PURE and APPLIED ANALYSIS
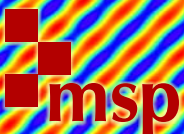
John Malik, Chao Shen, Hau-Tieng Wu and Nan Wu

CONNECTING DOTS: FROM LOCAL COVARIANCE TO
EMPIRICAL INTRINSIC GEOMETRY AND LOCALLY LINEAR
EMBEDDING

CONNECTING DOTS: FROM LOCAL COVARIANCE TO EMPIRICAL INTRINSIC GEOMETRY AND LOCALLY LINEAR EMBEDDING

John Malik, Chao Shen, Hau-Tieng Wu and Nan Wu

Local covariance structure under the manifold setup has been widely applied in the machine-learning community. Based on the established theoretical results, we provide an extensive study of two relevant manifold learning algorithms, empirical intrinsic geometry (EIG) and locally linear embedding (LLE) under the manifold setup. Particularly, we show that without an accurate dimension estimation, the geodesic distance estimation by EIG might be corrupted. Furthermore, we show that by taking the local covariance matrix into account, we can more accurately estimate the local geodesic distance. When understanding LLE based on the local covariance structure, its intimate relationship with the curvature suggests a variation of LLE depending on the "truncation scheme". We provide a theoretical analysis of the variation.

1. Introduction

Covariance is arguably one of the most important quantities in data analysis. It has been widely studied in the past century and is still an active research topic nowadays. In this paper, we focus on the *local covariance structure* under the manifold setup, which has been widely applied, explicitly or implicitly, to various applications in different fields; see, for example, a far-from-complete list [Kambhatla and Leen 1997; Roweis and Saul 2000; Donoho and Grimes 2003; Brand 2003; Zhang and Zha 2004; Kushnir et al. 2006; Goldberg et al. 2009; Salhov et al. 2012; Gong et al. 2012; Singer and Wu 2012; Pedagadi et al. 2013; Little et al. 2017; Arias-Castro et al. 2017]. In the past few years, its mathematical and statistical properties have been well-established [Singer and Wu 2012; Cheng and Wu 2013; Bernstein and Kuleshov 2014; Kaslovsky and Meyer 2014; Tyagi et al. 2013; Wu and Wu 2018] for different purposes. In this paper, based on the established theoretical foundation, we extensively discuss two topics in the manifold-learning community that are related to the local covariance structure — empirical intrinsic geometry (EIG) and locally linear embedding (LLE).

EIG [Talmon and Coifman 2012; 2013], or originally called nonlinear independent component analysis [Singer and Coifman 2008], is a technique aiming to deal with the distortion underlying the collected dataset that is caused by the observation process. In many applications, the manifold structure of interest can only be accessed via an observation and not directly. However, the observation process might nonlinearly deform the manifold of interest. As a result, the information inferred from the observed data point cloud might not faithfully reflect the intrinsic properties. The goal of EIG is correcting

this deformation by taking the local covariance matrix into account. From the statistical viewpoint, it is a nonlinear *latent space model*, and the local covariance structure leads to a generalization of the Mahalanobis distance. While it has been successfully applied to different problems [Wu et al. 2015; Mishne et al. 2015; Yair and Talmon 2017; Shemesh et al. 2017; Liu et al. 2018], to the best of our knowledge, besides an argument on the Euclidean space setup [Singer and Coifman 2008], a systematic evaluation of how the algorithm works under the manifold setup, and its sensitivity to the parameter choice, is missing. Due to its importance, the first contribution of this paper is providing a quantification of EIG under the manifold setup, and discussing how the chosen parameter influences the final result. In the special case that there is no deformation (that is, we can access the manifold directly), we show a more accurate geodesic distance estimator, called the *covariance-corrected geodesic distance estimator*, by correcting the Euclidean distance when the manifold is embedded in the Euclidean space.

LLE [Roweis and Saul 2000] is a widely applied nonlinear dimension-reduction technique in the manifold-learning community. Despite its wide application, its theoretical properties were studied only recently. See [Wu and Wu 2018] as an example. Based on the analysis, several peculiar behaviors of LLE have been better understood. While LLE depends on the barycentric coordinate to determine the affinity between pairs of points, it has a natural relationship with the local covariance matrix; the kernel associated with LLE is not symmetric, which is different from the kernel commonly used in graph Laplacian-based algorithms like Laplacian eigenmaps [Belkin and Niyogi 2003] or diffusion maps [Coifman and Lafon 2006]. The regularization plays an essential role in the algorithm. Different regularizations lead to different embedding results. Based on the intimate relationship between the curvature and regularization, the second contribution of this paper is studying a variation of LLE by directly truncating the local covariance matrix.

The paper is organized in the following way. In Section 2, we introduce the notation for the local covariance structure analysis and some relevant known results. In Section 3, we provide a theoretical argument of EIG under the manifold setup; when the observation process is trivial, we analyze the covariance-corrected geodesic distance estimator. In Section 4, we discuss the relationship of EIG and LLE, and provide a variation of LLE. Numerical results are shown in each section to support the theoretical findings. In Section 5, we provide some numerical results. In Section 6, discussion and conclusion are provided.

1.1. *Notation and mathematical setup.* Let X be a p-dimensional random vector with the range supported on a d-dimensional, compact, smooth Riemannian manifold (M, g) isometrically embedded in $\mathbb{R}^p$ via $\iota : M \hookrightarrow \mathbb{R}^p$. We assume that M is boundary-free in this work. Let $\exp_x : T_x M \to M$ be the exponential map at x. Unless otherwise specified, we will carry out calculations using normal coordinates. Let $T_x M$ denote the tangent space at $x \in M$, and let $\iota_* T_x M$ denote the embedded tangent space in $\mathbb{R}^p$. Write the normal space at $y = \iota(x)$ as $(\iota_* T_x M)^\perp$. Let II_x be the second fundamental form of ι at x. Let P be the probability density function (p.d.f.) associated with the random vector X [Cheng and Wu 2013, Section 4]. We assume that $P \in \mathcal{C}^5(\iota(M))$ and that there exist $0 < P_m \leq P_M$ such that $P_m \leq P(y) \leq P_M < \infty$ for all $y \in \iota(M)$. Let $e_i \in \mathbb{R}^p$ be the unit p-dimensional unitary vector with a 1 in the i-th entry. Let $B_h^{\mathbb{R}^p}(z)$ denote the p-dimensional Euclidean ball of radius $h > 0$ with center $z \in \mathbb{R}^p$, and let $\chi_A : \mathbb{R}^p \to \{0, 1\}$ denote the indicator function of the set $A \subset \mathbb{R}^p$. Define $\mathbb{O}(p)$ to be

the orthogonal group in dimension $p \in \mathbb{N}$. We define two main quantities, the *truncated inverse* and *regularized inverse*, of a symmetric matrix that are related to EIG and LLE respectively.

Definition 1. Let $A \in \mathbb{R}^{p \times p}$ be a real symmetric matrix and set $r = \text{rank}(A)$. Let $\lambda_1 \geq \cdots \geq \lambda_p$ be the eigenvalues of A, and let $u_1, \ldots, u_p$ be the corresponding normalized eigenvectors. For $0 < \alpha \leq r$, the α-truncated inverse of A is defined as

$$\mathcal{T}_\alpha[A] = \begin{bmatrix} u_1 & \cdots & u_\alpha \end{bmatrix} \begin{bmatrix} \lambda_1^{-1} & & \\ & \ddots & \\ & & \lambda_\alpha^{-1} \end{bmatrix} \begin{bmatrix} u_1^\top \\ \vdots \\ u_\alpha^\top \end{bmatrix}. \tag{1}$$

Choose a regularization constant $c > 0$. The c-regularized inverse of A is defined as

$$\mathcal{I}_c[A] = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} \begin{bmatrix} (\lambda_1 + c)^{-1} & & \\ & \ddots & \\ & & (\lambda_r + c)^{-1} \end{bmatrix} \begin{bmatrix} u_1^\top \\ \vdots \\ u_r^\top \end{bmatrix}. \tag{2}$$

Note that if $\alpha = r$, then $\mathcal{T}_\alpha[A]$ is the Penrose–Moore pseudoinverse. When $c \to 0$, $\mathcal{I}_c[A]$ becomes $\mathcal{T}_r[A]$.

2. Local covariance matrix and some facts

We start with the definition of the local covariance matrix.

Definition 2. For $x \in M$ and a measurable set $\mathcal{O} \subset \iota(M)$, the *local covariance matrix at* $\iota(x) \in \iota(M)$ *associated with* $\mathcal{O}$ is defined as

$$C_{x,\mathcal{O}} := \mathbb{E}[(X - \iota(x))(X - \iota(x))^\top \chi_\mathcal{O}(X)] \in \mathbb{R}^{p \times p}. \tag{3}$$

When $\mathcal{O}$ is $B_h^{\mathbb{R}^p}(\iota(x)) \cap \iota(M)$, where $h > 0$, we define

$$C_h(x) := C_{x, B_h^{\mathbb{R}^p}(\iota(x)) \cap \iota(M)}$$

and simply call $C_h(x)$ the *local covariance matrix at* $\iota(x)$.

The local covariance matrix and its relationship with the embedded tangent space of the manifold have been widely studied recently, including (but not exclusively) [Singer and Wu 2012; Cheng and Wu 2013; Tyagi et al. 2013; Bernstein and Kuleshov 2014; Kaslovsky and Meyer 2014; Little et al. 2017]. Recently, in order to systematically study the LLE algorithm, the higher-order structure of the local covariance matrix was explored in [Wu and Wu 2018]. We now summarize the result for our purpose. Since the local covariance matrix is invariant up to translation and rotation and the analysis is local at one point, to simplify the discussion, from now on, when we analyze the local covariance matrix at x, we always assume that the manifold is translated and rotated in $\mathbb{R}^p$ so that $\iota_* T_x M$ is spanned by $e_1, \ldots, e_d \in \mathbb{R}^p$.

Lemma 3 [Wu and Wu 2018, Proposition 3.2]. *Fix $x \in M$ and take $h > 0$. Write the eigendecomposition of $C_h(x)$ as $U_h(x) \Lambda_h(x) U_h(x)^\top$. Then, when h is sufficiently small, the diagonal matrix $\Lambda_h(x) \in \mathbb{R}^{p \times p}$ satisfies*

$$\Lambda_h(x) = \frac{|S^{d-1}| P(x) h^{d+2}}{d(d+2)} \begin{bmatrix} I_{d \times d} + O(h^2) & 0 \\ 0 & O(h^2) \end{bmatrix}, \tag{4}$$

and the orthogonal matrix $U_h(x) \in \mathbb{O}(p)$ satisfies

$$U_h(x) = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} (I_{p\times p} + h^2 \mathsf{S}) + O(h^4), \tag{5}$$

where $U_1 \in \mathbb{O}(d)$, $U_2 \in \mathbb{O}(p-d)$, and S is an antisymmetric matrix.

First, note that the local covariance matrix depends on the p.d.f. Particularly, when the sampling is nonuniform, the eigenvalues are deviated and the p.d.f. is responsible for this deviation. Equation (4) in this lemma says that the first d eigenvalues of $C_h(x)$ are of order h^{d+2}, while the rest of the eigenvalues are two orders higher, that is, of order $O(h^{d+4})$. Moreover, note that $I_{p\times p} + h^2 \mathsf{S}$ approximates a rotation, so (5) says that the first d normalized eigenvectors of $C_h(x)$ deviate from an orthonormal basis of $\iota_* T_x M$ by a rotational error of order $O(h^2)$, and the rest of the orthonormal eigenvectors of $C_h(x)$ deviate from an orthonormal basis of $(\iota_* T_x M)^\perp$ within a rotational error of order $O(h^2)$.

In practice, we are given a finite sampling of points from the embedded manifold and need to approximate the local covariance matrix. Since the finite convergence argument of the sample local covariance matrix to the local covariance matrix is standard (see, for example, [Wu and Wu 2018, Propositions 3.1 and 3.2 and Lemma E.4]), to focus on the main idea and simplify the discussion, below we work directly on the continuous setup; that is, we consider only the asymptotical case when $n \to \infty$.

It is well known that when a manifold M is isometrically embedded into another manifold M', then for two close points $x, y \in M$, the geodesic distance between x, y in M could be well-approximated by the geodesic distance between x, y in M', with the error depending on the second fundamental form of the embedding; see, for example [Smolyanov et al. 2007, Proposition 6]. We have the following lemma when M' is Euclidean space.

