TANGENT SPACE ESTIMATION FOR SMOOTH EMBEDDINGS OF RIEMANNIAN MANIFOLDS

HEMANT TYAGI, ELIF VURAL AND PASCAL FROSSARD

ABSTRACT. Numerous dimensionality reduction problems in data analysis involve the recovery of low-dimensional models or the learning of manifolds underlying sets of data. Many manifold learning methods require the estimation of the tangent space of the manifold at a point from locally available data samples. Local sampling conditions such as (i) the size of the neighborhood (*sampling width*) and (ii) the number of samples in the neighborhood (*sampling density*) affect the performance of learning algorithms. In this work, we propose a theoretical analysis of local sampling conditions for the estimation of the tangent space at a point P lying on a m-dimensional Riemannian manifold S in $\mathbb{R}^n$. Assuming a smooth embedding of S in $\mathbb{R}^n$, we estimate the tangent space $T_P S$ by performing a Principal Component Analysis (PCA) on points sampled from the neighborhood of P on S. Our analysis explicitly takes into account the second order properties of the manifold at P, namely the *principal curvatures* as well as the higher order terms. We consider a random sampling framework and leverage recent results from random matrix theory to derive conditions on the *sampling width* and the local *sampling density* for an accurate estimation of tangent subspaces. We measure the estimation accuracy by the angle between the estimated tangent space $\widehat{T}_P S$ and the true tangent space $T_P S$ and we give conditions for this angle to be bounded with high probability. In particular, we observe that the local sampling conditions are highly dependent on the correlation between the components in the second-order local approximation of the manifold. We finally provide numerical simulations to validate our theoretical findings.

1. INTRODUCTION

A data set that resides in a high-dimensional ambient space and that is locally homeomorphic to a lower-dimensional Euclidean space constitutes a manifold. For example, a set of signals that is representable by a parametric model, such as parametrizable visual signals or acoustic signals form a manifold. Data manifolds are however rarely given in an explicit form. The recovery of low-dimensional structures underlying a set of data, also known as manifold learning, has thus been a popular research problem in the recent years. This is typically achieved by constructing a mapping from the original data in the high-dimensional space to a space of much lower dimension. Importantly, most manifold learning methods rely on the assumption that the data has a locally linear structure. Of course, for such an assumption to be valid at some reference point on the manifold, one has to take into account (i) the size of the neighborhood from which the samples are chosen and also, (ii) the number of neighborhood points. For instance, if the manifold is a linear subspace, then the neighborhood can be chosen to be arbitrarily large and the number of samples needs to be simply greater than the dimension of the manifold. However, most manifolds are typically nonlinear, which prevents the selection of an arbitrarily large neighborhood size. Hence, one might expect the existence of an upper bound on the neighborhood size. Furthermore, the number of necessary samples is likely to vary according to the local characteristics of the manifold.

The purpose of this work is to analyze the relation between the sampling conditions of a manifold and the validity of the local linearity assumption of the data sampled from the manifold. We characterize the local linearity of the data with the accuracy of the tangent space estimation. We do a local analysis around a point P on a manifold S. We examine the deviation between the tangent space $\widehat{T}_P S$ estimated using manifold samples in a neighborhood of P, and the true tangent space $T_P S$ at P. This deviation is related to

the local geometric properties of the manifold around P and the local sampling conditions. In this paper, S is assumed to be an m-dimensional Riemannian manifold in $\mathbb{R}^n$ that can be locally represented with smooth $(\mathcal{C}^r, r > 2)$ mappings, where $m < n$. We consider a random sampling where the orthogonal projections of the samples to $T_P S$ in a neighborhood of P is uniform. We derive bounds on the size of the neighborhood and on the number of samples such that the deviation (i.e., the angle) between $\widehat{T}_P S$ and $T_P S$ is upper bounded with high probability. In particular, our analysis captures the dependency of the sampling conditions on the second-order properties of the manifold, namely the local curvature of S at P, and on the higher-order terms. Thus, broadly speaking, this work consists of a theoretical analysis of the manifold sampling problem that relates the *local* sampling conditions to the accuracy of the local linearity assumption. This paper builds on our preliminary work [1], where the sampling of manifolds represented with quadratic embeddings is examined, and extends the analysis to arbitrary smooth embeddings. We envisage two main applications where our study can prove to be useful. Firstly, our results can be used for deducing performance guarantees or for determining a good local subset of data samples that gives an accurate estimation of the tangent space in *manifold learning* applications. Secondly, our analysis can also be used in *manifold sampling* applications, i.e., for choosing samples from a manifold with a known parametric model. The discretization of a manifold can be achieved in various ways depending on the target application (see for example [2]); however, in certain cases one may want to sample the manifold in such a way that the local linearity of the data is preserved and the tangent space can be correctly recovered from data samples.

