varepsilon) J_{p,d}^\top e_d = \partial_l P(x) \mu_{2e_l + e_d}(x, \varepsilon). \quad (\text{E.24})$$

Hence,

$$\begin{aligned} e_i^\top J_{p,d} S_{11}(x, \varepsilon) J_{p,d}^\top e_j &= e_j^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_d \\ &\times \sum_{l=1}^t e_l^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_l \frac{\partial_l P(x) \mu_{2e_l + e_d}(x, \varepsilon)}{P(x) \mu_{2e_d}(x, \varepsilon) - P(x) \mu_{2e_l}(x, \varepsilon)}. \end{aligned} \quad (\text{E.25})$$

We substitute above equation into (E.21),

$$\begin{aligned} &e_k^\top J_{p,d} \bar{v} \\ &= \mu_{e_d}(x, \varepsilon) \sum_{j=t+1}^d (e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_j) (e_j^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_d) \left[\sum_{i=1}^t e_i^\top J_{p,d} X_1(x) J_{p,d}^\top e_i \right. \\ &\quad \times \left(\frac{1}{\mu_{2e_d}(x, \varepsilon)} - \frac{1}{\mu_{2e_i}(x, \varepsilon)} \right) \sum_{l=1}^t e_l^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_l \frac{\partial_l P(x) \mu_{2e_l + e_d}(x, \varepsilon)}{P(x) \mu_{2e_d}(x, \varepsilon) - P(x) \mu_{2e_l}(x, \varepsilon)} \Big] \\ &= \mu_{e_d}(x, \varepsilon) \left[\sum_{i=1}^t e_i^\top J_{p,d} X_1(x) J_{p,d}^\top e_i \left(\frac{1}{\mu_{2e_d}(x, \varepsilon)} - \frac{1}{\mu_{2e_i}(x, \varepsilon)} \right) \right. \\ &\quad \times \left. \sum_{l=1}^t e_l^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_l \frac{\partial_l P(x) \mu_{2e_l + e_d}(x, \varepsilon)}{P(x) \mu_{2e_d}(x, \varepsilon) - P(x) \mu_{2e_l}(x, \varepsilon)} \right] \\ &= \mu_{e_d}(x, \varepsilon) \left[\sum_{l=1}^t \sum_{i=1}^t (e_l^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) (e_i^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_l) \left(\frac{1}{\mu_{2e_d}(x, \varepsilon)} - \frac{1}{\mu_{2e_i}(x, \varepsilon)} \right) \right. \\ &\quad \times \left. \frac{\partial_l P(x) \mu_{2e_l + e_d}(x, \varepsilon)}{P(x) \mu_{2e_d}(x, \varepsilon) - P(x) \mu_{2e_l}(x, \varepsilon)} \right] \\ &= \mu_{e_d}(x, \varepsilon) \left[\left(\frac{1}{\mu_{2e_d}(x, \varepsilon)} - \frac{1}{\mu_{2e_k}(x, \varepsilon)} \right) \sum_{l=1}^t \sum_{i=1}^t (e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) (e_i^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_l) \right. \\ &\quad \times \left. \frac{\partial_l P(x) \mu_{2e_l + e_d}(x, \varepsilon)}{P(x) \mu_{2e_d}(x, \varepsilon) - P(x) \mu_{2e_l}(x, \varepsilon)} \right], \end{aligned}$$

where we use the fact that

$$\sum_{j=t+1}^d (e_d^\top J_{p,d} X_1(x) J_{p,d}^\top e_j) (e_j^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_d) = 1 \quad (\text{E.26})$$

in the second step and the fact that

$$(e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) (e_i^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_l) \neq 0 \quad (\text{E.27})$$

only if $e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i$ and $e_i^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_l$ are entries in the block $X_1^{(m)}$ in (E.9) corresponding to μ_{2e_k} in the fourth step. Note that

$$\sum_{i=1}^t (e_k^\top J_{p,d} X_1(x) J_{p,d}^\top e_i) (e_i^\top J_{p,d} X_1^\top(x) J_{p,d}^\top e_l) = 1, \quad (\text{E.28})$$

if $1 \leq k = l \leq t$ and is 0 otherwise.

Hence, we have

$$e_k^\top J_{p,d} \bar{v} = -\frac{\partial_k P(x) \mu_{e_d}(x, \varepsilon) \mu_{2e_k+e_d}(x, \varepsilon)}{P(x) \mu_{2e_k}(x, \varepsilon) \mu_{2e_d}(x, \varepsilon)}, \quad (\text{E.29})$$

for $1 \leq k \leq t$, and $e_k^\top J_{p,d} \bar{v} = 0$, for $t+1 \leq k \leq d$. If we sum up equations (E.16) (E.18) and (E.29), the tangent component of $\sum_{i=1}^d \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}}$ becomes

$$\begin{aligned} & \frac{\mu_{e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} J_{p,d}^\top e_d + \frac{\nabla P(x)}{P(x)} - \frac{\varepsilon^{d+3} \mu_{e_d}(x, \varepsilon)}{P(x) (\mu_{2e_d}(x, \varepsilon))^2} J_{p,d}^\top e_d \\ & - \sum_{i=t+1}^d \frac{\partial_i P(x) \mu_{e_d}(x, \varepsilon) \mu_{2e_i+e_d}(x, \varepsilon)}{P(x) (\mu_{2e_d}(x, \varepsilon))^2} J_{p,d}^\top e_i \\ & - \sum_{i=1}^t \frac{\partial_i P(x) \mu_{e_d}(x, \varepsilon) \mu_{2e_i+e_d}(x, \varepsilon)}{P(x) \mu_{2e_i}(x, \varepsilon) \mu_{2e_d}(x, \varepsilon)} J_{p,d}^\top e_i + O(\varepsilon) \\ & = \frac{\mu_{e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} J_{p,d}^\top e_d + \frac{\nabla P(x)}{P(x)} - \frac{\varepsilon^{d+3} \mu_{e_d}(x, \varepsilon)}{P(x) (\mu_{2e_d}(x, \varepsilon))^2} J_{p,d}^\top e_d \\ & - \sum_{i=1}^d \frac{\partial_i P(x) \mu_{e_d}(x, \varepsilon) \mu_{2e_i+e_d}(x, \varepsilon)}{P(x) \mu_{2e_i}(x, \varepsilon) \mu_{2e_d}(x, \varepsilon)} J_{p,d}^\top e_i + O(\varepsilon), \end{aligned} \quad (\text{E.30})$$

where we use $\mu_{2e_{t+1}}(x, \varepsilon) = \dots = \mu_{2e_d}(x, \varepsilon)$ in the last step.

We now finish calculating the tangent component of $\sum_{i=1}^d \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}}$. Next, we need to calculate both the tangent and the normal component of $\sum_{i=d+1}^r \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}}$. Note that for $i = d+1, \dots, r$,

$$\begin{aligned} & \mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \\ & = P(x) \mu_{e_d}(x, \varepsilon) (e_d^\top J_{p,d} X_1(x) S_{12}(x, \varepsilon) \tilde{J}_{p,r-d}^\top e_i) \\ & + \frac{P(x)}{2} \sum_{j=1}^d \mu_{2e_j}(x, \varepsilon) \mathfrak{N}_{jj}^\top(x) J_{p-d,r-d} X_2(x) \tilde{J}_{p,r-d}^\top e_i + O(\varepsilon^{d+3}), \end{aligned} \quad (\text{E.31})$$

where both terms are of order ε^{d+2} . Since $\lambda_i = O(\varepsilon^{d+4})$, ε^{d+3} dominates the eigenvalues. For $i = d+1, \dots, r$, we have

$$\begin{aligned} & \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i}{\lambda_i + \varepsilon^{d+3}} = \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i}{\varepsilon^{d+3} + O(\varepsilon^{d+4})} \\ & = P(x) \frac{\mu_{e_d}(x, \varepsilon)}{\varepsilon^{d+3}} (e_d^\top J_{p,d} X_1(x) S_{12}(x, \varepsilon) \tilde{J}_{p,r-d}^\top e_i) \\ & + \frac{P(x)}{2} \sum_{j=1}^d \frac{\mu_{2e_j}(x, \varepsilon)}{\varepsilon^{d+3}} \mathfrak{N}_{jj}^\top(x) J_{p-d,r-d} X_2(x) \tilde{J}_{p,r-d}^\top e_i + O(1). \end{aligned} \quad (\text{E.32})$$

Similarly, we need to express the above formula in terms of $\mu_{2e_i}(x, \varepsilon)$ and $\mu_{2e_i+e_d}(x, \varepsilon)$. The simplification here mainly relies on the perturbation formula equations (D.20) and (D.21) which relates

$\mathbb{S}_{12}(x, \varepsilon)$ with the second fundamental form of the manifold at x . First of all, a direct calculation shows that

$$\begin{aligned}
& \sum_{i=d+1}^r \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}} \\
&= \sum_{i=d+1}^r \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\varepsilon^{d+3} + O(\varepsilon^{d+4})} \\
&= \left[\sum_{i=d+1}^r \left[P(x) \frac{\mu_{e_d}(x, \varepsilon)}{\varepsilon^{d+3}} (e_d^\top J_{p,d} X_1(x) \mathbb{S}_{12}(x, \varepsilon) \mathfrak{J}_{p,r-d}^\top e_i) \right] X_1(x) \mathbb{S}_{12}(x, \varepsilon) \mathfrak{J}_{p,r-d}^\top e_i \right. \\
&\quad + \sum_{i=d+1}^r \left[\frac{P(x)}{2} \sum_{j=1}^d \frac{\mu_{2e_j}(x, \varepsilon)}{\varepsilon^{d+3}} \mathfrak{N}_{jj}^\top(x) J_{p-d,r-d} X_2(x) \mathfrak{J}_{p,r-d}^\top e_i \right] X_1(x) \mathbb{S}_{12}(x, \varepsilon) \mathfrak{J}_{p,r-d}^\top e_i + O(\varepsilon), \\
&\quad \left. + \sum_{i=d+1}^r \left[P(x) \frac{\mu_{e_d}(x, \varepsilon)}{\varepsilon^{d+3}} (e_d^\top J_{p,d} X_1(x) \mathbb{S}_{12}(x, \varepsilon) \mathfrak{J}_{p,r-d}^\top e_i) \right] J_{p-d,r-d} X_2(x) \mathfrak{J}_{p,r-d}^\top e_i \right. \\
&\quad + \sum_{i=d+1}^r \left[\frac{P(x)}{2} \sum_{j=1}^d \frac{\mu_{2e_j}(x, \varepsilon)}{\varepsilon^{d+3}} \mathfrak{N}_{jj}^\top(x) J_{p-d,r-d} X_2(x) \mathfrak{J}_{p,r-d}^\top e_i \right] J_{p-d,r-d} X_2(x) \mathfrak{J}_{p,r-d}^\top e_i + O(1) \right] \\
&= \left[\sum_{i=d+1}^r \left\{ \frac{P(x)}{2\varepsilon^{d+3}} \sum_{j=1}^d \left[\left(\mu_{2e_j}(x, \varepsilon) - \frac{\mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \right) \mathfrak{N}_{jj}^\top(x) \right] J_{p-d,r-d} X_2(x) \mathfrak{J}_{p,r-d}^\top e_i \right\} \right. \\
&\quad \times X_1(x) \mathbb{S}_{12}(x, \varepsilon) \mathfrak{J}_{p,r-d}^\top e_i + O(\varepsilon), \\
&\quad \left. \sum_{i=d+1}^r \left\{ \frac{P(x)}{2\varepsilon^{d+3}} \sum_{j=1}^d \left[\left(\mu_{2e_j}(x, \varepsilon) - \frac{\mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \right) \mathfrak{N}_{jj}^\top(x) \right] \right. \right. \\
&\quad \times J_{p-d,r-d} X_2(x) \mathfrak{J}_{p,r-d}^\top e_i \left. \right\} J_{p-d,r-d} X_2(x) \mathfrak{J}_{p,r-d}^\top e_i + O(1) \right],
\end{aligned}$$

where we use (D.21) in the last step. To simplify the tangent and normal components of