Lemma 4 [Wu and Wu 2018, Lemma B.2]. *Suppose that M is isometrically embedded in $\mathbb{R}^p$ through ι. Fix $x \in M$ and use polar coordinates $(t, \theta) \in [0, \infty) \times S^{d-1}$ to parametrize $T_x M$. For $y = \exp_x(\theta t)$ for sufficiently small t, we have*

$$\iota(y) - \iota(x) = (\iota_* \theta)t + \frac{\mathrm{II}_x(\theta, \theta)}{2}t^2 + \frac{\nabla_\theta \mathrm{II}_x(\theta, \theta)}{6}t^3 + O(t^4). \tag{6}$$

Moreover, when $h := \|\iota(y) - \iota(x)\|_{\mathbb{R}^p}$ is sufficiently small, we have

$$h = t - \frac{\|\mathrm{II}_x(\theta, \theta)\|^2}{24}t^3 - \frac{\nabla_\theta \mathrm{II}_x(\theta, \theta) \cdot \mathrm{II}_x(\theta, \theta)}{24}t^4 + O(t^5), \tag{7}$$

and hence

$$t = h + \frac{\|\mathrm{II}_x(\theta, \theta)\|^2}{24}h^3 + \frac{\nabla_\theta \mathrm{II}_x(\theta, \theta) \cdot \mathrm{II}_x(\theta, \theta)}{24}h^4 + O(h^5). \tag{8}$$

The proof of the lemma can be found in, for example, [Smolyanov et al. 2007, Proposition 6] when M' is a generic manifold or [Wu and Wu 2018, Lemma B.2] when M' is Euclidean space.
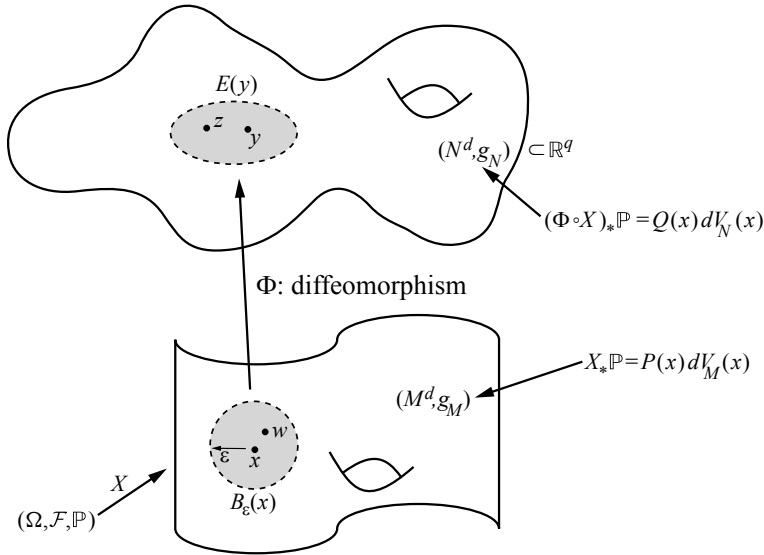
Figure 1. The illustration of the theoretical framework of EIG.

3. Empirical intrinsic geometry and local covariance matrix

EIG [Talmon and Coifman 2012; Singer and Coifman 2008] is a technique aiming to deal with the underlying distortion caused by the observation process. The basic idea of the technique is that the local covariance matrix captures the distortion, under suitable assumptions, and we can correct the distortion by manipulating the local covariance matrix and recover the local geodesic distance. From the statistical viewpoint, it is a generalization of the Mahalanobis distance. We now examine the intimate relationship between EIG and the covariance-corrected geodesic estimator.

Suppose that M is a d-dimensional closed Riemannian manifold that hosts the information in which we are interested but we cannot directly access. We assume that there is a method to *indirectly* access M via an *observation*, and hence collect a dataset that contains *indirect* information of M in which we are interested. We model the observation as a nonlinear function $\Phi : M \to N$, where Φ is a diffeomorphism of M and N is isometrically embedded in $\mathbb{R}^q$ via ι. Under this setup, for the point $x \in M$ that we cannot access, there is a corresponding point $\iota(y) \in \iota(N) \subset \mathbb{R}^q$ that we can access and collect as the dataset, where $y = \Phi(x)$. The mission is estimating the geodesic distance between two close points $x \in M$ and $w \in M$ through accessible data points $\iota(y) = \iota(\Phi(x))$ and $\iota(z) = \iota(\Phi(w))$. This situation is commonly encountered in data analysis; for example, the brain activity information contained in an electroencephalogram recorded from the scalp might be deformed due to the process of recording the electroencephalogram, the anatomical structure and physiological properties. Clearly, due to the diffeomorphism associated with the observation, the pairwise distance estimated from the collected database no longer faithfully reflects the pairwise distance of the inaccessible space. Hence, analysis tools depending on the pairwise distance are biased. The interest of EIG is correcting this bias.

The main idea of EIG can be summarized in the following way. For each hidden point $x \in M$, take the geodesic ball $B_\varepsilon(x) \subset M$ with the radius ε and centered at x. As is discussed in [Singer and Coifman 2008], *if we can determine* $E(y) := \Phi(B_\varepsilon(x)) \subset N$ around y that is associated with $B_\varepsilon(x)$, then up to a constant, the geodesic distance between two close points w and x can be well-approximated by $\iota(y)$, $\iota(z)$, and the local covariance matrix $C_{y,\iota(E(y))}$ associated with $\iota(E(y))$ as defined in (3). To simplify the notation, set $\mathcal{C}_\varepsilon(y) := C_{y,\iota(E(y))}$. See Figure 1 for an illustration of the setup. Note that when M is Euclidean and Φ is linear, $E(y)$ is an ellipsoid. In the following, although we consider the manifold model that is in general not Euclidean, we abuse the terminology and call $E(y)$ an ellipsoid. Numerically, $\mathcal{C}_\varepsilon(y)$ is estimated by taking points in $E(y)$ into account. Again, since the convergence proof is standard, to simplify the discussion, we skip the finite convergence step and focus on the continuous setup.

This main idea is carried out by noting that the local covariance matrix at y associated with $\mathcal{C}_\varepsilon(y)$ captures the Jacobian of the diffeomorphism associated with the observation; that is, $\mathcal{C}_\varepsilon(y) \approx \nabla\Phi|_x \Phi|_x^\top$ [Singer and Coifman 2008]. Under the assumption that the ellipsoid $E(y)$ is known, $\mathcal{C}_\varepsilon(y) \approx \nabla\Phi|_x \Phi|_x^\top$ is true and the dimension of the manifold d is known, authors in [Singer and Coifman 2008; Talmon and Coifman 2012] consider the following quantity, called the *EIG distance*, to estimate the geodesic distance between x and w:

$$(\iota(z) - \iota(y))^\top \left(\frac{\mathcal{T}_\alpha[\mathcal{C}_\varepsilon(y)] + \mathcal{T}_\alpha[\mathcal{C}_\varepsilon(z)]}{2} \right)(\iota(z) - \iota(y)), \tag{9}$$

where α is chosen to be the dimension of d.

3.1. *Challenges of EIG and analysis.* There are two challenges of applying the EIG idea: how to determine the ellipsoid $E(y)$ associated with $B_\varepsilon(x)$, and how to estimate the dimension d of the intrinsic manifold. In [Singer and Coifman 2008] and most of its citations, the state-space model combined with the stochastic differential equation (SDE) is considered, and the ellipsoid can be determined simply by taking the temporal relationship into account under this model. We refer readers with interest to [Singer and Coifman 2008; Talmon and Coifman 2012] for more details about this state-space model. If the state-space model and SDE cannot be directly applied, this task is most of time a big challenge and to the best of our knowledge not too much is known. This challenge is however out of the scope of this paper. On the other hand, most of time we do not have an access to the dimension of d, and a dimension estimation is needed. Although theoretically we can count on the spectral gap to estimate d, in practice it depends on the try-and-error process and little can be guaranteed. Particularly, when the manifold is highly distorted by the observation, faithfully estimating the dimension is another challenging task. To the best of our knowledge, although EIG has obtained several successes in different applications [Wu et al. 2015; Mishne et al. 2015; Yair and Talmon 2017; Shemesh et al. 2017], a systematic exploration of the approximation under the general manifold setup is lacking. It is also not clear what may happen when the dimension is not estimated correctly.

Below, we assume that we have the knowledge of the ellipsoid associated with $B_\varepsilon(x)$ for all x, and hence we can evaluate the local covariance matrix at $y = \Phi(x)$ associated with $E(y)$. We show how $\mathcal{C}_\varepsilon(y) \approx \nabla\Phi|_x \Phi|_x^\top$ holds under the manifold setup, and the influence of an erroneously estimated dimension. We start with the following definitions.

Definition 5. *For $y \in N$, define the normalized local covariance matrix at y associated with $E(y) \subset N$ as*

$$\bar{\mathcal{C}}_\varepsilon(y) := \frac{\mathcal{C}_\varepsilon(y)}{\varepsilon^2 \mathbb{E}[\chi_{E(y)}(Y)]}, \tag{10}$$

where $Y := \iota \circ \Phi \circ X$ is the induced random variable.

The normalization step in (10) is introduced to remove the impact of the nonuniform sampling. As we will see, without this normalization, when the sampling is nonuniform, the p.d.f. will play a role in the final analysis.

Lemma 6. *Let $x \in M$ and $B_\varepsilon(x) \subset M$ be the geodesic ball around x. The local covariance matrix at $y = \Phi(x)$ associated with the ellipsoid $E(y) = \Phi(B_\varepsilon(x)) \subset N$ satisfies*

$$\mathcal{C}_\varepsilon(y) = \frac{|S^{d-1}| P(x) \varepsilon^{d+2}}{d(d+2)} [\iota_*|_y \nabla\Phi(x)][\nabla\Phi(x)^\top \iota_*|_y^\top] + O(\varepsilon^{d+4}) \tag{11}$$

and the normalized local covariance matrix at y associated with $E(y)$ satisfies

$$\bar{\mathcal{C}}_\varepsilon(y) = \frac{1}{d+2} [\iota_*|_y \nabla\Phi(x)][\nabla\Phi(x)^\top \iota_*|_y^\top] + O(\varepsilon^2). \tag{12}$$

If $v_1, v_2 \in (\iota_* T_y N)^\perp$, then

$$v_1^\top \bar{\mathcal{C}}_\varepsilon(y) v_2 = \frac{d\varepsilon^2}{4|S^{d-1}|(d+4)} \int_{S^{d-1}} v_1^\top \mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x))(\mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x)))^\top v_2 \, d\theta + O(\varepsilon^4).$$

By this lemma, we see that the expansion of $\mathcal{C}_\varepsilon(y)$ depends not only on the p.d.f. but also on the Jacobian of the deformation, and the normalization step cancels this dependence. The proof is postponed until Appendix, page 531. Based on this lemma and the considered setup in [Singer and Coifman 2008; Talmon and Coifman 2012], we consider the following quantity.

Definition 7. *For $y, z \in N$, define the EIG distance of order α between y, z, where $1 \leq \alpha \leq q$, as*

$$\mathrm{EIG}_\alpha^2(y, z) = (\iota(y) - \iota(z))^\top \left(\frac{\mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(y)] + \mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(z)]}{2} \right) (\iota(y) - \iota(z)). \tag{13}$$

Note that this definition is slightly different from that considered in (9). Since the only difference is the normalization step, the result below can be directly translated for (9). Below we show that the EIG distance between $y = \Phi(x) \in N$ and $z = \Phi(w) \in N$ is a good estimator of the geodesic distance between $x \in M$ and $w \in M$ only when some conditions are satisfied.

Theorem 8. *Suppose two d-dimensional smooth closed Riemannian manifolds M and N are diffeomorphic via $\Phi : M \to N$, and suppose N is isometrically embedded in $\mathbb{R}^q$ via ι. Take $x \in M$ and $w \in B_\varepsilon(x) \subset M$, where $\varepsilon > 0$, and denote by t the geodesic distance between x and w. Let $y = \Phi(x)$, $z = \Phi(w)$. Take $\alpha \leq \min(\mathrm{rank}[\bar{\mathcal{C}}_\varepsilon(y)], \mathrm{rank}[\bar{\mathcal{C}}_\varepsilon(z)])$. When ε is sufficiently small, we have the following:*

(1) *When $\alpha = d$, we have*

$$\mathrm{EIG}_d(y, z) = \sqrt{d+2} t + O(t\varepsilon^2) + O(t^3).$$

(2) *When $1 \leq \alpha < d$, we have*

$$\text{EIG}_\alpha(y, z) = \sqrt{d + 2}\, t + O(t). \tag{14}$$

Let $V_\alpha(y)$ and $V_\alpha(z)$ be the subspaces of $\iota_* T_y N$ and $\iota_* T_z N$ generated by the first α eigenvectors of $\bar{\mathcal{C}}_\varepsilon(y)$ and $\bar{\mathcal{C}}_\varepsilon(z)$ respectively. Suppose $z = \exp_y \vartheta(y)$ for $\vartheta(y) \in T_y N$ and $y = \exp_z \vartheta(z)$ for $\vartheta(z) \in T_z N$. If $\iota_*|_y \vartheta(y) \in V_\alpha(y)$ and $\iota_*|_z \vartheta(z) \in V_\alpha(z)$, we have

$$\text{EIG}_\alpha(y, z) = \sqrt{d + 2}\, t + O(t\varepsilon^2) + O(t^3). \tag{15}$$

(3) *When $\alpha > d$, assume that the smallest nonzero eigenvalues of $\bar{\mathcal{C}}_\varepsilon(y)$ and $\bar{\mathcal{C}}_\varepsilon(z)$ are of order ε^4 and there are $l_y \geq 0$ and $l_z \geq 0$ such eigenvalues respectively. In general, we have*

$$\text{EIG}_\alpha(y, z) = \sqrt{d + 2}\, t + O(t). \tag{16}$$

When $l_y = 0$ or $\alpha \leq q - l_y$ hold, and $l_z = 0$ or $\alpha \leq q - l_z$ hold, for $t = \varepsilon^\beta$, where $\beta > 1$, we have

$$\begin{aligned}
\text{EIG}_\alpha(y, z) &= \sqrt{d + 2}\, t + O(t(t/\varepsilon + \varepsilon)^2) \\
&= \sqrt{d + 2}\, t + O(t^{1 + \min\{1 - 1/\beta, 1/\beta\}}).
\end{aligned} \tag{17}$$

The proof is postponed until Appendix A. We now have discussion of the theorem. In general, $\text{rank}[\bar{\mathcal{C}}_\varepsilon(y)]$ and $\text{rank}[\bar{\mathcal{C}}_\varepsilon(z)]$ are different; thus we need the condition $\alpha \leq \min(\text{rank}[\bar{\mathcal{C}}_\varepsilon(y)], \text{rank}[\bar{\mathcal{C}}_\varepsilon(z)])$. First of all, if we know the dimension of the manifold, then the EIG distance of order d between $y = \Phi(x) \in N$ and $z = \Phi(w) \in N$ is an accurate estimator of the geodesic distance between $x \in M$ and $w \in M$, up to a global constant $\sqrt{d + 2}$. This result coincides with the claim in [Singer and Coifman 2008; Talmon and Coifman 2012] when the sampling on N is uniform. However, when the sampling on N is nonuniform, the result is deviated by the p.d.f. if we replace the normalized local covariance matrix $\bar{\mathcal{C}}_\varepsilon(y)$ in (13) by the local covariance matrix $\mathcal{C}_\varepsilon(y)$. This deviation can be seen by comparing (11) and (12) in Lemma 6.

