Abstract

Differential Geometry Approach for Unsupervised Machine Learning Algorithms

Nan Wu
Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto
2018

Since its introduction in 2000, locally linear embedding (LLE) algorithm has been widely applied in data science. In this thesis, we provide an asymptotical analysis of LLE under the manifold setup. First, by study the regularized barycentric problem, we derive the corresponding kernel function of LLE. Second, we show that when the point cloud is sampled from a general closed manifold, asymptotically LLE algorithm does not always recover the Laplace-Beltrami operator, and the result may depend on the non-uniform sampling. We demonstrate that a careful choosing of the regularization is necessary to ensure the recovery of the Laplace-Beltrami operator. A comparison with the other commonly applied nonlinear algorithms, particularly the diffusion map, is provided. Moreover, we discuss the relationship between two common nearest neighbor search schemes and the relationship of LLE with the locally linear regression. At last, we consider the case when the point cloud is sampled from a manifold with boundary. We show that if the regularization is chosen correctly, LLE algorithm asymptotically recovers a linear second order differential operator with “free” boundary condition. Such operator coincides with Laplace-Beltrami operator in the interior of the manifold. We further modify LLE algorithm to the Dirichlet Graph Laplacian algorithm which can be used to recover the Laplace-Beltrami operator of the manifold with Dirichlet boundary condition.
Acknowledgements

First and foremost, I would like to express my deep gratitude to my advisors Professor Alex Nabutovsky and Professor Hau-tieng Wu for their valuable supervision of my work and countless hours of discussing mathematical problems with me. I greatly benefited from their knowledge and deep understanding of mathematics which they shared with me and from their invaluable advice on the development of my academic career. In my long journey to get a Ph.D. degree, it was their kind encouragement and support that helped me to overcome my difficulties and frustrations.

I am grateful to Professor Regina Rotman, my Ph.D. Committee member, who enriched my knowledge of Riemannian geometry and topology and provided thoughtful and patient guidance over the years. I also want to thank Professor Almut Burchard for teaching me analysis and her helpful advice.

I would like to thank my collaborator and friend Zhifei Zhu for our enlightening discussions and enjoyable time spent on the Riemannian geometry projects.

I also thank the staff of the Department of Mathematics at our university. Especially I want to thank the help of Jemima Merisca. And I would also like to thank the assistance from the late Ida Bulat.

Last, but not the least, I am grateful to my dear parents for their unconditional love and support without which I would have never been able to pursue a Ph.D. in mathematics.
# Contents

1 Introduction ................................................................. 1
   1.1 Notations ............................................................ 3
   1.2 LLE algorithm ....................................................... 3
   1.3 Barycentric coordinate on point cloud ......................... 4
   1.4 Perturbation analysis ............................................. 6

2 LLE on closed manifolds ................................................. 14
   2.1 Manifold Setup ..................................................... 14
   2.2 Some preliminary lemmas in Riemmanian geometry ............. 15
   2.3 Local covariance structure and local PCA ................. 24
   2.4 Integral kernel of LLE and variance analysis on closed manifolds 33
      2.4.1 Proof of Theorem 2.4.1 .................................... 36
   2.5 Biased analysis on closed manifolds .......................... 48
      2.5.1 Proof of Theorem 2.5.1 .................................... 51
   2.6 Conclusion of the chapter: convergence of LLE on closed manifolds 64
   2.7 ε-radius neighborhood v.s. K nearest neighborhood ........... 65
   2.8 LLE v.s. LLR ....................................................... 66
   2.9 Error in variable .................................................. 67
   2.10 Numerical examples ............................................... 68
      2.10.1 Sphere ........................................................ 68
      2.10.2 Examine the kernel ........................................ 72
      2.10.3 Two-dimensional random tomography example .......... 74

3 LLE on manifolds with boundary ...................................... 77
   3.1 Setup on manifolds with boundary and preliminary lemmas .... 77
   3.2 Variance analysis and biased analysis on manifolds with boundary, when ρ = 3 84
      3.2.1 Conclusion when ρ = 3 ...................................... 92
      3.2.2 Proof of Theorem 3.2.2 .................................... 93
      3.2.3 Proof of Theorem 3.2.1 .................................... 96
   3.3 Dirichlet Graph Laplacian ........................................ 105

Bibliography ................................................................. 109
Chapter 1

Introduction

Dimension reduction is a fundamental step in data analysis. In past decades, due to the high demand for analyzing the large scale, massive and complicated datasets accompanying technological advances, there have been many efforts to solve this problem from different angles. The resulting algorithms can be roughly classified into two types, linear and nonlinear. Linear methods include principal component analysis (PCA), multidimensional scaling, and others. Nonlinear methods include ISOMAP [65, 3], eigenmap [6], maximal variance unfolding [72], t-distributed stochastic neighbor embedding [69], commute time embedding [48], Patch-to-tensor embedding [51], locally linear embedding (LLE) [49, 1] and its variations such as, Hessian LLE [22], modified LLE [74], robust LLE [15] and weighted LLE [46], diffusion map (DM) [18] and its variations such as, local tangent space alignment [75], vector diffusion map [55, 57], horizontal diffusion map [27], ODM(orientable diffusion map) [54], magnetic diffusion map [26, 19], alternating diffusion [39, 64, 38, 37], multiview diffusion map [40], time coupled diffusion maps [42] and empirical intrinsic geometry (EIG) [53, 62, 63].

The subject of this thesis, LLE, was published in Science in 2000 [49]. It has been widely used and has been cited more than 10,000 times. The algorithm is designed to be intuitive and simple. It naturally merges two ideas “fit locally” and “think globally” together. Given a point cloud, first, for each data point, we determine its nearest neighbors, and catch the local geometric structure of the dataset through finding the barycenter coordinates for those neighboring points by using a regularization. The barycentric coordinates generalize the following notion: they are an assignment of weights to the points in the neighborhood so that under such assignment the center of the neighborhood is the center of mass. This is the “fit locally” part of LLE. Second, by extending the barycenter coordinates into LLE matrix, the eigenvectors the matrix are evaluated as coordinate maps to reduce the dimension of the point cloud. This is the “think globally” part of LLE. However, unlike the fruitful theoretical results from discussing eigenmap and the diffusion-based approach like DM [6, 44, 7, 32, 52, 30, 18, 58, 8, 70, 71, 28, 57, 24, 66, 55, 11, 10, 35, 47, 4, 14, 50, 13, 9, 31, 61, 33, 59], a systematic analysis of LLE algorithm has not been undertaken, except an ad hoc argument shown in [6] based on some conditions.

In this thesis, we work under the assumption that the point cloud is (non-) uniformly sampled from a low dimensional manifold isometrically embedded in the Euclidean space. Under such manifold setup, the “think globally” part of LLE algorithm could be understood from the spectral geometry viewpoint [10, 11], where they prove that the orthonormal eigenfunctions of the Laplace-Beltrami operator of a closed Riemannian manifold can be used as coordinate functions to embed the manifold into a Hilbert space with the Euclidean inner product. Hence, one should expect that LLE matrix approximates the Laplace-Beltrami operator and the eigenvectors of
LLE matrix approximate the eigenfunctions of the Laplace-Beltrami operator, so that if one uses the eigenvectors as coordinate maps to reduce the dimension of the point cloud, the topological structure of the underlying manifold is preserved.

One of the major contributions in this thesis is an asymptotic pointwise convergence analysis of LLE under the manifold (may have boundary) setup. Such analysis is achieved in the following steps. First, for the point cloud sampled from a manifold (may have boundary), we solve the regularized barycentric problem in the “fit locally” part of LLE algorithm. From the barycentric coordinate estimation, we establish the kernel function of LLE which depends on the regularization. Second, we prove that LLE matrix asymptotically converges to an integral operator associated with the kernel function almost surely. Third, although it is widely believed that under the closed manifold setup, asymptotically LLE matrix should lead to the Laplace-Beltrami operator, we show that this might not always be the case. Specifically, we show that the behavior of the integral operator relies on the regularization. If the regularization is chosen properly, the integral operator approximates the Laplace-Beltrami operator, even if the sampling is non-uniform. If the regularization is not chosen properly, the acquired information will be contaminated by the extrinsic information (the second fundamental form of the underlying manifold). At last, under the proper choice of the regularization, if the manifold has boundary, we show that the integral operator approximates a second order linear differential operator on the manifold with “free” boundary condition. The differential operator coincides with the Laplace-Beltrami operator in the interior of the manifold.

Given a point cloud that sampled from a manifold with boundary, it is an important problem in scientific computation that how to recover the Laplace-Beltrami operator of the underlying manifold with the Dirichlet boundary condition. For example, the paper [29] in studying molecular dynamics provides motivation to solve such problem. Based on the analysis of LLE algorithm on manifold with boundary, we propose an algorithm coined with the name Dirichlet Graph Laplacian Algorithm. Without boundary detection, the algorithm constructs a diagonal matrix which approximates the bump function concentrated on the boundary of the manifold. The Dirichlet Graph Laplacian is constructed by adding the diagonal matrix onto the LLE matrix with the proposed regularization. If the manifold is closed, then the Dirichlet Graph Laplacian recovers the Laplace-Beltrami operator. However, if the manifold has a boundary, then the diagonal matrix will force a Dirichlet boundary condition. Specifically, we show that the Dirichlet Graph Laplacian converges to an integral operator almost surely and the integral operator converges uniformly over finite dimensional eigensubspaces of the Laplace-Beltrami operator with the Dirichlet boundary condition.

The contents of this thesis can be summarized as follows. In Chapter 1, we review LLE algorithm and the barycentric coordinates on the point cloud. We introduce the regularization which stabilizes the barycentric problem and the solution to the regularized barycentric problem. At last, we provide a perturbation analysis to a special type of analytic symmetric matrix, which is used in the asymptotic analysis in later chapters. In Chapter 2, we establish an asymptotic analysis of LLE on closed manifold setup. Moreover, we have a direct comparison of LLE and other relevant nonlinear machine learning algorithms, for example, the eigenmap and DM. The relationship between two common nearest neighbor search schemes is discussed and we link LLE back to the widely applied kernel regression technique, the locally linear regression (LLR) and error in variable problem. In the end, we provide numerical simulations to support our theoretical findings. This chapter is based on the joint work with Hau-tieng Wu [73]. In chapter 3, we establish an asymptotic analysis of LLE on closed manifold with boundary setup. At last, we introduce the Dirichlet Graph Laplacian (DGL) algorithm by modifying the original LLE algorithm. This chapter is based a joint work with Hau-tieng Wu.
1.1 Notations

We fix the notations used in this thesis. For $d \in \mathbb{N}$, $I_{d \times d}$ means the identity matrix of size $d \times d$. For $n \in \mathbb{N}$, denote $I_n$ to be the $n$-dim vector with all entries 1. For $\varepsilon \geq 0$, denote $B_{\mathbb{R}^p}^\varepsilon(x) := \{ y \in \mathbb{R}^p | \| x - y \|_{\mathbb{R}^p} \leq \varepsilon \}$. Denote $e_i = [0, \cdots, 1, \cdots 0]^\top \in \mathbb{R}^p$ to be the unit $p$-dim vector with 1 in the i-th entry. For $p, r \in \mathbb{N}$ so that $r \leq p$, denote $J_{p,r} \in \mathbb{R}^{p \times r}$ so that the $(i,i)$ entry is 1 for $i = 1, \cdots, r$, and zeros elsewhere and denote $J^\top_{p,r} \in \mathbb{R}^{p \times r}$ so that the $(i,i)$-th entry is 1 for $i = 1, \cdots, r$, and zeros elsewhere. $I_{p,r} := J_{p,r}J^\top_{p,r}$ is a $p \times p$ matrix so that the $(i,i)$-th entry is 1 for $i = 1, \cdots, r$ and 0 elsewhere; and $I^\top_{p,r} := J^\top_{p,r}J_{p,r}$ is a $p \times p$ matrix so that the $(i,i)$-th entry is 1 for $i = p - r + 1, \cdots, p$ and 0 elsewhere. Denote $S(p)$ to be the set of real symmetric matrix of size $p \times p$, $O(p)$ to be the orthogonal group in dimension $p$, and $\mathfrak{s}(p)$ to be the set of anti-symmetric matrix of size $p \times p$. For $M \in \mathbb{R}^{p \times p}$, denote $M^\top$ to be the transpose of $M$ and $M^*$ to be the Moore-Penrose pseudo-inverse of $M$. For $a, b \in \mathbb{R}$, we use $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$ to simplify the notation. We summarize the commonly used notations for the asymptotical analysis in Table 1.1 for the convenience of the readers.

1.2 LLE algorithm

We start by summarizing LLE algorithm. Suppose $\mathcal{X} = \{z_i\}_{i=1}^p \subset \mathbb{R}^P$ is the provided dataset, or the point cloud.

1. Fix $\varepsilon > 0$. For each $z_k \in \mathcal{X}$, denote $\mathcal{N}_k := B_{\mathbb{R}^p}^\varepsilon(z_k) \cap (\mathcal{X} \setminus \{z_k\}) = \{z_{kj}\}_{j=1}^{n_k}$, where $n_k \in \mathbb{N}$ is the number of points in $\mathcal{N}_k$. $\mathcal{N}_k$ is called the $\varepsilon$-radius neighborhood of $z_k$. Alternatively, we can also fix a number $K$, and choose the $K$ nearest points of $z_k$. This is called the $K$ nearest neighbors (KNN) scheme.

2. For each $z_k \in \mathcal{X}$, find its barycentric coordinates associated with $\mathcal{N}_k$ by

$$w_{zk} = \arg \min \sum_{j=1}^{n_k} w_{zk,j} \| z_k - \sum_{j=1}^{n_k} w_{zk,j} z_{kj} \|^2 \in \mathbb{R}^{n_k}. \tag{1.2.1}$$

Notice that $w_{zk}$ satisfies $w_{zk}^\top 1_{n_k} = \sum_{j=1}^{n_k} w_{zk}(j) = 1$. 

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>Dimension of the ambient space</td>
</tr>
<tr>
<td>$d$</td>
<td>Dimension of the low-dimensional Riemannian manifold</td>
</tr>
<tr>
<td>$(M,g)$</td>
<td>$d$-dimensional smooth Riemannian manifold</td>
</tr>
<tr>
<td>$dV$</td>
<td>Riemannian volume form of $(M,g)$</td>
</tr>
<tr>
<td>$</td>
<td>S^{d-1}</td>
</tr>
<tr>
<td>$\exp_x$</td>
<td>Exponential map at $x$</td>
</tr>
<tr>
<td>$T_x M$</td>
<td>Tangent space of $M$ at $x$</td>
</tr>
<tr>
<td>$\text{Ric}_x$</td>
<td>Ricci curvature tensor of $(M,g)$ at $x$</td>
</tr>
<tr>
<td>$i, t, s$</td>
<td>Isometric embedding of $M$ into $\mathbb{R}^P$ and its differential</td>
</tr>
<tr>
<td>$\Pi_x$</td>
<td>Second fundamental form of the embedding $t$ at $x$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Probability density function on $t(M)$</td>
</tr>
<tr>
<td>$n \in \mathbb{N}$</td>
<td>Number of data points sampled from $M$</td>
</tr>
<tr>
<td>$\mathcal{X} = {z_i}_{i=1}^p$</td>
<td>Point cloud sampled from $t(M) \subset \mathbb{R}^P$</td>
</tr>
<tr>
<td>$w_{zk} \in \mathbb{R}^{n_k}$</td>
<td>Barycentric coordinates of $z_k$ with respect to data points in the $\varepsilon$-neighborhood</td>
</tr>
</tbody>
</table>
3. Define a \( n \times n \) matrix \( W \), called the **LLE matrix**, by

\[
W_{kj} = \begin{cases} 
    w_{kj}(j) & \text{if } z_l = z_{k,j} \in \mathcal{N}_z; \\
    0 & \text{otherwise.}
\end{cases}
\]  

(1.2.2)

4. To reduce the dimension of \( \mathcal{X} \), it is suggested in [49] to embed \( \mathcal{X} \) into a low dimension Euclidean space

\[
z_k \mapsto y_k = [v_1(k), \ldots, v_{\ell}(k)]^\top \in \mathbb{R}^\ell,
\]  

(1.2.3)

for each \( z_k \in \mathcal{X} \), where \( \ell \) is the dimension of the embedded points chosen by the user, and \( v_1, \ldots, v_{\ell} \in \mathbb{R}^n \) are eigenvectors of \((I - W)^\top(I - W)\) corresponding to the \( \ell \) smallest eigenvalues. Note that this is equivalent to minimizing the cost function \( \sum_{k=1}^p \| y_k - \sum_{l=1}^n W_{kj} y_l \|^2 \).

Although the algorithm looks relatively simple, there are actually several details that should be discussed prior to the asymptotical analysis. To simplify the discussion, we focus on one point \( z_k \in \mathcal{X} \) and assume that there are \( N \) data points in \( \mathcal{N}_{z_k} = \{z_{k,1}, \ldots, z_{k,N}\} \). To find the barycentric coordinates of \( z_k \), we define the **local data matrix** associated with \( \mathcal{N}_{z_k} \):

\[
G_n := \begin{bmatrix} 
    z_{k,1} - z_k & \cdots & z_{k,N} - z_k \\
\end{bmatrix} \in \mathbb{R}^{p \times N}.
\]  

(1.2.4)

It is important to note that \( G_n \) depends not only on \( n \), but also \( c \) and \( z_k \). However, we only keep \( n \) to make the notation easier. The other notations in this section are simplified in the same way. Minimizing (1.2.1) is equivalent to minimizing the functional \( w^\top G_n^\top G_n w \) over \( w \in \mathbb{R}^N \) under the constraint \( w^\top 1_N = 1 \). Here, \( G_n^\top G_n \) is the Gramian matrix associated with the dataset \( \{z_{k,1} - z_k, \ldots, z_{k,N} - z_k\} \). In general, \( G_n^\top G_n \) might be singular, and it is suggested in [49] to stabilize the algorithm by regularizing the equation by

\[
(G_n^\top G_n + c I_{N \times N}) y = 1_N,
\]  

(1.2.5)

where \( c > 0 \) is the regularizer chosen by the user. For example, in [49], \( c \) is suggested to be \( \frac{\delta}{\|G_n\|_F} \), where \( 0 < \delta < \|G_n\|_F \) is chosen by the user and \( \|G_n\|_F \) is the Frobenius norm of \( G_n \). It has been observed that LLE is sensitive to the choice of the regularizer (see, for example, [74]). We will later quantify this dependence under the manifold setup. Using the Lagrange multiplier method, the minimizer is

\[
w_n = \frac{y_n}{y_n^\top 1_N},
\]  

(1.2.6)

where \( y_n \) is the solution of (1.2.5). We are going to find \( w_n \) explicitly in next section.

### 1.3 Barycentric coordinate on point cloud

In this section, we explicitly express \( w_n \), which is the essential step toward the asymptotical analysis. Suppose \( \text{rank}(G_n^\top G_n) = r_n \). Note that \( r_n = \text{rank}(G_n G_n^\top) = \text{rank}(G_n) \leq p \), so \( G_n^\top G_n \) is singular when \( p < N \). Moreover, \( G_n^\top G_n \) is positive semidefinite. Denote the eigen-decomposition of \( G_n^\top G_n \) as \( V_n \Lambda_n V_n^\top \), where

\[
\Lambda_n = \text{diag}(\lambda_{n,1}, \lambda_{n,2}, \ldots, \lambda_{n,N}),
\]  

(1.3.1)
\[ \lambda_{n,1} \geq \lambda_{n,2} \geq \cdots \geq \lambda_{n,r_n} > \lambda_{n,r_n+1} = \cdots = \lambda_{n,N} = 0, \]  

and 

\[
V_n = \begin{bmatrix} v_{n,1} & \ldots & v_{n,N} \end{bmatrix} \in O(N).
\]  

(1.3.2)

Clearly, \( \{v_{n,i}\}_{i=r_n+1}^{N} \) form an orthonormal basis of the null space of \( \text{null}(G_n^T G_n) \), which is equivalent to \( \text{null}(G_n) \). Then (1.2.5) is equivalent to solving 

\[
V_n(\Lambda_n + cI_{N \times N})V_n^T y = 1_N,
\]

and the solution is 

\[
y_n = V_n(\Lambda_n + cI_{N \times N})^{-1}V_n^T 1_N
\]

\[
= c^{-1}1_N + V_n[(\Lambda_n + cI_{N \times N})^{-1} - c^{-1}I_{N \times N}]V_n^T 1_N.
\]

(1.3.4)

Therefore, 

\[
w_n^T = \frac{1_N + \frac{1}{c}V_n[V_n[(\Lambda_n + cI_{N \times N})^{-1} - c^{-1}I_{N \times N}]V_n^T]}{N + \frac{1}{c}V_n[V_n[(\Lambda_n + cI_{N \times N})^{-1} - c^{-1}I_{N \times N}]V_n^T]1_N}.
\]

(1.3.5)

Without recasting (1.3.5) into a proper form, it is not clear how to capture the geometric information contained in (1.3.5). Observe that while \( G_n^T G_n \) is the Gramian matrix, \( G_n^T G_n \) is related to the sample covariance matrix associated with \( \mathcal{N}_d \). We call \( \frac{1}{n}G_nG_n^T \) the local sample covariance matrix.\(^1\) Clearly, \( r_n \leq p \) and \( G_n^T G_n \) and \( G_n^T G_n \) share the same positive eigenvalues, \( \lambda_{n,1} \cdots \lambda_{n,r_n} \). Denote the eigen-decomposition of \( G_n^T G_n \) as \( U_n\Lambda_n U_n^T \), where \( U_n \in O(p) \) and \( \Lambda_n \) is a \( p \times p \) diagonal matrix. By a direct calculation, the first \( r_n \) columns of \( V_n \) are related to \( U_n \) by 

\[
V_n|_{N \times r_n} = G_n^T U_n(\Lambda_n^1)^{1/2}J_{p \times r_n},
\]

(1.3.6)

where \( V_n = [V_n|_{N \times r_n}V_n|_{N \times N-r_n}] \). Since \( (\Lambda_n + cI_{N \times N})^{-1} - c^{-1}I_{N \times N} \) has only \( r_n \) non-zero diagonal entries, based on (1.3.4), we have 

\[
y_n^T = c^{-1}1_N + \frac{1}{c}V_n]\left[(\Lambda_n + cI_{N \times N})^{-1} - c^{-1}I_{N \times N}\right]V_n^T
\]

\[
= c^{-1}1_N + \frac{1}{c}G_n^T U_n(\Lambda_n^1)^{1/2}J_{p \times r_n}^T (\Lambda_n + cI_{p \times p})^{-1} - c^{-1}I_{p \times p}J_{p \times r_n}^T U_n^T 1_N.
\]

Note that we have 

\[
U_n(\Lambda_n^1)^{1/2}J_{p \times r_n}^T (\Lambda_n + cI_{p \times p})^{-1} - c^{-1}I_{p \times p}J_{p \times r_n}^T U_n^T
\]

\[
= -c^{-1}U_nJ_{p \times r_n}^T (\Lambda_n + cI_{p \times p})^{-1}J_{p \times r_n}^T U_n^T,
\]

(1.3.7)

which could be understood as a “regularized pseudo-inverse”. Specifically, when \( c \) is small, we have 

\[
U_nJ_{p \times r_n}^T (\Lambda_n + cI_{p \times p})^{-1}J_{p \times r_n}^T U_n^T \approx (G_n^T G_n)^+.
\]

(1.3.8)

Denote 

\[
\mathcal{F}_c(G_n^T G_n) := U_nJ_{p \times r_n}^T (\Lambda_n + cI_{p \times p})^{-1}J_{p \times r_n}^T U_n^T.
\]

(1.3.9)

\(^1\)The usual sample covariance matrix associated with \( \mathcal{N}_d \) is defined as \( \frac{1}{n} \sum_{j=1}^{N}(z_{k,j} - \mu_k)(z_{k,j} - \mu_k)^T \), where \( \mu_k = \frac{1}{n} \sum_{j=1}^{N} z_{k,j} \).
Hence, we can recast (1.3.4) and (1.3.5) into
\[ y_n^T = c^{-1}1_N^T - c^{-1}1_N^T G_n^T \mathcal{F}_c(G_n G_n^T) G_n \] (1.3.10)
and
\[ w_n^T = \frac{1_N^T - 1_N^T G_n^T \mathcal{F}_c(G_n G_n^T) G_n}{N - 1_N G_n \mathcal{F}_c(G_n G_n^T) G_n 1_N} = \frac{1_N^T - T_{n,z_k} G_n}{N - T_{n,z_k} G_n 1_N}, \] (1.3.11)
where
\[ T_{n,z_k} := \mathcal{F}_c(G_n G_n^T) G_n 1_N \] (1.3.12)
is chosen in order to have a better geometric insight into LLE algorithm. We now summarize the expansion of the barycentric coordinate.

**Proposition 1.3.1.** Take a data set \( \mathcal{X} = \{ z_i \}_{i=1}^n \subset \mathbb{R}^p \). Suppose there are \( N \) data points in the \( \varepsilon \) neighborhood of \( z_k \), namely \( \{ z_k, 1 \leq k \leq N \} \subset B_\varepsilon(z_k) \cap (\mathcal{X} \setminus \{ z_k \}) \). Assume \( p < N \). Let \( G_n^T G_n \) be the Gramian matrix associated with \( \{ z_k, 1 \leq k \leq N \} \) and let \( \{ \lambda_i \}_{i=1}^r \) and \( \{ u_{n,i} \}_{i=1}^r \), where \( r \leq p \) is the rank of \( G_n^T G_n \), be the nonzero eigenvalues and the corresponding orthonormal eigenvectors of \( G_n G_n^T \) satisfying (1.3.6). With \( T_{n,z_k} \) defined in (1.3.12), the barycentric coordinates of \( z_k \) coming from the regularized equation (1.2.5) is
\[ w_n^T = \frac{1_N^T - T_{n,z_k} G_n}{N - T_{n,z_k} G_n 1_N}. \] (1.3.13)

**Remark 1.3.1.** The denominator \( N - T_{n,z_k} G_n 1_N \) is the sum of all entries of the numerator \( 1_N^T - T_{n,z_k} G_n \). We could thus view LLE matrix defined in (1.2.2) as a “normalized kernel” defined on the point cloud. However, we mention that while all entries of \( w_n \) are summed to 1, the vector \( 1_N^T - T_{n,z_k} G_n \) might have negative entries, depending on the vector \( T_{n,z_k}^T \). Therefore, in general, \( W \) cannot be understood as a transition matrix.

### 1.4 Perturbation analysis

In this section, we introduce a perturbation analysis on a special type of symmetric matrix. Such analysis is going to be used in the proof of the main theorem.

Suppose \( A : \mathbb{R} \to S(p) \), where \( S(p) \) is the set of real symmetric \( p \times p \) matrices, is an analytic function around 0. In this appendix, we are going to introduce an algorithm to calculate the eigenvalues and orthonormal eigenvectors of \( A(\varepsilon) \) when \( \varepsilon \) is small enough. The method introduced in this appendix follows the standard approach, like [2, 68]. For discussion of more general matrices, interested readers are referred to [68].

Suppose
\[ A(0) = \begin{bmatrix} \lambda I_{d \times d} & 0 \\ 0 & 0 \end{bmatrix}, \]
where \( 0 < d < p \) and \( \lambda \neq 0 \). Decompose \( A(0) \) by
\[ A(0)X(0) = X(0)A(0), \] (1.4.1)
where \( \Lambda(0) = A(0) \) is a diagonal matrix consisting of eigenvalues of \( A(0) \), and

\[
X(0) = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \in O(p),
\]

where \( X_1 \in O(d) \) and \( X_2 \in O(p-d) \). Note that due to the possible nontrivial multiplicity of eigenvalues, \( X(0) \) may not be uniquely determined. Take the Taylor expansion of \( A \) around 0 as

\[
A(\varepsilon) = A(0) + A'(0)\varepsilon + \frac{1}{2}A''(0)\varepsilon^2 + O(\varepsilon^3),
\]

where \( \varepsilon > 0 \) is sufficiently small, \( A'(0) \) and \( A''(0) \) are divided into blocks of the same size as those of \( A(0) \) by

\[
A'(0) = \begin{bmatrix} A'_{11} & A'_{12} \\ A'_{21} & A'_{22} \end{bmatrix}, \quad A''(0) = \begin{bmatrix} A''_{11} & A''_{12} \\ A''_{21} & A''_{22} \end{bmatrix},
\]

where \( A'_{11} \in S(d), A'_{22} \in S(p-d), A''_{11} \in S(d) \) and \( A''_{22} \in S(p-d) \). Let \( \Lambda(\varepsilon) \in \mathbb{R}^{p \times p} \) be the diagonal matrix consisting of eigenvalues of \( A(\varepsilon) \) and \( X(\varepsilon) \in \mathbb{R}^{p \times p} \) be the matrix formed by the corresponding eigenvectors, i.e.

\[
A(\varepsilon)X(\varepsilon) = X(\varepsilon)\Lambda(\varepsilon). \tag{1.4.2}
\]

Since \( A \) is symmetric, \( X \) and \( \Lambda \) are both analytic around 0 based on [5, Section 3.6.2, Theorem 1]. We thus have the following Taylor expansion when \( \varepsilon \) is sufficiently small:

\[
\Lambda(\varepsilon) = \Lambda(0) + \varepsilon A'(0) + \frac{1}{2}\varepsilon^2 A''(0) + O(\varepsilon^3),
\]

\[
X(\varepsilon) = X(0) + X'(0)\varepsilon + O(\varepsilon^2).
\]

Here \( \Lambda(0), A'(0) \) and \( A''(0) \) are all diagonal matrices and columns of \( X(\varepsilon) \) form an orthogonal set. Note that if we normalize \( X(\varepsilon) \) to be in \( O(p) \), then by the fact that the Lie algebra of \( O(p) \) is the set of anti-symmetric matrices, we know that \( X(0)^{-1}X'(0) \) is an anti-symmetric matrix. We discuss the eigendecomposition of \( \Lambda(\varepsilon) \) under two different setups, depending on the multiplicity of eigenvalues.

**When there is no repeated eigenvalue in both** \( A'_{11} \) **and** \( A'_{22} \). In the first case, we assume that the eigenvalues of \( A'_{11} \) are distinct and the eigenvalues of \( A'_{22} \) are distinct (but the eigenvalues of \( A'_{11} \) and the eigenvalues of \( A'_{22} \) could overlap). To get \( \Lambda(\varepsilon) \) up to the first order, we need to solve \( \Lambda'(0) \). To determined \( \Lambda'(0) \), we check the first order derivative of \( A(\varepsilon) \) at \( \varepsilon = 0 \). Differentiate (1.4.2) and we get

\[
A'(0)X(0) + A(0)X'(0) = X'(0)\Lambda(0) + X(0)A'(0).
\]

Denote

\[
A'(0) = \begin{bmatrix} A'_{11} & 0 \\ 0 & A'_{22} \end{bmatrix}
\]

and set

\[
X'(0) = X(0)C,
\]
If we substitute $X(0), X'(0)$, and $\Lambda'(0)$ into (1.4.3), we have the following linear equations by comparing blocks:

\begin{align}
A'_{11}X_1 &= X_1\Lambda'_{11}, \\
A'_{22}X_2 &= X_2\Lambda'_{22}, \\
A'_{12}X_2 &= -\lambda X_1C_{12}, \\
A'_{21}X_1 &= \lambda X_2C_{21},
\end{align}

where

\[
C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \in \mathbb{R}^{p \times p}.
\]

By (1.4.3) and (1.4.4), $A'_{11}$ and $A'_{22}$ are eigenvalue matrices of $A'_{11}$ and $A'_{22}$, and $X_1$ and $X_2$ are the corresponding eigenvector matrices, and we obtain the first order approximation of the eigenvalues. Note that above equations hold without assuming that the eigenvalues of $A'_{11}$ are distinct and the eigenvalues of $A'_{22}$ are distinct. Also note that although we could obtain the first order relationship between the eigenvectors of $A(\epsilon)$ and $A(0)$, without assuming distinct eigenvalues, the eigenvectors may not be unique.

If we want to further get $\Lambda(\epsilon)$ up to the second order and solve $X(\epsilon)$ uniquely up to the first order, we need to solve $A'(0), A''(0), X(0)$, and $X'(0)$. To solve $A'(0), A''(0), X(0)$, and $X'(0)$, we need the assumption that $A'_{11}$ and $A'_{22}$ have no repeated eigenvalues, while we allow eigenvalues of $A'_{11}$ to be same as those of $A'_{22}$. By (1.4.5) and (1.4.6) we have

\begin{align}
C_{12} &= -\lambda^{-1}X_1^TA'_{12}X_2, \\
C_{21} &= \lambda^{-1}X_2^TA'_{21}X_1.
\end{align}

Clearly, since $A'_{11}$ and $A'_{22}$ do not have repeated eigenvalues, $X_1$ and $X_2$ are uniquely defined and $C_{12}$ and $C_{21}$ can be uniquely determined. Since the information of $C_{11}$ and $C_{22}$ are not available from (1.4.3), we need the higher order derivative of $A(\epsilon)$. Differentiate $A(\epsilon)$ twice, we get

\[
A''(0)X(0) + 2A'(0)X'(0) + A(0)X''(0) = X''(0)\Lambda(0) + 2X'(0)\Lambda'(0) + X(0)\Lambda''(0).
\]

Now, we further substitute $X(0)$ and $X'(0) = X(0)C$ into (1.4.9), and get

\begin{align}
A''_{11}X_1 - X_1\Lambda''_{11} &= 2X_1(C_{11}\Lambda'_{11} - \Lambda'_{11}C_{11}) - 2A'_{12}X_2C_{21}, \\
A''_{22}X_2 - X_2\Lambda''_{22} &= 2X_2(C_{22}\Lambda'_{22} - \Lambda'_{22}C_{22}) - 2A'_{21}X_1C_{12}.
\end{align}

Since $X_1 \in O(d)$ and $X_2 \in O(p-d)$, we have

\begin{align}
X_1^T(A''_{11}X_1 + 2A'_{12}X_2C_{21}) &= 2(C_{11}\Lambda'_{11} - \Lambda'_{11}C_{11}) + \Lambda''_{11}, \\
X_2^T(A''_{22}X_2 + 2A'_{21}X_1C_{12}) &= 2(C_{22}\Lambda'_{22} - \Lambda'_{22}C_{22}) + \Lambda''_{22}.
\end{align}

Since that diagonal entries of $C_{11}\Lambda'_{11} - \Lambda'_{11}C_{11}$ and $C_{22}\Lambda'_{22} - \Lambda'_{22}C_{22}$ are zero, and the off-diagonal entries of $\Lambda''_{11}$ and $\Lambda''_{22}$ are zero, the off diagonal entries of $C_{11}$ and $C_{22}$, as well as $\Lambda''_{11}$ and $\Lambda''_{22}$, can be found from (1.4.12). Specially,
where $A'_{11}$ and $A'_{22}$ do not have repeated eigenvalues, we have

\[
(C_{11})_{mn} = \frac{-1}{(A'_{1})_{m,n} - (A'_{1})_{n,n}} e_m^T (X_1^T A''_{11} X_1 + \frac{2}{\Lambda} X_1^T A'_{12} A'_{21} X_1) e_n,
\]

\[
(A''_{11})_{mn} = e_m^T (X_1^T A''_{11} X_1 + \frac{2}{\Lambda} X_1^T A'_{12} A'_{21} X_1) e_n,
\]

where $1 \leq m \neq n \leq d$ and

\[
(C_{22})_{mn} = \frac{-1}{(A'_{2})_{m,n} - (A'_{2})_{n,n}} e_m^T (X_2^T A''_{22} X_2 - \frac{2}{\Lambda} X_2^T A'_{21} A'_{12} X_2) e_n,
\]

\[
(A''_{22})_{mn} = e_m^T (X_2^T A''_{22} X_2 - \frac{2}{\Lambda} X_2^T A'_{21} A'_{12} X_2) e_n,
\]

where $1 \leq m \neq n \leq p - d$. By the above evaluation, we know $C_{i,j} = -C_{j,i}$ for $1 \leq i \neq j \leq p$, and what is left unknown is the diagonal entries of $C$. To determine the diagonal entries of $C$, we normalize $X(\varepsilon) = X(0) + X(0)C\varepsilon + O(\varepsilon^2)$ so that $X(\varepsilon) \in O(p)$. We thus have

\[
I_{p \times p} = (X(0) + X'(0)\varepsilon + O(\varepsilon^2))^T (X(0) + X'(0)\varepsilon + O(\varepsilon^2))
\]

\[
= X(0)^T X(0) + (C^T X(0)^T X(0) + X(0)^T X(0)C)\varepsilon + O(\varepsilon^2)
\]

\[
= I_{p \times p} + 2\varepsilon \text{diag}(C) + O(\varepsilon^2),
\]

(1.4.14)

where the last equality holds since $C_{i,j} = -C_{j,i}$ when $i \neq j$, and $\text{diag}(C)$ is a diagonal matrix so that $\text{diag}(C)_{i,i} = C_{i,i}$ for $i = 1, \ldots, p$. As a result, we know that the diagonal entries of $C$ are of order $\varepsilon$. As a result, we have the following solution to the eigenvalues and eigenvectors of $A(\varepsilon)$:

\[
\Lambda(\varepsilon) = \begin{bmatrix}
\lambda I_{d \times d} + \varepsilon \Lambda_1' + \frac{1}{2} \varepsilon^2 \Lambda_1'' & 0 \\
0 & \varepsilon \Lambda_2' + \frac{1}{2} \varepsilon^2 \Lambda_2''
\end{bmatrix} + O(\varepsilon^3),
\]

(1.4.15)

\[
X(\varepsilon) = X(0)(I_{p \times p} + \varepsilon S) + O(\varepsilon^2) \in O(p),
\]

(1.4.16)

where $S := C - \text{diag}(C)$ and the last equality holds since the entries of $\text{diag}(C)$ are of order $\varepsilon$. Note that $S$ is an anti-symmetric matrix. This result could be understood from the fact that the Lie algebra of $O(p)$ is the set of anti-symmetric matrices, and the tangent vector at $X(0)$ leading to $X(\varepsilon)$ is $X(0)S$.

**When there exists a repeated eigenvalue in $A'_{22}$**. In this case, we assume that $A'_{22}$ may have repeated eigenvalues, and to simplify the discussion, we assume that $A'_{11}$ does not have a repeated eigenvalue. Recall (1.4.3) and (1.4.4). Write

\[
\Lambda_2' = \begin{bmatrix}
\Lambda_2'_{11} & 0 \\
0 & \Lambda_2'_{22}
\end{bmatrix},
\]

where $\Lambda_2'_{22} \in \mathbb{R}^{l \times l}$, $1 \leq l \leq p - d$, is a diagonal matrix with the same diagonal entries, denoted as $\gamma \in \mathbb{R}$. To simplify the discussion, we assume that the diagonal entries of $\Lambda_2'_{22} \in \mathbb{R}^{(p - d - l) \times (p - d - l)}$ are all distinct and are different from $\gamma$. Hence, we have

\[
\Lambda'(0) = \begin{bmatrix}
\Lambda_1' & 0 & 0 \\
0 & \Lambda_2'_{11} & 0 \\
0 & 0 & \Lambda_2'_{22}
\end{bmatrix}.
\]

Let $\Gamma_1 \in O(d)$ be the orthonormal eigenvector matrix of $A'_{11}$ and $\Gamma_2 \in O(p - d)$ be any orthonormal eigenvector.
Consider

$$\tilde{A}(\varepsilon) = \Gamma^{-1}A(\varepsilon)\Gamma.$$  \hfill (1.4.17)

Note that $A(\varepsilon)$ has the same eigenvalue matrix as $\tilde{A}(\varepsilon)$. By a direct expansion,

$$\tilde{A}(\varepsilon) = \Gamma^{-1}A(\varepsilon)\Gamma$$

$$= \Gamma^{-1}A(0)\Gamma + \Gamma^{-1}A'(0)\Gamma\varepsilon + \frac{1}{2}\Gamma^{-1}A''(0)\Gamma\varepsilon^2 + O(\varepsilon^3)$$

$$= \tilde{A}(0) + \tilde{A}'(0)\varepsilon + \frac{1}{2}\tilde{A}''(0)\varepsilon^2 + O(\varepsilon^3),$$

where $\tilde{A}(0) := \Gamma^{-1}A(0)\Gamma$, $\tilde{A}'(0) := \Gamma^{-1}A'(0)\Gamma$, and $\tilde{A}''(0) := \Gamma^{-1}A''(0)\Gamma$. By the assumption of $A(0)$, we have

$$\tilde{A}(0) = \Gamma^{-1}A(0)\Gamma = A(0).$$

Furthermore, we have

$$\tilde{A}'(0) = \Gamma^{-1}A'(0)\Gamma = \begin{bmatrix} \Gamma_1^{-1}A_{11}'\Gamma_1 & \Gamma_1^{-1}A_{12}'\Gamma_2 \\ \Gamma_2^{-1}A_{21}'\Gamma_1 & \Gamma_2^{-1}A_{22}'\Gamma_2 \end{bmatrix} = \begin{bmatrix} \Lambda_1' & \Gamma_1^{-1}A_{12}'\Gamma_2 \\ \Gamma_2^{-1}A_{21}'\Gamma_1 & \Lambda_2' \end{bmatrix},$$

where the last equality holds since $\Gamma_1$ and $\Gamma_2$ are eigenvector matrices of $A_{11}'$ and $A_{22}'$. Then, we divide $\tilde{A}(0)$, $\tilde{A}'(0)$ and $\tilde{A}''(0)$ and $\Lambda''(0)$ into blocks in the same way as that of $\Lambda'(0)$:

$$\tilde{A}(0) = \begin{bmatrix} \lambda I_{d \times d} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tilde{A}'(0) = \begin{bmatrix} \Lambda_1' & \tilde{A}_{12,1}' & \tilde{A}_{12,2}' \\ \tilde{A}_{21,1}' & \Lambda_2' & 0 \\ \tilde{A}_{21,2}' & 0 & \Lambda_2' \end{bmatrix},$$

$$\tilde{A}''(0) = \begin{bmatrix} \tilde{A}_{11}'' & \tilde{A}_{12,1}'' & \tilde{A}_{12,2}'' \\ \tilde{A}_{211}'' & \tilde{A}_{21,1}'' & \tilde{A}_{21,2}'' \\ \tilde{A}_{221}'' & \tilde{A}_{22,1}'' & \tilde{A}_{22,2}'' \end{bmatrix} \Lambda''(0) = \begin{bmatrix} \Lambda''_1 & 0 & 0 \\ 0 & \Lambda''_2 & 0 \\ 0 & 0 & \Lambda''_2 \end{bmatrix},$$

where we use the following notations for the blocks of $\tilde{A}'(0)$:

$$\Gamma_2^{-1}A_{21}'\Gamma_1 = \begin{bmatrix} \tilde{A}_{21}' \tilde{A}_{21}' \end{bmatrix} \Gamma_1^{-1}A_{12}'\Gamma_2 = \begin{bmatrix} \tilde{A}_{12,1}' \tilde{A}_{12,2}' \end{bmatrix}.$$
Since \( \tilde{X} = X \) and \( \tilde{X}' = X' \). Since
\[
\tilde{A}'_{22} = \Gamma^{-1}_{2} A'_{22} \Gamma_{2} = \begin{bmatrix} \Lambda'_{2,1} & 0 \\ 0 & \Lambda'_{2,2} \end{bmatrix}
\]
is a diagonal matrix after the conjugation with \( \Gamma \), by the assumption about the eigenvalues and (1.4.3) and (1.4.4), we have
\[
\tilde{X}(0) = \begin{bmatrix} \tilde{x}_{1} & 0 & 0 \\ 0 & \tilde{x}_{2,1} & 0 \\ 0 & 0 & \tilde{x}_{2,2} \end{bmatrix}.
\]
Similarly, define \( \tilde{X}'(0) = \tilde{X}(0) C \), where we divide \( C \) into blocks in the same way as that of \( \Lambda'(0) \):
\[
C = \begin{bmatrix} C_{11} & C_{12,1} & C_{12,2} \\ C_{21,1} & C_{22,11} & C_{22,12} \\ C_{21,2} & C_{22,21} & C_{22,22} \end{bmatrix}.
\]
Under such a block decomposition, we apply (1.4.3) to \( \tilde{A}'(0) \), and we have
\[
\Lambda'_{1} \tilde{x}_{i} = \tilde{x}_{1} \Lambda'_{1}, \quad (1.4.18)
\]
\[
\Lambda'_{2,1} \tilde{x}_{2,1} = \tilde{x}_{2,1} \Lambda'_{2,1}, \quad (1.4.19)
\]
\[
\Lambda'_{2,2} \tilde{x}_{2,2} = \tilde{x}_{2,2} \Lambda'_{2,2}, \quad (1.4.20)
\]
\[
\tilde{A}'_{12,1} \tilde{x}_{2,1} = -\lambda \tilde{x}_{1} C_{12,1}, \quad (1.4.21)
\]
\[
\tilde{A}'_{12,2} \tilde{x}_{2,2} = -\lambda \tilde{x}_{1} C_{12,2}, \quad (1.4.22)
\]
\[
\tilde{A}'_{21,1} \tilde{x}_{1} = \lambda \tilde{x}_{2,1} C_{21,1}, \quad (1.4.23)
\]
\[
\tilde{A}'_{21,2} \tilde{x}_{1} = \lambda \tilde{x}_{2,2} C_{21,2}. \quad (1.4.24)
\]
Then, we apply (1.4.9) to \( \tilde{A}''(0) \), we have
\[
\tilde{A}_{11}'' \tilde{x}_{1} - \tilde{x}_{1} \Lambda''_{1} = 2 \tilde{x}_{1} (C_{11} \Lambda'_{1} - \Lambda'_{1} C_{11}) - 2 \tilde{A}'_{12,1} \tilde{x}_{2,1} C_{21,1} - 2 \tilde{A}'_{12,2} \tilde{x}_{2,2} C_{21,2}, \quad (1.4.25)
\]
\[
\tilde{A}_{22,11}'' \tilde{x}_{2,1} - \tilde{x}_{2,1} \Lambda''_{2,1} = 2 \tilde{x}_{2,1} (C_{22,11} \Lambda'_{2,1} - \Lambda'_{2,1} C_{22,11}) - 2 \tilde{A}'_{21,1} \tilde{x}_{1} C_{12,1}, \quad (1.4.26)
\]
\[
\tilde{A}_{22,22}'' \tilde{x}_{2,2} - \tilde{x}_{2,2} \Lambda''_{2,2} = 2 \tilde{x}_{2,2} (C_{22,22} \Lambda'_{2,2} - \Lambda'_{2,2} C_{22,22}) - 2 \tilde{A}'_{21,2} \tilde{x}_{1} C_{12,2}, \quad (1.4.27)
\]
\[
\tilde{A}_{22,12}'' \tilde{x}_{2,1} = 2 \tilde{x}_{2,1} (C_{22,12} \Lambda'_{2,1} - \Lambda'_{2,1} C_{22,12}) - 2 \tilde{A}'_{21,2} \tilde{x}_{1} C_{12,2}, \quad (1.4.28)
\]
\[
\tilde{A}_{22,21}'' \tilde{x}_{2,1} = 2 \tilde{x}_{2,1} (C_{22,21} \Lambda'_{2,1} - \Lambda'_{2,1} C_{22,21}) - 2 \tilde{A}'_{21,2} \tilde{x}_{1} C_{12,1}. \quad (1.4.29)
\]
Since \( \Lambda'_{1} \) and \( \Lambda'_{2,1} \) both have distinct diagonal entries, by (1.4.18) and (1.4.19), we have
\[
\tilde{x}_{1} = I_{d \times d}
\]
and
\[
\tilde{x}_{2,1} = I_{(p-d-l) 	imes (p-d-l)}.
\]
In this case, \( C_{12,1} \) and \( C_{21,1} \) can be uniquely determined by (1.4.21) and (1.4.23), and we have

\[
C_{12,1} = \frac{-1}{\lambda} \tilde{A}_{12,1}, \quad (1.4.30)
\]
\[
C_{21,1} = \frac{1}{\lambda} \tilde{A}_{21,1}. \quad (1.4.31)
\]

Similarly, by (1.4.22) and (1.4.24), we have

\[
C_{12,2} = \frac{-1}{\lambda} \tilde{A}_{12,2}; \quad (1.4.32)
\]
\[
C_{21,2} = \frac{1}{\lambda} \tilde{A}_{21,2}. \quad (1.4.33)
\]

By plugging (1.4.32) into (1.4.27), and use the assumption that \( \Lambda'_{2,2} = \gamma l_{i,l} \), we can solve \( \Lambda''_{2,2} \). Indeed, since \( \Lambda'_{2,2} \) is a scalar multiple of the identity matrix, \( C_{22,22} \Lambda''_{2,2} - \Lambda'_{2,2} C_{22,22} = 0 \) in (1.4.27). Thus, (1.4.27) becomes

\[
(\tilde{A}_{22,22}'' - 2\lambda^{-1} \tilde{A}_{21,2} \tilde{A}_{12,2}) \tilde{x}_{2,2} = \tilde{x}_{2,2} \Lambda''_{2,2},
\]

and \( \Lambda''_{2,2} \) and \( \tilde{x}_{2,2} \) are eigenvalue and orthonormal eigenvector matrices of \( \tilde{A}_{22,22}'' - 2\lambda^{-1} \tilde{A}_{21,2} \tilde{A}_{12,2} \). Thus, we have obtained the eigenvalue information. However, note that in general \( \tilde{x}_{2,2} \) cannot be uniquely determined.

Suppose we want to uniquely determine the eigenvectors, \( \tilde{x}(e) \), we have to further assume that \( \Lambda'_{2,2} \) does not have repeated diagonal entries; that is, eigenvalues of \( \tilde{A}_{22,22}'' - 2\lambda^{-1} \tilde{A}_{21,2} \tilde{A}_{12,2} \) do not repeat. Under this assumption, \( \tilde{x}_{2,2} \) is uniquely determined, and we can proceed to solve \( C \). With \( \tilde{x}_{2,2} \), from (1.4.28) and (1.4.29) we can solve \( C_{22,12} \) and \( C_{22,21} \) since \( \Lambda'_{2,2} \) is a scalar multiple of the identity matrix and the diagonal entries of \( \Lambda'_{2,2} \) are assumed to be different from \( \Lambda_{2,2} \). In fact, we have

\[
C_{22,12} = (\gamma l_{(p-d-l)\times(p-d-l)} - \Lambda'_{2,2})^{-1} (1/2 \tilde{A}_{22,12}' \tilde{x}_{2,2} + \tilde{A}_{21,1} C_{12,2}), \quad (1.4.35)
\]
\[
C_{22,21} = \tilde{x}_{2,2}^\top (1/2 \tilde{A}_{21,2}' C_{12,1}) (\Lambda''_{2,2} - \gamma l_{(p-d-l)\times(p-d-l)})^{-1} . \quad (1.4.36)
\]

Next, \( \Lambda''_{2,2} \) and the off-diagonal entries of \( C_{11} \) and \( C_{22,11} \) are solved by rewriting (1.4.25) and (1.4.26) as

\[
2(C_{11} \Lambda''_{1} - \Lambda'(C_{11})) + \Lambda''_{1} = (\tilde{A}_{11}'' + 2\tilde{A}_{12,1} C_{21,1} + 2\tilde{A}_{12,2} \tilde{x}_{2,2} C_{21,2}) \quad (1.4.37)
\]
\[
2(C_{22,11} \Lambda''_{2,1} - \Lambda'(C_{22,11})) + \Lambda''_{2,1} = (\tilde{A}_{22,11}'' + 2\tilde{A}_{21,1} C_{12,1}) . \quad (1.4.38)
\]

Therefore, with the assumption that \( \Lambda''_{2,2} \) does not have repeated diagonal entries, we have

\[
(C_{11})_{m,n} = -\frac{1}{(\Lambda''_{1})_{m,n} - (\Lambda'(C_{11}))_{n,m}} e_m^\top ((1/2 \tilde{A}_{11,2}'' + \tilde{A}_{12,1} C_{21,1} + 2\tilde{A}_{12,2} \tilde{x}_{2,2} C_{21,2})) e_n,
\]
\[
(C_{22,11})_{m,n} = -\frac{1}{(\Lambda''_{2,1})_{m,n} - (\Lambda'(C_{22,11}))_{n,m}} e_m^\top ((1/2 \tilde{A}_{22,11,2}'' + \tilde{A}_{21,1} C_{12,1})) e_n,
\]

where \( 1 \leq m \neq n \leq d \) and \( d + 1 \leq i \neq j \leq p - l \). However, the problem cannot be solved and more information is needed. Indeed, note that (1.4.27) can be rewritten as

\[
\Lambda''_{2,2} = \tilde{x}_{2,2}^\top (\tilde{A}_{22,22}'' \tilde{x}_{2,2} + 2\tilde{A}_{21,2} C_{21,2}) = \tilde{x}_{2,2}^\top (\tilde{A}_{22,22}'' - 2\lambda \tilde{A}_{21,2} \tilde{A}_{12,2}) \tilde{x}_{2,2} \quad (1.4.39)
\]

since \( C_{22,22} \Lambda''_{2,2} - \Lambda''_{2,2} C_{22,22} = 0 \), which is the same as (1.4.34). Thus, it is not informative and we need higher
order derivatives of $A(\varepsilon)$ at 0 to solve $C_{22,22}$.