$\sum_{i=d+1}^r \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}}$, we need the following formula. Suppose $v \in \mathbb{R}^{r-d}$, $G \in \mathbb{R}^{d \times (r-d)}$ with $e_i^\top J_{p,d} G = w_i^\top$ for $i = 1, \dots, d$. By Lemma 31, $X_2(x) \in O(r-d)$. We can represent the inner product between v and w_i in orthonormal basis formed by the column vectors of $X_2(x)$.

$$\sum_{i=d+1}^r [v^\top X_2(x) \mathfrak{J}_{p,r-d}^\top e_i] G X_2(x) \mathfrak{J}_{p,r-d}^\top e_i = \sum_{i=1}^d v^\top w_i J_{p,d}^\top e_i. \quad (\text{E.33})$$

By (D.20), the tangent component of $\sum_{i=d+1}^r \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_{\varepsilon}^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}}$ is

$$\begin{aligned}
 & \sum_{i=d+1}^r \left\{ \frac{P(x)}{2\varepsilon^{d+3}} \sum_{j=1}^d \left[\left(\mu_{e_d}(x, \varepsilon) - \frac{\mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \right) \mathfrak{N}_{jj}^\top(x) \right] J_{p-d, r-d} X_2(x) \mathfrak{J}_{p, r-d}^\top e_i \right\} \\
 & \quad \times X_1(x) \mathfrak{S}_{12}(x, \varepsilon) \mathfrak{J}_{p, r-d}^\top e_i \\
 & = \frac{P(x)}{2\varepsilon^{d+3}} \sum_{i=1}^{d-1} \sum_{j=1}^d \left[\left(\frac{\mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} - \mu_{2e_j}(x, \varepsilon) \right) \frac{\mu_{2e_i+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \mathfrak{N}_{jj}^\top(x) \right] \\
 & \quad \times J_{p-d, r-d} J_{p-d, r-d}^\top \mathfrak{N}_{id}(x) J_{p, d}^\top e_i \\
 & \quad + \frac{P(x)}{4\varepsilon^{d+3}} \sum_{i=1}^d \sum_{j=1}^d \left[\left(\frac{\mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} - \mu_{2e_j}(x, \varepsilon) \right) \frac{\mu_{2e_i+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \mathfrak{N}_{jj}^\top(x) \right] \\
 & \quad \times J_{p-d, r-d} J_{p-d, r-d}^\top \mathfrak{N}_{ii}(x) J_{p, d}^\top e_d \\
 & = \frac{P(x)}{2\varepsilon^{d+3}} \sum_{i=1}^{d-1} \sum_{j=1}^d \left[\left(\frac{\mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} - \mu_{2e_j}(x, \varepsilon) \right) \frac{\mu_{2e_i+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \mathfrak{N}_{jj}^\top(x) \right] \mathfrak{N}_{id}(x) J_{p, d}^\top e_i \\
 & \quad + \frac{P(x)}{4\varepsilon^{d+3}} \sum_{i=1}^d \sum_{j=1}^d \left[\left(\frac{\mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} - \mu_{2e_j}(x, \varepsilon) \right) \frac{\mu_{2e_i+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \mathfrak{N}_{jj}^\top(x) \right] \mathfrak{N}_{ii}(x) J_{p, d}^\top e_d,
 \end{aligned}$$

where in the first step we apply equations (D.20), (D.21) and (E.33). In the last step, $e_m^\top \mathbb{I}(e_i, e_j) = 0$ for $m = r+1, \dots, p$, and $i, j = 1, \dots, d$. Hence,

$$\mathfrak{N}_{jj}^\top(x) J_{p-d, r-d} J_{p-d, r-d}^\top \mathfrak{N}_{ii}(x) = \mathfrak{N}_{jj}^\top(x) \mathfrak{N}_{ii}(x). \quad (\text{E.34})$$

By Lemma 31, we have $X_2(x) \in O(r-d)$, and $e_m^\top \mathbb{I}(e_j, e_j) = 0$ for $m = r+1, \dots, p$, and $j = 1, \dots, d$. Hence, (E.33) implies that

$$\sum_{i=d+1}^r [\mathfrak{N}_{jj}^\top(x) J_{p-d, r-d} X_2(x) \mathfrak{J}_{p, r-d}^\top e_i] J_{p-d, r-d} X_2(x) \mathfrak{J}_{p, r-d}^\top e_i = \mathfrak{N}_{jj}. \quad (\text{E.35})$$

We use it to simplify the normal component $\sum_{i=d+1}^r \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_{\varepsilon}^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}}$. We have

$$\begin{aligned}
 & \sum_{i=d+1}^r \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_{\varepsilon}^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}} \\
 & = \left[\frac{P(x)}{2\varepsilon^{d+3}} \sum_{i=1}^{d-1} \sum_{j=1}^d \left[\left(\frac{\mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} - \mu_{2e_j}(x, \varepsilon) \right) \frac{\mu_{2e_i+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \mathfrak{N}_{jj}^\top(x) \right] \mathfrak{N}_{id}(x) J_{p, d}^\top e_i \right. \\
 & \quad + \frac{P(x)}{4\varepsilon^{d+3}} \sum_{i=1}^d \sum_{j=1}^d \left[\left(\frac{\mu_{e_d}(x, \varepsilon) \mu_{2e_j+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} - \mu_{2e_j}(x, \varepsilon) \right) \frac{\mu_{2e_i+e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} \mathfrak{N}_{jj}^\top(x) \right] \\
 & \quad \times \mathfrak{N}_{ii}(x) J_{p, d}^\top e_d + O(\varepsilon), \\
 & \quad \left. \frac{P(x)}{2\varepsilon^{d+3}} \sum_{j=1}^d \left(\mu_{2e_j}(x, \varepsilon) - \frac{\mu_{e_d} \mu_{2e_j+e_d}}{\mu_{2e_d}} \right) \mathfrak{N}_{jj} + O(1) \right].
 \end{aligned}$$

By summing up $\sum_{i=1}^d \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}}$ and $\sum_{i=d+1}^r \frac{\mathbb{E}[(\iota(X) - \iota(x)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon^{d+3}}$, we have the conclusion. \blacksquare

Appendix F. Bias analysis on the kernel of LLE and the associated integral operator

F.1 Proof of Proposition 10

1. When $x \in M \setminus M_\varepsilon$, $\mu_{e_d} = 0$. By Lemma 32, $\mathbf{T}(x) = [O(1), O(\varepsilon^{-1})]$. If $\iota(y) \in B_\varepsilon^{\mathbb{R}^p}(\iota(x))$, $\iota(y) - \iota(x) = [O(\varepsilon), O(\varepsilon^2)]$. So, $(\iota(y) - \iota(x))^\top \mathbf{T}(x) = O(\varepsilon)$ and $K_\varepsilon(x, y) = 1 - O(\varepsilon) > 0$ when ε is small enough.
2. When $x \in M_\varepsilon$, $\mathbf{T}(x) = [\frac{\mu_{e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} J_{p,d}^\top e_d + O(1), O(\varepsilon^{-1})]$ and

$$\iota(y) - \iota(x) = [\sum_{i=1}^d u_i e_i + O(\|u\|^3), O(\|u\|^2)] = [\sum_{i=1}^d u_i e_i + O(\varepsilon^3), O(\varepsilon^2)]. \quad (\text{F.1})$$

Therefore, by Corollary 28,

$$K_\varepsilon(x, y) = 1 - \frac{\sigma_{1,d}(\tilde{\varepsilon}_x) u_d}{\sigma_{2,d}(\tilde{\varepsilon}_x) \varepsilon} + O(\varepsilon). \quad (\text{F.2})$$

By definition, $-\frac{\sigma_{1,d}(\tilde{\varepsilon}_x)}{\sigma_{2,d}(\tilde{\varepsilon}_x)} > 0$ and it is a decreasing function of $\tilde{\varepsilon}_x$. Therefore, to discuss the infimum of $K_\varepsilon(x, y)$, it is sufficient to consider the case when $x \in \partial M$, i.e. when $\tilde{\varepsilon}_x = 0$. If $\tilde{\varepsilon}_x = 0$, then

$$K_\varepsilon(x, y) = 1 + \left[\frac{2d(d+2)|S^{d-2}|}{(d^2-1)|S^{d-1}|\varepsilon} + O(1) \right] u_d + O(\varepsilon). \quad (\text{F.3})$$

Hence, let $u_d^* = \inf u_d$ where the infimum is taken over $x \in \partial M$ and $\iota(y) \in B_\varepsilon^{\mathbb{R}^p}(\iota(x))$, then if ε is small enough,

$$\inf_{x,y} K_\varepsilon(x, y) = 1 + \left[\frac{2d(d+2)|S^{d-2}|}{(d^2-1)|S^{d-1}|\varepsilon} + O(1) \right] u_d^* + O(\varepsilon). \quad (\text{F.4})$$

Obviously, $u_d^* = -\varepsilon + O(\varepsilon^2)$. Therefore, $\inf_{x,y} K_\varepsilon(x, y) = 1 - \frac{2d(d+2)|S^{d-2}|}{(d^2-1)|S^{d-1}|} + O(\varepsilon)$. It is worth to note that $\frac{2d(d+2)|S^{d-2}|}{(d^2-1)|S^{d-1}|} > 1$ by Lemma 29.

3. By Lemma 25 and part (1)

$$\begin{aligned} & \mathbb{E} K_\varepsilon(x, X) \\ &= \int_{D(x)} \left(1 - \frac{\mu_{e_d}(x, \varepsilon) u_d}{\mu_{2e_d}(x, \varepsilon)} + O(\varepsilon) \right) (P(x) + O(u))(1 + O(u^2)) du \\ &= \int_{\tilde{D}(x)} \left(1 - \frac{\mu_{e_d}(x, \varepsilon) u_d}{\mu_{2e_d}(x, \varepsilon)} + O(\varepsilon) \right) (P(x) + O(u))(1 + O(u^2)) du + O(\varepsilon^{d+2}) \\ &= P(x) \int_{\tilde{D}(x)} 1 - \frac{\sigma_{1,d}(\tilde{\varepsilon}_x) u_d}{\sigma_{2,d}(\tilde{\varepsilon}_x) \varepsilon} du + O(\varepsilon^{d+1}) \end{aligned}$$

Since $-\frac{\sigma_{1,d}(\tilde{\varepsilon}_x)}{\sigma_{2,d}(\tilde{\varepsilon}_x)} > 0$ and it is a decreasing function of $\tilde{\varepsilon}_x$, it suffice to show that if $x \in \partial M$, then $\int_{\tilde{D}(x)} 1 - \frac{\mu_{e_d}(x, \varepsilon) u_d}{\mu_{2e_d}(x, \varepsilon)} du \geq C(d) \varepsilon^d$.