Second, if the dimension is unknown and wrongly estimated, the result depends on the situation. When the dimension is underestimated, in general the estimator is wrong. In a nongeneric situation where the geodesic direction from y to z and that from z to y are both located on the first α eigenvectors of the associated local covariant matrices, we may still obtain an accurate estimator. Note that due to the curvature, there is no guarantee that the first α eigenvectors of $\bar{\mathcal{C}}_\varepsilon(y)$ will contain $\vartheta(y)$; even if they do, there is no guarantee that the first α eigenvectors of $\bar{\mathcal{C}}_\varepsilon(z)$ will contain $\vartheta(z)$. When the dimension is overestimated and some assumptions are satisfied, we still can obtain a reasonably good estimate of the intrinsic geodesic distance when t/ε is sufficiently small, but with a slow convergence rate, as is shown in (17).

Third, it is important to note that this estimate for the geodesic distance is valid only for points that are "not too far away". The estimate might be degenerate if two points are far away. For example, if $y \neq z \in N$ are two points such that the vector $\iota(z) - \iota(y)$ is orthogonal to the column space of $\mathcal{T}_\alpha[\mathcal{C}_\varepsilon(y)] + \mathcal{T}_\alpha[\mathcal{C}_\varepsilon(z)]$, then the quantity (9) is zero, and hence the degeneracy. This obviously destroys the topology of the manifold in which we are interested. A concrete example of this happening is when y and z are conjugate

points on the sphere $N = S^d$. In practice, we thus should only evaluate EIG when two points are "close" in N.

The assumption "the smallest nonzero eigenvalues of $\bar{\mathcal{C}}_\varepsilon(y)$ and $\bar{\mathcal{C}}_\varepsilon(z)$ are of order ε^4" needs some discussion. The rule of thumb is that eigenvalues of $\bar{\mathcal{C}}_\varepsilon(y)$ associated with the tangent directions are of order 1, but the eigenvalues associated with the normal directions are of order ε^2 or higher. The assumption that the smallest eigenvalues are of order ε^4 means that when the principle curvature is zero, the higher-order curvature is not too small. As is shown in the proof of this theorem, under this assumption, we can control the error induced by the interaction between $\iota(y) - \iota(z)$ and $\mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(y)]$. However, when there is an eigenvalue of even higher order, the interaction between $\iota(y) - \iota(z)$ and $\mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(y)]$ becomes more complicated, and a more delicate analysis involving a systematic recursive formula of Taylor expansion of $\iota(y) - \iota(z)$ and $\mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(y)]$ is needed. We will explore this issue in our future work.

To sum up, in practice if we are not confident about the estimated dimension, we may want to choose a larger α. Although we do not discuss it in this paper, we mention that this conservative approach might not be preferred if noise exists, since the noise will contaminate the eigenvector associated with the small eigenvalues. A more statistical approach to solve these issues will be studied in our future work.

3.2. A special EIG setup: covariance-corrected geodesic estimator. When Φ is the identity, EIG is reduced to the ordinary manifold learning setup; that is, we have samples from the manifold and we want to estimate its geometric structure. By Lemma 4, when two points on the smooth manifold are close enough, we can accurately estimate their geodesic distance t by the ambient Euclidean distance $h = \|\iota(y) - \iota(x)\|_{\mathbb{R}^p}$ up to the *third* order $O(t^3)$, and the third-order term is essentially the second fundamental form. We call h the *Euclidean-distance-based geodesic distance estimator*. In general, it is not an easy task to directly estimate the second fundamental form from the point cloud if M' is not Euclidean space. However, when M' is Euclidean, like the setup in Lemma 4, the second fundamental form information could be well-approximated by the local covariance matrix. We define the following projection operator associated with the local covariance matrix.

Definition 9. Suppose M is a closed, smooth d-dimensional Riemannian manifold isometrically embedded in $\mathbb{R}^p$ through ι. Fix $x \in M$. For $\bar{h} > 0$, let $C_{\bar{h}}(x) := C_{x, B_{\bar{h}}^{\mathbb{R}^p}(\iota(x)) \cap \iota(M)}$ as defined in (3). Let $\lambda_1 \geq \cdots \geq \lambda_p$ be the eigenvalues of $C_{\bar{h}}(x)$, and $u_1, \ldots, u_p$ be the corresponding normalized eigenvectors. Define $P_{\bar{h}}^\perp$ to be the orthogonal projection from $\mathbb{R}^p$ to the subspace spanned by $u_{d+1}, \ldots, u_p$.

The name orthogonal projection follows from Lemma 3, where we prove that $u_{d+1}, \ldots, u_p$ deviate from an orthonormal basis of $(\iota_* T_x M)^\perp$ within a rotational error of order $O(\bar{h}^2)$. Inspired by Lemmas 3 and 4, we have the following theorem. The proof is postponed until Appendix B.

Theorem 10. *Fix $x \in M$. Suppose that $y \in M$ and $\iota(y) \in B_{\bar{h}}^{\mathbb{R}^p}(\iota(x))$. We use the polar coordinate $(t, \theta) \in [0, \infty) \times S^{d-1}$ to parametrize $T_x M$ so that $y = \exp_x(\theta t)$ for $t > 0$. Define $h := \|\iota(y) - \iota(x)\|_{\mathbb{R}^p} \leq \bar{h}$. When $\bar{h}$ is small enough,*

$$\left| t - \left(h + \frac{\|P_{\bar{h}}^\perp(\iota(y) - \iota(x))\|_{\mathbb{R}^p}^2}{6h} \right) \right| = O(h^2 \bar{h}^2). \tag{18}$$

Based on this theorem, we have the following definition.

Definition 11. Following the notation in Theorem 10, the *covariance-corrected geodesic distance estima-tor* between points $x, y \in M$ is defined as

$$h + \frac{\| P_{\vec{h}}^{\perp} (\iota(y) - \iota(x)) \|_{\mathbb{R}^p}^2}{6h}. \tag{19}$$

Clearly, the covariance-corrected geodesic distance estimator accurately estimates the geodesic distance t up to the *fourth* order. It is worth mentioning that a better estimation for the geodesic distance cannot be directly achieved with only the local covariance matrix. Indeed, to estimate $d(x, y)$ within an error of $O(h^5)$, by Lemma 4, we need to estimate $\nabla_\theta \mathrm{II}_x(\theta, \theta)^{\perp}$ within an error of order $O(h)$. Thus, more information is needed if a more accurate geodesic distance estimate is needed.

4. Locally linear embedding and a variation by truncation

LLE [Roweis and Saul 2000] is a widely applied dimension-reduction technique in the manifold-learning community. Some of its theoretical properties have been explored in [Wu and Wu 2018]. For example, under the manifold setup, the barycentric coordinate is intimately related to the local covariance structure of the manifold. By reformulating the barycentric coordinate in the local covariance structure framework, the natural kernel associated with LLE is discovered, and it is very different from the kernel in graph Laplacian-based algorithms like Laplacian eigenmaps [Belkin and Niyogi 2003], diffusion maps [Coifman and Lafon 2006] or commute time embeddings [Qiu and Hancock 2007]. The regularization step in LLE is crucial to the whole algorithm, and it might provide different embeddings with different regularizations. In this section, we connect LLE and EIG via the local covariance structure, and use the geometric structure to explore a variation of LLE via truncation.

4.1. *A summary of the LLE algorithm.* We start with recalling the LLE algorithm in the finite sampling setup. Let $\mathcal{X} = \{x_i\}_{i=1}^n \subset \mathbb{R}^p$ denote a set of point clouds. Fix one point $x_k \in \mathcal{X}$. For $h > 0$, assume $\mathcal{N}_{x_k} := \mathcal{X} \cap (B_h^{\mathbb{R}^p}(x_k) \setminus \{x_k\}) = \{x_{k,1}, \ldots, x_{k,N_k}\}$, where $h > 0$. Here we use the h-radius ball to determine neighbors to be consistent with the following discussion. In practice we can use the k-nearest neighbors, where $k \in \mathbb{N}$ is determined by the user. The relationship between the h-radius ball and k-nearest neighbors is discussed in [Wu and Wu 2018, Section 5]. Define the *local data matrix associated with x_k* as

$$G_{n,h}(x_k) := \begin{bmatrix} | & & | \\ x_{k,1} - x_k & \cdots & x_{k,N_k} - x_k \\ | & & | \end{bmatrix} \in \mathbb{R}^{p \times N_k}.$$

With the local data matrix, the LLE algorithm is composed of three steps. First, for each $x_k \in \mathcal{X}$, find its barycentric coordinate associated with $\mathcal{N}_{x_k}$ by solving

$$w_{x_k} = \operatorname*{argmin}_{w \in \mathbb{R}^{N_k}, \, w^\top \mathbf{1}_{N_k} = 1} \left\| x_k - \sum_{j=1}^{N_k} w(j) x_{k,j} \right\|^2 \in \mathbb{R}^{N_k}. \tag{20}$$

Solving (20) is equivalent to minimizing $w^\top G_{n,h}(x_k)^\top G_{n,h}(x_k)w$ over $w \in \mathbb{R}^{N_k}$ under the constraint $w^\top \mathbf{1}_{N_k} = 1$. Since in general $G_{n,h}(x_k)^\top G_{n,h}(x_k)$ is singular, it is recommended in [Roweis and Saul 2000] to solve $(G_{n,h}(x_k)^\top G_{n,h}(x_k) + c I_{N_k \times N_k})y = \mathbf{1}_{N_k}$, where $c > 0$ is the regularizer chosen by the user, and hence obtain

$$w_{x_k} = \frac{(G_{n,h}(x_k)^\top G_{n,h}(x_k) + c I_{N_k \times N_k})^{-1} \mathbf{1}_{N_k}}{((G_{n,h}(x_k)^\top G_{n,h}(x_k) + c I_{N_k \times N_k})^{-1} \mathbf{1}_{N_k})^\top \mathbf{1}_N}. \tag{21}$$

By viewing w_{x_k} as the "affinity" of x_k and its neighbors, the second step is defining an *LLE matrix* $W \in \mathbb{R}^{n \times n}$ by

$$W_{k,l} = \begin{cases} w_{x_k}(j) & \text{if } x_l = x_{k,j} \in \mathcal{N}_{x_k}, \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

Finally, to reduce the dimension of $\mathcal{X}$ or to visualize $\mathcal{X}$, it is suggested in [Roweis and Saul 2000] to embed $\mathcal{X}$ into a low-dimensional Euclidean space

$$x_k \mapsto [v_1(k), \dots, v_\ell(k)]^\top \in \mathbb{R}^\ell \tag{23}$$

for each $x_k \in \mathcal{X}$, where ℓ 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. Note that in general W is not symmetric, and this is why the spectrum of $(I - W)^\top (I - W)$ is suggested but not that of $I - W$.

It is shown in [Wu and Wu 2018, Section 2] that by taking the relationship between the *sample covariance matrix* $C_{n,h}(x_k) := G_{n,h}(x_k)G_{n,h}(x_k)^\top$ and the Gramian matrix $G_{n,h}(x_k)^\top G_{n,h}(x_k)$ into account, we can rewrite (21) as

$$w_{x_k}^\top = \frac{\mathbf{1}_{N_k}^\top - \mathbf{1}_{N_k}^\top G_{n,h}(x_k)^\top \mathcal{I}_c(C_{n,h}(x_k))G_{n,h}(x_k)}{N_k - \mathbf{1}_{N_k}^\top G_{n,h}(x_k)^\top \mathcal{I}_c(C_{n,h}(x_k))G_{n,h}(x_k)\mathbf{1}_{N_k}}, \tag{24}$$

and view $\mathbf{1}_{N_k}^\top - \mathbf{1}_{N_k}^\top G_{n,h}(x_k)^\top \mathcal{I}_c(C_{n,h}(x_k))G_{n,h}(x_k)$ as the "kernel" associated with the LLE algorithm. It is clear that in general this kernel has negative values, and the LLE matrix is not a transition matrix. By defining

$$T_{n,x_k} := \mathcal{I}_c(C_{n,h}(x_k))G_{n,h}(x_k)\mathbf{1}_{N_k}, \tag{25}$$

we obtain the result claimed in [Wu and Wu 2018, Proposition 2.1].

4.2. *LLE and Mahalanobis distance in the continuous setup.* The above is for the general dataset. When the dataset $\mathcal{X} = \{\iota(x_i)\}_{i=1}^n \subset \iota(M) \subset \mathbb{R}^p$ is sampled from a manifold M, with (24) and (25), the asymptotical behavior of LLE under the manifold setup is explored in [Wu and Wu 2018]. As is shown in [loc. cit., (3.12)], in the continuous setup, the unnormalized kernel associated with LLE at $x \in M$ is

$$K_h^{\text{LLE}}(x, y) = [1 - T_{\iota(x)}^\top (\iota(y) - \iota(x))]\chi_{B_h^{\mathbb{R}^p}(\iota(x)) \cap \iota(M)}(\iota(y)), \tag{26}$$

where $y \in M$ and

$$T_{\iota(x)} := \mathcal{I}_c(C_h(x))[\mathbb{E}(X - \iota(x))\chi_{B_h^{\mathbb{R}^p}(x)}(X)] \in \mathbb{R}^p, \tag{27}$$

and hence the normalized kernel is defined by

$$P_h^{\mathrm{LLE}}(x, y) := \frac{K_h^{\mathrm{LLE}}(x, y)}{\int_M K_h^{\mathrm{LLE}}(x, y) P(y)\, dV(y)}. \tag{28}$$

Clearly, (24) and (25) are discretizations of $P_h^{\mathrm{LLE}}(x, y)$ and $\boldsymbol{T}_{\iota(x)}$, and their convergence behavior can be found in [Wu and Wu 2018, (3.7) and Theorem 3.1].