Suppose we know $A''(0)$, and denote $\tilde{A}'''(0) = \Gamma^{-1}A'''(0)\Gamma$, which is divided correspondingly as

$$
\tilde{A}'''(0) = \begin{bmatrix}
\tilde{A}'''_{11} & \tilde{A}'''_{12,1} & \tilde{A}'''_{12,2} \\
\tilde{A}'''_{21,1} & \tilde{A}'''_{22,11} & \tilde{A}'''_{22,12} \\
\tilde{A}'''_{21,2} & \tilde{A}'''_{22,21} & \tilde{A}'''_{22,22}
\end{bmatrix}.
$$

Then, if we differentiate (1.4.2) three times and use the similar method as before, we get

$$(C_{22,22})_{m,n} = -\frac{1}{(\Lambda''_{2,2})_{m,m} - (\Lambda''_{2,2})_{n,n}} e_m^T [\tilde{X}_{2,2}^T \tilde{A}'''_{22,22} \tilde{X}_{2,2} - \frac{2}{\lambda^2} \tilde{X}_{2,2}^T \tilde{A}'_{21,2} (\tilde{A}'_{11} - \gamma I_{d \times d}) \tilde{A}'_{12,2} \tilde{X}_{2,2}]
$$

$$
+ \frac{1}{\Lambda} \tilde{X}_{2,2}^T \tilde{A}'_{21,2} \tilde{A}'_{22,2} \tilde{X}_{2,2} + \frac{1}{\Lambda} \tilde{X}_{2,2}^T \tilde{A}'_{21,2} \tilde{A}'_{21,2} \tilde{A}'_{12,2} \tilde{X}_{2,2}
$$

$$
- \frac{4}{\Lambda^2} \tilde{X}_{2,2}^T (\tilde{A}'_{21,2} \tilde{A}'_{12,1}) (\gamma I_{(p-d-I) \times (p-d-I)} - \Lambda'_{2,1})^{-1} (\tilde{A}'_{21,2} \tilde{A}'_{12,2}) \tilde{X}_{2,2} e_n.
$$

By normalizing $\tilde{X}(\varepsilon)$, we can get the diagonal terms of $C_{11}$, $C_{22,11}$ and $C_{22,22}$, which are of order $\varepsilon$. As a result, we have

$$
\Lambda(\varepsilon) = \begin{bmatrix}
\lambda I_{d \times d} + \varepsilon A'_1 + \varepsilon^2 A''_1 & 0 & 0 \\
0 & \varepsilon A'_{2,1} + \varepsilon^2 A''_{2,1} & 0 \\
0 & 0 & \varepsilon A'_{2,2} + \varepsilon^2 A''_{2,2}
\end{bmatrix} + O(\varepsilon^3),
$$

$$
\tilde{X}(\varepsilon) = \tilde{X}(0) (I_{p \times p} + \varepsilon (C - \text{diag}(C))) + O(\varepsilon^2) \in O(p),
$$

where the last equality holds since the entries of $\text{diag}(C)$ are of order $\varepsilon$. Finally, we can find $X(0)$ and $X'(0)$ by using

$$
X(0) = \Gamma \tilde{X}(0), \quad X'(0) = \Gamma \tilde{X}'(0).
$$

**General cases.** In general, if $A'_1$ or $A'_{2,1}$ has repeated diagonal entries, we divide them into more blocks, and the block with the same diagonal entries can be treated in the same way as we treated $A'_{2,2}$ above. We skip details here.
Chapter 2

LLE on closed manifolds

2.1 Manifold Setup

Let $X$ be a $p$-dimensional random vector. Assume that the distribution of $X$ is supported on a $d$-dimensional compact, smooth Riemannian manifold $(M, g)$ isometrically embedded in $\mathbb{R}^p$ via $\iota: M \hookrightarrow \mathbb{R}^p$, where we assume that $M$ is boundary-free to simplify the discussion. Denote $d(\cdot, \cdot)$ to be the geodesic distance associated with $g$.

For the tangent space $T_yM$ on $y \in M$, denote $\iota^*T_yM$ to be the embedded tangent space in $\mathbb{R}^p$. Denote the normal space at $\iota(y)$ as $(\iota^*T_yM)^\perp$. Denote $\mathbb{II}_x$ to be the second fundamental form of $\iota$ at $x$. Denote $\exp_y: T_yM \rightarrow M$ to be the exponential map at $y$. Denote $S^{d-1}$ to be the $(d-1)$-dim unit sphere embedded in $\mathbb{R}^p$, and denote $|S^{d-1}|$ to be the volume. Unless otherwise stated, in this paper we will carry out the calculation with the normal coordinate.

We now define the probability density function (p.d.f.) associated with $X$. The random vector $X: \Omega \rightarrow \mathbb{R}^p$ is a measurable function with respect to the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where $\mathcal{P}$ is the probability measure defined on the sigma algebra $\mathcal{F}$ in $\Omega$. By assumption, the range of $X$ is supported on $\iota(M)$. Let $\mathcal{B}$ be the Borel sigma algebra of $\iota(M)$, and denote by $\mathcal{P}_X$ the probability measure defined on $\mathcal{B}$, induced from $\mathcal{P}$. When $d < p$, the p.d.f. of $X$ is not well-defined as a function on $\mathbb{R}^p$. Thus, for an integrable function $\zeta: \iota(M) \rightarrow \mathbb{R}$, we have

$$E\zeta(X) = \int_\Omega \zeta(X(\omega))d\mathcal{P}(\omega) = \int_{\iota(M)} \zeta(z)d\mathcal{P}_X(z), \quad (2.1.1)$$

where the second equality follows from the fact that $\mathcal{P}_X$ is the induced probability measure. If $\mathcal{P}_X$ is absolutely continuous with respect to the volume density on $\iota(M)$, by the Radon-Nikodym theorem, $d\mathcal{P}_X(z) = P(z)\iota_*dV(z)$, where $\iota_*dV(z)$ is the induced measure on $\iota(M)$ via $\iota$, $P$ is a non-negative measurable function defined on $\iota(M)$ and $dV$ is the volume form associated with the metric $g$. Thus, (2.1.1) becomes

$$E\zeta(X) = \int_{\iota(M)} \zeta(z)P(x)\iota_*dV(z) = \int_M \zeta(t(y))P(t(y))dV(y), \quad (2.1.2)$$

where the second equality comes from the change of variable $z = t(y)$. We thus call $P$ as the p.d.f. of $X$ on $M$. When $P$ is a constant function, we call $X$ a uniform random sampling scheme; otherwise it is nonuniform. Since $\iota$ is an isometric embedding, when we do the calculation, we will abuse the notation and write

$$E\zeta(X) = \int_M \zeta(y)dV(y). \quad (2.1.3)$$
In other words, we will not distinguish either a function is defined on \( t(M) \) or \( M \).

With the above discussion, to facilitate the discussion and the upcoming analysis, we make the following assumption about the random vector \( X \) and the regularity of the associated p.d.f..

**Assumption 2.1.1.** Assume \( \tilde{P}_X \) is absolutely continuous with respect to the volume density on \( t(M) \) so that \( d\tilde{P}_X = P_t, dV \), where \( P \) is a measurable function. We further assume that \( P \in C^5(t(M)) \) and there exist \( P_m > 0 \) and \( P_M \geq P_m \) so that \( P_m \leq P(x) \leq P_M < \infty \) for all \( x \in t(M) \).

Let \( \mathcal{X} = \{ t(x_i) \}_{i=1}^n \subset t(M) \subset \mathbb{R}^p \) denote a set of identical and independent (i.i.d.) random samples from \( X \), where \( x_i \in M \). We could then run LLE on \( \mathcal{X} \). For \( t(x_i) \in \mathcal{X} \) and \( \epsilon > 0 \), we have \( \mathcal{M}_{t(x_i)} := \{ t(x_{i,1}), \ldots, t(x_{i,N}) \} \subset B_{\epsilon}^{\mathbb{R}^p}(t(x_i)) \cap (\mathcal{X} \setminus \{ t(x_i) \}) \). Take \( G_n \in \mathbb{R}^{p \times N} \) to be the local data matrix associated with \( \mathcal{M}_{t(x_i)} \) and evaluate the barycentric coordinates \( w_n = [w_{n,1}, \ldots, w_{n,N}]^\top \in \mathbb{R}^N \). Again, while \( G_n \) and \( w_n \) depend on \( \epsilon, n, \) and \( x_k \), to ease the notation, we only keep \( n \) to indicate that we have a finite sampling points.

In our asymptotic analysis of LLE under the manifold setup, we make the following assumption about the tangent space of the manifold.

**Assumption 2.1.2.** Since the barycentric coordinates are rotational and translational invariant, without loss of generality, we assume that the manifold is translated and rotated properly, so that \( t, t_\ast M \) is spanned by \( e_1, \ldots, e_d \).

## 2.2 Some preliminary lemmas in Riemannian geometry

In this section we prepare several technical lemmas. For \( \nu \in \mathbb{R}^p \), we use the following notation to simplify the proof:

\[
\nu = [\nu_1, \nu_2] \in \mathbb{R}^p,
\]

where \( \nu_1 \in \mathbb{R}^d \) forms the first \( d \) coordinates of \( \nu \) and \( \nu_2 \in \mathbb{R}^{p-d} \) forms the last \( p-d \) coordinates of \( \nu \). Thus, under Assumptions 2.1.2, for \( \nu = [\nu_1, \nu_2] \in T_{t(x)} \mathbb{R}^p \), \( \nu_1 = J_{\nu, d}^\top \nu \) is tangential to \( t_\ast T_x M \) and \( \nu_2 = J_{\nu, p-d}^\top \nu \) is the coordinate of the normal component of \( \nu \) associated with a chosen basis of the normal bundle. The first three lemmas are basic facts about the exponential map, the normal coordinate, and the volume form. The proofs of Lemmas 2.2.1 and 2.2.2 are standard and we skip the proof. Interested readers are referred to [55].

**Lemma 2.2.1.** Fix \( x \in M \). If we use the polar coordinate \( (t, \theta) \in [0, \infty) \times S^{d-1} \) to parametrize \( T_x M \), the volume form has the following expansion:

\[
dV = \left( t^{d-1} - \frac{1}{6} \nabla_\theta R_{c_i}(\theta, \theta) t^{d+1} - \frac{1}{12} \nabla_\theta^2 R_{c_i}(\theta, \theta) t^{d+2} \right. \\
- \left. \left( \frac{1}{40} \nabla^2_{\theta \theta} R_{c_i}(\theta, \theta) + \frac{1}{180} \sum_{a,b=1}^d R_a(\theta, a, \theta, b) R_b(\theta, a, \theta, b) - \frac{1}{72} R_{c_i}(\theta, \theta)^2 \right) t^{d+3} + O(t^{d+4}) \right) dt d\theta,
\]

where \( R_{c_i} \) is the Riemannian curvature of \( (M, g) \) at \( x \). If we use the Cartesian coordinate to parametrize \( T_x M \), the
Moreover, since \( \|t\| \) sufficiently small, we have the following Taylor expansion:

\[
dV = \left(1 - \frac{d}{12} \sum_{i,j=1}^{d} R \mathring{c}_i(\partial_i, \partial_j) u^i u^j - \frac{1}{12} \sum_{i,j,k,l=1}^{d} \frac{1}{12} \mathring{R} \mathring{c}_i(\partial_i, \partial_j) u^i u^j u^k\right)
- \sum_{i,j,k,l=1}^{d} \left[ \frac{1}{80} \mathring{R} \mathring{c}_i(\partial_i, \partial_j) + \frac{1}{180} \sum_{a,b=1}^{d} \mathring{R}_a(\partial_i, \partial_j, \partial_k, \partial_l) \mathring{R}_b(\partial_i, \partial_j, \partial_k, \partial_l)\right] u^i u^j u^k u^l + O(\|u\|^5) \right) du,
\]

where \( u = u^{\partial_i} \in T_\circ M \).

**Lemma 2.2.2.** Fix \( x \in M \). For \( u \in T_\circ M \) with \( \|u\| \) sufficiently small, we have the following Taylor expansion:

\[
t \circ \exp_x(u) - t(x) = t_u^1 + \frac{1}{2} \Pi_x(u, u) + \frac{1}{6} \nabla_x \Pi_x(u, u) + \frac{1}{24} \nabla^2 u^2 \Pi_x(u, u) + \frac{1}{120} \nabla^3 u^3 \Pi_x(u, u) + O(\|u\|^6).
\]

**Lemma 2.2.3.** Fix \( x \in M \). If we use the polar coordinate \( (t, \theta) \in [0, \infty) \times S^{d-1} \) to parametrize \( T_\circ M \), when \( \tilde{t} = \|t \circ \exp_x(\theta t) - t(x)\|_{\mathcal{R}^p} \) is sufficiently small, we have

\[
\tilde{t} = t + \frac{1}{24} \|\Pi_x(t, \theta, \theta)\|^2 r^3 + \frac{1}{24} \|\nabla \Pi_x(t, \theta, \theta)\| r^4 - \left( \frac{1}{80} \|\nabla^2 \Pi_x(t, \theta, \theta)\| r^5 + O(t^6) \right)
+ \frac{1}{90} \|\nabla^3 \Pi_x(t, \theta, \theta)\| r^5 + O(t^6)
\]

Hence, \( (t \circ \exp_x)^{-1}(B_t^p(\pi(x)) \cap t(M)) \subset T_\circ M^d \) is star shaped.

**Proof.** Let \( \gamma(t) \) be the geodesic in \( t(M) \) with \( \gamma(0) = t(x) \). If \( \gamma^{(i)}(0) \) denotes the \( i \)-th derivative of \( \gamma(t) \) with respect to \( t \) at 0, then we have

\[
\gamma(t) = \gamma(0) + \frac{1}{2} \gamma^{(2)}(0) t^2 + \frac{1}{6} \gamma^{(3)}(0) t^3 + \frac{1}{24} \gamma^{(4)}(0) t^4 + \frac{1}{120} \gamma^{(5)}(0) t^5 + O(t^6).
\]

Moreover, since \( \gamma(t) \) is a geodesic, if we apply the product rule, we have

\[
\gamma^{(1)}(0) \cdot \gamma^{(1)}(0) = 1, \quad (2.2.2)
\]
\[
\gamma^{(2)}(0) \cdot \gamma^{(1)}(0) = 0, \quad (2.2.3)
\]
\[
\gamma^{(2)}(0) \cdot \gamma^{(0)}(0) = -\gamma^{(3)}(0) \cdot \gamma^{(1)}(0),
\]
\[
3\gamma^{(3)}(0) \cdot \gamma^{(2)}(0) = -\gamma^{(4)}(0) \cdot \gamma^{(1)}(0),
\]
\[
4\gamma^{(4)}(0) \cdot \gamma^{(2)}(0) + 3\gamma^{(3)}(0) \cdot \gamma^{(1)}(0) = -\gamma^{(5)}(0) \cdot \gamma^{(1)}(0),
\]
where $\gamma^{(l)}$ is the $l$-th derivative of $\gamma$ and $l \in \mathbb{N}$. From (2.2.2), we have

$$\|\gamma(t) - \gamma(0)\|^2_{\mathbb{R}^p} = \gamma^{(1)}(0) \cdot \gamma^{(1)}(0) t^2 + (\gamma^{(2)}(0) \cdot \gamma^{(1)}(0)) t^3$$

\[\begin{align*}
&+ \left(\frac{1}{3}\gamma^{(3)}(0) \cdot \gamma^{(1)}(0) + \frac{1}{4}\gamma^{(2)}(0) \cdot \gamma^{(2)}(0)) t^4 \\
&+ \left(\frac{1}{12}\gamma^{(4)}(0) \cdot \gamma^{(1)}(0) + \frac{1}{6}\gamma^{(3)}(0) \cdot \gamma^{(2)}(0)) t^5 \\
&+ \left(\frac{1}{60}\gamma^{(5)}(0) \cdot \gamma^{(1)}(0) + \frac{1}{24}\gamma^{(4)}(0) \cdot \gamma^{(2)}(0) + \frac{1}{36}\gamma^{(3)}(0) \cdot \gamma^{(3)}(0)) t^6 + O(t^7).
\end{align*}\]

If we substitute (2.2.3) into (2.2.4), we have

$$\|\gamma(t) - \gamma(0)\|^2_{\mathbb{R}^p} = t^2 - \frac{1}{12}\gamma^{(2)}(0) \cdot \gamma^{(2)}(0) t^4 - \frac{1}{12}\gamma^{(3)}(0) \cdot \gamma^{(2)}(0) t^5$$

\[\begin{align*}
&- \left(\frac{1}{40}\gamma^{(4)}(0) \cdot \gamma^{(2)}(0) + \frac{1}{45}\gamma^{(3)}(0) \cdot \gamma^{(3)} + \frac{1}{1152}(\gamma^{(2)}(0) \cdot \gamma^{(2)}(0))^2\right) t^6 + O(t^7).
\end{align*}\]

Therefore

$$\tilde{t} = \|\gamma(t) - \gamma(0)\|^2_{\mathbb{R}^p} = t - \frac{1}{24}\gamma^{(2)}(0) \cdot \gamma^{(2)}(0) t^3 - \frac{1}{24}\gamma^{(3)}(0) \cdot \gamma^{(2)}(0) t^4$$

\[\begin{align*}
&- \left(\frac{1}{80}\gamma^{(4)}(0) \cdot \gamma^{(2)}(0) + \frac{1}{90}\gamma^{(3)}(0) \cdot \gamma^{(3)} + \frac{1}{1152}(\gamma^{(2)}(0) \cdot \gamma^{(2)}(0))^2\right) t^5 + O(t^6).
\end{align*}\]

By comparing the order, we have

$$t = \tilde{t} + \frac{1}{24}\gamma^{(2)}(0) \cdot \gamma^{(2)}(0) t^3 + \frac{1}{24}\gamma^{(3)}(0) \cdot \gamma^{(2)}(0) t^4$$

\[\begin{align*}
&+ \left(\frac{1}{80}\gamma^{(4)}(0) \cdot \gamma^{(2)}(0) + \frac{1}{90}\gamma^{(3)}(0) \cdot \gamma^{(3)} + \frac{7}{1152}(\gamma^{(2)}(0) \cdot \gamma^{(2)}(0))^2\right) t^5 + O(t^6).
\end{align*}\]

Finally, by applying Lemma 2.2.2 to (2.2.2) with $\gamma(t) = t \circ \exp_\gamma(\theta t)$ and substituting corresponding terms for $\gamma^{(l)}(0)$, the conclusion follows. \(\square\)

The essence of Lemma 2.2.3 is describing how well we could estimate the local geodesic distance by the ambient space metric. When the manifold setup is considered in an algorithm, this Lemma could be helpful in the analysis since most of time we only have an access to the ambient space metric, but not the intrinsic Riemannian metric.

To alleviate the notational load, we denote

$$\tilde{B}_\varepsilon(x) := t^{-1}(B_{\tilde{B}_\varepsilon}(t(x)) \cap t(M)) \subset M, \quad (2.2.8)$$

and for a sufficiently small $\varepsilon$, by Lemma 2.2.3, denote

$$\varepsilon = \varepsilon - \frac{1}{24}\|\Pi_\varepsilon(\theta, \theta)\|^2\varepsilon \cdot \varepsilon - \frac{1}{24}\nabla_\theta \Pi_\varepsilon(\theta, \theta) \cdot \Pi_\varepsilon(\theta, \theta) \varepsilon^4 - \left(\frac{1}{80}\nabla_\theta \Pi_\varepsilon(\theta, \theta) \cdot \Pi_\varepsilon(\theta, \theta)ight)\varepsilon^4$$

\[\begin{align*}
&+ \frac{1}{90}\nabla_\theta \Pi_\varepsilon(\theta, \theta) \cdot \nabla_\theta \Pi_\varepsilon(\theta, \theta) + \frac{1}{1152}\|\Pi_\varepsilon(\theta, \theta)\|^4\varepsilon^5 + O(\varepsilon^6).
\end{align*}\]

To have a more succinct proof, we prepare the following integration, which comes from a direct expansion and the proof is skipped.
Lemma 2.2.4. For \( d \in \mathbb{N} \), \( \gamma > -d \) and \( h_1, h_2, h_3 \in \mathbb{R} \), we have the following asymptotical expansion when \( \varepsilon \) is sufficiently small:

\[
\int_0^{\varepsilon + h_1 \varepsilon^3 + h_2 \varepsilon^4 + h_3 \varepsilon^5 + O(\varepsilon^6)} t^{d-1+\gamma} dt = \frac{\varepsilon^{d+\gamma}}{d+\gamma} \left( 1 + (d+\gamma)h_1 \varepsilon^2 + (d+\gamma)h_2 \varepsilon^3 + \left( (d+\gamma)h_3 + \frac{(d+\gamma)(d+\gamma-1)}{2} h_1^2 \right) \varepsilon^4 \right) + O(\varepsilon^{d+\gamma+5}).
\]

In the next lemma, we calculate the asymptotical expansion of few quantities that we are going to use in proving the main theorem. Note that in order to capture the extra terms introduced by the barycentric coordinate, we calculate the normal term 2 orders higher than those for the tangential direction. To handle the normal component is the main reason we need the \( C^5 \) regularity for \( P \).

Lemma 2.2.5. Fix \( x \in M \) and assume Assumptions 2.1.2 hold. When \( \varepsilon \) is sufficiently small, we have the following expansion for \( \mathbb{E}[f(X) \chi_{B^p \varepsilon \gamma}^{\varepsilon}(t(x))(X)] \) :

\[
\mathbb{E}[f(X) \chi_{B^p \varepsilon \gamma}^{\varepsilon}(t(x))(X)] = \frac{|S^{d-1}|}{d} f(x) \varepsilon^d + \frac{|S^{d-1}|}{d(d+2)} \varepsilon f(x) \Delta f(x) + \frac{1}{2} f(x) \Delta P(x) + \nabla f(x) \cdot \nabla P(x) + \frac{s(x) f(x) P(x)}{6} + \frac{d(d+2) \omega(x) f(x) P(x)}{24} \varepsilon^{d+2} + O(\varepsilon^{d+3})
\]

and the following expansion for \( \mathbb{E}[X - t(x)) f(X) \chi_{B^p \varepsilon \gamma}^{\varepsilon}(t(x))(X)] \in \mathbb{R}^p :\)

\[
\mathbb{E}[X - t(x)) f(X) \chi_{B^p \varepsilon \gamma}^{\varepsilon}(t(x))(X)] = \|v_1, v_2\| + O(\varepsilon^{d+5}),
\]

where \( v_1 \in \mathbb{R}^d \) and \( v_2 \in \mathbb{R}^{p-d} \) are defined in (2.2.14) and (2.2.15) respectively, which contain terms of order \( \varepsilon^{d+2} \) and \( \varepsilon^{d+4} \).

Proof. By Lemma 2.2.1, Lemma 2.2.2 and Lemma 2.2.3,

\[
\mathbb{E}[f(X) \chi_{B^p \varepsilon \gamma}^{\varepsilon}(t(x))(X)] = \int_{B(x)} f(y) P(y) dV(y) = \int_{S^{d-1}} \int_0^{\varepsilon} (f(x) + \nabla \theta f(x) t + \frac{1}{2} \nabla_{\theta} \theta f(x) t^2 + O(t^3)) (P(x) + \nabla \theta P(x) t + \nabla_{\theta} \theta P(x) t^2 + O(t^3))(t^{d-1} - \frac{1}{6} \text{Ric}_{\varepsilon}(\theta, \theta) t^{d+1} + O(t^{d+2})) dt d\theta = A_1 + B_1 + C_1 + O(\varepsilon^{d+4}),
\]
where

\[ A_1 := \int_{S^{d-1}} \int_0^\varepsilon f(x) P(x) t^{d-1} dt d\theta \quad (2.2.11) \]

\[ B_1 := \int_{S^{d-1}} \int_0^\varepsilon \left( \nabla f(x) P(x) + f(x) \nabla P(x) \right) t^d dt d\theta \]

\[ C_1 := \int_{S^{d-1}} \int_0^\varepsilon \left[ \frac{1}{6} f(x) P(x) R \tilde{c}_x(\theta, \theta) + \nabla f(x) \nabla P(x) + \frac{1}{2} \nabla^2 \tilde{\theta} f(x) P(x) + \frac{1}{2} \nabla^2 \tilde{\theta} P(x) f(x) \right] t^{d+1} dt d\theta, \]

the second equality holds by Lemma 2.2.3 and the last equality holds due to the symmetry of sphere. Indeed, the symmetry forces all terms of odd order contribute to the \( \varepsilon^{d+4} \) term; for example,

\[
B_1 = \int_{S^{d-1}} \int_0^\varepsilon (\nabla f(x) P(x) + f(x) \nabla P(x)) t^d dt d\theta \\
= \frac{1}{d+1} \int_{S^{d-1}} (\nabla f(x) P(x) + f(x) \nabla P(x)) \left( \varepsilon + \frac{1}{24} \| \Pi_x(\theta, \theta) \|^2 \varepsilon^3 + O(\varepsilon^4) \right)^{d+1} d\theta \\
= \frac{\varepsilon^{d+1}}{d+1} \int_{S^{d-1}} (\nabla f(x) P(x) + f(x) \nabla P(x)) \left( 1 + \frac{d+1}{24} \| \Pi_x(\theta, \theta) \|^2 \varepsilon^2 + O(\varepsilon^3) \right) d\theta \\
= O(\varepsilon^{d+4})
\]

since \( \int_{S^{d-1}} (\nabla f(x) P(x) + f(x) \nabla P(x)) \left( 1 + \frac{d+1}{24} \| \Pi_x(\theta, \theta) \|^2 \varepsilon^2 \right) d\theta = 0 \). The other even order terms could be expanded by a direct calculation. We have

\[
A_1 = \int_{S^{d-1}} \int_0^\varepsilon f(x) P(x) t^{d-1} dt d\theta \\
= \frac{f(x) P(x)}{d} \int_{S^{d-1}} \left( \varepsilon + \frac{1}{24} \| \Pi_x(\theta, \theta) \|^2 \varepsilon^3 + O(\varepsilon^4) \right)^d d\theta \\
= \varepsilon^d \frac{f(x) P(x)}{d} \int_{S^{d-1}} \left( 1 + \frac{d}{24} \| \Pi_x(\theta, \theta) \|^2 \varepsilon^2 + O(\varepsilon^3) \right) d\theta \\
= \varepsilon^d |S^{d-1}| f(x) P(x) \left[ \frac{1}{d} + \frac{\omega_1(x)}{24} \varepsilon^2 \right] + O(\varepsilon^{d+3}).
\]

A similar argument holds for \( B_1 \). By denoting \( R_2(\theta) := \frac{1}{6} f(x) P(x) R \tilde{c}_x(\theta, \theta) + \nabla f(x) \nabla P(x) + \frac{1}{2} \nabla \tilde{\theta} f(x) P(x) + \frac{1}{2} \nabla^2 \tilde{\theta} P(x) f(x) \), we have

\[
C_1 = \int_{S^{d-1}} \int_0^\varepsilon R_2(\theta) t^{d+1} dt d\theta \\
= \frac{1}{d+2} \int_{S^{d-1}} \left( \varepsilon + \frac{1}{24} \| \Pi_x(\theta, \theta) \|^2 \varepsilon^3 + O(\varepsilon^4) \right)^{d+2} R_2(\theta) d\theta \\
= \frac{\varepsilon^{d+2}}{d+2} \int_{S^{d-1}} \left( 1 + \frac{d+2}{24} \| \Pi_x(\theta, \theta) \|^2 \varepsilon^2 + O(\varepsilon^3) \right) R_2(\theta) d\theta \\
= \frac{\varepsilon^{d+2}}{d+2} \int_{S^{d-1}} R_2(\theta) d\theta + O(\varepsilon^{d+4}).
\]
To proceed, note that by expressing \( \theta \) in the local coordinate as \( \theta^i \partial_i \), we have, for example,

\[
\int_{S^{d-1}} \nabla_{\theta} f(x) \nabla_{\theta} P(x) d\theta = \sum_{ij} \int_{S^{d-1}} \partial_i f(x) \partial_j P(x) \theta^i \theta^j d\theta
\]

\[
= \sum_{i} \int_{S^{d-1}} \partial_i f(x) \partial_i P(x) (\theta^i)^2 d\theta = \frac{|S^{d-1}|}{d} \nabla f(x) \cdot \nabla P(x),
\]

where the second equality holds since odd order terms disappear when integrated over the sphere, and the last equality holds since \( \int_{S^{d-1}} (\theta^i)^2 d\theta = \frac{1}{d} \int_{S^{d-1}} (\theta^1)^2 d\theta = \frac{|S^{d-1}|}{d} \) due to again the symmetry of the sphere. The same argument leads to

\[
\int_{S^{d-1}} f(x) P(x) \mathrm{Rc}(\theta, \theta) d\theta = \frac{|S^{d-1}|}{d} f(x) P(x) s(x)
\]

\[
\int_{S^{d-1}} \nabla^2_{\theta, \theta} f(x) P(x) d\theta = \frac{|S^{d-1}|}{d} f(x) \Delta P(x)
\]

\[
\int_{S^{d-1}} \nabla^2_{\theta, \theta} P(x) f(x) d\theta = \frac{|S^{d-1}|}{d} P(x) \Delta f(x),
\]

where \( s(x) \) is the scalar curvature of \((M, g)\) at \( x \). As a result, we have

\[
C_1 = \frac{|S^{d-1}|}{d(d+2)} \left[ \frac{1}{2} P(x) \Delta f(x) + \frac{1}{2} f(x) \Delta P(x) + \nabla f(x) \cdot \nabla P(x) + \frac{s(x) f(x) P(x)}{6} \right] e^{d+2} + O(\varepsilon^{d+4}).
\]

By putting all the above together, we have

\[
\mathbb{E}[f(X) \chi_{B^p}(t(x))](X)] = \frac{|S^{d-1}|}{d} f(x) P(x) e^d + \frac{|S^{d-1}|}{d(d+2)} \left[ \frac{1}{2} P(x) \Delta f(x) + \frac{1}{2} f(x) \Delta P(x) \right] e^{d+2} + O(\varepsilon^{d+3}).
\]

Next, we evaluate \( \mathbb{E}[(X - t(x)) f(X) \chi_{B^p}(t(x))](X) \). Again, by Lemma 2.2.1, Lemma 2.2.2 and Lemma 2.2.3, we have

\[
\mathbb{E}[(X - t(x)) f(X) \chi_{B^p}(t(x))](X)] = \int_{B(x)} (t(y) - t(x)) f(y) P(y) dV(y)
\]

\[
= \int_{S^{d-1}} \int_0^\epsilon (t(1) + t(x)) + \frac{1}{2} \nabla_{\theta} \nabla \nabla_{\theta} (\theta, \theta) r^2 + \frac{1}{6} \nabla_{\theta} \nabla \nabla_{\theta} (\theta, \theta) r^3 + \frac{1}{24} \nabla_{\theta} \nabla \nabla_{\theta} (\theta, \theta) r^4 + O(r^5))
\]

\[
\times (f(x) + \nabla_{\theta} f(x) r + \frac{1}{2} \nabla^2_{\theta, \theta} f(x) r^2 + O(r^3))
\]

\[
\times (P(x) + \nabla_{\theta} P(x) r + \frac{1}{2} \nabla^2_{\theta, \theta} P(x) r^2 + O(r^3))
\]

\[
\times ((r^{d-1} - \frac{1}{6} \mathrm{Ric}(\theta, \theta)) r^{d+1} + O(r^{d+2})) d\theta dt d\theta
\]

\[
= A_2 + B_2 + C_2 + D_2 + O(\varepsilon^{d+6}),
\]
where

\[ A_2 := \int_{S^{d-1}} \int_{0}^{E} [t_s \theta f(x) P(x)] P(x) t^d dt d\theta \]

\[ B_2 := \int_{S^{d-1}} \int_{0}^{E} [t_s \theta (\nabla \theta f(x) P(x) + \nabla \theta P(x) f(x)) + \frac{1}{2} \Pi_{\epsilon}(\theta, \theta) f(x) P(x)] t^d+1 dt d\theta, \]

\[ C_2 := \int_{S^{d-1}} \int_{0}^{E} [t_s \theta (\nabla \theta f(x) \nabla \theta P(x) + \nabla \theta \theta P(x) f(x) + \nabla \theta \theta f(x) P(x)) + \frac{1}{6} \nabla \theta \Pi_{\epsilon}(\theta, \theta) f(x) P(x) - \frac{1}{6} f(x) P(x) \text{Ric}_\epsilon(\theta, \theta)] t^d+2 dt d\theta \]

and

\[ D_2 := \int_{S^{d-1}} \int_{0}^{E} [t_s \theta (\frac{1}{6} \nabla \theta \theta P(x) + \frac{1}{6} \nabla \theta \theta f(x) + \frac{1}{2} \Pi_{\epsilon}(\theta, \theta) f(x) P(x)) + \frac{1}{6} \Pi_{\epsilon}(\theta, \theta) (\nabla \theta f(x) \nabla \theta P(x) + \nabla \theta \theta P(x) f(x) + \nabla \theta \theta f(x) P(x)) - \frac{1}{6} \Pi_{\epsilon}(\theta, \theta) f(x) P(x)) + \frac{1}{6} \nabla \theta \Pi_{\epsilon}(\theta, \theta) f(x) P(x) + \frac{1}{24} f(x) P(x) \text{Ric}_\epsilon(\theta, \theta)] t^d+3 dt d\theta \]

and the \(O(\epsilon^{d+5})\) term disappears in the last equality due to the symmetry of the sphere. The main difference between evaluating \(\mathbb{E}[(X - t(x)) f(X) \chi_{B^d_{\epsilon}(t(x))}(X)]\) and \(\mathbb{E}[f(X) \chi_{B^d_{\epsilon}(t(x))}(X)]\) is the existence of \(t(y)\) in the integrand in (2.2.13). Clearly, \(\mathbb{E}[(X - t(x)) f(X) \chi_{B^d_{\epsilon}(t(x))}(X)]\) is a vector while \(\mathbb{E}[f(X) \chi_{B^d_{\epsilon}(t(x))}(X)]\) is a scalar. Due to the curvature, \(t(y) - t(x)\) does not always exist on \(t_s T_s M\) for all \(y \in B_{\epsilon}\), and we need to carefully trace the normal components. By Lemma 2.2.4,

\[ A_2 = \int_{S^{d-1}} \int_{0}^{E} f(x) P(x) t_s \theta t^d dt d\theta \]

\[ = \frac{f(x) P(x)}{d + 1} \int_{S^{d-1}} t_s \theta (\epsilon + \frac{1}{24} \Pi_{\epsilon}(\theta, \theta)) \frac{\epsilon^d}{d} d\theta \]

\[ = \frac{f(x) P(x)}{d + 1} \int_{S^{d-1}} [t_s \theta (\epsilon + \frac{1}{24} \Pi_{\epsilon}(\theta, \theta)) \frac{\epsilon^d}{d} d\theta + \frac{1}{24} t_s \theta |\Pi_{\epsilon}(\theta, \theta)| \frac{\epsilon^d}{d} + O(\epsilon^{d+5})] d\theta \]

\[ = \frac{f(x) P(x)}{24} [t_s \int_{S^{d-1}} \theta \Pi_{\epsilon}(\theta, \theta) \cdot \Pi_{\epsilon}(\theta, \theta) d\theta] \frac{\epsilon^d}{d} + O(\epsilon^{d+5}) \]

\[ = \frac{f(x) P(x) |S^{d-1}|}{24} t_s \text{R}_{\epsilon}(x) \frac{\epsilon^d}{d} + O(\epsilon^{d+5}), \]

where the second last equality holds due to the symmetry of the sphere. We could see that \(A_2 = O(\epsilon^{d+4})\) and \(A_2 \in t_s T_s M\). Similarly, we have \(C_2 = O(\epsilon^{d+6})\), but \(C_2\) might not be on \(t_s T_s M\) due to the term \(\nabla \theta \Pi_{\epsilon}(\theta, \theta) f(x) P(x)\).
$B_2$ could be evaluated by a similar direct expansion.

$$B_2 = \int_{S^{d-1}} \int_0^2 \left[ P(x) \triangledown \theta (\triangledown f(x) \cdot \theta) + f(x) \triangledown \theta (\triangledown P(x) \cdot \theta) \ight. \\
+ \frac{1}{2} \Pi_s(\theta, \theta) f(x) P(x) \right] t^{d+1} dt d\theta \\
= \frac{\varepsilon^{d+2}}{d+2} \int_{S^{d-1}} \left[ P(x) \triangledown \theta \triangledown f(x) + f(x) \triangledown \theta \triangledown P(x) + \frac{P(x)}{2} \Pi_s(\theta, \theta) f(x) \right] d\theta \\
+ \frac{\varepsilon^{d+4}}{24} \int_{S^{d-1}} \left( \frac{1}{2} \Pi_s(\theta, \theta) \right)^2 \left[ P(x) \triangledown \theta \triangledown f(x) + f(x) \triangledown \theta \triangledown P(x) + \frac{P(x)}{2} \Pi_s(\theta, \theta) f(x) \right] d\theta + O(\varepsilon^{d+5}),$$

which becomes

$$\frac{\varepsilon^{d+2}}{d+2} \int_{S^{d-1}} \left[ \frac{1}{2} \Pi_s(\theta, \theta) \right] \left[ P(x) \triangledown \theta \triangledown f(x) + f(x) \triangledown \theta \triangledown P(x) \right] + \frac{|S^{d-1}|}{2(d+2)} f(x) P(x) \Omega_0(x) \varepsilon^{d+2} \\
+ \frac{\varepsilon^{d+4}}{24} \left( \frac{1}{2} \Pi_s(\theta, \theta) \right)^2 \left[ P(x) \triangledown \theta \triangledown f(x) + f(x) \triangledown \theta \triangledown P(x) \right] \frac{f(x) P(x)}{2} \Omega_0(x) d\theta + O(\varepsilon^{d+5}) \\
= \frac{|S^{d-1}|}{d+2} \left[ \frac{1}{2} \Pi_s(\theta, \theta) \right] \left[ P(x) \triangledown \theta \triangledown f(x) + f(x) \triangledown \theta \triangledown P(x) \right] + \frac{f(x) P(x) \Omega_0(x)}{2} \varepsilon^{d+2} + O(\varepsilon^{d+5}),$$

where the second equality holds by the same argument as that for $B_1$ and the fourth equality holds since $\int_{S^{d-1}} \theta \triangledown \theta \triangledown f(x) d\theta = \frac{|S^{d-1}|}{d-1} f(x) \in 3$.

For $D_2$, we only need to explicitly write down the $\varepsilon^{d+4}$ term. By the same argument as that for $B_1$, we have

$$D_2 = \int_{S^{d-1}} \int_0^2 \left[ \Pi_s(\theta, \theta) \right] \left[ P(x) \triangledown \theta \triangledown f(x) + \frac{1}{6} \triangledown^3 \theta \triangledown \triangledown f(x) + \frac{1}{2} \triangledown^2 \theta \triangledown f(x) \right] + \frac{1}{2} \Pi_s(\theta, \theta) \left[ P(x) \triangledown \theta \triangledown f(x) + \frac{1}{2} \triangledown^2 \theta \triangledown f(x) \right] \right] t^{d+3} dt d\theta,$$

which becomes

$$\frac{\varepsilon^{d+4}}{d+4} \int_{S^{d-1}} \left[ \Pi_s(\theta, \theta) \right] \left[ P(x) \triangledown \theta \triangledown f(x) + \frac{1}{6} \triangledown^3 \theta \triangledown \triangledown f(x) + \frac{1}{2} \triangledown^2 \theta \triangledown f(x) \right] + \frac{1}{2} \Pi_s(\theta, \theta) \left[ P(x) \triangledown \theta \triangledown f(x) + \frac{1}{2} \triangledown^2 \theta \triangledown f(x) \right] \right] t^{d+3} dt d\theta + O(\varepsilon^{d+6}).$$

We now simplify this complicated expression. The first term on the right hand side of $D_2$ becomes $|S^{d-1}| f(x) \in$
As a result, by putting the above together, when expressing which is in
\(\frac{1}{6}\text{Ric}_{\theta}(\theta,\theta) f(x)P(x)\) \(d\theta\)
\[= \nabla f(x)^T \int_{S^{d-1}} \Pi_{\theta}(\theta,\theta) \theta^T d\theta \nabla P(x) + \frac{1}{2} P(x) \nabla P(x) \]
\[+ \frac{1}{2} f(x) \nabla \left( \int_{S^{d-1}} \Pi_{\theta}(\theta,\theta) \theta^T d\theta \nabla^2 f(x) \right) \]
\[= \left| S^{d-1} \right| \left[ f(x) \nabla P(x) \right] + \frac{1}{2} P(x) \nabla P(x) \]
\[+ \frac{1}{2} f(x) \nabla \left( \nabla^2 f(x) \right) \]
\[+ \frac{1}{6} f(x) P(x) \text{Ric}_{\theta}(\theta,\theta) \]
\[= \left| S^{d-1} \right| \left[ f(x) \nabla P(x) \right] + \frac{1}{2} P(x) \nabla P(x) \]
\[+ \frac{1}{2} f(x) \nabla \left( \nabla^2 f(x) \right) \]
\[+ \frac{1}{6} f(x) P(x) \text{Ric}_{\theta}(\theta,\theta) \]
which is in \((t,T,M)^\perp\), where we use the equality \(u^T M v = \text{tr}(uv^T M)\), where \(M\) is a \(d \times d\) matrix and \(u,v \in \mathbb{R}^d\). For the third term on the right hand side of \(D_2\), it simply becomes
\[= \left| S^{d-1} \right| \left[ f(x) \nabla P(x) \right] + \frac{1}{2} P(x) \nabla P(x) \]
\[+ \frac{1}{2} f(x) \nabla \left( \nabla^2 f(x) \right) \]
\[+ \frac{1}{6} f(x) P(x) \text{Ric}_{\theta}(\theta,\theta) \]
which might or might not in \((t,T,M)^\perp\). Therefore, we have
\[D_2 = \left| S^{d-1} \right| \left[ f(x) \nabla P(x) \right] + \frac{1}{2} P(x) \nabla P(x) \]
\[+ \frac{1}{2} f(x) \nabla \left( \nabla^2 f(x) \right) \]
\[+ \frac{1}{6} f(x) P(x) \text{Ric}_{\theta}(\theta,\theta) \]
As a result, by putting the above together, when expressing \(E[(X - t(X)) f(X) \chi_{B_{R}(\{t(x)\})}] \) as \([v_1, v_2]\), we have
\[v_1 = \frac{1}{d+4} \left[ \int_{\mathbb{R}^d} (X - t(x)) f(X) \chi_{B_{R}(\{t(x)\})} \right] \]
\[= \frac{1}{d+2} \left[ \int_{\mathbb{R}^d} (X - t(x)) f(X) \chi_{B_{R}(\{t(x)\})} \right] \]
\[+ \frac{1}{24} \left[ \int_{\mathbb{R}^d} \left( \text{Ric}_{\theta}(\theta,\theta) \right) f(X) \right] \]
\[= \frac{1}{d+4} \left[ \int_{\mathbb{R}^d} (X - t(x)) f(X) \chi_{B_{R}(\{t(x)\})} \right] \]
\[+ \frac{1}{24} \left[ \int_{\mathbb{R}^d} \left( \text{Ric}_{\theta}(\theta,\theta) \right) f(X) \right] \]
and

\[ v_2 = \tilde{J}_{p,p-d}^T \mathbb{E}[(X - t(x))f(X)\chi_{\mathbb{R}^p}(t(x))(X)]\]

\[ = \frac{|S^{d-1}| f(x)P(x)\tilde{J}_{p,p-d}^T \mathcal{F}_0(x)}{2d+2} \varepsilon^{d+2} + \frac{|S^{d-1}| f(x)P(x)\tilde{J}_{p,p-d}^T \mathcal{F}_1(x)}{24} \varepsilon^{d+4} \]

\[ + \frac{|S^{d-1}| \tilde{J}_{p,p-d}^T}{2d+4} \left( \frac{1}{2} \nabla f(x)^T \mathcal{R}_2(x) \nabla P(x) + \frac{1}{4} \left[ P(x) \nabla \mathcal{R}_2(x) \nabla^2 f(x) + f(x) \nabla (\mathcal{R}_2(x) \nabla^2 P(x)) \right] \right) \]

\[ - \frac{1}{12} f(x)P(x)\mathcal{R}_2(x) \varepsilon^{d+4} + \frac{|S^{d-1}| \tilde{J}_{p,p-d}^T}{6(d+4)} \left[ P(x)\mathcal{R}_1(x) \nabla f(x) + f(x)\mathcal{R}_1(x) \nabla P(x) + \frac{1}{4} f(x)P(x)\mathcal{R}_2(x) \right] \varepsilon^{d+4} + O(\varepsilon^{d+5}). \]

\[ \square \]

### 2.3 Local covariance structure and local PCA

We call

\[ C_x := \mathbb{E}[(X - t(x))(X - t(x))^T \chi_{\mathbb{R}^p}(t(x))](X) \in \mathbb{R}^{p \times p} \]

the local covariance matrix at \( t(x) \in \mathcal{M} \), which is the covariance matrix associated with the local PCA [55, 17, 67, 12, 36, 41]. In the proof of LLE under the manifold setup, the eigen-structure of \( C_x \) plays an essential role due to its relationship with the barycentric coordinate. Geometrically, for a \( d \)-dim manifold, the first \( d \) eigenvectors of \( C_x \) corresponding to the largest \( d \) eigenvalues provide an estimated basis for the embedded tangent space \( T_x \mathcal{M} \), and the remaining eigenvectors form an estimated basis for the normal space at \( t(x) \). To be more precise, a smooth manifold can be well-approximated locally by an affine subspace. However, this approximation cannot be perfect, if the curvature exists. It is well-known that the contribution of curvature is of high order. For the purpose of fitting the manifold, we can ignore its contribution. For example, in [55, 17] the local PCA is applied to estimate the tangent space. However, in LLE, the curvature plays an essential role and a careful analysis is needed to understand its role. In Lemma 2.2.5, we show a generalization of the result shown in [55, 17] by expanding the \( C_x \) up to the third order for the sake of capturing LLE behavior. The third order term is needed for analyzing the regularization step shown in (1.2.5).

**Proposition 2.3.1.** Fix \( x \in \mathcal{M} \) and suppose Assumption 2.1.2 holds. When \( \varepsilon \) is sufficiently small, we have

\[ C_x = \frac{|S^{d-1}| P(x)}{d(d+2)} \varepsilon^{d+2} \left( \begin{bmatrix} L_{d \times d} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} M_{11}^{(2)} & M_{12}^{(2)} \\ M_{12}^{(2)} & M_{22}^{(2)} \end{bmatrix} \varepsilon^2 + \begin{bmatrix} M_{11}^{(4)} & M_{12}^{(4)} \\ M_{12}^{(4)} & M_{22}^{(4)} \end{bmatrix} \varepsilon^4 + O(\varepsilon^6) \right), \]

where \( M_{11}^{(2)}, M_{11}^{(4)} \in \mathbb{S}(d) \), \( M_{22}^{(2)}, M_{22}^{(4)} \in \mathbb{S}(p-d) \), \( M_{12}^{(2)} = M_{21}^{(2)} \), \( M_{12}^{(4)} = M_{21}^{(4)} \), and \( M_{12}^{(2)} = M_{21}^{(2)} \). These matrices are defined in (2.3.5), (2.3.7), (2.3.9), and (2.3.10). \( M_{22}^{(2)} \) depends on \( \Pi_x \) but does not depend on the p.d.f. \( P \), and \( M_{22}^{(4)} \) depends on the \( \Pi_x \) and its derivatives, the Ricci curvature, and \( P \).