If $x \in \partial M$, then $1 - \frac{\sigma_{1,d}(\tilde{\varepsilon}_x) u_d}{\sigma_{2,d}(\tilde{\varepsilon}_x) \varepsilon} = 1 + \frac{2d(d+2)|S^{d-2}|u_d}{(d^2-1)|S^{d-1}|\varepsilon}$, and

$$\begin{aligned} & \int_{\tilde{D}(x)} 1 - \frac{\sigma_{1,d}(\tilde{\varepsilon}_x) u_d}{\sigma_{2,d}(\tilde{\varepsilon}_x) \varepsilon} du \\ & \geq \frac{|S^{d-2}|}{d-1} \int_{-\varepsilon}^0 [1 + \frac{2d(d+2)|S^{d-2}|u_d}{(d^2-1)|S^{d-1}|\varepsilon}] (\varepsilon^2 - u_d^2)^{\frac{d-1}{2}} du_d \\ & = \varepsilon^d \frac{|S^{d-2}|}{d-1} \int_0^1 [1 - \frac{2d(d+2)|S^{d-2}|a}{(d^2-1)|S^{d-1}|}] (1-a^2)^{\frac{d-1}{2}} da \\ & = \varepsilon^d \left[\frac{|S^{d-2}|}{d-1} \int_0^1 (1-a^2)^{\frac{d-1}{2}} da - \frac{|S^{d-2}|}{d-1} \frac{2d(d+2)|S^{d-2}|}{(d^2-1)|S^{d-1}|} \int_0^1 a(1-a^2)^{\frac{d-1}{2}} da \right] \\ & = \varepsilon^d \left[\frac{|S^{d-1}|}{2d} - \frac{2d(d+2)|S^{d-2}|^2}{(d^2-1)^2|S^{d-1}|} \right]. \end{aligned}$$

We have thus finished the proof since $\frac{|S^{d-1}|}{2d} - \frac{2d(d+2)|S^{d-2}|^2}{(d^2-1)^2|S^{d-1}|} > 0$ for any d following from Lemma 29.

F.2 Proof of Proposition 13

When $d = 1$, the differentiability follows from the direct calculation. For $d > 1$, the differentiability follows from the fundamental theorem of calculus. The rest of the statements follow directly from the definition of σ , except $\phi_1(\tilde{\varepsilon}_x) > 0$ and $\phi_2(\tilde{\varepsilon}_x) < 0$ when $\tilde{\varepsilon}_x = 0$.

We now prove $\phi_1(\tilde{\varepsilon}_x) > 0$. When $\tilde{\varepsilon}_x = 0$,

$$\sigma_{2,d}(0)\sigma_2(0) - \sigma_3(0)\sigma_{1,d}(0) = \frac{|S^{d-1}|^2}{4d^2(d+2)^2} - \frac{|S^{d-2}|^2}{(d^2-1)^2(d+3)}, \quad (\text{F.5})$$

which is positive since we have proved $\frac{|S^{d-2}|^2}{|S^{d-1}|^2} < \frac{(d^2-1)^2}{4d^2(d+2)}$ in Lemma 29. Note that $\sigma_{2,d}(\tilde{\varepsilon}_x)\sigma_2(\tilde{\varepsilon}_x) - \sigma_3(\tilde{\varepsilon}_x)\sigma_{1,d}(\tilde{\varepsilon}_x)$ is increasing when $0 \leq \tilde{\varepsilon}_x \leq \varepsilon$. Hence, $\sigma_{2,d}(\tilde{\varepsilon}_x)\sigma_2(\tilde{\varepsilon}_x) - \sigma_3(\tilde{\varepsilon}_x)\sigma_{1,d}(\tilde{\varepsilon}_x) > 0$. Similarly, we can show that $\sigma_{2,d}(\tilde{\varepsilon}_x)\sigma_0(\tilde{\varepsilon}_x) - \sigma_{1,d}^2(\tilde{\varepsilon}_x) > 0$. Therefore, we conclude that $\phi_1(\tilde{\varepsilon}_x) > 0$.

Next, we study ϕ_2 . To prove $\phi_2(\tilde{\varepsilon}_x) < 0$ when $\tilde{\varepsilon}_x = 0$, it suffices to show $\sigma_{2,d}^2(0) - \sigma_{3,d}(0)\sigma_{1,d}(0) < 0$ since we have shown $\sigma_{2,d}(\tilde{\varepsilon}_x)\sigma_0(\tilde{\varepsilon}_x) - \sigma_{1,d}^2(\tilde{\varepsilon}_x) > 0$ above. When $\tilde{\varepsilon}_x = 0$, we have

$$\sigma_{2,d}^2(0) - \sigma_{3,d}(0)\sigma_{1,d}(0) = \frac{|S^{d-1}|^2}{4d^2(d+2)^2} - \frac{2|S^{d-2}|^2}{(d^2-1)^2(d+3)}, \quad (\text{F.6})$$

which is negative due to $\frac{|S^{d-2}|^2}{|S^{d-1}|^2} > \frac{(d^2-1)^2(d+3)}{8d^2(d+2)^2}$ proved in Lemma 29. We now check that $\sigma_{2,d}^2(\tilde{\varepsilon}_x) - \sigma_{3,d}(\tilde{\varepsilon}_x)\sigma_{1,d}(\tilde{\varepsilon}_x) > 0$ when $\tilde{\varepsilon}_x = \varepsilon$. Since $\sigma_{2,d}^2(\tilde{\varepsilon}_x) - \sigma_{3,d}(\tilde{\varepsilon}_x)\sigma_{1,d}(\tilde{\varepsilon}_x)$ is an increasing continuous function of $\tilde{\varepsilon}_x$, there is a unique $t^* = t^*(x) \in (0, \tilde{\varepsilon}_x)$ such that $\sigma_{2,d}^2(\tilde{\varepsilon}_x) - \sigma_{3,d}(\tilde{\varepsilon}_x)\sigma_{1,d}(\tilde{\varepsilon}_x) = 0$, and

hence $\phi_2(t^*) = 0$. We thus have

$$\begin{aligned} & \left[\frac{|S^{d-1}|}{2d(d+2)} + \frac{|S^{d-2}|}{d-1} \int_0^{\frac{t^*}{\varepsilon}} (1-z^2)^{\frac{d-1}{2}} z^2 dz \right]^2 \\ &= \frac{|S^{d-2}|^2}{(d^2-1)^2(d+3)} \left(2 + (d+1) \left(\frac{t^*}{\varepsilon} \right)^2 \right) \left(1 - \left(\frac{t^*}{\varepsilon} \right)^2 \right)^{d+1}. \end{aligned} \quad (\text{F.7})$$

Since t^* does not depend on x , the set \mathcal{S} is diffeomorphic to ∂M when ε is sufficiently small. Since $0 < \frac{t^*}{\varepsilon} < 1$, (F.7) becomes

$$\left[\frac{|S^{d-1}|}{2d(d+2)} + \frac{|S^{d-2}|}{d-1} \int_0^{\frac{t^*}{\varepsilon}} (1-z^2)^{\frac{d-1}{2}} z dz \right]^2 > \frac{2|S^{d-2}|^2}{(d^2-1)^2(d+3)} \left(1 - \left(\frac{t^*}{\varepsilon} \right)^2 \right)^{d+1}, \quad (\text{F.8})$$

which is equivalent to

$$\frac{|S^{d-1}|}{2d(d+2)} + \frac{|S^{d-2}|}{d^2-1} \left(1 - \left(1 - \left(\frac{t^*}{\varepsilon} \right)^2 \right)^{\frac{d+1}{2}} \right) > \sqrt{\frac{2}{d+3}} \frac{|S^{d-2}|}{d^2-1} \left(1 - \left(\frac{t^*}{\varepsilon} \right)^2 \right)^{\frac{d+1}{2}}. \quad (\text{F.9})$$

If we isolate t^* in the above equation, we have the lower bound for t^* :

$$\left(1 - \left[\frac{1 + \frac{(d^2-1)|S^{d-1}|}{2d(d+2)|S^{d-2}|}}{1 + \sqrt{\frac{2}{d+3}}} \right]^{\frac{2}{d+1}} \right)^{\frac{1}{2}} \varepsilon < t^*. \quad (\text{F.10})$$

Note that by Lemma 29, $\frac{(d^2-1)|S^{d-1}|}{2d(d+2)|S^{d-2}|} < \sqrt{\frac{2}{d+3}}$, so $1 - \left[\frac{1 + \frac{(d^2-1)|S^{d-1}|}{2d(d+2)|S^{d-2}|}}{1 + \sqrt{\frac{2}{d+3}}} \right]^{\frac{2}{d+1}} > 0$.

Next, we find the upper bound of t^* . Since $\frac{t^*}{\varepsilon} < 1$, by (F.7), we have,

$$\begin{aligned} & \left[\frac{|S^{d-1}|}{2d(d+2)} + \frac{|S^{d-2}|}{d-1} \int_0^{\frac{t^*}{\varepsilon}} (1-z^2)^{\frac{d-1}{2}} z^3 dz \right]^2 \\ & < \frac{|S^{d-2}|^2}{(d^2-1)^2(d+3)} \left(2 + (d+1) \left(\frac{t^*}{\varepsilon} \right)^2 \right) \left(1 - \left(\frac{t^*}{\varepsilon} \right)^2 \right)^{d+1}, \end{aligned} \quad (\text{F.11})$$

which is equivalent to

$$\begin{aligned} & \frac{|S^{d-1}|}{2d(d+2)} + \frac{|S^{d-2}|}{(d^2-1)(d+3)} \left(2 - \left(1 - \left(\frac{t^*}{\varepsilon} \right)^2 \right)^{\frac{d+1}{2}} \right) \left(2 + (d+1) \left(\frac{t^*}{\varepsilon} \right)^2 \right) \\ & < \frac{|S^{d-2}|}{(d^2-1)\sqrt{d+3}} \left(2 + (d+1) \left(\frac{t^*}{\varepsilon} \right)^2 \right)^{\frac{1}{2}} \left(1 - \left(\frac{t^*}{\varepsilon} \right)^2 \right)^{\frac{d+1}{2}}. \end{aligned} \quad (\text{F.12})$$

If we isolate t^* in the above equation, we have the lower bound,

$$t^* < \left(1 - \left[\frac{(d^2-1)|S^{d-1}|}{4d(d+2)|S^{d-2}|} + \frac{1}{d+3} \right]^{\frac{2}{d+1}} \right)^{\frac{1}{2}} \varepsilon. \quad (\text{F.13})$$

By Lemma 29 and the upper bound, $t^* \rightarrow 0$ as $d \rightarrow \infty$.

F.3 Proof of Theorem 15

In this proof, we calculate the first two order terms in $R_\varepsilon f(x)$. First, we are going to calculate $\mathbb{E}[\chi_{B_\varepsilon^{\mathbb{R}^p}(l(x))}(l(X))] - \mathbb{E}[(l(X) - l(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(l(x))}(l(X))]^\top \mathbf{T}(x)$ and show that it is dominated by the order ε^d terms. Then we are going to calculate $\mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(l(x))}(l(X))] - \mathbb{E}[(l(X) - l(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(l(x))}(l(X))]^\top \mathbf{T}(x)$ and show that it is dominated by the order ε^{d+2} terms. Hence their ratio is dominated by the order ε^2 terms.