The manifold learning problem has been largely studied and we provide now a brief overview of the literature, with a special focus on locally linear approximation methods. The manifold structure of data can be retrieved in various ways, from a global parameterization based on geodesic distances as in ISOMAP [3], or via locally linear representations as in LLE [4] and Hessian Eigenmaps [5]. The LLE algorithm considers the locally linear structure of the manifold, where each data sample is approximated by a weighted linear combination of its nearest neighbors. Then, the key idea in computing a mapping of the data is the preservation of these weights in the embedded low-dimensional space. Moreover, there are other algorithms such as [6] which employ the locally linearity of the data by expressing the tangent plane as a linear combination of the manifold samples in a local neighborhood. The Hessian Eigenmaps algorithm is similar to LLE in the sense that it is based on locally linear approximations of the manifold. However, it has been seen to be more robust than LLE as it also takes more detailed geometric characteristics of the manifold into account. With similar ideas, an adaptive manifold learning algorithm is presented in [7], where the authors propose an adaptive local neighborhood size selection strategy. Finally, the work in [8] examines the conditions under which manifold learning algorithms are able to recover true global parameters from local structures computed with data samples. In particular, the authors show that the error in the global parameterization depends on the local approximation errors, as well as the null space and eigenvalue separation properties of the global parameterization.

Among the dimensionality reduction methods, one can find many examples of algorithms such as [5], [9], [10], [11], which apply a local Principal Component Analysis (PCA) for the computation of the tangent space of the manifold like we do in this work. In other words, the tangent space is estimated by computing the eigenvectors of the covariance of the data matrix, where the data samples come from a set of neighbor points on the manifold. This step can be seen as an analysis of PCA under data perturbations, where the perturbation of the data is caused by the nonlinear geometry of the embedding, i.e., the deviation of the manifold samples from the tangent space as a result of nonzero curvature. The performance of Singular Value Decomposition (SVD) or PCA in case of stochastic perturbations is a well-studied topic. There are many results in the literature that examine the perturbation on the singular vectors of a data matrix in the presence of noise. The Davis-Kahan theorem [12] is a classical result that examines how much the subspace spanned by the eigenvectors of a Hermitian matrix is rotated upon the perturbation of the matrix. The Wedin theorem [13] generalizes the analysis to non-Hermitian operators by bounding the angle between the estimated and true singular vectors in terms of the separation between the eigenvalues of the data matrix. A recent result in [14] addresses the singular vector estimation problem under assumptions of random perturbation noise and low-rank matrix. Finally, the work in [15] examines the bias of random measurement error on PCA and relates the bias to the SNR of the observed data. However, above studies do not involve the geometric structure of the data. There are also many studies that analyze the performance of PCA for a set of data generated by a specific model. For instance, the works such as [16], [17], [18] address the analysis of the

eigenvalues and eigenvectors of the covariance matrix of some data conforming to a multivariate normal distribution. These works however do not specifically consider any manifold data model either.

Only a few recent works have studied the relation between the PCA performance and the data geometry. The work in [19] presents an interesting study that generalizes the idea of diffusion maps in dimensionality reduction [20] to vector diffusion maps, where the new vector diffusion distance involves the similarity between the tangent spaces on different manifold points. In their analysis, the authors also provide a soft bound for the deviation of the locally estimated tangent space at a reference point (using local PCA) from the true tangent space, for a probabilistic sampling of the manifold. In particular, it shows that, when the size ε of the local area for tangent estimation is set to $\varepsilon = O(K^{-\frac{1}{m+2}})$ with K being the number of samples on the *whole* manifold, the deviation between the estimated and the true tangent space is typically of $O(\varepsilon^{3/2})$. This work however considers a global sampling from a compact manifold while we focus on the local manifold geometry. Finally, the accuracy of tangent space estimation from noisy manifold samples is analyzed in a work parallel to ours [21]. The manifold is assumed to be embedded with exactly quadratic forms (similarly to [1]) and the data consists of manifold samples corrupted with Gaussian noise. The work optimizes the number of samples (from a fixed sets of candidates) that is used for estimating the tangent space by considering the effect of noise and curvature on the accuracy of estimation. In particular, the optimal number of samples is selected as a trade-off between the error due to noise and the error caused by the curvature that respectively decreases and increases as the number of samples grows. This study however focuses on manifolds that are embedded with exactly quadratic forms and characterized with a subset of noisy samples given a priori. On the contrary, we are interested in more generic embeddings with arbitrary smooth functions and we aim at characterizing a sampling strategy in terms of the sampling width and density for noiseless manifold samples.