There are two geometric facts we should mention. First, rewriting (21) as (24) is equivalent to representing (21) in the frame-bundle setup, and hence an explicit form of the kernel underlying LLE. Second, the unnormalized kernel at x involves projecting $\iota(y) - \iota(x)$ onto the vector space spanned by $u_1, \dots, u_r$, and scaling the resulting vector componentwise by $(\lambda_1 + c)^{-1/2}, \dots, (\lambda_r + c)^{-1/2}$. When $r > d$, the involvement of the regularizer c and the eigenvalues $\lambda_{d+1}, \dots, \lambda_r$ are associated with the normal bundle. An important intuition is that when $c \to 0$, if we put aside the expectation in $\boldsymbol{T}_{\iota(x)}^{\top}(\iota(y) - \iota(x))$ and replace $\mathbb{E}(X - \iota(x))$ by $(\iota(y) - \iota(x))$, we obtain the term $(\iota(y) - \iota(x))^{\top} \mathcal{T}_r(C_h(x))(\iota(y) - \iota(x))$, which can be understood as a variation of the Mahalanobis distance. Therefore, we can interpret LLE as *mixing* together the neighbor information (the 0-1 kernel in (26), that is, $\chi_{B_h^{\mathbb{R}^p}(\iota(x)) \cap \iota(M)}(\iota(y))$) and the Mahalanobis distance of neighboring points (that is, $\boldsymbol{T}_{\iota(x)}^{\top}(\iota(y) - \iota(x)) \chi_{B_h^{\mathbb{R}^p}(\iota(x)) \cap \iota(M)}(\iota(y))$ in (26)) to design the kernel.

Lemma 3 guarantees that the first d eigenvalues of $C_h(x)$ are sufficiently large. However, the remaining $r - d$ nontrivial eigenvalues are small. Moreover, the remaining $r - d$ nontrivial eigenvalues are related to the curvature of the manifold. The importance of choosing the regularization constant c has been extensively discussed in [Wu and Wu 2018, Theorem 3.2], and it is known that c is to adjust the curvature information in these scaling factors $(\lambda_{d+1} + c)^{-1/2}, \dots, (\lambda_r + c)^{-1/2}$ for different purposes. To be more precise, the regularization c plays a role of "radio tuner". The larger the c is, the more enhanced the p.d.f. information will be; the smaller the c is, the more dominated the curvature information will be [Wu and Wu 2018, Theorem 3.2]. It is shown that if we want to obtain the Laplace–Beltrami operator of the Riemannian manifold, c should be chosen properly so as to *suppress* the curvature information contained in the $(d+1)$-th to the r-th eigenvalues.

4.3. *A variation of LLE and EIG.* Under the manifold setup, the above discussion suggests that the curvature information associated with the $d + 1, \dots, r$ eigenvalues and eigenvectors is *not needed* if we want to obtain the Laplace–Beltrami operator. Therefore, an alternative to choosing a regularization parameter is a direct truncation by taking only the largest d eigenvalues into account; that is, replace $\mathcal{I}_c[C_h(x)]$ in (26) by $\mathcal{T}_d[C_h(x)]$, and define the *normalized truncated LLE kernel* at $x \in M$ by

$$P_h^{\mathrm{tLLE}}(x, y) := \frac{K_h^{\mathrm{tLLE}}(x, y)}{\int_M K_h^{\mathrm{tLLE}}(x, y) P(y)\, dV(y)}, \tag{29}$$

where

$$K_h^{\mathrm{tLLE}}(x, y) := [1 - \widetilde{\boldsymbol{T}}_{\iota(x)}^{\top}(\iota(y) - \iota(x))] \chi_{B_h^{\mathbb{R}^p}(\iota(x)) \cap \iota(M)}(\iota(y)), \tag{30}$$

$y \in M$ and

$$\widetilde{\boldsymbol{T}}_{\iota(x)} := \mathcal{T}_d(C_h(x))[\mathbb{E}(X - \iota(x)) \chi_{B_h^{\mathbb{R}^p}(x)}(X)] \in \mathbb{R}^p. \tag{31}$$

Numerically when we have only finite data points, we run the same LLE algorithm, but replace $w_{x_k}^\top$ and T_{n,x_k} by the terms

$$\tilde{w}_{x_k}^\top := \frac{\mathbf{1}_{N_k}^\top - \tilde{T}_{n,x_k}^\top G_{n,h}(x_k)}{N_k - \tilde{T}_{n,x_k}^\top G_{n,h}(x_k)\mathbf{1}_{N_k}}, \tag{32}$$

and

$$\tilde{T}_{n,x_k} := \mathcal{T}_d(G_{n,h}(x_k)G_{n,h}(x_k)^\top)G_{n,h}(x_k)\mathbf{1}_{N_k}; \tag{33}$$

that is, instead of running a regularized pseudoinversion, we run a truncated inversion. *LLE with low-dimensional neighborhood representation* (LDR-LLE) proposed in [Goldberg and Ritov 2008] is an algorithm related to this truncated idea. While the geometric structure of the local covariance matrix is not specifically discussed in LDR-LLE, it can be systematically studied in our framework. For simplicity, we also call the above truncation scheme LDR-LLE.

Geometrically, $(\iota(y) - \iota(x))^\top \mathcal{T}_d[C_h(x)](\iota(y) - \iota(x))$ evaluates geodesic distances between points in the h-neighborhood of x using a method directly related to EIG in (9) when Φ is the identity. With the normalized truncated LLE kernel, we can proceed with the standard LLE dimension-reduction step. Therefore, LDR-LLE takes the neighbor information, $1 \times \chi_{B_h^{\mathbb{R}^p}(\iota(x)) \cap \iota(M)}(\iota(y)))$, and the EIG distance of neighboring points into account to design the kernel.

Under the same condition specified in [Wu and Wu 2018, Theorem 3.2], LDR-LLE has the same asymptotical behavior as that in [loc. cit., Theorem 3.3] with the properly chosen regularization; that is:

Theorem 12. *Assume the setup in Section 1.1. Take $f \in C^4(M)$. If $h = h(n)$ so that*

$$\frac{\sqrt{\log(n)}}{n^{1/2}h^{d/2+1}} \to 0$$

and $h \to 0$ when $n \to \infty$, with probability greater than $1 - n^{-2}$, for all $x_k \in \mathcal{X}$ we have

$$\sum_{j=1}^{N_k} \tilde{w}_{n,x_k}(j)f(x_{k,j}) = \int_M P_h^{\text{tLLE}}(x_k, y)f(y)P(y)\,dV(y) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2-1}}\right)$$

and

$$\int_M P_h^{\text{tLLE}}(x, y)f(y)P(y)\,dV(y) = f(y) + \frac{h^2}{2(d+2)}\Delta f(x) + O(h^3).$$

As a result, with probability greater than $1 - n^{-2}$, for all $x_k \in \mathcal{X}$ we have

$$\frac{1}{h^2}\left[\sum_{j=1}^{N_k} \tilde{w}_{n,x_k}(i)f(x_{k,j}) - f(x_k)\right] = \frac{1}{2(d+2)}\Delta f(x) + O(h) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\varepsilon^{d/2+1}}\right).$$

In other words, by LDR-LLE, we recover the Laplace–Beltrami operator of the manifold even if the second fundamental form is nontrivial and the sampling is nonuniform. While the proof is a direct modification of the proof of [Wu and Wu 2018, Theorem 3.2], we provide the proof in Appendix C for the sake of self-containedness. We mention that the finite sample convergence has the same statement and follows exactly the same line as that of [loc. cit., Theorem 3.1], so we skip it.

Based on the theoretical development, we identify two benefits of LDR-LLE. First, the normal bundle information is automatically removed by the truncation, and we obtain the intrinsic geometric quantity Δ. Second, the nonuniform sampling effect is automatically removed, and no kernel density estimation is needed. This comes from the fact that the EIG part of the kernel $\widetilde{T}_{\iota(x)}$ (or the Mahalanobis distance part of the kernel $T_{\iota(x)}$) automatically provides the density information (see Lemma 18). Compared with diffusion maps (DM), since we do not need normalization as is proposed in [Coifman and Lafon 2006] to eliminate the nonuniform sampling effect, the convergence rate is faster, as is stated in [Wu and Wu 2018, Theorem 3.3].

However, we emphasize that although this approach leads to aesthetic asymptotical properties under the manifold assumption, it may not work well when we do not have a manifold structure. For a general graph or point cloud, we found it more reliable to apply the original LLE, and a model or analysis explaining why is needed. Last but not least, note that this action requires that we know the dimension of the manifold, so even if we know the point cloud is sampled from a manifold, we need to estimate its dimension before applying LDR-LLE. This requirement might render LDR-LLE less applicable in nonmanifold data.

5. Numerical results

We demonstrate some numerical results associated with the algorithms studied in this paper. Uniformly sample $n = 8000$ points from the logarithmic spiral $L \subset \mathbb{R}^2$, a 1-dimensional manifold parametrized as $x(s) = \left(\frac{s}{\sqrt{2}} + 1\right) \cos \log\left(\frac{s}{\sqrt{2}} + 1\right)$ and $y(s) = \left(\frac{s}{\sqrt{2}} + 1\right) \sin \log\left(\frac{s}{\sqrt{2}} + 1\right)$, and choose $\varepsilon = 0.2$. We fix $x_k \in L$ and plot the absolute error between the true geodesic distance and the Euclidean and covariance-corrected distances. The result is shown in Figure 2. Note that the two "branches" correspond to differences in curvature on either side of x_k. In some manifold-learning algorithms, knowing the local
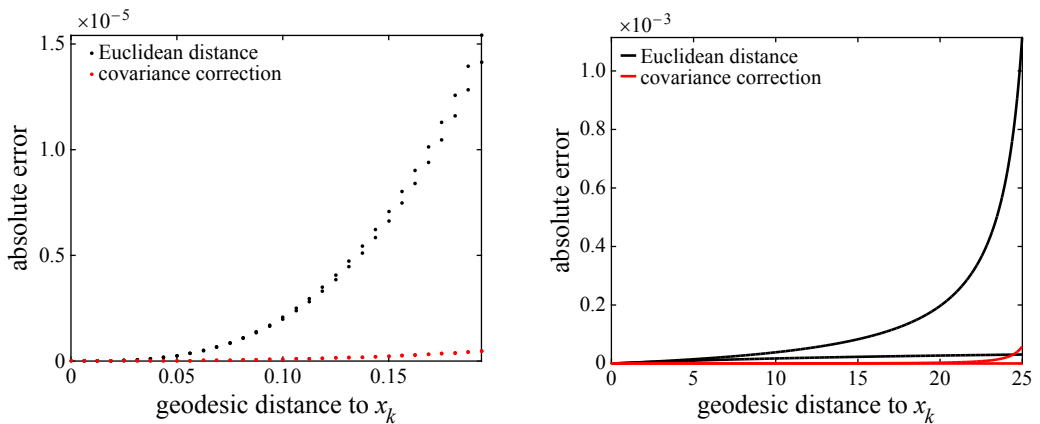

Figure 2. Comparison of the Euclidean-distance-based geodesic distance estimator (black) and the covariance-corrected geodesic distance estimator (red) in the logarithmic spiral L. Left: local distance for $\varepsilon = 0.2$. Right: global geodesic distance estimation by applying Dijkstra's algorithm with different local geodesic distance estimators.
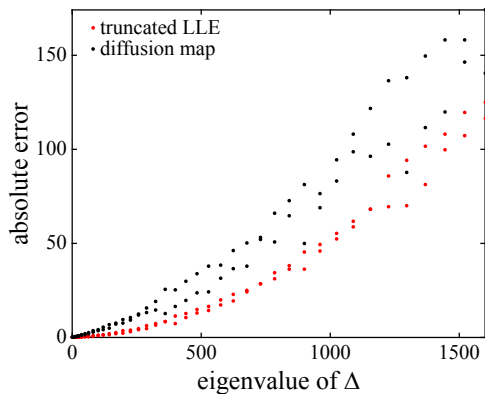
Figure 3. The eigenvalues of the Laplace–Beltrami operator on S^1 are well-recovered by LDR-LLE in the presence of a nonuniform sampling, compared with the α-normalized DM.

geodesic distance is not enough, and we need to know the global geodesic distance between any two points. ISOMAP [Tenenbaum et al. 2000] is a typical example. Usually, we estimate the global geodesic distance by applying Dijkstra's algorithm to find the shortest-path distance between all pairs, where the geodesic distance between neighboring points is estimated by the Euclidean distance. By combining the covariance-corrected geodesic distance estimator and Dijkstra's algorithm, we achieve a more accurate global geodesic distance, as is shown in Figure 2.

We show in Figure 3 that on $S^1 \subset \mathbb{R}^2$ with a nonuniform sampling, the eigenvalues of the Laplace–Beltrami operator can be accurately recovered using LDR-LLE. When compared with the eigenvalues obtained from the 1-normalized DM, we see that LDR-LLE performs better, which comes from the faster convergence rate of LDR-LLE. We use $n = 8000$ points and we choose $h = 0.03$. The nonuniform sampling is obtained as follows. Sample n points $\{\theta_i\}_{i=1}^n$ uniformly from $[0, 1]$ and set

$$x_i = \big(\cos(2\pi(\theta_i + 0.3\sin(\theta_i))), \sin(2\pi(\theta_i + 0.3\sin(\theta_i)))\big) \in S^1,$$

where $i = 1, \ldots, n$.