By Lemma 30 and Lemma 32, we have

$$\begin{aligned}\mathbb{E}[\chi_{B_\varepsilon^{\mathbb{R}^p}(l(x))}(l(X))] &= P(x)\mu_0(x, \varepsilon) + O(\varepsilon^{d+1}), \\ \mathbb{E}[(l(X) - l(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(l(x))}(l(X))] &= [[P(x)\mu_{e_d}(x, \varepsilon)J_{p,d}^\top e_d + O(\varepsilon^{d+2}), O(\varepsilon^{d+2})]],\end{aligned}\quad (\text{F.14})$$

and $\mathbf{T}(x) = [v_1^{(-1)} + v_{1,1}^{(0)} + v_{1,2}^{(0)} + v_{1,3}^{(0)} + v_{1,4}^{(0)}, v_2^{(-1)}] + [O(\varepsilon), O(1)]$, where

$$v_1^{(-1)} = \frac{\mu_{e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)} J_{p,d}^\top e_d, \quad v_{1,1}^{(0)} = \frac{\nabla P(x)}{P(x)}, \quad (\text{F.15})$$

and $v_{1,2}^{(0)}, v_{1,3}^{(0)}, v_{1,4}^{(0)}$ and $v_2^{(-1)}$ are defined in Lemma 32. Moreover, $v_{1,2}^{(0)}, v_{1,3}^{(0)}$ and $v_{1,4}^{(0)}$ are of order 1 and $v_2^{(-1)}$ is of order ε^{-1} . Hence,

$$\begin{aligned}\mathbb{E}[\chi_{B_\varepsilon^{\mathbb{R}^p}(l(x))}(l(X))] - \mathbb{E}[(l(X) - l(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(l(x))}(l(X))]^\top \mathbf{T}(x) \\ = P(x) \left[\mu_0(x, \varepsilon) - \frac{\mu_{e_d}(x, \varepsilon)^2}{\mu_{2e_d}(x, \varepsilon)} \right] + O(\varepsilon^{d+1}), \\ = P(x) \left[\frac{\mu_0(x, \varepsilon)\mu_{2e_d}(x, \varepsilon) - \mu_{e_d}(x, \varepsilon)^2}{\mu_{2e_d}(x, \varepsilon)} \right] + O(\varepsilon^{d+1}),\end{aligned}\quad (\text{F.16})$$

where the leading term in above expression is of order ε^d by Lemma 27. Based on Lemma 30, we have

$$\begin{aligned}\mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(l(x))}(l(X))] \\ = P(x)\partial_d f(x)\mu_{e_d}(x, \varepsilon) + \sum_{i=1}^d \left[\frac{P(x)}{2} \partial_{ii}^2 f(x) + \partial_i f(x)\partial_i P(x) \right] \mu_{2e_i}(x, \varepsilon) + O(\varepsilon^{d+3}),\end{aligned}\quad (\text{F.17})$$

and

$$\mathbb{E}[(l(X) - l(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(l(x))}(l(X))] = [[v_1, v_2]], \quad (\text{F.18})$$

where

$$\begin{aligned}v_1 &= P(x) \sum_{i=1}^d (\partial_i f(x)\mu_{2e_i}(x, \varepsilon)) J_{p,d}^\top e_i \\ &\quad + \sum_{i=1}^{d-1} [\partial_i f(x)\partial_d P(x) + \partial_d f(x)\partial_i P(x) + P(x)\partial_{id}^2 f(x)] \mu_{2e_i+e_d}(x, \varepsilon) J_{p,d}^\top e_i \\ &\quad + \sum_{i=1}^d \left([\partial_i f(x)\partial_i P(x) + \frac{P(x)}{2} \partial_{ii}^2 f(x)] \mu_{2e_i+e_d}(x, \varepsilon) \right) J_{p,d}^\top e_d + O(\varepsilon^{d+4}), \\ v_2 &= P(x) \sum_{i=1}^{d-1} \partial_i f(x)\mathfrak{N}_{id}(x)\mu_{2e_i+e_d}(x, \varepsilon) + \frac{P(x)}{2} \partial_d f(x) \sum_{i=1}^d \mathfrak{N}_{ii}(x)\mu_{2e_i+e_d}(x, \varepsilon) + O(\varepsilon^{d+4}).\end{aligned}$$

Therefore, we have

$$\begin{aligned}
 & \mathbb{E}[(\iota(X) - \iota(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \mathbf{T}(x) \\
 &= P(x) \sum_{i=1}^d (\partial_i f(x) \mu_{2e_i}(x, \varepsilon)) v_1^{(-1)\top} J_{p,d}^\top e_i \\
 &\quad + P(x) \sum_{i=1}^d (\partial_i f(x) \mu_{2e_i}(x, \varepsilon)) v_{1,1}^{(0)\top} J_{p,d}^\top e_i \\
 &\quad + P(x) \sum_{i=1}^d (\partial_i f(x) \mu_{2e_i}(x, \varepsilon)) [v_{1,2}^{(0)} + v_{1,3}^{(0)} + v_{1,4}^{(0)}]^\top J_{p,d}^\top e_i \\
 &\quad + \sum_{i=1}^{d-1} [\partial_i f(x) \partial_d P(x) + \partial_d f(x) \partial_i P(x) + P(x) \partial_{id}^2 f(x)] \mu_{2e_i+e_d}(x, \varepsilon) v_1^{(-1)\top} J_{p,d}^\top e_i \\
 &\quad + \sum_{i=1}^d [\partial_i f(x) \partial_i P(x) + \frac{P(x)}{2} \partial_{ii}^2 f(x)] \mu_{2e_i+e_d}(x, \varepsilon) v_1^{(-1)\top} J_{p,d}^\top e_d \\
 &\quad + P(x) \sum_{i=1}^{d-1} \partial_i f(x) \mu_{2e_i+e_d}(x, \varepsilon) v_2^{(-1)\top} \mathfrak{N}_{id}(x) \\
 &\quad + \frac{P(x)}{2} \partial_d f(x) \sum_{i=1}^d \mu_{2e_i+e_d}(x, \varepsilon) v_2^{(-1)\top} \mathfrak{N}_{ii}(x) + O(\varepsilon^{d+3}).
 \end{aligned} \tag{F.19}$$

Note that by Lemma 27, the first term is of order ε^{d+1} and the second to seventh terms are of order ε^{d+2} . Furthermore, we can simplify the first and the second term as:

$$\begin{aligned}
 & P(x) \sum_{i=1}^d (\partial_i f(x) \mu_{2e_i}(x, \varepsilon)) v_1^{(-1)\top} J_{p,d}^\top e_i = P(x) \partial_d f(x) \mu_{e_d}(x, \varepsilon) \\
 & P(x) \sum_{i=1}^d (\partial_i f(x) \mu_{2e_i}(x, \varepsilon)) v_{1,1}^{(0)\top} J_{p,d}^\top e_i = \sum_{i=1}^d \partial_i f(x) \partial_i P(x) \mu_{2e_i}(x, \varepsilon).
 \end{aligned} \tag{F.20}$$

Next we calculate $\mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] - \mathbb{E}[(\iota(X) - \iota(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \mathbf{T}(x)$. Clearly, the common terms, $P(x) \partial_d f(x) \mu_{e_d}(x, \varepsilon)$ and $\sum_{i=1}^d \partial_i f(x) \partial_i P(x) \mu_{2e_i}(x, \varepsilon)$, are canceled, and

hence only terms of order ε^{d+2} are left in the difference; that is, we have

$$\begin{aligned}
 & \mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] - \mathbb{E}[(\iota(X) - \iota(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \mathbf{T}(x) \\
 &= \frac{P(x)}{2} \sum_{i=1}^d \partial_{ii}^2 f(x) \mu_{2e_i}(x, \varepsilon) - P(x) \sum_{i=1}^d (\partial_i f(x) \mu_{2e_i}(x, \varepsilon)) [v_{1,2}^{(0)} + v_{1,3}^{(0)} + v_{1,4}^{(0)}]^\top J_{p,d}^\top e_i \\
 &\quad - \sum_{i=1}^{d-1} [\partial_i f(x) \partial_d P(x) + \partial_d f(x) \partial_i P(x) + P(x) \partial_{id}^2 f(x)] \mu_{2e_i+e_d}(x, \varepsilon) v_1^{(-1)\top} J_{p,d}^\top e_i \\
 &\quad - \sum_{i=1}^d [\partial_i f(x) \partial_i P(x) + \frac{P(x)}{2} \partial_{ii}^2 f(x)] \mu_{2e_i+e_d}(x, \varepsilon) v_1^{(-1)\top} J_{p,d}^\top e_d \\
 &\quad - P(x) \sum_{i=1}^{d-1} \partial_i f(x) \mu_{2e_i+e_d}(x, \varepsilon) v_2^{(-1)\top} \mathfrak{N}_{id}(x) \\
 &\quad - \frac{P(x)}{2} \partial_d f(x) \sum_{i=1}^d \mu_{2e_i+e_d}(x, \varepsilon) v_2^{(-1)\top} \mathfrak{N}_{ii}(x) + O(\varepsilon^{d+3}).
 \end{aligned}$$

Next, we simplify the above expression. Note that $v_1^{(-1)\top} J_{p,d}^\top e_i = \frac{\mu_{e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)}$ if $i = d$, and it is 0 otherwise. Hence,

$$- \sum_{i=1}^{d-1} [\partial_i f(x) \partial_d P(x) + \partial_d f(x) \partial_i P(x) + P(x) \partial_{id}^2 f(x)] \mu_{2e_i+e_d}(x, \varepsilon) v_1^{(-1)\top} J_{p,d}^\top e_i = 0$$

and by definition of $v_{1,3}^{(0)}$ and $v_1^{(-1)}$, we have

$$P(x) \mu_{2e_i}(x, \varepsilon) v_{1,3}^{(0)\top} J_{p,d}^\top e_i + \partial_i P(x) \mu_{2e_i+e_d}(x, \varepsilon) v_1^{(-1)\top} J_{p,d}^\top e_d = 0. \quad (\text{F.21})$$

For $i = 1, \dots, d-1$, by definition of $v_{1,4}^{(0)}$ and $v_2^{(-1)}$, we have

$$P(x) \mu_{2e_i}(x, \varepsilon) v_{1,4}^{(0)\top} J_{p,d}^\top e_i + P(x) \mu_{2e_i+e_d}(x, \varepsilon) v_2^{(-1)\top} \mathfrak{N}_{id}(x) = 0 \quad (\text{F.22})$$

and

$$P(x) \mu_{2e_d}(x, \varepsilon) v_{1,4}^{(0)\top} J_{p,d}^\top e_d + \frac{P(x)}{2} \sum_{i=1}^d \mu_{2e_i+e_d}(x, \varepsilon) v_2^{(-1)\top} \mathfrak{N}_{ii}(x) = 0. \quad (\text{F.23})$$

Moreover, we have $v_{1,2}^{(0)\top} J_{p,d}^\top e_i = -\frac{\mu_{e_d}(x, \varepsilon) \varepsilon^{d+3}}{P(x)(\mu_{2e_d}(x, \varepsilon))^2}$ if $i = d$, and it is 0 otherwise. Therefore,

$$\begin{aligned}
 & \mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] - \mathbb{E}[(\iota(X) - \iota(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]^\top \mathbf{T}(x) \\
 &= \frac{P(x)}{2} \sum_{i=1}^d \partial_{ii}^2 f(x) [\mu_{2e_i}(x, \varepsilon) - \mu_{2e_i+e_d}(x, \varepsilon) \frac{\mu_{e_d}(x, \varepsilon)}{\mu_{2e_d}(x, \varepsilon)}] + \partial_d f(x) \left(\frac{\mu_{e_d}(x, \varepsilon) \varepsilon^{d+3}}{\mu_{2e_d}(x, \varepsilon)} \right).
 \end{aligned}$$

Therefore, the ratio

$$\begin{aligned}
 & \frac{\mathbb{E}[(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] - \mathbb{E}[(\iota(X) - \iota(x))(f(X) - f(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]\top \mathbf{T}(x)}{\mathbb{E}[\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))] - \mathbb{E}[(\iota(X) - \iota(x))\chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x))}(\iota(X))]\top \mathbf{T}(x)} \\
 &= \sum_{i=1}^d \partial_{ii}^2 f(x) \left[\frac{\mu_{2e_i}(x, \varepsilon)\mu_{2e_d}(x, \varepsilon) - \mu_{2e_i+e_d}(x, \varepsilon)\mu_{e_d}(x, \varepsilon)}{2\mu_0(x, \varepsilon)\mu_{2e_d}(x, \varepsilon) - 2\mu_{e_d}(x, \varepsilon)^2} \right] \\
 &\quad + \partial_d f(x) \frac{\mu_{e_d}(x, \varepsilon)\varepsilon^{d+3}}{P(x)(\mu_0(x, \varepsilon)\mu_{2e_d}(x, \varepsilon) - \mu_{e_d}(x, \varepsilon)^2)} + O(\varepsilon^3).
 \end{aligned} \tag{F.24}$$

And the conclusion follows by substituting terms and in Corollary 28.