In our paper, we propose to characterize the local linearity of a manifold by studying the accuracy of the tangent space estimation from a local set of randomly selected manifold samples. We propose the following contributions. First, we determine a suitable upper bound on the neighborhood size within which random manifold sampling can be done. In the derivation of this bound, we consider the asymptotic case $K \to \infty$ so that the neighborhood size purely depends on the manifold geometry. In particular, our analysis depends on (i) the maximum principal curvature of the manifold and (ii) the deviation of the manifold from its second-order approximation. Our main results are stated precisely in Lemma 2 for the quadratic embedding case and in Lemma 4 for the more general smooth embedding case. They show the dependency of the neighborhood size on the correlation between the components in the second-order local approximation of the manifold. Second, we compute a bound on the *minimum* number of samples for accurate tangent space estimation, given that the sampling is performed randomly in a neighborhood whose size conforms with Lemmas 2 and 4. We utilize recent results from random matrix theory [22], [23] in our analysis. We state the precise expression for this bound on the number of samples in Lemma 3. Combining the two above results, we give a complete characterization of the local sampling conditions in the form of main theorems, namely Theorem 1 for the quadratic embedding case and Theorem 2 for the more general smooth embedding case. We finally discuss potential applications of the new theoretical results proposed in this paper, in respectively manifold learning and manifold sampling problems.

The rest of the paper is organized as follows. In Section 2, we first define the notations used in the paper and then give a formal statement of the problem along with the assumptions made. For ease of readability, the main results of the paper are presented in Section 3. We then present in Section 4 a detailed analysis of the local sampling conditions for tangent space estimation at a reference point P on S. In particular, Sections 4.2 and 4.3 contain the sampling analysis for the case when the embedding is assumed to be exactly quadratic at P. In Sections 4.4 and 4.5, we analyze the more general scenario of m-dimensional smooth embeddings in $\mathbb{R}^n$. Section 5 presents simulation results on synthetically generated smooth manifolds. In Section 6, we provide a discussion regarding the usage of our theoretical results in practical applications. Finally, in Section 7, we provide concluding remarks along with possible directions for future work.

2. Problem Formulation

In this section we first define the notations used in the paper. We then define the our manifold approximation framework. We finally state formally the problem of tangent space estimation that is studied in this paper.

2.1. **Notations.** Let $S \subset \mathbb{R}^n$ be a manifold and $P \in S$ be a reference point on the manifold where the local sampling analysis is performed. We denote the dimension of the manifold S by m. The tangent space at
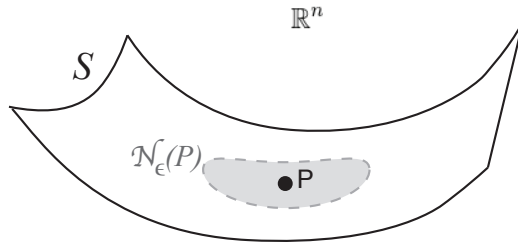
FIGURE 1. The manifold S and the ε-neighborhood of a manifold point $P \in S$.

$P \in S$ is represented by $T_P S$ and $T_P S^\perp$ is used to denote the orthogonal complement of $T_P S$ in $\mathbb{R}^n$. The notation $\mathcal{C}^r$ is used for denoting r times continuous differentiability.

We denote the ℓ_p-norm of a vector $\bar{x} \in \mathbb{R}^n$, $1 \leq p \leq \infty$, by $\| \bar{x} \|_p := \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$ and its ℓ_∞-norm by $\| \bar{x} \|_\infty := \max_i |x_i|$. The inner product between $\bar{x}, \bar{y} \in \mathbb{R}^n$ is denoted by $\langle \bar{x}, \bar{y} \rangle := \bar{x}^T \bar{y}$. Furthermore, we represent a canonical vector in $\mathbb{R}^n$ by $\bar{e}_j$ for $j = 1, \ldots, n$, where $\bar{e}_j$ has a 1 at the jth position and 0 at all other positions.

Given a matrix $X \in \mathbb{R}^{p \times q}$, we have by its (reduced) singular value decomposition (SVD) [24] the factorization $X = U \Sigma V^T$ where $U \in \mathbb{R}^{p \times s}$ and $V \in \mathbb{R}^{q \times s}$ are the singular vector matrices with orthonormal columns. The dimension $s \leq \min(p, q)$ corresponds to the rank of X. The matrix $\Sigma = \mathrm{diag}(\sigma_1(X), \ldots, \sigma_s(X))$ is a diagonal matrix where $\sigma_1(X) \geq \cdots \geq \sigma_s(X) > 0$ are the singular values of X. We denote the Frobenius norm of X (the ℓ_2-norm of its vector of singular values) by $\| X \|_F := (\mathrm{Tr}(X^T X))^{1/2}$ and its operator norm (the largest singular value) by $\| X \|$. For any square matrix $X \in \mathbb{R}^{p \times p}$, we denote the trace by $\mathrm{Tr}(X)$ and the determinant by $\det(X)$.

For a symmetric matrix $X \in \mathbb{R}^{p \times p}$, $X = X^T$ we have the eigenvalue decomposition $X = U \Lambda U^T$. Here $\Lambda = \mathrm{diag}(\lambda_1(X), \ldots, \lambda_p(X))$ denotes the eigenvalue matrix with $\lambda_1(X) \geq \cdots \geq \lambda_p(X)$ and $U \in \mathbb{R}^{p \times p}$ is a unitary matrix so that $U U^T = U^T U = I$. If X is symmetric and positive semidefinite we then have $\lambda_i(X) \geq 0$ for $i = 1, \ldots, p$. We denote the spectral radius of a symmetric matrix X by $\rho(X) = \max_i(|\lambda_i(X)|)$.