6. Conclusion

In this paper, we extend the knowledge of local covariance matrix, particularly the higher-order expansion of the local covariance matrix, to study two commonly applied manifold-learning algorithms, EIG and LLE. We provide a theoretical analysis of EIG under different situations under the manifold setup, and emphasize the importance of correctly estimating the dimension and its sensitivity to chosen parameters. The curvature effect is specifically carefully discussed. Under the trivial EIG setup when there is no deformation, that is, Φ is the identity, we show that the local covariance matrix structure allows us to obtain a more accurate geodesic distance for neighboring points. The geometric relationship between the local covariance matrix and LLE leads to a natural generalization of LLE by taking EIG into account. We provide a theoretical justification of LDR-LLE and compare its pros and cons with the original LLE.

Appendix A: Proof of Theorem 8

Suppose that M and N are d-dimensional closed Riemannian manifolds and $\Phi : M \to N$ is a diffeomorphism. Moreover, suppose that N is isometrically embedded in $\mathbb{R}^q$ via ι. In this section, all the derivatives of Φ are calculated in normal coordinates. Hence for $x \in M$,

$$\nabla^k \Phi(x) : \underbrace{T_x M \times \cdots \times T_x M}_{k} \to T_{\Phi(x)} N$$

is a multilinear map. We set $\nabla^k \Phi(x)(\theta, \ldots, \theta) = \nabla^k_{\theta \cdots \theta} \Phi(x)$ for $\theta \in S^{d-1} \subset T_x M$. For any $y \in \mathbb{R}^q$, we identify $T_y \mathbb{R}^q$ with $\mathbb{R}^q$. The proof of the theorem is composed of three lemmas, including Lemma 6.

Lemma 13. *Fix $x \in M$ and use polar coordinates $(t, \theta) \in [0, \infty) \times S^{d-1}$ to parametrize $T_x M$. For $w = \exp_x(\theta t)$ for sufficiently small t, let $y = \Phi(x)$ and $z = \Phi(w)$; then we have*

$$\iota(z) - \iota(y) = \iota_*|_y \vartheta t + \tfrac{1}{2}[\mathrm{II}_y(\vartheta, \vartheta) + \iota_*|_y \nabla^2_{\theta\theta} \Phi(x)]t^2$$
$$+ \tfrac{1}{6}[\nabla_\vartheta \mathrm{II}_y(\vartheta, \vartheta) + 3\mathrm{II}_y(\nabla^2_{\theta\theta} \Phi(x), \vartheta) + \iota_*|_y \nabla^3_{\theta\theta\theta} \Phi(x)]t^3 + O(t^4), \quad (34)$$

where $\vartheta := \nabla_\theta \Phi(x) \in T_y N$, II_y is the second fundamental form of $\iota(N)$ at $\iota(y)$, and $\nabla \mathrm{II}_y$ is the covariant derivative of the second fundamental form. Moreover, we have

$$2t^2 = (\iota(z) - \iota(y))^\top \iota_*|_y [\nabla \Phi(x) \nabla \Phi(x)^\top]^{-1} \iota_*|_y^\top (\iota(z) - \iota(y))$$
$$+ (\iota(y) - \iota(z))^\top \iota_*|_z [\nabla \Phi(w) \nabla \Phi(w)^\top]^{-1} \iota_*|_z^\top (\iota(y) - \iota(z)) + O(t^4). \quad (35)$$

The lemma above can be regarded as a generalization of Lemma 4. Indeed, when $N = M$ and Φ is the identity, $\nabla \Phi$ is the identity and we recover Lemma 4.

Proof. Equation (34) follows from a direct Taylor expansion. Since t is small enough, let $\gamma : [0, t] \mapsto M$ to be the unique unit speed geodesic from x to w; that is, $\gamma(0) = x$ and $\gamma(t) = w$. By (34), we have

$$\iota(z) - \iota(y) = \iota_*|_y \vartheta t + \tfrac{1}{2}[\mathrm{II}_y(\vartheta, \vartheta) + \iota_*|_y \nabla^2_{\gamma'(0)\gamma'(0)} \Phi(x)]t^2 + O(t^3),$$

where $\vartheta := \nabla_\theta \Phi(x) = \nabla \Phi|_x(\theta)$. Since the calculation is made in normal coordinates, we have $\iota_*|_y = \mathcal{O}_y J_{p,d}$, where $\mathcal{O}_y \in O(q)$ and $J_{q,d} \in \mathbb{R}^{q \times d}$ with 1 on the diagonal entries and 0 on the others. Therefore, we have $\iota_*|_y^\top \iota_*|_y = I_{d \times d}$. Moreover, $\iota_*|_y^\top \mathrm{II}_y(v, v) = 0$ for any $v \in S^{d-1}$. Hence, if we multiply by $[\nabla \Phi(x)]^{-1} \iota_*|_y^\top$ on both sides of above equation, we have

$$[\nabla \Phi(x)]^{-1} \iota_*|_y^\top (\iota(z) - \iota(y)) = \gamma'(0)t + \tfrac{1}{2}[\nabla \Phi(x)]^{-1} \nabla^2_{\gamma'(0)\gamma'(0)} \Phi(x)t^2 + O(t^3).$$

Consequently,

$$(\iota(z) - \iota(y))^\top \iota_*|_y [\nabla \Phi(x) \nabla \Phi(x)^\top]^{-1} \iota_*|_y^\top (\iota(z) - \iota(y))$$
$$= (\iota(z) - \iota(y))^\top \iota_*|_y [\nabla \Phi(x)^{-1}]^\top \nabla \Phi(x)^{-1} \iota_*|_y^\top (\iota(z) - \iota(y))$$
$$= \gamma'(0)^\top \gamma'(0)t^2 + \gamma'(0)^\top [\nabla \Phi(x)]^{-1} \nabla^2_{\gamma'(0)\gamma'(0)} \Phi(x)t^3 + O(t^4)$$
$$= t^2 + \gamma'(0)^\top [\nabla \Phi(x)]^{-1} \nabla^2_{\gamma'(0)\gamma'(0)} \Phi(x)t^3 + O(t^4). \quad (36)$$

Note that the last step follows from $\gamma'(0)^\top \gamma'(0) = 1$.

Let $c : [0, t] \mapsto M$ be the unique geodesic from w to x so that $c(0) = w$ and $c(t) = x$. Similarly, we obtain

$$
(\iota(y) - \iota(z))^\top \iota_*|_z [\nabla\Phi(w)\nabla\Phi(w)^\top]^{-1} \iota_*|_z^\top (\iota(y) - \iota(z))
$$
$$
= t^2 + c'(0)^\top [\nabla\Phi(w)]^{-1} \nabla^2_{c'(0)c'(0)} \Phi(w) t^3 + O(t^4)
$$
$$
= t^2 - \gamma'(t)^\top [\nabla\Phi(\gamma(t))]^{-1} \nabla^2_{\gamma'(t)\gamma'(t)} \Phi(\gamma(t)) t^3 + O(t^4)
$$
$$
= t^2 - \gamma'(0)^\top [\nabla\Phi(x)]^{-1} \nabla^2_{\gamma'(0)\gamma'(0)} \Phi(x) t^3 + O(t^4), \tag{37}
$$

where the second-to-last equality comes from $w = \gamma(t)$, $c'(0) = -\gamma'(t)$ and the last equality comes from the Taylor expansion of $\gamma'(t)^\top [\nabla\Phi(\gamma(t))]^{-1} \nabla^2_{\gamma'(t)\gamma'(t)} \Phi(\gamma(t))$ at $t = 0$. The conclusion follows from adding (36) and (37) together. $\qquad\square$

Proof of Lemma 6. Fix $x \in M$ and use polar coordinates $(t, \theta) \in [0, \infty) \times S^{d-1}$ to parametrize $T_x M$. Set $y' = \Phi(x')$ and $x' = \exp_x(\theta t)$. Let X be the random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the range M, where $\mathbb{P}$ is the probability measure defined on the sigma algebra $\mathcal{F}$ in the event space Ω. Define $\mathcal{P} = X_* \mathbb{P}$ to be the induced measure defined on the Borel sigma algebra on M. By the Radon–Nikodym theorem, $\mathcal{P}(x') = P(x')\, dV_M(x')$, where P is the probability density function of X associated with the Riemannian volume measure, dV_M, defined on M. Furthermore, define $\mathcal{Q}$ to be the induced measure by Φ defined on the Borel sigma algebra on N; that is, $\mathcal{Q} = \Phi_* \mathcal{P} = (\Phi \circ X)_* \mathbb{P}$. By Radon–Nikodym chain rule,

$$
Q(y') = \frac{d\mathcal{Q}(y')}{dV_N(y')} = \frac{d\mathcal{P}(\Phi^{-1}(y'))}{dV_M(\Phi^{-1}(y'))} \frac{dV_M(\Phi^{-1}(y'))}{dV_N(y')} = \frac{P(\Phi^{-1}(y'))}{|\nabla\Phi(\Phi^{-1}(y'))|} \tag{38}
$$

is the probability density function of the random variable $Y := \Phi \circ X$ associated with the Riemannian volume measure defined on N. The regularity of Q is thus the same as that of P.

By definition we have

$$
\mathcal{C}_\varepsilon(y) = \int_{E(y)} (\iota(y') - \iota(y))(\iota(y') - \iota(y))^\top Q(y')\, dV_N(y') \in \mathbb{R}^{q \times q}. \tag{39}
$$

By a direct Taylor expansion we have

$$
\mathcal{C}_\varepsilon(y) = \int_{B_\varepsilon(x)} (\iota\circ\Phi(x') - \iota\circ\Phi(x))(\iota\circ\Phi(x') - \iota\circ\Phi(x))^\top \frac{P(x')}{|\nabla\Phi(x')|} |\nabla\Phi(x')|\, dV_M(x')
$$
$$
= \int_{S^{d-1}} \int_0^\varepsilon (\iota_*|_y \nabla\Phi(x)\theta t + O(t^2))(\iota_*|_y \nabla\Phi(x)\theta t + O(t^2))^\top (P(x) + O(t))(t^{d-1} + O(t^{d+1}))\, dt\, d\theta
$$
$$
= P(x) \int_{S^{d-1}} \int_0^\varepsilon (\iota_*|_y \nabla\Phi(x)\theta)(\iota_*|_y \nabla\Phi(x)\theta)^\top t^{d+1} + O(t^{d+2})\, dt\, d\theta
$$
$$
= \frac{P(x)\varepsilon^{d+2}}{d+2} [\iota_*|_y \nabla\Phi(x)] \left\{ \int_{S^{d-1}} \theta\theta^\top\, d\theta \right\} [\iota_*|_y \nabla\Phi(x)]^\top + O(\varepsilon^{d+4}).
$$

Note that all terms of order t^{d+2} above contain a factor $\theta\theta^\top\theta$. Due to the symmetry of the sphere, $\int_{S^{d-1}} \theta\theta^\top\theta\, d\theta = 0$. Hence, the error in the last step above is of order $O(\varepsilon^{d+4})$ rather than $O(\varepsilon^{d+3})$.

Moreover, we have $\int_{S^{d-1}} \theta\theta^\top \, d\theta = (|S^{d-1}|/d)I_{d\times d}$; therefore

$$C_\varepsilon(y) = \frac{|S^{d-1}|P(x)\varepsilon^{d+2}}{d(d+2)} [\iota_*|_y \nabla\Phi(x)][\nabla\Phi(x)^\top \iota_*|_y{}^\top] + O(\varepsilon^{d+4}). \tag{40}$$

By a similar calculation, we have

$$\mathbb{E}[\chi_{E_\varepsilon(y)}(Y)] = \int_{E(y)} Q(y') \, dV_N(y') = \int_{B_\varepsilon(x)} P(x') \, dV_M(x')$$

$$= \int_{S^{d-1}} \int_0^\varepsilon (P(x) + O(t))(t^{d-1} + O(t^{d+1})) \, dt \, d\theta$$

$$= \frac{|S^{d-1}|P(x)\varepsilon^d}{d} + O(\varepsilon^{d+2}). \tag{41}$$

Note that all the terms of order t^d above contain a factor θ. Due to the symmetry of the sphere, $\int_{S^{d-1}} \theta \, d\theta = 0$. Hence, the error in the last step above is of order $O(\varepsilon^{d+2})$ rather than $O(\varepsilon^{d+1})$. The first result follows from taking the quotient of the above two equations.

By Lemma 13, we have

$$\iota \circ \Phi(x') - \iota \circ \Phi(x) = \iota_*|_y \nabla_\theta \Phi(x)t + \tfrac{1}{2}[\mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x)) + \iota_*|_y \nabla^2_{\theta\theta}\Phi(x)]t^2 + O(t^3).$$

If $v_1, v_2 \in (\iota_* T_y N)^\perp$, since $\iota_*|_y \nabla_\theta\Phi(x)$ and $\iota_*|_y \nabla^2_{\theta\theta}\Phi(x)$ are in $\iota_* T_y N$, then

$$v_1^\top(\iota \circ \Phi(x') - \iota \circ \Phi(x)) = \tfrac{1}{2}v_1^\top \mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x))t^2 + O(t^3), \tag{42}$$

$$v_2^\top(\iota \circ \Phi(x') - \iota \circ \Phi(x)) = \tfrac{1}{2}v_2^\top \mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x))t^2 + O(t^3), \tag{43}$$

and we have

$$v_1^\top C_\varepsilon(y) v_2$$

$$= \int_{B_\varepsilon(x)} v_1^\top(\iota \circ \Phi(x') - \iota \circ \Phi(x))(\iota \circ \Phi(x') - \iota \circ \Phi(x))^\top v_2 \, P(x') \, dV_M(x')$$

$$= \int_{S^{d-1}} \int_0^\varepsilon \tfrac{1}{4}(v_1^\top \mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x))t^2 + O(t^3))(v_2^\top \mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x))t^2 + O(t^3))^\top$$

$$\qquad\qquad\qquad\qquad\qquad\qquad \times (P(x) + O(t))(t^{d-1} + O(t^{d+1})) \, dt \, d\theta$$

$$= \frac{P(x)}{4} \int_{S^{d-1}} \int_0^\varepsilon v_1^\top \mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x))(\mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x)))^\top v_2 t^{d+3} + O(t^{d+4}) \, dt \, d\theta$$

$$= \frac{P(x)\varepsilon^{d+4}}{4(d+4)} \int_{S^{d-1}} v_1^\top \mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x))(\mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x)))^\top v_2 \, d\theta + O(\varepsilon^{d+6}).$$

Due to the symmetry of the sphere, the error in the last step above is of order $O(\varepsilon^{d+6})$ rather than $O(\varepsilon^{d+5})$. The result follows from taking the quotient. $\qquad\square$

To further control the influence of the truncated pseudoinverse, we need the following lemma to explore the higher-order eigenvalue structure of the local covariance matrix.