Appendix G. Variance analysis on LLE

For simplicity of notations, for each x_k , denote

$$\mathbf{f} := (f(x_{k,1}), f(x_{k,2}), \dots, f(x_{k,N}))^\top \in \mathbb{R}^N.$$

By a direct expansion of equations (2.5), (2.6), (2.8) and $c = n\varepsilon^{d+3}$, we have

$$\sum_{j=1}^n [W - I_{n \times n}]_{kj} f(x_j) = \frac{\mathbf{1}_N^\top \mathbf{f} - \mathbf{1}_N^\top G_{n,k}^\top U_n I_{p,r_n} (\Lambda_n + n\varepsilon^{d+3} I_{p \times p})^{-1} U_n^\top G_{n,k} \mathbf{f}}{N - \mathbf{1}_N^\top G_{n,k}^\top U_n I_{p,r_n} (\Lambda_n + n\varepsilon^{d+3} I_{p \times p})^{-1} U_n^\top G_{n,k} \mathbf{1}_N} - f(x_k), \tag{G.1}$$

which can be rewritten as $\frac{g_{n,1}}{g_{n,2}}$, where

$$\begin{aligned}
 g_{n,1} &:= \frac{1}{n\varepsilon^d} \sum_{j=1}^N (f(x_{k,j}) - f(x_k)) - \left[\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) \right]^\top U_n I_{p,r_n} \left(\frac{\Lambda_n}{n\varepsilon^d} + \varepsilon^3 I_{p \times p} \right)^{-1} \\
 &\quad \times U_n^\top \left[\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) (f(x_{k,j}) - f(x_k)) \right] \\
 g_{n,2} &:= \frac{N}{n\varepsilon^d} - \left[\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) \right]^\top U_n I_{p,r_n} \left(\frac{\Lambda_n}{n\varepsilon^d} + \varepsilon^3 I_{p \times p} \right)^{-1} U_n^\top \left[\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) \right].
 \end{aligned}$$

The goal is to relate the finite sum quantity $\frac{g_{n,1}}{g_{n,2}}$ to $Q_\varepsilon f(x_k) := \frac{g_1}{g_2}$, where

$$\begin{aligned}
 g_1 &= \mathbb{E} \left[\frac{1}{\varepsilon^d} \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x_k))}(X) (f(X) - f(x_k)) \right] - \mathbb{E} \left[\frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x_k))}(X) \right]^\top \\
 &\quad \times \left(U I_{p,r} \left(\frac{\Lambda}{\varepsilon^d} + \varepsilon^3 I_{p \times p} \right)^{-1} U^\top \right) \mathbb{E} \left[\frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x_k))}(X) (f(X) - f(x_k)) \right]
 \end{aligned} \tag{G.2}$$

and

$$\begin{aligned}
 g_2 &= \mathbb{E} \left[\frac{1}{\varepsilon^d} \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x_k))}(X) \right] - \mathbb{E} \left[\frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x_k))}(X) \right]^\top \\
 &\quad \times \left(U I_{p,r} \left(\frac{\Lambda}{\varepsilon^d} + \varepsilon^3 I_{p \times p} \right)^{-1} U^\top \right) \mathbb{E} \left[\frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(\iota(x_k))}(X) \right].
 \end{aligned} \tag{G.3}$$

We now control the size of the fluctuation of the following four terms

$$\frac{1}{n\epsilon^d} \sum_{j=1}^N 1 \quad (\text{G.4})$$

$$\frac{1}{n\epsilon^d} \sum_{j=1}^N (f(x_{k,j}) - f(x_k)) \quad (\text{G.5})$$

$$\frac{1}{n\epsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) \quad (\text{G.6})$$

$$\frac{1}{n\epsilon^d} \sum_{j=1}^N (x_{k,j} - x_k)(f(x_{k,j}) - f(x_k)) \quad (\text{G.7})$$

as a function of n and ϵ by the Bernstein type inequality. Here, we put ϵ^{-d} in front of each term to normalize the kernel so that the computation is consistent with the existing literature, like (Cheng and Wu, 2013; Singer and Wu, 2017).

The size of the fluctuation of these terms are controlled in the following Lemmas. The term (G.4) is the usual kernel density estimation, so we have the following lemma.

Lemma 33 Suppose $\epsilon = \epsilon(n)$ so that $\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2+1}} \rightarrow 0$ and $\epsilon \rightarrow 0$ as $n \rightarrow \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \dots, n$,

$$\left| \frac{1}{n\epsilon^d} \sum_{j=1}^N 1 - \mathbb{E} \frac{1}{\epsilon^d} \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) \right| = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2}}\right).$$

Denote Ω_0 to be the event space that above Lemma is satisfied. The behavior of (G.5) is summarized in the following Lemma.

Lemma 34 Suppose $\epsilon = \epsilon(n)$ so that $\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2+1}} \rightarrow 0$ and $\epsilon \rightarrow 0$ as $n \rightarrow \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \dots, n$,

$$\left| \frac{1}{n\epsilon^d} \sum_{j=1}^N (f(x_{k,j}) - f(x_k)) - \mathbb{E} \frac{1}{\epsilon^d} (f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) \right| = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-1}}\right).$$

Proof By denoting

$$F_{1,j} = \frac{1}{\epsilon^d} (f(x_j) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(x_j), \quad (\text{G.8})$$

we have

$$\frac{1}{n\epsilon^d} \sum_{j=1}^N (f(x_{k,j}) - f(x_k)) = \frac{1}{n} \sum_{j \neq k, j=1}^n F_{1,j}. \quad (\text{G.9})$$

Define a random variable

$$F_1 := \frac{1}{\epsilon^d} (f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)). \quad (\text{G.10})$$

Clearly, when $j \neq k$, $F_{1,j}$ can be viewed as randomly sampled i.i.d. from F_1 . Note that we have

$$\frac{1}{n} \sum_{j \neq k, j=1}^n F_{1,j} = \frac{n-1}{n} \left[\frac{1}{n-1} \sum_{j \neq k, j=1}^n F_{1,j} \right]. \quad (\text{G.11})$$

Since $\frac{n-1}{n} \rightarrow 1$ as $n \rightarrow \infty$, the error incurred by replacing $\frac{1}{n}$ by $\frac{1}{n-1}$ is of order $\frac{1}{n}$, which is negligible asymptotically, we can simply focus on analyzing $\frac{1}{n-1} \sum_{j=1, j \neq i}^n F_{1,j}$. We have by Lemma 27 and Lemma 30,

$$\begin{aligned} \mathbb{E}[F_1] &= O(\varepsilon) \quad \text{if } x \in M_\varepsilon \\ \mathbb{E}[F_1] &= O(\varepsilon^2) \quad \text{if } x \notin M_\varepsilon \end{aligned} \quad (\text{G.12})$$

and

$$\mathbb{E}[F_1^2] = \sum_{i=1}^d P(x_k) (\partial_i f(x_k))^2 \mu_{2e_i}(x_k, \varepsilon) \varepsilon^{-2d} + O(\varepsilon^{-d+3}), \quad (\text{G.13})$$

By Lemma 27, $\frac{|S^{d-1}|}{2d(d+2)} \varepsilon^{-d+2} + O(\varepsilon^{-d+3}) \leq \mu_{2e_i}(x_k, \varepsilon) \varepsilon^{-2d} \leq \frac{|S^{d-1}|}{d(d+2)} \varepsilon^{-d+2}$, therefore, in any case,

$$\sigma_1^2 := \text{Var}(F_1) \leq \frac{|S^{d-1}| \|P\|_{L^\infty}}{d(d+2)} \varepsilon^{-d+2} + O(\varepsilon^{-d+3}). \quad (\text{G.14})$$

With the above bounds, we could apply the large deviation theory. First, note that the random variable F_1 is uniformly bounded by

$$c_1 = 2 \|f\|_{L^\infty} \varepsilon^{-d}, \quad (\text{G.15})$$

so we apply Bernstein's inequality to provide a large deviation bound. Recall Bernstein's inequality

$$\Pr \left\{ \frac{1}{n-1} \sum_{j \neq k, j=1}^n (F_{1,j} - \mathbb{E}[F_1]) > \eta_1 \right\} \leq e^{-\frac{n\eta_1^2}{2\sigma_1^2 + \frac{2}{3}c_1\eta_1}}, \quad (\text{G.16})$$

where $\eta_1 > 0$. Note that $\mathbb{E}[F_1] = O(\varepsilon)$, if $x_k \in M_\varepsilon$ and $\mathbb{E}[F_1] = O(\varepsilon^2)$, if $x_k \notin M_\varepsilon$. Hence, we assume $\eta_1 = O(\varepsilon^{2+s})$, where $s > 0$. Then $c_1 \eta_1 = O(\varepsilon^{-d+2+s})$. If ε is small enough, $2\sigma_1^2 + \frac{2}{3}c_1\eta_1 \leq C\varepsilon^{-d+2}$ for some constant C which depends on f and P . We have,

$$\frac{n\eta_1^2}{2\sigma_1^2 + \frac{2}{3}c_1\eta_1} \geq \frac{n\eta_1^2 \varepsilon^{d-2}}{C}. \quad (\text{G.17})$$

Suppose n is chosen large enough so that

$$\frac{n\eta_1^2 \varepsilon^{d-2}}{C} \geq 3 \log(n); \quad (\text{G.18})$$

that is, the deviation from the mean is set to

$$\eta_1 \geq O\left(\frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2-1}}\right). \quad (\text{G.19})$$

Note that by the assumption that $\eta_1 = O(\varepsilon^{2+s})$, we know that $\eta_1/\varepsilon^2 = \frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2+1}} \rightarrow 0$. It implies that the deviation greater than η_1 happens with probability less than

$$\exp\left(-\frac{n\eta_1^2}{2\sigma_1^2 + \frac{2}{3}c_1\eta_1}\right) \leq \exp\left(-\frac{n\eta_1^2\varepsilon^{d-2}}{C}\right) = \exp(-3\log(n)) = 1/n^3. \quad (\text{G.20})$$

As a result, by a simple union bound, we have

$$\Pr\left\{\frac{1}{n-1} \sum_{j \neq k, j=1}^n (F_{1,j} - \mathbb{E}[F_1]) > \eta_1 \mid k = 1, \dots, n\right\} \leq ne^{-\frac{n\eta_1^2}{2\sigma_1^2 + \frac{2}{3}c_1\eta_1}} \leq 1/n^2. \quad (\text{G.21})$$

■

Denote Ω_1 to be the event space that the deviation $\frac{1}{n-1} \sum_{j \neq k, j=1}^n (F_{1,j} - \mathbb{E}[F_1]) \leq \eta_1$ for all $i = 1, \dots, n$, where η_1 is chosen in (G.19) is satisfied.

Lemma 35 Suppose $\varepsilon = \varepsilon(n)$ so that $\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2+1}} \rightarrow 0$ and $\varepsilon \rightarrow 0$ as $n \rightarrow \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \dots, n$,

$$e_i^\top \left[\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) - \mathbb{E} \frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) \right] = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-1}}\right), \quad (\text{G.22})$$

where $i = 1, \dots, d$. And

$$e_i^\top \left[\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) - \mathbb{E} \frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) \right] = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-2}}\right), \quad (\text{G.23})$$

where $i = d+1, \dots, p$.