**Proof.** We use the notation \( \langle \cdot, \cdot \rangle \) to mean the inner product and use the notation (2.2.8). The \((m,n)\)-th entry of \( C_x = \mathbb{E}[(X - t(x))(X - t(x))^T \chi_{\mathbb{R}^p}(t(x))](X) \) is

\[ e_m^T C_x e_n = \int_{\tilde{B}_e(x(t(y)) - t(x), e_m)} (t(y) - t(x), e_n) P(y) dV(y). \]
The quantities \( t \circ \exp_s(\theta t) \), \( \varepsilon \) and \( dV \) need to be expanded up to higher order terms. By the change of variable \( y = \exp_s(t \theta) \), where \((t, \theta) \in [0, \infty) \times S^{d-1} \) constitutes the polar coordinate, we have the following expressions:

\[
\begin{align*}
t \circ \exp_s(\theta t) - t(x) & = K_1(\theta)t + K_2(\theta)t^2 + K_3(\theta)t^3 + K_4(\theta)t^4 + K_5(\theta)t^5 + O(t^6) \\
\varepsilon & = \varepsilon + H_1(\theta)\varepsilon^3 + H_2(\theta)\varepsilon^4 + H_3(\theta)\varepsilon^5 + O(\varepsilon^6) \\
dV(\exp_s(t \theta)) & = t^{d-1} + R_1(\theta)t^{d+1} + R_2(\theta)t^{d+2} + R_3(\theta)t^{d+3} + O(t^{d+4}) \\
P(\exp_s(t \theta)) & = P_0 + P_1(\theta) + P_2(\theta)t^2 + P_3(\theta)t^3 + P_4(\theta)t^4 + O(t^5),
\end{align*}
\]

where

\[
\begin{align*}
K_1(\theta) & = t, \\
K_2(\theta) & = \frac{1}{2}\|\Pi_s(\theta, \theta)\|^2, \\
K_3(\theta) & = \frac{1}{6}\nabla_\theta \Pi_s(\theta, \theta), \\
K_4(\theta) & = \frac{1}{24}\nabla_{\theta \theta \theta} \Pi_s(\theta, \theta), \\
K_5(\theta) & = \frac{1}{120}\nabla_{\theta \theta \theta \theta} \Pi_s(\theta, \theta),
\end{align*}
\]

by Lemma (2.2.1),

\[
\begin{align*}
H_1(\theta) & = \frac{1}{24}\|\Pi_s(\theta, \theta)\|^2, \\
H_2(\theta) & = \frac{1}{24}\nabla_\theta \Pi_s(\theta, \theta) \cdot \Pi_s(\theta, \theta), \\
H_3(\theta) & = \frac{1}{80}\nabla_{\theta \theta} \Pi_s(\theta, \theta) \cdot \Pi_s(\theta, \theta) \\
& \quad + \frac{1}{90}\nabla_\theta \Pi_s(\theta, \theta) \cdot \nabla_\theta \Pi_s(\theta, \theta) + \frac{7}{1152}\|\Pi_s(\theta, \theta)\|^4,
\end{align*}
\]

by Lemma (2.2.2),

\[
\begin{align*}
R_1(\theta) & = -\frac{1}{6}\text{Ric}_s(\theta, \theta), \\
R_2(\theta) & = -\frac{1}{12}\nabla_\theta \text{Ric}_s(\theta, \theta), \\
R_3(\theta) & = -\frac{1}{40}\nabla_{\theta \theta \theta} \text{Ric}_s(\theta, \theta) \\
& \quad - \frac{1}{180}\sum_{a,b=1}^d \text{R}_s(\theta, a, \theta, b) \text{R}_s(\theta, a, \theta, b) + \frac{1}{72}\text{Ric}_s(\theta, \theta)^2,
\end{align*}
\]

by Lemma (2.2.3), and

\[
\begin{align*}
P_0 & := P(x), \\
P_1(\theta) & := \nabla_\theta P(x), \\
P_2(\theta) & := \frac{1}{2}\nabla_{\theta \theta} P(x), \\
P_3(\theta) & := \frac{1}{6}\nabla_{\theta \theta \theta} P(x), \\
P_4(\theta) & := \frac{1}{24}\nabla_{\theta \theta \theta \theta} P(x).
\end{align*}
\]

Note that \( H_1, H_2, R_1, R_3, P_0, P_2, \) and \( P_4 \) are even functions on \( S^{d-1} \) and \( H_2, R_2, P_1 \) and \( P_3 \) are odd on \( S^{d-1} \). Similarly, for \( m, n = 1, \ldots, p \), we have

\[
\begin{align*}
\langle t \circ \exp_s(\theta t) - t(x), e_m \rangle & \langle t \circ \exp_s(\theta t) - t(x), e_n \rangle \\
= & A_{m,n}(\theta)t^2 + B_{m,n}(\theta)t^3 + C_{m,n}(\theta)t^4 + D_{m,n}(\theta)t^5 + E_{m,n}(\theta)t^6 + O(t^7),
\end{align*}
\]
where

\[
A_{m,n}(\theta) = \langle K_1(\theta), e_m \rangle \langle K_1(\theta), e_n \rangle,
\]
\[
B_{m,n}(\theta) = \langle K_2(\theta), e_m \rangle \langle K_1(\theta), e_n \rangle + \langle K_1(\theta), e_m \rangle \langle K_2(\theta), e_n \rangle,
\]
\[
C_{m,n}(\theta) = \langle K_2(\theta), e_m \rangle \langle K_2(\theta), e_n \rangle + \langle K_1(\theta), e_m \rangle \langle K_3(\theta), e_n \rangle + \langle K_3(\theta), e_m \rangle \langle K_1(\theta), e_n \rangle,
\]
\[
D_{m,n}(\theta) = \langle K_1(\theta), e_m \rangle \langle K_4(\theta), e_n \rangle + \langle K_2(\theta), e_m \rangle \langle K_3(\theta), e_n \rangle + \langle K_3(\theta), e_m \rangle \langle K_1(\theta), e_n \rangle,
\]
\[
E_{m,n}(\theta) = \langle K_1(\theta), e_m \rangle \langle K_5(\theta), e_n \rangle + \langle K_2(\theta), e_m \rangle \langle K_4(\theta), e_n \rangle + \langle K_3(\theta), e_m \rangle \langle K_2(\theta), e_n \rangle + \langle K_4(\theta), e_m \rangle \langle K_1(\theta), e_n \rangle.
\]

Observe that \(A_{m,n}, C_{m,n}\) and \(E_{m,n}\), for \(m, n = 1, \ldots, p\), are even functions on \(S^{d-1}\), while \(B_{m,n}\) and \(D_{m,n}\) are odd functions on \(S^{d-1}\). But plugging these expressions into (2.3.2), we have

\[
e_m^T C_s e_n = \int_{S^{d-1}} \int_0^\varepsilon (A_{m,n}(\theta)t^2 + B_{m,n}(\theta)t^3 + C_{m,n}(\theta)t^4 + D_{m,n}(\theta)t^5 + E_{m,n}(\theta)t^6 + O(t^7))
\]
\[
\times \left( P_0 + P_1(\theta)t + P_2(\theta)t^2 + P_3(\theta)t^3 + P_4(\theta)t^4 + O(t^5) \right)
\]
\[
\times \left( t^{d-1} + R_1(\theta)t^{d+1} + R_2(\theta)t^{d+2} + R_3(\theta)t^{d+3} + O(t^{d+4}) \right) dt d\theta.
\]

We now collect terms of the same order to simplify the calculation. We focus on those terms with the order less than or equal to \(\varepsilon^{d+6}\).

\[
e_m^T C_s e_n = \int_{S^{d-1}} \int_0^\varepsilon P_0 A_{m,n}(\theta)t^{d+1} + \left( P_0 B_{m,n}(\theta) + P_1(\theta) A_{m,n}(\theta) \right) t^{d+2}
\]
\[
+ \left( P_0 A_{m,n}(\theta) R_1(\theta) + P_0 B_{m,n}(\theta) R_1(\theta) + P_1(\theta) A_{m,n}(\theta) R_1(\theta) + P_0 D_{m,n}(\theta) \right) t^{d+3}
\]
\[
+ \left( P_0 A_{m,n}(\theta) R_2(\theta) + P_0 B_{m,n}(\theta) R_2(\theta) + P_1(\theta) A_{m,n}(\theta) R_2(\theta) + P_0 D_{m,n}(\theta) \right) t^{d+4}
\]
\[
+ \left( P_0 A_{m,n}(\theta) R_3(\theta) + P_0 B_{m,n}(\theta) R_3(\theta) + P_1(\theta) A_{m,n}(\theta) R_3(\theta) + P_0 C_{m,n}(\theta) R_3(\theta) \right) t^{d+5}
\]
\[
+ \left( P_0 A_{m,n}(\theta) R_4(\theta) + P_0 B_{m,n}(\theta) R_4(\theta) + P_1(\theta) A_{m,n}(\theta) R_4(\theta) + P_0 D_{m,n}(\theta) \right) t^{d+6}
\]
\[
+ \left( P_0 A_{m,n}(\theta) R_5(\theta) + P_0 B_{m,n}(\theta) R_5(\theta) + P_1(\theta) A_{m,n}(\theta) R_5(\theta) + P_0 C_{m,n}(\theta) R_5(\theta) \right) t^{d+7}
\]
\[
+ O(\varepsilon^{d+6})
\]

By further expanding the integration of \(t\) over \([0, \varepsilon]\) = \([0, \varepsilon + H_1(\theta)\varepsilon^3 + H_2(\theta)\varepsilon^4 + H_3(\theta)\varepsilon^5 + O(\varepsilon^6)]\) by Lemma 2.2.4, we have

\[
e_m^T C_s e_n = \varepsilon^{d+2} Q_m^{(0)}(x) + \varepsilon^{d+4} Q_m^{(2)}(x) + \varepsilon^{d+6} Q_m^{(4)}(x) + O(\varepsilon^{d+7}),
\]
where

\[ Q_{m,n}^{(0)}(x) = P_0 \int_{S^{d-1}} \frac{A_{m,n}(\theta)}{d+2} d\theta, \]

\[ Q_{m,n}^{(2)}(x) = \int_{S^{d-1}} \left( P_0 A_{m,n}(\theta) H_1(\theta) + \frac{1}{d+4} \left[ P_0 A_{m,n}(\theta) R_1(\theta) + P_0 C_{m,n}(\theta) \right] \right) d\theta, \]

\[ + P_1(\theta) B_{m,n}(\theta) + P_2(\theta) A_{m,n}(\theta) \]

and

\[ Q_{m,n}^{(4)}(x) = \int_{S^{d-1}} \left( P_0 A_{m,n}(\theta) H_3(\theta) + \frac{d+1}{2} P_0 A_{m,n}(\theta) H_2^2(\theta) \right) d\theta, \]

\[ + \left[ P_0 B_{m,n}(\theta) + P_1(\theta) A_{m,n}(\theta) \right] H_2(\theta) \]

\[ + \left[ P_0 A_{m,n}(\theta) R_1(\theta) + P_1(\theta) C_{m,n}(\theta) \right] H_1(\theta) \]

\[ + \frac{1}{d+6} \left( P_0 A_{m,n}(\theta) R_3(\theta) + P_0 B_{m,n}(\theta) R_2(\theta) + P_1(\theta) A_{m,n}(\theta) R_2(\theta) + P_1(\theta) C_{m,n}(\theta) R_1(\theta) \right) \]

\[ + P_1(\theta) B_{m,n}(\theta) R_1(\theta) + P_2(\theta) A_{m,n}(\theta) R_1(\theta) + P_2(\theta) C_{m,n}(\theta) \]

\[ + P_1(\theta) B_{m,n}(\theta) + P_3(\theta) B_{m,n}(\theta) + P_4(\theta) A_{m,n}(\theta) \right) d\theta. \]

To finish the proof, we evaluate \( Q_{m,n}^{(0)}, Q_{m,n}^{(2)}, \) and \( Q_{m,n}^{(4)} \), for \( 1 \leq m, n \leq p \). Due to Assumptions 2.1.2 and 2.3.1, \( \{e_1, \ldots, e_d\} \) is an orthonormal basis of \( t, T_s M \) and \( \{e_{d+1}, \ldots, e_p\} \) is an orthonormal basis of \( (t, T_s M)^\perp \). There, we have \( \langle K_i(\theta), e_i \rangle = \langle t, e_i \rangle = 0 \) for \( i = d+1, \ldots, p \). Using Lemma 2.2.2 and symmetry of sphere, we can evaluate the term of order \( d+2 \) in \( C_x \). For \( 1 \leq m = n \leq d \), we have

\[ Q_{m,n}^{(0)} = \frac{P_0}{d+2} \int_{S^{d-1}} A_{m,n}(\theta) d\theta = \frac{P(x)}{d+2} \int_{S^{d-1}} |\langle t, e_1 \rangle|^2 d\theta = \frac{|S^{d-1}|P(x)}{d(d+2)}; \]  

(2.3.3)

for other \( m \) and \( n \), \( \int_{S^{d-1}} A_{m,n}(\theta) d\theta = 0 \). Thus, the coefficient of the \( e^{d+2} \) term is

\[ \frac{|S^{d-1}|P(x)}{d(d+2)} \begin{bmatrix} I_{d \times d} & 0 \\ 0 & 0 \end{bmatrix}. \]

Denote

\[ M_{11}^{(0)} = I_{d \times d} \in \mathbb{R}^{d \times d} \quad \text{and} \quad M_{12}^{(0)} = 0 \in \mathbb{R}^{d \times (p-d)}, \quad M_{21}^{(0)} = M_{22}^{(0)\top} \quad \text{and} \quad M_{22}^{(0)} = 0 \in \mathbb{R}^{(p-d) \times (p-d)}. \]

Next, we evaluate the term of order \( d+4 \) in \( C_x \). Note that \( \langle \Pi_x(\theta, \theta), e_m \rangle = 0 \), for \( m = 1, \ldots, d \), so \( B_{m,n}(\theta) = 0 \). Thus, for \( 1 \leq m, n \leq d \), by a direct calculation,

\[ Q_{m,n}^{(4)} = \frac{P(x)}{24} \int_{S^{d-1}} \langle t, \theta, e_m \rangle \langle t, \theta, e_n \rangle \| \Pi_x(\theta, \theta) \| d\theta \]

\[ - \frac{P(x)}{6(d+4)} \int_{S^{d-1}} \langle t, \theta, e_1 \rangle \langle t, \theta, e_n \rangle \Re c_x(\theta, \theta) d\theta \]

\[ - \frac{P(x)}{6(d+4)} \int_{S^{d-1}} \langle t, \theta, e_1 \rangle \langle \Pi_x(e_1, \theta), \Pi_x(\theta, \theta) \rangle + \langle t, \theta, e_n \rangle \langle \Pi_x(e_n, \theta), \Pi_x(\theta, \theta) \rangle d\theta \]

\[ + \frac{1}{2(d+4)} \int_{S^{d-1}} \nabla_{\theta, \theta}^2 \Pi_x(\theta, \theta) \| P(x) \| \langle t, \theta, e_m \rangle \langle t, \theta, e_n \rangle d\theta, \]

where we use the fact that \( \langle \nabla_{\theta} \Pi_x(\theta, \theta), e_m \rangle = -\langle \Pi_x(e_m, \theta), \Pi_x(\theta, \theta) \rangle \) when \( m = 1, \ldots, d \). By definition, it is clear
that $A_{m,n}(\theta) = 0$ when $1 \leq m \leq d$ and $d+1 \leq n \leq p$. Thus, for $1 \leq m \leq d$ and $d+1 \leq n \leq p$, we have

$$Q_{m,n}^{(2)} = \frac{P(x)}{6(d+4)} \int_{S^{d-1}} \langle \tau, \theta, e_m \rangle \langle \nabla_\theta I_\tau(\theta, \theta), e_n \rangle d\theta$$

$$+ \frac{1}{2(d+4)} \int_{S^{d-1}} \nabla_\theta P(x) \langle \tau, \theta, e_m \rangle \langle I_\tau(\theta, \theta), e_n \rangle d\theta. \tag{2.3.5}$$

By definition, for $d+1 \leq m, n \leq p$, $A_{m,n}(\theta) = B_{m,n}(\theta) = 0$, and hence

$$Q_{m,n}^{(2)} = \frac{P(x)}{4(d+4)} \int_{S^{d-1}} \langle I_\tau(\theta, \theta), e_m \rangle \langle I_\tau(\theta, \theta), e_n \rangle d\theta. \tag{2.3.6}$$

Finally, we evaluate the $\varepsilon^{d+6}$ term. Again, recall the fact that when $d+1 \leq m, n \leq p$, $A_{m,n}(\theta) = 0$ and $B_{m,n}(\theta) = 0$. Therefore, $Q_{m,n}^{(4)}$, where $d+1 \leq m, n \leq p$, consists of only

$$Q_{m,n}^{(4)} = \int_{S^{d-1}} P_0 C_{m,n}(\theta) H_1(\theta)$$

$$+ \frac{1}{d+6} \left( P_0 C_{m,n}(\theta) R_1(\theta) + P_0 E_{m,n}(\theta) + P_1(\theta) D_{m,n}(\theta) + P_2(\theta) C_{m,n}(\theta) \right) d\theta.$$ 

Based on Lemmas 2.2.1, 2.2.2 and 2.2.3, for $d+1 \leq m, n \leq p$, we have

$$Q_{m,n}^{(4)} = \frac{P(x)}{96} \int_{S^{d-1}} \langle I_\tau(\theta, \theta), e_m \rangle \langle I_\tau(\theta, \theta), e_n \rangle \| I_\tau(\theta, \theta) \|^2 d\theta$$

$$- \frac{P(x)}{24(d+6)} \int_{S^{d-1}} \langle I_\tau(\theta, \theta), e_m \rangle \langle I_\tau(\theta, \theta), e_n \rangle \text{Re} c_\tau(\theta, \theta) d\theta$$

$$+ \frac{P(x)}{48(d+6)} \int_{S^{d-1}} \langle I_\tau(\theta, \theta), e_m \rangle \langle \nabla^2_\theta I_\tau(\theta, \theta), e_n \rangle$$

$$+ \langle \nabla^2_\theta I_\tau(\theta, \theta), e_m \rangle \langle I_\tau(\theta, \theta), e_n \rangle d\theta$$

$$+ \frac{P(x)}{36(d+6)} \int_{S^{d-1}} (\nabla_\theta P(x) \langle I_\tau(\theta, \theta), e_m \rangle \langle \nabla_\theta I_\tau(\theta, \theta), e_n \rangle$$

$$+ \langle \nabla_\theta I_\tau(\theta, \theta), e_m \rangle \langle I_\tau(\theta, \theta), e_n \rangle \rangle d\theta$$

$$+ \frac{1}{12(d+6)} \int_{S^{d-1}} \nabla_\theta P(x) \langle I_\tau(\theta, \theta), e_m \rangle \langle \nabla_\theta I_\tau(\theta, \theta), e_n \rangle d\theta$$

$$+ \frac{1}{4(d+6)} \int_{S^{d-1}} \nabla^2_\theta P(x) \langle I_\tau(\theta, \theta), e_m \rangle \langle I_\tau(\theta, \theta), e_n \rangle d\theta.$$ 

Since we only need to evaluate $Q_{m,n}^{(4)}$, where $d+1 \leq m, n \leq p$, for LLE analysis, we omit the calculation of the other pairs of $m,n$. We thus conclude that

$$C_x = \varepsilon^{d+2} \frac{\| S^{d-1} P(x) \|}{d(d+2)} \left( I_{d \times d} 0 \begin{bmatrix} M_1^{(2)} & M_2^{(2)} \\ M_2^{(2)} & M_2^{(2)} \end{bmatrix} \varepsilon^2 + \begin{bmatrix} M_1^{(4)} & M_2^{(4)} \\ M_2^{(4)} & M_2^{(4)} \end{bmatrix} \varepsilon^4 + O(\varepsilon^6) \right), \tag{2.3.8}$$

where $M^{(j)}_{ij} \in \mathbb{R}^{d \times d}$ is defined as

$$e_m^T M^{(j)}_{11} e_n = \frac{d(d+2)}{\| S^{d-1} P(x) \|} Q_{m,n}^{(j)}, \tag{2.3.9}$$
for \( m, n = 1, \ldots, d \) and \( j = 2, 4 \), \( M_{22}^{(j)} \in \mathbb{R}^{(p-d) \times (p-d)} \) is defined as
\[
e_m^T M_{22}^{(j)} e_n = \frac{d(d+2)}{|S^{d-1}|P(x)} Q_{m,n}^{(j)}
\]
for \( m, n = 1, \ldots, p-d \) and \( j = 2, 4 \), \( M_{12}^{(2)} \in \mathbb{R}^{d \times (p-d)} \) is defined as
\[
e_m^T M_{12}^{(2)} e_n = \frac{d(d+2)}{|S^{d-1}|P(x)} Q_{m,n}^{(2)}
\]
\[= \frac{d(d+2)}{6|S^{d-1}|(d+4)} \int_{S^{d-1}} \langle 1, \theta, e_m \rangle \langle \nabla_\theta \Pi_x(\theta, \theta), e_n \rangle d\theta\]
\[+ \frac{d(d+2)}{2(d+4)|S^{d-1}|P(x)} \int_{S^{d-1}} \nabla_\theta P(x) \langle 1, \theta, e_m \rangle \langle \Pi_x(\theta, \theta), e_n \rangle d\theta\]
for \( m = 1, \ldots, d \) and \( n = 1, \ldots, p-d \), and \( M_{21}^{(2)} = M_{12}^{(2)\top} \). 

Since \( P \) is bounded by \( P_m \) from below, when \( \varepsilon \) is sufficiently small, the \( \varepsilon^{d+2} \) term is dominant and the largest \( d \) eigenvalues of \( C_x \) are of order \( \varepsilon^{d+2} \). The other eigenvalues of \( C_x \) are of higher order and depend on the \( \varepsilon^{d+4} \) term or even the \( \varepsilon^{d+6} \) term. The behavior of eigenvectors is more complicated, due to the possible multiplicity of the corresponding eigenvalues.

To simplify the statement of the eigen-structure, following Assumption 2.1.2, we make one more assumption.

**Assumption 2.3.1.** Following Assumption 2.1.2, without loss of generality, we assume that the manifold is translated and rotated properly, so that \( e_d, \ldots, e_p \) “diagonalize” the second fundamental form; that is, \( M_{22}^{(2)} \) in Proposition 2.3.1 is diagonalized to \( \Lambda_2^{(2)} = \text{diag}(\lambda_{d+1}^{(2)}, \ldots, \lambda_p^{(2)}) \).

The eigen-structure of the local covariance matrix is summarized in the following Proposition.

**Proposition 2.3.2.** Fix \( x \in M \). Suppose \( \varepsilon \) is sufficiently small and Assumptions 2.1.2 and 2.3.1 hold. Let \( r = \text{rank}(C_x) \). The eigen-decomposition of \( C_x = U_x \Lambda_x U_x^\top \), where \( U_x \in O(p) \) and \( \Lambda_x \in \mathbb{R}^{p \times p} \) is a diagonal matrix, is summarized below.

**Case 0:** When \( r = d \), we have
\[
C_x := \frac{|S^{d-1}|P(x)\varepsilon^{d+2}}{d(d+2)} \left( \begin{bmatrix} I_{d \times d} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} O(\varepsilon^2) & 0 \\ 0 & 0 \end{bmatrix} \right)
\]
and
\[
\Lambda_x = \frac{|S^{d-1}|P(x)\varepsilon^{d+2}}{d(d+2)} \left[ I_{d \times d} + O(\varepsilon^2) \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right] + O(\varepsilon^4),
\]
\[
U_x(\varepsilon) = U_x(0)(I_{p \times p} + \varepsilon^2 S) + O(\varepsilon^4) \in O(p),
\]
where \( U_x(0) = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \), \( X_1 \in O(d) \), \( X_2 \in O(p-d) \), and \( S \in o(p) \).

**Case 1:** When all diagonal entries of \( \Lambda_x^{(2)} \) are nonzero, we have:
\[ \Lambda_{x} = \frac{|S^{d-1}|p(x)e^{d+2}}{d(d+2)} \begin{bmatrix} I_{d\times d} + \epsilon^2 \Lambda_{1}^{(2)} + \epsilon^4 \Lambda_{1}^{(4)} & \mathbf{0} \\ \mathbf{0} & \epsilon^2 \Lambda_{2}^{(2)} + \epsilon^4 \Lambda_{2}^{(4)} \end{bmatrix} + O(\epsilon^6), \]

\[ U_x = U_x(0)(I_{p \times p} + \epsilon^2 S) + O(\epsilon^4) \in O(p), \]

where \( \Lambda_{1}^{(2)}, \Lambda_{1}^{(4)} \in \mathbb{R}^{d \times d} \) and \( \Lambda_{2}^{(2)}, \Lambda_{2}^{(4)} \in \mathbb{R}^{(p-d) \times (p-d)} \) are diagonal matrices with diagonal entries of order 1, \( U_x(0) = \begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix} \in O(p) \), \( x_1 \in O(d) \), \( x_2 \in O(p-d) \), and \( S \in o(p) \). The explicit expression of these matrices are listed in (2.3.14)-(2.3.21).

**Case 2:** When 1 diagonal entries for \( \Lambda_{2}^{(2)} \) are 0, where 1 \( \leq l \leq p-d \), we have the following eigen-decomposition under some conditions. Divide \( C_x \) into blocks corresponding to the multiplicity \( l \) as

\[ C_x = \frac{|S^{d-1}|p(x)e^{d+2}}{d(d+2)} \begin{bmatrix} I_{d \times d} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} M_{11}^{(2)} & M_{12}^{(2)} & M_{12}^{(2)} \\ M_{21}^{(2)} & M_{21}^{(2)} & M_{21}^{(2)} \\ M_{22}^{(2)} & M_{22}^{(2)} & M_{22}^{(2)} \end{bmatrix} \epsilon^2 + O(\epsilon^6), \]

(2.3.12)

where \( M_{11}^{(2)}, M_{12}^{(2)} \in \mathbb{R}^{d \times (p-d-1)} \), \( M_{21}^{(2)}, M_{22}^{(2)} \in \mathbb{R}^{d \times l} \), \( M_{12}^{(2)} = M_{21}^{(2)} \), \( M_{12}^{(2)} = M_{21}^{(2)} \), \( M_{12}^{(2)} = M_{21}^{(2)} \), \( M_{12}^{(2)} = M_{21}^{(2)} \), \( M_{12}^{(2)} = M_{21}^{(2)} \), \( M_{12}^{(2)} = M_{21}^{(2)} \), \( M_{12}^{(2)} = M_{21}^{(2)} \)

Denote the eigen-decomposition of the matrix \( M_{22}^{(4)} - 2M_{12}^{(2)}M_{12}^{(2)} \) as

\[ M_{22}^{(4)} - 2M_{12}^{(2)}M_{12}^{(2)} = U_{22}\Lambda_{22}^{(4)}U_{22}^T, \]

(2.3.13)

where \( U_{22} \in O(l) \) and \( \Lambda_{22}^{(4)} = \text{diag}[\lambda_{p-l+1}^{(4)}, \ldots, \lambda_{p}^{(4)}] \) is a diagonal matrix. If we further assume that all diagonal entries of \( \Lambda_{22}^{(4)} \) are nonzero, we have

\[ \Lambda_{x} = \frac{|S^{d-1}|p(x)e^{d+2}}{d(d+2)} \begin{bmatrix} I_{d \times d} + \epsilon^2 \Lambda_{1}^{(2)} + \epsilon^4 \Lambda_{1}^{(4)} & \mathbf{0} \\ \mathbf{0} & \epsilon^2 \Lambda_{2,1}^{(2)} + \epsilon^4 \Lambda_{2,1}^{(4)} \end{bmatrix} + O(\epsilon^6), \]

\[ U_x = U_x(0)(I_{p \times p} + \epsilon^2 S) + O(\epsilon^4) \in O(p), \]

where \( \Lambda_{1}^{(4)} \) and \( \Lambda_{2,1}^{(4)} \) are diagonal matrices, \( U_{22}(0) = \begin{bmatrix} x_1 & 0 & 0 \\ 0 & x_{2,1} & 0 \\ 0 & 0 & x_{2,2} \end{bmatrix} \in O(p) \), \( x_1 \in O(d) \), \( x_{2,1} \in O(p-d-l) \), \( x_{2,2} \in O(l) \), and \( S \in o(p) \). The explicit formula for these matrices are listed in (2.3.22)-(2.3.24).

**Proof.** We now evaluate the eigenvalue and eigenvectors of \( C_x \) shown in (2.3.8) based on the technique introduced in Section 1.4.
For Case 0, when \( \varepsilon \) is sufficiently small, we have
\[
C_\varepsilon := \frac{|S^{d-1}|P(x)\varepsilon^{d+2}}{d(d+2)} \left( \begin{bmatrix} I_{d \times d} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} O(\varepsilon^2) & 0 \\ 0 & 0 \end{bmatrix} \right)
\]
and hence the \( d \) non-zero eigenvalues satisfies
\[
\Lambda_\varepsilon = \frac{|S^{d-1}|P(x)\varepsilon^{d+2}}{d(d+2)} \begin{bmatrix} I_{d \times d} + O(\varepsilon^2) & 0 \\ 0 & 0 \end{bmatrix} + O(\varepsilon^4),
\]
\[
U_\varepsilon(\varepsilon) = U_\varepsilon(0)(I_{p \times p} + \varepsilon^2 S) + O(\varepsilon^4) \in O(p),
\]
where \( U_\varepsilon(0) = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}, X_1 \in O(d), X_2 \in O(p - d), \) and \( S \in o(p) \). Note that in this case \( S, X_1 \) and \( X_2 \) cannot be uniquely determined by the order \( \varepsilon^{d+2} \) part of \( C_\varepsilon \).

For Case 1, when \( \varepsilon \) is sufficiently small, we have
\[
\Lambda_\varepsilon = \frac{|S^{d-1}|P(x)\varepsilon^{d+2}}{d(d+2)} \begin{bmatrix} I_{d \times d} + \varepsilon^2 \Lambda_1^{(2)} + \varepsilon^4 \Lambda_1^{(4)} & 0 \\ 0 & \varepsilon^2 \Lambda_2^{(2)} + \varepsilon^4 \Lambda_2^{(4)} \end{bmatrix} + O(\varepsilon^3),
\]
\[
U_\varepsilon(\varepsilon) = U_\varepsilon(0)(I_{p \times p} + \varepsilon^2 S) + O(\varepsilon^4) \in O(p),
\]
where \( U_\varepsilon(0) = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}, X_1 \in O(d), \) and \( X_2 \in O(p - d), \) and
\[
S := \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \in o(p). \tag{2.3.14}
\]

Since \( M_{22}^{(2)} \) is a diagonal matrix under Assumptions 2.1.2 and 2.3.1, (1.4.4) implies that it is \( M_{22}^{(2)} = \Lambda_2^{(2)} \). From (1.4.7) and (1.4.8), we have
\[
S_{12} = -X_1^T M_{12}^{(2)} X_2, \tag{2.3.15}
\]
\[
S_{21} = X_2^T M_{21}^{(2)} X_1. \tag{2.3.16}
\]

If all eigenvalues of \( M_{11}^{(2)} \) are distinct, then \( X_1 \) could be uniquely determined; if all eigenvalues of \( M_{22}^{(2)} \) are distinct, since it is a diagonal matrix, \( X_2 \) is identity matrix. Moreover, \( \Lambda_1', \Lambda_2'^n \) and \( S \) can be uniquely determined:
\[
\Lambda_1' = \text{diag}(X_1^T M_{11}^{(2)} X_1 + 2X_1^T M_{12}^{(2)} M_{21}^{(2)} X_1), \tag{2.3.17}
\]
\[
\Lambda_2' = \text{diag}(M_{22}^{(2)} - 2M_{21}^{(2)} M_{21}^{(2)}), \tag{2.3.18}
\]
\[
(S_{11})_{m,n} = -\frac{1}{(\Lambda_1^{(2)})_{m,m} - (\Lambda_1^{(2)})_{n,n}} e_m^T (X_1^T M_{11}^{(2)} X_1 + 2X_1^T M_{12}^{(2)} M_{21}^{(2)} X_1) e_n, \tag{2.3.19}
\]
\[
(S_{22})_{i,j} = -\frac{1}{(\Lambda_2^{(2)})_{i,i} - (\Lambda_2^{(2)})_{j,j}} e_i^T (M_{22}^{(2)} - 2M_{21}^{(2)} M_{21}^{(2)}) e_j, \tag{2.3.20}
\]
where \( 1 \leq m \neq n \leq d \) and \( 1 \leq i \neq j \leq p - d \). On the other hand, if \( M_{22}^{(2)} \) has \( q + t \) distinct eigenvalues, where
\( q, t \geq 0 \), and \( q \) eigenvalues are simple, then based on the perturbation theory in the introduction, we have

\[
X_2 = \begin{bmatrix}
I_{q \times q} & 0 & \cdots & 0 \\
0 & X_1^1 & \cdots & 0 \\
0 & 0 & \ddots & 0 \\
0 & 0 & \cdots & X_2^1
\end{bmatrix}
\tag{2.3.21}
\]

since \( M_{22}^{(2)} \) is diagonal under Assumption 2.3.1. Each of \( X_1^1 \cdots X_2^1 \) corresponds to a repeated eigenvalue, and each of them is an orthogonal matrix whose dimension depends on the multiplicity of the repeated eigenvalue. We mention that they may be uniquely determined by higher order terms in \( C_t \) as described in the perturbation theory in the introduction.

For Case 2, when \( \epsilon \) is sufficiently small, by dividing all matrices into blocks of the same size, we have

\[
\Lambda_2 = \frac{|S_{d-1}P(x)|}{d(d+2)} \left[ I_{d \times d} + \epsilon^2 \Lambda_4^{(2)} + \epsilon^4 \Lambda_4^{(4)} \right] + O(\epsilon^6),
\]

\[
U_2(\epsilon) = U_2(0)(I_{p \times p} + \epsilon^2 S) + O(\epsilon^4) \in O(p),
\tag{2.3.22}
\]

\[
U_2(0) = \begin{bmatrix}
X_1 & 0 & 0 \\
0 & X_{2,1} & 0 \\
0 & 0 & X_{2,2}
\end{bmatrix} \in O(p), \quad S = \begin{bmatrix}
S_{11} & S_{12,1} & S_{12,2} \\
S_{21,1} & S_{22,1} & S_{22,2} \\
S_{21,2} & S_{22,1} & S_{22,2}
\end{bmatrix} \in O(p),
\]

by Assumptions 2.1.2 and 2.3.1, where \( \Lambda_4^{(2)} \) is the eigenvalue matrix of \( M_{11}^{(2)} \), diagonal entries of \( \Lambda_2^{(2)} \) are nonzero, \( X_1 \in O(d) \), \( X_{2,1} \in O(p-d-l) \) and \( X_{2,2} \in O(l) \). By (1.4.7) and (1.4.8), we have

\[
S_{12,1} = -X_1^T M_{12,1}^{(2)} X_{2,1},
\]

\[
S_{21,1} = X_{2,1}^T M_{21,1}^{(2)} X_1,
\]

\[
S_{12,2} = -X_1^T M_{12,2}^{(2)} X_{2,2},
\tag{2.3.23}
\]

\[
S_{21,2} = X_{2,2}^T M_{21,2}^{(2)} X_1.
\]

If the eigenvalues of \( M_{11}^{(2)} \) are distinct, then \( X_1 \) is the corresponding orthonormal eigenvector matrix. \( \Lambda_2^{(2)} = 0 \) by the assumption of Case 2. Recall that \( \Lambda_2^{(4)} \) is the eigenvalue matrix of \( M_{22,2}^{(2)} - 2 \frac{d(d+2)}{|S_{d-1}P(x)|} M_{21,2}^{(2)} M_{12,2}^{(2)} \). If \( \Lambda_2^{(4)} \) has different diagonal entries then \( X_{2,2} \) is the corresponding orthonormal eigenvector matrix. Recall that if \( \Lambda_1^{(2)}, \Lambda_2^{(2)} \)
and \( \Lambda_{2,2}^{(4)} \), each has distinct diagonal entries, then \( X(0) \) and \( S \) can be determined uniquely, and we have

\[
S_{22,12} = (-\Lambda_{2,2}^{(4)})^{-1}\left(\frac{1}{2}M_{22,12}^{(4)}X_{2,2} + M_{21,1}^{(2)}X_{1,1}S_{12,2}\right),
\]

\[
S_{22,21} = X_{2,2}^\top\left(\frac{1}{2}M_{22,21}^{(4)} + M_{21,1}^{(2)}X_{1,1}S_{12,1}\right)(\Lambda_{2,1}^{(2)})^{-1},
\]

\[
\Lambda_{1,1}^{(4)} = \text{diag}\left[\left(M_{11}^{(4)}X_1 + 2M_{12,1}^{(2)}S_{21,1} + 2M_{12,2}^{(2)}X_{2,2}S_{22,1}\right)\right],
\]

\[
\Lambda_{2,1}^{(4)} = \text{diag}\left[\left(M_{22,11}^{(4)} + 2M_{21,1}^{(2)}X_{1,1}S_{12,1}\right)\right],
\]

\[
(S_{11})_{m,n} = \frac{-1}{(\Lambda_{2,1}^{(2)})_{m,m} - (\Lambda_{1,1}^{(2)})_{n,n}} e_n^\top \left(\frac{1}{2}M_{11}^{(4)}X_1 + M_{12,1}^{(2)}S_{21,1} + M_{12,2}^{(2)}X_{2,2}S_{22,1}\right) e_n,
\]

where \( 1 \leq m \neq n \leq d \), and

\[
(S_{22,11})_{m,n} = \frac{-1}{(\Lambda_{2,1}^{(2)})_{m,m} - (\Lambda_{1,1}^{(2)})_{n,n}} e_m^\top \left[\left(M_{22,11}^{(4)} + M_{21,1}^{(2)}X_{1,1}S_{12,1}\right)\right] e_n.
\]  

(2.3.24)

where \( d + 1 \leq m \neq n \leq p - 1 \). However, we need higher order derivative of \( C \) to solve \( S_{22,22} \) following the same step as evaluating (1.4.41). We skip the details here. Finally, if diagonal entries of \( \Lambda_{2,1}^{(2)} \) are distinct, then \( X_{2,1} \) is the identity matrix. If \( \Lambda_{2,1}^{(2)} \) or \( \Lambda_{2,2}^{(4)} \) contains repeated eigenvalues, then it can be described as (2.3.21). We also skip the details here.

In general, the eigen-structure of \( C \) may be more complicated than the two cases considered in Proposition 2.3.2. In this general case, we could apply the same perturbation theory to evaluate the eigenvalues. Since the proof is similar but there is extensive notational loading, and it does not bring further insight to LLE, we skip details of these more general situations.

### 2.4 Integral kernel of LLE and variance analysis on closed manifolds

We now study the asymptotic behavior of LLE. Under the manifold setup, from now on, we fix

\[
c = ne^{d+p},
\]  

(2.4.1)

and we call \( p \) the regularization order. By (1.3.13), for \( \mathbf{v} \in \mathbb{R}^N \), we have

\[
\sum_{j=1}^{N} w_k(j) \mathbf{v}(j) = \frac{1}{N} \mathbf{v} - \frac{1}{N} G_n^\top \mathcal{J}_{ne^{d+p}}(G_n G_n^\top) G_n \mathbf{v} - \frac{1}{N} G_n^\top \mathcal{J}_{ne^{d+p}}(G_n G_n^\top) G_n 1_N.
\]  

(2.4.2)

Before proceeding, we provide a geometric interpretation of this formula. By the eigen-decomposition \( G_n G_n^\top = U_n \tilde{\Lambda}_n U_n^\top \) and the fact that \( \mathcal{J}_{ne^{d+p}}(G_n G_n^\top) = U_n J_{p,d+p}^{\top} (\tilde{\Lambda}_n + ne^{d+p} I_{p \times p})^{-1} J_{p,d+p} U_n^\top = U_n \mathcal{J}_{ne^{d+p}}(\tilde{\Lambda}_n) U_n^\top \) by the definition of \( \mathcal{J}_{p} \) in (1.3.9), we have

\[
1_N^\top G_n^\top \mathcal{J}_{ne^{d+p}}(G_n G_n^\top) G_n \mathbf{v} = 1_N^\top G_n^\top \mathcal{J}_{ne^{d+p}}(\tilde{\Lambda}_n) U_n^\top G_n \mathbf{v}
\]

and

\[
1_N^\top G_n^\top \mathcal{J}_{ne^{d+p}}(G_n G_n^\top) G_n 1_N = 1_N^\top G_n^\top U_n \mathcal{J}_{ne^{d+p}}(\tilde{\Lambda}_n) U_n^\top G_n 1_N.\]

By the discussion of the local PCA in Section 2.3, \( U_n^\top G_n \) means evaluating the coordinates of all neighboring points of \( t(x_i) \) with the basis composed of the column vectors of \( U_n \), \( U_n^\top G_n 1 \) means the mean coordinate of all neighboring points, \( \mathcal{J}_{ne^{d+p}}(\tilde{\Lambda}_n) \) means a regularized weighting of the coordinates that helps to enhance the nonlinear geometry of the point cloud, and \( G_n^\top U_n \mathcal{J}_{ne^{d+p}}(\tilde{\Lambda}_n) U_n^\top G_n \) is a quadratic form of the averaged coordinates of all neighboring points. We could thus view the “kernel” part, \( 1_N^\top G_n^\top U_n \mathcal{J}_{ne^{d+p}}(\tilde{\Lambda}_n) U_n^\top G_n \), as preserving the geometry of the point cloud, by evaluating how strongly the weighted coordinates of neighboring
points are related to the mean coordinate of all neighboring points by the inner product.

Asymptotically, by the law of large numbers, when conditional on \( t(x_k) \),

\[
\frac{1}{n} G_n 1_N = \frac{1}{n} \sum_{j=1}^{N} (t(x_{k,j}) - t(x_k)) \xrightarrow{n \to \infty} \mathbb{E}[(X - t(x_k))X_{B^{E^p}_{t(x_k)}}(X)]
\]

and we “expect” the following holds

\[
n \mathcal{F}_{ne^{d+p}}(G_n G_n^\top) = \mathcal{F}_{e^{d+p}}(\frac{1}{n} G_n G_n^\top) \xrightarrow{n \to \infty} \mathcal{F}_{e^{d+p}}(C_{x_k}).
\]

Also, we would “expect” to have

\[
n \mathcal{F}_{ne^{d+p}}(G_n G_n^\top) \frac{1}{n} G_n 1_N \xrightarrow{n \to \infty} \mathcal{F}_{e^{d+p}}(C_{x_k}) \left[ \mathbb{E}(X - t(x_k))X_{B^{E^p}_{t(x_k)}} \right] =: T_{t(x_k)}.
\]

Hence, for \( f \in C(t(M)) \), for \( t(x_k) \) and its corresponding \( \mathcal{N}_{t(x_k)} \), we would “expect” to have

\[
\sum_{j=1}^{N} w_n(j) f(x_{k,j}) \xrightarrow{n \to \infty} \frac{\mathbb{E}[X_{B^{E^p}_{t(x_k)}}(X)f(X)] - T_{t(x_k)} \mathbb{E}[(X - t(x_k))X_{B^{E^p}_{t(x_k)}}(X)f(X)]}{\mathbb{E}[X_{B^{E^p}_{t(x_k)}}(X)] - T_{t(x_k)} \mathbb{E}[(X - t(x_k))X_{B^{E^p}_{t(x_k)}}(X)]}
\]

\[
= \frac{\mathbb{E}[f(X)(1 - T_{t(x_k)}(X - t(x)))X_{B^{E^p}_{t(x)}}(X)]}{\mathbb{E}[(1 - T_{t(x_k)}(X - t(x)))X_{B^{E^p}_{t(x)}}(X)]}.
\]

However, it is not possible to directly see how the convergence happens, due to the dependence among different terms and how the regularized pseudo-inverse converges. The dependence on the regularization order is also not clear. A careful theoretical analysis is needed.

To proceed with the proof, we need to discuss a critical observation. Note that the term \( C_x \) might be ill-conditioned for the pseudo-inverse procedure, and the regularized pseudo inverse depends on how the regularization penalty \( p \) is chosen. As we will see later, the choice of \( p \) is critical for the outcome. The ill-conditionedness depends on the manifold geometry, and can be complicated. In this paper we focus on the following three cases.

**Condition 2.4.1.** Follow the notations used in Proposition 2.3.2. For the local covariance matrix \( C_x \) with the rank \( r \), without loss of generality, we consider the following three cases:

- **Case 0:** \( r = d \);
- **Case 1:** \( r = p > d \), and \( \lambda_{d+1}^{(2)}, \ldots, \lambda_{p}^{(2)} \) are nonzero;
- **Case 2:** \( r = p > d \), \( \lambda_{d+1}^{(2)}, \ldots, \lambda_{p-l}^{(2)} \) are nonzero, where \( 1 \leq l \leq p-d \), \( \lambda_{p-l+1}^{(2)} = \ldots = \lambda_{p}^{(2)} = 0 \), and \( \lambda_{p-l+1}^{(4)}, \ldots, \lambda_{p}^{(4)} \) are nonzero.

At first glance, it is limited to assume that when \( r > d \), we have \( r = p \) in Cases 1 and 2. However, it is general enough in the following sense. In Cases 1 and 2, if \( C_x \) is degenerate, that is, \( d < r < p \), it means that locally the manifold only occupies a lower dimensional affine subspace. Therefore, the sampled data are constrained to this affine subspace, and hence the rank of the local sample covariance matrix satisfies \( r_n \leq r \). As a result, the analysis can be carried out only on this affine subspace without changing the outcome. More general situations could be studied by the same analysis techniques shown below, but they will not provide more insights about our
understanding of the algorithm and will introduce additional notational burdens. For \( f \in C(\mathcal{S}(M)) \), define
\[
Qf(x) := \frac{E[f(X)(1 - T_{t(x)}^T(X - t(x)))\chi_{B^R_{\rho^n}(x)}(X)]}{E[(1 - T_{t(x)}^T(X - t(x)))\chi_{B^R_{\rho^n}(x)}(X)]},
\]
(2.4.4)

The following theorem summarizes the relationship between LLE and \( Qf \) under these three cases.

**Theorem 2.4.1.** Fix \( f \in C(\mathcal{S}(M)) \). Suppose the regularization order is \( \rho \in \mathbb{R}, \epsilon = \epsilon(n) \) so that \( \sqrt{n \log n} \rightarrow 0 \) and \( \epsilon \rightarrow 0 \) as \( n \rightarrow \infty \). With probability greater than \( 1 - n^{-2} \), for all \( x_k \in \mathcal{X} \), under different conditions listed in Condition 2.4.1, we have:
\[
\sum_{j=1}^{N} w_k(j)f(x_{k,j}) - f(x_k)
\]
(2.4.5)
\[
= \begin{cases} 
Qf(x_k) - f(x_k) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right) & \text{in Case 0} \\
Qf(x_k) - f(x_k) + O\left(\sqrt{\frac{\log(n)}{n^{1/2}e^{d/2} - 1/2}}\right) & \text{in Cases 1,2}
\end{cases}
\]

Particularly, when \( \rho \leq 3 \), with probability greater than \( 1 - n^{-2} \), for all \( x_k \in \mathcal{X} \), for all Cases listed in Condition 2.4.1, we have:
\[
\sum_{j=1}^{N} w_k(j)f(x_{k,j}) - f(x_k) = Qf(x_k) - f(x_k) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right).
\]
(2.4.6)

Note that the convergence rate of Case 0 is fast, no matter what regularization order \( \rho \) is chosen, while the convergence rate of Case 1 and Case 2 depends on \( \rho \). This theorem echoes several practical findings of LLE that the choice of regularization is critical in the performance, and it suggests that we should choose \( \rho = 3 \).

**Remark 2.4.1.** We should compare the convergence rate of LLE with that of the DM. The convergence rate of Case 0 is the same as that of the eigenmap or the DM without any normalization [57], while the convergence rate of Case 1 and Case 2 is the same as that of the \( \alpha \)-normalized DM [18] when \( \rho \geq 4 \) [57]. Note that the main convergence rate bottleneck for the \( \alpha \)-normalized DM comes from the probability density function estimation, while the convergence bottleneck for LLE is the regularized pseudo-inverse.

Theorem 2.4.1 describes how LLE could be viewed as a “diffusion process” on the dataset. Note that
\[
E[f(X)(1 - T_{t(x)}^T(X - t(x)))\chi_{B^R_{\rho^n}(x)}(X)]
\]
(2.4.7)
\[
= \int_M (1 - T_{t(x)}^T(Y - t(x)))\chi_{B^R_{\rho^n}(x)}(Y)f(t(Y))P(y)dV(y)
\]

Therefore, we can view \( w_n \) as a “zero-one” kernel supported on \( B^R_{\rho^n}(x_k) \cap t(M) \) with the correction depending on \( T_{t(x_k)} \). Note that after the correction, the whole operator may no longer be a diffusion.

**Corollary 2.4.1.** The integral kernel associated with LLE when the regularization order is \( \rho \in \mathbb{R} \) is
\[
K_{\text{LLE}}(x,y) = [1 - T_{t(x)}^T(t(y) - t(x))]\chi_{B^R_{\rho^n}(t(x)) \cap t(M)}(t(y)),
\]
(2.4.8)

where \( x, y \in M \) and
\[
T_{t(x)} := \mathcal{K}_{d+\rho}^1(C_x)\left[E(X - t(x))\chi_{B^R_{\rho^n}(x)}\right] \in \mathbb{R}^p.
\]
(2.4.9)
Note that $K_{\text{LLE}}$ depends on $\epsilon$, the geometry of the manifold near $x$, and $\rho$ via $T_{i}(x)$. We provide some properties of the kernel function $K_{\text{LLE}}$. By a direct expansion, we have $T_{i}(x) = \sum_{j=1}^{d} \frac{E[(X-t(x))X_{B_{j}}^{\rho}(x)]}{\lambda_{j}} u_{j}$, where $u_{j}$ and $\lambda_{j}$ are the $j$-th eigen-pair of $C_{x}$. Since $|\mathbb{E}(X-t(x))X_{B_{j}}^{\rho}(x)|$ is bounded above by $\text{vol}(M)\epsilon$, $\lambda_{i} + \epsilon^{d+p}$ is bounded below by $\epsilon^{d+p}$ and each $u_{j}$ is a unit vector, $|T_{x}|$ is bounded above by $\sum_{j=1}^{d} \frac{\epsilon^{\text{vol}(M)}}{\lambda_{j}} \epsilon^{d+p}$. Consequently, we have the following proposition.

**Proposition 2.4.1.** The kernel $K_{\text{LLE}}$ is compactly supported and is in $L^{2}(M \times M)$. Thus, the linear operator $A : L^{2}(M, PdV) \rightarrow L^{2}(M, PdV)$ defined by

$$Af(x) := \mathbb{E}[f(X)(1-T_{i}(x)(X-t(x))X_{B_{j}}^{\rho}(x))] \quad (2.4.10)$$

is Hilbert-Schmidt.

Note that the kernel function $K_{\text{LLE}}(x, \cdot)$ depends on $x$ and hence the manifold, and the kernel is dominated by normal bundle information, due to the regularized pseudo-inverse procedure. For example, if $M$ is an affine subspace of $\mathbb{R}^{p}$ and the data is uniformly sampled, then $\mathbb{E}[(X-x)X_{B_{j}}^{\rho}(x)] = 0$. Consequently, $T_{x} = 0$ and $K(x, y) = 1$. If $M$ is $S^{p-1}$, a unit sphere centered at origin embedded in $\mathbb{R}^{p}$ and the data is uniformly sampled, the first dominant $p - 1$ eigenvectors are perpendicular to $x$ and the last eigenvector is parallel to $x$. By a direct calculation, $\mathbb{E}[(X-x)X_{B_{j}}^{\rho}(x)]$ is parallel to $x$ and hence $K(x, y)$ behaves like a quadratic function $1 - cx^\top(y-x) = 1 - cx^\top(y-x)$, where $c$ is the constant depending on the eigenvalues.

### 2.4.1 Proof of Theorem 2.4.1

For each $x_k$, denote $f = (f(x_{k,1}), f(x_{k,2}), \ldots, f(x_{k,N}))^\top \in \mathbb{R}^{N}$. By the expansion

$$\sum_{j=1}^{N} w_{k}(j)f(x_{k,j}) = \frac{1}{N} f - \frac{1}{N} G_{n}^{\top} \mathcal{J}_{n_{d+p}}(G_{n}G_{n}^{\top})G_{n}f,$$

we can write $\sum_{j=1}^{N} w_{k}(j)f(x_{k,j}) - f(x_{k})$ as

$$\frac{1}{n} \sum_{j=1}^{N} (f(x_{k,j}) - f(x_{k})) - \frac{1}{n} \sum_{j=1}^{N} (x_{k,j} - x_{k}) \frac{1}{n} \sum_{j=1}^{N} (x_{k,j} - x_{k}) (f(x_{k,j}) - f(x_{k})) = \frac{1}{n} \sum_{j=1}^{N} (x_{k,j} - x_{k}) (f(x_{k,j}) - f(x_{k})) = \frac{1}{n} \sum_{j=1}^{N} (x_{k,j} - x_{k}) \mathcal{J}_{n_{d+p}}(G_{n}G_{n}^{\top})G_{n}G_{n}^{\top} \frac{1}{n} \sum_{j=1}^{N} (x_{k,j} - x_{k}) \] \quad (2.4.11)

Note that we have

$$n \mathcal{J}_{n_{d+p}}(G_{n}G_{n}^{\top}) = \mathcal{J}_{d+p} \left( \frac{1}{n} G_{n}G_{n}^{\top} \right).$$

Thus, the goal is to relate the finite sum quantity (2.4.11) with the following “expectation”

$$Af(x_{k}) - f(x_{k}) = Qf(x_{k}) - f(x_{k}), \quad (2.4.12)$$

where $A$ is defined in (2.4.10). Note that LLE is a ratio of two dependent random variables, and the denominator and numerator both involve complicated mixup of sampling points. Therefore, the convergence fluctuation cannot
be simply computed. We control the size of the fluctuation of the following five terms

\[
\frac{1}{ne^d} \sum_{j=1}^{N} 1 \tag{2.4.13}
\]
\[
\frac{1}{ne^d} \sum_{j=1}^{N} (f(x_{k,j}) - f(x_k)) \tag{2.4.14}
\]
\[
\frac{1}{ne^d} \sum_{j=1}^{N} (x_{k,j} - x_k) \tag{2.4.15}
\]
\[
\frac{1}{ne^d} \sum_{j=1}^{N} (x_{k,j} - x_k)(f(x_{k,j}) - f(x_k)) \tag{2.4.16}
\]
\[
\frac{1}{ne^d} (G_n G_n^\top + \varepsilon^p I_{p \times p}) \tag{2.4.17}
\]

as functions of \( n \) and \( \varepsilon \) by the Bernstein type inequality. Here, we put \( \varepsilon^{-d} \) in front of each term to normalize the kernel so that the computation is consistent with the existing literature, like \([17, 57]\). The size of the fluctuation of these terms are controlled in the following Lemmas. The term (2.4.13) is the usual kernel density estimation, so we have the following lemma.

**Lemma 2.4.1.** When \( n \) is large enough, we have with probability greater than \( 1 - n^{-2} \) that for all \( k = 1, \ldots, n \) that

\[
\left| \frac{1}{ne^d} \sum_{j=1}^{N} 1 - \mathbb{E} \frac{1}{e^d} \chi_{B_{\varepsilon}^p(x_k)}(X) \right| = O\left( \frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}} \right).
\]

The behavior of (2.4.14) is summarized in the following Lemma. Although the proof is standard, we provide it for the sake of self-containment.

**Lemma 2.4.2.** When \( n \) is large enough, we have with probability greater than \( 1 - n^{-2} \) that for all \( k = 1, \ldots, n \) that

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

**Proof.** By denoting

\[
F_{1,j} = \frac{1}{e^d} (f(x_j) - f(x_k)) \chi_{B_{\varepsilon}^p(x_k)}(x_j),
\]

we have

\[
\frac{1}{ne^d} \sum_{j=1}^{N} (f(x_{k,j}) - f(x_k)) = \frac{1}{n} \sum_{j \neq k, j = 1}^{n} F_{1,j}.
\]

Define a random variable

\[
F_1 := \frac{1}{e^d} (f(X) - f(x_k)) \chi_{B_{\varepsilon}^p(x_k)}(X).
\]

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

\[
\frac{1}{n} \sum_{j \neq k, j = 1}^{n} F_{1,j} = \frac{n - 1}{n} \left[ \frac{1}{n - 1} \sum_{j \neq k, j = 1}^{n} F_{1,j} \right].
\]

Since \( \frac{n - 1}{n} \to 1 \) as \( n \to \infty \), the error incurred by replacing \( \frac{1}{n} \) by \( \frac{1}{n - 1} \) is of order \( \frac{1}{n} \), which is negligible asymptotically.
Thus, we can simply focus on analyzing $\frac{1}{n} \sum_{j=1,j\neq 1}^{n} F_{1,j}$. We have by Lemma 2.2.5

$$E[F_1] = \frac{|S^{d-1}|}{2d(d+2)} [\Delta((f(y) - f(x_k))P(y))|_{y=x_k}] \epsilon^2 + O(\epsilon^3)$$

$$E[\sigma_1^2] = \frac{|S^{d-1}|}{2d(d+2)} [\Delta((f(y) - f(x_k))^2P(y))|_{y=x_k}] \epsilon^{-d+2} + O(\epsilon^{-d+3}),$$

where $\Delta$ acts on $y$ and we apply the Lemma by viewing $f(y)P(y)$ as a function and evaluate the integration over the uniform measure. Thus, we conclude that

$$\sigma_1^2 := \text{Var}(F_1) = \frac{|S^{d-1}|}{2d(d+2)} [\Delta((f(y) - f(x_k))^2P(y))|_{y=x_k}] \epsilon^{-d+2} + O(\epsilon^{-d+3}). \quad (2.4.18)$$

To simplify the discussion, we assume that $\Delta((f(y) - f(x_k))^2P(y))|_{y=x_k} \neq 0$ so that $\sigma_1^2 = O(\epsilon^{-d+2})$ when $\epsilon$ is small enough. In the case that $\Delta((f(y) - f(x_k))^2P(y))|_{y=x_k} = 0$, the variance is of higher order, and the proof is the same.

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

$$c_1 = 2\|f\|_{L^\infty} \epsilon^{-d}$$

and

$$\sigma_1^2/c_1 \to 0 \text{ as } \epsilon \to 0,$$

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

$$\Pr \left\{ \frac{1}{n-1} \sum_{j\neq k,j=1}^{n} (F_{1,j} - E[F_1]) > \beta_1 \right\} \leq e^{-\frac{n\beta_1^2}{3\sigma_1^2 + \frac{2}{c_1} \beta_1}},$$

where $\beta_1 > 0$. Since our goal is to estimate a quantity of order $\epsilon^2$, which is the order that the Laplace-Beltrami operator lives, we need to take $\beta_1 = \beta_1(\epsilon)$ much smaller than $\epsilon^2$ in the sense that $\beta_1/\epsilon^2 \to 0$ as $\epsilon \to 0$. In this case, $c_1 \beta_1$ is much smaller than $\sigma_1^2$, and hence $2\sigma_1^2 + \frac{2}{c_1} \beta_1 \leq 3\sigma_1^2$ when $\epsilon$ is smaller enough. Thus, when $\epsilon$ is smaller enough, the exponent in Bernstein’s inequality is bounded from below by

$$\frac{n\beta_1^2}{2\sigma_1^2 + \frac{1}{3} c_1 \beta_1} \geq \frac{n\beta_1^2}{3\sigma_1^2} \geq \frac{n\beta_1^2 \epsilon^{d-2}}{3 \frac{|S^{d-1}|}{d(d+2)} [\Delta((f(y) - f(x_k))^2P(y))|_{y=x_k}]}.$$  

Suppose $n$ is chosen large enough so that

$$\frac{n\beta_1^2 \epsilon^{d-2}}{3 \frac{|S^{d-1}|}{d(d+2)} [\Delta((f(y) - f(x_k))^2P(y))|_{y=x_k}]} = 3\log(n);$$

that is, the deviation from the mean is set to

$$\beta_1 = \frac{3 \sqrt{\log(n)} \sqrt{\frac{|S^{d-1}|}{d(d+2)} [\Delta((f(y) - f(x_k))^2P(y))|_{y=x_k}]}}{n^{1/2} \epsilon^{d/2-1}} = O\left(\frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2-1}}\right), \quad (2.4.19)$$

where the implied constant in $O\left(\frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2-1}}\right)$ is $\sqrt{\frac{|S^{d-1}|}{d(d+2)} [\Delta((f(y) - f(x_k))^2P(y))|_{y=x_k}]}$. Note that by the assump-
tion that $\epsilon = \epsilon(n)$ so that $\frac{\sqrt{\log n}}{n^{1/2} \varepsilon^2 / n} \to 0$ as $n \to 0$, we know that $\beta_1 / \epsilon^2 = \frac{\sqrt{\log n}}{n^{1/2} \varepsilon^2 / n} \to 0$. It implies that the deviation greater than $\beta_1$ happens with probability less than

$$\exp \left( -\frac{n\beta_1^2}{2\sigma_i^2 + \frac{1}{3}c_1 \beta_1} \right) \leq \exp \left( -\frac{n\beta_1^2 \epsilon^{d-2}}{3 \frac{\log \frac{1}{\delta}}{d(d+2)} \Delta(f(y) - f(x)) \sigma(y)} \right)$$

$$= \exp(-3 \log(n)) = 1/n^3.$$

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

$$\Pr \left\{ \frac{1}{n-1} \sum_{j \neq k, j=1}^n (F_{i,j} - \mathbb{E}[F_i]) > \beta_1 \right\} \leq \exp\left(-\frac{n\beta_1^2 \epsilon^d}{3 \left( \frac{\log \frac{1}{\delta}}{d(d+2)} \Delta(f(y) - f(x)) \sigma(y) \right)} \right) \leq \frac{1}{n^2}.$$

Denote $\Omega_1$ to be the event space that the deviation $\frac{1}{n-1} \sum_{j \neq k, j=1}^n (F_{i,j} - \mathbb{E}[F_i]) \leq \beta_1$ for all $i = 1, \ldots, n$, where $\beta_1$ is chosen in (2.4.19) is satisfied. We now proceed to (2.4.15). In this case, we need to discuss different cases indicated by Condition 2.4.1.