Throughout the paper, $\mathbb{E}[\cdot]$ is used for denoting the expectation and $\mathbb{P}(\cdot)$ for denoting the probability.

2.2. **Framework.** We consider an m-dimensional submanifold S of $\mathbb{R}^n$ with a smooth embedding in $\mathbb{R}^n$, $n \geq m + 1$. Let $\mathcal{N}_\varepsilon(P)$ denote a ε-neighbourhood of P for some $\varepsilon > 0$, where

$$\mathcal{N}_\varepsilon(P) = \{ M \in S : \| M - P \|_2 \leq \varepsilon \} .$$

The neighborhood of P on S is illustrated in Fig. 1. In this work, as we represent points in $\mathcal{N}_\varepsilon(P)$ via tangent space parameterization using local functions $f_l : T_P S \to \mathbb{R}$, we are interested in the mapping that orthogonally projects the manifold points in a neighborhood of P to $T_P S$. In [25], Niyogi et al. provide a characterization of the neighborhood of P within which this mapping is one-to-one, through the condition number of the manifold. Therefore, there exists an ε such that all points $M \in \mathcal{N}_\varepsilon(P)$ can be uniquely represented in the form

$$[\bar{x}^T \ f_1(\bar{x}) \ldots f_{n-m}(\bar{x})]^T. \tag{2.1}$$

Here $\bar{x} = [x_1 \ldots x_m]^T$ denotes the coordinates of the orthogonal projection of a point on $T_P S$. Note that, in (2.1), the coordinates are with respect to the point P that is the reference point, i.e., the local origin. Furthermore, the tangent space $T_P S$ at P can be represented as

$$T_P S = \mathrm{span} \{ \bar{e}_1, \ldots, \bar{e}_m \} ,$$

where $\bar{e}_j \in \mathbb{R}^n$ denote the canonical vectors.

Now, we further assume the smoothness of the embedding to be $\mathcal{C}^r$, $r > 2$, implying that each

$$f_l : T_P S \to \mathbb{R}, \quad l = 1, \ldots, n - m,$$

is a $\mathcal{C}^r$-smooth function in the variables $(x_1, \ldots, x_m)$. Since $\nabla f_l(\bar{0}) = \bar{0}$ we have by the Taylor expansion of f_l around the origin (i.e., P) the following identity:

$$f_l(\bar{x}) = f_{q,l}(\bar{x}) + O(\| \bar{x} \|_2^3); \quad l = 1, \ldots, n - m \tag{2.2}$$
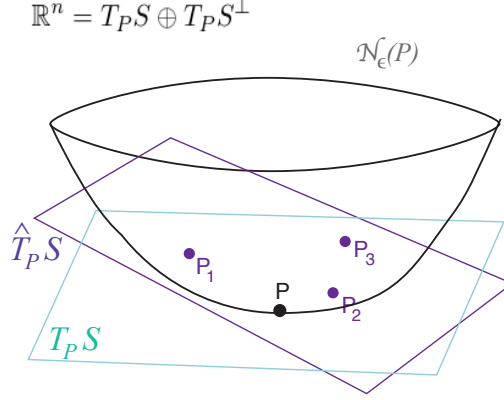
FIGURE 2. The true tangent space $T_P S$ and the estimated tangent space $\widehat{T}_P S$ at point P.

where $f_{q,l}$ is a quadratic form. As a special case, we have a *quadratic embedding* at P when each f_l is an exact quadratic form, i.e.,

$$f_l(\cdot) = f_{q,l}(\cdot); \quad l = 1, \ldots, n - m.$$

Consider the Hessian of f_l at the local origin P, which is given as

$$\nabla^2 f_l(\bar{0}) = V_l \Lambda_l V_l^T,$$

where $\Lambda_l = \text{diag}(\mathcal{K}_{l,1}, \mathcal{K}_{l,2}, \ldots, \mathcal{K}_{l,m})$. Here $\mathcal{K}_{l,1}, \mathcal{K}_{l,2}, \ldots, \mathcal{K}_{l,m}$ are the principal curvatures of the hypersurface

$$\mathcal{S}_l = \left\{ [x_1 \ \ldots \ x_m \ f_l(x_1, \ldots, x_m)] : [x_1 \ \ldots \ x_m]^T \in T_P S \right\} \subset \mathbb{R}^{m+1}$$

defined by f_l. We then define the maximum principal curvature at P as

$$\mathcal{K}_{max} := \mathcal{K}_{l',j'} \quad \text{where} \quad (l', j') = \underset{l,j}{\text{argmax}} |\mathcal{K}_{l,j}|.$$