Proof Fix x_k . By denoting

$$\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) = \frac{1}{n} \sum_{j \neq k, j=1}^n \sum_{\ell=1}^p F_{2,\ell,j} e_\ell. \quad (\text{G.24})$$

where

$$F_{2,\ell,j} := \frac{1}{\varepsilon^d} e_\ell^\top (x_j - x_k) \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(x_j), \quad (\text{G.25})$$

and we know that when $j \neq k$, $F_{2,\ell,j}$ is randomly sampled i.i.d. from

$$F_{2,\ell} := \frac{1}{\varepsilon^d} e_\ell^\top (\iota(X) - \iota(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(\iota(X)). \quad (\text{G.26})$$

Similarly, we can focus on analyzing $\frac{1}{n-1} \sum_{j=1, j \neq i}^n F_{2,\ell,j}$ since $\frac{n-1}{n} \rightarrow 1$ as $n \rightarrow \infty$. By Lemma 30 we have

$$\mathbb{E}[F_{2,\ell}] = \begin{cases} (P(x)\mu_{e_d}(x, \varepsilon)\varepsilon^{-d})e_\ell^\top e_d + \sum_{i=1}^d (\partial_i P(x)\mu_{2e_i}(x, \varepsilon)\varepsilon^{-d})e_\ell^\top e_i + O(\varepsilon^{d+3}) & \text{when } \ell = 1, \dots, d \\ \frac{P(x)\varepsilon^{-d}}{2} e_\ell^\top \sum_{i=1}^d \mathfrak{N}_{ii}(x)\mu_{2e_i} + O(\varepsilon^{d+3}) & \text{when } \ell = d+1, \dots, p. \end{cases}$$

In other words, by Lemma 27, for $\ell = 1, \dots, d$ we have $\mathbb{E}[F_{2,\ell}] = O(\varepsilon)$ if $x_k \in M_\varepsilon$, and $\mathbb{E}[F_{2,\ell}] = O(\varepsilon^2)$ if $x_k \notin M_\varepsilon$. Moreover, $\mathbb{E}[F_{2,\ell}] = O(\varepsilon^2)$ for $\ell = d+1, \dots, p$. By (D.8) we have, for $\ell = 1, \dots, d$

$$\mathbb{E}[F_{2,\ell}^2] \leq C_\ell \varepsilon^{-d+2} + O(\varepsilon^{-d+3}), \quad (\text{G.27})$$

and C_ℓ depends on $\|P\|_{L^\infty}$. For $\ell = d+1, \dots, p$,

$$\mathbb{E}[F_{2,\ell}^2] \leq C_\ell \varepsilon^{-d+4} + O(\varepsilon^{-d+5}), \quad (\text{G.28})$$

and C_ℓ depends on $\|P\|_{L^\infty}$ and second fundamental form of M .

Thus, we conclude that

$$\begin{aligned} \sigma_{2,\ell}^2 &\leq C_\ell \varepsilon^{-d+2} + O(\varepsilon^{-d+3}) \text{ when } \ell = 1, \dots, d \\ \sigma_{2,\ell}^2 &\leq C_\ell \varepsilon^{-d+4} + O(\varepsilon^{-d+5}) \text{ when } \ell = d+1, \dots, p. \end{aligned} \quad (\text{G.29})$$

Note that for $\ell = d+1, \dots, p$, the variance is of higher order than that of $\ell = 1, \dots, d$.

With the above bounds, we could apply the large deviation theory. For $\ell = 1, \dots, d$, the random variable $F_{2,\ell}$ is uniformly bounded by $c_{2,\ell} = 2\varepsilon^{-d+1}$. Since $\mathbb{E}[F_{2,\ell}] = O(\varepsilon)$ if $x_k \in M_\varepsilon$, and $\mathbb{E}[F_{2,\ell}] = O(\varepsilon^2)$ if $x_k \notin M_\varepsilon$, we assume $\eta_{2,\ell} = O(\varepsilon^{2+s})$, where $s > 0$. Then $c_{2,\ell}\eta_{2,\ell} = O(\varepsilon^{-d+3+s})$. If ε is small enough, $2\sigma_{2,\ell}^2 + \frac{2}{3}c_{2,\ell}\eta_{2,\ell} \leq C\varepsilon^{-d+2}$ for some constant C which depends on P and manifold M . We have

$$\frac{n\eta_{2,\ell}^2}{2\sigma_{2,\ell}^2 + \frac{2}{3}c_{2,\ell}\eta_{2,\ell}} \geq \frac{n\eta_{2,\ell}^2 \varepsilon^{d-2}}{C}. \quad (\text{G.30})$$

Suppose n is chosen large enough so that

$$\frac{n\eta_{2,\ell}^2 \varepsilon^{d-2}}{C} \geq 3\log(n); \quad (\text{G.31})$$

that is, the deviation from the mean is set to

$$\eta_{2,\ell} \geq O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-1}}\right). \quad (\text{G.32})$$

Note that by the assumption that $\eta_{2,\ell} = O(\varepsilon^{2+s})$, we know that $\eta_{2,\ell}/\varepsilon^2 = \frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2+1}} \rightarrow 0$. Thus, when ε is sufficiently smaller and n is sufficiently large, the exponent in Bernstein's inequality

$$\Pr\left\{\frac{1}{n-1} \sum_{j \neq k, j=1}^n (F_{2,\ell,j} - \mathbb{E}[F_{2,\ell}]) > \eta_{2,\ell}\right\} \leq \exp\left(-\frac{n\eta_{2,\ell}^2}{2\sigma_{2,\ell}^2 + \frac{2}{3}c_{2,\ell}\eta_{2,\ell}}\right) \leq \frac{1}{n^3}. \quad (\text{G.33})$$

By a simple union bound, for $\ell = 1, \dots, d$, we have

$$\Pr\left\{\left|\frac{1}{n} \sum_{j \neq k, j=1}^n F_{2,\ell,j} - \mathbb{E}[F_{2,\ell}]\right| > \eta_{2,\ell} \mid k = 1, \dots, n\right\} \leq 1/n^2. \quad (\text{G.34})$$

For $\ell = d+1, \dots, p$, the random variable $F_{2,\ell}$ is uniformly bounded by $c_{2,\ell} = 2\varepsilon^{-d+1}$. Since $\mathbb{E}[F_{2,\ell}] = O(\varepsilon^2)$ for $\ell = d+1, \dots, p$, we assume $\eta_{2,\ell} = O(\varepsilon^{3+s})$, where $s > 0$. Then $c_{2,\ell}\eta_{2,\ell} =$

$O(\varepsilon^{-d+4+s})$. If ε is small enough, $2\sigma_{2,\ell}^2 + \frac{2}{3}c_{2,\ell}\eta_{2,\ell} \leq C\varepsilon^{-d+4}$ for some constant C which depends on M and P . We have,

$$\frac{n\eta_{2,\ell}^2}{2\sigma_{2,\ell}^2 + \frac{2}{3}c_{2,\ell}\eta_{2,\ell}} \geq \frac{n\eta_{2,\ell}^2\varepsilon^{d-4}}{C}. \quad (\text{G.35})$$

Suppose n is chosen large enough so that

$$\frac{n\eta_{2,\ell}^2\varepsilon^{d-4}}{C} = 3\log(n); \quad (\text{G.36})$$

that is, the deviation from the mean is set to

$$\eta_{2,\ell} = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-2}}\right). \quad (\text{G.37})$$

Note that by the assumption that $\beta_1 = O(\varepsilon^{3+s})$, we know that $\eta_{2,\ell}/\varepsilon^3 = \frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2+1}} \rightarrow 0$.

By a similar argument, for $\ell = d+1, \dots, p$, we have

$$\Pr\left\{\left|\frac{1}{n}\sum_{j \neq k, j=1}^n F_{2,\ell,j} - \mathbb{E}[F_{2,\ell}]\right| > \eta_{2,\ell} \mid k = 1, \dots, n\right\} \leq 1/n^2. \quad (\text{G.38})$$

■

Denote Ω_2 to be the event space that the deviation $\left|\frac{1}{n}\sum_{j \neq k, j=1}^n F_{2,\ell,j} - \mathbb{E}[F_{2,\ell}]\right| \leq \eta_{2,\ell}$ for all $\ell = 1, \dots, p$ and $k = 1, \dots, n$, where $\eta_{2,\ell}$ are chosen in (G.32) and (G.37). Next Lemma summarizes behavior of (G.7) and can be proved similarly as Lemma 35.

Lemma 36 Suppose $\varepsilon = \varepsilon(n)$ so that $\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2+1}} \rightarrow 0$ and $\varepsilon \rightarrow 0$ as $n \rightarrow \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \dots, n$,

$$e_i^\top \left[\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k)(f(x_{k,j}) - f(x_k)) - \mathbb{E} \frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k))(f(X) - f(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) \right] = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-2}}\right),$$

where $i = 1, \dots, d$, and

$$e_i^\top \left[\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k)(f(x_{k,j}) - f(x_k)) - \mathbb{E} \frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k))(f(X) - f(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) \right] = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-3}}\right),$$

where $i = d+1, \dots, p$.

Denote Ω_3 to be the event space that Lemma 36 is satisfied. In the next two lemmas, we describe the behavior of $\frac{1}{n\varepsilon^d} G_{n,k} G_{n,k}^\top$. The proofs are the same as Lemma E.4 in Wu and Wu (2018) with $\rho = 3$.

Lemma 37 Suppose $\varepsilon = \varepsilon(n)$ so that $\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2+1}} \rightarrow 0$ and $\varepsilon \rightarrow 0$ as $n \rightarrow \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \dots, n$,

$$\left| e_i^\top \left(\frac{1}{n\varepsilon^d} G_{n,k} G_{n,k}^\top - \frac{1}{\varepsilon^d} C_{x_k} \right) e_j \right| = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-2}}\right), \quad (\text{G.39})$$

where $i, j = 1, \dots, d$.

$$\left| e_i^\top \left(\frac{1}{n\varepsilon^d} G_{n,k} G_{n,k}^\top - \frac{1}{\varepsilon^d} C_{x_k} \right) e_j \right| = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-4}}\right), \quad (\text{G.40})$$

where $i, j = 1 + 1, \dots, p$.

$$\left| e_i^\top \left(\frac{1}{n\varepsilon^d} G_{n,k} G_{n,k}^\top - \frac{1}{\varepsilon^d} C_{x_k} \right) e_j \right| = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-3}}\right), \quad (\text{G.41})$$

otherwise.

Lemma 38 $r_n \leq r$ and r_n is a non decreasing function of n . If n is large enough, $r_n = r$. Suppose $\varepsilon = \varepsilon(n)$ so that $\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2+1}} \rightarrow 0$ and $\varepsilon \rightarrow 0$ as $n \rightarrow \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \dots, n$,

$$\left| e_i^\top \left[I_{p,r_n} \left(\frac{\Lambda_n}{n\varepsilon^d} + \varepsilon^3 I_{p \times p} \right)^{-1} - I_{p,r} \left(\frac{\Lambda}{\varepsilon^d} + \varepsilon^3 I_{p \times p} \right)^{-1} \right] e_i \right| = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2+2}}\right) \quad (\text{G.42})$$

for $i = 1, \dots, r$ and

$$U_n = U\Theta + \frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-2}} U\Theta S + O\left(\frac{\log(n)}{n\varepsilon^{d-4}}\right), \quad (\text{G.43})$$

where $S \in \mathfrak{o}(p)$, and $\Theta \in O(p)$. Θ commutes with $I_{p,r}(\frac{\Lambda}{\varepsilon^d} + \varepsilon^3 I_{p \times p})^{-1}$.

Denote Ω_4 to be the event space that Lemma 38 is satisfied. In the proofs of Lemma 32 and Theorem 15, we need the order ε^{d+3} terms of the eigenvalues $\{\lambda_i\}$ of C_x for $i = 1, \dots, d$ and we need the order ε term of the eigenvectors $\{\beta_i\}$ of C_x for $i = 1, \dots, p$. We also use the fact that $\{\lambda_i\}$ of C_x for $i = d+1, \dots, p$ are of order ε^{d+4} , so that we can calculate the leading terms (order ε^2) of $Q_\varepsilon f(x)$ for all $x \in M$. Since $\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2+1}} \rightarrow 0$, the above two lemmas imply that the differences between the first d eigenvalues of $\frac{1}{n\varepsilon^d} G_{n,k} G_{n,k}^\top$ and $\frac{1}{\varepsilon^d} C_{x_k}$ are less than $O(\varepsilon^3)$. The differences between the rest of the eigenvalues of $\frac{1}{n\varepsilon^d} G_{n,k} G_{n,k}^\top$ and $\frac{1}{\varepsilon^d} C_{x_k}$ are less than $O(\varepsilon^4)$. In other words, we can make sure that the rest of the eigenvalues of $\frac{1}{n\varepsilon^d} G_{n,k} G_{n,k}^\top$ are of order ε^4 . Moreover U_n and $U\Theta$ differ by a matrix of order ε^3 . Consequently, in the following proof, we can show that the deviation between $\sum_{j=1}^n [W - I_{n \times n}]_{kj} f(x_{k,j})$ and $Q_\varepsilon f(x_k)$ is less than ε^2 for all x_k .