**Lemma 2.4.3.** Suppose Case 0 in Condition 2.4.1 holds. When $n$ is large enough, we have with probability greater than $1 - n^{-2}$ that for all $k = 1, \ldots, n$,

$$e_i^T \left[ \frac{1}{n \epsilon d} \sum_{j=1}^N (x_{k,j} - x_k) - \mathbb{E}\left[ \frac{1}{\epsilon d} (X - x_k) X_{B_k^p}(x_k) X \right] \right] = O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2-1}} \right),$$

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

Suppose Case 1 in Condition 2.4.1 holds. When $n$ is large enough, we have with probability greater than $1 - n^{-2}$ that for all $k = 1, \ldots, n$,

$$e_i^T \left[ \frac{1}{n \epsilon d} \sum_{j=1}^N (x_{k,j} - x_k) - \mathbb{E}\left[ \frac{1}{\epsilon d} (X - x_k) X_{B_k^p}(x_k) X \right] \right] = O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2-1}} \right),$$

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

$$e_i^T \left[ \frac{1}{n \epsilon d} \sum_{j=1}^N (x_{k,j} - x_k) - \mathbb{E}\left[ \frac{1}{\epsilon d} (X - x_k) X_{B_k^p}(x_k) X \right] \right] = O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2-1}} \right),$$

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

Suppose Case 2 in Condition 2.4.1 holds. When $n$ is large enough, we have with probability greater than $1 - n^{-2}$ that for all $k = 1, \ldots, n$,

$$e_i^T \left[ \frac{1}{n \epsilon d} \sum_{j=1}^N (x_{k,j} - x_k) - \mathbb{E}\left[ \frac{1}{\epsilon d} (X - x_k) X_{B_k^p}(x_k) X \right] \right] = O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2-1}} \right),$$

where $i = 1, \ldots, d$,
where \( i = d + 1, \ldots, p - l \), and
\[
e_i^\top \left[ \frac{1}{n \epsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) - \frac{1}{\epsilon^d} \frac{1}{n} \sum_{j=1}^n \mathbb{E} F_{2,i,j} e_i \right] = O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^d/2 - 3} \right),
\]
where \( i = p - l + 1, \ldots, p \).

**Proof.** First, we prove Case 1. Case 0 is a special case of Case 1. Suppose Case 1 holds. Fix \( x_k \). By denoting
\[
\frac{1}{n \epsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) = \frac{1}{n} \sum_{j=1}^n \sum_{\ell=1}^p F_{2,i,j} e_i,
\]
where
\[
F_{2,i,j} := \frac{1}{\epsilon^d} e_i^\top (x_j - x_k) \mathbb{I}_{B^{p,i}}(x_j),
\]
we know that when \( j \neq k \), \( F_{2,i,j} \) is randomly sampled i.i.d. from the random variable
\[
F_{2,i} := \frac{1}{\epsilon^d} e_i^\top (x_i - x_k) \mathbb{I}_{B^{p,i}}(x_i).
\]
Similarly, we can focus on analyzing \( \frac{1}{n} \sum_{j=1, j \neq i}^n F_{2,i,j} \) since \( \frac{n-1}{n} \to 1 \) as \( n \to \infty \). By plugging \( f = 1 \) in (2.2.10), we have
\[
\mathbb{E}[F_{2,i}] = \frac{|S^{d-1}| P(x) e^{-d+2}}{d(d+2)} + O(\epsilon^{-d+4}) \quad \text{when } \ell = 1, \ldots, d
\]
and by (2.3.8) we have
\[
\mathbb{E}[F_{2,i}] = \begin{cases} 
\frac{|S^{d-1}| P(x) e^{-d+2}}{d(d+2)} + O(\epsilon^{-d+4}) & \text{when } \ell = 1, \ldots, d \\
\frac{P(x) e^{-d+4}}{4(d+4)} \int_{S^{d-1}} |\langle \Pi_{\ell}(\theta, \theta), e_i \rangle|^2 d\theta + O(\epsilon^{-d+6}) & \text{when } \ell = d + 1, \ldots, p.
\end{cases}
\]
Thus, we conclude that
\[
\sigma_{2,i}^2 := \text{Var}(F_{2,i}) = \begin{cases} 
\frac{|S^{d-1}| P(x) e^{-d+2}}{d(d+2)} + O(\epsilon^{-d+4}) & \text{when } \ell = 1, \ldots, d \\
\frac{P(x) e^{-d+4}}{4(d+4)} \int_{S^{d-1}} |\langle \Pi_{\ell}(\theta, \theta), e_i \rangle|^2 d\theta + O(\epsilon^{-d+6}) & \text{when } \ell = d + 1, \ldots, p.
\end{cases}
\]
Note that for \( \ell = d + 1, \ldots, p \), the variance is of higher order than that of \( \ell = 1, \ldots, d \). By the same argument, Case 0 satisfies \( \mathbb{E}[F_{2,i}] = \sigma_{2,i}^2 = 0 \) for \( \ell = d + 1, \ldots, p \).

With the above bounds, we could apply the large deviation theory. For \( \ell = 1, \ldots, d \), the random variable \( F_{2,i} \) is uniformly bounded by \( c_{2,i} = 2e^{-d+1} \) and \( \sigma_{2,i}^2/c_{2,i} \to 0 \) as \( \epsilon \to 0 \), so when \( \epsilon \) is sufficiently smaller and \( n \) is sufficiently large, the exponent in Bernstein’s inequality,
\[
\text{Pr} \left\{ \frac{1}{n - 1} \sum_{j \neq k}^n \left( F_{2,i,j} - \mathbb{E}[F_{2,i}] \right) > \beta_{2,i} \right\} \leq \exp \left( - \frac{n \beta_{2,i}^2}{2 \sigma_{2,i}^2 + \frac{2}{3} c_{2,i} \beta_{2,i}^2} \right),
\]
which
where $\beta_{2,\ell} > 0$, satisfies
\[
\frac{n\beta_{2,\ell}^2}{2\sigma_{2,\ell}^2 + \frac{1}{2} \epsilon_{2,\ell} \beta_{2,\ell}} \geq \frac{n\beta_{2,\ell}^2 e^{\ell-2}}{3 \sigma_{2,\ell}^2} = 3\log(n);
\]
that is, the deviation from the mean is set to
\[
\beta_{2,\ell} = 3\sqrt{\log(n)} \sqrt{\frac{3 \|S^{t-1}P(s)\|_{d(d+2)}}{d(d+2)}} = O\left(\frac{\sqrt{\log(n)}}{n^{1/2} e^{\ell-2}}\right),
\]
(2.4.20)

For $\ell = d + 1, \ldots, p$, since the variance is of higher order, by the same argument, we have
\[
\beta_{2,\ell} = 3\sqrt{\log(n)} \sqrt{\frac{3 \|S^{t-1}P(s)\|_{d(d+2)}}{d(d+2)}} = O\left(\frac{\sqrt{\log(n)}}{n^{1/2} e^{\ell-2}}\right),
\]
(2.4.21)

As a result, in both Case 0 and Case 1, by a simple union bound, for $\ell = 1, \ldots, d$, we have
\[
\Pr\left\{ \frac{1}{n} \sum_{j \neq k, j = 1}^{n} F_{2,\ell,j} - \mathbb{E}[F_{2,\ell}] \geq \beta_{2,\ell} \mid k = 1, \ldots, n \right\} \leq 1/n^2.
\]
where
\[
\beta_{2,\ell} = 3\sqrt{\log(n)} \sqrt{\frac{3 \|S^{t-1}P(s)\|_{d(d+2)}}{d(d+2)}} = O\left(\frac{\sqrt{\log(n)}}{n^{1/2} e^{\ell-2}}\right),
\]
(2.4.22)

and in Case 1, for $\ell = d + 1, \ldots, p$, we have
\[
\Pr\left\{ \frac{1}{n} \sum_{j \neq k, j = 1}^{n} F_{2,\ell,j} - \mathbb{E}[F_{2,\ell}] \geq \beta_{2,\ell} \mid k = 1, \ldots, n \right\} \leq 1/n^2.
\]
where
\[
\beta_{2,\ell} = 3\sqrt{\log(n)} \sqrt{\frac{3 \|S^{t-1}P(s)\|_{d(d+2)}}{d(d+2)}} = O\left(\frac{\sqrt{\log(n)}}{n^{1/2} e^{\ell-2}}\right),
\]
(2.4.23)

For Case 2, by plugging $f = 1$ in (2.2.10), we have
\[
\mathbb{E}[F_{2,\ell}] = \left\{ \begin{array}{ll}
\epsilon^2 \epsilon_{\ell} \frac{|S^{t-1}|_{t} \nabla P(x)}{d(d+2)} + O(\epsilon^4) & \text{when } \ell = 1, \ldots, d \\
\epsilon^2 \epsilon_{\ell} \frac{|S^{t-1}|_{t} \nabla P(x)^{T} J_{p,p-d}^{T} \mathcal{H}_{\ell}(x)}{d(d+2)} + O(\epsilon^4) & \text{when } \ell = d + 1, \ldots, p - l \\
\epsilon^4 \epsilon_{\ell} \frac{\mathcal{H}_{\ell}(x) \nabla P(x)}{6(d+4)} + O(\epsilon^5) & \text{when } \ell = p - l + 1, \ldots, p.
\end{array} \right.
\]
Lemma 2.4.4. \( \quad \)

then the eigen-decomposition of \( \mathbf{P} \) under Case 0 in Condition 2.4.1, (2.4.22) and (2.4.23) under Case 2. Consequently, we can conclude that

\[
\mathbf{P} = \begin{pmatrix} \mathbf{P}_{1,1} & \mathbf{P}_{1,2} \\ \mathbf{P}_{2,1} & \mathbf{P}_{2,2} \end{pmatrix},
\]

Denote \( \Omega_2 \) to be the event space that the deviation \( \frac{1}{n} \sum_{j \neq k, \ell, \mathbf{F}_2, j} \mathbf{F}_2, j - \mathbb{E}[\mathbf{F}_2, \ell] \leq \beta_{2, \ell} \) for all \( \ell = 1, \ldots, p \) and \( k = 1, \ldots, n \), where \( \beta_{2, \ell} \) are chosen in (2.4.22) under Case 0 in Condition 2.4.1, (2.4.22) and (2.4.23) under Case 1, and (2.4.24) under Case 2.

Denote the eigen-decomposition of \( \frac{1}{n} G_n G_n^\top \) as \( U_n \tilde{\mathbf{A}}_n U_n^\top \), where \( U_n \in O(p) \) and \( \tilde{\mathbf{A}}_n \in \mathbb{R}^{p \times p} \) a diagonal matrix, and the eigen-decomposition of \( \frac{1}{p} \mathbf{C}_n \) as \( U \tilde{\mathbf{A}} U^\top \), where \( U \in O(p) \) and \( \tilde{\mathbf{A}} \in \mathbb{R}^{p \times p} \) a diagonal matrix. Note that

\[
ne^{d,\mathcal{F}_{\text{ne}^{d,\mathcal{F}}} (G_n G_n^\top)} = \mathcal{F}_{\text{ne}^{d,\mathcal{F}}} (\frac{1}{n} G_n G_n^\top).
\]

We first control \( \mathcal{F}_{\mathcal{E}^p} (\tilde{\mathbf{A}}_n) - \mathcal{F}_{\mathcal{E}^p} (\tilde{\mathbf{A}}) = \mathcal{I}_{p, r_n} (\tilde{\mathbf{A}}_n + \mathcal{E}^p)^{-1} - \mathcal{I}_{p, r} (\tilde{\mathbf{A}} + \mathcal{E}^p)^{-1} \mathcal{I}_{p, r} \) based on the three cases listed in Condition 2.4.1. By Proposition 2.3.1, the first \( d \) eigenvalues of \( \mathbb{E} \mathbf{F} \) are of order \( \mathcal{E}^2 \). In Case 0, all the remaining eigenvalues are 0; in Case 1, all the remaining eigenvalues are nonzero and of order \( \mathcal{E}^4 \); in Case 2, there are \( l \) nonzero eigenvalues of order \( \mathcal{E}^4 \) and \( p - d - l \) remaining eigenvalues of order \( \mathcal{E}^4 \).

Lemma 2.4.4. When \( n \) is large enough, with probability greater than \( 1 - n^{-2} \), for Case 0 in Condition 2.4.1, we have

\[
|e_{i, j}^\top (\mathcal{F}_{\mathcal{E}^p} (\tilde{\mathbf{A}}_n) - \mathcal{F}_{\mathcal{E}^p} (\tilde{\mathbf{A}})) e_{i, j}| = O\left( \frac{\sqrt{\log(n)}}{n^{1/2} e^{d/2 - 2} e^{2/2}} \right).
\]

Thus, we conclude that

\[
\mathbb{E}[F_{2, \ell}^2] = \frac{1}{n} \sum_{j \neq k, \ell, \mathbf{F}_2, j} \mathbf{F}_2, j - \mathbb{E}[\mathbf{F}_2, \ell] \leq \beta_{2, \ell} \quad \text{when} \quad \ell = 1, \ldots, d
\]

By the same large deviation argument that we skip the details, we conclude the claim with

\[
\beta_{2, \ell} = \begin{cases} O\left( \frac{\sqrt{\log(n)}}{n^{1/2} e^{d/2 - 1}} \right) & \text{when} \quad \ell = 1, \ldots, d \\
O\left( \frac{\sqrt{\log(n)}}{n^{1/2} e^{d/2 - 2}} \right) & \text{when} \quad \ell = d + 1, \ldots, p - l \\
O\left( \frac{\sqrt{\log(n)}}{n^{1/2} e^{d/2 - 3}} \right) & \text{when} \quad \ell = p - l + 1, \ldots, p.
\end{cases}
\]
for \( i = 1, \ldots, d \); for Case 1 in Condition 2.4.1, we have

\[
|e_i^T [ \mathcal{J}_{EP}(\bar{\Lambda}_n) - \mathcal{J}_{EP}(\bar{\Lambda})]e_i| = \begin{cases}
O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{\ell(2-2\gamma/2+2/p)}}\right) & \text{for } i = 1, \ldots, d \\
O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{\ell(6-2\gamma/2+2/p)}}\right) & \text{for } i = d + 1, \ldots, p.
\end{cases}
\]

for Case 2 in Condition 2.4.1, we have

\[
|e_i^T [ \mathcal{J}_{EP}(\bar{\Lambda}_n) - \mathcal{J}_{EP}(\bar{\Lambda})]e_i| = \begin{cases}
O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{\ell(2-2\gamma/2+2/p)}}\right) & \text{for } i = 1, \ldots, d \\
O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{\ell(6-2\gamma/2+2/p)}}\right) & \text{for } i = d + 1, \ldots, p - 1 \\
O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{\ell(6-2\gamma/2+2/p)}}\right) & \text{for } i = p - l + 1, \ldots, p.
\end{cases}
\]

Moreover, for each case in Condition 2.4.1, when \( n \) is sufficiently large, with probability greater than \( 1 - n^{-2} \), we have \( U_n = U\Theta + \frac{\log(n)}{n^{1/2}e^{\ell(2-2\gamma/2+2/p)}} U\Theta S + O\left(\frac{\log(n)}{n^{1/2}e^{\ell(2-2\gamma/2+2/p)}}\right) \), where \( S \in o(p) \), and \( \Theta \in O(p) \). \( \Theta \) commutes with \( \mathcal{J}_{EP}(\bar{\Lambda}) \)

Note that \( \frac{\log(n)}{n^{1/2}e^{\ell(2-2\gamma/2+2/p)}} \) is asymptotically bounded by \( \varepsilon^6 \) due to the assumption that \( \frac{\sqrt{\log(n)}}{n^{1/2}e^{\ell(2-2\gamma/2+2/p)}} \) is asymptotically approaching zero as \( n \to \infty \).

**Proof.** We start from analyzing \( \frac{1}{n^d} G_n G_n^T \). The proof can be found in [57, (6.12)-(6.19)], and here we summarize the results with our notations. Denote

\[
F_{a,b,i} := \frac{1}{\varepsilon^2} e_a^T (x_{k,i} - x_k)(x_{k,i} - x_k)^T e_b,
\]

so that

\[
\frac{1}{n^d} G_n G_n^T = \frac{1}{n^d} \sum_{a,b,i=1}^n F_{a,b,i} e_a e_b^T.
\]

Note that for each \( a, b = 1, \ldots, p \), \( \{F_{a,b,i}\}_{i=1}^n \) are i.i.d. realizations of the random variable \( F_{a,b} = \frac{1}{\varepsilon^2} e_a^T (X - x_k)(X - x_k)^T e_b X_{B_k}X_{B_k}^T(X) \). Denote \( F_3 \in \mathbb{R}^{p \times p} \) so that the \((a,b)\)-th entry of \( F_3 \) is \( F_{a,b} \). Note that \( C_a = \varepsilon^2 \mathbb{E} F_3 \).

The random variable \( F_{a,b} \) is bounded by \( c_{a,b} = 2\varepsilon^{-d+2} \) when \( a, b = 1, \ldots, d \), by \( c_{a,b} = c_{a,b} \varepsilon^{-d+4} \) when \( a, b = d + 1, \ldots, p \), and by \( c_{a,b} = c_{a,b} \varepsilon^{-d+3} \) for other pairs of \( a, b \), where \( c_{a,b} \), when \( a > d \) or \( b > d \), are constants depending on the second fundamental form [55, (B.33)-(B.34)].

The variance of \( F_{a,b} \), denoted as \( \sigma^2_{a,b} \), is \( s_{a,b} \varepsilon^{-d+1} \) when \( a, b = 1, \ldots, d \), \( s_{a,b} \varepsilon^{-d+8} \) when \( a, b = d + 1, \ldots, p \), and \( s_{a,b} \varepsilon^{-d+6} \) for other pairs of \( a, b \) (see [55, (B.33)-(B.35)], or [57]), where \( s_{a,b} \) are constants depending on the second fundamental form. Again, to simplify the discussion, we assume that \( c_{a,b} \) and \( s_{a,b} \) are not zero for all \( a, b = 1, \ldots, p \). When the variance is of higher order, the deviation could be evaluated similarly and we skip the details.\(^1\) Thus, for \( \beta_{3,1}, \beta_{3,2}, \beta_{3,3} > 0 \), by Bernstein’s inequality, we have

\[
\Pr \left\{ \frac{n}{n} \sum_{i \neq k, i=1} F_{a,b,i} - \mathbb{E} F_{a,b} > \beta_{3,1} \right\} \leq \exp \left\{ -\frac{(n-1)\beta_{3,1}^2}{s_{a,b} \varepsilon^{-d+4} + c_{a,b} \varepsilon^{-d+2} \beta_{3,1}} \right\}
\]

\(^1\):For example, when the manifold is flat around \( x_k \) and \( \varepsilon \) is sufficiently small, \( c_{a,b} = s_{a,b} = 0 \) when \( a > d \) or \( b > d \), and the proof of the bound is trivial.
when \(a, b = 1, \ldots, d\),
\[
\Pr \left\{ \frac{1}{n} \sum_{i=1}^{n} F_{a, b, i} - \mathbb{E} F_{a, b} \mid > \beta_{1, 2} \right\} \leq \exp \left\{ -\frac{(n-1)\beta_{1, 2}^2}{s_{a, b}e^{-d+8} + c_{a, b}e^{-d+4}\beta_{1, 2}} \right\} \quad (2.4.26)
\]
when \(a, b = d + 1, \ldots, p\), and
\[
\Pr \left\{ \frac{1}{n} \sum_{i=1}^{n} F_{a, b, i} - \mathbb{E} F_{a, b} \mid > \beta_{3, 3} \right\} \leq \exp \left\{ -\frac{(n-1)\beta_{3, 3}^2}{s_{a, b}e^{-d+6} + c_{a, b}e^{-d+3}\beta_{3, 3}} \right\} \quad (2.4.27)
\]
for the other cases.

Choose \(\beta_{3, 1}, \beta_{3, 2}\) and \(\beta_{3, 3}\) so that \(\beta_{3, 1}/\varepsilon^2 \to 0, \beta_{3, 2}/\varepsilon^4 \to 0\) and \(\beta_{3, 3}/\varepsilon^3 \to 0\) as \(\varepsilon \to 0\) so that when \(\varepsilon\) is sufficiently small,
\[
\begin{align*}
s_{a, b}e^{-d+4} + c_{a, b}e^{-d+2}\beta_{3, 1} & \leq 2s_{a, b}e^{-d+4} \quad \text{for all } k, l = 1, \ldots, d, \\
s_{a, b}e^{-d+8} + c_{a, b}e^{-d+4}\beta_{3, 2} & \leq 2s_{a, b}e^{-d+8} \quad \text{for all } k, l = d + 1, \ldots, p, \\
s_{a, b}e^{-d+6} + c_{a, b}e^{-d+3}\beta_{3, 3} & \leq 2s_{a, b}e^{-d+6} \quad \text{for other } k, l.
\end{align*}
\]

To guarantee that the deviation of (2.4.25), (2.4.26) and (2.4.27) greater than \(\beta_{3, 1}\) (respectively \(\beta_{3, 2}\) and \(\beta_{3, 3}\)) happens with probability less than \(\frac{1}{n}\), \(n\) should satisfy \(\frac{n^{\beta_{3, 1}}}{\log(n)} \geq 6s_{a, b}e^{-d+4}\) (respectively \(\frac{n^{\beta_{3, 2}}}{\log(n)} \geq 6s_{a, b}e^{-d+8}\) and \(\frac{n^{\beta_{3, 3}}}{\log(n)} \geq 6s_{a, b}e^{-d+6}\)). By setting \(\beta_{3, 1} = \sqrt{6s_{a, b}\log(n)/n^{d/2}}\), \(\beta_{3, 2} = \sqrt{6s_{a, b}\log(n)/n^{d/2}}\), and \(\beta_{3, 3} = \sqrt{6s_{a, b}\log(n)/n^{d/2}}\)
the conditions \(\beta_{3, 1}/\varepsilon^3 \to 0, \beta_{3, 2}/\varepsilon^4 \to 0\) and \(\beta_{3, 3}/\varepsilon^3 \to 0\) as \(\varepsilon \to 0\) hold by the assumed relationship between \(n\) and \(\varepsilon\) and the deviations of (2.4.25), (2.4.26) and (2.4.27) are well controlled by \(\beta_{3, 1}, \beta_{3, 2}\) and \(\beta_{3, 3}\) respectively, with probability greater than \(1 - n^{-3}\). Define the deviation of \(\frac{1}{n\varepsilon^d}G_nG_n^\top\) from \(\mathbb{E} F_3\) as
\[
E := \frac{1}{n\varepsilon^d}G_nG_n^\top - \mathbb{E} F_3 \in \mathbb{R}^{p \times p}.
\]

As a result, again by a trivial union bound, with probability greater than \(1 - n^{-2}\), for all \(x_k\), we have
\[
\begin{align*}
|E_{a, b}| & \leq \frac{c\sqrt{\log(n)}}{n^{1/2}e^{d/2-2}} \quad \text{when } a, b = 1, \ldots, d, \\
|E_{a, b}| & \leq \frac{c\sqrt{\log(n)}}{n^{1/2}e^{d/2-4}} \quad \text{when } a, b = d + 1, \ldots, p, \\
|E_{a, b}| & \leq \frac{c\sqrt{\log(n)}}{n^{1/2}e^{d/2-3}} \quad \text{otherwise},
\end{align*}
\]
where
\[
c := \max_{a, b = 1, \ldots, p} \sqrt{6s_{a, b}}.
\]
Denote \(\Omega_3\) to be the event space that the deviation (2.4.29) is satisfied. With the above preparation, we now start the proof of Lemma 2.4.4 case by case.

**Case 0 in Condition 2.4.1.** Note that both \(\mathbb{E} F\) and \(\frac{1}{n\varepsilon^d}G_nG_n^\top\) are of rank \(r = d\) due to the geometric constraints. By the calculation in Section 1.4, when conditional on \(\Omega_3\), (2.4.28) holds, and the nonzero eigenvalues of
There are only \( d \) such eigenvalues) are deviated from the nonzero eigenvalues of \( EF \) by \( O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-2}}\right) \), which is smaller than \( \varepsilon^3 \) by the assumed relationship between \( n \) and \( \varepsilon \). Thus, \( r_n = d \) when \( \varepsilon \) is sufficiently small, and we have

\[
I_{p,r}(\tilde{\Lambda}_n + \varepsilon^p)^{-1}I_{p,r} - I_{p,d}(\tilde{\Lambda} + \varepsilon^p)^{-1}I_{p,d} = I_{p,d}[(\tilde{\Lambda}_n + \varepsilon^p)^{-1} - (\tilde{\Lambda} + \varepsilon^p)^{-1}]I_{p,d}.
\]

Denote the \( i \)-th eigenvalue of \( EF = \frac{1}{\varepsilon^p}C_i \) as \( \tilde{\lambda}_i \), where \( i = 1, \ldots, d \). By a direct calculation, we have

\[
|e_{i}^\top [\mathcal{F}_{e^p}(\tilde{\Lambda}_n) - \mathcal{F}_{e^p}(\tilde{\Lambda})]e_{i}| = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-2+2(2/\rho)}}\right)
\]

for \( i = 1, \ldots, d \) when \( \varepsilon \) is sufficiently small since we have

\[
\frac{1}{\tilde{\lambda}_i} + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-2}}\right) + \varepsilon^p - \frac{1}{\lambda_i} = \frac{1}{\lambda_i + \varepsilon^p} \left( O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-2}(\lambda_i + \varepsilon^p)}\right) + 1 \right)
\]

\[
= O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-2}(\lambda_i + \varepsilon^p)^2}\right) = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-2+2(2/\rho)}}\right)
\]

doing to the fact that \( \tilde{\lambda}_i \) is of order \( \varepsilon^2 \) for \( i = 1, \ldots, d \), \( \tilde{\lambda}_i + \varepsilon^p = O(\varepsilon^{2+p}) \) and \( n^{1/2}e^{d/2+1} \to \infty \) as \( n \to \infty \). Suppose there are \( 1 \leq l \leq d \) distinct eigenvalues, and the multiplicity of the \( j \)-th distinct eigenvalue is \( p_j \in \mathbb{N} \). Clearly, \( \Sigma_{j=1}^l p_j = p \). By the calculation in Section 1.4 that we skip the details, when conditional on \( \Omega_3 \), \( U_n = U\Theta + \frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-2}}U\Theta S + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-4+2(4/\rho)}}\right), \) where \( S \in o(p) \),

\[
\Theta = \begin{bmatrix}
X^{(1)} & 0 & \cdots & 0 \\
0 & X^{(2)} & \cdots & 0 \\
0 & 0 & \ddots & 0 \\
0 & 0 & \cdots & X^{(l)}
\end{bmatrix} \in O(p), \quad (2.4.31)
\]

and \( X^{(j)} \in O(p_j), \) \( j = 1, \ldots, l \), comes from the \( j \)-th distinct eigenvalue. Note that \( \Theta \) commutes with \( \tilde{\Lambda} \) and \( \mathcal{F}_{e^p}(\tilde{\Lambda}) \).

**Case 1 in Condition 2.4.1.** By the calculation in Section 1.4, when conditional on \( \Omega_3 \), the first \( d \) eigenvalues of \( \frac{1}{\varepsilon^p}G_n G_n^\top \) are deviated from the first \( d \) eigenvalues of \( EF \) by \( O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-2}}\right) \), which is smaller than \( \varepsilon^3 \), and the left \( p-d \) eigenvalues of \( \frac{1}{\varepsilon^p}G_n G_n^\top \) are deviated from the left \( p-d \) eigenvalues of \( EF \) by \( O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-2}}\right) \), which is smaller than \( \varepsilon^5 \). Thus, again, when \( \varepsilon \) is sufficiently small, \( r_n = r = p \), and \( I_{p,r}(\tilde{\Lambda}_n + \varepsilon^p)^{-1}I_{p,r} - I_{p,r}(\tilde{\Lambda} + \varepsilon^p)^{-1}I_{p,r} = [(\tilde{\Lambda}_n + \varepsilon^p)^{-1} - (\tilde{\Lambda} + \varepsilon^p)^{-1}] \). Therefore,

\[
|e_{i}^\top [\mathcal{F}_{e^p}(\tilde{\Lambda}_n) - \mathcal{F}_{e^p}(\tilde{\Lambda})]e_{i}| = \begin{cases}
O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-2+2(2/\rho)}}\right) & \text{for } i = 1, \ldots, d \\
O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-4+2(4/\rho)}}\right) & \text{for } i = d+1, \ldots, p
\end{cases}
\]

when \( \varepsilon \) is sufficiently small. Again, by the calculation in Section 1.4, when conditional on \( \Omega_3 \), we have \( U_n = U\Theta + \frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-2}}U\Theta S + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-4+2(4/\rho)}}\right), \) where \( S \in o(p) \) and \( \Theta \in O(p) \) is defined in (2.4.31).
Case 2 in Condition 2.4.1. A similar discussion holds. In this case, when conditional on \(\Omega_3\), we have

\[
|e_i^\top [\mathcal{I}_{n \varepsilon}(\tilde{\Lambda}_n) - \mathcal{I}_{n \varepsilon}(\tilde{\Lambda})] e_i| = \begin{cases} 
O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2 - 2} + 2(2/\rho)} \right) & \text{for } i = 1, \ldots, d, \\
O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2 - 4} + 2(4/\rho)} \right) & \text{for } i = d + 1, \ldots, p - l, \\
O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2 - 6} + 2(6/\rho)} \right) & \text{for } i = p - l + 1, \ldots, p 
\end{cases}
\]

when \(\varepsilon\) is sufficiently small. Similarly, when conditional on \(\Omega_3\), \(U_n = U \Theta + \frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2 - 2}} U \Theta S + O\left( \frac{\log(n)}{n \varepsilon^{d - 4}} \right)\), where \(S \in \alpha(p)\) and \(\Theta \in O(p)\) is defined in (2.4.31).

\[\square\]

Back to finish the proof of Theorem 2.4.1. Denote \(\Omega := \cap_{i=1, \ldots, d} \Omega_i\). It is clear that the probability of the event space \(\Omega\) is greater than \(1 - 4n^{-2}\). Below, all arguments are conditional on \(\Omega\). When \(\varepsilon\) is sufficiently small, based on Lemma (2.4.4), we have

\[
\mathcal{I}_{n \varepsilon}(\frac{1}{n \varepsilon^d} G_n G_n^\top) = \mathcal{I}_{n \varepsilon}(EF) + \delta_3,
\]

(2.4.32)

where

\[
\delta_3 := U_n \mathcal{I}_{n \varepsilon}(\tilde{\Lambda}_n) U_n^\top - U \mathcal{I}_{n \varepsilon}(\tilde{\Lambda}_n) U^\top
\]

\[
= \left( U \Theta + \frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2 - 2}} U \Theta S + O\left( \frac{\log(n)}{n \varepsilon^{d - 4}} \right) \right) \mathcal{I}_{n \varepsilon}(\tilde{\Lambda}) + \delta_{3,1}
\]

\[
\times \left( U \Theta + \frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2 - 2}} U \Theta S + O\left( \frac{\log(n)}{n \varepsilon^{d - 4}} \right) \right)^\top - U \mathcal{I}_{n \varepsilon}(\tilde{\Lambda}) U^\top.
\]

\[
= \frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2 - 2}} U \Theta S \mathcal{I}_{n \varepsilon}(\tilde{\Lambda}) + \mathcal{I}_{n \varepsilon}(\tilde{\Lambda}) S^\top U^\top + U \Theta \delta_{3,1} \Theta^\top U^\top + \text{[higher order terms]},
\]

and \(\delta_{3,1} := \mathcal{I}_{n \varepsilon}(\tilde{\Lambda}_n) - \mathcal{I}_{n \varepsilon}(\tilde{\Lambda})\), which bound is provided in Lemma (2.4.4). Define

\[
\delta_3 := \frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2 - 2}} U \Theta S \mathcal{I}_{n \varepsilon}(\tilde{\Lambda}) + \mathcal{I}_{n \varepsilon}(\tilde{\Lambda}) S^\top U^\top + U \Theta \delta_{3,1} \Theta^\top U^\top.
\]

(2.4.33)

By (2.4.3), we have

\[
\frac{1}{n \varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k) = EF_2 + \delta_2,
\]

(2.4.34)

where the bound of \(\delta_2\) is provided in Lemma 2.4.3. Similarly, we have

\[
\frac{1}{n \varepsilon^d} \sum_{j=1}^N (x_{k,j} - x_k)(f(x_{k,j}) - f(x_k)) = EF_4 + \delta_4,
\]

(2.4.35)

where the bound of \(\delta_4\) is the same as that in Lemma 2.4.3.
Table 2.1: The relevant items in each error term in $E_2^T \mathcal{J}_{\text{ep}}(E_F)E_F + E_F^T \mathcal{J}_{\text{ep}}(E_F)E_F + E_F^T G \mathcal{J}_{\text{ep}}(E_F)E_F$. The bounds are for entrywise errors. T means the tangential components in all Cases, N means the normal components in Case 1, and N1 means the first $p - d - l$ normal components of order $e^{d+4}$, and N2 means the last $l$ normal components of order $e^{d+6}$ in Case 2. “Total” means the overall bound of $E_2^T \mathcal{J}_{\text{ep}}(E_F)E_F + E_F^T \mathcal{J}_{\text{ep}}(E_F)E_F + E_F^T G \mathcal{J}_{\text{ep}}(E_F)E_F$, where only the major terms depending on $n$ and $\varepsilon$ in the leading order terms are shown.

<table>
<thead>
<tr>
<th>Case 0</th>
<th>Case 1</th>
<th>Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_F$</td>
<td>$E_2^T$</td>
<td>$E_2^T$</td>
</tr>
<tr>
<td>$\mathcal{J}_{\text{ep}}(E_F)$</td>
<td>$\varepsilon^{-(2,\rho)}$</td>
<td>$\varepsilon^{-(2,\rho)}$</td>
</tr>
<tr>
<td>$E_F^T$</td>
<td>$\varepsilon^2$</td>
<td>$\varepsilon^2$</td>
</tr>
<tr>
<td>$\mathcal{J}_{\text{ep}}(E_F)$</td>
<td>$\varepsilon^{-(2,\rho)}$</td>
<td>$\varepsilon^{-(4,\rho)}$</td>
</tr>
<tr>
<td>$E_F^T$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\mathcal{J}_{\text{ep}}(E_F)$</td>
<td>$\varepsilon^{-(2,\rho)}$</td>
<td>$\varepsilon^{-(4,\rho)}$</td>
</tr>
<tr>
<td>$E_F^T$</td>
<td>$\varepsilon^4$</td>
<td>$\varepsilon^4$</td>
</tr>
</tbody>
</table>

We could therefore recast $\frac{1}{n^d} \sum_{j=1}^N (x_{kj} - x_k)^T \mathcal{J}_{\text{ep}} \left( \frac{1}{n^d} G \mathcal{J}_{\text{ep}} G^T \right) \left[ \frac{1}{n^d} \sum_{j=1}^N (x_{kj} - x_k) (f(x_{kj}) - f(x_k)) \right]$ as

$$\frac{1}{n^d} \sum_{j=1}^N (x_{kj} - x_k)^T \mathcal{J}_{\text{ep}} \left( \frac{1}{n^d} G \mathcal{J}_{\text{ep}} G^T \right) \left[ \frac{1}{n^d} \sum_{j=1}^N (x_{kj} - x_k) (f(x_{kj}) - f(x_k)) \right]$$

We now control the error term $E_2^T \mathcal{J}_{\text{ep}}(E_F)E_F + E_F^T \mathcal{J}_{\text{ep}}(E_F)E_F + E_F^T \mathcal{J}_{\text{ep}}(E_F)E_F + E_F^T \mathcal{J}_{\text{ep}}(E_F)E_F + E_F^T \mathcal{J}_{\text{ep}}(E_F)E_F + E_F^T \mathcal{J}_{\text{ep}}(E_F)E_F$, which depends on the tangential and normal components. Since the errors are of different orders in the tangential and normal directions, we should evaluate the total error separately.

To avoid tedious description of each Case, we summarize the main order of each term for different Cases in Table 2.1. We mention that in Case 2, if the N1 part is zero; that is, the non-trivial eigenvalues corresponding to the normal bundle are all of order $e^{d+6}$, the final error rate is $\frac{\sqrt{\log(n)}}{n^{d/2}}$, which is the same as Case 0.

We only carry out the calculation for Case 1, and skip the details for the other cases since the calculation is the same. By checking the error order in Table 2.1, the leading order error term of $E_2^T \mathcal{J}_{\text{ep}}(E_F)E_F + E_F^T \mathcal{J}_{\text{ep}}(E_F)E_F + E_F^T \mathcal{J}_{\text{ep}}(E_F)E_F$ is controlled by $\frac{\sqrt{\log(n)}}{n^{d/2}} + \frac{\log(n)}{n^{d/2}}$, where $\frac{\log(n)}{n^{d/2}}$ comes from the tangential part, and $\frac{\sqrt{\log(n)}}{n^{d/2}}$ comes from the normal part. Note that the sizes of $(2 \land \rho) - 3$ and $(4 \land \rho) - 4$ depend on the chosen $\rho$, so we keep both. On the other hand, by (2.4.33) and Table 2.1, the error $E_F^2 \mathcal{J}_{\text{ep}}(E_F)$ is controlled by $\frac{\sqrt{\log(n)}}{n^{d/2}} + \frac{\log(n)}{n^{d/2}}$. By a direct comparison, it is clear that when $\varepsilon$ is sufficiently small, no matter which $\rho$ is chosen, $E_F^2 \mathcal{J}_{\text{ep}}(E_F)$ is dominated by $\frac{\sqrt{\log(n)}}{n^{d/2}} + \frac{\log(n)}{n^{d/2}}$, and hence the total error term is controlled by $\frac{\sqrt{\log(n)}}{n^{d/2}} + \frac{\log(n)}{n^{d/2}}$.

Therefore, when conditional on $\Omega$, for all $k = 1, \ldots, n$, the deviation of the nominator of (2.4.11) from $E[F_1] - E[F_2]^T \mathcal{J}_{\text{ep}}(E_F)E_F$ depends on $\rho$ and different Cases in Condition 2.4.1; for Case 0, it is controlled by $O(\frac{\log(n)}{n^{d/2}} + \frac{\log(n)}{n^{d/2}})$, which comes from the fact that $(2 \land \rho) - 3 \leq -1$ and $(4 \land \rho) - 4 = 0 \land (\rho - 4)$. Similarly, the deviation of the denominator of
(2.4.11) from $\mathbb{E}[F_0] - \mathbb{E}F_2^\top\mathcal{L}_\rho\mathbb{E}F_2$ depends on $\rho$ and different Cases in Condition 2.4.1; for Case 0, it is controlled by $O(n^{\log(n)/\sqrt{n^{1/2}d/2}}) + O(n^{\log(n)/\sqrt{n^{1/2}d/2}}) = O(n^{\log(n)/\sqrt{n^{1/2}d/2}})$ since $(2 \wedge \rho) - 3 \leq -1 < 0$; for Case 1 and Case 2, it is controlled by $O(n^{\log(n)/\sqrt{n^{1/2}d/2}}) + O(n^{\log(n)/\sqrt{n^{1/2}d/2}}) + O(n^{\log(n)/\sqrt{n^{1/2}d/2}}) = O(n^{\log(n)/\sqrt{n^{1/2}d/2}})$, which comes from the fact that $(4 \wedge \rho) - 4 = 0 \wedge (\rho - 4) \leq 0$.

As a result, when conditional on $\Omega$, for all $k = 1, \ldots, n$, we have

$$
\sum_{j=1}^{N} w_k(f(x_{k,j}) - f(x_k))
$$

(2.4.36)

which leads to

$$
\sum_{j=1}^{N} w_k(f(x_{k,j}) - f(x_k))
$$

(2.4.37)

where the equality comes from rewriting (2.4.12) as

$$
Qf(x_k) - f(x_k) = \frac{\mathbb{E}F_1 - \mathbb{E}F_2^\top\mathcal{L}_\rho\mathbb{E}F_2}{\mathbb{E}(F_0 - \mathbb{E}F_2^\top\mathcal{L}_\rho\mathbb{E}F_2)}
$$

and the fact that $\mathbb{E}F_0$ is of order 1 and $\mathbb{E}F_1$ is of order $\varepsilon^2$. Hence, we finish the proof.

### 2.5 Biased analysis on closed manifolds

For $f \in C(t(M))$, by the definition of $A$, we have

$$
Qf(x) = \frac{(Af)(x)}{(A1)(x)},
$$

(2.5.1)

where 1 means the constant function. We now provide an approximation of identity expansion of the $Q$ operator. By a direct expansion, we have

$$
Af(x) = \int_M K_{\text{LLE}}(x,y)f(t(y))P(y)dV(y).
$$

(2.5.2)
While the formula of the $Q$ operator looks like the diffusion process commonly encountered in the graph Laplacian based approach, like the DM [18], the proof and the result are essentially different. To ease the notation, define

$$
\mathcal{M}_0(x) := \frac{1}{|S^{d-1}|} \int_{S^{d-1}} \Pi_x(\theta, \theta) d\theta,
$$

$$
\mathcal{M}_2(x) := \frac{1}{|S^{d-1}|} \int_{S^{d-1}} \Pi_x(\theta, \theta) \theta \theta^\top d\theta,
$$

where $f \in C^3(M)$.

**Theorem 2.5.1.** Suppose $f \in C^3(M)$ and $P \in C^5(M)$ and fix $x \in M$. Assume that Assumptions 2.1.2 and 2.3.1 hold and the regularization order is $\rho \in \mathbb{R}$. Following the same notations used in Proposition 2.3.2, we have the following result

$$
Qf(x) - f(x) = (\mathcal{C}_1(x) + \mathcal{C}_2(x)) \varepsilon^2 + O(\varepsilon^3),
$$

where $\mathcal{C}_1(x)$ and $\mathcal{C}_2(x)$ depend on different cases stated in Condition 2.4.1.

- **Case 0.** In this case,

$$
\mathcal{C}_1(x) = \frac{1}{d+2} \left[ \frac{1}{2} \Delta f(x) + \frac{\nabla f(x) \cdot \nabla P(x)}{P(x)} - \frac{\nabla f(x) \cdot \nabla P(x)}{P(x) + \frac{d+2}{|S^{d-1}|} \varepsilon^2} \right],
$$

$$
\mathcal{C}_2(x) = 0.
$$

- **Case 1.** In this case,

$$
\mathcal{C}_1(x) = \frac{1}{d+2} \left[ \frac{1}{2} \Delta f(x) + \frac{\nabla f(x) \cdot \nabla P(x)}{P(x)} - \frac{\nabla f(x) \cdot \nabla P(x)}{P(x) + \frac{d+2}{|S^{d-1}|} \varepsilon^2} \right]
\left(1 - \frac{d}{2(d+2)} \sum_{i=d+1}^p \frac{(\mathcal{A}_i(x) e_i)^2}{\lambda_i^{(2)}} + \frac{d(d+2)}{P(x) |S^{d-1}|} \varepsilon^4 \right),
$$

$$
\mathcal{C}_2(x) = -\frac{1}{2(d+2)} \sum_{i=d+1}^p \lambda_i^{(1)} \frac{(\mathcal{A}_i(x) e_i)^2}{\lambda_i^{(2)}} + \frac{d(d+2)}{P(x) |S^{d-1}|} \varepsilon^4 \left(1 - \frac{d}{2(d+2)} \sum_{i=d+1}^p \frac{(\mathcal{A}_i(x) e_i)^2}{\lambda_i^{(2)}} + \frac{d(d+2)}{P(x) |S^{d-1}|} \varepsilon^4 \right).
$$

- **Case 2.** In this case,

$$
\mathcal{C}_1(x) = \frac{1}{d+2} \left[ \frac{1}{2} \Delta f(x) + \frac{\nabla f(x) \cdot \nabla P(x)}{P(x)} - \frac{\nabla f(x) \cdot \nabla P(x)}{P(x) + \frac{d+2}{|S^{d-1}|} \varepsilon^2} \right]
\left(1 - \frac{d}{2(d+2)} \sum_{i=d+1}^p \frac{(\mathcal{A}_i(x) e_i)^2}{\lambda_i^{(2)}} + \frac{d(d+2)}{P(x) |S^{d-1}|} \varepsilon^4 \right),
$$

$$
\mathcal{C}_2(x) = -\frac{1}{2(d+2)} \sum_{i=d+1}^p \lambda_i^{(1)} \frac{(\mathcal{A}_i(x) e_i)^2}{\lambda_i^{(2)}} + \frac{d(d+2)}{P(x) |S^{d-1}|} \varepsilon^4 \left(1 - \frac{d}{2(d+2)} \sum_{i=d+1}^p \frac{(\mathcal{A}_i(x) e_i)^2}{\lambda_i^{(2)}} + \frac{d(d+2)}{P(x) |S^{d-1}|} \varepsilon^4 \right).
$$

Intuitively, based on the approximation of the identity, the kernel representation of the $Q$ operator suggests that
asymptotically we get the function value back, with the second order derivative popping out in the second order error term. In the GL setup, it has been well known that the second order derivative term is the Laplace-Beltrami operator when the p.d.f. is constant [18]. However, due to the interaction between the geometric structure and the barycentric coordinate, LLE usually does not lead to the Laplace-Beltrami operator, unless under special situations. Note that while we could still see the Laplace-Beltrami operator in \( C_1 \), it is contaminated by other quantities, including \( \eta_l(x), \tilde{\eta}_f(x) \) and \( \lambda_l^{(2)} \). These terms all depend on the second fundamental form. When \( \rho > 4 \), the curvature term appears in the \( \varepsilon^2 \) order term.

This theorem states that the asymptotic behavior of LLE is sensitive to the choice of \( \rho \). We discuss each case based on different choices of \( \rho \). If \( \rho < 2 \), for all cases,

\[
\mathcal{C}_1(x) = \frac{1}{(d+2)} \left[ \frac{1}{2} \Delta f(x) + \frac{\nabla f(x) \cdot \nabla P(x)}{P(x)} \right] \quad \text{and} \quad \mathcal{C}_2(x) = 0, \tag{2.5.11}
\]

which comes from the fact that when \( \varepsilon^p \) is large, \( T_{1(i)} \) is small, and hence \( K_{\text{LLE}} \) is dominated by 1. Note that not only the Laplacian-Beltrami operator but also the p.d.f are involved, if the sampling is non-uniform. Therefore, when \( \rho \) is chosen to be too small, the resulting asymptotic operator is the Laplace-Beltrami operator, only when the sampling is uniform. If \( \rho = 3 \), for all cases we have

\[
\mathcal{C}_1(x) = \frac{1}{2(d+2)} \Delta f(x) \quad \text{and} \quad \mathcal{C}_2(x) = 0. \tag{2.5.12}
\]

In this case, we recover the Laplacian-Beltrami operator, and the asymptotic result of LLE is independent of the non-uniform p.d.f.. This theoretical finding partially explains why such regularization could lead to a good result. If \( \rho > 4 \), since \( \varepsilon^{d+p} \) is smaller than all eigenvalues of the local covariance matrix, asymptotically \( \varepsilon^{d+p} \) is negligible and the result depends on different cases considered in Condition 2.4.1: for Case 0, we have

\[
\mathcal{C}_1(x) = \frac{1}{2(d+2)} \Delta f(x) \quad \text{and} \quad \mathcal{C}_2(x) = 0,
\]

for Case 1, we have

\[
\mathcal{C}_1(x) = \frac{1}{2(d+2)} \Delta f(x) \quad \text{and} \quad \mathcal{C}_2(x) = -\frac{d}{8(d+4)} \sum_{i=d+1}^p \frac{(\eta_l(0)(x)\tilde{\eta}_f(0)(x)\eta_l(0)(x))}{\lambda_l^{(2)}},
\]

and for Case 2, we have

\[
\mathcal{C}_1(x) = \frac{1}{2(d+2)} \Delta f(x) \quad \text{and} \quad \mathcal{C}_2(x) = -\frac{d}{8(d+4)} \sum_{i=d+1}^{p-l} \frac{(\eta_l(0)(x)\tilde{\eta}_f(0)(x)\eta_l(0)(x))}{\lambda_l^{(2)}}.
\]

Note that when \( \rho > 4 \), we do not get the Laplace-Beltrami operator asymptotically in Cases 1 and 2. Furthermore, the behavior of LLE is dominated by the curvature and is independent of the p.d.f..

It is worth mentioning a specific situation when \( \rho > 4 \). Suppose the principal curvatures are equal to \( p \in \mathbb{R} \) in the direction \( e_i \), where \( i = d + 1, \ldots, p \), and vanish in the other directions. Then, there is a choice of basis \( e_1, \ldots, e_d \) so that \( \| x(\theta, \tilde{\theta}) \cdot e_i = \sum_{j=1}^d p \theta_j^2 = p \), where \( \theta = (\theta_1, \ldots, \theta_d) \in S^{d-1} \). Under this specific situation, by a
direct expansion, we have a simplification that
\[
d\frac{(\Omega_0^T(x)e_i)(\Omega_f^T(x)e_i)}{8(d+4)} = \frac{1}{2(d+2)}\Delta f(x),
\]
which leads to \(c_4(x) + c_5(x) = 0\). Therefore, asymptotically we obtain a fourth order term.

We mention that the statement “suppose \(\varepsilon\) is sufficiently small” in Proposition 2.3.1, Proposition 2.3.2 and Theorem 2.5.1 is a technical condition needed in the proof of Lemma 2.2.3, which describes how well we could estimate the local geodesic distance by the ambient space metric. This technical condition depends on the fact that the exponential map is a diffeomorphism only if it is restricted to a subset of \(T_xM\) that is bounded by the injectivity radius of the manifold. That is, \(\varepsilon\) needs to be less than the injectivity radius. For any closed (compact without boundary) and smooth manifold, it is clear that different kinds of curvatures are bounded and the injectivity radius is strictly positive, so there exists \(\varepsilon_0 > 0\) less than the injectivity radius, so that for all \(\varepsilon \leq \varepsilon_0\), the statement “suppose \(\varepsilon\) is sufficiently small” is satisfied. The relationship between the curvature and the \(\varepsilon_0\) could be further elaborated by quoting the well known result in \([16]\): for a closed Riemannian manifolds of dimension \(d\) with the sectional curvature bounded by \(K\), where \(K \geq 0\), and with the volume lower bound \(v\), where \(v > 0\), the injectivity radius is bounded below by \(i(d,K,v) > 0\), where \(i(d,K,v)\) can be expressed explicitly in terms of \(d, K\) and \(v\). Hence, \(\varepsilon_0\) needs to satisfy \(\varepsilon_0 < i(d,K,v)\).

### 2.5.1 Proof of Theorem 2.5.1

We need the following Proposition for the proof.

**Proposition 2.5.1.** Suppose \(l = \text{nullity}(M_{22}^{(2)}) > 0\) and Assumptions 2.1.2 and 2.3.1 hold. Then \(\langle II_x(\theta, \theta), e_i \rangle = 0\) for \(p - l + 1 \leq i \leq p\). Moreover, for \(m,n = p - l + 1, \ldots, p\), we have
\[
\begin{align*}
\left[M_{22,22}^{(4)} - 2M_{21,21}^{(2)}M_{12,22}^{(2)}\right]_{m-p+l,n-p+l} &= \frac{d(d+2)}{36(d+6)|S^{d-1}|} \int_{S^{d-1}} \langle \nabla_\theta II_x(\theta, \theta), e_m \rangle \langle \nabla_\theta II_x(\theta, \theta), e_n \rangle d\theta \\
&- \frac{d^2(d+2)^2}{18|S^{d-1}|^2(d+4)^2} \sum_{k=1}^d \int_{S^{d-1}} \langle \nabla_\theta II_x(\theta, \theta), e_m \rangle \langle e_i, e_k \rangle d\theta \int_{S^{d-1}} \langle e_i, e_k \rangle \langle \nabla_\theta II_x(\theta, \theta), e_n \rangle d\theta.
\end{align*}
\]

This Proposition essentially says that if \(\text{nullity}(M_{22}^{(2)}) = l > 0\) and \(M_{22}^{(2)}\) is diagonalized as in (2.3.12), then geometrically \(e_{p-l+1}, \ldots, e_p\) are perpendicular to the second fundamental form \(II_x(\theta, \theta)\). Furthermore, the eigenvalues of order \(e^{d+l+6}\) in Case 2 of Proposition 2.3.2 depend only on the third order derivative of the embedding, \(\nabla_\theta II_x(\theta, \theta)\), in those directions.

**Proof.** Suppose \(l = \text{nullity}(M_{22}^{(2)}) > 0\). By Assumption 2.3.1, \(M_{22}^{(2)}\) is diagonalized as in (2.3.12). Therefore, based on (2.3.6) and (2.3.10), we have
\[
\int_{S^{d-1}} \langle II_x(\theta, \theta), e_m \rangle \langle II_x(\theta, \theta), e_m \rangle d\theta = 0, \tag{2.5.13}
\]
where \(m = p - l + 1, \ldots, p\).

If we denote \(\theta = \theta^i \partial_i \in S^{d-1} \subset T_xM\), the following expression for the second fundamental form holds:
\[
\langle II_x(\theta, \theta), e_m \rangle = \sum_{i=1}^d p_{ii}^m \theta^i + 2 \sum_{i<j} p_{ij}^m \theta^i \theta^j, \tag{2.5.14}
\]
where \( p^m_{ij} = \langle \Pi_i(\partial_i, \partial_j), e_m \rangle \in \mathbb{R}, i, j = 1, \ldots, d \), are the corresponding coefficients. Note that \( \iota_i, \partial_i = e_i \) for \( i = 1, \ldots, d \). By plugging (2.5.14) into (2.5.13), we have

\[
0 = \int_{S^{d-1}} \left[ \sum_{i=1}^d (p^m_{ii})^2 + 4 \sum_{k=1}^d p^m_{kk} \sum_{i<j} p^m_{ij} \theta^i \theta^j + 4 \left( \sum_{i<j} p^m_{ij} \theta^i \theta^j \right)^2 \right] d\theta \\
= \frac{1}{d(d+2)} |S^{d-1}| \left( \sum_{i=1}^d (p^m_{ii})^2 + 2 \sum_{i<j} p^m_{ii} p^m_{jj} + 4 \sum_{i<j} (p^m_{ij})^2 \right) \\
= 2 \sum_{i=1}^d (p^m_{ii})^2 + \sum_{i<j} (p^m_{ij})^2,
\]

which leads to the conclusion that \( p^m_{ij} = 0 \) for all \( i \) and \( j \). To get the expansion of \( [M^{(4)}_{22,22} - 2M^{(2)}_{21,2}M^{(2)}_{12,2}]_{m-p+l,n-p+l} \), we directly plug the above formula to (2.3.5) and (2.3.7) and get the claim.