Lemma 14. *Up to a rotation of $\mathbb{R}^q$, we have*

$$\bar{\mathcal{C}}_\varepsilon(y) = \begin{bmatrix} \frac{1}{d+2}\nabla\Phi(x)\nabla\Phi(x)^\top + O(\varepsilon^2) & O(\varepsilon^2) \\ O(\varepsilon^2) & \frac{d}{4(d+4)}M\varepsilon^2 + O(\varepsilon^4) \end{bmatrix}, \tag{44}$$

where M is a $(q-d)\times(q-d)$ diagonal matrix satisfying

$$M_{i,i} = \frac{1}{|S^{d-1}|}\int_{S^{d-1}} e_{q-d+i}^\top \mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x))(\mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x)))^\top e_{q-d+i}\, d\theta. \tag{45}$$

Moreover, if $M_{i,i} = 0$, then $e_{q-d+i}^\top \mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x)) = 0$.

Proof. First, we can rotate $\iota(N)$ so that $\iota_* T_y N$ is generated by $\{e_1, e_2, \ldots, e_d\}$. In other words,

$$\iota_*|_y = \begin{bmatrix} \mathcal{O}_1 \\ 0 \end{bmatrix} \in \mathbb{R}^{q\times d},$$

where $\mathcal{O}_1 \in O_d$. Then, by the previous lemma we have

$$\bar{\mathcal{C}}_\varepsilon(y) = \begin{bmatrix} \frac{1}{d+2}\mathcal{O}_1\nabla\Phi(x)\nabla\Phi(x)^\top\mathcal{O}_1^\top + O(\varepsilon^2) & O(\varepsilon^2) \\ O(\varepsilon^2) & \overline{M}\varepsilon^2 + O(\varepsilon^4) \end{bmatrix}, \tag{46}$$

where $\overline{M}$ is a $(q-d)\times(q-d)$ symmetric matrix. If $\mathcal{O}_2$ diagonalizes $\overline{M}$, let

$$\mathcal{O} = \begin{bmatrix} \mathcal{O}_1 & 0 \\ 0 & \mathcal{O}_2 \end{bmatrix},$$

and we have

$$\mathcal{O}^\top\bar{\mathcal{C}}_\varepsilon(y)\mathcal{O} = \begin{bmatrix} \frac{1}{d+2}\nabla\Phi(x)\nabla\Phi(x)^\top + O(\varepsilon^2) & O(\varepsilon^2) \\ O(\varepsilon^2) & \frac{d}{4(d+4)}M\varepsilon^2 + O(\varepsilon^4) \end{bmatrix},$$

where M is a $(q-d)\times(q-d)$ diagonal matrix. And by the previous lemma

$$M_{i,i} = \frac{1}{|S^{d-1}|}\int_{S^{d-1}} e_{q-d+i}^\top \mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x))(\mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x)))^\top e_{q-d+i}\, d\theta.$$

Note that if $\theta = \sum_{j=1}^d \theta_j\partial_j \in S^{d-1} \subset T_y N$, then $e_{q-d+i}^\top\mathrm{II}_y(\nabla_\theta\Phi(x), \nabla_\theta\Phi(x)) = p_i(\theta_1, \theta_2, \ldots, \theta_d)$, where p_i is a quadratic form. If $M_{i,i} = 0$, then $\int_{S^{d-1}} p_i^2\, d\theta = 0$. Hence $p_i = 0$, and the conclusion follows. $\square$

We are now ready to finish the proof of Theorem 8.

Proof of Theorem 8. Note that $(\iota(z) - \iota(y))^\top \mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(y)](\iota(z) - \iota(y))$ is invariant under any rotation of $\mathbb{R}^q$. Therefore, by Lemma 14 we can assume that

$$\bar{\mathcal{C}}_\varepsilon(y) = \begin{bmatrix} \frac{1}{d+2}\nabla\Phi(x)\nabla\Phi(x)^\top + O(\varepsilon^2) & O(\varepsilon^2) \\ O(\varepsilon^2) & \frac{d}{4(d+4)}M\varepsilon^2 + O(\varepsilon^4) \end{bmatrix}, \tag{47}$$

where M is a $(q - d) \times (q - d)$ diagonal matrix shown in (45). Suppose that M has $l \geq 0$ zero diagonal entries. By assumption, we have $q - d - l$ eigenvalues of order ε^2 that are related to principle curvatures, and l eigenvalues of order ε^4. Write the eigendecomposition of $\bar{\mathcal{C}}_\varepsilon(y)$ as

$$\bar{\mathcal{C}}_\varepsilon(y) = U_\varepsilon(x)\Lambda_\varepsilon(x)U_\varepsilon(x)^\top, \tag{48}$$

where $U_\varepsilon(x) \in O(q)$ and $\Lambda_\varepsilon(x)$ is a $q \times q$ diagonal matrix. By the perturbation argument (see Appendix A of [Wu and Wu 2018] for details), $\Lambda_\varepsilon(x)$ and $U_\varepsilon(x)$ satisfy

$$\Lambda_\varepsilon(x) = \begin{bmatrix} \frac{1}{d+2}\Lambda_1(x) + O(\varepsilon^2) & 0 & 0 \\ 0 & \frac{d}{4(d+4)}\Lambda_2(x)\varepsilon^2 + O(\varepsilon^4) & 0 \\ 0 & 0 & O(\varepsilon^4) \end{bmatrix}, \tag{49}$$

where $\Lambda_1(x)$ is the eigenvalue matrix of $\nabla\Phi(x)\nabla\Phi(x)^\top$, which is of order 1, and $\Lambda_2(x)$ is a $(q-d-l) \times (q-d-l)$ diagonal matrix which consists of nonzero diagonal entries of M, which is also of order 1, and

$$U_\varepsilon(x) = \begin{bmatrix} U_1(x) & 0 & 0 \\ 0 & U_2(x) & 0 \\ 0 & 0 & U_3(x) \end{bmatrix} (I_{p \times p} + \varepsilon^2 \mathsf{S}(x)) + O(\varepsilon^4), \tag{50}$$

where $U_1(x) \in \mathbb{O}(d)$ is the orthonormal eigenvector matrix of $\Phi(x)\nabla\Phi(x)^\top$, $U_2(x) \in O(q-d-l)$, $U_3(x) \in O(l)$ and $\mathsf{S}(x)$ is antisymmetric.[1] We now finish the proof.

Case 1: $\alpha = d$. In this case,

$$\mathcal{T}_d[\bar{C}_\varepsilon(y)] = (d+2)\begin{bmatrix} U_1(x) + O(\varepsilon^2) \\ O(\varepsilon^2) \end{bmatrix} [\Lambda_{1,\varepsilon}^{-1}(x) + O(\varepsilon^2)] \begin{bmatrix} U_1(x)^\top + O(\varepsilon^2) & O(\varepsilon^2) \end{bmatrix}$$

$$= (d+2)\begin{bmatrix} [\nabla\Phi(x)\nabla\Phi(x)^\top]^{-1} & 0 \\ 0 & 0 \end{bmatrix} + O(\varepsilon^2)$$

$$= (d+2)\iota_*|_y[\nabla\Phi(x)\nabla\Phi(x)^\top]^{-1}\iota_*|_y^\top + O(\varepsilon^2).$$

Therefore,

$$(\iota(z) - \iota(y))^\top \mathcal{T}_d[\bar{C}_\varepsilon(y)](\iota(z) - \iota(y))$$
$$= (d+2)(\iota(z) - \iota(y))^\top\iota_*|_y[\nabla\Phi(x)\nabla\Phi(x)^\top]^{-1}\iota_*|_y^\top(\iota(z) - \iota(y)) + O(\varepsilon^2\|\iota(y) - \iota(z)\|_{\mathbb{R}^q}^2)$$
$$= (d+2)(\iota(z) - \iota(y))^\top\iota_*|_y[\nabla\Phi(x)\nabla\Phi(x)^\top]^{-1}\iota_*|_y^\top(\iota(z) - \iota(y)) + O(\varepsilon^2 t^2), \tag{51}$$

where the last step follows from $O(\|\iota(y) - \iota(z)\|_{\mathbb{R}^q}^2) = O(t^2)$ by Lemma 13. Similarly,

$$(\iota(y) - \iota(z))^\top\mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(z)](\iota(y) - \iota(z))$$
$$= (d+2)(\iota(y) - \iota(z))^\top\iota_*|_z[\nabla\Phi(w)\nabla\Phi(w)^\top]^{-1}\iota_*|_z^\top(\iota(y) - \iota(z)) + O(\varepsilon^2 t^2). \tag{52}$$

---

[1]We mention that if the eigenvalues of $\Phi(x)\nabla\Phi(x)^\top$ repeat, $U_1(x)$ may not be uniquely determined, or if the nonzero diagonal terms of $\Lambda_2(x)$ repeat, $U_2(x)$ may not be the identity matrix. In this case, if one wants to fully determine $U_1(x)$, $U_2(x)$ and $U_3(x)$, he/she needs to further explore the higher-order expansion of $\bar{\mathcal{C}}_\varepsilon(y)$. Since it is irrelevant to this proof, we refer readers with interest to Appendix A of [Wu and Wu 2018].

By Lemma 13, we have

$$\mathrm{EIG}_d^2(y,z) = (\iota(z)-\iota(y))^\top \left[\frac{\mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(y)] + \mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(z)]}{2} \right] (\iota(z)-\iota(y))$$

$$= (d+2)t^2 + O(t^2\varepsilon^2) + O(t^4). \tag{53}$$

<u>Preparation for $\alpha \neq d$.</u> To show the case when $\alpha \neq d$, we need some preparations. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q$ denote the eigenvalues of $\bar{\mathcal{C}}_\varepsilon(y)$, and let $\{u_i\}$ be the corresponding normalized eigenvectors. Note that u_i is the i-th column of $[U_1(x)^\top \ 0 \ 0]^\top + O(\varepsilon^2)$ when $1 \leq i \leq d$ in (50). Based on (50), for $1 \leq i \leq d$, the first d entries of u_i are of order 1 and the other entries are of order ε^2. Note that the first d entries of u_i are associated with the tangent space and the others are associated with the normal space. For $d+1 \leq i \leq q-l$, the tangent components of u_i (the first d entries) are of order ε^2 and the normal components of u_i (the remaining $p-d$ entries) are of order 1. On the other hand, based on Lemma 13, the tangent component of $\iota(z)-\iota(y)$ is of order t and the normal component of $\iota(z)-\iota(y)$ is of order t^2. Following the notation in Lemma 13 and putting the above together, for $1 \leq i \leq p-l$, we have

$$(\iota(z)-\iota(y))^\top u_i = O(t) \qquad \text{for } 1 \leq i \leq d, \tag{54}$$

$$(\iota(z)-\iota(y))^\top u_i = O(t^2 + t\varepsilon^2) \quad \text{for } d+1 \leq i \leq q-l. \tag{55}$$

For $q-l+1 \leq i \leq q$, we have $u_i = \bar{u}_i + O(\varepsilon^2)$, where $\bar{u}_i$ is $(i-q-l)$'s column of $[0 \ 0 \ U_3(x)^\top]^\top$ shown in (50). By Lemma 13, $\bar{u}_i^\top(\iota(z)-\iota(y))$ becomes

$$\bar{u}_i^\top \iota_*|_y \nabla_\theta \Phi(x)t + \frac{1}{2}[\bar{u}_i^\top \mathrm{II}_y(\nabla_\theta \Phi(x), \nabla_\theta \Phi(x)) + \bar{u}_i^\top \iota_*|_y \nabla_{\theta\theta}^2 \Phi(x)]t^2 + O(t^3).$$

Note that $\bar{u}_i^\top \iota_*|_y \nabla_\theta \Phi(x) = 0$ and $\bar{u}_i^\top \iota_*|_y \nabla_{\theta\theta}^2 \Phi(x) = 0$, since $\bar{u}_i$ is in the normal direction. By Lemma 14, we have $\bar{u}_i^\top \mathrm{II}_y(\nabla_\theta \Phi(x), \nabla_\theta \Phi(x)) = 0$. Therefore $\bar{u}_i^\top(\iota(z)-\iota(y)) = O(t^3)$. Since $u_i - \bar{u}_i$ is of order $O(\varepsilon^2)$ and $\iota(z)-\iota(y)$ is of order $O(t)$, we have $(u_i - \bar{u}_i)^\top(\iota(z)-\iota(y)) = O(t\varepsilon^2)$. Putting the above together gives

$$(\iota(z)-\iota(y))^\top u_i = O(t^3 + t\varepsilon^2) \quad \text{for } q-l+1 \leq i \leq q. \tag{56}$$

With (54), (55) and (56), we can finish the proof for $\alpha \neq d$.