Proof [Proof of Theorem 11.] Denote $\Omega := \cap_{i=0, \dots, 4} \Omega_i$. By a direct union bound, the probability of the event space Ω is great than $1 - n^{-2}$. Below, all arguments are conditional on Ω . Based on previous lemmas, we have, for $k = 1, \dots, n$,

$$\frac{1}{n\varepsilon^d} \sum_{j=1}^N 1 = \mathbb{E} \frac{1}{\varepsilon^d} \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2}}\right), \quad (\text{G.44})$$

$$\frac{1}{n\epsilon^d} \sum_{j=1}^N (f(x_{k,j}) - f(x_k)) = \mathbb{E} \frac{1}{\epsilon^d} (f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-1}}\right), \quad (\text{G.45})$$

and

$$\frac{1}{n\epsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) = \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) + \mathcal{E}_1, \quad (\text{G.46})$$

where $\mathcal{E}_1 \in \mathbb{R}^p$, $e_i^\top \mathcal{E}_1 = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-1}}\right)$ for $i = 1, \dots, d$, and $e_i^\top \mathcal{E}_1 = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-2}}\right)$ for $i = d+1, \dots, p$. Moreover, we have

$$\begin{aligned} & \frac{1}{n\epsilon^d} \sum_{j=1}^N (x_{k,j} - x_k)(f(x_{k,j}) - f(x_k)) \\ &= \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k))(f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) + \mathcal{E}_2, \end{aligned} \quad (\text{G.47})$$

where $\mathcal{E}_2 \in \mathbb{R}^p$. $e_i^\top \mathcal{E}_2 = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-2}}\right)$ for $i = 1, \dots, d$, and $e_i^\top \mathcal{E}_2 = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-3}}\right)$ for $i = d+1, \dots, p$. Therefore, we have

$$\begin{aligned} & U_n I_{p,r_n} \left(\frac{\Lambda_n}{n\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} U_n^\top - U I_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} U^\top \\ &= \left(U \Theta + \frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-2}} U \Theta S + O\left(\frac{\log(n)}{n\epsilon^{d-4}}\right) \right) \left(I_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2+2}}\right) \right) \\ &\quad \times \left(U \Theta + \frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-2}} U \Theta S + O\left(\frac{\log(n)}{n\epsilon^{d-4}}\right) \right)^\top - U I_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} U^\top. \\ &= \frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-2}} U \Theta \left(S I_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} + I_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} S^\top \right) \Theta^\top U^\top \\ &\quad + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2+2}}\right) I_{p \times p} + [\text{higher order terms}]. \end{aligned}$$

Define a $p \times p$ matrix

$$\begin{aligned} \mathcal{E}_3 &= \frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-2}} U \Theta \left[S I_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} + I_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} S^\top \right] \Theta^\top U^\top \\ &\quad + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2+2}}\right) I_{p \times p}. \end{aligned} \quad (\text{G.48})$$

We have

$$\begin{aligned}
 & \left[\frac{1}{n\epsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) \right]^\top U_n I_{p,r_n} \left(\frac{\Lambda_n}{n\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} U_n^\top \left[\frac{1}{n\epsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) (f(x_{k,j}) - f(x_k)) \right] \\
 &= [\mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) + \mathcal{E}_1]^\top [UI_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} U^\top + \mathcal{E}_3 + \text{higher order terms}] \\
 &\quad \times [\mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) (f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) + \mathcal{E}_2] \\
 &= \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X))^\top [UI_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} U^\top] \\
 &\quad \times \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) (f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) \\
 &\quad + \mathcal{E}_1^\top UI_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} U^\top \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) (f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) \\
 &\quad + \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X))^\top \mathcal{E}_3 \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) (f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) \\
 &\quad + \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X))^\top UI_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} U^\top \mathcal{E}_2 + \text{higher order terms}.
 \end{aligned}$$

Note that

$$\mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X))^\top UI_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} U^\top \mathcal{E}_2 = \mathbf{T}_{\iota(x_k)} \mathcal{E}_2. \quad (\text{G.49})$$

When $x \in M_\epsilon$

$$\mathbf{T}_{\iota(x_k)} \mathcal{E}_2 = [\mathbb{O}(\epsilon^{-1}), \mathbb{O}(\epsilon^{-1})] \cdot \left[O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-2}}\right), O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-3}}\right) \right] = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-1}}\right).$$

When $x \notin M_\epsilon$

$$\mathbf{T}_{\iota(x_k)} \mathcal{E}_2 = [\mathbb{O}(1), \mathbb{O}(\epsilon^{-1})] \cdot \left[O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-2}}\right), O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-3}}\right) \right] = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-2}}\right).$$

Moreover, when $x_k \in M_\epsilon$ or $x_k \in M \setminus M_\epsilon$ by a similar calculation as in Lemma 32, $UI_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} U^\top \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) (f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) = [\mathbb{O}(1), \mathbb{O}(1)]$. Hence,

$$\mathcal{E}_1^\top UI_{p,r} \left(\frac{\Lambda}{\epsilon^d} + \epsilon^3 I_{p \times p} \right)^{-1} U^\top \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) (f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-1}}\right).$$

Next, we calculate $\mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X))^\top \mathcal{E}_3 \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) (f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X))$. By a straightforward calculation, we can show that it is dominated by

$$O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2+2}}\right) \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X))^\top \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) (f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)).$$

Hence, when $x_k \in M_\epsilon$,

$$\mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) \mathcal{E}_3 \mathbb{E} \frac{1}{\epsilon^d} (\iota(X) - \iota(x_k)) (f(X) - f(x_k)) \chi_{B_\epsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\epsilon^{d/2-1}}\right).$$

When $x_k \notin M_\varepsilon$,

$$\mathbb{E} \frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) \mathcal{E}_3 \mathbb{E} \frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k)) (f(X) - f(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) = O\left(\frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2-2}}\right).$$

In conclusion for $k = 1, \dots, n$, we have

$$\begin{aligned} & \left[\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) \right]^\top U_n I_{p,r_n} \left(\frac{\Lambda_n}{n\varepsilon^d} + \varepsilon^3 I_{p \times p} \right)^{-1} U_n^\top \left[\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) (f(x_{k,j}) - f(x_k)) \right] \\ &= \mathbb{E} \frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(\iota(X))^\top \left[U I_{p,r} \left(\frac{\Lambda}{\varepsilon^d} + \varepsilon^3 I_{p \times p} \right)^{-1} U^\top \right] \\ &\quad \times \mathbb{E} \frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k)) (f(X) - f(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2-1}}\right). \end{aligned}$$

A similar argument shows that for $k = 1, \dots, n$,

$$\begin{aligned} & \left[\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) \right]^\top U_n I_{p,r_n} \left(\frac{\Lambda_n}{n\varepsilon^d} + \varepsilon^3 I_{p \times p} \right)^{-1} U_n^\top \left[\frac{1}{n\varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) \right] \\ &= \mathbb{E} \frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(\iota(X))^\top \left[U I_{p,r} \left(\frac{\Lambda}{\varepsilon^d} + \varepsilon^3 I_{p \times p} \right)^{-1} U^\top \right] \mathbb{E} \frac{1}{\varepsilon^d} (\iota(X) - \iota(x_k)) \chi_{B_\varepsilon^{\mathbb{R}^p}(x_k)}(\iota(X)) \\ &\quad + O\left(\frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2}}\right). \end{aligned} \tag{G.50}$$

By Theorem 15, g_1 has order $O(\varepsilon^2)$ and g_2 has order 1. Hence, we have

$$\sum_{j=1}^n [W - I_{n \times n}]_{kj} f(x_{k,j}) = \frac{g_1 + O\left(\frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2-1}}\right)}{g_2 + O\left(\frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2}}\right)} = Q_\varepsilon f(x_k) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2-1}}\right). \tag{G.51}$$

■

Appendix H. Detailed calculation of (4.19)

We show that the operator $\bar{Q}_\varepsilon f(x)$ may not approximate the Laplace-Beltrami operator in the interior of $\iota(M) = [-1, 1] \subset \mathbb{R}$. Let $x, y \in M$ such that $\iota(x) \in [-1 + \varepsilon, 1 - \varepsilon]$. Since M is 1-dimensional and is embedded in \mathbb{R} , the local covariance matrix C_x has rank 1. Suppose $P \in C^2(M)$ and $\varepsilon > 0$ is small enough. By (Wu and Wu, 2018, Lemma SI.6, Case 0) and the fact that $K(x, y) = 0$ when $|\iota(x) - \iota(y)| \geq \varepsilon$, we have

$$K(x, y) = \left(1 - \frac{P'(x)}{P(x)} (\iota(y) - \iota(x)) + O(\varepsilon^2) \right) \chi_{B_\varepsilon^{\mathbb{R}}(\iota(x))}. \tag{H.1}$$

Hence,

$$K(y, x) = \left(1 - \frac{P'(y)}{P(y)} (\iota(x) - \iota(y)) + O(\varepsilon^2) \right) \chi_{B_\varepsilon^{\mathbb{R}}(\iota(x))}. \tag{H.2}$$

Without loss of generality, consider $\iota(x) = 0$. Let $\gamma(t) : [-1, 1] \rightarrow M$ be the arclength parametrization of M with $\gamma(0) = x$. Suppose $f \in C^3(M)$. Then, by a straightforward expansion in the parametrization $\gamma(t)$, we have

$$\begin{aligned}\bar{Q}_\varepsilon f(x) &:= \frac{\mathbb{E}[K_\varepsilon(X, x)[f(X) - f(x)]]}{\mathbb{E} K_\varepsilon(X, x)} \\ &= \frac{\int_{-\varepsilon}^\varepsilon \left(1 + \frac{(P \circ \gamma)'(t)}{(P \circ \gamma)(t)} t + O(\varepsilon^2)\right) \left((f \circ \gamma)'(0)t + \frac{(f \circ \gamma)''(0)}{2} t^2 + O(t^3)\right) (P \circ \gamma)(t) dt}{\int_{-\varepsilon}^\varepsilon \left(1 + \frac{(P \circ \gamma)'(t)}{(P \circ \gamma)(t)} t + O(\varepsilon^2)\right) (P \circ \gamma)(t) dt} \quad (\text{H.3})\end{aligned}$$