We consider that the tangent space can be estimated from sample points in $\mathcal{N}_\varepsilon(P)$ through a PCA decomposition. More precisely, let us consider K points $\{P_i\}_{i=1}^K$ sampled from $\mathcal{N}_\varepsilon(P)$. Let $M^{(K)}$ denote the local covariance matrix where

$$M^{(K)} = \sum_{i=1}^{K} \frac{1}{K} P_i P_i^T = U \Lambda U^T.$$

It is a common preprocessing step to subtract the empirical mean of the data from data samples in usual PCA. However, in our application, the linear subspace computed with PCA is an estimation of the tangent space, which is restricted to pass from the local origin P. Therefore, we omit the mean subtraction step in our analysis and assume that the principal components are computed with respect to the reference point P. The matrices U and $\Lambda \in \mathbb{R}^n$ represent the eigenvector and eigenvalue matrices respectively of $M^{(K)}$ where

$$U = [\bar{u}_1 \ldots \bar{u}_m \ldots \bar{u}_n]; \quad \Lambda = \text{diag}(\lambda_1, \ldots \lambda_m, \ldots \lambda_n),$$

with the ordering $\lambda_1 \geq \cdots \geq \lambda_m \geq \cdots \geq \lambda_n$. The optimal m-dimensional linear subspace at P in the least squares sense is then given by the span of the m largest eigenvectors of $M^{(K)}$, i.e.,

$$\widehat{T}_P S := \text{span} \{\bar{u}_1, \ldots, \bar{u}_m\}.$$

The tangent space $T_P S$ and its estimation $\widehat{T}_P S$ are illustrated in Fig. 2. Finally, we characterize the accuracy of our estimation with the angle between the estimated and the true tangent spaces. The notion of 'angle' between two linear subspaces as defined in [26] is given in Definition 1.

Definition 1. *The angle $\angle A, B$ between two subspaces $A = \text{span}\{\bar{a}_1, \ldots, \bar{a}_p\}$ and $B = \text{span}\{\bar{b}_1, \ldots, \bar{b}_q\}$ of a Euclidean space $\mathbb{R}^n$, where $\bar{a}_i$'s and $\bar{b}_i$'s are orthonormal vectors, is defined as*

$$\cos^2 \angle A, B := \det(W^T W),$$

where $[W^T]_{i,k} := \langle \bar{a}_i, \bar{b}_k \rangle$ is a $p \times q$ matrix, with $1 \leq p \leq q < \infty$.

Observe that the definition can be applied to subspaces that are not necessarily of the same dimension. Geometrically speaking,

$$\cos \angle A, B \ := \ \frac{V_1}{V_2}$$

where V_1 is the volume of the p-dimensional parallelepiped spanned by the projection of $\{\bar{a}_1, \ldots, \bar{a}_p\}$ on B and V_2 is the volume of the p-dimensional parallelepiped spanned by $\{\bar{a}_1, \ldots, \bar{a}_p\}$. Therefore, in order to compare $T_P S$ and $\widehat{T}_P S$, one could also consider the distance between the respective projection matrices EE^T and $U^{(m)} U^{(m)^T}$ through

$$\| EE^T - U^{(m)} U^{(m)^T} \|_F^2, \tag{2.3}$$

where $E = [\bar{e}_1 \ldots \bar{e}_m]$ and $U^{(m)} = [\bar{u}_1 \ldots \bar{u}_m]$. Note that an upper bound on $|\angle \widehat{T}_P S, T_P S|$ implies a corresponding upper bound on $\| EE^T - U^{(m)} U^{(m)^T} \|_F^2$. Finally, our choice of using Definition 1 for estimating angles is motivated by the measure of the geometric deviation of $\widehat{T}_P S$ from $T_P S$. However, one could also work with the error criteria of Eq. (2.3) with no change in the analysis and sampling conditions.[1]

2.3. **Problem statement.** Given the above settings, we want to describe the conditions on the manifold samples $\{P_i\}_{i=1}^K$ such that for a given error bound $\phi \in (0, \frac{\pi}{2})$ on the tangent space estimation,

$$|\angle \widehat{T}_P S, T_P S| \ < \ \phi \ < \frac{\pi}{2}$$

is ensured. In particular, for a given error bound ϕ, we would like to answer the following questions:

- *Question 1:* What would be a suitable upper bound on the *sampling distance*; i.e., the distance of P_i from P? In particular, for large embeddding dimensions n, what is the nature of the dependency of this bound on n, m and $\mathcal{K}_{max}$?
- *Question 2:* Given that the points $\{P_i\}_{i=1}^K$ are sampled such that the sampling distance satisfies the sampling distance bound, what would be a suitable lower bound on the *sampling density* K? In particular, for large embeddding dimensions n, what is the nature of the dependency of this bound on n, m and $\mathcal{K}_{max}$?