Recall the definition of \( T_{i(x)} = \mathcal{J}_{\sqrt{d-1}} \left( C_x \right) \left[ \mathbb{E}(X - \iota(x)) \chi_{\mathbb{R}^p} \right] \) in (2.4.9), which could be expanded as \( \sum_{i=1}^r \frac{\mathbb{E}(X - \iota(x)) \chi_{\mathbb{R}^p} \left( C_x \right)}{\lambda_i + \sqrt{d-1}^p} \in \mathbb{R}^p \), where \( r \) is the rank of \( C_x \) and \( \lambda_i \) form the \( i \)-th eigen-pair of \( C_x \). Clearly, \( T_{i(x)} \) is dominated by those “small” eigenvalues of \( C_x \).

Define the notation to simplify the statement of the next lemma:

\[
\bar{J} := \mathcal{J}_{p-p-d} \mathcal{J}_{p-d-p-d-l} \in \mathbb{R}^{p \times (p-d-l)}.
\]

**Lemma 2.5.1.** Fix \( x \in M \) and assume Assumptions 2.1.2 and 2.3.1 hold. Suppose \( \epsilon \) is sufficiently small. Following the same notations used in Proposition 2.3.2, under three conditions shown in Condition 2.4.1, \( T_{i(x)} \) satisfies:

**Case 0.** \( T_{i(x)} = [v_1, v_2] + [O(\epsilon^2), 0] \), where

\[
v_1 = \frac{\mathcal{J}_{p-d} \iota_i \nabla P(x)}{P(x) + \left( \frac{d(d+2)}{|S^{d-1}|} \right) \epsilon^{d-2} + O(\epsilon^2)}, \quad v_2 = 0.
\]

**Case 1.** \( T_{i(x)} = [v_1, v_2] + [O(\epsilon^2), O(1)] \), where

\[
v_1 = \frac{\mathcal{J}_{p-d} \iota_i \nabla P(x)}{P(x) + \left( \frac{d(d+2)}{|S^{d-1}|} \right) \epsilon^{d-2} + \sum_{i=d+1}^p \frac{\mathcal{N}_0(x) \mathcal{J}_{p-d} \mathcal{I}_x \mathcal{J}_{p-d}^T \mathcal{I}_x \mathcal{I}_x \mathcal{J}_{p-d}^T d \lambda_i}{2 \lambda_i} + \frac{2(d+2)}{P(x)|S^{d-1}|} \epsilon^{d-4} X_1 S_{12} \mathcal{J}_{p-d}^T d \lambda_i},
\]

\[
v_2 = \frac{1}{\epsilon^2} \sum_{i=d+1}^p \frac{\mathcal{N}_0(x) \mathcal{J}_{p-d} \mathcal{I}_x \mathcal{J}_{p-d}^T \mathcal{I}_x \mathcal{I}_x \mathcal{J}_{p-d}^T d \lambda_i}{2 \lambda_i} + \frac{2(d+2)}{P(x)|S^{d-1}|} \epsilon^{d-4} X_2 \mathcal{J}_{p-d}^T d \lambda_i,
\]

and \( \mathcal{N}_0(x) \) is defined in (2.5.3).
\textbf{Case 2.} \( T_{1(x)} = [v_1, v_2] + [O(\varepsilon^2), O(1)] \), where

\[
v_1 = \frac{J_{\rho,d}^\top t_s \nabla P(x)}{P(x)} + \sum_{i = d+1}^{p-1} \frac{\gamma_i^{(2)}}{\lambda_i(\rho)} + \frac{\gamma_i^{(4)}}{\lambda_i(\rho)} - \frac{\gamma_i^{(2)}}{\lambda_i(\rho)} - \frac{2^{(2)}}{\lambda_i(\rho)} \varepsilon^{\rho - 2} - \frac{\gamma_i^{(4)}}{\lambda_i(\rho)} \varepsilon^{\rho - 2}
\]

\[
v_2 = \frac{1}{\varepsilon^2} \sum_{i = d+1}^{p-1} \frac{\gamma_i^{(2)}}{\lambda_i(\rho)} + \frac{\gamma_i^{(4)}}{\lambda_i(\rho)} - \frac{2^{(2)}}{\lambda_i(\rho)} \varepsilon^{\rho - 2} - \frac{\gamma_i^{(4)}}{\lambda_i(\rho)} \varepsilon^{\rho - 2}
\]

where \( \alpha_i \in \mathbb{R} \) is defined in (2.5.19).

\[ \text{Proof.} \] We show the lemma case by case, and we will recycle the equations shown in Lemma 2.2.5. Note that although the eigenvectors of \( C_x \) might not be unique, we will see that the result is independent of the choice of the eigenvectors.

\textbf{Case 0 in Condition 2.4.1.} In this case, by Proposition 2.3.2, denote the \( i \)-th eigenvector of \( C_x \) as \( u_i = \begin{bmatrix} X_i J_{\rho,d}^\top e_i + O(\varepsilon^2) \\ 0 \end{bmatrix} \), where \( i = 1, \ldots, d \) and \( X_i \in O(d) \), and the corresponding eigenvalue \( \lambda_i = \frac{\|S_{x(d)}\|}{d(x)} \varepsilon^{d+2} + O(\varepsilon^{d+4}) \). By Lemma 2.2.5 and Lemma 2.3.2, we have \( 1 \leq i \leq d \)

\[
\mathbb{E}[X_i J_{\rho,d}^\top e_i] = \frac{\|S_{x(d)}\|}{d(x)} \varepsilon^{d+2} + O(\varepsilon^{d+4})
\]

where the last equality comes from the fact that \( [J_{\rho,d}^\top t_s \nabla P(x), X_i J_{\rho,d}^\top e_i] = e_i^\top J_{\rho,d}^\top X_i^\top t_s \nabla P(x) = u_i^\top t_s \nabla P(x) \). Thus,

\[
\mathbb{E}[X_i J_{\rho,d}^\top e_i] = \frac{\|S_{x(d)}\|}{d(x)} \varepsilon^{d+2} + O(\varepsilon^{d+4})
\]

where the last expansion holds for all chosen regularization order \( \rho \). Specifically, when \( \rho > 2 \), it is trivial; when \( \rho \leq 2 \),

\[
\frac{u_i^\top t_s \nabla P(x)}{P(x)} + \frac{d(x)}{\|S_{x(d)}\| d(x)} \varepsilon^{d+2} + O(\varepsilon^{d+4})
\]

is of order smaller than \( \varepsilon^2 \) since the denominator is dominated by
The eigenvalues of $\varepsilon^p$. Hence, since $u_i$ for an orthonormal set, we have

$$ T_{t(x)} = \sum_{i=1}^{d} \frac{\mathbb{E}[(X - t(x)) \chi_{B_2^p (x_k)} (X)] \cdot u_i}{\lambda_i + \varepsilon^{d + p}} u_i $$

$$ = \sum_{i=1}^{d} \left( \frac{u_i^\top t \nabla P(x)}{P(x) + \frac{d(d+2)}{\varepsilon^p - 1} \varepsilon^{d + 2} + O(\varepsilon^2)} + O(\varepsilon^2) \right) \left[ X_i J_{p,d}^\top e_i + O(\varepsilon^2), 0 \right] $$

$$ = \left[ \frac{J_{p,d}^\top t \nabla P(x)}{P(x) + \frac{d(d+2)}{\varepsilon^p - 1} \varepsilon^{d + 2} + O(\varepsilon^2)}, 0 \right] + \| O(\varepsilon^2), 0 \|. $$

Case 1 in Condition 2.4.1. The eigenvalues of $C$, are $\lambda_i = \frac{\langle \xi \rangle (\varepsilon^{d+2} + \lambda_i^{(2)} \varepsilon^{d+4} + O(\varepsilon^{d+6}))}{d(d+2)}$ for $i = 1, \ldots, d$ and $\lambda_i = \frac{\langle \xi \rangle (\varepsilon^{d+4} + O(\varepsilon^{d+6}))}{d(d+2)}$ for $i = d + 1, \ldots, p$. The eigenvectors of $C$ are

$$ u_i = \begin{bmatrix} X_i J_{p,d}^\top e_i \\ 0_{(p-d) \times 1} \end{bmatrix} + \varepsilon^2 U_\varepsilon(0) S e_i + O(\varepsilon^4) = \left[ X_i J_{p,d}^\top e_i + O(\varepsilon^2), O(\varepsilon^2) \right] $$

for $i = 1, \ldots, d$, where $U_\varepsilon(0) = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \in O(p)$, and

$$ u_i = \begin{bmatrix} 0_{d \times 1} \\ X_2 J_{p,d}^\top e_i \end{bmatrix} + \varepsilon^2 U_\varepsilon(0) S e_i + O(\varepsilon^4) = \left[ J_{p,d}^\top U_\varepsilon(0) S e_i + O(\varepsilon^4), X_2 J_{p,d}^\top e_i + O(\varepsilon^2) \right] $$

for $i = d + 1, \ldots, p$. X_{1} \in O(d)$ and $X_2 \in O(p - d)$. For $1 \leq i \leq d$, we have

$$ \mathbb{E}[(X - t(x)) \chi_{B_2^p (x_k)} (X)] \cdot u_i $$

$$ = \frac{\langle \xi \rangle}{d + 2} \left[ \frac{t \nabla P(x)}{d} e^{d+2} + O(\varepsilon^{d+4}), \frac{P(x) J_{p,d}^\top \delta(x)}{2} e^{d+2} + O(\varepsilon^{d+4}) \right] \cdot \left[ X_i J_{p,d}^\top e_i + O(\varepsilon^2), O(\varepsilon^2) \right] $$

$$ = \frac{\langle \xi \rangle}{d(d+2)} \frac{t \nabla P(x)}{d} e^{d+2} + O(\varepsilon^{d+4}) $$

and hence

$$ \mathbb{E}[(X - t(x)) \chi_{B_2^p (x_k)} (X)] \cdot u_i $$

$$ = \frac{\langle \xi \rangle \varepsilon^2}{d(d+2)} \frac{t \nabla P(x)}{d} e^{d+2} + O(\varepsilon^{d+4}) $$

$$ = \left[ \frac{\langle \xi \rangle \varepsilon^2}{d(d+2)} \frac{t \nabla P(x)}{d} e^{d+2} + O(\varepsilon^{d+4}), \right] $$

$$ = \left[ \frac{\langle \xi \rangle \varepsilon^2}{d(d+2)} \frac{t \nabla P(x)}{d} e^{d+2} + O(\varepsilon^{d+4}), O(\varepsilon^2) \right] $$
Since columns of $J_{p,d}X_1$ form an orthonormal basis of $t, T_xM$, we have

$$\sum_{i=1}^{d} \frac{\mathbb{E}[(X - t(x))X_{\mathbb{B}_{2}^d, \langle x \rangle}(X)] \cdot u_i}{\lambda_i + \varepsilon^{d+p}} u_i = \sum_{i=1}^{d} \left[ \frac{\left( t, \nabla P(x) \right) J_{p,d}X_1 J_{p,d}^\top e_i}{P(x) + \frac{d(d+2)}{(d+1)^2} \varepsilon^{p-2}} X_1 J_{p,d}^\top e_i, 0 \right] + \left[ O(\varepsilon^2), O(\varepsilon^2) \right]$$

$$= \left[ \frac{J_{p,d}^\top \nabla P(x)}{P(x) + \frac{d(d+2)}{(d+1)^2} \varepsilon^{p-2}} + O(\varepsilon^2), O(\varepsilon^2) \right].$$

For $d + 1 \leq i \leq p$, similarly we have

$$\mathbb{E}[(X - t(x))X_{\mathbb{B}_{2}^d, \langle x \rangle}(X)] \cdot u_i = \frac{|S^{d-1}|}{d+2} \left[ \frac{J_{p,d}^\top U_1(0)X_1(0)Se_i \varepsilon^2 + O(\varepsilon^4), X_2J_{p,d}^\top e_i + O(\varepsilon^2)}{2} \right]$$

$$= \left[ \frac{|S^{d-1}|}{2(d+2)} P(x) \eta_0(x) J_{p,d}^\top X_2 J_{p,d}X_2 J_{p,d}^\top e_i + O(\varepsilon^4) \right],$$

and hence

$$\mathbb{E}[(X - t(x))X_{\mathbb{B}_{2}^d, \langle x \rangle}(X)] \cdot u_i = \frac{\lambda_i + \varepsilon^{d+p}}{\lambda_i + \varepsilon^{d+p}} u_i$$

$$= \left[ \frac{|S^{d-1}|}{2} \left( \frac{1}{\lambda_i} + \frac{2(d+2)}{(d+1)^2} \varepsilon^{p-4} \right) + \frac{2(d+2)}{(d+1)^2} \varepsilon^{p-4} \right] \left[ O(\varepsilon^2), O(1) \right].$$

As a result, by the fact that $J_{p,d}^\top U_1(0)Se_i = X_1S_1J_{p,d}^\top e_i$ when $i = d + 1, \ldots, p$, in this case we have

$$T_{1(x)} = \left[ \frac{J_{p,d}^\top \nabla P(x)}{P(x) + \frac{d(d+2)}{(d+1)^2} \varepsilon^{p-2}} + \sum_{i=d+1}^{p} \frac{\eta_0(x) J_{p,d}^\top X_2 J_{p,d}^\top e_i}{2} \lambda_i \left( \frac{1}{\lambda_i} + \frac{2(d+2)}{(d+1)^2} \varepsilon^{p-4} \right) X_1S_1J_{p,d}^\top e_i \right] + \left[ O(\varepsilon^2), O(1) \right].$$

**Case 2 in Condition 2.4.1.** In this case, the eigenvalues of $C_\delta$ are

$$\lambda_i = \begin{cases} 
\frac{|S^{d-1}|}{d(d+2)}, \varepsilon^{d+2} + O(\varepsilon^{d+4}) & \text{for } i = 1, \ldots, d \\
\frac{|S^{d-1}|}{d(d+2)} \lambda_i \left( \frac{1}{\lambda_i} + \frac{2(d+2)}{(d+1)^2} \varepsilon^{p-4} \right) & \text{for } i = d + 1, \ldots, p - 1 \\
\frac{|S^{d-1}|}{d(d+2)} \lambda_i \left( \frac{1}{\lambda_i} + \frac{2(d+2)}{(d+1)^2} \varepsilon^{p-4} \right) & \text{for } i = p - 1, \ldots, p.
\end{cases}$$
Adapt notations from Proposition 2.3.2 and use

\[ U_\lambda(0) = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_{2,1} & 0 \\ 0 & 0 & X_{2,2} \end{bmatrix} \in O(p), \quad S = \begin{bmatrix} S_{11} & S_{12,1} & S_{12,2} \\ S_{21,1} & S_{22,1} & S_{22,2} \\ S_{21,2} & S_{22,1} & S_{22,2} \end{bmatrix} \in o(p), \]

where \( X_1 \in O(d), X_{2,1} \in O(p - d - l) \) and \( X_{2,2} \in O(l) \). The eigenvectors of \( C_x \), on the other hand, are \( u_i = \begin{bmatrix} X_1 J_\lambda^p \epsilon_i \\ 0_{(p-d) \times 1} \end{bmatrix} + \epsilon^2 U_\lambda(0) S \epsilon_i + O(\epsilon^4) \) for \( i = 1, \ldots, d, \) \( u_i = \begin{bmatrix} 0_{d \times 1} \\ X_{2,1} J_\lambda^1 \epsilon_i \end{bmatrix} + \epsilon^2 U_\lambda(0) S \epsilon_i + O(\epsilon^4) \) for \( i = d + 1, \ldots, p - l, \) and \( u_i = \begin{bmatrix} 0_{d \times 1} \\ X_{2,2} J_\lambda^p \epsilon_i \end{bmatrix} + \epsilon^2 U_\lambda(0) S \epsilon_i + O(\epsilon^4) \) for \( i = p - l + 1, \ldots, p. \)

Similar to Case 1, we could evaluate \( \frac{\mathbb{E}[(X - t(x)) X_{p}^{\lambda \epsilon} \epsilon \{X\}] \cdot u_i}{\lambda + \epsilon^{d+p}} u_i \) for \( i = 1, \ldots, p - l, \) and have

\[
\frac{p-l}{\lambda + \epsilon^{d+p}} \mathbb{E}[(X - t(x)) X_{p}^{\lambda \epsilon} \epsilon \{X\}] \cdot u_i = \frac{\frac{1}{\epsilon^2} \sum_{i=d+1}^{p-l} \frac{\mathbb{E}[(X - t(x)) X_{2,1} J_\lambda^1 \epsilon_i \epsilon e_i]}{2 \mathcal{A}_i(2)} + \frac{2(d+2)}{p(x) \mathcal{S}^{d-1}} \epsilon e_i}{\mathcal{A}_i(2)} + \frac{1}{2} \mathcal{A}_i(2) + \frac{2(d+2)}{p(x) \mathcal{S}^{d-1}} \epsilon e_i} \left[ \mathcal{O}(\epsilon^2), \mathcal{O}(1) \right].
\]

For \( p - l + 1 \leq i \leq p, \) base on Proposition 3.2.3, we have

\[ \mathbb{E}[(X - t(x)) X_{p}^{\lambda \epsilon} \epsilon \{X\}] \cdot u_i = \frac{\mathbb{E}[(X - t(x)) X_{0}^{\lambda \epsilon} \epsilon \{X\}] \cdot u_i}{\lambda + \epsilon^{d+p}} u_i \]

and hence

\[
\mathbb{E}[(X - t(x)) X_{p}^{\lambda \epsilon} \epsilon \{X\}] \cdot u_i = \frac{\mathcal{A}_i(2)}{2} \left[ \frac{\mathcal{A}_i(2)}{p(x) \mathcal{S}^{d-1}} \epsilon e_i^2 + O(\epsilon^2) \right] + \frac{1}{2} \mathcal{A}_i(2) + \frac{2(d+2)}{p(x) \mathcal{S}^{d-1}} \epsilon e_i}
\]

where we use the fact that \( J_\lambda^p U_\lambda(0) S \epsilon_i = X_1 S_{12,2} J_\lambda^p \epsilon_i \) when \( i = p - l + 1, \ldots, p, \) and \( \alpha_i \in \mathbb{R} \) is the coefficient of the order \( \epsilon^{d+4} \) term. Note that the \( \epsilon^{d+2} \) term disappears due to (2.5.18). We mention that since \( \alpha_i \) will be canceled
out in the main Theorem, we do not spell it out explicitly. Therefore,

\[
\sum_{i=p-l+1}^{p} \frac{\mathbb{E}[(X - t(x))X_{\mathcal{G}^p t(x)}(X)] \cdot u_i}{\lambda_i + \varepsilon^{d+p} u_i} = \sum_{i=p-l+1}^{p} \frac{\alpha_i \varepsilon^{d+4} + O(\varepsilon^{d+6})}{\frac{\lvert d(x) \rvert}{d(d+2)} \lambda_i^{(4)} + \varepsilon^{d+p} + O(\varepsilon^{d+8})} \times \left[ X_1 S_{12,2} J^\top_{p,j} e_i e^2 + O(\varepsilon^4), X_{2,2} J^\top_{p,j} e_i + O(\varepsilon^2) \right] \\
= \sum_{i=p-l+1}^{p} \frac{\alpha_i}{\frac{\lvert d(x) \rvert}{d(d+2)} \lambda_i^{(4)} + \varepsilon^{p-6}} \left[ X_1 S_{12,2} J^\top_{p,j} e_i + X_{2,2} J^\top_{p,j} e_i \frac{1}{\varepsilon^2} \right] + [O(\varepsilon^2), O(1)].
\]

As a result, in this case we have

\[
T_{i(x)} = \|v_1, v_2\| + \|O(\varepsilon^2), O(1)\|,
\]

where

\[
v_1 = \frac{J_{p,d}^\top \nabla P(x)}{P(x)} + \sum_{i=d+1}^{p-l} \frac{Q_{0,i}^\top(x) J X_{2,1} J^\top e_i}{\frac{\lvert d(x) \rvert}{d(d+2)} \lambda_i^{(2)} + \frac{2\varepsilon^{d+2}}{\mathcal{P}(m) \mathcal{R}(x)} \varepsilon^{p-4}} X_{2,1} S_{12,1} J^\top e_i + \sum_{i=p-l+1}^{p} \frac{\alpha_i}{\frac{\lvert d(x) \rvert}{d(d+2)} \lambda_i^{(4)} + \varepsilon^{p-6}} X_1 S_{12,2} J^\top_{p,j} e_i
\]

and

\[
v_2 = \frac{1}{\varepsilon^2} \sum_{i=d+1}^{p-l} \frac{Q_{0,i}^\top(x) J X_{2,1} J^\top e_i}{\frac{\lvert d(x) \rvert}{d(d+2)} \lambda_i^{(2)} + \frac{2\varepsilon^{d+2}}{\mathcal{P}(m) \mathcal{R}(x)} \varepsilon^{p-4}} X_{2,1} J^\top e_i + \frac{1}{\varepsilon^2} \sum_{i=p-l+1}^{p} \frac{\alpha_i}{\frac{\lvert d(x) \rvert}{d(d+2)} \lambda_i^{(4)} + \varepsilon^{p-6}} X_{2,2} J^\top_{p,j} e_i.
\]

\[\square\]
We introduce the following notations to simplify the proof:

\[
\omega(x) := \frac{1}{|S^{d-1}|} \int_{S^{d-1}} \| \Pi_x(\theta, \theta) \|^2 d\theta
\]

\[
\mathcal{M}_1(x) := \frac{1}{|S^{d-1}|} \int_{S^{d-1}} \| \Pi_x(\theta, \theta) \|^2 \Pi_x(\theta, \theta) d\theta
\]

\[
\mathcal{M}_2(x) := \frac{1}{|S^{d-1}|} \int_{S^{d-1}} \Pi_x(\theta, \theta) \mathcal{R}_x(\theta, \theta) d\theta
\]

\[
\mathcal{M}_3(x) := \frac{1}{|S^{d-1}|} \int_{S^{d-1}} \Pi_x(\theta, \theta) \Pi_x(\theta, \theta) d\theta
\]

\[
\mathcal{M}_4(x) := \frac{1}{|S^{d-1}|} \int_{S^{d-1}} \Pi_x(\theta, \theta) \Pi_x(\theta, \theta) \Pi_x(\theta, \theta) d\theta
\]

\[
\mathcal{M}_5(x) := \frac{1}{|S^{d-1}|} \int_{S^{d-1}} \Pi_x(\theta, \theta) \Pi_x(\theta, \theta) \Pi_x(\theta, \theta) \Pi_x(\theta, \theta) d\theta
\]

\[
\mathcal{M}_6(x) := \frac{1}{|S^{d-1}|} \int_{S^{d-1}} \Pi_x(\theta, \theta) \Pi_x(\theta, \theta) \Pi_x(\theta, \theta) \Pi_x(\theta, \theta) \Pi_x(\theta, \theta) d\theta
\]

For \( f \in C^3(\mathcal{M}) \) and \( P \in C^5(\mathcal{M}) \), define

\[
\Omega_f := \frac{1}{2} \nabla f(x) \nabla \mathcal{M}_2(x) \mathcal{P}(x) + \frac{1}{4} P(x) \nabla (\mathcal{M}_2(x) \nabla^2 f(x)) + \frac{1}{6} P(x) \mathcal{M}_1(x) \nabla f(x)
\]

\[
\mathcal{I}_f := \frac{1}{|S^{d-1}|} t_s \int_{S^{d-1}} \theta (\frac{1}{6} \nabla_{\theta, \theta, \theta} f(x) P(x) + \frac{1}{6} \nabla_{\theta, \theta, \theta} P(x) f(x) + \frac{1}{2} \nabla_{\theta, \theta} f(x) \nabla_{\theta} P(x)
\]

\[
+ \frac{1}{2} \nabla_{\theta, \theta} P(x) \nabla_{\theta} f(x) - \frac{1}{6} \mathcal{R}_x(\theta, \theta)[f(x) \nabla_{\theta} P(x) + \nabla_{\theta} f(x) P(x)])] d\theta.
\]

We prepare some calculations. By Lemma 2.2.5, we have

\[
\mathbb{E}[\mathcal{X}_{B^p_g(t(x))}(X)] = \frac{|S^{d-1}| P(x) e^d + |S^{d-1}| \Delta P(x)}{d(d+2)} \left[ \frac{1}{2} P(x) \omega(x) \right] e^{d+2} + O(e^{d+3}),
\]

and hence

\[
\mathbb{E}[(f(X) - f(x)) \mathcal{X}_{B^p_g(t(x))}(X)] = \mathbb{E}[f(X) \mathcal{X}_{B^p_g(t(x))}(X)] - f(x) \mathbb{E}[\mathcal{X}_{B^p_g(t(x))}(X)]
\]

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

Similarly, by Lemma 2.2.5, we have

\[
\mathbb{E}[(X - t(x)) \mathcal{X}_{B^p_g(t(x))}(X)] = [v_1, v_2],
\]
where
\[
  v_1 = \frac{|S^{d-1}|}{d+2} J_{p,d}^\top t_4 \nabla P(x) \epsilon^{d+2} + \frac{|S^{d-1}|}{24} J_{p,d}^\top t_4 (\Omega_1(x) \nabla P(x) + P(x) \Omega_0(x)) \epsilon^{d+4} \\
  + \frac{|S^{d-1}|}{d+4} J_{p,d}^\top [3_1(x) + \frac{1}{6} \Omega_1(x) \nabla P(x) + \frac{1}{24} P(x) \Omega_2(x)] \epsilon^{d+4} + O(\epsilon^{d+5}),
\]
and
\[
  v_2 = \frac{|S^{d-1}|}{d+2} \frac{2}{P(x)} J_{p,d}^\top P(x) \epsilon^{d+2} + \frac{|S^{d-1}|}{24} \frac{2}{P(x)} J_{p,d}^\top P(x) \Omega_1(x) \epsilon^{d+4} \\
  + \frac{|S^{d-1}|}{(d+2)(d+4)} f_{p,d}^\top \left[ \frac{1}{4} \text{tr}(\Omega_2(x) \nabla^2 P(x)) - \frac{1}{12} f(x) P(x) \Omega_2(x) \right] \epsilon^{d+4} \\
  + \frac{|S^{d-1}|}{6(d+4)} J_{p,d}^\top [\Omega_1(x) \nabla P(x) + \frac{1}{4} P(x) \Omega_2(x)] \epsilon^{d+4} + O(\epsilon^{d+5}).
\]

Again, by Lemma 2.2.5, we have
\[
  \mathbb{E}\left[ (X - t(x))(f(X) - f(x)) \chi_{B_2^p(t(x))}(x) \right] \\
  = \mathbb{E}\left[ (X - t(x)) f(X) \chi_{B_2^p(t(x))}(x) \right] - f(x) \mathbb{E}\left[ (X - t(x)) \chi_{B_2^p(t(x))}(x) \right] \\
  = \left\| v_1, v_2 \right\|,
\]

where
\[
  v_1 = \frac{|S^{d-1}|}{d+2} \frac{2}{P(x)} J_{p,d}^\top P(x) \epsilon^{d+2} + \frac{|S^{d-1}|}{24} \frac{2}{P(x)} J_{p,d}^\top P(x) \Omega_1(x) \epsilon^{d+4} \\
  + \frac{|S^{d-1}|}{d+4} J_{p,d}^\top [3_1(f(x) - f(x)) \nabla f(x) + \frac{1}{6} P(x) \nabla f(x)] \epsilon^{d+4} + O(\epsilon^{d+5})
\]
and
\[
  v_2 = \frac{|S^{d-1}|}{d+2} \frac{2}{P(x)} J_{p,d}^\top P(x) \epsilon^{d+2} + \frac{|S^{d-1}|}{24} \frac{2}{P(x)} J_{p,d}^\top P(x) \Omega_1(x) \epsilon^{d+4} \\
  + \frac{|S^{d-1}|}{d+4} J_{p,d}^\top [\Omega_1(x) \nabla f(x) + \frac{1}{4} P(x) \nabla f(x)] \epsilon^{d+4} + O(\epsilon^{d+5})
\]

With the above preparation, we are ready to prove Theorem 2.5.1.

**Proof of Theorem 2.5.1.** The proof is straightforward, and we show it case by case.

**Case 0 in Condition 2.4.1.** In this case, by Lemma 2.5.1, (2.5.24), and (2.5.23),
\[
  T_{t(x)}^\top \mathbb{E}\left[ X(f(X) - f(x)) \chi_{B_2^p(t(x))}(x) \right] = \frac{|S^{d-1}|}{d+2} \frac{2}{P(x)} \frac{d(d+2)}{|S^{d-1}|} \epsilon^{d+2} + O(\epsilon^{d+4})
\]
and
\[
  T_{t(x)}^\top \mathbb{E}\left[ (X - t(x)) \chi_{B_2^p(t(x))}(x) \right] = \frac{|S^{d-1}|}{d+2} \frac{2}{P(x)} \frac{d(d+2)}{|S^{d-1}|} \epsilon^{d+2} + O(\epsilon^{d+4}),
\]
and hence
\[
\mathbb{E}[(f(X) - f(x))X_{B_p}^{(i)}(X)] - T_{t(x)}^\top \mathbb{E}[(X-t(x))(f(X) - f(x))X_{B_p}^{(i)}(X)] = \frac{|S^{d-1}|}{d(d+2)} \left[ \frac{1}{2} P(x) \Delta f(x) + \nabla f(x) \cdot \nabla P(x) - \frac{P(x) \nabla f(x) \cdot \nabla P(x)}{P(x)} \right] \varepsilon^{d+2} + O(\varepsilon^{d+4}).
\]

Note that \(T_{t(x)}^\top \mathbb{E}[(X-t(x))X_{B_p}^{(i)}(X)]\) is of order \(O(\varepsilon^{d+2})\) for any \(p\), therefore
\[
\mathbb{E}[X_{B_p}^{(i)}(X)] - T_{t(x)}^\top \mathbb{E}[(X-t(x))X_{B_p}^{(i)}(X)] = \frac{|S^{d-1}|}{d} P(x) \varepsilon^{d} + O(\varepsilon^{d+2}).
\]

As a result, we conclude that
\[
Q_f(x) - f(x) = \frac{1}{d+2} \left( \frac{1}{2} \Delta f(x) + \frac{\nabla f(x) \cdot \nabla P(x)}{P(x)} \right) \varepsilon^{d} + O(\varepsilon^{d}).
\]

**Case 1 in Condition 2.4.1.** Observe that by Lemma 2.5.1, the tangential component of \(T_{t(x)}\) is of order \(O(1)\) and the normal component of \(T_{t(x)}\) is of order \(O(\frac{1}{\varepsilon})\). Hence, by (2.5.23)
\[
T_{t(x)}^\top \mathbb{E}[(X-t(x))X_{B_p}^{(i)}(X)] = \frac{|S^{d-1}|}{d} P(x) \sum_{i=d+1}^p \left( \frac{\mathcal{N}_{t(x)}^0 f_p^{(i)} \left( X, f_p - d X_{12} f_p^T e_i \right) \right)^2}{2(d+2)} \frac{2 \lambda_i^{(2)}}{p(x) |S^{d-1}|} \varepsilon^{d} + O(\varepsilon^{d+2}),
\]

and hence
\[
\mathbb{E}[X_{B_p}^{(i)}(X)] - T_{t(x)}^\top \mathbb{E}[(X-t(x))X_{B_p}^{(i)}(X)] = \left[ \frac{|S^{d-1}|}{d} P(x) - \frac{|S^{d-1}|}{d} P(x) \sum_{i=d+1}^p \left( \frac{\mathcal{N}_{t(x)}^0 f_p^{(i)} \left( X, f_p - d X_{12} f_p^T e_i \right) \right)^2}{2(d+2)} \frac{2 \lambda_i^{(2)}}{p(x) |S^{d-1}|} \varepsilon^{d} + O(\varepsilon^{d+2}).
\]

Similarly, by (2.5.24)
\[
T_{t(x)}^\top \mathbb{E}[(X-t(x))(f(X) - f(x))X_{B_p}^{(i)}(X)] = \left[ \frac{|S^{d-1}|}{d} P(x) \left( \frac{\nabla f(x) \cdot \nabla P(x)}{P(x)} \right) + \sum_{i=d+1}^p \frac{\mathcal{N}_{t(x)}^0 f_p^{(i)} \left( X, f_p - d X_{12} f_p^T e_i \right) \right)^2}{2(d+2)} \frac{2 \lambda_i^{(2)}}{p(x) |S^{d-1}|} \varepsilon^{d} + O(\varepsilon^{d+3}),
\]

which could be significantly simplified. Since \(M_{12}^{(2)}\) satisfies (2.3.11), by a direct expansion we have that
\[

\nabla f(x)^\top J_{p,d} M_{12}^{(2)} = \frac{d(d+2)}{P(x)(d+4)} \left( \frac{1}{2} \nabla f(x)^\top \mathcal{M}_2(x) \nabla P(x) + \frac{1}{6} P(x) \mathcal{R}_1(x) \nabla f(x) \right)^\top J_{p,d}.
\]
By (2.3.15), we have $X_1S_{12} = -M_{12}^{(2)}X_2$, and hence
\[
\frac{|S^{d-1}|P(x)}{d(d+2)} \frac{\mathfrak{M}_1(x)J_{p,p-d}X_2J_{p,p-d}^\top e_i}{2\lambda_i^{(2)} + \frac{2(d+2)}{P(x)|S^{d-1}|} \epsilon^{p-d-4}} \nabla f(x) \nabla f(x)^\top J_{p,d}X_1S_{12}J_{p,d}^\top e_i = - \frac{|S^{d-1}|}{d+4} \frac{\mathfrak{M}_1(x)J_{p,p-d}X_2J_{p,p-d}^\top e_i}{2\lambda_i^{(2)} + \frac{2(d+2)}{P(x)|S^{d-1}|} \epsilon^{p-d-4}} \left( \nabla f(x) \nabla f(x) + \frac{1}{6} P(x) \frac{\mathfrak{M}_1(x) \nabla f(x)}{\partial x} \right)^\top J_{p,p-d}X_2J_{p,p-d}^\top e_i.
\]

Combining this with $\Omega_f$ defined in (2.5.20), the second and third terms in $T_{(t(x))}^\top [(X - t(x))(f(X) - f(x))X_B^{(p)}(t(x))](X)$ are simplified. As a result, we have
\[
E[(f(X) - f(x))X_B^{(p)}(t(x))(X)] - T_{(t(x))}^\top E[(X - t(x))(f(X) - f(x))X_B^{(p)}(t(x))](X)] = \left[ \frac{|S^{d-1}|}{d(d+2)} \left( \frac{1}{2} P(x) \Delta f(x) + \nabla f(x) \cdot \nabla P(x) - \frac{P(x) \nabla f(x) \cdot \nabla P(x)}{P(x) + \frac{2(d+2)}{|S^{d-1}|} \epsilon^{p-2}} \right) \right] \left( \sum_{i=d+1}^p \frac{\mathfrak{M}_1(x)J_{p,p-d}X_2J_{p,p-d}^\top e_i}{\epsilon^{p-d-4}} \epsilon^{d+2} + O(\epsilon^{d+3}) \right).
\]

To finish the proof for Case 1, we claim that
\[
\sum_{i=d+1}^p \frac{\mathfrak{M}_1(x)J_{p,p-d}X_2J_{p,p-d}^\top e_i}{\epsilon^{p-d-4}} \epsilon^{d+2} = \sum_{i=d+1}^p \frac{\mathfrak{M}_1(x)e_i\delta_j^i(x)e_i}{\epsilon^{p-d-4}}.
\]

Recall (2.3.21). Suppose there are $q + t$ eigenvalues of $M_{12}^{(2)}$, where $q, t \in \mathbb{N} \cup \{0\}$, so that $q$ eigenvalues are simple. We have
\[
X_2 = \begin{bmatrix} I_{q \times q} & 0 & \cdots & 0 \\ 0 & X_2^1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & X_2^t \end{bmatrix},
\]

where $X_j^i$, where $j = 1, \ldots, t$ are orthogonal matrices whose size is the multiplicity of the associated eigenvalue. Suppose $X_2^1 \in O(\alpha)$, where $\alpha > 1$. Then
\[
\sum_{i=d+q}^{d+q+\alpha} (\mathfrak{M}_1(x)J_{p,p-d}X_2J_{p,p-d}^\top e_i)(\delta_j^i(x)J_{p,p-d}X_2J_{p,p-d}^\top e_i) = \sum_{i=d+q}^{d+q+\alpha} (\mathfrak{M}_1(x)e_i)(\delta_j^i(x)e_i)
\]

since the left hand side is the inner product between the projections of $\mathfrak{M}_1(x)$ and $\delta_j^i(x)$ onto the eigenspace. By a similar argument for the other blocks, we conclude the claim. By exactly the same argument we have
\[
\sum_{i=d+1}^p \frac{\mathfrak{M}_1(x)J_{p,p-d}X_2J_{p,p-d}^\top e_i}{\epsilon^{p-d-4}} = \sum_{i=d+1}^p \frac{\mathfrak{M}_1(x)e_i^2}{\epsilon^{p-d-4}}.
\]

In conclusion, we have
\[
Qf(x) - f(x) = (\mathcal{C}_1(x) + \mathcal{C}_2(x))e^2 + O(\epsilon^3),
\]
where
\[
\mathcal{C}_1(x) = \frac{1}{d(d+2)} \left[ \frac{1}{2} \Delta f(x) + \frac{\nabla f(x) \nabla P(x)}{P(x)} - \frac{\nabla f(x) \nabla P(x)}{P(x) + \frac{d(d+2)}{d+1} \epsilon^{d+2}} \right]
\]
\[
\frac{1}{d} - \frac{1}{2} \sum_{i=d+1}^{n} \frac{(\Omega_i^0 v_i)^2}{\lambda_i^2 + \frac{2(d+2)}{P(x)} \epsilon^{d+4}}
\]
and
\[
\mathcal{C}_2(x) = -\frac{1}{4d+4} \sum_{i=d+1}^{n} \frac{(\Omega_i^0 v_i)(\Omega_i^0 v_i)}{\lambda_i^4 + \frac{2(d+2)}{P(x)} \epsilon^{d+4}}
\]
\[
\frac{1}{d} - \frac{1}{2} \sum_{i=d+1}^{n} \frac{(\Omega_i^0 v_i)^2}{\lambda_i^4 + \frac{2(d+2)}{P(x)} \epsilon^{d+4}}.
\]

**Case 2 in Condition 2.4.1.** In this case, by (2.5.16) and (2.5.17), we rewrite $T_{t(x)}$ as
\[
T_{t(x)} = [v_1, v_2] + [O(\epsilon^2), O(1)],
\]
where
\[
v_1 = \frac{J_{p,d+1}^T \nabla P(x)}{P(x) + \frac{d(d+2)}{d+1} \epsilon^{d+2}} + \frac{1}{\epsilon^2} \sum_{i=d+1}^{n} \frac{\Omega_i^0 v_i}{\lambda_i^2 + \frac{2(d+2)}{P(x)} \epsilon^{d+4}} \epsilon^{d+4} - X_{2,1} S_{12,1} \epsilon^{d+4} v_i
\]
\[
+ \frac{1}{\epsilon^2} \sum_{i=p-l+1}^{n} \frac{\alpha_i}{\lambda_i^4 + \epsilon^{d+6}} X_{2,1} S_{12,1} (J_{p,l})^T v_i
\]
and
\[
v_2 = \frac{1}{\epsilon^2} \sum_{i=d+1}^{n} \frac{\Omega_i^0 v_i}{\lambda_i^2 + \frac{2(d+2)}{P(x)} \epsilon^{d+4}} \epsilon^{d+4} - X_{2,2} \epsilon^{d+4} v_i
\]
\[
+ \frac{1}{\epsilon^2} \sum_{i=p-l+1}^{n} \frac{\alpha_i}{\lambda_i^4 + \epsilon^{d+6}} X_{2,2} \epsilon^{d+4} v_i.
\]

Note that $\frac{\alpha_i}{\lambda_i^4 + \epsilon^{d+4}}$ is of order 1 or smaller, no matter what regularization order $\rho$ is chosen. Rewrite (2.5.23) up to $O(\epsilon^{d+4})$ as
\[
E[(X - t(x)) X_{B_{G_{\rho}}(t(x))}^T(X)] = \left[ \frac{d+1}{d+2} J_{p,d+1}^T \nabla P(x) \epsilon^{d+2} + O(\epsilon^{d+4}) \right],
\]
\[
\left[ \frac{d+1}{d} \Omega_i v_i^T \epsilon^{d+4} \right] + O(\epsilon^{d+4}).
\]

We claim that in $E[(X - t(x)) X_{B_{G_{\rho}}(t(x))}], T_{t(x)}$, the “fourth order” terms, i.e., the terms with $\sum_{i=p-l+1}^{n} \frac{\alpha_i}{\lambda_i^4 + \epsilon^{d+4}}$, do not have dominant contribution asymptotically by showing that for each $i = p-l+1, \ldots, p$, we have
\[
E[(X - t(x)) X_{B_{G_{\rho}}(t(x))}^T(X)] = O(\epsilon^{d+1})
\]
(2.5.25)
Based on Lemma 2.5.1 and (2.3.23), the first part of (2.5.27) becomes
\[
\mathbb{E}[\{(f(X) - f(x))(X - t(x))X^{B_\beta}_{(t(x))}(X)\}^\top \left[ X_1 S_{12,2} J_{p,l}^T e_i, \frac{1}{\varepsilon^2} X_{2,2} J_{p,l}^T e_i \right] = O(\varepsilon^{d+3}). \tag{2.5.26}
\]
Since the tangential direction of $T_{t(x)}$ is of order 1 and the normal direction of $T_{t(x)}$ is of order $\varepsilon^{-2}$, it is sufficient to focus on the normal direction in order to show (2.5.25). By Proposition 3.2.3, the dominant term in the normal direction satisfies
\[
\mathfrak{R}_0^T (X) J_{p,l} X_{2,2} J_{p,l}^T e_i = 0,
\]
and hence (2.5.25) follows.

To show (2.5.26), for each $p - l + 1 \leq i \leq p$, by a direct expansion we have
\[
\mathbb{E}[\{(f(X) - f(x))(X - t(x))X^{B_\beta}_{(t(x))}(X)\}^\top \left[ X_1 S_{12,2} J_{p,l}^T e_i, \frac{1}{\varepsilon^2} X_{2,2} J_{p,l}^T e_i \right]
= \left( \frac{|S^{d-1}|}{d(d+2)} P(x)_t \nabla f(x)^\top J_{p,d} X_1 S_{12,2} J_{p,l}^T e_i + \frac{|S^{d-1}|}{d+4} \Omega^1_{J_{p,l}} X_{2,2} J_{p,l}^T e_i \right) \varepsilon^{d+2} + O(\varepsilon^{d+3}).
\]
Again, it is sufficient to focus on the normal direction. We now claim that
\[
\frac{|S^{d-1}|}{d(d+2)} P(x)_t \nabla f(x)^\top J_{p,d} X_1 S_{12,2} J_{p,l}^T e_i + \frac{|S^{d-1}|}{d+4} \Omega^1_{J_{p,l}} X_{2,2} J_{p,l}^T e_i = 0. \tag{2.5.27}
\]
Based on Lemma 2.5.1 and (2.3.23), the first part of (2.5.27) becomes
\[
\frac{|S^{d-1}|}{d(d+2)} P(x)_t \nabla f(x)^\top J_{p,d} X_1 S_{12,2} J_{p,l}^T e_i
= - \frac{|S^{d-1}|}{d(d+2)} P(x)_t \nabla f(x)^\top J_{p,d} M_{12}^{(2)} X_{2,2} J_{p,l}^T e_i
= - \frac{|S^{d-1}|}{d(d+2)} P(x)_t \nabla f(x)^\top J_{p,d} M_{12}^{(2)} J_{p,p-d} J_{p,l} X_{2,2} J_{p,l}^T e_i
= - \frac{|S^{d-1}|}{d+4} \left( \frac{1}{2} \nabla f(x)^\top \mathfrak{R}_2(x) \nabla P(x) + \frac{1}{6} P(x) \mathfrak{R}_1(x) \nabla f(x) \right)^\top J_{p,l} X_{2,2} J_{p,l}^T e_i,
\]
where the second equality comes from the direct expansion that
\[
M_{12}^{(2)} = M_{12}^{(2)} J_{p,p-d} J_{p,l}
\]
and the last equality comes from (2.3.11). For the second part of (2.5.27), based on Proposition 3.2.3, for $p - l + 1 \leq i \leq p$, we have
\[
\frac{|S^{d-1}|}{d+4} \Omega^1_{J_{p,l}} X_{2,2} J_{p,l}^T e_i
= \frac{|S^{d-1}|}{d+4} \left( \frac{1}{2} \nabla f(x)^\top \mathfrak{R}_2(x) \nabla P(x) + \frac{1}{6} P(x) \mathfrak{R}_1(x) \nabla f(x) \right)^\top J_{p,l} X_{2,2} J_{p,l}^T e_i.
\]
Thus, two terms in (2.5.27) cancel each other and (2.5.26) follows. Based on the above discussion, we know that
\[ \mathbb{E}[(X - t(x))X_{B}^{\rho}(t(x))(X)]^{T}T_{t(x)} \] is dominated by
\[
\mathbb{E}[(X - t(x))X_{B}^{\rho}(t(x))(X)]^{T} \left[ \frac{J_{p,d+4}^{\top} \nabla P(x)}{P(x) + \frac{d(d+2)}{8} e^{-2}} + \sum_{i=d+1}^{p-l} \frac{\gamma_{l}^{\top}(x)JX_{2,1}J_{i}^{\top} e_{i}}{\frac{d}{2} \lambda_{i}^{(2)} + \frac{2(d+2)}{P(x) |S_{0}|} e^{-4}}X_{2,1}S_{1,2}J_{i}^{\top} e_{i}, \right.
\]
\[
\left. \frac{1}{\epsilon^{2}} \sum_{i=d+1}^{p-l} \frac{\gamma_{l}^{\top}(x)JX_{2,1}J_{i}^{\top} e_{i}}{\frac{d}{2} \lambda_{i}^{(2)} + \frac{2(d+2)}{P(x) |S_{0}|} e^{-4}} \right].
\]

which is of order \(O(\epsilon^{d})\) by a similar argument as in Case 1, and \[ \mathbb{E}[(f(X) - f(x))(X - t(x))X_{B}^{\rho}(t(x))(X)]^{T}T_{t(x)} \] is dominated by
\[
\mathbb{E}[(f(X) - f(x))(X - t(x))X_{B}^{\rho}(t(x))(X)]^{T} \left[ \frac{J_{p,d+4}^{\top} \nabla P(x)}{P(x) + \frac{d(d+2)}{8} e^{-2}} + \sum_{i=d+1}^{p-l} \frac{\gamma_{l}^{\top}(x)JX_{2,1}J_{i}^{\top} e_{i}}{\frac{d}{2} \lambda_{i}^{(2)} + \frac{2(d+2)}{P(x) |S_{0}|} e^{-4}}X_{2,1}S_{1,2}J_{i}^{\top} e_{i}, \right.
\]
\[
\left. \frac{1}{\epsilon^{2}} \sum_{i=d+1}^{p-l} \frac{\gamma_{l}^{\top}(x)JX_{2,1}J_{i}^{\top} e_{i}}{\frac{d}{2} \lambda_{i}^{(2)} + \frac{2(d+2)}{P(x) |S_{0}|} e^{-4}} \right].
\]

which is of order \(O(\epsilon^{d+2})\) by using a similar argument in Case 1. By putting the above together, we conclude that

\[ Qf(x) - f(x) = (\mathcal{C}_{1}(x) + \mathcal{C}_{2}(x)) \epsilon^{2} + O(\epsilon^{3}), \]

where

\[ \mathcal{C}_{1}(x) = \frac{1}{d(d+2)} \left[ \frac{1}{2} \Delta f(x) + \frac{\nabla f(x) \nabla P(x)}{P(x) + \frac{d(d+2)}{8} e^{-2}} - \frac{\nabla f(x) \nabla P(x)}{P(x) + \frac{d(d+2)}{8} e^{-2}} \right], \]

\[ \mathcal{C}_{2}(x) = -\frac{1}{d(d+2)} \sum_{i=d+1}^{p-l} \frac{\gamma_{l}^{\top}(x) e_{i} e_{i}}{\frac{d}{2} \lambda_{i}^{(2)} + \frac{2(d+2)}{P(x) |S_{0}|} e^{-4}} \left( \gamma_{l}^{\top}(x) e_{i} e_{i} \right), \]

and hence we finish the proof.

### 2.6 Conclution of the chapter: convergence of LLE on closed manifolds

By combining the variation analysis and the bias analysis shown above, we conclude the following pointwise convergence theorem for LLE, when we have a proper choice of \(\rho\).

**Theorem 2.6.1.** Take \(f \in C(\mathcal{C}(M))\), \(\rho = 3\), and \(\epsilon = \epsilon(n)\) so that \(\frac{\sqrt{\log(n)}}{n^{1/2} \log^{2/3} n} \to 0\) and \(\epsilon \to 0\) as \(n \to \infty\). With probability greater than \(1 - n^{-2}\), for all \(x_{k} \in \mathcal{X}\),

\[ \frac{1}{\epsilon^{2}} \left[ \sum_{j=1}^{N} w_{k}(j) f(x_{k,j}) - f(x_{k}) \right] = \frac{1}{2(d+2)} \Delta f(x) + O(\epsilon^{2}) + O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \log^{2/3} n} \right). \]

Based on the Borel-Cantelli Lemma, it is clear that asymptotically LLE converges almost surely. For practical
purposes, we need to discuss the bandwidth choice when \( \rho = 3 \). Based on the assumption about the relationship between \( n \) and \( \varepsilon \), we have \( \frac{\sqrt{\log(n)}}{n^{1/4}\varepsilon^{d/2+1}} \to 0 \) as \( n \to \infty \), but the convergence rate of \( \frac{\sqrt{\log(n)}}{n^{1/4}\varepsilon^{d/2+1}} \) might be slower than \( \varepsilon \to 0 \). Suppose we call a bandwidth “optimal”, if it balances the standard deviation and the bias for all cases in Condition 2.4.1; that is, \( \frac{\sqrt{\log(n)}}{n^{1/4}\varepsilon^{d/2+1}} \approx \varepsilon \). We then have \( n \frac{\log(n)}{\varepsilon^d} \approx \frac{1}{\varepsilon^{d+1}} \), and we can estimate the optimal bandwidth from \( n \).

### 2.7 \( \varepsilon \)-radius neighborhood v.s. \( K \) nearest neighborhood

In the original article [49], the KNN scheme was proposed for LLE algorithm. However, the analysis in this paper has been based on the \( \varepsilon \)-radius neighborhood scheme. These two schemes are closely related asymptotically from the viewpoint of density function estimation [43]. The following argument shows that the developed theorems are actually transferrable to the KNN scheme under the manifold setup.

For \( t(x_k) \in \mathcal{M} \), take \( K \) nearest neighbors of \( t(x_k) \), namely \( t(x_{k_1}), \ldots, t(x_{k_K}) \), with respect to the Euclidean distance. Intuitively, \( K \) is closely related to the volume of the minimal ball centered at \( x_k \) with the radius \( \varepsilon(x_k) \) containing the \( K \) nearest neighbors of \( x_k \), where \( \varepsilon(x) \) depends on \( K \) and the p.d.f.; that is, we expect to have

\[
n P(x_k) \text{vol}(D_{x_k}) \approx K,
\]

where \( D_x := B_{\varepsilon(x)}^\rho(t(x) \cap t(M)) \) is the minimal ball centered at \( x \in M \) with the radius \( \varepsilon(x) > 0 \) so that \( D_x \) contains the \( K \) nearest neighbors of \( x \). Under the smoothness assumption of the p.d.f. and the manifold setup, we claim that asymptotically when \( n \to \infty \), this relationship holds uniformly over the manifold a.s., if \( K = K(n) \), \( K/\log(n) \to \infty \) and \( K/n \to 0 \) as \( n \to \infty \). This claim could be achieved by slightly modifying the argument for the Theorem in [20] to obtain the large deviation bound for (2.7.1) when \( n \) is finite. To bound \( \Pr\{\sup_{x \in M} \left| \frac{K}{\text{vol}(D_x)} - P(x) \right| > \alpha \} \), where \( \alpha > 0 \), it is sufficient to bound the two terms on the right hand side of [20, equation (10)]. By a straightforward calculation of the equations on page 539 in [20], we achieve the bound \( \Pr\{\sup_{x \in M} \left| \frac{K}{\text{vol}(D_x)} - P(x) \right| > \alpha \} \leq \text{poly}(n) e^{-cK\alpha^2} \), where \( c \) is a constant depending on \( d \) and the upper bounds of \( P(x) \) on \( M \), and \( \text{poly}(n) = 3(1 + 2^{n+3}n^{n+3}) \).\(^2\) Therefore, if we choose \( \alpha = \left(\frac{2^p+10^p}{c}\right)^{1/3} \left(\frac{\log n}{n}\right)^{1/3} \), with probability greater than \( 1 - n^{-2} \), we have uniformly \( \frac{K}{\text{vol}(D_x)} = P(x) + O(\alpha) \). Note that by the assumption, \( \alpha \to 0 \) as \( n \to \infty \). We conclude that with probability greater than \( 1 - n^{-2} \),

\[
\varepsilon(x) = \left(\frac{d}{|8^{d-1}|}\right)^{1/d} \left(\frac{K}{nP(x)}\right)^{1/d} \left(1 + O\left(\left(\frac{\log n}{K}\right)^{1/3}\right)\right),
\]

where we use the fact that \( \text{vol}(D_x) = \frac{(\varepsilon(x)^d + O(\varepsilon(x)^{d+1}))}{d} \) when \( \varepsilon(x) \) is sufficiently small. It is transparent that \( \varepsilon(x) \) depends on \( n \) and \( \varepsilon(x) \) \( \to 0 \) a.s. as \( n \to \infty \) since \( K(n)/n \to 0 \) by assumption. In other words, \( \varepsilon \) is not a constant value. It is a function depending on the p.d.f.. If we require \( K = K(n) \) to additionally satisfy \( \frac{K(n)}{n^{3/2} \log(n)^{d/2}} \to \infty \), then \( \varepsilon(x_k) \) satisfies \( \frac{\sqrt{n}}{n^{1/4}\varepsilon^{d/2+1}} \to 0 \) a.s.. On the other hand, notice that the statement of Theorem 2.6.1 is pointwise. Therefore, its proof could be directly employed to the case when \( \varepsilon \) is chosen pointwisely, and hence the KNN scheme. As a result, if we take \( \rho = 3 \) and is \( K/n \to 0 \), \( K/\log(n) \to \infty \), and \( (K/n)(K/\log(n))^{d/2} \to \infty \) when

\(^2\)This can be observed by combining [20, equations (6) (7) (9) and (10)]. The second term on the right hand side of [20, equation (10)] is dominated by the first term. To bound the first term, we can substitute \( \delta = \frac{K}{n^{3/2} \log(n)^{d/2}} < M = \frac{\beta n^{3/2}}{\log(n)^{d/2}} \) into the fourth unlabeled equation on page 539 in [20], where \( P_k \) is the upper bound of p.d.f. In the fourth unlabeled equation, \( \alpha \) is the upper bound of the volume ratio of \( B_{\varepsilon(x)}^\rho(t(x) \cap t(M)) \) and \( B_{\varepsilon(x)}^\rho(t(x) \cap t(M)) \), which can be chosen as \( 3^d \) when \( \varepsilon(x) \) is sufficiently small. Finally, we use the fact that when \( \beta \) is small, the equation follows.
n \rightarrow \infty$, by plugging (2.7.2) into Theorem 2.6.1, when $n$ is sufficiently large, the following convergence holds for all $x_k$ with probability greater than $1 - 2n^{-2}$:

$$
\sum_{j=1}^{K} w_k(j) f(x_k,j) - f(x_k) = \left( \frac{d}{(d+1)^{1/d}} \right) ^{1/d} \frac{\Delta f(x_k)}{P(x_k)^{2/d}} \left( \frac{K}{n} \right) ^{2/d} + O \left( \left( \frac{\log(n)}{K} \right) ^{1/3} \left( \frac{K}{n} \right) ^{2/d} \right) + O \left( \left( \frac{\log(n)}{K} \right) ^{1/2} \left( \frac{K}{n} \right) ^{1/d} \right).
$$

(2.7.3)

In summary, unless the sampling is uniform, we do not obtain the Laplace-Beltrami operator with the KNN scheme. Based on the expansion (2.7.3), to obtain the Laplace-Beltrami operator with the KNN scheme, we could numerically consider a "normalized LLE matrix"; that is, find the eigen-structure of $\tilde{L}$ := $\delta^{-1}(W - I)$, where $W$ is the ordinary LLE matrix, and $\delta$ $\in \mathbb{R}^{n \times n}$ is a diagonal matrix so that $\delta_{ii} = \varepsilon(x_i)^2$. Since the analysis of the pointwise convergence of $\tilde{L}$ is similar to that of Theorem 2.6.1, we skip the details here.