The numerator of (H.3) can be expanded as

$$\begin{aligned}&\int_{-\varepsilon}^\varepsilon \left(1 + \frac{(P \circ \gamma)'(t)}{(P \circ \gamma)(t)} t + O(\varepsilon^2)\right) \left((f \circ \gamma)'(0)t + \frac{(f \circ \gamma)''(0)}{2} t^2 + O(t^3)\right) (P \circ \gamma)(t) dt \\ &= (f \circ \gamma)'(0) \int_{-\varepsilon}^\varepsilon (P \circ \gamma)(t) dt + (f \circ \gamma)'(0) \int_{-\varepsilon}^\varepsilon \frac{(P \circ \gamma)'(t)}{(P \circ \gamma)(t)} (P \circ \gamma)(t) t^2 dt \\ &\quad + \frac{(f \circ \gamma)''(0)}{2} \int_{-\varepsilon}^\varepsilon (P \circ \gamma)(t) t^2 dt + O(\varepsilon^4) \\ &= (f \circ \gamma)'(0) \int_{-\varepsilon}^\varepsilon (P \circ \gamma)(t) dt + (f \circ \gamma)'(0) \int_{-\varepsilon}^\varepsilon (P \circ \gamma)'(t) t^2 dt \\ &\quad + \frac{(f \circ \gamma)''(0)}{2} \int_{-\varepsilon}^\varepsilon (P \circ \gamma)(t) t^2 dt + O(\varepsilon^4) \\ &= (f \circ \gamma)'(0) \int_{-\varepsilon}^\varepsilon \left((P \circ \gamma)(0)t + (P \circ \gamma)'(0)t^2 + O(t^3)\right) dt + (f \circ \gamma)'(0) \int_{-\varepsilon}^\varepsilon \left((P \circ \gamma)'(0)t^2 + O(t^3)\right) dt \\ &\quad + \frac{(f \circ \gamma)''(0)}{2} \int_{-\varepsilon}^\varepsilon ((P \circ \gamma)(0)t^2 O(t^3)) dt + O(\varepsilon^4) \\ &= \frac{2}{3}(f \circ \gamma)'(0)(P \circ \gamma)'(0)\varepsilon^3 + \frac{2}{3}(f \circ \gamma)'(0)(P \circ \gamma)'(0)\varepsilon^3 + \frac{1}{3}(f \circ \gamma)''(0)(P \circ \gamma)(0)\varepsilon^3 + O(\varepsilon^4) \\ &= \left(\frac{4}{3}(f \circ \gamma)'(0)(P \circ \gamma)'(0) + \frac{1}{3}(f \circ \gamma)''(0)(P \circ \gamma)(0)\right)\varepsilon^3 + O(\varepsilon^4).\end{aligned}$$

The denominator of (H.3) can be expanded as

$$\int_{-\varepsilon}^\varepsilon \left(1 + \frac{(P \circ \gamma)'(t)}{(P \circ \gamma)(t)} t + O(\varepsilon^2)\right) (P \circ \gamma)(t) dt = 2(P \circ \gamma)(0)\varepsilon + O(\varepsilon^2).$$

Hence, we have

$$\begin{aligned}\bar{Q}_\varepsilon f(x) &= \frac{\left[\frac{4}{3}(f \circ \gamma)'(0)(P \circ \gamma)'(0) + \frac{1}{3}(f \circ \gamma)''(0)(P \circ \gamma)(0)\right]\varepsilon^3 + O(\varepsilon^4)}{2(P \circ \gamma)(0)\varepsilon + O(\varepsilon^2)} \quad (\text{H.4}) \\ &= \left(\frac{2}{3} \frac{(f \circ \gamma)'(0)(P \circ \gamma)'(0)}{P \circ \gamma}(0) + \frac{1}{6}(f \circ \gamma)''(0)\right)\varepsilon^2 + O(\varepsilon^3) \\ &= \left(\frac{1}{6}f''(x) + \frac{2}{3} \frac{f'(x)P'(x)}{P(x)}\right)\varepsilon^2 + O(\varepsilon^3)\end{aligned}$$

Recall that $Q_\varepsilon f(x) = \frac{1}{6}f''(x)\varepsilon^2 + O(\varepsilon^3)$. However, since the kernel is not symmetric, \bar{Q}_ε does not approximate the Laplace-Beltrami operator Δ . Moreover, when $f \in C^5(M)$, we conclude that

$$\bar{Q}_\varepsilon Q_\varepsilon f(x) = \frac{1}{9} \left(\frac{1}{4}f'''(x)P'(x) + \frac{f'''(x)P'(x)}{P(x)} \right) \varepsilon^4 + O(\varepsilon^5).$$

References

- F.V. Atkinson. Discrete and continuous boundary value problems. *Acad. Press., New York. MR*, 176141, 1964.
- Jonathan Bates. The embedding dimension of Laplacian eigenfunction maps. *Applied and Computational Harmonic Analysis*, 37(3):516–530, 2014.
- M. Belkin and P. Niyogi. Laplacian Eigenmaps for dimensionality reduction and data representation. *Neural. Comput.*, 15(6):1373–1396, 2003.
- M. Belkin and P. Niyogi. Convergence of Laplacian eigenmaps. In *Adv. Neur. In.: Proceedings of the 2006 Conference*, volume 19, page 129. The MIT Press, 2007.
- Pierre Bérard, Gérard Besson, and Sylvain Gallot. Embedding Riemannian manifolds by their heat kernel. *Geometric & Functional Analysis GAFA*, 4(4):373–398, 1994.
- T. Berry and T. Sauer. Density estimation on manifolds with boundary. *Computational Statistics and Data Analysis*, 107:1–17, 2017.
- Jeff Calder, Nicolas Garcia Trillos, and Marta Lewicka. Lipschitz regularity of graph Laplacians on random data clouds. *SIAM Journal on Mathematical Analysis*, 54(1):1169–1222, 2022.
- Sun-Yung A Chang, Lihe Wang, and Paul C Yang. A regularity theory of biharmonic maps. *Communications on Pure and Applied Mathematics*, 52(9):1113–1137, 1999.
- M.-Y. Cheng and H.-T. Wu. Local linear regression on manifolds and its geometric interpretation. *J. Am. Stat. Assoc.*, 108:1421–1434, 2013.
- Xiuyuan Cheng and Nan Wu. Eigen-convergence of Gaussian kernelized graph Laplacian by manifold heat interpolation. *Applied and Computational Harmonic Analysis*, 61:132–190, 2022.
- R. R. Coifman and S. Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal.*, 21(1):5–30, 2006.
- Fabrizio Cuccu and Giovanni Porru. Maximization of the first eigenvalue in problems involving the bi-Laplacian. *Nonlinear Analysis: Theory, Methods & Applications*, 71(12):e800–e809, 2009.
- Xiucui Ding and Hau-Tieng Wu. Impact of signal-to-noise ratio and bandwidth on graph Laplacian spectrum from high-dimensional noisy point cloud. *IEEE Transactions on Information Theory*, 2022.
- D. L. Donoho and C. Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *P. Natl. Acad. Sci. USA*, 100(10):5591–5596, 2003.

- David B Dunson and Nan Wu. Inferring manifolds from noisy data using Gaussian processes. *arXiv preprint arXiv:2110.07478*, 2021.
- David B Dunson, Hau-Tieng Wu, and Nan Wu. Spectral convergence of graph Laplacian and heat kernel reconstruction in L^∞ from random samples. *Applied and Computational Harmonic Analysis*, 55:282–336, 2021.
- C. L. Epstein and R. Mazzeo. *Degenerate diffusion operators arising in population biology*. Number 185. Princeton University Press, 2013.
- Tingran Gao. The diffusion geometry of fibre bundles: Horizontal diffusion maps. *Applied and Computational Harmonic Analysis*, 50:147–215, 2021.
- A. S. Georgiou, J. M. Bello-Rivas, C. W. Gear, H.-T. Wu, E. Chiavazzo, and I. G. Kevrekidis. An exploration algorithm for stochastic simulators driven by energy gradients. *Entropy*, 19(7):294, 2017.
- D. N. Kaslovsky and F. G. Meyer. Non-asymptotic analysis of tangent space perturbation. *Information and Inference: a Journal of the IMA*, 3(2):134–187, 2014.
- D. Kershaw. Some extensions of W. Gautschi’s inequalities for the gamma function. *Mathematics of Computation*, pages 607–611, 1983.
- John M Lee. *Smooth manifolds*. Springer, 2012.
- Chen-Yun Lin, Arin Minasian, Xin Jessica Qi, and Hau-Tieng Wu. Manifold learning via the principle bundle approach. *Frontiers in Applied Mathematics and Statistics*, 4:21, 2018.
- L. Lin, B. St. Thomas, H. Zhu, and D. B. Dunson. Extrinsic local regression on manifold-valued data. *Journal of the American Statistical Association*, 112(519):1261–1273, 2017.
- A. V. Little, M. Maggioni, and L. Rosasco. Multiscale geometric methods for data sets I: Multiscale SVD, noise and curvature. *Applied and Computational Harmonic Analysis*, 43(3):504–567, 2017.
- J. Malik, C. Shen, H.-T. Wu, and N. Wu. Connecting dots – from local covariance to empirical intrinsic geometry and locally linear embedding. *Pure and Applied Analysis*, 1(4):515 – 542, 2019.
- S. Mukherjee, Q. Wu, and D.-X. Zhou. Learning gradients on manifolds. *Bernoulli*, 16(1):181–207, 2010.
- B. Nadler. Finite sample approximation results for principal component analysis: A matrix perturbation approach. *The Annals of Statistics*, 36(6):2791–2817, 2008.
- M. A. Naïmark. *Linear differential operators*. F. Ungar Publishing Company, 1967.
- S. Osher, Z. Shi, and W. Zhu. Low dimensional manifold model for image processing. *SIAM Journal on Imaging Sciences*, 10(4):1669–1690, 2017.
- J Wilson Peoples and John Harlim. Spectral convergence of symmetrized graph Laplacian on manifolds with boundary. *arXiv preprint arXiv:2110.06988*, 2021.

- Jacobus W Portegies. Embeddings of Riemannian manifolds with heat kernels and eigenfunctions. *Communications on Pure and Applied Mathematics*, 69(3):478–518, 2016.
- S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000. doi: 10.1126/science.290.5500.2323.
- T. Shnitzer, M. Ben-Chen, L. Guibas, R. Talmon, and H.-T. Wu. Recovering hidden components in multimodal data with composite diffusion operators. *arXiv preprint arXiv:1808.07312*, 2018.
- A. Singer and H.-T. Wu. Vector diffusion maps and the connection Laplacian. *Communications on Pure and Applied Mathematics*, 65(8):1067–1144, 2012.
- A. Singer and H.-T. Wu. Spectral convergence of the connection Laplacian from random samples. *Information and Inference: A Journal of the IMA*, 6(1):58–123, 2017.
- J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- N. G. Trillos, M. Gerlach, M. Hein, and D. Slepcev. Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace–Beltrami operator. *Foundations of Computational Mathematics*, 20(4):827–887, 2020.
- H. Tyagi, E. Vural, and P. Frossard. Tangent space estimation for smooth embeddings of Riemannian manifolds. *Information and Inference: A Journal of the IMA*, 2(1):69–114, 2013.
- L. van der Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.
- R. Vaughn, T. Berry, and H. Antil. Diffusion maps for embedded manifolds with boundary with applications to pdes. *arXiv preprint arXiv:1912.01391*, 2019.
- Ryan Vaughn. *Diffusion maps for manifolds with boundary*. PhD thesis, George Mason University, 2020.
- U. Von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, pages 555–586, 2008.
- K.Q. Weinberger and L.K. Saul. An introduction to nonlinear dimensionality reduction by maximum variance unfolding. *Aaaai*, pages 1683–1686, 2006.
- Hassler Whitney. Analytic extensions of differentiable functions defined in closed sets. In *Hassler Whitney Collected Papers*, pages 228–254. Springer, 1992.
- J. Wong. An extension procedure for manifolds with boundary. *Pacific Journal of Mathematics*, 235(1):173–199, 2008.
- Caroline L Wormell and Sebastian Reich. Spectral convergence of diffusion maps: Improved error bounds and an alternative normalization. *SIAM Journal on Numerical Analysis*, 59(3):1687–1734, 2021.

H.-T. Wu and N Wu. Think globally, fit locally under the manifold setup: Asymptotic analysis of locally linear embedding. *Annals of Statistics*, 46(6B):3805–3837, 2018.

Hau-Tieng Wu and Nan Wu. Strong uniform consistency with rates for kernel density estimators with general kernels on manifolds. *Information and Inference: A Journal of the IMA*, 11(2):781–799, 2022.

Y. Yang and D. B. Dunson. Bayesian manifold regression. *The Annals of Statistics*, 44(2):876–905, 2016.