<u>Case 2:</u> $1 \leq \alpha < d$. In this case, we have

$$(\iota(z)-\iota(y))^\top(\mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(y)] - \mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(y)])(\iota(z)-\iota(y)) = (\iota(z)-\iota(y))^\top \left[\sum_{j=\alpha+1}^{d} \frac{u_j u_j^\top}{\lambda_j} \right] (\iota(z)-\iota(y)). \tag{57}$$

Recall that λ_j is of order 1 for $1 \leq j \leq d$ and also (54). As a result,

$$(\iota(z)-\iota(y))^\top \mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(y)](\iota(z)-\iota(y)) = (\iota(z)-\iota(y))^\top \mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(y)](\iota(z)-\iota(y)) + O(t^2).$$

Similarly,

$$(\iota(z)-\iota(y))^\top \mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(z)](\iota(z)-\iota(y)) = (\iota(z)-\iota(y))^\top \mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(z)](\iota(z)-\iota(y)) + O(t^2).$$

We then achieve the claim

$$\text{EIG}_\alpha^2(y, z) = (d + 2)t^2 + O(t^2). \tag{58}$$

We now show a special case when $1 \le \alpha < d$. Let the geodesic distance between y and z be $d(y, z)$; then by Lemma 13, we have $d(y, z) = \|\iota(y) - \iota(z)\|_{\mathbb{R}^q} + O(\|\iota(y) - \iota(z)\|_{\mathbb{R}^q}^3) = O(t)$. In a special case when $\iota_*|_y \vartheta(y) \in V_\alpha(y)$, where $V_\alpha(y)$ is the subspace of $\iota_* T_y N$ generated by the first α eigenvectors of $\bar{\mathcal{C}}_\varepsilon(y)$, we have

$$(\iota(z) - \iota(y))^\top (\mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(y)] - \mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(y)])(\iota(z) - \iota(y))$$
$$= (\iota(z) - \iota(y))^\top \left[\sum_{j=\alpha+1}^{d} \frac{u_j u_j^\top}{\lambda_j} \right] (\iota(z) - \iota(y)) = O(d(y, z)^4) = O(t^4). \tag{59}$$

If furthermore $\iota_*|_z \vartheta(z) \in V_\alpha(z)$, where $V_\alpha(z)$ is the subspace of $\iota_* T_z N$ generated by the first α eigenvectors of $\bar{\mathcal{C}}_\varepsilon(z)$, we have

$$(\iota(z) - \iota(y))^\top (\mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(z)] - \mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(z)])(\iota(z) - \iota(y)) = O(t^4) \tag{60}$$

and hence

$$\text{EIG}_\alpha^2(y, z) = (d + 2)t^2 + O(t^2 \varepsilon^2) + O(t^4). \tag{61}$$

Case 3: $\alpha > d$. In this case, we have $0 \le l_y \le q - d$. By a similar calculation, we have

$$(\iota(z) - \iota(y))^\top (\mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(y)] - \mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(y)])(\iota(z) - \iota(y))$$
$$= (\iota(z) - \iota(y))^\top \left[\sum_{j=d+1}^{q-l_y} \frac{u_j u_j^\top}{\lambda_j} + \sum_{j=q-l_y+1}^{\alpha} \frac{u_j u_j^\top}{\lambda_j} \right] (\iota(z) - \iota(y)). \tag{62}$$

Since λ_j is of order ε^2 for $d + 1 \le j \le q - l_y$, by (55)

$$(\iota(z) - \iota(y))^\top \sum_{j=d+1}^{q-l_y} \frac{u_j u_j^\top}{\lambda_j} (\iota(z) - \iota(y)) = O((t^2/\varepsilon + t\varepsilon)^2). \tag{63}$$

When $q - l_y + 1 \le j \le \alpha$, the eigenvalue λ_j is of order ε^4 by assumption. By (56)

$$(\iota(z) - \iota(y))^\top \sum_{j=q-l_y+1}^{\alpha} \frac{u_j u_j^\top}{\lambda_j} (\iota(z) - \iota(y)) = O((t^3/\varepsilon^2 + t)^2).$$

Hence, when $\alpha \ge q - l_y + 1$,

$$(\iota(z) - \iota(y))^\top (\mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(y)] - \mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(y)])(\iota(z) - \iota(y)) = O((t^2/\varepsilon + t\varepsilon)^2 + (t^3/\varepsilon^2 + t)^2);$$

otherwise we have

$$(\iota(z) - \iota(y))^\top (\mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(y)] - \mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(y)])(\iota(z) - \iota(y)) = O((t^2/\varepsilon + t\varepsilon)^2).$$

By the same argument, when $\alpha \ge q - l_z + 1$, we have

$$(\iota(z) - \iota(y))^\top (\mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(z)] - \mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(z)])(\iota(z) - \iota(y)) = O((t^2/\varepsilon + t\varepsilon)^2 + (t^3/\varepsilon^2 + t)^2);$$

otherwise we have

$$(\iota(z) - \iota(y))^\top (\mathcal{T}_\alpha[\bar{\mathcal{C}}_\varepsilon(z)] - \mathcal{T}_d[\bar{\mathcal{C}}_\varepsilon(z)])(\iota(z) - \iota(y)) = O((t^2/\varepsilon + t\varepsilon)^2).$$

Summing above two equations together leads to the fact that when $l_y > 0$ and $\alpha \geq q - l_y + 1$ hold, or $l_z > 0$ and $\alpha \geq q - l_z + 1$ hold, we have

$$\text{EIG}_\alpha^2(y, z) = (d + 2)t^2 + O((t^2/\varepsilon + t\varepsilon)^2 + (t^3/\varepsilon^2 + t)^2); \tag{64}$$

otherwise we have

$$\text{EIG}_\alpha^2(y, z) = (d + 2)t^2 + O((t^2/\varepsilon + t\varepsilon)^2).$$

Set $t = \varepsilon^\beta$, where $\beta \geq 1$. By a straightforward calculation, when $l_y = 0$ or $\alpha \leq q - l_y$ holds, and $l_z = 0$ or $\alpha \leq q - l_z$ holds, t^2 dominates $O((t^2/\varepsilon + t\varepsilon)^2)$ asymptotically when $\beta > 1$; otherwise t^2 and $O((t^2/\varepsilon + t\varepsilon)^2 + (t^3/\varepsilon^2 + t)^2)$ are asymptotically of the same order. We thus conclude the claim. $\quad\square$

Remark 15. We would like to make a comment when $l_y > 0$ and $\alpha > q - l_y$ hold, or $l_z > 0$ and $\alpha > q - l_z$ hold. To simplify the discussion, assume $l_y = l_z = l > 0$ and $\alpha > q - l$. For $u \in \mathbb{R}^q$, we write $u = [u_1^\top, u_2^\top, u_3^\top]^\top$, where $u_1 \in \mathbb{R}^d$. $u_2 \in \mathbb{R}^{q-d-l}$ and $u_3 \in \mathbb{R}^l$. As we stated in the proof of (55), $u_j = [O(\varepsilon^2), O(1), O(\varepsilon^2)]^\top$ for $j = d+1, \ldots, q-l$. On the other hand, $\iota(z) - \iota(y) = [O(t), O(t^2), O(t^3)]^\top$. Here, based on the structure of $\bar{\mathcal{C}}_\varepsilon(y)$, the first d components of $\iota(z) - \iota(y)$ are in the tangent direction of $\iota(N)$ at $\iota(y)$, and they are of order $O(t)$. The next $q - d - l$ components are in the direction of the second fundamental form of ι at $\iota(y)$, and they are of order $O(t^2)$. The last l components are in the normal direction and perpendicular to the second fundamental form of ι at $\iota(y)$, and they are of order $O(t^3)$. When we calculate the product $(\iota(z) - \iota(y))^\top u_j$, the products of components in the tangent directions and the products of components in the normal directions cannot be canceled in general. The argument holds for (56). Hence, the order estimation for the case $\alpha > d$ cannot be improved for an arbitrary vector $\iota(z) - \iota(y)$ without imposing more conditions. In other words, an expansion of u_j or $\iota(z) - \iota(y)$ into higher orders might not improve the results.

Appendix B: Proof of Theorem 10

By Lemma 3, we have

$$U_{\bar{h}}(x) = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} + O(\bar{h}^2),$$

where $U_1 \in \mathbb{O}(d)$ and $U_2 \in \mathbb{O}(p - d)$. By (6),

$$\iota(y) - \iota(x) = (\iota_* \theta)t + \frac{\text{II}_x(\theta, \theta)}{2} t^2 + O(t^3).$$

By (7),

$$t = h + \frac{\|\text{II}_x(\theta, \theta)\|^2}{24} h^3 + O(h^4).$$

Hence

$$\iota(y) - \iota(x) = (\iota_* \theta)h + \frac{\text{II}_x(\theta, \theta)}{2} h^2 + O(h^3).$$

Note that $(\iota_*\theta)h$ is in $\iota_* T_x M$. Therefore

$$P_h^{\perp}(\iota(y) - \iota(x)) = \frac{\mathrm{II}_x(\theta, \theta)}{2}h^2 + O(h^3 + h\bar{h}^2) = \frac{\mathrm{II}_x(\theta, \theta)}{2}h^2 + O(h\bar{h}^2),$$

where we use the fact that $h < \bar{h}$ in the last step. Hence,

$$\frac{\|P_h^{\perp}(\iota(y) - \iota(x))\|^2_{\mathbb{R}^p}}{6h} = \frac{\|\mathrm{II}_x(\theta, \theta)\|^2}{24}h^3 + O(h^2\bar{h}^2).$$

The conclusion follows.　　　　　　　　　　　　　　　　　　　　　　　　　　　　$\square$

Appendix C: Proof of Theorem 12

The proof of Theorem 12 consists of two steps, the variance analysis and bias analysis. The variance analysis of LDR-LLE is similar to Case 0 in [Wu and Wu 2018, Theorem 3.1], so we only provide the result and the proof is omitted.

Proposition 16. *Fix $f \in C(\iota(M))$. Suppose $h = h(n)$ so that*

$$\frac{\sqrt{\log(n)}}{n^{1/2}h^{d/2+1}} \to 0$$

and $h \to 0$ as $n \to \infty$. With probability greater than $1 - n^{-2}$, for all $x_k \in \mathcal{X}$,

$$\sum_{j=1}^{N} \tilde{w}_{n,x_k}(i) f(x_{k,j}) = \int_M P_h^{\mathrm{tLLE}}(x_k, y) f(y) P(y) \, dV(y) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}h^{d/2-1}}\right).$$

The bias analysis part of Theorem 12 depends on the following two technical lemmas. We use the following notation to simplify the proof. For $p, d \in \mathbb{N}$ such that $d \leq p$, let $J_{p,d} \in \mathbb{R}^{p \times d}$ be such that the (i, i) entry is 1 for $i = 1, \ldots, d$, and the other entries are 0. For $v \in \mathbb{R}^p$,

$$v = [\![v_1, \, v_2]\!] \in \mathbb{R}^p, \tag{65}$$

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, by a proper translation and rotation of $\iota(M)$ in $\mathbb{R}^p$ so that $\iota(x) = 0$ and $\iota_* T_x$ occupies the first d axes of $\mathbb{R}^p$, for $v = [\![v_1, \, v_2]\!] \in T_{\iota(x)}\mathbb{R}^p$, $v_1 = J_{p,d}^{\top} v$ is tangential to $\iota_* T_x M$ and $[\![0, \, v_2]\!]$ is normal to $\iota_* T_x M$.

Lemma 17 [Wu and Wu 2018, Lemma B.5]. *Fix $x \in M$ and $f \in C^3(M)$. When $h > 0$ is sufficiently small, the following expansions hold*:

(1) *The scalar $\mathbb{E}[\chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)]$ satisfies*

$$\mathbb{E}[\chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)] = P(x)\frac{|S^{d-1}|}{d}h^d + O(h^{d+2}).$$

(2) *The scalar* $\mathbb{E}[(f(X) - f(x))\chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)]$ *satisfies*

$$\mathbb{E}[(f(X) - f(x))\chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)] = \frac{|S^{d-1}|}{d(d+2)}\left[\frac{1}{2}P(x)\Delta f(x) + \nabla f(x) \cdot \nabla P(x)\right]h^{d+2} + O(h^{d+3}).$$

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

$$\mathbb{E}[(X - \iota(x))\chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)] = \left[\!\!\left[\frac{|S^{d-1}|}{d(d+2)}J_{p,d}^{\top}\iota_*\nabla P(x)h^{d+2} + O(h^{d+3}),\ O(h^{d+2})\right]\!\!\right].$$

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

$$\mathbb{E}[(X-\iota(x))(f(X)-f(x))\chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)] = \left[\!\!\left[\frac{|S^{d-1}|}{d(d+2)}J_{p,d}^{\top}P(x)\iota_*\nabla f(x)h^{d+2} + O(h^{d+3}),\ O(h^{d+4})\right]\!\!\right].$$

The proof of the above lemma can be found in [Wu and Wu 2018, Lemma B.5]. The next lemma describes the vector T when h is sufficiently small. The main significance is that the vector $\widetilde{T}_{\iota(x)}$ almost recovers $\nabla \log P(x)$ in the tangent direction.

Lemma 18. *Fix $x \in M$. When h is sufficiently small, we have*

$$\widetilde{T}_{\iota(x)} = \left[\!\!\left[\frac{J_{p,d}^{\top}\iota_*\nabla P(x)}{P(x)} + O(h^2),\ O(h^2)\right]\!\!\right].$$

Proof. Note that

$$\widetilde{T}_{\iota(x)} := \mathcal{T}_d[C_h(x)][\mathbb{E}(X - \iota(x))\chi_{B_h^{\mathbb{R}^p}(\iota(x))}] = \sum_{i=1}^{d}\frac{u_i u_i^{\top}[\mathbb{E}(X - \iota(x))\chi_{B_h^{\mathbb{R}^p}(\iota(x))}]}{\lambda_i},$$

where u_i and λ_i form the i-th eigenpair of $C_h(x)$.