### 2.8 LLE v.s. LLR

Based on the above theoretical study under the manifold setup, we could link LLE to the locally linear regression (LLR) [25, 17]. Recall that in the LLR, we locally fit a linear function to the response, and the associated kernel depends on the inverse of a variation of the covariance matrix. We summarize how the LLR is operated. Consider the following regression model

$$
Y = m(X) + \sigma(X) \xi,
$$

(2.8.1)

where $\xi$ is a random error independent of $X$ with $\mathbb{E}(\xi) = 0$ and $\text{Var}(\xi) = 1$, and both the regression function $m$ and the conditional variance function $\sigma^2$ are defined on $\mathbb{R}^d$. Let $\{(X_i, Y_i)\}_{i=1}^n$ denote a random sample observed from model (2.8.1) with $\mathcal{X} := \{X_i\}_{i=1}^n$ being sampled from $X$. Given $\{(X_i, Y_i)\}_{i=1}^n$ and $x \in \mathbb{R}^d$, the problem is then to estimate $m(x)$ assuming enough smoothness of $m$. Choose a smooth kernel function with fast decay $K : [0, \infty) \rightarrow \mathbb{R}$ and a bandwidth $\varepsilon > 0$. The LLR estimator for $m(x)$ is defined as $\hat{e}_i^\top \hat{\beta}_x$, where

$$
\hat{\beta}_x = \arg \min_{\beta \in \mathbb{R}^{d+1}} (Y - X_i \beta)^\top W_x (Y - X_i \beta),
$$

(2.8.2)

$$
Y = (Y_1, \ldots, Y_n)^\top, \quad X_i = \left[ \begin{array}{ccc} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{array} \right] ^\top \in \mathbb{R}^{n \times (d+1)},
$$

$$
W_x = \text{diag}(K_\varepsilon(X_1, x), \ldots, K_\varepsilon(X_n, x)) \in \mathbb{R}^{n \times n},
$$

and $K_\varepsilon(X_i, x) := \varepsilon^{-d}K(||X_i - x||_{\mathbb{R}^d}/\varepsilon)$. By a direct expansion, (2.8.2) becomes

$$
\hat{\beta}_x = (X_i^\top W_x X_i)^{-1} X_i^\top W_x Y
$$

(2.8.3)

if $(X_i^\top W_x X_i)^{-1}$ exists. We have $X_x = \left[ \begin{array}{c} 1_n \\ G_x \end{array} \right]$, where $G_x$ is the data matrix associated with $\{X_i\}_{i=1}^n$ centered at $x$. By yet another direct expansion by the block inversion,

$$
e_i^\top \hat{\beta}_x = w_x^{(LLR)}^\top Y,
$$

(2.8.4)
where \(w^{(\text{LLR})}_x\) is called the “smoothing kernel” and satisfies

\[
w^{(\text{LLR})}_x := \frac{1_n^\top W_x - 1_n^\top W_x G_x^\top (G_x W_x G_x^\top)^{-1} G_x W_x}{1_n^\top W_x 1_n - 1_n^\top W_x G_x^\top (G_x W_x G_x^\top)^{-1} G_x W_x 1_n}.
\] (2.8.5)

Through a direct comparison, we see that the vector \(w^{(\text{LLR})}_x\) is almost the same as the weight matrix in LLE algorithm shown in (1.3.11), except the weighting by the chosen kernel – in LLE, the kernel function and its support are both determined by the data, while in the LLR the kernel is selected in the beginning and the data points are weighted by the chosen kernel like \(G_x W_x\). If we choose the kernel to be a zero-one kernel with the support on the ball centered at \(x\) with the radius \(\epsilon\), then we “recover” (1.3.11).

Under the low dimensional manifold setup, \(G_x W_x G_x^\top\) might not be of full rank. Note that the term \(G_x W_x G_x^\top\) is the weighted local covariance matrix, which is considered in [55] to estimate the tangent space. Unlike the regularized pseudo-inverse (1.3.9) in LLE, to handle this degeneracy issue, in LLR the data matrix \(G_x\) is constructed by projecting the point cloud to the estimated tangent plane. This projection step could be understood as taking the Moore-Penrose pseudo-inverse approach to handle the degeneracy. We mention that in [17, Section 6], the relationship between the LLR and the manifold learning under the manifold setup is established. It is shown that asymptotically, the smooth matrix from the kernel \(w^{(\text{LLR})}_x\) leads to the Laplace-Beltrami operator. The result is parallel to the reported result in this paper.

These relationships between LLE and the LLR suggest the possibility of fitting the data locally by taking the locally polynomial regression into account, and generalizing the barycentric coordinates by fitting a polynomial function locally. This might lead to a variation of LLE that catches more delicate structure of the manifold, in a different adaptive way. Since this direction is outside the scope of this paper, the study of this possibility is left to future studies.

### 2.9 Error in variable

In this work, we analyze LLE under the assumption that the dataset is randomly sampled directly from a manifold, without any influence of the noise. However, the noise is inevitable and a further study is needed. By the analysis, we observe that LLE takes care of the error in variable challenge “in some sense”.

Suppose the dataset is \(\{y_i\}_{i=1}^n \subseteq \mathbb{R}^p\), where \(y_i = z_i + \xi_i, z_i\) is supported on a manifold and \(\xi_i\) is an i.i.d. noise with good properties. The question is to ask how much information LLE could recover from \(\{z_i\}_{i=1}^n\). A parallel problem for the GL, or the more general graph connection Laplacian (GCL), has been studied in [23, 24]. It shows that the spectral properties of the GL and GCL are robust to noise. For LLE, while a similar analysis could be applied, if we view LLE as a kernel method and show a similar result, we mention that we might benefit by taking the special algorithmic structure of LLE into account.

When the dimension of the dataset is high, the noise might have a nontrivial behavior. For example, when the dimension of the database \(p = p(n)\) satisfies \(p(n)/n \to \gamma > 0\) when \(n \to \infty\) (known as the large \(p\) and large \(n\) setup), it is problematic to even estimate the covariance matrix. Note that the covariance matrix is directly related to LLE algorithm

since the covariance matrix appears in the regularized pseudo inverse, \(\mathcal{F}_{(\gamma \sigma \epsilon p)} (\tilde{G}_n \tilde{G}_n^\top)\), where \(\tilde{G}_n\) is the local data matrix associated with \(y_k\) determined from the noisy database \(\{y_i\}_{i=1}^n\), and \(\tilde{G}_n \tilde{G}_n^\top\) is the covariance matrix.

Under the large \(p\) and large \(n\) setup, the eigenvalues and eigenvectors of the covariance matrix will both be biased, depending on the “signal-to-noise ratio” and \(\gamma\) [34]. A careful manipulation of the noise, or a modification of the covariance matrix estimator, is needed in order to address these introduced biases. For example, the
“shrinkage technique” was introduced to correct the eigenvalue bias with a theoretical guarantee [56, 21]. The covariance matrix estimator based on the shrinkage technique is \( \tilde{C}_n := \sum_{i=1}^{p} f(\lambda_i) u_i u_i^\top \), where \( u_i \) and \( \lambda_i \) form the \( i \)-th eigenpair of \( \tilde{G}_n \tilde{G}_n^\top \) and \( f \) is the designed shrinkage function.

A direct comparison shows that the regularized pseudo inverse in LLE behaves like a shrinkage technique. Recall that \( \mathcal{S}_{n, \epsilon} \left( \tilde{G}_n \tilde{G}_n^\top \right) = \sum_{i=1}^{p} \frac{1}{\lambda_i + n \epsilon} u_i u_i^\top \) \( (1.3.9) \), where \( r_n \) is the rank of \( \tilde{G}_n \tilde{G}_n^\top \), the shrinkage function is \( f(x) = \frac{1}{x + n \epsilon} \chi_\{x \geq 0\}(x) \), and \( \chi \) is the indicator function. Although how \( f \) corrects the noise impact is outside the scope of this paper, it would be potential to carefully improve the regularized pseudo inverse by taking the shrinkage technique into account. In other words, by modifying the barycentric coordinate evaluation and applying the technique discussed in [23, 24], it is possible to improve LLE algorithm.

### 2.10 Numerical examples

#### 2.10.1 Sphere

Suppose that \( S^{p-1} \in \mathbb{R}^p \) is the unit sphere in \( \mathbb{R}^p \). Denote \( H_k \) to be the space of homogeneous polynomials in \( \mathbb{R}^p \) restricted on \( S^{p-1} \). We have that the space \( H_k \) is the eigenspace of the Laplace-Beltrami operator on \( S^{p-1} \) corresponding to eigenvalue \(-k(p+2)\), and the dimension of \( H_k \) is \( \binom{p+k+1}{p+1} - \binom{p+k}{p+1} [60] \). In this example, we show that if we choose a \( \epsilon^{d+p} \) that is too small, then we are not going to get the Laplace-Beltrami operator. When \( p = 8 \), which is much greater than 3, by Theorem 2.5.1, we have

\[
Q f(x_k) - f(x_k) = \left( \frac{-(p-1)}{8(p+3)(p+5)} \sum_{i=1}^{p-1} \partial_i^4 f(x_k) - \frac{(p-1)}{24(p+3)(p+5)} \sum_{i \neq j} \partial_i^2 \partial_j^2 f(x_k) \right. \\
\left. - \frac{p+1}{24(p+3)(p+5)} \sum_{i=1}^{p-1} \partial_i^2 f(x_k) \right) \epsilon^4 + O(\epsilon^6). \tag{2.10.1}
\]

It is obvious that asymptotically, we get the fourth order differential operator, instead of the Laplace-Beltrami operator. Specifically, when \( p = 2 \), or \( S^1 \),

\[
Q f(x_k) - f(x_k) = -\frac{1}{280} (f'''(x_k) + f''(x_k)) \epsilon^4 + O(\epsilon^6). \tag{2.10.2}
\]

We mention that if the data set \{\( x_i \)\}_{i=1}^{n} is non-uniformly sampled based on the p.d.f. \( P \) from \( S^1 \), then for any \( x_k \) we have \( Q f(x_k) - f(x_k) = C \epsilon^4 + O(\epsilon^6) \), where \( C \) depends on the first four order differentiation of \( f \) at \( x_k \) and the first three order differentiations of \( P \) at \( x_k \).

**Calculation of the sphere case.** The calculation flow could serve as a simplified proof of Theorem 2.5.1 under the special manifold setup, so we provide the details here. Consider the unit sphere \( S^{p-1} \subset \mathbb{R}^p \). We assume that the center of the sphere is at \( [0, \cdots, 0, 1] \) the data set \{\( x_i \)\}_{i=1}^{n} is uniformly sampled from \( S^{p-1} \), and \( x_k \) is at the origin. To simplify the calculation, for \( v \in \mathbb{R}^p \), denote \( v_1 \in \mathbb{R}^{p-1} \) to be the first \( p-1 \) coordinates of \( v \) and \( v_2 \in \mathbb{R} \) to be the last coordinate of \( v \), and use the notation \( v = [v_1, v_2] \). We parametrize \( S^{p-1} \setminus [0, \cdots, 0, 2] \) by the normal coordinates at \( x_k \) via

\[
\theta t \to [\theta \sin(t), 1 - \cos(t)] \in S^{p-1} \setminus [0, \cdots, 0, 2],
\]
where \( \theta \in S^{p-2} \subset T_{x_k}S^{p-1} \approx \mathbb{R}^{p-1} \) and \( r \in [0, \pi) \) is the geodesic distance. The volume form is

\[
dV = \sin^{p-2}(t)dt d\theta.
\]

Denote \( r := r(\varepsilon) \) to be the radius of the ball \( \exp_{x_k}^{-1}(B^p_\varepsilon(x_k) \cap S^{p-1}) \) in \( T_{x_k}S^{p-1} \), where \( \varepsilon \) is assumed to be sufficiently small. By a direct calculation, we have

\[
E[XX^T X_{B^p_\varepsilon}(x_k)] = \frac{\int_{S^{p-2}} \int_0^r \theta \sin^2(t) \sin^{p-2}(t) dt d\theta}{\int_{S^{p-2}} \int_0^r \theta (\sin(t) - \sin(t) \cos(t)) \sin^{p-2}(t) dt d\theta}
\]

which is a diagonal matrix containing the eigenvalues of \( C_{x_k} \), and we can choose \( \{e_i\}_{i=1}^p \) to be its orthonormal eigenvectors. Next, we have

\[
E[X X^T X_{B^p_\varepsilon}(x_k)] = \int_{S^{p-2}} \int_0^r \theta \sin^{p-1}(t) dt d\theta, \int_{S^{p-2}} \int_0^r (1 - \cos(t)) \sin^{p-2}(t) dt d\theta
\]

\[
= [0, \, |S^{p-2}| \int_0^r (1 - \cos(t)) \sin^{p-2}(t) dt]
\]

We now choose \( \rho = 8 \). Therefore, by definition,

\[
T_{x_k} = \mathcal{F}_{\varepsilon^{p+5}}(C_{x_k}) = \mathcal{F}_{\varepsilon^{p+5}}(C_{x_k}) = \frac{\int_0^r (1 - \cos(t)) \sin^{p-2}(t) dt}{\int_0^r (1 - \cos(t))^2 \sin^{p-2}(t) dt + \varepsilon^{p+7}}
\]

where the last equality holds since \( r = r(\varepsilon) = \varepsilon + O(\varepsilon^3) \) and hence \( \varepsilon^{p+7} = r^{p+7} + O(r^{p+9}) \). Thus, the kernel centered at \( x_k = 0 \) and evaluated at \( y = \[\theta \sin(t), 1 - \cos(t)\] \in \mathbb{R}^p \) satisfies

\[
K_{\text{LLE}}(x_k, y) = 1 - [\theta \sin(t), 1 - \cos(t)] \cdot T_{x_k}
\]

\[
= 1 - (1 - \cos(t)) \frac{\int_0^r (1 - \cos(t)) \sin^{p-2}(t) dt}{\int_0^r (1 - \cos(t))^2 \sin^{p-2}(t) dt + \varepsilon^{p+7} + O(r^{p+9})}
\]

Suppose \( f \in C^5(S^{p-1}) \), we are going to calculate \( \frac{K_{\text{LLE}}(x_k, y)(y) dV(y)}{K_{\text{LLE}}(x_k, y) dV(y)} - f(x_k) \). The evaluation of \( \int K_{\text{LLE}}(x_k, y) dV(y) \)
is direct, and we have
\[
\int K_{LLE}(x_k, y) dV(y)
= \int_{S^{p-2}} \int_0^r \left( 1 - (1 - \cos(t)) \left( \frac{2(p+3)}{(p+1)r^2} + \left( \frac{p^2 + 14p - 3}{6(p+1)(p+3)} + O(r^2) \right) \right) \right) \sin^{p-2}(t) dt d\theta \\
= \left( \frac{4 |S^{p-2}|}{(p+1)(p^2-1)} \right) r^{p-1} + O(r^{p+1}).
\]

On the other hand, we have
\[
\int K_{LLE}(x_k, y)(f(y) - f(x_k)) dV(y)
= \int_{S^{p-2}} \int_0^r \left( \nabla_\theta f(x_k) t + \frac{1}{2} \nabla_\theta^2 f(x_k) t^2 + \frac{1}{6} \nabla_\theta^3 f(x_k) t^3 + \frac{1}{24} \nabla_\theta^4 f(x_k) t^4 + O(r^5) \right) \sin^{p-2}(t) dt d\theta \\
\times \left( 1 - (1 - \cos(t)) \left( \frac{2(p+3)}{(p+1)r^2} + \left( \frac{p^2 + 14p - 3}{6(p+1)(p+3)} + O(r^2) \right) \right) \right) \sin^{p-2}(t) dt d\theta.
\]

We calculate each part in the above integration by using the symmetry of sphere $S^{p-2}$ in the tangent space. Specifically, we have
\[
\int_{S^{p-2}} \int_0^r \left( \nabla_\theta f(x_k) t + \frac{1}{2} \nabla_\theta^2 f(x_k) t^2 + \frac{1}{6} \nabla_\theta^3 f(x_k) t^3 + \frac{1}{24} \nabla_\theta^4 f(x_k) t^4 + O(r^5) \right) \sin^{p-2}(t) dt d\theta \\
= \int_{S^{p-2}} \frac{\nabla_\theta^4 f(x_k) d\theta}{2(p+1)} r^{p+1} + \left( \frac{\int_{S^{p-2}} \nabla_\theta^2 f(x_k) d\theta}{24(p+3)} - \frac{(p-2) \int_{S^{p-2}} \nabla_\theta^3 f(x_k) d\theta}{24(p+3)} \right) r^{p+3} + O(r^{p+5})
\]
and
\[
\int_{S^{p-2}} \int_0^r \left( \nabla_\theta f(x_k) t + \frac{1}{2} \nabla_\theta^2 f(x_k) t^2 + \frac{1}{6} \nabla_\theta^3 f(x_k) t^3 + \frac{1}{24} \nabla_\theta^4 f(x_k) t^4 + O(r^5) \right) (1 - \cos(t)) \sin^{p-2}(t) dt d\theta \\
= \int_{S^{p-2}} \frac{\nabla_\theta^4 f(x_k) d\theta}{4(p+1)} r^{p+3} + \frac{\int_{S^{p-2}} \nabla_\theta^3 f(x_k) d\theta}{48(p+5)} r^{p+5} + O(r^{p+7}).
\]

Due to $\frac{2(p+3)}{(p+1)^2}$, the term of order $r^{p+1}$ is cancelled and we obtain
\[
\int K_{LLE}(x_k, y)(f(y) - f(x_k)) dV(y)
= \frac{-1}{6(p+1)(p+3)(p+5)} \left( \int_{S^{p-2}} \nabla_\theta^4 f(x_k) d\theta + \int_{S^{p-2}} \nabla_\theta^3 f(x_k) d\theta \right) r^{p+3} + O(r^{p+5}).
\]

We use the fact that $r = r(\varepsilon) = \varepsilon + O(\varepsilon^3)$ and summarize the result as follows:
\[
\frac{\int K_{LLE}(x_k, y)f(y) dV(y)}{\int K_{LLE}(x_k, y) dV(y)} - f(x_k) = \frac{\int K_{LLE}(x_k, y)(f(y) - f(x_k)) dV(y)}{\int K_{LLE}(x_k, y) dV(y)} \\
= \frac{-1}{24 |S^{p-2}| (p+1)(p+3)(p+5)} \left( \int_{S^{p-2}} \nabla_\theta^4 f(x_k) d\theta + \int_{S^{p-2}} \nabla_\theta^3 f(x_k) d\theta \right) \varepsilon^4 + O(\varepsilon^6).
\]
Finally, if we use formulas

\[
\begin{align*}
\int_{S^{p-2}} x_i^4 d\theta &= \frac{3}{p+1}|S^{p-2}| \\
\int_{S^{p-2}} x_i^2 x_j^2 d\theta &= \frac{1}{p+1}|S^{d-2}| \quad i \neq j \\
\int_{S^{p-2}} x_i x_j x_k d\theta &= 0 \quad i \neq j,
\end{align*}
\]  

(2.10.3)

we obtain the expansion (2.10.1).

We now numerically show the relationship between the non-uniform sampling scheme and the regularization term. Fix \( n = 30,000 \). Take non-uniform sampling points \( \theta_i := 2\pi U_i + 0.3 \sin(2\pi i/n) \) on \((0, 2\pi]\), where \( i = 1, \ldots, n \) and \( U_i \) is the uniform distribution on \([0, 1]\), and construct \( \mathcal{X}_2 = \{ (\cos(\theta_i), \sin(\theta_i)) \}^{n}_{i=1} \subset \mathbb{R}^2 \). Run LLE with \( \varepsilon = 0.0002 \) and different \( \rho \)'s, and evaluate the first 400 eigenvalues. Based on the theory, we know that when \( \rho < 3 \), the asymptotic depends on the non-uniform density function; when \( \rho = 3 \), we recover the Laplace-Beltrami operator in the \( \varepsilon^2 \) order; when \( \rho > 3 \), we get the fourth order differential operator in the \( \varepsilon^3 \), which depends on the non-uniform density function. See Figure 2.1 for a comparison of the estimated eigenvalues and the predicted eigenvalues under different setups. We clearly see that the eigenvalues are well predicted under different \( \rho \). When \( \rho = 8 \), we get the fourth order term that depends on the non-uniform density function; when \( \rho = 3 \), LLE is independent of the non-uniform density function and we recover the Laplace-Beltrami operator in the second order term, as is predicted by the developed theory; when \( \rho = -5 \), the non-uniform density function comes into play, and the eigenvalues are slightly shifted. To enhance the visualization, the difference between the estimated eigenvalues of \( S^1 \) and the theoretical values are shown on the middle subplot. The eigenfunctions provide more information. When \( \rho = -5 \) and \( \rho = 8 \), the dependence of the eigenfunctions on the non-uniform density function could be clearly seen.

Next, we show the results on \( S^2 \) with different radii under the non-uniform sampling scheme with \( \rho = 3 \) and different \( \varepsilon \)'s. Fix \( n = 30,000 \). Take uniform sampling points \( x_i = (x_{i1}, x_{i2}, x_{i3})^\top \in S^2 \subset \mathbb{R}^3 \), where \( i = 1, \ldots, n \), randomly choose \( n/10 \) points, randomly perturb those \( n/10 \) points by setting \( \bar{x}_i := x_i + 1 - \cos(2\pi U_i) \), where \( U_i \) is the uniform distribution on \([0, 1]\), and \( y_i := \frac{(x_{i1}, x_{i2}, x_{i3})}{\| (x_{i1}, x_{i2}, x_{i3}) \|} \). As a result, \( \mathcal{Y} := \{ y_i \}^{n}_{i=1} \subset S^2 \) is nonuniformly distributed on \( S^2 \). Denote \( r\mathcal{Y} \) to be the scaled sampling points on the sphere with radius \( r > 0 \). Run LLE on \( r\mathcal{Y} \) with different \( \varepsilon \)'s, and evaluate the first 400 eigenvalues. We consider \( r = 0.5, 1, 2 \). For \( r = 1 \), consider \( \varepsilon = 0.02 \); for \( r = 0.5 \), consider \( \varepsilon = 0.02/4 \) and 0.02/6; for \( r = 2 \), consider \( \varepsilon = 0.02 \times 4 \) and 0.02 \times 3. Based on the theory, when \( \rho = 3 \), LLE is independent of the non-uniform density function and we obtain the eigenvalues of the Laplace-Beltrami operator in all cases. See Figure 2.2 for the results under different setups. Theoretically, the eigenvalues of \( S^2 \) without counting multiplicities are \( \nu_i = -i(i + 1) \), where \( i = 0, 1, \ldots \). The multiplicity of \( \nu_i \) is \( 2i+1 \). When the radius is \( r > 0 \), the eigenvalues are scaled by \( r^2 \). The eigenvalues, as is shown in Figure 2.2, can be well estimated by LLE, and the gap between the eigenvalues of spheres with different radii is predicted. The sawtooth behavior of the error comes from the spectral convergence behavior of eigenvalues with multiplicities. Note that there are 19 eigenvalues with multiplicity greater than 1 in the first 400 eigenvalues, which match the 19 oscillations found in Figure 2.2(b). The eigenfunctions are shown in Figure 2.2(c). As is predicted, the first eigenfunction is constant, as is shown in \( \psi_1 \). The eigenspace of \( \nu_1 \) is spanned by three linear functions \( x, y, \) and \( z \), restricted on \( S^2 \). Therefore, \( \psi_1 \) is a linear. The eigenspace of \( \nu_2 \) is spanned by spherical harmonics of order \( \ell \), and its oscillation is illustrated in \( \psi_2 \) associated with \( \nu_2 \) and \( \psi_{16} \) associated with \( \nu_3 \).
CHAPTER 2. LLE ON CLOSED MANIFOLDS

72

(a) $S^1$ Eigenvalues

(b) $S^1$ Eigenvalues error

(c) $S^1$ Eigenfunctions

Figure 2.1: The first 400 eigenvalues of LLE on 30,000 points sampled from $S^1$ under a non-uniform sampling scheme with $\rho = -5, 3, 8$. $\lambda_k$ and $\psi_k$ are the $k$-th largest eigenvalue and the corresponding eigenfunction of LLE under different situations. $\lambda_k$ denotes the estimated $k$-th smallest eigenvalue of the Laplace-Beltrami operator or the fourth order differential operator under different situations. The theoretical value, $L_k := \left\lceil \frac{k-1}{2} \right\rceil^2$ for the Laplace-Beltrami operator and $\bar{L}_k := \left\lceil \frac{k-1}{4} \right\rceil^4 - \left\lceil \frac{k-1}{2} \right\rceil^2$ for the fourth order differential operator $f''' + f''$, where $\lceil x \rceil$ means the the least integer greater than or equal to $x$, are provided for a comparison. The eigenvalues and the theoretical values under different setups are shown in 2.1(a), with $L_k$ shown as the black crosses and $\bar{L}_k$ as the black circles. To enhance the visualization, the deviation of the evaluated eigenvalues from the theoretical values under different setups are shown in 2.1(b). The tenth eigenfunctions associated with the tenth largest eigenvalues of LLE under different setups are shown in 2.1(c). Note that when $\rho = 3$, we recover the spectrum of the Laplace-Beltrami operator when the sampling is non-uniform; when $\rho = -5$, the non-uniform density function comes into play, and the eigenvalues are shifted from the theoretical value. For the non-uniform sampling scheme and $\rho = 8$, theoretically the first three eigenvalues come from the sixth order term and depend on the non-uniform density function. Therefore, numerically the first three eigenvalues are non-zero. When $\rho = -5$ and $\rho = 8$, the eigenfunctions are the same (up to the global rotation), and depend on the non-uniform density function.

2.10.2 Examine the kernel

We now show the numerical simulations of the corresponding kernel on the unit circle $S^1$ embedded in $\mathbb{R}^2$. We take a uniform grid $\theta_i := 2\pi i/n$ on $(0, 2\pi]$, where $n \in \mathbb{N}$ and $i = 1, \ldots, n$, and construct $\mathcal{X} = \{x_i := (\cos(\theta_i), \sin(\theta_i))^\top \}_{i=1}^n \subset \mathbb{R}^2$, which could be viewed as a uniform sampled set from the unit circle. We fix $n = 10,000$. We then run LLE with $\varepsilon = [(\cos(\theta_{K/2}) - 1)^2 + \sin(\theta_{K/2})^2]^{1/2}$, where $K \in \mathbb{N}$. See Figure 2.3 for an example of the corresponding kernels when $K = 80$, and $K = 320$. Note that the constructed normalized kernel, $K_{\text{LLE}}(x_{1000}, y) \mathcal{X}_{\text{LLE}}(x_{1000}, y) dV(y)$, is non-positive.
2. LLE on Closed Manifolds

(a) $S^2$ Eigenvalues

(b) $S^2$ Eigenvalues Error

(c) $S^2$ Eigenfunctions

Figure 2.2: 2.2(a): the first 400 eigenvalues of LLE with $\rho = 3$ but different $\epsilon$, over a $n = 30,000$ non-uniform sampling points on $S^2$ with different radii $r > 0$. $\tilde{\lambda}_k$ is the $k$-th smallest eigenvalue of the Laplace-Beltrami operator estimated by LLE under different situations. When $r = 0.5$ (respectively $r = 1$ and $r = 2$), $\tilde{\lambda}_k$ are shown in the black (respectively blue and gray) curve. The results with different $\epsilon$ are shown as the red dash (respectively blue dash) when $r = 0.5$ (respectively $r = 2$). The theoretical eigenvalues for the canonical $S^2$ (with the radius 1), denoted as $L_k$, $k = 1, \ldots$, are provided for a comparison (superimposed as black circles). 2.2(b): to enhance the visualization, the difference between the theoretical values and numerical values, $\log_{10}(\tilde{\lambda}_k) - \log_{10}(L_k)$, are shown with the same color and line properties as those shown on 2.2(a). Some eigenfunctions evaluated when $r = 0.5$ are shown on 2.2(c).

Next, we show the numerical simulations of the corresponding kernel on the 1-dim flat torus $\mathbb{T}^1 \sim \mathbb{R}/\mathbb{Z}$ with the induced metric from the canonical metric on $\mathbb{R}^1$. We take a uniform grid on $\mathbb{T}^1$ as $\{ \theta_i = 2\pi i/n \}_{i=1}^n$, and take $\mathcal{X} = \{ x_i := (\cos(\theta_i), \sin(\theta_i)) \}_{i=1}^n \subset \mathbb{R}^2$ to illustrate the flat torus. Fix $n = 10,000$ and run LLE with $\epsilon = |\theta_K/2|$, where $K \in \mathbb{N}$. See Figure 2.3 for an example of the corresponding kernels when $K = 80$ and $K = 320$. The constructed normalized kernel, as the theory predicts, is constant. Note that in this case, we can view the flat 1-dim flat torus as the unit circle, when we have the access to the geodesic distance information on the manifold.

Finally, we take a look at the unit sphere $S^2$ embedded in $\mathbb{R}^3$ with the center at $(0, 0, 1)$, and its corresponding kernel. We uniformly sample $n$ points, $\mathcal{X} = \{ x_i \}_{i=1}^n \subset \mathbb{R}^2$, from $S^2$. Fix $n = 10,000$ and run LLE with 400 nearest neighbors. See Figure 2.3 for the corresponding kernel. Note that the normalized kernel is not positive. These examples show that even with the simple manifolds, the corresponding kernels might be complicated.
2.10.3 Two-dimensional random tomography example

To further examine the capability of LLE from the viewpoint of nonlinear dimension reduction, we consider the two-dimensional random tomography problem [56]. It is chosen because its geometrical structure is well known and complicated.

We briefly describe the dataset and refer the reader with interest to [56]. The classical two-dimensional transmission computerized tomography problem is to recover the function $f: \mathbb{R}^2 \to \mathbb{R}$ from its Radon transform. In the parallel beam model, the Radon transform of $f$ is given by the line integral $R_{\theta} f(s) = \int_{x \cdot \theta = s} f(x) dx$, where $\theta \in S^1$ is perpendicular to the beaming direction $\theta^\perp \in S^1$, where $S^1$ is the unit circle, and $s \in \mathbb{R}$. We call $\theta$ the projection direction and $R_{\theta} f$ the projected image. There are cases, however, in which we only have the projected images and the projection directions are unknown. In such cases, the problem at hand is to estimate $f$ from these projected images without knowing their corresponding projection directions. To better study this random projection problem, we need the following facts and assumptions. First, we know that for $f \in L^2(\mathbb{R}^2)$ with a compact support
within $B_1(0)$, the map $R \cdot f : \theta \in S^1 \mapsto L^2([-1,1])$ is continuous [56]. To simplify the discussion, we assume that there is no symmetry in $f$; that is, $R_{\theta_1}f$ and $R_{\theta_2}f$ are different for all pairs of $\theta_1 \neq \theta_2$. Next, take $S := \{s_i\}_{i=1}^p$ to be the chosen set of sampling points on $[-1,1]$, where $p \in \mathbb{N}$. In this example, we assume that $S$ is a uniform grid on $[-1,1]$; that is, $s_i = -1 + 2(i-1)/(p-1)$. For $\theta \in S^1$, denote the discretization of the projection image $R_{\theta}f$ as $D_S : L^2([-1,1]) \to \mathbb{R}^p$, which is defined by $D_S : R_{\theta}f \mapsto (R_{\theta f} \ast h_{\varepsilon}(s_1), R_{\theta f} \ast h_{\varepsilon}(s_2), \ldots, R_{\theta f} \ast h_{\varepsilon}(s_p))^\top \in \mathbb{R}^p$, where $h_{\varepsilon}(x) := \frac{1}{\varepsilon} h(\frac{x}{\varepsilon})$, $h$ is a Schwartz function, $h_{\varepsilon}$ converges weakly to the Dirac delta measure at 0 as $\varepsilon \to 0$. Note that, in general, $R_{\theta}f$ is a $L^2$ function when $f$ is a $L^2$ function. Therefore, we need a convolution to model the sampling step. We assume that the discretization $D_S$ is dense enough, so that $M^1 := \{D_p \circ R_{\theta} f\}_{\theta \in S^1}$ is also simple. In other words, we assume that $p$ is large enough so that $M^1$ is a one-dimensional closed simple curved embedded in $\mathbb{R}^p$ and $M^1$ is diffeomorphic to $S^1$. Finally, we sample finite points from $S^1$ uniformly and obtain the simulation.

With the above facts and assumptions, we sample the Radon transform $\mathcal{X} := \{x_i := D_S \circ R_{\theta} f\}_{i=1}^n \subset \mathbb{R}^p$ with finite projection directions $\{\theta_i\}_{i=1}^n$, where $\{\theta_i\}_{i=1}^n$ is a finite uniform grid on $S^1$; that is, $\mathcal{X}$ is sampled from the one-dimensional manifold $M^1$. For the simulations with the Shepp-Logan phantom, we take $n = 4096$, and the number of discretization points was $p = 128$. It has been shown in [56], that the DM could recover the $M^1$ up to diffeomorphism, that is, we could achieve the nonlinear dimensional reduction. In order to avoid distractions, we do not consider any noise as is considered in [56], and focus our analysis on the clean dataset. The Shepp-Logan image, some examples of the projections and the results of PCA, DM and LLE, are shown in Figure 2.4. As is shown in [56], the PCA fails to embed $\mathcal{X}$ with only the first three principal components, while the DM succeeded. There can be additional discussion for the DM, particularly its robustness to the noise and metric design. They have been extensively discussed in [56], so they are not discussed here. For LLE, we take $\varepsilon = 0.004$. The embedding results of LLE with different regularization orders, $\rho = 8, 3, -5$, are shown. Due to the complicated geometrical structure, we encounter difficulty even to recover the topology of $M^1$ by LLE, if the regularization order is not chosen properly.

To examine whether the sign of the kernel corresponding to LLE is indeterminate in this database, we fixed $x_{3555} \in \mathcal{X}$, and apply the PCA to visualize its $K = 150$ neighbors. The kernel function is shown in Figure 2.4 as the color encoded on the embedded points. The sign of the kernel is indeterminate, as is predicted by the above theory due to the existence of curvature. In summary, we should be careful when we apply LLE to a complicated real database.
Figure 2.4: Top row: the left panel is the Shepp-Logan phantom, the middle panel shows two projection images from two different projection directions, and the right panel shows the linear dimension reduction of the dataset by the first three principal components, $u_1$, $u_2$, and $u_3$. Middle row: the left panel shows the diffusion map (DM) of the dataset, where the embedding is done by choosing the first two non-trivial eigenvectors of the graph Laplacian, $\phi_2$ and $\phi_3$, and we simply take the Gaussian kernel to design the affinity without applying the $\alpha$-normalization technique [18], the middle panel shows the DM of the dataset, where we apply the $\alpha$-normalization technique when $\alpha = 1$, and the right panel shows that the sign of the kernel corresponding to the locally linear embedding (LLE) is indeterminate, where the black cross indicates $x_{3555}$, and the kernel value on its neighbors are encoded by color (the neighbors are visualized by the top three principal components, $v_1$, $v_2$, and $v_3$). Bottom row: the embedding using the second and third eigenvectors of LLE, $\psi_2$ and $\psi_3$, under different setups are shown. The left panel shows the result with $\rho = -5$, the middle panel shows the result with $\rho = 3$, and the right panel shows the result with $\rho = 8$. The results shows the importance of choosing the regularization and are explained by the theory.
Chapter 3

LLE on manifolds with boundary

3.1 Setup on manifolds with boundary and preliminary lemmas

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 $t:M \mapsto \mathbb{R}^p$. When the boundary of $M$ is not empty, we assume that it is smooth. When $\partial M \neq \emptyset$, we define the $\varepsilon$-neighborhood of $\partial M$ as

$$M_\varepsilon = \{x \in M|d(x, \partial M) < \varepsilon\}. \quad (3.1.1)$$

Let $P \in C^2(t(M))$ be the probability density function (p.d.f.) associated with $X$, and there exist $P_m > 0$ and $P_M \geq P_m$ so that $P_m \leq P(x) \leq P_M < \infty$ for all $x \in t(M)$. Let $\mathcal{X} = \{t(x_i)\}_{i=1}^n \subset t(M) \subset \mathbb{R}^p$ denote a set of identical and independent (i.i.d.) random samples from $X$, where $x_i \in M$. For $t(x_i) \in \mathcal{X}$ and $\varepsilon > 0$, we have $\mathcal{M}(x_i) := \{t(x_{i,1}), \ldots, t(x_{i,N})\} \subset B^\varepsilon(x_i) \cap (\mathcal{X} \setminus \{t(x_i)\})$. Take $G_n \in \mathbb{R}^{p \times N}$ to be the local data matrix associated with $\mathcal{M}(x_i)$ and evaluate the barycentric coordinates $w_n = [w_{n,1}, \ldots, w_{n,N}]^\top \in \mathbb{R}^N$.

For $x \in M_\varepsilon$, define

$$D_\varepsilon(x) = (t \circ \exp_x)^{-1}(B^\varepsilon_t \cap tM) \subset T_xM \approx \mathbb{R}^d. \quad (3.1.2)$$

Denote $x' = \arg\min_{y \in \partial M} d(y, x)$ and $\varepsilon(x) = \min_{y \in \partial M} d(y, x)$. Clearly, we have $0 \leq \varepsilon(x) \leq \varepsilon$. Choose the normal coordinates $\{\partial_i\}_{i=1}^d$ around $x$, so that $x' = t \circ \exp_x(\varepsilon(x)\partial_d)$. If $\varepsilon$ is sufficiently small, such $x'$ is unique.

In this chapter, we again assume the Assumption 2.1.2.

Note that $(t \circ \exp_x)^{-1}(B^\varepsilon_t \cap t(\partial M))$ can be regarded as the graph of a function depending on the curvature; that is, there is a domain $K \in \mathbb{R}^{d-1}$ and a smooth function $q$ defined on $K$, such that

$$(t \circ \exp_x)^{-1}(B^\varepsilon_t \cap t(\partial M)) = \{(u_1, \ldots, u_d) \in T_xM|\{u_1, \ldots, u_{d-1}\} \in K, u_d = q(u_1, \ldots, u_{d-1})\},$$

(3.1.3)

where $q(u_1, \ldots, u_{d-1})$ can be approximated by $\tilde{\varepsilon}(x) + \sum_{i=1}^{d-1} a_{ij}u_iu_j$ up to the error depending on a cubic function of $u_1, \ldots, u_{d-1}$ and $a_{ij}$ are the second fundamental form of the embedding of $\partial M$ to $\mathbb{R}^p$. Note that in general $(t \circ \exp_x)^{-1}(B^\varepsilon_t (t(x)) \cap t(\partial M))$ is not symmetric across the axes $\partial_1, \ldots, \partial_{d-1}$.

Now we define the symmetric region associated with $(t \circ \exp_x)^{-1}(B^\varepsilon_t (t(x)) \cap t(\partial M))$.

Definition 3.1.1. For $x \in M_\varepsilon$ and $\varepsilon$ sufficiently small, choose a normal coordinate $\{\partial_i\}_{i=1}^d$ around $x$ so that
arg min_{x \in \partial M} d(y, x) = 1 \circ \exp_x (\varepsilon(x) \partial_d). The symmetric region associated with \((1 \circ \exp_x)^{-1}(B^{R^p}_\varepsilon(t(x)))\) is defined as

\[ D_\varepsilon(x) = \{(u_1, \cdots, u_d) \in T_x M \mid \sum_{i=1}^d u_i^2 \leq \varepsilon^2 \text{ and } u_d \leq \tilde{\varepsilon}(x) + \sum_{i,j=1}^{d-1} a_{ij} u_i u_j \} \subset T_x M. \quad (3.1.4) \]

For \(x \notin M_{\varepsilon}\) and \(\varepsilon\) sufficiently small, choose a normal coordinate \(\{\partial_i\}_{i=1}^d\) around \(x\) and define the symmetric region associated with \((1 \circ \exp_x)^{-1}(B^{R^p}_\varepsilon(t(x)))\) as

\[ \tilde{D}_\varepsilon(x) = \{(u_1, \cdots, u_d) \in T_x M \mid \sum_{i=1}^d u_i^2 \leq \varepsilon^2 \} \subset T_x M. \quad (3.1.5) \]

Clearly, \(\tilde{D}_\varepsilon(x)\) is an approximation of \(D_\varepsilon(x)\) up to the third order error term. When \(x \in M_{\varepsilon}\), it is symmetric across \(\partial_1, \ldots, \partial_{d-1}\) since if \((u_1, \cdots, u_i, \cdots, u_d) \in D_\varepsilon(x)\), then \((u_1, \cdots, -u_i, \cdots, u_d) \in D_\varepsilon(x)\) for \(i = 1, \cdots, d-1\).

Next Lemma describes the error between \(\int_{D_\varepsilon(x)} f(u) du\) and \(\int_{\tilde{D}_\varepsilon(x)} f(u) du\).

**Lemma 3.1.1.** Fix \(x \in M\). When \(\varepsilon > 0\) is sufficiently small, we have

\[ \left| \int_{D_\varepsilon(x)} du - \int_{\tilde{D}_\varepsilon(x)} du \right| = O(\varepsilon^{d+2}). \quad (3.1.6) \]

**Proof.** Based on (2.2.3) 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 \(O(\varepsilon^3)\). The volume of the boundary \(\tilde{D}_\varepsilon(x)\) is of order \(O(\varepsilon^{d-1})\). Hence the volume difference between \(\tilde{D}_\varepsilon(x)\) and \(D_\varepsilon(x)\) is of order \(O(\varepsilon^{d-2} \varepsilon^3) = O(\varepsilon^{d+2})\). The conclusion follows. \(\square\)

For \(x \in M\), consider the following quantities, which could be understood as moments capturing the geometric asymmetry:

\[ \mu_v(x, \varepsilon) := \int_{D_\varepsilon(x)} \prod_{i=1}^d u_i^{v_i} du, \quad (3.1.7) \]

where \(v = [v_1, \ldots, v_d]^{\top}\) describes the moment order.

We summarize the behavior of moments that we need here:

**Lemma 3.1.2.** Suppose \(\varepsilon\) is sufficiently small. Then \(\mu_0(x, \varepsilon)\), \(\mu_{\varepsilon_1}(x, \varepsilon)\), \(\mu_{2\varepsilon}\), and \(\mu_{2\varepsilon + \varepsilon_1}(x, \varepsilon)\) are continuous functions of \(x\) on \(M\) for all \(i = 1, \cdots, d\). Define \(\frac{|S^{d-1}|}{d-1} = 1\) when \(d = 1\). Then, those functions can be quantitatively described as follows.

1. If \(x \in M_{\varepsilon}\), \(\mu_0\) is an increasing function of \(\varepsilon(x)\) and

\[ \mu_0(x, \varepsilon) = \frac{|S^{d-1}|}{2d} \varepsilon^d + \int_0^{\tilde{\varepsilon}(x)} \frac{|S^{d-2}|}{d-1} (\varepsilon^2 - h^2)^{d-1} dh + O(\varepsilon^{d+1}). \quad (3.1.8) \]

If \(x \notin M_{\varepsilon}\), then

\[ \mu_0(x, \varepsilon) = \frac{|S^{d-1}|}{d} \varepsilon^d. \quad (3.1.9) \]

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

\[ \frac{|S^{d-1}|}{2d} \varepsilon^d + O(\varepsilon^{d+1}) \leq \mu_0(x, \varepsilon) \leq \frac{|S^{d-1}|}{d} \varepsilon^d. \quad (3.1.10) \]
2. If \( x \in M_c \), \( \mu_{c_d} \) is an increasing function of \( \varepsilon(x) \) and

\[
\mu_{c_d}(x, \varepsilon) = -\frac{|S|^{d-2}}{d^2 - 1} \left( \varepsilon^2 - \bar{\varepsilon}(x)^2 \right)^{d+1} + O(\varepsilon^{d+2}).
\]  

(3.1.11)

If \( x \notin M_c \), then

\[
\mu_{c_d}(x, \varepsilon) = 0.
\]

(3.1.12)

In general, \( \mu_{c_d}(x, \varepsilon) \) is of order \( O(\varepsilon^{d+1}) \).

3. If \( x \in M_c \), \( \mu_{2e_i} \) is an increasing function of \( \bar{\varepsilon}(x) \) for \( i = 1, \ldots, d \). We have

\[
\mu_{2e_i}(x, \varepsilon) = \frac{|S|^{d-1}}{2d(d+2)} \varepsilon^{d+2} + \int_0^{\bar{\varepsilon}(x)} \frac{|S|^{d-2}}{(d-1)^2} \left( \varepsilon^2 - h^2 \right)^{d+1} dh + O(\varepsilon^{d+3}),
\]

for \( i = 1, \ldots, d - 1 \), and

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

(3.1.14)

If \( x \notin M_c \), then

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

(3.1.15)

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

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

(3.1.16)

4. If \( x \in M_c \), \( \mu_{2e_i, + e_d} \) is an increasing function of \( \bar{\varepsilon}(x) \) and

\[
\mu_{2e_i, + e_d}(x, \varepsilon) = -\frac{|S|^{d-2}}{(d-1)(d+3)} \left( \varepsilon^2 - \bar{\varepsilon}(x)^2 \right)^{d+1} + O(\varepsilon^{d+4}),
\]

for \( i = 1, \ldots, d - 1 \), and

\[
\mu_{3e_d}(x, \varepsilon) = -\frac{|S|^{d-2}}{(d^2 - 1)(d+3)} \left( \varepsilon^2 - \bar{\varepsilon}(x)^2 \right)^{d+1} (2\varepsilon^2 + (d+1)\bar{\varepsilon}(x)^2) + O(\varepsilon^{d+4}).
\]

(3.1.17)

If \( x \notin M_c \), then

\[
\mu_{2e_i, + e_d}(x, \varepsilon) = 0.
\]

(3.1.19)

In general, \( \mu_{2e_i, + e_d}(x, \varepsilon) \) is of order \( O(\varepsilon^{d+3}) \).

The proof is a straightforward integration, we omit it here. This Lemma tells us that when \( x \notin M_c \), that is, when \( x \) is far away from the boundary, all odd order moments disappear due to the symmetry of the integration domain. However, when \( x \in M_c \), it no longer holds – the integration domain becomes asymmetric, and the odd moments no longer disappear.

For \( v \in \mathbb{R}^p \), denote

\[
v = [v_1, v_2] \in \mathbb{R}^p,
\]

(3.1.20)
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 \( v = [v_1, v_2] \in T_1(M) \mathbb{R}^p \), \( v_1 = J^T_{p,d} v \) is the coordinate of the tangential component of \( v \) on \( t_1 M \) and \( v_2 = J^T_{p,p-d} v \) is the coordinate of the normal component of \( v \) associated with a chosen basis of the normal bundle.

Define \( \mathcal{N}_{ij}(x) = J^T_{p,p-d} \mathbb{1}(e_i, e_j) \). Note that \( \mathcal{N}_{ij}(x) \) is symmetric.

For \( d \leq r \leq p \), we define

\[
\mathfrak{J}_{p,r-d} := J_{p,p-d} J_{p-d,r-d} \in \mathbb{R}^{p \times (r-d)}.
\] (3.1.21)

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 \( \mathbb{E}[\chi_{B^p_{\varepsilon} (1(x))}(X)] \), \( \mathbb{E}[(f(X) - f(x))\chi_{B^p_{\varepsilon} (1(x))}(X)] \), and the first two order terms in the tangent component of \( \mathbb{E}[(X - t(x))\chi_{B^p_{\varepsilon} (1(x))}(X)] \) and \( \mathbb{E}[(X - t(x))(f(X) - f(x))\chi_{B^p_{\varepsilon} (1(x))}(X)] \).

**Lemma 3.1.3.** Fix \( x \in M \) and \( f \in C^3(M) \). When \( \varepsilon > 0 \) is sufficiently small, the following expansions hold.

1. \( \mathbb{E}[\chi_{B^p_{\varepsilon} (1(x))}(X)] \) satisfies

\[
\mathbb{E}[\chi_{B^p_{\varepsilon} (1(x))}(X)] = P(x) \mu_0(x, \varepsilon) + \partial_d P(x) \mu_d(x, \varepsilon) + O(\varepsilon^{d+2}) \] (3.1.22)

2. \( \mathbb{E}[(f(X) - f(x))\chi_{B^p_{\varepsilon} (1(x))}(X)] \) satisfies

\[
\mathbb{E}[(f(X) - f(x))\chi_{B^p_{\varepsilon} (1(x))}(X)] = P(x) \partial_d f(x) \mu_d(x, \varepsilon)
+ \sum_{i=1}^d \frac{P(x)}{2} \partial^2_{dd} f(x) + \partial_i f(x) \partial_d P(x) \mu_{2d}(x, \varepsilon) + O(\varepsilon^{d+3}),
\] (3.1.23)

3. The vector \( \mathbb{E}[(X - t(x))\chi_{B^p_{\varepsilon} (1(x))}(X)] \) satisfies

\[
\mathbb{E}[(X - t(x))\chi_{B^p_{\varepsilon} (1(x))}(X)] = [v_1, v_2],
\] (3.1.24)

where

\[
v_1 = P(x) \mu_0(x, \varepsilon) J^T_{p,d} e_d + \sum_{i=1}^d (\partial_i P(x) \mu_{2d}(x, \varepsilon)) J^T_{p,d} e_i + O(\varepsilon^{d+3})
\]

\[
v_2 = \frac{P(x)}{2} \sum_{i=1}^d \mathcal{N}_{ii}(x) \mu_{2d} + O(\varepsilon^{d+3}).
\] (3.1.25)

4. The vector \( \mathbb{E}[(X - t(x))(f(X) - f(x))\chi_{B^p_{\varepsilon} (1(x))}(X)] \) satisfies

\[
\mathbb{E}[(X - t(x))(f(X) - f(x))\chi_{B^p_{\varepsilon} (1(x))}(X)] = [v_1, v_2],
\] (3.1.26)
where

\[ v_1 = P(x) \sum_{i=1}^{d} \left( \partial_i f(x) \mu_{2e_i}(x, \varepsilon) \right) J^{T}_{p,d} e_i \]

\[ + \sum_{i=1}^{d-1} \left[ \partial_i f(x) \partial_d P(x) + \partial_d f(x) \partial_i P(x) + P(x) \partial_{e_i}^2 f(x) \right] \mu_{2e_i+e_d}(x, \varepsilon) J^{T}_{p,d} e_i \]

\[ + \sum_{i=1}^{d} \left( [\partial_i f(x) \partial_d P(x) + \frac{P(x)}{2} \partial_{e_i}^2 f(x)] \mu_{2e_i+e_d}(x, \varepsilon) \right) J^{T}_{p,d} e_d + O(\varepsilon^{d+4}), \] (3.1.27)

\[ v_2 = P(x) \sum_{i=1}^{d-1} \partial_i f(x) \mu_{2e_i}(x, \varepsilon) + \frac{P(x)}{2} \partial_d f(x) \mu_{e_d}(x, \varepsilon) + O(\varepsilon^{d+4}). \] (3.1.28)

Note that this Lemma could be viewed as the generalization of Lemma 2.2.5 to the boundary. In particular, when \( x \notin M_\varepsilon \), we recover Lemma 2.2.5.