In order to answer the above questions, we consider a random sampling framework where we assume that the coordinates of the orthogonal projections of manifold samples on $T_P S$ are distributed uniformly in the region $[-\nu, \nu]^m \in T_P S$. In other words, denoting the coordinates of the projection of P_i on $T_P S$ by $\bar{x}_i = [x_1^{(i)} \ldots x_m^{(i)}]$, we assume that

$$x_j^{(i)} \sim \mathcal{U}[-\nu, \nu] \quad \text{i.i.d.} \quad i = 1, \ldots, K; j = 1, \ldots, m$$

where $\mathcal{U}$ denotes the uniform distribution. Therefore, we characterize the *sampling distance* in Question 1 by the parameter ν, which we shall refer to as the sampling width in our analysis.[2]

3. Main results

We summarize in this section the main results of the paper. We provide sampling conditions for tangent space estimation in two different cases; namely, quadratic embeddings and generic smooth embeddings.

3.1. **Quadratic embedding at P.** We first consider as a special case the scenario where the manifold S has a quadratic embedding at P in $\mathbb{R}^n$. We present the main sampling theorem in the form of Theorem 1 below. The main purpose of this result is to gain some intuition about the sampling conditions when the local functions f_l's involved in the tangent space parametrization have a purely quadratic form and they are not 'perturbed' by higher-order terms. We refer the reader to Section 4.2 for details regarding the proof and for a more rigorous analysis.

Theorem 1 (Quadratic manifold sampling). *Consider $\{P_i\}_{i=1}^K$ to be formed by sampling uniformly at random from the region $[-\nu, \nu]^m$ around P in $T_P S$, i.e.,*

$$x_j^{(i)} \sim \mathcal{U}[-\nu, \nu] \quad i.i.d., \quad i = 1, \ldots, K, j = 1, \ldots, m.$$

---

[1]This is explained in more detail in Lemma 1 and Remark 2.

[2]See Section 6 for a discussion on how the bound on ν relates to the distance in the ambient space.

Let $D \in \mathbb{R}^{n-m \times n-m}$ denote the local correlation matrix for the mappings $\{f_{q,l}\}_{l=1}^{n-m}$ such that

$$[D]_{l,k} = \mathbb{E}[f_{q,l}(\bar{x})f_{q,k}(\bar{x})]; \quad l, k = 1, \dots, n - m.$$

We then have the following sufficient sampling conditions that guarantee a bound on $|\angle \widehat{T}_P S, T_P S|$. For any $\tau \in (0,1)$, the choices

$$\nu = O(n^{-1/2}m^{-1}|\mathcal{K}_{max}|^{-1}) \quad and \quad K = O(\tau^{-2}m^2 \log n), \quad as\ n \to \infty$$

ensure that $|\angle \widehat{T}_P S, T_P S| < \cos^{-1} \sqrt{(1 - \tau^2)^m}$ holds w.h.p.

Interpretation of Theorem 1. We see that the bound on ν behaves as $O(n^{-1/2}m^{-1}|\mathcal{K}_{max}|^{-1})$, indicating that the sampling region needs to shrink with the increase in ambient space dimension. Furthermore, we observe that the bound on the sampling width depends linearly on the reciprocal $|\mathcal{K}_{max}|^{-1}$ of the maximum curvature. The decrease in ν with respect to the increase in n and $\mathcal{K}_{max}$ can be explained as follows. Assuming that m is fixed, as n increases, the number $(n - m)$ of normal components that increase the nonlinearity of the manifold increases, which causes the deviation of S from the tangent space. Furthermore, the magnitude of each normal component increases with the increase in the curvatures associated with that normal direction. The sampling width ν must be selected sufficiently small such that the tangential components of the data have larger magnitudes than the normal components in PCA, in order to ensure the correct identification of the tangent space. Hence, the largest admissible value of the sampling width ν depends on the overall magnitude of the normal components. This is affected by both the codimension $(n - m)$ of S and the curvature parameter $\mathcal{K}_{max}$, which is used as a uniform bound on the individual curvatures in this work. In the derivation of these main results, the magnitude of the normal components is captured by the spectral norm of the $(n - m) \times (n - m)$ correlation matrix D, which increases with both the dimension of the ambient space, and the curvature.

Lastly, we remark that the approximation error term τ^2 arises on account of finite sampling and can be interpreted as the variance error. In particular, provided that the sampling width ν is chosen to satisfy the appropriate bound, then we have that $|\angle \widehat{T}_P S, T_P S| \to 0$ in the limit where $K \to \infty$.

3.2. **Smooth embedding of S in $\mathbb{R}^n$.** We now present our main sampling theorem for the general case of smooth embeddings of S in $\mathbb{R}^n$ in the form of Theorem 2. For details regarding the proof and for a rigorous analysis we refer the reader to Section 4.4.