By Lemma 3,

$$u_i = \begin{bmatrix} U_1 J_{p,d}^{\top}e_i + O(h^2) \\ O(h^2) \end{bmatrix}, \quad i = 1, \ldots, d$$

and $U_1 \in \mathbb{O}(d)$, and

$$\lambda_i = \frac{|S^{d-1}|P(x)}{d(d+2)}h^{d+2} + O(h^{d+4}).$$

By Lemma 17 , we have for $1 \leq i \leq d$

$$u_i^{\top}\mathbb{E}[(X - \iota(x))\chi_{B_h^{\mathbb{R}^p}(x_k)}(X)]$$

$$= \frac{|S^{d-1}|}{d+2}\left[\!\!\left[\frac{J_{p,d}^{\top}\iota_*\nabla P(x)}{d}h^{d+2} + O(h^{d+4}),\ O(h^{d+2})\right]\!\!\right] \cdot \left[\!\!\left[U_1 J_{p,d}^{\top}e_i + O(h^2),\ O(h^2)\right]\!\!\right]$$

$$= \frac{|S^{d-1}|}{d(d+2)}(\iota_*\nabla P(x))^{\top}J_{p,d}U_1 J_{p,d}^{\top}e_i h^{d+2} + O(h^{d+4}).$$

Thus, for $1 \le i \le d$,

$$
\frac{u_i^\top \mathbb{E}[(X - \iota(x)) \chi_{B_h^{\mathbb{R}^p}(x_k)}(X)]}{\lambda_i} = \frac{\dfrac{|S^{d-1}|}{d(d+2)}(\iota_* \nabla P(x))^\top J_{p,d} U_1 J_{p,d}^\top e_i h^{d+2} + O(h^{d+4})}{\dfrac{|S^{d-1}| P(x)}{d(d+2)} h^{d+2} + O(h^{d+4})}
$$

$$
= \frac{(\iota_* \nabla P(x))^\top J_{p,d} U_1 J_{p,d}^\top e_i}{P(x)} + O(h^2).
$$

Hence, we have

$$
\widetilde{T}_{\iota(x)} = \sum_{i=1}^d \frac{u_i^\top \mathbb{E}[(X - \iota(x)) \chi_{B_h^{\mathbb{R}^p}(x_k)}(X)]}{\lambda_i} u_i
$$

$$
= \sum_{i=1}^d \left(\frac{(\iota_* \nabla P(x))^\top J_{p,d} U_1 J_{p,d}^\top e_i}{P(x)} + O(h^2) \right) \left[\!\left[U_1 J_{p,d}^\top e_i + O(h^2),\ O(h^2) \right]\!\right]
$$

$$
= \left[\!\left[\frac{J_{p,d}^\top \iota_* \nabla P(x)}{P(x)} + O(h^2),\ O(h^2) \right]\!\right],
$$

where the last step follows from the fact that $U_1 \in \mathbb{O}(d)$. $\square$

We are ready to prove Theorem 12.

Proof of Theorem 12. Note that

$$
\int_M P_{\mathrm{tLLE}}(x, y) f(y) P(y)\, dV(y) - f(x)
$$

$$
= \frac{\mathbb{E}[(f(X) - f(x)) \chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)] - \widetilde{T}_{\iota(x)}^\top \mathbb{E}[(X - \iota(x))(f(X) - f(x)) \chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)]}{\mathbb{E}[\chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)] - \widetilde{T}_{\iota(x)}^\top \mathbb{E}[(X - \iota(x)) \chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)]}.
$$

By Lemmas 17 and 18, we have

$$
\widetilde{T}_{\iota(x)}^\top \mathbb{E}[(X - \iota(x)) \chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)] = O(h^{d+2}), \tag{66}
$$

and

$$
\widetilde{T}_{\iota(x)}^\top \mathbb{E}[(X - \iota(x))(f(X) - f(x)) \chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)] = \frac{|S^{d-1}|}{d(d+2)} \nabla P(x) \cdot \nabla f(x) h^{d+2} + O(h^{d+3}).
$$

Hence,

$$
\mathbb{E}[\chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)] - T_{\iota(x)}^\top \mathbb{E}[(X - \iota(x)) \chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)] = P(x) \frac{|S^{d-1}|}{d} h^d + O(h^{d+2}),
$$

and

$$
\mathbb{E}[(f(X) - f(x)) \chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)] - T_{\iota(x)}^\top \mathbb{E}[(X - \iota(x))(f(X) - f(x)) \chi_{B_h^{\mathbb{R}^p}(\iota(x))}(X)]
$$

$$
= \frac{|S^{d-1}|}{d(d+2)} \left[\tfrac{1}{2} P(x) \Delta f(x) + \nabla f(x) \cdot \nabla P(x) \right] h^{d+2} - \frac{|S^{d-1}|}{d(d+2)} \nabla P(x) \cdot \nabla f(x) h^{d+2} + O(h^{d+3})
$$

$$
= \frac{|S^{d-1}|}{2d(d+2)} P(x) \Delta f(x) h^{d+2} + O(h^{d+3}).
$$

If we combine the above two equations, we conclude the claim that

$$\int_M P_{\text{tLLE}}(x, y) f(y) P(y) \, dV(y) - f(x) = \frac{1}{2(d+2)} \Delta f(x) h^2 + O(h^3). \qquad \square$$

Acknowledgement

References

[Arias-Castro et al. 2017] E. Arias-Castro, G. Lerman, and T. Zhang, "Spectral clustering based on local PCA", *J. Mach. Learn. Res.* **18** (2017), art. id. 9. MR Zbl

[Belkin and Niyogi 2003] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation", *Neural Comp.* **15**:6 (2003), 1373–1396. Zbl

[Bernstein and Kuleshov 2014] A. Bernstein and A. Kuleshov, "Data-based manifold reconstruction via tangent bundle manifold learning", poster from ICML-2014, Topological Methods for Machine Learning Workshop, 2014, available at http://topology.cs.wisc.edu/KuleshovBernstein.pdf.

[Brand 2003] M. Brand, "Charting a manifold", pp. 961–968 in *Advances in neural information processing systems, 15* (*NIPS 2002*) (British Columbia, 2002), edited by S. Becker et al., MIT press, Cambridge, MA, 2003.

[Cheng and Wu 2013] M.-Y. Cheng and H.-T. Wu, "Local linear regression on manifolds and its geometric interpretation", *J. Amer. Statist. Assoc.* **108**:504 (2013), 1421–1434. MR Zbl

[Coifman and Lafon 2006] R. R. Coifman and S. Lafon, "Diffusion maps", *Appl. Comput. Harmon. Anal.* **21**:1 (2006), 5–30. MR Zbl

[Donoho and Grimes 2003] D. L. Donoho and C. Grimes, "Hessian eigenmaps: locally linear embedding techniques for high-dimensional data", *Proc. Natl. Acad. Sci. USA* **100**:10 (2003), 5591–5596. MR Zbl

[Goldberg and Ritov 2008] Y. Goldberg and Y. Ritov, "LDR-LLE: LLE with low-dimensional neighborhood representation", pp. 43–54 in *ISVC 2008: Advances in visual computing* (Las Vegas, 2008), edited by R. Boyle et al., Lecture Notes in Computer Science **5259**, Springer, 2008.

[Goldberg et al. 2009] A. B. Goldberg, X. Zhu, A. Singh, Z. Xu, and R. Nowak, "Multi-manifold semi-supervised learning", pp. 169–176 in *Artificial intelligence and statistics* (Clearwater Beach, FL, 2009), edited by D. van Dyk and M. Welling, Proceedings of Machine Learning Research **5**, PMLR, 2009.

[Gong et al. 2012] D. Gong, X. Zhao, and G. Medioni, "Robust multiple manifolds structure learning", in *Proceedings of the 29th International Conference on Machine Learning* (Edinburgh, 2012), edited by J. Langford and J. Pineau, IMLS, Madison, WI, 2012.

[Kambhatla and Leen 1997] N. Kambhatla and T. K. Leen, "Dimension reduction by local principal component analysis", *Neural Comp.* **9**:7 (1997), 1493–1516.

[Kaslovsky and Meyer 2014] D. N. Kaslovsky and F. G. Meyer, "Non-asymptotic analysis of tangent space perturbation", *Inf. Inference* **3**:2 (2014), 134–187. MR Zbl

[Kushnir et al. 2006] D. Kushnir, M. Galun, and A. Brandt, "Fast multiscale clustering and manifold identification", *Pattern Recognition* **39**:10 (2006), 1876–1891.

[Little et al. 2017] A. V. Little, M. Maggioni, and L. Rosasco, "Multiscale geometric methods for data sets, I: Multiscale SVD, noise and curvature", *Appl. Comput. Harmon. Anal.* **43**:3 (2017), 504–567. MR Zbl

[Liu et al. 2018] G.-R. Liu, Y.-L. Lo, Y.-C. Sheu, and H.-T. Wu, "Diffuse to fuse EEG spectra: intrinsic geometry of sleep dynamics for classification", preprint, 2018. arXiv

[Mishne et al. 2015] G. Mishne, R. Talmon, and I. Cohen, "Graph-based supervised automatic target detection", *IEEE Trans. Geosci. Remote Sensing* **53**:5 (2015), 2738–2754.

[Pedagadi et al. 2013]   S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local fisher discriminant analysis for pedestrian re-identification", pp. 3318–3325 in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Portland, OR, 2013), IEEE, Piscataway, NJ, 2013.

[Qiu and Hancock 2007]   H. J. Qiu and E. R. Hancock, "Clustering and embedding using commute times", *IEEE Trans. Patt. Anal. Mach. Intell.* **29**:11 (2007), 1873–1890. Zbl

[Roweis and Saul 2000]   S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding", *Science* **290**:5500 (2000), 2323–2326.

[Salhov et al. 2012]   M. Salhov, G. Wolf, and A. Averbuch, "Patch-to-tensor embedding", *Appl. Comput. Harmon. Anal.* **33**:2 (2012), 182–203. MR Zbl

[Shemesh et al. 2017]   A. Shemesh, R. Talmon, O. Karp, I. Amir, M. Bar, and Y. J. Grobman, "Affective response to architecture: investigating human reaction to spaces with different geometry", *Architectural Sci. Rev.* **60**:2 (2017), 116–125.

[Singer and Coifman 2008]   A. Singer and R. R. Coifman, "Non-linear independent component analysis with diffusion maps", *Appl. Comput. Harmon. Anal.* **25**:2 (2008), 226–239. MR Zbl

[Singer and Wu 2012]   A. Singer and H.-T. Wu, "Vector diffusion maps and the connection Laplacian", *Comm. Pure Appl. Math.* **65**:8 (2012), 1067–1144. MR Zbl

[Smolyanov et al. 2007]   O. G. Smolyanov, H. v. Weizsäcker, and O. Wittich, "Chernoff's theorem and discrete time approximations of Brownian motion on manifolds", *Potential Anal.* **26**:1 (2007), 1–29. MR Zbl

[Talmon and Coifman 2012]   R. Talmon and R. R. Coifman, "Empirical intrinsic modeling of signals and information geometry", research report YALEU/DCS/TR-1467, 2012, available at http://www.cs.yale.edu/publications/techreports/tr1467.pdf.

[Talmon and Coifman 2013]   R. Talmon and R. R. Coifman, "Empirical intrinsic modeling of signals and information geometry", *Proc. Natl. Acad. Sci. USA* **110**:31 (2013), 12535–12540.

[Tenenbaum et al. 2000]   J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction", *Science* **290**:5500 (2000), 2319–2323.

[Tyagi et al. 2013]   H. Tyagi, E. Vural, and P. Frossard, "Tangent space estimation for smooth embeddings of Riemannian manifolds", *Inf. Inference* **2**:1 (2013), 69–114. MR Zbl

[Wu and Wu 2018]   H.-T. Wu and N. Wu, "Think globally, fit locally under the manifold setup: asymptotic analysis of locally linear embedding", *Ann. Statist.* **46**:6B (2018), 3805–3837. MR Zbl

[Wu et al. 2015]   H.-T. Wu, R. Talmon, and Y.-L. Lo, "Assess sleep stage by modern signal processing techniques", *IEEE Trans. Biomed. Eng.* **62**:4 (2015), 1159–1168.

[Yair and Talmon 2017]   O. Yair and R. Talmon, "Local canonical correlation analysis for nonlinear common variables discovery", *IEEE Trans. Signal Process.* **65**:5 (2017), 1101–1115. MR Zbl

[Zhang and Zha 2004]   Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment", *SIAM J. Sci. Comput.* **26**:1 (2004), 313–338. MR Zbl

JOHN MALIK: john.malik@duke.edu
Department of Mathematics, Duke University, Durham, NC, United States

CHAO SHEN: chao.shen@duke.edu
Department of Mathematics, Duke University, Durham, NC, United States

HAU-TIENG WU: hauwu@math.duke.edu
Department of Mathematics and Department of Statistical Science, Duke University, Durham, NC, United States
and
Mathematics Division, National Center for Theoretical Sciences, Taipei, Taiwan

NAN WU: nan.wu@duke.edu
Department of Mathematics, Duke University, Durham, NC, United States

PURE and APPLIED ANALYSIS

msp.org/paa

Cover image: The figure shows the outgoing scattered field produced by scattering a plane wave, coming from the northwest, off of the (stylized) letters P A A. The total field satisfies the homogeneous Dirichlet condition on the boundary of the letters. It is based on a numerical computation by Mike O'Neil of the Courant Institute.

See inside back cover or msp.org/paa for submission instructions.

PURE and APPLIED ANALYSIS

vol. 1   no. 4   2019