**Proof.** First, we calculate \( \mathbb{E}[X^B_{\varepsilon}(\{x\})](X) \).

\[ \mathbb{E}[X^B_{\varepsilon}(\{x\})](X) \] (3.1.29)

\[ = \int_{D_\varepsilon(x)} (P(x) + \sum_{i=1}^{d} \partial_i P(x) u_i + O(u^2)) \left( 1 - \frac{1}{6} \sum_{i,j=1}^{d} \text{Ric}_\varepsilon(i, j) u_i u_j + O(u^3) \right) du \]

\[ = P(x) \int_{D_\varepsilon(x)} du + \int_{D_\varepsilon(x)} \sum_{i=1}^{d} \partial_i P(x) u_i du + O(\varepsilon^{d+2}) \]

\[ = P(x) \mu_0(x, \varepsilon) + \partial_d P(x) \mu_{e_d}(x, \varepsilon) + O(\varepsilon^{d+2}), \]

where the second equality holds by applying Lemma 3.1.1 that the error of changing domain from \( D_\varepsilon(x) \) to \( \bar{D}_\varepsilon(x) \) is of order \( O(\varepsilon^{d+2}) \). Note that \( P(x) \) is bounded away from 0.

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

\[ f \circ \exp_\varepsilon(u) - f(x) = \sum_{i=1}^{d} \partial_i f(x) u_i + \frac{1}{2} \sum_{i,j=1}^{d} \partial_{i,j}^2 f(x) u_i u_j + O(u^3), \] (3.1.30)

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

\[ \mathbb{E}[(f(X) - f(x))X^B_{\varepsilon}(\{x\})](X) \] (3.1.31)

\[ = \int_{D_\varepsilon(x)} \left( \sum_{i=1}^{d} \partial_i f(x) u_i + \frac{1}{2} \sum_{i,j=1}^{d} \partial_{i,j}^2 f(x) u_i u_j + O(u^3) \right) (P(x) + \sum_{i=1}^{d} \partial_i P(x) u_i + O(u^2)) \]

\[ \times \left( 1 - \frac{1}{6} \sum_{i,j=1}^{d} \text{Ric}_\varepsilon(i, j) u_i u_j + O(u^3) \right) du, \]
which by Lemma 3.1.1 and the symmetry of $D_\varepsilon(x)$ becomes

$$\int_{D_\varepsilon(x)} \left[ P(x) \sum_{i=1}^{d} \partial_i f(x) u_i + \frac{P(x)}{2} \sum_{i,j=1}^{d} \partial^2_{ij} f(x) u_i u_j + \sum_{i=1}^{d} \partial_i f(x) u_i \sum_{j=1}^{d} \partial_j P(x) u_j + O(u^2) \right] du$$

$$= P(x) \partial_d f(x) \int_{D_\varepsilon(x)} u_d du + \frac{P(x)}{2} \sum_{i,j=1}^{d} \partial^2_{ij} f(x) + \partial_i f(x) \partial_d P(x) \int_{D_\varepsilon(x)} u_i d\mu + O(\varepsilon^{d+3})$$

$$= P(x) \partial_d f(x) \mu_{d\varepsilon}(x, \varepsilon) + \frac{P(x)}{2} \sum_{i,j=1}^{d} \partial^2_{ij} f(x) + \partial_i f(x) \partial_d P(x) \mu_{2\varepsilon}(x, \varepsilon) + O(\varepsilon^{d+3}),$$

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

Third, by a direct expansion, we have

$$\mathbb{E}[ (X - \iota(x)) X_{B_P^\varepsilon(t)(x)}(X) ] = \int_{D_\varepsilon(x)} (t, u + O(u^3)) (P(x) + \sum_{i=1}^{d} \partial_i P(x) u_i + O(u^2))$$

$$\times \left[ 1 - \frac{1}{6} \sum_{i,j=1}^{d} \partial \varepsilon_{(i,j)} u_i u_j + O(u^3) \right] du$$

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

$$\int_{D_\varepsilon(x)} (t, u + O(u^3)) (P(x) + \sum_{i=1}^{d} \partial_i P(x) u_i + O(u^2))$$

$$\times \left[ 1 - \frac{1}{6} \sum_{i,j=1}^{d} \partial \varepsilon_{(i,j)} u_i u_j + O(u^3) \right] du$$

$$= \int_{D_\varepsilon(x)} (t, u + O(u^3)) (P(x) + \sum_{i=1}^{d} \partial_i P(x) u_i + O(u^2))$$

$$\times \left[ 1 - \frac{1}{6} \sum_{i,j=1}^{d} \partial \varepsilon_{(i,j)} u_i u_j + O(u^3) \right] du + O(\varepsilon^{d+3}),$$

where the equality holds by Lemma 3.1.1. Similarly, by Lemma 3.1.1, the normal part is

$$\int_{D_\varepsilon(x)} \left( \frac{1}{2} \| \iota \| (u, u) + O(u^3) \right) (P(x) + \sum_{i=1}^{d} \partial_i P(x) u_i + O(u^2))$$

$$\times \left[ 1 - \frac{1}{6} \sum_{i,j=1}^{d} \partial \varepsilon_{(i,j)} u_i u_j + O(u^3) \right] du$$

$$= \int_{D_\varepsilon(x)} \left( \frac{1}{2} \| \iota \| (u, u) + O(u^3) \right) (P(x) + \sum_{i=1}^{d} \partial_i P(x) u_i + O(u^2))$$

$$\times \left[ 1 - \frac{1}{6} \sum_{i,j=1}^{d} \partial \varepsilon_{(i,j)} u_i u_j + O(u^3) \right] du + O(\varepsilon^{d+4})$$

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

$$v_1 = J_{p,d}^T \left[ P(x) \int_{D_e(x)} t_u du + \int_{D_e(x)} t_u \sum_{i=1}^d \partial_i P(x) u_i du + O(\varepsilon^{d+3}) \right]$$

$$= \left( P(x) \int_{D_e(x)} u_d du \right) J_{p,d}^T e_d + \sum_{i=1}^d \left( \partial_i P(x) \int_{D_e(x)} u_i^2 du \right) J_{p,d}^T e_i + O(\varepsilon^{d+3})$$

$$= P(x) \mu_d(x, \varepsilon) J_{p,d}^T e_d + \sum_{i=1}^d \partial_i P(x) \mu_{2c_i}(x, \varepsilon) J_{p,d}^T e_i + O(\varepsilon^{d+3})$$

$$v_2 = \frac{P(x)}{2} \int_{D_e(x)} \Pi_x(u, u) du + O(\varepsilon^{d+3})$$

$$= \frac{P(x)}{2} \sum_{i=1}^d \gamma_{11}(x) \mu_{2c_i} + O(\varepsilon^{d+3}).$$

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

$$\mathbb{E}[(X - t(x))(f(X) - f(x))X_{B_{\mathbb{R}^p}^p(t(x))}(X)] = \int_{D_e(x)} (t_u + \frac{1}{2} \Pi_x(u, u) + O(u^3)) \left( \sum_{i=1}^d \partial_i f(x) u_i + \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 f(x) u_i u_j + O(u^3) \right)$$

$$\times \left( P(x) + \sum_{i=1}^d \partial_i P(x) u_i + O(u^2) \right) \left( 1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i, j) u_i u_j + O(u^3) \right) du.$$

The tangential part is

$$\int_{D_e(x)} (t_u + O(u^3)) \left( \sum_{i=1}^d \partial_i f(x) u_i + \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 f(x) u_i u_j + O(u^3) \right)$$

$$\times \left( P(x) + \sum_{i=1}^d \partial_i P(x) u_i + O(u^2) \right) \left( 1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i, j) u_i u_j + O(u^3) \right) du.$$  \hspace{1cm} (3.1.38)

The leading term $P(x) t_u \sum_{i=1}^d \partial_i f(x) u_i$ is of order $O(\varepsilon^2)$ on $D_e(x)$, therefore the error of changing domain from $D_e(x)$ to $D_e(x)$ is of order $O(\varepsilon^{d+4})$. The normal part is

$$\int_{D_e(x)} \left( \frac{1}{2} \Pi_x(u, u) + O(u^3) \right) \left( \sum_{i=1}^d \partial_i f(x) u_i + \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 f(x) u_i u_j + O(u^3) \right)$$

$$\times \left( P(x) + \sum_{i=1}^d \partial_i P(x) u_i + O(u^2) \right) \left( 1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i, j) u_i u_j + O(u^3) \right) du.$$  \hspace{1cm} (3.1.39)

The leading term $P(x) \Pi_x(u, u) \sum_{i=1}^d \partial_i f(x) u_i$ is of order $O(\varepsilon^3)$ on $D_e(x)$. Therefore, the error of changing domain from $D_e(x)$ to $D_e(x)$ is of order $O(\varepsilon^{d+5})$. Putting the above together, $\mathbb{E}[(X - t(x))(f(X) - f(x))X_{B_{\mathbb{R}^p}^p(t(x))}(X)] = \|v_1, v_2\|$. 

[[v_1, v_2]], where by the symmetry of \( D_x(x) \) we have

\[
v_1 = J_{p,d}^{\top} \left[ P(x) \int_{D_x(x)} u_i \sum_{i=1}^d \partial_i f(x) u_i du + \int_{\partial D_x(x)} \mu_i u_i \sum_{i=1}^d \partial_i f(x) u_i du \right] = P(x) \sum_{i=1}^d \left( \partial_i f(x) \int_{D_x(x)} u_i^2 du \right) J_{p,d}^{\top} e_i
\]

\[
+ P(x) \sum_{i=1}^{d-1} \left[ \partial_i f(x) \partial_d P(x) + \partial_d f(x) \partial_i P(x) + P(x) \partial_d^2 f(x) \right] \int_{D_x(x)} u_i^2 u_d du J_{p,d}^{\top} e_i
\]

\[
+ \frac{P(x)}{2} \sum_{i=1}^{d-1} \left[ \partial_d f(x) \partial_i P(x) + P(x) \partial_d^2 f(x) \right] \int_{D_x(x)} u_i^2 u_d du J_{p,d}^{\top} e_i + O(\varepsilon^{d+4})
\]

\[
v_2 = P(x) \sum_{i=1}^{d-1} \partial_i f(x) \mu_{2\varepsilon_d}(x, \varepsilon) J_{p,d}^{\top} e_i + O(\varepsilon^{d+4}),
\]

\[
\text{and}
\]

\[
v_2 = P(x) \sum_{i=1}^d \partial_i f(x) \eta_{ud}(x) \mu_{2\varepsilon_d}(x, \varepsilon) + \frac{P(x)}{2} \partial_d f(x) \eta_{dd}(x) \mu_{3\varepsilon_d}(x, \varepsilon) + O(\varepsilon^{d+4}).
\]

\\]

3.2 Variance analysis and biased analysis on manifolds with boundary, when \( \rho = 3 \)

From now on, we fix the regularizer in (1.2.5) to be

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

Recall that \( C_x = \mathbb{E}[(X - t(x))(X - t(x))^\top] \mathbb{R}^p \). Suppose \( \text{rank}(C_x) = r \). For \( x \in M \), let \( C_x = U A U^\top \) be the eigendecomposition of \( C_x \). Let \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_p = 0 \) be the eigenvalues of \( C_x \) and let \( \beta_i \) be the corresponding orthonormal eigenvector.

In next lemma we calculate \( C_x \). This Lemma could be viewed as the generalization of Proposition 2.3.1 in the sense that when \( x \notin M_e \), the result is reduced to that of Proposition 2.3.1. To handle the boundary effect, we only need to calculate the first two order terms in eigenvalues and orthonormal eigenvectors of \( C_x \).
Lemma 3.2.1. Fix $x \in M$.

$$
C_x = P(x) \begin{bmatrix}
M^{(0)}(x, \varepsilon) & 0 \\
0 & 0
\end{bmatrix} + \begin{bmatrix}
M^{(11)}(x, \varepsilon) & M^{(12)}(x, \varepsilon) \\
M^{(21)}(x, \varepsilon) & 0
\end{bmatrix} + O(\varepsilon^{d+4})
$$

(3.2.2)

where $M^{(0)}$ is a $d \times d$ diagonal matrix with the $m$-th diagonal entry $\mu_{2m}(x, \varepsilon)$. In particular, when $x \notin M_{\varepsilon}$, \[ M^{(11)}(x, \varepsilon) \begin{bmatrix}
M^{(21)}(x, \varepsilon) & 0
\end{bmatrix} = 0. \]

The first $d$ eigenvalues of $C_x$ are

$$
\lambda_i = P(x) \mu_{2i}(x, \varepsilon)[1 + \gamma_i(x)\varepsilon] + O(\varepsilon^{d+4}),
$$

(3.2.3)

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

Suppose that $\text{rank}(C_x) = r$, then the corresponding orthonormal eigenvector matrix is

$$
X(x, \varepsilon) = X(x, 0) + \varepsilon X(x, 0) S(x) + O(\varepsilon^2),
$$

(3.2.4)

where

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

(3.2.5)

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

The proof is essentially the same as that of Proposition 2.3.1, except that we need to handle the fact that the integral domain is no longer symmetric when $x$ is close to the boundary.

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

$$
e_m^T C_x e_n = \int_{D_{\varepsilon}(x)} (t(y) - t(x))^T e_m(t(y) - t(x))^T e_n P(y) dV(y).
$$

(3.2.6)

By the expression

$$
t \circ \exp_x(u) - t(x) = t_u + \frac{1}{2} \Pi_x(u, u) + O(u^3),
$$

(3.2.7)

we have

$$
(t(y) - t(x))^T e_m(t(y) - t(x))^T e_n = (e_m^T t_u)(e_n^T t_u) + \frac{1}{2} (e_m^T t_u)(\Pi_x(u, u)) + \frac{1}{2} (e_m^T \Pi_x(u, u))(e_n^T t_u) + O(u^4).
$$

(3.2.8)
Thus, (3.2.6) is reduced to
\[
e^{m\top}C_se_n = \int_{D_\varepsilon(x)} \left( (e^{m\top}_n u)(e^{n\top}_n u) + \frac{1}{2} (e^{m\top}_n u)(e^{n\top}_n \Pi_x(u,u)) + \frac{1}{2} (e^{m\top}_n \Pi_x(u,u))(e^{n\top}_n u) + O(u^4) \right)
\times \left( P(x) + \nabla_u P(x) + O(u^2) \right) \left( 1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i,j) u_i u_j + O(u^3) \right) du.
\]

For \(1 \leq m, n \leq d\), \( (e^{m\top}_n u)(e^{n\top}_n u) = u_m u_n \). Moreover \( e^{n\top}_n \Pi_x(u,u) \) and \( e^{m\top}_n \Pi_x(u,u) \) are zero, so
\[
e^{m\top}C_se_n = \int_{D_\varepsilon(x)} (u_m u_n + O(u^4)) \left( P(x) + \nabla_u P(x) + O(u^2) \right) \left( 1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i,j) u_i u_j + O(u^3) \right) du
\]

(3.2.10)

where we use Lemma 3.1.1 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 \( 1 \leq m = n \leq d \),
\[
M^{(0)}_{m,n} = \int_{D_\varepsilon(x)} u_m^2 du = \mu_{2e_m}(x, \varepsilon)
\]

(3.2.11)

and \( M^{(0)}_{m,n} \) is 0 otherwise.

Next,
\[
M^{(1)}_{m,n} = \int_{D_\varepsilon(x)} u_m u_n \sum_{k=1}^d u_k \partial_k P(x) du
\]

(3.2.12)

So, by the symmetry of domain \( \tilde{D}_\varepsilon(x) \), we have
\[
M^{(1)}_{m,n} = \begin{cases} 
\partial_d P(x) \mu_{2e_m+e_d} & 1 \leq m = n \leq d, \\
\partial_n P(x) \mu_{2e_m+e_d} & m = d, 1 \leq n \leq d, \\
\partial_m P(x) \mu_{2e_m+e_d} & n = d, 1 \leq m \leq d, \\
0 & \text{otherwise}.
\end{cases}
\]

For \( 1 \leq m \leq d \) and \( n \geq d \),
\[
e^{m\top}C_se_n = \int_{D_\varepsilon(x)} \left( \frac{1}{2} (e^{m\top}_n u)(e^{n\top}_n \Pi_x(u,u)) + O(u^4) \right) \left( P(x) + \nabla_u P(x) + O(u^2) \right)
\times \left( 1 - \sum_{i,j=1}^d \frac{1}{6} \text{Ric}_x(i,j) u_i u_j + O(u^3) \right) du
\]

(3.2.13)

\[
= \frac{P(x)}{2} \int_{D_\varepsilon(x)} u_m (e^{n\top}_n \Pi_x(u,u)) du + O(\varepsilon^{d+4})
\]

(3.2.14)

We use Lemma 3.1.1 to handle the error of changing domain from \( D_\varepsilon(x) \) to \( \tilde{D}_\varepsilon(x) \), which is \( O(\varepsilon^{d+5}) \), hence
\[
M^{(12)}_{m,n-d} = \frac{P(x)}{2} \int_{D_\varepsilon(x)} u_m (e^{n\top}_n \Pi_x(u,u)) du.
\]

(3.2.15)
Similarly, for \( 1 \leq n \leq d \) and \( m \geq d \),
\[
e_m^\top C_x e_n = \frac{P(x)}{2} \int_{D_\varepsilon(x)} u_n (e_m^\top \Pi_x(u,u)) du + O(\varepsilon^{d+4}).
\] (3.2.16)

And,
\[
M^{(21)}(x)_{m-d,n} = \frac{P(x)}{2} \int_{D_\varepsilon(x)} u_m (e_n^\top \Pi_x(u,u)) du.
\] (3.2.17)

At last, for \( n \geq d \) and \( m \geq d \), \( e_m^\top t.u \) and \( e_n^\top t.u \) are 0, therefore, \( e_m^\top C_x e_n = O(\varepsilon^{d+4}) \).

Based on Lemma 3.1.2, \( \begin{bmatrix} M^{(0)}(x,\varepsilon) & 0 \\ 0 & 0 \end{bmatrix} \) is of order \( O(\varepsilon^{d+2}) \) and \( \begin{bmatrix} M^{(11)}(x,\varepsilon) & M^{(12)}(x,\varepsilon) \\ M^{(21)}(x,\varepsilon) & 0 \end{bmatrix} \) is of order \( O(\varepsilon^{d+3}) \). Note that the entries of \( \begin{bmatrix} M^{(11)}(x,\varepsilon) & M^{(12)}(x,\varepsilon) \\ M^{(21)}(x,\varepsilon) & 0 \end{bmatrix} \) are integral of odd order polynomial over \( D_\varepsilon(x) \), hence, the matrix is 0 when \( x \notin M_\varepsilon \).

We can apply the perturbation theory in the introduction, the first \( d \) eigenvalues of \( C_x \) are
\[
\lambda_i = P(x) \mu_{2\varepsilon}(x,\varepsilon) + \lambda_i^{(1)}(x) + O(\varepsilon^{d+4}),
\] (3.2.18)
for \( i = 1, \ldots, d \) and any \( x \in M \), where \( \{\lambda_i^{(1)}(x)\} \) are eigenvalues of \( M^{(11)}(x,\varepsilon) \). And \( \{\lambda_i^{(1)}(x)\} \) are of order \( O(\varepsilon^{d+3}) \).

Note that \( \mu_{2\varepsilon}(x,\varepsilon) \) is of order \( \varepsilon^{d+2} \), thus we define an order \( O(1) \) term,
\[
\gamma(x) = \frac{\lambda_i^{(1)}(x)}{P(x) \mu_{2\varepsilon}(x,\varepsilon) \varepsilon}.
\] (3.2.19)

Here \( \gamma(x) \) depends on \( P \). When \( x \notin M_\varepsilon \), since \( \{\lambda_i^{(1)}(x)\} \) are zero, we have \( \gamma(x) = 0 \). Therefore,
\[
\lambda_i = P(x) \mu_{2\varepsilon}(x,\varepsilon) [1 + \gamma(x) \varepsilon] + O(\varepsilon^{d+4}),
\] (3.2.20)
where \( i = 1, \ldots, d \). Moreover, \( \lambda_i = O(\varepsilon^{d+4}) \) for \( i = d+1, \ldots, p \).

Suppose that \( \text{rank}(C_x) = r \), based on the perturbation theory in the introduction, the orthonormal eigenvector matrix of \( C_x \) is in the form
\[
X(x,\varepsilon) = \begin{bmatrix} X_1(x) & 0 & 0 \\ 0 & X_2(x) & 0 \\ 0 & 0 & X_3(x) \end{bmatrix} + \varepsilon \begin{bmatrix} X_1(x) & 0 & 0 \\ 0 & X_2(x) & 0 \\ 0 & 0 & X_3(x) \end{bmatrix} S(x) + O(\varepsilon^2),
\] (3.2.21)
where \( X_1(x) \in O(d) \), \( X_2(x) \in O(r-d) \) and \( X_3(x) \in O(p-r) \).

\( S(x) \) is an antisymmetric matrix depending on \( \begin{bmatrix} M^{(11)}(x,\varepsilon) & M^{(12)}(x,\varepsilon) \\ M^{(21)}(x,\varepsilon) & 0 \end{bmatrix} \) and the higher order terms of \( C_x \).
In particular, if \( x \notin M_\varepsilon \), \( S(x) = 0 \). Moreover, if among first \( d \) eigenvalues of \( C_x \), there are \( 1 \leq t \leq d \) distinct ones,
then there is a choice of basis \( \{ e_1, \cdots, e_d \} \) in the tangent space of \( M \) so that

\[
X_1(x) = \begin{bmatrix}
X_1^{(1)}(x) & 0 & \cdots & 0 \\
0 & X_1^{(2)}(x) & \cdots & 0 \\
0 & 0 & \ddots & 0 \\
0 & 0 & \cdots & X_1^{(d)}(x)
\end{bmatrix},
\]

(3.2.22)

where each \( X_1^{(i)}(x) \) is an orthogonal matrix corresponding to the same eigenvalue of \( C_x \). The conclusion follows.

Recall that

\[
T_{t(x)}^\top = \mathbb{E}[(X - t(x))X_{\beta^p(t(x))}(X)]^\top U_{p,r}(A + \varepsilon^{d+3} I_{p \times p})^{-1} U^\top
\]

(3.2.23)

\[
= \sum_{i=1}^r \frac{\mathbb{E}[(X - t(x))X_{\beta^p(t(x))}(X)]^\top \beta_i^\top}{\lambda_i + \varepsilon^{d+3}} \in \mathbb{R}^p.
\]

(3.2.24)

We are going to calculate \( T_{t(x)} \) in the next lemma. For our purpose, we only need to calculate the first two order terms in the tangent component of \( T_{t(x)} \) and the first term in the normal direction.

**Lemma 3.2.2.** \( T_{t(x)} = [v_1^{(-1)} + v_{1,1}^{(0)} + v_{1,2}^{(0)}, v_2^{(-1)}] + \|O(\varepsilon), O(1)\| \), where

\[
v_1^{(-1)} = \frac{\mu_{e_d}(x, \varepsilon)}{\mu_{e_d}(x, \varepsilon)} J_{p,d}^\top e_d
\]

(3.2.25)

\[
v_{1,1}^{(0)} = \frac{\nabla P(x)}{P(x)},
\]

(3.2.26)

\[
v_{1,2}^{(0)} = -\frac{\varepsilon \mu_{e_d}(x, \varepsilon) y_d(x)}{\mu_{e_d}(x, \varepsilon)} J_{p,d}^\top e_d - \frac{\mu_{e_d}(x, \varepsilon) \varepsilon^{d+3}}{P(x)} J_{p,d}^\top e_d
\]

(3.2.27)

\[
+ \frac{\varepsilon \mu_{e_d}(x, \varepsilon)}{\mu_{e_d}(x, \varepsilon)} \sum_{i=1}^d \left[ \frac{\mu e_d}{\mu_{e_d}(x, \varepsilon)} \left( e_d J_{p,d} X_1(x) S_{11}(x) J_{p,d}^\top e_i \right) X_1(x) J_{p,d}^\top e_i \right]
\]

\[
+ \frac{\sum_{i=d+1}^r \frac{P(x)}{P(x)} \sum_{j=1}^d \frac{\mu e_d}{\mu_{e_d}(x, \varepsilon)} \left( e_d J_{p,d} X_1(x) S_{12}(x) J_{p,d}^\top e_i \right) X_1(x) S_{12}(x) J_{p,d}^\top e_i \right]
\]

and

\[
v_2^{(-1)} = \sum_{i=d+1}^r \left[ \frac{P(x)}{2} \frac{\mu_{e_d}(x, \varepsilon)}{\varepsilon^{d+2}} \left( e_d J_{p,d} X_1(x) S_{12}(x) J_{p,d}^\top e_i \right) X_1(x) J_{p,d}^\top e_i \right]
\]

(3.2.28)

\[
+ \sum_{i=d+1}^r \frac{P(x)}{2} \sum_{j=1}^d \frac{\mu e_d}{\mu_{e_d}(x, \varepsilon)} \left( e_d J_{p,d} X_1(x) S_{12}(x) J_{p,d}^\top e_i \right) X_1(x) J_{p,d}^\top e_i \right]
\]

(3.2.28)

Note that by Lemma 3.1.2, \( v_1^{(-1)} \) is of order \( \varepsilon^{-1} \) 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 \( \varepsilon^{d+1} \) and \( \mu_{e_{d'}}(x, \varepsilon) \) is of order \( \varepsilon^{d+2} \) for \( i = 1, \ldots, d \). Moreover, when \( x \notin M_\varepsilon \), we have \( \mu_{e_d}(x, \varepsilon) = 0 \) and \( S_{12}(x) = 0 \), hence \( v_{1,2}^{(0)} = 0 \). Similarly, \( v_2^{(-1)} \) is of order \( \varepsilon^{-1} \).
Proof. Recall that
\[
T_{t(x)}^\top = \sum_{i=1}^r \frac{\mathbb{E}(\langle X - t(x) \rangle X_{B^{ep}_{E}(t;i)}(X))\beta_i\beta_i^\top}{\lambda_i + \epsilon^{d+3}}.
\] (3.2.29)

To show the proof, we evaluate the terms in \( T_{t(x)} \) one by one.

Based on Lemma 3.2.1, the first \( d \) eigenvalues are \( \lambda_i = P(x)\mu_{2\epsilon}r(x, \epsilon) [1 + \gamma_i(x)\epsilon] + O(\epsilon^{d+4}) \), where \( i = 1, \ldots, d \), and the corresponding eigenvectors are
\[
\beta_i = \begin{bmatrix} X_1(x)J^\top_{p,d}e_i \\ 0_{(p-d)\times 1} \end{bmatrix} + \begin{bmatrix} \epsilon X_1(x)S_{11}(x)J^\top_{p,d}e_i + O(\epsilon^2) \\ O(\epsilon) \end{bmatrix},
\] (3.2.30)

where \( X_1(x) \in O(d) \).

For \( i = d + 1, \ldots, r \), \( \lambda_i = O(\epsilon^{d+4}) \), and the corresponding eigenvectors are
\[
\beta_i = \begin{bmatrix} 0_{d\times 1} \\ J_{p-d,r-d}X_2(x)J^\top_{p-r-d}e_i \end{bmatrix} + \begin{bmatrix} \epsilon X_1(x)S_{12}(x)J^\top_{p-r-d}e_i + O(\epsilon^2) \\ O(\epsilon) \end{bmatrix},
\] (3.2.31)

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

By Lemma 3.1.3, we have
\[
\mathbb{E}(\langle X - t(x) \rangle X_{B^{ep}_{E}(t;i)}(X)) = [v_1, v_2],
\] (3.2.32)

where
\[
v_1 = P(x)\mu_{d}(x, \epsilon)J^\top_{p,d}e_d + \sum_{i=1}^d \partial_t P(x)\mu_{2\epsilon}r(x, \epsilon)J^\top_{p,d}e_i + O(\epsilon^{d+3}),
\]
\[
v_2 = \frac{P(x)}{2} \sum_{i=1}^d \gamma_i(x)\mu_{2\epsilon}r_i + O(\epsilon^{d+3}).
\] (3.2.33)

Next, we calculate \( \mathbb{E}(\langle X - t(x) \rangle X_{B^{ep}_{E}(t;i)}(X))\beta_i \), for \( i = 1, \ldots, d \). Note that the normal component of \( \beta_i \) is of order \( O(\epsilon) \) and the normal component of \( \mathbb{E}(\langle X - t(x) \rangle X_{B^{ep}_{E}(t;i)}(X)) \) is of order \( O(\epsilon^{d+2}) \), so they will only contribute in the \( O(\epsilon^{d+3}) \) term.

Therefore, for \( i = 1, \ldots, d \), the first two order terms of \( \mathbb{E}(\langle X - t(x) \rangle X_{B^{ep}_{E}(t;i)}(X))\beta_i \) are
\[
\mathbb{E}(\langle X - t(x) \rangle X_{B^{ep}_{E}(t;i)}(X))\beta_i = (P(x)\mu_{d}(x, \epsilon)) (e_d J_{p,d}X_1(x)J^\top_{p,d}e_i) + \epsilon (P(x)\mu_{d}(x, \epsilon)) (e_d J_{p,d}X_1(x)S_{11}(x)J^\top_{p,d}e_i)
+ \sum_{j=1}^d (\partial_t P(x)\mu_{2\epsilon}r_j(x, \epsilon)) (e_d J_{p,d}X_1(x)J^\top_{p,d}e_i) + O(\epsilon^{d+3}).
\] (3.2.34)

By putting the above expressions together, a direct calculation shows that the normal component of
\[
\frac{\sum_{i=1}^d \mathbb{E}(\langle X - t(x) \rangle X_{B^{ep}_{E}(t;i)}(X))\beta_i\beta_i^\top}{\lambda_i + \epsilon^{d+3}}
\] is of order \( O(1) \) and the tangent component of
\[
\frac{\sum_{i=1}^d \mathbb{E}(\langle X - t(x) \rangle X_{B^{ep}_{E}(t;i)}(X))\beta_i\beta_i^\top}{\lambda_i + \epsilon^{d+3}}
\]
is of order $O(\varepsilon^{-1})$:

\[
P(x)\mu_{e_1}(x, \varepsilon)\sum_{i=1}^{d} \frac{(e_{d_i}^T J_{p,d} X_1(x) J_{p,d}^T e_i)X_1(x)J_{p,d}^T e_i}{\lambda_i + \varepsilon^{d+3}} + \varepsilon P(x)\mu_{e_d}(x, \varepsilon) \sum_{i=1}^{d} \frac{(e_{d_i}^T J_{p,d} X_1(x) S_{11}(x) J_{p,d}^T e_i)X_1(x) J_{p,d}^T e_i}{\lambda_i + \varepsilon^{d+3}} + \frac{d}{i=1} \frac{\sum_{j=1}^{d} (dp_j(x)\mu_{2e_j}(x, \varepsilon))(e_{d_i}^T J_{p,d} X_1(x) J_{p,d}^T e_i)}{\lambda_i + \varepsilon^{d+3}} + \varepsilon P(x)\mu_{e_d}(x, \varepsilon) \sum_{i=1}^{d} \frac{(e_{d_i}^T J_{p,d} X_1(x) J_{p,d}^T e_i)X_1(x) S_{11}(x) J_{p,d}^T e_i + O(\varepsilon)},
\]

where the first term is of order $\varepsilon^{-1}$, the second to the fourth terms are of order 1 since $\mu_{e_d}(x, \varepsilon)$ is of order $\varepsilon^{d+1}$, $\mu_{2e_j}(x, \varepsilon)$ is of order $\varepsilon^{d+2}$ for $i = 1, \ldots, d$ and $\lambda_i$ is of order $\varepsilon^{d+2}$ for $i = 1, \ldots, d$.

To simplify the formula of tangent component of $T_{i(x)}$, recall that from (3.2.22),

\[
X_1(x) = \begin{bmatrix}
X_1^{(1)}(x) & 0 & \cdots & 0 \\
0 & X_1^{(2)}(x) & \cdots & 0 \\
0 & 0 & \ddots & 0 \\
0 & 0 & \cdots & X_1^{(t)}(x)
\end{bmatrix},
\]

(3.36)

$1 \leq t \leq d$. Here different $X_1^{(k)}$ corresponds to different $\lambda_k$. In particular $X_1^{(k)} \in O(d_k)$ corresponds to the eigenvalue $\lambda_k$ of $C_x$, where $d_k \in \mathbb{N}$ is the multiplicity of $\lambda_k$. Note that $\sum_{i=1}^{d} (e_{j}^T J_{p,d} X_1(x) J_{p,d}^T e_i)X_1(x) J_{p,d}^T e_i$, for $j = 1, \ldots, d$, is the projection of $J_{p,d}^T e_j$ onto the space spanned by the columns of $[0, \ldots, X_1^{(k)}, \ldots, 0]^T \in \mathbb{R}^{d \times d_k}$, where $X_1^{(k)}$ corresponds to $\lambda_k = \lambda_j$. In other words, if $\lambda_j \neq \lambda_i$, then

\[
(e_{j}^T J_{p,d} X_1(x) J_{p,d}^T e_i)X_1(x) J_{p,d}^T e_i = 0.
\]

(3.37)

Thus we have

\[
\sum_{i=1}^{d} \frac{(e_{j}^T J_{p,d} X_1(x) J_{p,d}^T e_i)X_1(x) J_{p,d}^T e_i}{\lambda_i + \varepsilon^{d+3}} = \frac{J_{p,d}^T e_j}{\lambda_j + \varepsilon^{d+3}}.
\]

(3.38)

We conclude that

\[
P(x)\mu_{e_1}(x, \varepsilon)\sum_{i=1}^{d} \frac{(e_{d_i}^T J_{p,d} X_1(x) J_{p,d}^T e_i)X_1(x)J_{p,d}^T e_i}{\lambda_i + \varepsilon^{d+3}} = P(x)\frac{\mu_{e_d}(x, \varepsilon) J_{p,d}^T e_d}{\lambda_d + \varepsilon^{d+3}} - \mu_{2e_d}(x, \varepsilon) \frac{\mu_{2e_d}(x, \varepsilon) [1 + \gamma_d(x) \varepsilon + \frac{\varepsilon^{d+3}}{P(x)\mu_{2e_d}(x, \varepsilon)}]}{[1 - \gamma_d(x) \varepsilon - \frac{\varepsilon^{d+3}}{P(x)\mu_{2e_d}(x, \varepsilon)}] + O(\varepsilon),
\]

where $\frac{\mu_{e_d}(x, \varepsilon) J_{p,d}^T e_d}{\mu_{2e_d}(x, \varepsilon)}$ is of order $\varepsilon^{-1}$ since $\mu_{e_d}(x, \varepsilon)$ is of order $\varepsilon^{d+1}$ and $\mu_{2e_d}(x, \varepsilon)$ is of order $\varepsilon^{d+2}$.
Similarly,

\[
\sum_{j=1}^{d} \frac{\partial P_j(x) \mu_{2j}(x, \varepsilon)}{\lambda_j + \varepsilon^{d+3}} (e_j^T J_{p,d} X_1(x) J_{p,d}^T e_i) X_1(x) J_{p,d}^T e_i
\]

\[
= \sum_{j=1}^{d} \frac{(\partial P_j(x) \mu_{2j}(x, \varepsilon)) J_{p,d}^T e_j}{\lambda_j + \varepsilon^{d+3}}
\]

\[
= \sum_{j=1}^{d} \frac{\partial P_j(x) \mu_{2j}(x, \varepsilon) (e_j^T J_{p,d} e_i) X_1(x) J_{p,d}^T e_i}{P(x) \mu_{2j}(x, \varepsilon)} [1 + \gamma_j(x) \varepsilon] + \varepsilon^{d+3} + O(\varepsilon^{d+4})
\]

At last,

\[
\varepsilon P(x) \mu_{2j}(x, \varepsilon) \sum_{i=1}^{d} \left( \frac{e_i^T J_{p,d} X_1(x) S_{11}(x) J_{p,d}^T e_i}{\lambda_i} \right) X_1(x) J_{p,d}^T e_i
\]

\[
+ \varepsilon P(x) \mu_{2j}(x, \varepsilon) \sum_{i=1}^{d} \left( \frac{e_i^T J_{p,d} X_1(x) J_{p,d}^T e_i}{\mu_{2j}(x, \varepsilon)} \right) X_1(x) S_{11}(x) J_{p,d}^T e_i
\]

\[
= \varepsilon \mu_{2j}(x, \varepsilon) \sum_{i=1}^{d} \left( \frac{e_i^T J_{p,d} X_1(x) S_{11}(x) J_{p,d}^T e_i}{\mu_{2j}(x, \varepsilon)} \right) + \varepsilon \mu_{2j}(x, \varepsilon) \right) X_1(x) S_{11}(x) J_{p,d}^T e_i
\]

\[
= \varepsilon \mu_{2j}(x, \varepsilon) \sum_{i=1}^{d} \left( \frac{e_i^T J_{p,d} X_1(x) S_{11}(x) J_{p,d}^T e_i}{\mu_{2j}(x, \varepsilon)} \right) + \varepsilon \mu_{2j}(x, \varepsilon) \right) X_1(x) S_{11}(x) J_{p,d}^T e_i
\]

By combining above equations, the tangent component of \( \sum_{i=1}^{d} \frac{\varepsilon^{(X-1)(x) \chi^{B} \beta_i}(X)^T \beta_i}{\lambda_i + \varepsilon^{d+3}} \) can be simplified.

Next, we calculate \( \mathbb{E}[(X - \tau(x)) \chi^{B} \beta_i(X)^T \beta_i] \), for \( i = d+1, \ldots, r \).

\[
\mathbb{E}[(X - \tau(x)) \chi^{B} \beta_i(X)^T \beta_i] = P(x) \varepsilon \mu_{2j}(x, \varepsilon) \left( e_i^T J_{p,d} X_1(x) S_{12}(x) \beta_i^T \right) + \frac{P(x)}{2} \sum_{j=1}^{d} \mu_{2j} \nabla \nabla^T \beta_i \beta_i \beta_i \beta_i + O(\varepsilon^{d+3}),
\]

where both terms are of order \( O(\varepsilon^{d+2}) \).

Note that \( \lambda_i = O(\varepsilon^{d+4}) \), for \( i = d+1, \ldots, r \), so we have

\[
\mathbb{E}[(X - \tau(x)) \chi^{B} \beta_i(X)^T \beta_i] = \frac{P(x) \varepsilon \mu_{2j}(x, \varepsilon) \left( e_i^T J_{p,d} X_1(x) S_{12}(x) \beta_i^T \right) + \frac{P(x)}{2} \sum_{j=1}^{d} \mu_{2j} \nabla \nabla^T \beta_i \beta_i \beta_i \beta_i}{\varepsilon^{d+2} + O(\varepsilon^{d+4})}
\]

\[
\mathbb{E}[(X - \tau(x)) \chi^{B} \beta_i(X)^T \beta_i] = \frac{P(x) \varepsilon \mu_{2j}(x, \varepsilon)}{\varepsilon^{d+2}} \left( e_i^T J_{p,d} X_1(x) S_{12}(x) \beta_i^T \right) + \frac{P(x)}{2} \sum_{j=1}^{d} \mu_{2j} \nabla \nabla^T \beta_i \beta_i \beta_i \beta_i + O(1).
\]
A direct calculation shows that
\[
\sum_{i=d+1}^{r} \frac{\mathbb{E}[(X - t(x)) X_{B^P_i(t(x))}]^\top \beta \beta_i^\top}{\lambda_i + \varepsilon d + 3} = \sum_{i=d+1}^{r} \frac{\mathbb{E}[(X - t(x)) X_{B^P_i(t(x))}]^\top \beta_i \beta_i^\top}{\varepsilon d + 3 + O(\varepsilon d + 4)}
\]
\[
= \sum_{i=d+1}^{r} \left[ P(x) \frac{\mu_{e_i}(x, e)}{\varepsilon d + 1} (e_d J_p, d X_1(x) S_1(x) 3_1^\top p, r, d e_i) \right] X_1(x) S_1(x) 3_1^\top p, r, d e_i + O(\varepsilon),
\]
\[
+ \sum_{i=d+1}^{r} \left[ P(x) \frac{\mu_{e_i}(x, e)}{\varepsilon d + 1} (e_d J_p, d X_1(x) S_1(x) 3_1^\top p, r, d e_i) \right] J_p, d r, d X_2(x) 3_2^\top p, r, d e_i + O(1)].
\]

Note that the tangent component is of order $O(1)$ and the normal component is of order $\varepsilon^{-1}$. We sum up
\[
\sum_{i=1}^{d} \frac{\mathbb{E}[(X - t(x)) X_{B^P_i(t(x))}]^\top \beta \beta_i^\top}{\lambda_i + \varepsilon d + 3}
\]
and
\[
\sum_{i=d+1}^{r} \frac{\mathbb{E}[(X - t(x)) X_{B^P_i(t(x))}]^\top \beta_i \beta_i^\top}{\lambda_i + \varepsilon d + 3},
\]
then we have the conclusion.

Recall that
\[
K_{\text{LLE}}(x, y) = [1 - T_{t(x)}(t(y) - t(x))] X_{B^P_i(t(x))} \cap (M)(t(y)),
\]
and
\[
Q f(x) := \frac{\mathbb{E}[f(X) K(x, X)]}{\mathbb{E}[K(x, X)]},
\]
(3.2.46)

### 3.2.1 Conclusion when $\rho = 3$

We fix the regularizer in (1.2.5) to be $c = n \varepsilon d + 3$. The following Theorem describes the convergence behavior and the convergence rate of the matrix $W$

**Theorem 3.2.1.** Suppose $f \in C^3(M)$. Suppose $\varepsilon = \varepsilon(n)$ so that $\frac{\log(n)}{n^{1/2} \varepsilon^{d/2 - 1}} \to 0$ and $\varepsilon \to 0$ as $n \to \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \ldots, n$,
\[
\sum_{j=1}^{n} [W - I_{n \times n}]_{k, j} f(x_j) = Q f(x_k) - f(x_k) + O \left( \frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2 - 1}} \right)
\]

We have the following theorem describing $Q$ when $\varepsilon$ is sufficiently small.

**Theorem 3.2.2.** Let $(M, g)$ be a $d$-dimensional compact, smooth Riemannian manifold isometrically embedded in $\mathbb{R}^p$, where $M$ may have smooth boundary. Suppose $f \in C^3(M)$ and $P \in C^2(M)$.

\[
Q f(x) - f(x) = \sum_{i=1}^{d} \phi_i(x, \varepsilon) \partial^2_x f(x) + g(V(x, \varepsilon), \nabla f(x)) + O(\varepsilon^3),
\]
(3.2.47)

where $V$ is a vector field. $V$ and $\phi_i$ are defined in Notation 3.2.1. If $x \not\in M_\varepsilon$, then
\[
\phi_i(x, \varepsilon) = \frac{1}{2(d + 2)} \varepsilon^2
\]
(3.2.48)
for \( i = 1, \cdots, d \) and \( V(x, \varepsilon) = 0 \). If \( x \in M_{\varepsilon} \), then \( \phi_i(x, \varepsilon) \) is of order \( \varepsilon^2 \), for \( i = 1, \cdots, d \) and \( V(x, \varepsilon) = O(\varepsilon^2) \).

**Remark 3.2.1.** Since \( V(x, \varepsilon) = 0 \) if \( x \notin M_{\varepsilon} \); \( \frac{V(x, \varepsilon)}{\varepsilon} = O(1) \) for \( x \in M_{\varepsilon} \) and \( \frac{\phi_i(x, \varepsilon)}{\varepsilon} \) is of order \( O(1) \) for all \( x \), the operator \( \frac{Q(f)}{\varepsilon^2} \) converges to a second order differential operator when \( \varepsilon \to 0 \). Hence, we conclude that when \( c = n^{d+3} \), LLE matrix \( \frac{1}{\varepsilon}[W - I_{n \times n}] \) recovers a differential operator with free boundary condition.

### 3.2.2 Proof of Theorem 3.2.2

To simplify the proof of the main theorems, we introduce following notations,

**Notation 3.2.1.** Define functions

\[
\phi_i(x, \varepsilon) = \frac{\mu_{2i}\varepsilon(x, \varepsilon)\mu_{2i}(x, \varepsilon) - \mu_{2i+1}\varepsilon(x, \varepsilon)\mu_{2i+1}(x, \varepsilon)}{2\mu_{2i}(x, \varepsilon)\mu_0(x, \varepsilon) - 2\mu_{2i}(x, \varepsilon)^2}
\]

for \( i = 1, \cdots, d \).

Define a vector field

\[
V(x, \varepsilon) = \sum_{i=1}^{d} V_i(x, \varepsilon) \partial_i,
\]

where

\[
V_i(x, \varepsilon) = -\left[ \frac{\mu_{2i}(x, \varepsilon)\mu_{2i+1}(x, \varepsilon)\nu_{1,2}^{(0)}\top J_{p,d} \varepsilon_i + \frac{\partial p(x)}{p(x)} \mu_{2i+1}(x, \varepsilon)\mu_{2i+1}(x, \varepsilon) + \mu_{2i+1}(x, \varepsilon)\mu_{2i+2}(x, \varepsilon)\nu_{2}^{(-1)}\top \gamma(x)}{\mu_0(x, \varepsilon)\mu_{2i}(x, \varepsilon) - \mu_{2i}(x, \varepsilon)^2} \right],
\]

for \( i = 1 \cdots d - 1 \), and

\[
V_d(x, \varepsilon) = -\left[ \frac{\mu_{2d}(x, \varepsilon)^2\nu_{1,2}^{(0)}\top J_{p,d} \varepsilon_d + \frac{\partial p(x)}{p(x)} \mu_{3d}(x, \varepsilon)\mu_{3d}(x, \varepsilon) + \mu_{3d}(x, \varepsilon)\mu_{3d+1}(x, \varepsilon)\nu_{2}^{(-1)}\top \gamma(x)}{\mu_0(x, \varepsilon)\mu_{2d}(x, \varepsilon) - \mu_{2d}(x, \varepsilon)^2} \right],
\]

where \( \nu_{1,2}^{(0)} \) is defined in (3.2.26) and \( \nu_{2}^{(-1)} \) is defined in (3.2.28).

Next lemma describes the behavior of the above seemingly complicated notations. The proof follows from a direct calculation based on Lemma 3.1.2 and Lemma 3.2.2.

**Lemma 3.2.3.** If \( x \in M_{\varepsilon} \), then \( \phi_i(x, \varepsilon) \) is of order \( \varepsilon^2 \), for \( i = 1, \cdots, d \).

\[
V(x, \varepsilon) = O(\varepsilon^2)
\]

If \( x \notin M_{\varepsilon} \), then

\[
\phi_i(x, \varepsilon) = \frac{1}{2(d+2)}\varepsilon^2
\]

for \( i = 1, \cdots, d \),

\[
V(x, \varepsilon) = 0.
\]

In this proof of the main theorem, we calculate the first two order terms in \( Q(f) \). First, we are going to calculate \( \mathbb{E}[X(t(x)) X^\top(t(x))] - \mathbb{E}[X - t(x)] X^\top(t(x)) \) and show that it is dominated by the order \( \varepsilon^d \) terms.
Then we are going to calculate \( \mathbb{E}[m(X) - f(x)] \mathcal{X}_{\mathcal{H}_p}^p_{(t(x))}(X) \) and show that it is dominated by the order \( \epsilon^{d+2} \) terms. Hence their ratio is dominated by the order \( \epsilon^2 \) terms.

By Lemma 3.1.3 and Lemma 3.2.2, we have

\[
\mathbb{E}[\mathcal{X}_{\mathcal{H}_p}^p_{(t(x))}(X)] = P(x) \mu_0(x, \epsilon) + O(\epsilon^{d+1}),
\]

\[
\mathbb{E}[(X - t(x)) \mathcal{X}_{\mathcal{H}_p}^p_{(t(x))}(X)] = \left[ P(x) \mu_{e_1}(x, \epsilon) J^\top_{p,d} e_d + O(\epsilon^{d+2}), O(\epsilon^{d+2}) \right],
\]

and \( T_{t(x)} = [v_1^{(-1)} + v_2^{(0)} + v_1^{(0)} + v_2^{(-1)}] + [O(\epsilon), O(1)] \), where

\[
v_1^{(-1)} = \frac{\mu_{e_1}(x, \epsilon)}{\mu_{2e_1}(x, \epsilon)} J^\top_{p,d} e_d,
\]

\[
v_2^{(0)} = \frac{\nabla_\epsilon (P(x))}{P(x)},
\]

\[
v_1^{(1)} = \frac{\mu_{e_1}(x, \epsilon)}{\mu_{2e_1}(x, \epsilon)} J^\top_{p,d} e_d,
\]

\[
v_2^{(-1)} = \frac{\mu_{e_1}(x, \epsilon)}{\mu_{2e_1}(x, \epsilon)} J^\top_{p,d} e_d.
\]

\( v_1^{(0)} \) and \( v_2^{(-1)} \) are defined in Lemma 3.2.2. Moreover, \( v_1^{(0)} \) is of order \( O(1) \) and \( v_2^{(-1)} \) is of order \( O(\epsilon^{-1}) \). Hence,

\[
\mathbb{E}[\mathcal{X}_{\mathcal{H}_p}^p_{(t(x))}(X)] = \mathbb{E}[(X - t(x)) \mathcal{X}_{\mathcal{H}_p}^p_{(t(x))}(X)]^\top T_{t(x)}
\]

\[
= P(x) \left[ \mu_0(x, \epsilon) - \frac{\mu_{e_1}(x, \epsilon)^2}{\mu_{2e_1}(x, \epsilon)} \right] + O(\epsilon^{d+1}),
\]

\[
= P(x) \left[ \frac{\mu_0(x, \epsilon) \mu_{e_1}(x, \epsilon)}{\mu_{2e_1}(x, \epsilon)} + O(\epsilon^{d+1}) \right],
\]

where the leading term in above expression is of order \( \epsilon^d \) by Lemma 3.1.2.