Theorem 2 (Smooth manifold sampling). *Consider $\{P_i\}_{i=1}^K$ to be formed by sampling uniformly at random from the region $[-\nu, \nu]^m$ in $T_P S$, i.e.,*

$$x_j^{(i)} \sim \mathcal{U}[-\nu, \nu] \quad i.i.d., \quad i = 1, \dots, K, j = 1, \dots, m.$$

Let $D \in \mathbb{R}^{n-m \times n-m}$ denote the local correlation matrix for the mappings $\{f_{q,l}\}_{l=1}^{n-m}$, such that

$$[D]_{l,k} = \mathbb{E}[f_{q,l}(\bar{x})f_{q,k}(\bar{x})], \quad l, k = 1, \dots, n - m.$$

We then have the following sufficient sampling conditions that guarantee a bound on $|\angle \widehat{T}_P S, T_P S|$. For any $\tau \in (0,1)$, the choices

$$\nu = O(n^{-1/2}m^{-1}|\mathcal{K}_{max}|^{-1}) \quad and \quad K = O(\tau^{-2}m^2 \log n), \quad as\ n \to \infty$$

ensure that $|\angle \widehat{T}_P S, T_P S| < \cos^{-1} \sqrt{(1 - \tau^2 - O(n^{-1}m|\mathcal{K}_{max}|^{-4}))^m}$ holds w.h.p.

Interpretation of Theorem 2. As the manifold S is now smoothly embedded in $\mathbb{R}^n$, the local functions f_l's involved in the tangent space parametrization are arbitrary smooth functions of the form (2.2). Hence, in this case, the deviation of the manifold from the tangent space is due to both the second-order terms $\{f_{q,l}\}_{l=1}^{n-m}$ and the higher-order terms in the Taylor series of f_l (which are $O(\| \bar{x} \|_2^3)$). Observe that the bound on the sampling density K has a similar order of dependence on n, m, and $|\mathcal{K}_{max}|$ as in Theorem 1. In the bounds on $|\angle \widehat{T}_P S, T_P S|$, the error term represented by τ^2 corresponds to the variance due to finite sampling as in the quadratic embedding case. On the other hand, we see that there is an additional error term of $O(n^{-1}m|\mathcal{K}_{max}|^{-4})$ for smooth embeddings, which does not exist in the quadratic embeddings. This term arises on account of the higher-order terms in the Taylor expansion of f_l, and can be interpreted as a bias term due to a nonzero sampling width ν. This bias goes to zero as $\nu \to 0$. For smooth embeddings, in particular, for a fixed ν that is chosen to satisfy the appropriate bound, $|\angle \widehat{T}_P S, T_P S|$ approaches a constant

bias error term as the variance error vanishes in the limit where $K \to \infty$. The reason why the tangent space estimation is non-biased for quadratic embeddings and biased for arbitrary smooth embeddings can be explained as follows. The normal components $f_{q,l}$ in quadratic embeddings have a symmetry around the origin; i.e., $f_{q,l}(\bar{x}) = f_{q,l}(-\bar{x})$. However, for smooth embeddings we have $f_l(\bar{x}) \neq f_l(-\bar{x})$ in general because of the higher-order terms, which create an asymmetry in the orientation of the manifold points around the origin. This leads to a perturbation in the estimation of the tangent space with PCA and thus constitutes a bias.

Remark 1. *In the above results, we have considered the general case where the functions $\{f_{q,l}(\bar{x})\}_{l=1}^{n-m}$ are all correlated; i.e., D is a dense matrix with all nonzero entries. However, in a practical application, one can possibly encounter a setting where some of the $\left(f_{q,l}(\bar{x}), f_{q,k}(\bar{x})\right)$ pairs are uncorrelated, or weakly correlated. Therefore, D may typically be a sparse matrix or most of its entries may be close to zero in certain applications. In this case, some of the restrictions on the sampling conditions can be loosened. In order to demonstrate to what extent the sampling conditions may change with respect to the correlation of the normal components of the data, we focus throughout the analysis on the two extreme cases where the random variables $\{f_{q,l}(\bar{x})\}_{l=1}^{n-m}$ are all correlated (D is a dense matrix) and where $\{f_{q,l}(\bar{x})\}_{l=1}^{n-m}$ are mutually uncorrelated (D is a diagonal matrix)[3]. We now give an overview of how the above sampling conditions change when D is diagonal. In this case, for quadratic embeddings, the sufficient sampling conditions that guarantee the angle bound given in Theorem 1 can be replaced by*

$$\nu = O(m^{-1}|\mathcal{K}_{max}|^{-1}) \quad and \quad K = O(mn\tau^{-2}\log n).$$

Similarly, for a smooth embedding, the sampling conditions become

$$\nu = O(n^{-1/3}m^{-5/6}|\mathcal{K}_{max}|^{-1/3}) \quad and \quad K = O(n\log n),$$

which ensure that $|\angle \widehat{T}_P S, T_P S| < \cos^{-1}\sqrt{(1 - O(n^{-1/3}m^{5/3}|\mathcal{K}_{max}|^{-4/3}))^m}$ w.h.p. as $n \to \infty$.