Based on Lemma 3.1.3, we have

\[
\mathbb{E}[(f(X) - f(x)) \mathcal{X}_{\mathcal{H}_p}^p_{(t(x))}(X)]
\]

\[
= P(x) \partial_d f(x) \mu_{e_1}(x, \epsilon) + \sum_{i=1}^d \left[ \frac{P(x)}{2} \partial_d^2 f(x) + \partial_d f(x) \partial_i P(x) \right] \mu_{2e_1}(x, \epsilon) + O(\epsilon^{d+3}),
\]

and

\[
\mathbb{E}[(X - t(x))(f(X) - f(x)) \mathcal{X}_{\mathcal{H}_p}^p_{(t(x))}(X)] = [v_1, v_2],
\]

where

\[
v_1 = P(x) \sum_{i=1}^d \left( \partial_i f(x) \mu_{2e_1}(x, \epsilon) \right) J^\top_{p,d} e_i
\]

\[
+ \sum_{i=1}^{d-1} \left[ \partial_d f(x) \partial_i P(x) + \partial_d f(x) \partial_i P(x) + P(x) \partial_i^2 f(x) \right] \mu_{2e_1+e_d}(x, \epsilon) J^\top_{p,d} e_i
\]

\[
+ \sum_{i=1}^{d-1} \left( \partial_d f(x) \partial_i P(x) + \frac{P(x)}{2} \partial_i^2 f(x) \right) \mu_{2e_1+e_d}(x, \epsilon) J^\top_{p,d} e_i + O(\epsilon^{d+4}),
\]

\[
v_2 = P(x) \sum_{i=1}^d \partial_i f(x) \partial_{d} f(x) \mu_{2e_1+e_d}(x, \epsilon) + \frac{P(x)}{2} \partial_i^2 f(x) \partial_{d} f(x) \mu_{3e_1}(x, \epsilon) + O(\epsilon^{d+4}).
\]
Therefore, we have

\[
\mathbb{E}[(X - t(x))(f(X) - f(x))\mathcal{X}_{B^{p}_{\mathcal{L}}(\{t(x)\})(X)]^\top T_{t(x)}
\]

\[
= P(x) \sum_{i=1}^{d} (\partial_i f(x) \mu_{2\epsilon_i}(x, \epsilon)) v_1^{(-1)^\top} J_{p,d}e_i + P(x) \sum_{i=1}^{d} (\partial_i f(x) \mu_{2\epsilon_i}(x, \epsilon)) v_1^{(0)^\top} J_{p,d}e_i
\]

\[
+ P(x) \sum_{i=1}^{d} (\partial_i f(x) \mu_{2\epsilon_i}(x, \epsilon)) v_1^{(0)^\top} J_{p,d}e_i + \sum_{i=1}^{d-1} [\partial_i f(x) \partial_d P(x) + \partial_d f(x) \partial_i P(x) + P(x) \partial_i^2 f(x)] \mu_{2\epsilon_i+\epsilon_d}(x, \epsilon) v_1^{(-1)^\top} J_{p,d}e_i
\]

\[
+ \sum_{i=1}^{d-1} \partial_i f(x) \partial_d P(x) + \frac{P(x)}{2} \partial_d^2 f(x) \mu_{2\epsilon_i+\epsilon_d}(x, \epsilon) v_1^{(-1)^\top} J_{p,d}e_d
\]

\[
+ P(x) \sum_{i=1}^{d-1} \partial_i f(x) \mu_{2\epsilon_i+\epsilon_d}(x, \epsilon) v_2^{(-1)^\top} \mathcal{N}_{d}(x) + \frac{P(x)}{2} \partial_d f(x) \mu_{3\epsilon_d}(x, \epsilon) v_2^{(-1)^\top} \mathcal{N}_{d}(x) + O(\epsilon^{d+3})
\]

Note that by Lemma 3.1.2, the first term is of order $\epsilon^{d+1}$ and the second to seventh terms are of order $\epsilon^{d+2}$. Furthermore, we can simplify the first and the second term as:

\[
P(x) \sum_{i=1}^{d} (\partial_i f(x) \mu_{2\epsilon_i}(x, \epsilon)) v_1^{(-1)^\top} J_{p,d}e_i = P(x) \partial_d f(x) \mu_{\epsilon_d}(x, \epsilon)
\]

\[
P(x) \sum_{i=1}^{d} (\partial_i f(x) \mu_{2\epsilon_i}(x, \epsilon)) v_1^{(0)^\top} J_{p,d}e_i = \sum_{i=1}^{d} \partial_i f(x) \partial_d P(x) \mu_{2\epsilon_i}(x, \epsilon)
\]

Next we calculate $\mathbb{E}[(f(X) - f(x))\mathcal{X}_{B^{p}_{\mathcal{L}}(\{t(x)\})(X)] - \mathbb{E}[(X - t(x))(f(X) - f(x))\mathcal{X}_{B^{p}_{\mathcal{L}}(\{t(x)\})(X)]^\top T_{t(x)}$. Clearly, the common terms, $P(x) \partial_d f(x) \mu_{\epsilon_d}(x, \epsilon)$ and $\sum_{i=1}^{d-1} \partial_i f(x) \partial_d P(x) \mu_{2\epsilon_i}(x, \epsilon)$, are canceled, and hence only terms of order $\epsilon^{d+2}$ are left in the difference; that is, we have

\[
\mathbb{E}[(f(X) - f(x))\mathcal{X}_{B^{p}_{\mathcal{L}}(\{t(x)\})(X)] - \mathbb{E}[(X - t(x))(f(X) - f(x))\mathcal{X}_{B^{p}_{\mathcal{L}}(\{t(x)\})(X)]^\top T_{t(x)}
\]

\[
= \frac{1}{2} \sum_{i=1}^{d} \partial_i^2 f(x) \mu_{2\epsilon_i}(x, \epsilon) - \frac{1}{2} \sum_{i=1}^{d} (\partial_i f(x) \partial_d P(x) + \partial_d f(x) \partial_i P(x) + P(x) \partial_i^2 f(x)] \mu_{2\epsilon_i+\epsilon_d}(x, \epsilon) v_1^{(-1)^\top} J_{p,d}e_i
\]

\[
- \sum_{i=1}^{d-1} \partial_i f(x) \partial_d P(x) + \frac{P(x)}{2} \partial_d^2 f(x) \mu_{2\epsilon_i+\epsilon_d}(x, \epsilon) v_1^{(-1)^\top} J_{p,d}e_d
\]

\[
- P(x) \sum_{i=1}^{d-1} \partial_i f(x) \mu_{2\epsilon_i+\epsilon_d}(x, \epsilon) v_2^{(-1)^\top} \mathcal{N}_{d}(x) - \frac{P(x)}{2} \partial_d f(x) \mu_{3\epsilon_d}(x, \epsilon) v_2^{(-1)^\top} \mathcal{N}_{d}(x) + O(\epsilon^{d+3})
\]

Therefore, the ratio

\[
\frac{\mathbb{E}[(f(X) - f(x))\mathcal{X}_{B^{p}_{\mathcal{L}}(\{t(x)\})(X)] - \mathbb{E}[(X - t(x))(f(X) - f(x))\mathcal{X}_{B^{p}_{\mathcal{L}}(\{t(x)\})(X)]^\top T_{t(x)}}{\mathbb{E}[(f(X) - f(x))\mathcal{X}_{B^{p}_{\mathcal{L}}(\{t(x)\})(X)] - \mathbb{E}[(X - t(x))(f(X) - f(x))\mathcal{X}_{B^{p}_{\mathcal{L}}(\{t(x)\})(X)]^\top T_{t(x)}}
\]

(3.2.67)
could be expanded to

\[
\frac{1}{2} \sum_{i=1}^{d} \frac{\partial_{ij}^2 f(x)}{\mu_2(x, e) \mu_{2e_d}(x, e)} - \frac{d}{d} \frac{\partial_i f(x)}{\mu_0(x, e) \mu_{2e_d}(x, e) - \mu_{e_d}(x, e)^2} + \frac{d}{d} \frac{\partial_j f(x)}{\mu_0(x, e) \mu_{2e_d}(x, e) - \mu_{e_d}(x, e)^2} v_1^{(0\top)} J_{p,d}^e d_i
\]

\[
- \frac{d}{d} \frac{\partial f(x)}{P(x)} + \frac{1}{2} \frac{\partial_{i}^2 f(x)}{P(x)} - \frac{d}{d} \frac{\partial f(x)}{\mu_0(x, e) \mu_{2e_d}(x, e) - \mu_{e_d}(x, e)^2} v_1^{(0\top)} J_{p,d}^e d_i
\]

\[
- \frac{d}{d} \frac{\partial f(x)}{P(x)} + \frac{1}{2} \frac{\partial_{i}^2 f(x)}{P(x)} - \frac{d}{d} \frac{\partial f(x)}{\mu_0(x, e) \mu_{2e_d}(x, e) - \mu_{e_d}(x, e)^2} v_1^{(0\top)} J_{p,d}^e d_i
\]

\[
- \frac{d}{d} \frac{\partial f(x)}{\mu_0(x, e) \mu_{2e_d}(x, e) - \mu_{e_d}(x, e)^2} v_1^{(0\top)} J_{p,d}^e d_i
\]

Note that \( v_1^{(0\top)} J_{p,d}^e d_i = \frac{\mu_{e_d}(x, e)}{\mu_{2e_d}(x, e)} \) if \( i = d \), and it is 0 otherwise. Hence, above expression can be further simplified into

\[
E[(f(X) - f(x)) X_{BP}^{(i)(x)}(X)] - E[(X - x)(f(X) - f(x)) X_{BP}^{(i)(x)}(X)]^\top T_{t(x)}
\]

\[
= \sum_{i=1}^{d} \frac{\partial_{ii}^2 f(x)}{2\mu_0(x, e) \mu_{2e_d}(x, e) - 2\mu_{e_d}(x, e)^2} + \frac{d}{d} \frac{\partial f(x)}{P(x)} - \frac{d}{d} \frac{\partial f(x)}{\mu_0(x, e) \mu_{2e_d}(x, e) - \mu_{e_d}(x, e)^2} v_1^{(0\top)} J_{p,d}^e d_i
\]

\[
- \frac{d}{d} \frac{\partial f(x)}{\mu_0(x, e) \mu_{2e_d}(x, e) - \mu_{e_d}(x, e)^2} v_1^{(0\top)} J_{p,d}^e d_i
\]

\[
= \sum_{i=1}^{d} \phi_i(x, e) \partial_{ii}^2 f(x) + g(V(x, e), \nabla f(x)) + O(e^3)
\]

The conclusion of the theorem follows from applying Lemma 3.2.3 to simplify the expressions.

### 3.2.3 Proof of Theorem 3.2.1

The proof is similar to the proof of Theorem 2.4.1.

For each \( x_k \), denote \( f = (f(x_1), f(x_2), \ldots, f(x_n))^\top \in \mathbb{R}^n \). By a direct expansion, we have

\[
\sum_{j=1}^{N} [W - I_{n \times n}] x_j f(x_j) = \frac{1_N^f - 1_N^f \sum_{n}^{N} U_{n}^f I_{P_{n}}(A_n + n e^{d+3}I_{P_{n}})^{-1} U_{n}^f G_n f}{N - 1_N^f \sum_{n}^{N} U_{n}^f I_{P_{n}}(A_n + n e^{d+3}I_{P_{n}})^{-1} U_{n}^f G_n I_{N}} - f(x_k),
\]

which can be rewritten as

\[
\frac{1}{n} \sum_{j=1}^{N} (f(x_k) - f(x_j)) \left[ \frac{1}{n} \sum_{j=1}^{N} (x_{k,j} - x_j) \right] U_{n}^f I_{P_{n}}(A_n + n e^{d+3}I_{P_{n}})^{-1} U_{n}^f G_n f
\]

\[
\left( \frac{1}{n} \sum_{j=1}^{N} (x_{k,j} - x_j) \right) U_{n}^f I_{P_{n}}(A_n + n e^{d+3}I_{P_{n}})^{-1} U_{n}^f G_n I_{N}
\]

The goal is to relate the finite sum quantity (3.2.70) to \( Qf(x_k) = \frac{g_1}{g_2} \).
where

\[ g_1 = \mathbb{E}\left[ \frac{1}{\varepsilon d} X_{\mathcal{B}_p^p(t(x_k))}(X)(f(X) - f(x_k)) \right] - \mathbb{E}\left[ \frac{1}{\varepsilon d} (X - x_k) X_{\mathcal{B}_p^p(t(x_k))}(X) \right]^\top \] (3.2.71)

\[ (U_{I,p} - \varepsilon^2 I_p) U^{-1} \mathbb{E}\left[ \frac{1}{\varepsilon d} (X - x_k) X_{\mathcal{B}_p^p(t(x_k))}(X)(f(X) - f(x_k)) \right] \]

and

\[ g_2 = \mathbb{E}\left[ \frac{1}{\varepsilon d} X_{\mathcal{B}_p^p(t(x_k))}(X) \right] - \mathbb{E}\left[ \frac{1}{\varepsilon d} (X - x_k) X_{\mathcal{B}_p^p(t(x_k))}(X) \right]^\top \] (3.2.72)

\[ (U_{I,p} - \varepsilon^2 I_p) U^{-1} \mathbb{E}\left[ \frac{1}{\varepsilon d} (X - x_k) X_{\mathcal{B}_p^p(t(x_k))}(X) \right]. \]

Moreover, we relate

\[ B_{kk} f(x_k) = \frac{\|G_d \mathbf{1}_N\|_{\mathcal{B}_p^p}^2}{N^2 \varepsilon^2} f(x_k) = \frac{1}{\varepsilon} \left( \frac{1}{n^2} \sum_{j=1}^N (x_{k,j} - x_k) \right)^2 f(x_k) \] (3.2.73)

to

\[ B_{e}(x_k) f(x_k) = \frac{1}{\varepsilon} \left( \frac{\mathbb{E}\left[ \frac{1}{\varepsilon d} (X - t(x_k)) X_{\mathcal{B}_p^p(t(x_k))}(X) \right]_{\mathbb{B}_p}}{\mathbb{E}\left[ \frac{1}{\varepsilon d} X_{\mathcal{B}_p^p(t(x_k))}(X) \right]} \right) f(x_k). \] (3.2.74)

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

\[ \frac{1}{n^2 d} \sum_{j=1}^N \mathbf{1} \] (3.2.75)

\[ \frac{1}{n^2 d} \sum_{j=1}^N (f(x_{k,j}) - f(x_k)) \] (3.2.76)

\[ \frac{1}{n^2 d} \sum_{j=1}^N (x_{k,j} - x_k) \] (3.2.77)

\[ \frac{1}{n^2 d} \sum_{j=1}^N (x_{k,j} - x_k)(f(x_{k,j}) - f(x_k)) \] (3.2.78)

as a function of \( n \) and \( \varepsilon \) by the Bernstein type inequality. Here, we put \( \varepsilon^{-d} \) in front of each term to normalize the kernel so that the computation is consistent with the existing literature, like [17, 57].

The size of the fluctuation of these terms are controlled in the following Lemmas. The term (3.2.75) is the usual kernel density estimation, so we have the following lemma.

**Lemma 3.2.4.** Suppose \( \varepsilon = \varepsilon(n) \) so that \( \sqrt{\log(n)/n^{d+2/3}} \to 0 \) and \( \varepsilon \to 0 \) as \( n \to \infty \). We have with probability greater than \( 1 - n^{-2} \) that for all \( k = 1, \ldots, n, \)

\[ \left| \frac{1}{n^2 d} \sum_{j=1}^N \mathbf{1} - \mathbb{E}\left[ \frac{1}{\varepsilon d} X_{\mathcal{B}_p^p(t(x_k))}(X) \right] \right| = O\left( \frac{\sqrt{\log(n)}}{n^{1/2} d^{1/2}} \right). \] (3.2.79)

Denote \( \Omega_0 \) to be the event space that above Lemma is satisfied. The behavior of (3.2.76) is summarized in the following Lemma.

**Lemma 3.2.5.** Suppose \( \varepsilon = \varepsilon(n) \) so that \( \sqrt{\log(n)/n^{d+2/3}} \to 0 \) and \( \varepsilon \to 0 \) as \( n \to \infty \). We have with probability greater
than $1 - n^{-2}$ that for all $k = 1, \ldots, n$,

$$
\left| \frac{1}{n \epsilon^d} \sum_{j=1}^{N} (f(x_k) - f(x_j)) - \frac{1}{\epsilon^d} (f(X) - f(x_k)) \mathcal{X}_{B^p}(x_j)(X) \right| = O \left( \frac{\sqrt{\log(n)}}{n^{1/2} \epsilon^{d/2 - 1}} \right) .
$$

(3.2.80)

**Proof.** By denoting

$$
F_{1,j} = \frac{1}{\epsilon^d} (f(x_j) - f(x_k)) \mathcal{X}_{B^p}(x_k)(X),
$$

we have

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

(3.2.82)

Define a random variable

$$
F_1 := \frac{1}{\epsilon^d} (f(X) - f(x_k)) \mathcal{X}_{B^p}(x_k)(X).
$$

(3.2.83)

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

$$
\frac{1}{n} \sum_{j \neq k, j=1}^{n} F_{1,j} = \frac{n-1}{n} \left[ \frac{1}{n-1} \sum_{j \neq k, j=1}^{n} F_{1,j} \right].
$$

(3.2.84)

Since $\frac{n-1}{n} \to 1$ as $n \to \infty$, 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 $\frac{1}{n} \sum_{j=1, j \neq k}^{n} F_{1,j}$. We have by Lemma 3.1.2 and Lemma 3.1.3,

$$
\mathbb{E}[F_1] = O(\varepsilon) \quad \text{if } x \in M_e
$$

(3.2.85)

$$
\mathbb{E}[F_1] = O(\varepsilon^2) \quad \text{if } x \notin M_e
$$

(3.2.86)

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

(3.2.87)

By Lemma 3.1.2, $\frac{|y_t|}{2d(d+2)} \varepsilon^{-d+2} + O(\varepsilon^{-d+3}) \leq \mu_{2\varepsilon^2}(x_k, \varepsilon) \varepsilon^{-2d} \leq \frac{|y_t|}{2d(d+2)} \varepsilon^{-d+2}$, therefore, in any case,

$$
\sigma_1^2 := \text{Var}(F_1) \leq \frac{|y_t|}{d(d+2)} \varepsilon^{-d+2} + O(\varepsilon^{-d+3}).
$$

(3.2.88)

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

$$
c_1 = 2\|f\|_{L^\infty} \varepsilon^{-d}
$$

(3.2.89)

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

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

(3.2.90)

where $\eta_1 > 0$.

Note that $\mathbb{E}[F_1] = O(\varepsilon)$, if $x_k \in M_e$ and $\mathbb{E}[F_1] = O(\varepsilon^2)$, if $x_k \notin M_e$. Hence, we assume $\eta_1 = O(\varepsilon^{2+s})$, where $s > 0$. Then $c_1 \eta_1 = O(\varepsilon^{-d+2+s})$. If $\varepsilon$ 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 $P$. We have,

$$
\frac{n\eta_1^2}{2\sigma_1^2 + \frac{2}{3}c_1\eta_1} \geq \frac{n\eta_1^2 \varepsilon^{d-2}}{C}.
$$

(3.2.91)
Suppose \( n \) is chosen large enough so that

\[
\frac{n\eta^2}{C} \geq 3\log(n);
\]  

(3.2.92)

that is, the deviation from the mean is set to

\[
\eta \geq O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-1}}\right),
\]  

(3.2.93)

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

\[
\exp\left(-\frac{n\eta^2}{2\sigma^2 + \frac{\varepsilon}{2}^2 \eta_1}\right) \leq \exp\left(-\frac{n\eta^2}{C}\right) = \exp(-3\log(n)) = 1/n^3.
\]

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

\[
\Pr\left\{ \frac{1}{n-1} \sum_{j \neq k, j=1}^{n} (F_{1,j} - \mathbb{E}[F_1]) > \eta_1 \mid k = 1, \ldots, n \right\} \leq ne^{-\frac{n\eta^2}{2\sigma^2 + \frac{\varepsilon}{2}^2 \eta_1}} \leq 1/n^2.
\]  

(3.2.94)

Denote \( \Omega_1 \) 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 \( i = 1, \ldots, n \), where \( \eta_1 \) is chosen in (3.2.93) is satisfied.

**Lemma 3.2.6.** Suppose \( \varepsilon = \varepsilon(n) \) so that \( \frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-1}} \to 0 \) and \( \varepsilon \to 0 \) as \( n \to \infty \). We have with probability greater than \( 1-n^{-2} \) that for all \( k = 1, \ldots, n \).

\[
e_i^\top \left[ - \frac{n e^{d}}{n^{e^{d}}} \sum_{j=1}^{N} (x_{k,j} - x_k) - \frac{1}{\varepsilon^d} \mathbb{E}[(X-x_k)\mathcal{B}_\varepsilon^{d,p}(x_k)](X) \right] = O\left( \frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-1}} \right),
\]  

(3.2.95)

where \( i = 1, \ldots, d \). And

\[
e_i^\top \left[ - \frac{n e^{d}}{n^{e^{d}}} \sum_{j=1}^{N} (x_{k,j} - x_k) - \frac{1}{\varepsilon^d} \mathbb{E}[(X-x_k)\mathcal{B}_\varepsilon^{d,p}(x_k)](X) \right] = O\left( \frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2-2}} \right),
\]  

(3.2.96)

where \( i = d+1, \ldots, p \).

**Proof.** Fix \( x_k \). By denoting

\[
\frac{1}{n e^{d}} \sum_{j=1}^{N} (x_{k,j} - x_k) = \frac{1}{n} \sum_{j \neq k, j=1}^{n} \sum_{\ell=1}^{p} F_{2,\ell,j} e^\ell.
\]  

(3.2.97)

where

\[
F_{2,\ell,j} := \frac{1}{\varepsilon^d} e^\ell (x_j - x_k) \mathcal{B}_\varepsilon^{d,p}(x_k),
\]  

(3.2.98)

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

\[
F_{2,\ell} := \frac{1}{\varepsilon^d} e^\ell (X-x_k) \mathcal{B}_\varepsilon^{d,p}(x_k).
\]  

(3.2.99)
Similarly, we can focus on analyzing \( \frac{1}{n-1} \sum_{j=1}^{n} F_{2,\ell, j} \) since \( \frac{2-1}{n} \to 1 \) as \( n \to \infty \). By Lemma 3.1.3 we have

\[
\mathbb{E}[F_{2,\ell}] = \begin{cases} 
(P(x)\mu_{\varepsilon}(x, \varepsilon)\varepsilon^{-d})e_{\ell}^T e_{\ell} + \sum_{i=1}^{d} (\hat{g}_{i}(x, \varepsilon)\varepsilon^{-d})e_{\ell}^T e_{\ell} + O(\varepsilon^{d+3}) & \text{when } \ell = 1, \ldots, d \\
\frac{P(x)}{2}\varepsilon^{-d}e_{\ell}^T e_{\ell} + \sum_{i=1}^{d} \gamma_{i}(x, \varepsilon) + O(\varepsilon^{d+3}) & \text{when } \ell = d + 1, \ldots, p.
\end{cases}
\]

In other words, by Lemma 3.1.2, for \( \ell = 1, \ldots, d \) 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 \( \ell = d + 1, \ldots, p \).

By (3.2.10) we have, for \( \ell = 1, \ldots, d \)

\[
\mathbb{E}[F_{2,\ell}^2] \leq C_\ell \varepsilon^{-d+2} + O(\varepsilon^{-d+3}),
\]

and \( C_\ell \) depends on \( \|P\|_{L^2} \). For \( \ell = d + 1, \ldots, p \),

\[
\mathbb{E}[F_{2,\ell}^2] \leq C_\ell \varepsilon^{-d+4} + O(\varepsilon^{-d+5}),
\]

and \( C_\ell \) depends on \( \|P\|_{L^2} \) and second fundamental form of \( M \).

Thus, we conclude that

\[
\sigma_{2,\ell}^2 \leq C_\ell \varepsilon^{-d+2} + O(\varepsilon^{-d+3}) \quad \text{when } \ell = 1, \ldots, d \quad (3.2.102)
\]

\[
\sigma_{2,\ell}^2 \leq C_\ell \varepsilon^{-d+4} + O(\varepsilon^{-d+5}) \quad \text{when } \ell = d + 1, \ldots, p. \quad (3.2.103)
\]

Note that for \( \ell = d + 1, \ldots, p \), the variance is of higher order than that of \( \ell = 1, \ldots, d \).

With the above bounds, we could apply the large deviation theory. For \( \ell = 1, \ldots, d \), 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 \( c_{2,\ell} \eta_{2,\ell} = O(\varepsilon^{2d+3+s}) \). If \( \varepsilon \) is small enough, \( 2\sigma_{2,\ell}^2 + \frac{3}{2} c_{2,\ell} \eta_{2,\ell} \leq C \varepsilon^{-d+2} \) for some constant \( C \) which depends on \( P \) and manifold \( M \). We have,

\[
\frac{n\eta_{2,\ell}^2}{2\sigma_{2,\ell}^2 + \frac{3}{2} c_{2,\ell} \eta_{2,\ell}} \geq \frac{n\eta_{2,\ell}^2 \varepsilon^{d-2}}{C}. \quad (3.2.104)
\]

Suppose \( n \) is chosen large enough so that

\[
\frac{n\eta_{2,\ell}^2 \varepsilon^{d-2}}{C} \geq 3\log(n); \quad (3.2.105)
\]

that is, the deviation from the mean is set to

\[
\eta_{2,\ell} \geq O\left(\sqrt{\frac{\log(n)}{n^{1/2} \varepsilon^{d/2-1}}}\right), \quad (3.2.106)
\]

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

Thus, when \( \varepsilon \) is sufficiently smaller and \( n \) is sufficiently large, the exponent in Bernstein’s inequality

\[
\Pr\left\{ \frac{1}{n-1} \sum_{j \neq k, j=1}^{n} (F_{2,\ell, j} - \mathbb{E}[F_{2,\ell}]) > \eta_{2,\ell} \right\} \leq \exp\left( -\frac{n\eta_{2,\ell}^2}{2\sigma_{2,\ell}^2 + \frac{3}{2} c_{2,\ell} \eta_{2,\ell}} \right) \leq \frac{1}{n^3}. \quad (3.2.107)
\]
By a simple union bound, for $\ell = 1, \ldots, d$, we have
\[
\Pr \left\{ \frac{1}{n} \sum_{j \neq k, j=1}^{n} F_{2,\ell,j} - \mathbb{E}[F_{2,\ell}] \geq \eta_{2,\ell} \mid k = 1, \ldots, n \right\} \leq 1/n^2.
\]

For $\ell = d + 1, \ldots, p$, 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 $\ell = d + 1, \ldots, p$, we assume $\eta_{2,\ell} = O(\varepsilon^{3+s})$, where $s > 0$. Then $c_{2,\ell}^2 \eta_{2,\ell} = O(\varepsilon^{d+4+s})$. If $\varepsilon$ is small enough, $2\sigma_{2,\ell}^2 + \frac{1}{2} c_{2,\ell}^2 \eta_{2,\ell} \leq C \varepsilon^{-d+4}$ for some constant $C$ which depends on $M$ and $P$. We have,
\[
\frac{n\eta_{2,\ell}^2 \varepsilon^{d-4}}{2\sigma_{2,\ell}^2 + \frac{1}{2} c_{2,\ell}^2 \eta_{2,\ell}} \geq \frac{n\eta_{2,\ell}^2 \varepsilon^{d-4}}{C}.
\]

Suppose $n$ is chosen large enough so that
\[
\frac{n\eta_{2,\ell}^2 \varepsilon^{d-4}}{C} = 3\log(n);
\]
that is, the deviation from the mean is set to
\[
\eta_{2,\ell} = O\left(\frac{\sqrt[3]{\log(n)}}{n^{\frac{1}{3}} \varepsilon^{d/3-2}}\right),
\]
Note that by the assumption that $\beta_1 = O(\varepsilon^{3+s})$, we know that $\eta_{2,\ell} / \varepsilon^3 = \frac{\sqrt[3]{\log(n)}}{n^{\frac{1}{3}} \varepsilon^{d/3}} \to 0$.

By a similar argument, for $\ell = d + 1, \ldots, p$, we have
\[
\Pr \left\{ \frac{1}{n} \sum_{j \neq k, j=1}^{n} F_{2,\ell,j} - \mathbb{E}[F_{2,\ell}] \geq \eta_{2,\ell} \mid k = 1, \ldots, n \right\} \leq 1/n^2.
\]

Denote $\Omega_2$ to be the event space that the deviation $\frac{1}{n} \sum_{j \neq k, j=1}^{n} F_{2,\ell,j} - \mathbb{E}[F_{2,\ell}] \leq \eta_{2,\ell}$ for all $\ell = 1, \ldots, p$ and $k = 1, \ldots, n$, where $\eta_{2,\ell}$ are chosen in (3.2.106) and (3.2.110).

Next Lemma summarizes behavior of (3.2.78) and can be proved similarly as Lemma 3.2.6.

**Lemma 3.2.7.** Suppose $\varepsilon = \varepsilon(n)$ so that $\frac{\sqrt[3]{\log(n)}}{n^{\frac{1}{3}} \varepsilon^{d/3}} \to 0$ and $\varepsilon \to 0$ as $n \to \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \ldots, n$,
\[
ei \left[ \frac{1}{n^d} \sum_{j=1}^{N} (x_{k,j} - x_k)(f(x_{k,j}) - f(x_k)) - \frac{1}{\varepsilon d} (X - x_k)(f(X) - f(x_k)) \chi_{\beta_1 \rho_1^p}(x_k)(X) \right] = O\left(\frac{\sqrt[3]{\log(n)}}{n^{\frac{1}{3}} \varepsilon^{d/3}}\right),
\]
where $i = 1, \ldots, d$. And
\[
ei \left[ \frac{1}{n^d} \sum_{j=1}^{N} (x_{k,j} - x_k)(f(x_{k,j}) - f(x_k)) - \frac{1}{\varepsilon d} (X - x_k)(f(X) - f(x_k)) \chi_{\beta_2 \rho_2^p}(x_k)(X) \right] = O\left(\frac{\sqrt[3]{\log(n)}}{n^{\frac{1}{3}} \varepsilon^{d/3}}\right),
\]
where $i = d + 1, \ldots, p$.

Denote $\Omega_3$ to be the event space that Lemma 3.2.7 is satisfied.
In the next two lemmas, we describe the behavior of $\frac{1}{n^{d}} G_{n}G_{n}^{\top}$. The proofs are the same as Lemma 2.4.4 with $\rho = 3$.

**Lemma 3.2.8.** Suppose $\varepsilon = \varepsilon(n)$ so that $\frac{\sqrt{\log(n)}}{n^{1/2}d^{2/3+1}} \rightarrow 0$ and $\varepsilon \rightarrow 0$ as $n \rightarrow \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \ldots, n$,

$$|e_{i}^{\top}(\frac{1}{n}\varepsilon d G_{n}G_{n}^{\top} - \frac{1}{\varepsilon} C_{x_{k}})e_{j}| = O(\frac{\sqrt{\log(n)}}{n^{1/2}d^{2/3-2}}),$$

(3.2.113)

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

$$|e_{i}^{\top}(\frac{1}{n}\varepsilon d G_{n}G_{n}^{\top} - \frac{1}{\varepsilon} C_{x_{k}})e_{j}| = O(\frac{\sqrt{\log(n)}}{n^{1/2}d^{2/3-4}}),$$

(3.2.114)

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

$$|e_{i}^{\top}(\frac{1}{n}\varepsilon d G_{n}G_{n}^{\top} - \frac{1}{\varepsilon} C_{x_{k}})e_{j}| = O(\frac{\sqrt{\log(n)}}{n^{1/2}d^{2/3-3}}),$$

(3.2.115)

otherwise.

**Lemma 3.2.9.** $r_{n} \leq r$ and $r_{n}$ is a non decreasing function of $n$. If $n$ is large enough, $r_{n} = r$. Suppose $\varepsilon = \varepsilon(n)$ so that $\frac{\sqrt{\log(n)}}{n^{1/2}d^{2/3+1}} \rightarrow 0$ and $\varepsilon \rightarrow 0$ as $n \rightarrow \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \ldots, n$,

$$|e_{i}^{\top}[I_{p,r_{n}}(\Lambda_{d}^{\varepsilon} + \varepsilon^{3} I_{p \times p})^{-1} - I_{p,r}(\Lambda_{d}^{\varepsilon} + \varepsilon^{3} I_{p \times p})^{-1}]e_{i}| = O(\frac{\sqrt{\log(n)}}{n^{1/2}d^{2/3-2}}) \quad \text{for } i = 1, \ldots, r$$

(3.2.116)

$$U_{n} = U\Theta + \frac{\sqrt{\log(n)}}{n^{1/2}d^{2/3-2}} U\Theta S + O(\frac{\log(n)}{n^{d-4}}),$$

(3.2.117)

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

Denote $\Omega_{4}$ to be the event space that Lemma 3.2.9 is satisfied.

In the proofs of Lemma 3.2.2 and Theorem 3.2.2, we need the order $\varepsilon^{d+3}$ terms of the eigenvalues $\{\lambda_{i}\}$ of $C_{x}$ for $i = 1, \ldots, d$ and we need the order $\varepsilon$ term of the eigenvectors $\{\beta_{i}\}$ of $C_{x}$ for $i = 1, \ldots, p$. We also use the fact that $\{\lambda_{i}\}$ of $C_{x}$ for $i = d + 1, \ldots, p$ are of order $O(\varepsilon^{d+3})$, so that we can calculate the leading terms (order $\varepsilon^{2}$) of $Qf(x)$ for all $x \in M$. Since $\frac{\sqrt{\log(n)}}{n^{1/2}d^{2/3+1}} \rightarrow 0$, the above two lemmas imply that the differences between the first $d$ eigenvalues of $\frac{1}{n}\varepsilon d G_{n}G_{n}^{\top}$ and $\frac{1}{\varepsilon} C_{x_{k}}$ are less than $O(\varepsilon^{2})$. The differences between the rest of the eigenvalues of $\frac{1}{n^{d}} G_{n}G_{n}^{\top}$ and $\frac{1}{\varepsilon} C_{x_{k}}$ are less than $O(\varepsilon^{4})$. In other words, we can make sure that the rest of the eigenvalues of $\frac{1}{n^{d}} G_{n}G_{n}^{\top}$ are of order $O(\varepsilon^{4})$. Moreover $U_{n}$ and $U\Theta$ differ by an order $O(\varepsilon^{3})$ matrix. Consequently, in the following proof, we can show that the deviation between $\sum_{j=1}^{N}[W-I_{n \times n}1]f(x_{j})$ and $Qf(x_{k})$ is less than $\varepsilon^{2}$ for all $x_{k}$.

**Proof of Theorem 3.2.1.** Denote $\Omega := \cap_{i=0, \ldots, 4}\Omega_{i}$. By a direct union bound, the probability of the event space $\Omega$ is greater than $1 - n^{-2}$. Below, all arguments are conditional on $\Omega$. Based on previous lemmas, we have, for $k = 1, \ldots, n$,

$$\frac{1}{n^{d}} \sum_{j=1}^{N} 1 = \mathbb{E} \frac{1}{\varepsilon d} x_{l^{\varepsilon p}(x_{k})}^{\varepsilon d} + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}d^{2/3}}\right),$$

(3.2.118)

$$\frac{1}{n^{d}} \sum_{j=1}^{N} (f(x_{j,k}) - f(x_{k})) = \mathbb{E} \frac{1}{\varepsilon d} (f(X) - f(x_{k})) x_{l^{\varepsilon p}(x_{k})}^{\varepsilon d} + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}d^{2/3-1}}\right),$$

(3.2.119)
\[
\frac{1}{n^{d}} \sum_{j=1}^{N} (x_{k,j} - x_{k}) = \mathbb{E} \frac{1}{\varepsilon^{d}} (X - x_{k}) X_{B^{d}} (x_{k}) + \mathcal{E}_1,
\]

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

\[
\frac{1}{n^{d}} \sum_{j=1}^{N} (x_{k,j} - x_{k})(f(x_{k,j}) - f(x_{k})) = \mathbb{E} \frac{1}{\varepsilon^{d}} (X - x_{k})(f(X) - f(x_{k})) X_{B^{d}} (x_{k}) + \mathcal{E}_2,
\]

where \( \mathcal{E}_2 \in \mathbb{R}^{p} \). \( e_i^\top \mathcal{E}_2 = O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2 - 1}} \right) \) for \( i = 1, \ldots, d \), and \( e_i^\top \mathcal{E}_2 = O\left( \frac{\sqrt{\log(n)}}{n^{1/2} \varepsilon^{d/2 - 1}} \right) \) for \( i = d + 1, \ldots, p \).

Define a \( p \times p \) matrix

\[
\mathcal{E}_3 = \frac{1}{n^{d}} \sum_{j=1}^{N} (x_{k,j} - x_{k})^\top U_{n, l_p, r_n} \frac{\Lambda}{n^{d}} + \varepsilon^{3} I_{p \times p})^{-1} U_n + \mathcal{E}_3 + \text{higher order terms},
\]

\[
U_n l_p, r_n \left( \frac{\Lambda}{n^{d}} + \varepsilon^{3} I_{p \times p} \right)^{-1} U_n^{-1} - U_l p, r \left( \frac{\Lambda}{n^{d}} + \varepsilon^{3} I_{p \times p} \right)^{-1} U_n^{-1} U_l p, r \left( \frac{\Lambda}{n^{d}} + \varepsilon^{3} I_{p \times p} \right)^{-1} U_n^{-1}
\]

\[
= \frac{1}{n^{d}} \sum_{j=1}^{N} (x_{k,j} - x_{k}) (f(x_{k,j}) - f(x_{k})) X_{B^{d}} (x_{k}) + \mathcal{E}_3 + \text{higher order terms}
\]

Note that

\[
\mathbb{E} \frac{1}{\varepsilon^{d}} (X - x_{k}) X_{B^{d}} (x_{k})^\top U_l p, r \left( \frac{\Lambda}{n^{d}} + \varepsilon^{3} I_{p \times p} \right)^{-1} U_n \mathcal{E}_2 = T_{E}(x_{k}) \mathcal{E}_2
\]
When $x \in M_e$

$$T_{(x_k)} d_2 = \left[ O(\varepsilon^{-1}), O(\varepsilon^{-1}) \right] : \left[ O(\sqrt{\log(n)}), O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right) \right] = \left[ O(\sqrt{\log(n)}), O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right) \right].$$

(3.2.125)

When $x \notin M_e$

$$T_{(x_k)} d_2 = \left[ O(1), O(\varepsilon^{-1}) \right] : \left[ O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right), O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right) \right] = \left[ O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right), O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right) \right].$$

(3.2.126)

Moreover, when $x_k \in M_e$ or $x_k \in M \setminus M_e$ by a similar calculation as in Lemma 3.2.2, $UI_{p,r}(\frac{\Lambda}{\varepsilon d} + \varepsilon^3 I_{p \times p})^{-1} U^T \frac{1}{\varepsilon d} (X - x_k)(f(X) - f(x_k))\chi_{B_E^{\varepsilon,p}}(x_k)(X) = O(1, O(1))$. Hence,

$$\delta_1^T UI_{p,r}(\frac{\Lambda}{\varepsilon d} + \varepsilon^3 I_{p \times p})^{-1} U^T \frac{1}{\varepsilon d} (X - x_k)(f(X) - f(x_k))\chi_{B_E^{\varepsilon,p}}(x_k)(X) = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right).$$

(3.2.127)

Next, we calculate $E \frac{1}{\varepsilon d} (X - x_k)\chi_{B_E^{\varepsilon,p}}(x_k)(X)^T \delta_1^T E \frac{1}{\varepsilon d} (X - x_k)(f(X) - f(x_k))\chi_{B_E^{\varepsilon,p}}(x_k)(X)$. By a straightforward calculation, we can show that it is dominated by

$$O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right) E \frac{1}{\varepsilon d} (X - x_k)\chi_{B_E^{\varepsilon,p}}(x_k)(X)^T E \frac{1}{\varepsilon d} (X - x_k)(f(X) - f(x_k))\chi_{B_E^{\varepsilon,p}}(x_k)(X).$$

(3.2.128)

Hence, when $x_k \in M_e$,

$$E \frac{1}{\varepsilon d} (X - x_k)\chi_{B_E^{\varepsilon,p}}(x_k)(X)^T \delta_1 E \frac{1}{\varepsilon d} (X - x_k)(f(X) - f(x_k))\chi_{B_E^{\varepsilon,p}}(x_k)(X) = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right).$$

(3.2.129)

When $x_k \notin M_e$,

$$E \frac{1}{\varepsilon d} (X - x_k)\chi_{B_E^{\varepsilon,p}}(x_k)(X)^T \delta_1 E \frac{1}{\varepsilon d} (X - x_k)(f(X) - f(x_k))\chi_{B_E^{\varepsilon,p}}(x_k)(X) = O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right).$$

(3.2.130)

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

$$\frac{1}{n\varepsilon d} \sum_{j=1}^N (x_{k,j} - x_k)\sum_{j=1}^N \frac{\Lambda_n}{n\varepsilon d} E \frac{1}{\varepsilon d} (X - x_k)(f(X) - f(x_k))\chi_{B_E^{\varepsilon,p}}(x_k)(X)$$

$$+ O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right).$$

(3.2.131)

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

$$\frac{1}{n\varepsilon d} \sum_{j=1}^N (x_{k,j} - x_k)\sum_{j=1}^N \frac{\Lambda_n}{n\varepsilon d} E \frac{1}{\varepsilon d} (X - x_k)(f(X) - f(x_k))\chi_{B_E^{\varepsilon,p}}(x_k)(X)$$

$$+ O\left(\frac{\sqrt{\log(n)}}{n^{1/2}e^{d/2}}\right).$$

(3.2.131)

By Theorem 3.2.2, $g_1$ has order $O(\varepsilon^2)$ and $g_2$ has order 1. Hence, (3.2.118),(3.2.119),(3.2.130) and (3.2.131)
implies that
\[
\sum_{j=1}^{n} [W - I_{n\times n}]_{kj} f(x_j) = g_1 + O\left(\frac{\log(n)}{n^{1/2}d^{2/3}}\right) + \frac{g_2 + O\left(\frac{\log(n)}{n^{1/2}d^{2/3}}\right)}{n^{1/2}d^{2/3}} f(x_k) = Qf(x_k) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}d^{2/3}}\right).
\]

3.3 Dirichlet Graph Laplacian

Motivated by the spirit of LLE, we introduce our algorithm to estimate the Laplace-Beltrami operator with the Dirichlet boundary condition from a given point cloud. Fix \( \varepsilon \) and let \( \mathcal{N}_k := B_{\varepsilon}^{p} (x_k) \cap \mathcal{X} \setminus \{x_k\} = \{x_j, j \neq k\} \). Recall that \( G_n \) is the local data matrix associated with \( \mathcal{N}_k \):

\[
G_n := \begin{bmatrix}
    x_{k, 1} - x_k & \cdots & x_{k, N_k} - x_k \\
    \vdots & \ddots & \vdots \\
    \vdots & \cdots & \vdots \\
    x_{k, 1} - x_k & \cdots & x_{k, N_k} - x_k
\end{bmatrix} \in \mathbb{R}^{p \times N_k}.
\]

Define a \( n \times n \) diagonal matrix \( B \) such that

\[
B_{kk} = \frac{\|G_n 1_{N_k}\|_{\mathbb{R}^{p}}^2}{N_k \varepsilon}.
\]

We call \( B \) the bumping matrix.

With the bumping matrix, Dirichlet Graph Laplacian (DGL) is given by

\[
L := I_{n \times n} - W - B,
\]

where \( W \) is the weighted Laplacian matrix. Note that we do not detect boundary in the algorithm; instead, based on the intrinsic geometric nature of the truncated barycentric coordinate, the bumping matrix \( B \) takes care of the boundary automatically. This renders the DGL an approximation of the Laplace-Beltrami operator with the Dirichlet boundary condition.

We define a “boundary function” \( B_\varepsilon(x) \) on \( M \) as

\[
B_\varepsilon(x) = \frac{1}{\varepsilon} \left[ \frac{\|\mathbb{E}[X - 1(X(x))] B_{\varepsilon}^{p} (x (X(x))[X])\|_{\mathbb{R}^{p}}}{\mathbb{E}[\mathbb{E}[X B_{\varepsilon}^{p} (x (X(x))[X])]} \right]^2.
\]

The bump matrix and the boundary function can be related by the following proposition.

**Proposition 3.3.1.**

1. When \( \varepsilon \) is sufficiently small, the boundary function satisfies

\[
B_\varepsilon(x) = \frac{1}{\varepsilon} \left[ \frac{\mu_{c_d}(x, \varepsilon)^2}{\mu_0(x, \varepsilon)^2} + \frac{2\partial_d P(x)}{P(x)} (\frac{\mu_{c_d}(x, \varepsilon)\mu_{2c_d}(x, \varepsilon)}{\mu_0(x, \varepsilon)^2} - \frac{\mu_{c_d}(x, \varepsilon)^3}{\mu_0(x, \varepsilon)^3}) \right] + O(\varepsilon^3).
\]

If \( x \in M_\varepsilon \),

\[
B_\varepsilon(x) = O(\varepsilon).
\]

In particular, when \( x \in \partial M, B_\varepsilon(x) = \frac{4[d^{d-2}/2]}{(d-1)^{d-1}[\sqrt{\log(n)}}] \varepsilon + O(\varepsilon^3) \). If \( x \notin M_\varepsilon \),

\[
B_\varepsilon(x) = O(\varepsilon^3).
\]

2. Suppose \( f \in C^2(M) \). Suppose \( \varepsilon = \varepsilon(n) \) so that \( \sqrt{\log(n)} \rightarrow 0 \) and \( \varepsilon \rightarrow 0 \) as \( n \rightarrow \infty \). We have with probability
greater than $1 - n^{-2}$ that

if $x_k \in M$, then

$$B_{k} (x_k) = B_{k} (x_k) f (x_k) + O \left( \frac{\sqrt{\log (n)}}{n^{1/2} \varepsilon^{d/2 - 1}} \right),$$

If $x_k \not\in M$, then

$$B_{k} (x_k) = B_{k} (x_k) f (x_k) + O \left( \frac{\sqrt{\log (n)}}{n^{1/2} \varepsilon^{d/2 - 2}} \right). \quad (3.3.8)$$

**Proof.** By Lemma 3.1.3, we have

$$\mathbb{E}[X_{B_{k}^p (t(x))} (X)] = P (x) \mu_0 (x, \varepsilon) + O (\varepsilon^{d + 1}), \quad (3.3.9)$$

$$\mathbb{E} [(X - t (x)) X_{B_{k}^p (t(x))} (X)] = [P (x) \mu_{d} (x, \varepsilon)] + O (\varepsilon^{d + 2}), O (\varepsilon^{d + 2})], \quad (3.3.10)$$

Hence,

$$\left\| \mathbb{E} [(X - t (x)) X_{B_{k}^p (t(x))} (X)] \right\|_{p}^2 = \left( P (x) \mu_{d} (x, \varepsilon) + \partial_{d} P (x) \mu_{2d} (x, \varepsilon) \right)^2 + O (\varepsilon^{2d + 4})$$

$$= P (x)^2 \mu_{d} (x, \varepsilon)^2 + 2P (x) \partial_{d} P (x) \mu_{d} (x, \varepsilon) \mu_{2d} (x, \varepsilon) + O (\varepsilon^{2d + 4}), \quad (3.3.11)$$

and

$$\mathbb{E} [X_{B_{k}^p (t(x))} (X)] = P (x)^2 \mu_{0} (x, \varepsilon)^2 + 2P (x) \partial_{d} P (x) \mu_{0} (x, \varepsilon) \mu_{2d} (x, \varepsilon) + O (\varepsilon^{2d + 2}) \quad (3.3.12)$$

Therefore, by the definition (3.3.4), we have

$$B_{k} (x) = \frac{1}{\varepsilon} \left( \frac{\left\| \mathbb{E} [(X - t (x)) X_{B_{k}^p (t(x))} (X)] \right\|_{p}}{\mathbb{E} [X_{B_{k}^p (t(x))} (X)]} \right)^2$$

$$= \frac{1}{\varepsilon} \left( \frac{P (x)^2 \mu_{d} (x, \varepsilon)^2 + 2P (x) \partial_{d} P (x) \mu_{d} (x, \varepsilon) \mu_{2d} (x, \varepsilon) + O (\varepsilon^{2d + 4})}{P (x)^2 \mu_{0} (x, \varepsilon)^2 + 2P (x) \partial_{d} P (x) \mu_{0} (x, \varepsilon) \mu_{2d} (x, \varepsilon) + O (\varepsilon^{2d + 2})} \right)^2$$

$$= \frac{\mu_{d} (x, \varepsilon)^2 + \partial_{d} P (x) \mu_{d} (x, \varepsilon) \mu_{2d} (x, \varepsilon) - \mu_{d} (x, \varepsilon)^3 \mu_{2d} (x, \varepsilon)}{\mu_{0} (x, \varepsilon)^2} \left( 1 - \frac{\mu_{d} (x, \varepsilon)^3 \mu_{2d} (x, \varepsilon)}{\mu_{0} (x, \varepsilon)^2} \right) + O (\varepsilon^3). \quad (3.3.13)$$

Recall that by Lemma 3.2.6

$$\frac{1}{n \varepsilon^d} \sum_{j=1}^{N} (x_{k,j} - x_k) = \mathbb{E} \frac{1}{\varepsilon^d} (X - x_k) X_{B_{k}^p (x_k)} (X) + \varepsilon_1, \quad (3.3.14)$$

where $\varepsilon_1 \in \mathbb{R}^p$. $e_i^\top \varepsilon_1 = O \left( \frac{\sqrt{\log (n)}}{n^{1/2} \varepsilon^{d/2 - 1}} \right)$ for $i = 1, \ldots, d$, and $e_i^\top \varepsilon_1 = O \left( \frac{\sqrt{\log (n)}}{n^{1/2} \varepsilon^{d/2 - 2}} \right)$ for $i = d + 1, \ldots, p.$
Theorem 3.3.1. 1. Suppose $f \in C^3(M)$. Suppose $\varepsilon = \varepsilon(n)$ so that $\frac{\sqrt{\log(n)}}{n^{1/2} \sqrt{d/2 + 1}} \to 0$ and $\varepsilon \to 0$ as $n \to \infty$. We have with probability greater than $1 - n^{-2}$ that for all $k = 1, \ldots, n$,

$$- \sum_{j=1}^{n} L_{kj} f(x_j) = R_{\varepsilon} f(x_k) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2} \sqrt{d/2 - 1}}\right).$$

2. We have for any $x \in M$,

$$R_{\varepsilon} f(x) = \sum_{i=1}^{d} \phi_i(x, \varepsilon) \partial_i^2 f(x) + g(V(x, \varepsilon), \nabla f(x)) + B_{\varepsilon}(x) f(x) + O(\varepsilon^3).$$

When $x \not\in M_{\varepsilon}$, we have

$$R_{\varepsilon} f(x) = \frac{1}{2(d + 2)} \Delta f(x) \varepsilon^2 + O(\varepsilon^3);$$

where $f \in C(M)$.
When \( x \in M_{\varepsilon} \), we have
\[
R_{\varepsilon} f(x) = B_{\varepsilon}(x)f(x) + O(\varepsilon^2).
\] (3.3.21)

In particular, if \( x \in \partial M \),
\[
R_{\varepsilon} f(x) = C f(x) \varepsilon + O(\varepsilon^2),
\] (3.3.22)

where \( C \) is a constant which only depends on the dimension of \( M \).

Next, we show the convergence of \( \frac{R_{\varepsilon} f(x)}{\varepsilon^2} \) to the Laplace-Beltrami operator with the Dirichlet boundary condition, \( \Delta^{(D)} \), in the \( L^2 \) sense as \( \varepsilon \to 0 \). Suppose \(-\lambda_i, i = 1, \ldots, \) are the eigenvalues of \( \Delta^{(D)} \) so that \( 0 < \lambda_1 \leq \lambda_2 \leq \ldots \). Since \( \Delta^{(D)} \) is not a bounded operator with respect to \( L^2 \) norm, we cannot show the convergence in the operator norm. Instead, we show the proof via the uniform convergence over finite dimensional subspaces composed of finite eigenspace of \( \Delta^{(D)} \). For \( k \in \mathbb{N} \), define
\[
E_k := \{ f \in C^0(M)|\Delta^{(D)} f = -\lambda_k f, f|_{\partial M} = 0 \text{ and } ||f||_{L^2} = 1 \}.
\] (3.3.23)

The convergence result can be stated as follows,

**Theorem 3.3.2.** Fix \( K \in \mathbb{N} \). There exists a constant \( C \) which only depends on \( M \) and \( P \) such that if \( \varepsilon \) satisfies
\[
\varepsilon \leq \frac{C}{(1 + \lambda_k^{d/4+1})^d},
\] (3.3.24)
we have for all \( f \in \bigoplus_{l=1}^K E_l \) that
\[
\left| \frac{R_{\varepsilon} f(x)}{\varepsilon^2} - \frac{1}{2(d+2)} \Delta^{(D)} f(x) \right|_{L^2} \leq C \varepsilon^{1/2}.
\] (3.3.25)

**Proof.** In this proof, to simplify the notation, we use \( \lesssim \) to denote less than or equal up to a multiplicative constant which is independent of \( \varepsilon \) and \( f \). Based on (3.3.19), for \( f \in \bigoplus_{l=1}^K E_l \) and any \( x \in M \), we have
\[
\frac{R_{\varepsilon} f(x)}{\varepsilon^2} = \frac{1}{2(d+2)} \Delta f(x) - \frac{1}{2(d+2)} \sum_{i=1}^d \left[ \frac{\phi_i(x, \varepsilon)}{\varepsilon^2} \right] \partial_i^2 f(x) + g \left( \frac{V(x, \varepsilon)}{\varepsilon^2}, \nabla f(x) \right) + B_{\varepsilon}(x) f(x) + O(\varepsilon).
\] (3.3.26)

In particular, if \( x \notin M_{\varepsilon} \), by applying Theorem 3.3.1
\[
\frac{R_{\varepsilon} f(x)}{\varepsilon^2} = \frac{1}{2(d+2)} \Delta f(x) = O(\varepsilon).
\] (3.3.27)

We first bound the \( O(\varepsilon) \) term in (3.3.26) and (3.3.27). Denote \( \alpha = (\alpha_1, \ldots, \alpha_d) \) to be a multi-index. Then the term \( O(\varepsilon) \) in (3.3.26) and (3.3.27) depends on \( D^{(\alpha)} f \) for \( |\alpha| = 1, \ldots, 3 \). Note that
\[
||f||_{C^1(M)} \lesssim ||f||_{H^{d/2+1}/M} \lesssim 1 + ||\Delta^{d/4+1/2} f||_{L^2(M)} \lesssim 1 + \lambda_k^{d/4+1/2}
\]
where the first inequality is the Sobolev Embedding Theorem [45] and the third inequality is by the assumption that \( f \in \oplus_{i=1}^{K} E_i \). Note that \( \|f\|_{C^1(M)} \varepsilon^{1/2} \lesssim \varepsilon^{1/4} \) if and only if \( \varepsilon \approx (1 + \lambda_{K}^{d/4+1/2+1/2})^{-4/4} \), which is sufficient when

\[
\varepsilon \approx \frac{1}{(1 + \lambda_{K}^{d/4+1/2})^4}.
\]  

(3.3.28)

Combining above discussions, we conclude that for a given \( K \in \mathbb{N} \), if \( \varepsilon \approx \frac{1}{(1 + \lambda_{K}^{d/4+1/2})^4} \), the \( O(\varepsilon) \) terms in (3.3.26) and (3.3.27) are bounded above by \( \varepsilon^{1/2} \).

Next, we bound \( \left\| \frac{R_{\varepsilon}f(x)}{\varepsilon^2} - \frac{1}{2(d+2)} \Delta f(x) \right\|_{L^2} \). Fix \( x \in M_{\varepsilon} \), it follows from Proposition 3.3.1 and a straightforward calculation that \( \frac{B_{\varepsilon}(x)}{\varepsilon^2} \lesssim \frac{1}{\varepsilon} \) for \( \varepsilon \) sufficiently small. Similarly, the following expressions,

\[
\frac{\phi_i(x, \varepsilon)}{\varepsilon^2} \quad \text{for} \quad i = 1, \cdots, d \tag{3.3.29}
\]

\[
\frac{V(x, \varepsilon)}{\varepsilon^2} \tag{3.3.30}
\]

are of order \( O(1) \). Hence, they are uniformly bounded from above on \( M_{\varepsilon} \) for \( \varepsilon \) sufficiently small. Finally, note the fact that \( \text{Vol}(M_{\varepsilon}) \lesssim \varepsilon \) since the boundary is smooth and \( \text{Vol}(M_{\varepsilon}) \) is of order 1. Thus, by putting the above together, we have the following estimation:

\[
\left\| \frac{R_{\varepsilon}f(x)}{\varepsilon^2} - \frac{1}{2(d+2)} \Delta f(x) \right\|_{L^2} \leq \left\| \frac{R_{\varepsilon}f(x)}{\varepsilon^2} - \frac{1}{2(d+2)} \Delta f(x) \right\|_{L^2(M_{\varepsilon})} + \left\| \frac{R_{\varepsilon}f(x)}{\varepsilon^2} - \frac{1}{2(d+2)} \Delta f(x) \right\|_{L^2(M_{\varepsilon})} + \varepsilon^{1/2}
\]

\[
\lesssim \max_{1 \leq |\alpha| \leq 2} \|D^{\alpha} f\|_{L^2(M_{\varepsilon})} \text{Vol}(M_{\varepsilon}) + \left\| \frac{B_{\varepsilon}(x)}{\varepsilon^2} \right\|_{L^2(M_{\varepsilon})} \|f\|_{L^2(M_{\varepsilon})} \text{Vol}(M_{\varepsilon}) + \varepsilon^{1/2}
\]

\[
\lesssim \|f\|_{C^2(M_{\varepsilon})} \varepsilon + \|f\|_{L^2(M_{\varepsilon})} + \varepsilon^{1/2},
\]

where we use the fact that \( \max_{|\alpha|=2} \|D^{\alpha} f\|_{L^2(M_{\varepsilon})} \leq 2 \|f\|_{C^2(M_{\varepsilon})} \varepsilon \) and \( \|f\|_{C^2(M_{\varepsilon})} \varepsilon \leq \|f\|_{C^2(M_{\varepsilon})} \varepsilon \lesssim \varepsilon^{1/2} \). To control \( \|f\|_{L^2(M_{\varepsilon})} \), we use the assumption that \( \Delta \) is the Laplace-Beltrami operator with the Dirichlet boundary and \( f \in \oplus_{i=1}^{K} E_i \) is smooth. As a result, we have \( \|f\|_{L^2(M_{\varepsilon})} \lesssim \varepsilon \) when \( \varepsilon \) is sufficiently small. In conclusion, when \( \varepsilon \approx \frac{1}{(1 + \lambda_{K}^{d/4+1/2})^4} \), for all \( f \in \oplus_{i=1}^{K} E_i \) we have

\[
\left\| \frac{R_{\varepsilon}f(x)}{\varepsilon^2} - \frac{1}{2(d+2)} \Delta f(x) \right\|_{L^2} \lesssim \varepsilon^{1/2}
\]  

(3.3.32)
Bibliography