Correlation of $\{f_{q,l}\}_{l=1}^{n-m}$	Smooth embedding	Quadratic embedding
Correlated	$\nu = O(n^{-1/2}m^{-1}|\mathcal{K}_{max}|^{-1})$	$\nu = O(n^{-1/2}m^{-1}|\mathcal{K}_{max}|^{-1})$
	$K = O(\tau^{-2}m^2\log n)$	$K = O(\tau^{-2}m^2\log n)$
Uncorrelated	$\nu = O(n^{-1/3}m^{-5/6}|\mathcal{K}_{max}|^{-1/3})$	$\nu = O(m^{-1}|\mathcal{K}_{max}|^{-1})$
	$K = O(n\log n)$	$K = O(mn\tau^{-2}\log n)$

TABLE 1. Summary of the sampling conditions at a point P on an m-dimensional manifold S in $\mathbb{R}^n$ ($m < n$) such that $|\angle \widehat{T}_P S, T_P S| < \cos^{-1}\sqrt{(1-\tau^2)^m}$ for some $\tau \in (0,1)$. The two following cases are compared: (i) S has a smooth embedding in $\mathbb{R}^n$, (ii) S has a quadratic form in $\mathbb{R}^n$ w.r.t. the point P.

These dependences are summarized in Table 1 in comparison with the results obtained for the general case where the normal components are correlated. One can observe the following. For both quadratic and smooth embeddings, the admissible sampling width can be chosen larger if the functions $\{f_{q,l}(\bar{x})\}_{l=1}^{n-m}$ are uncorrelated; i.e., the dependence of ν on n, m and $\mathcal{K}_{max}$ is loosened if D is diagonal. This can be intuitively explained as follows. If D is dense and there is a high correlation between two functions $f_{q,l}(\bar{x})$ and $f_{q,k}(\bar{x})$, the projection $[0 \ldots 0\, f_{q,l}(\bar{x})\, 0 \ldots\, f_{q,k}(\bar{x})\, \ldots 0] \subset T_P S^{\perp}$ of S onto the normal plane generated by the normal directions l and k has a strong orientation along a certain direction on this normal plane. This creates a dominant normal direction along which the data is concentrated. Therefore, when the mutual correlations of the functions $\{f_{q,l}(\bar{x})\}_{l=1}^{n-m}$ are high, more dominant normal directions are generated. This creates a bigger challenge for the recovery of the tangent directions with PCA and necessitates the selection of a smaller sampling width. However, if $\{f_{q,l}(\bar{x})\}_{l=1}^{n-m}$ are uncorrelated or weakly correlated, there are less dominant normal directions. This results in looser constraints on the sampling width. This phenomenon reveals itself in the derivations through the spectral norm of the D matrix. When the nonzero entries of the D matrix are restricted to the diagonals, the spectral norm of the matrix grows at a slower rate with respect to the increase in n and $\mathcal{K}_{max}$, in comparison with the case where D is dense.

---

[3]More details about the feasibility of D being diagonal are given in Section 4.2, Remark 3.

Meanwhile, sampling from a wider region requires the selection of more samples; i.e., since ν is greater for the uncorrelated case, the required sampling density K is higher when D is diagonal. The bound on K is logarithmic in n for D dense, and loglinear in n for D diagonal. This makes sense, since reducing the sampling width ν causes $\mathcal{N}_\varepsilon(P)$ to be more linear and hence loosens the restrictions on the number of samples required to achieve a given approximation bound on $|\angle \widehat{T}_P S, T_P S|$. Of course, the case where D is dense is general and hence the corresponding bounds for ν and K can be used even if D is in fact diagonal. If D is diagonal, one can however afford to sample the manifold from a larger neighborhood.

4. ANALYSIS

We now present a detailed analysis of our local sampling results for smooth m-dimensional Riemannian manifolds in $\mathbb{R}^n$. To begin with, we first define the framework for our analysis by introducing the tangent space parameterization for points in $\mathcal{N}_\varepsilon(P)$.

4.1. Framework for tangent space estimation.
We discuss first the parametrization that we use in our analysis. Let $[x_1 \ \ldots \ x_m \ f_1(\bar{x}) \ \ldots \ f_{n-m}(\bar{x})]^T$ be a point in $\mathcal{N}_\varepsilon(P)$ and let $\bar{x} = [x_1 \ \ldots \ x_m]^T$ denote its orthogonal projection on $T_P S$. The region $\mathcal{N}_\varepsilon(P)$ can be represented in terms of $(n-m)$ hypersurfaces of dimension m in $\mathbb{R}^{m+1}$, where the lth hypersurface is given by

$$\mathcal{S}_l = \left\{ [x_1 \ \ldots \ x_m \ f_l(x_1, \ldots, x_m)] : \ [x_1 \ \ldots \ x_m]^T \ \in T_P S \right\} \subset \mathbb{R}^{m+1}.$$

Due to the assumption that the embedding is $\mathcal{C}^r$, where $r > 2$, the functions f_l have the following form $\forall l = 1, \ldots, n-m$

$$f_l(\bar{x}) = f_l(\bar{0}) + \nabla f_l(\bar{0})^T \bar{x} + \frac{1}{2} \bar{x}^T \nabla^2 f_l(\bar{0}) \bar{x} + R_l(\bar{\xi}_l),$$

$$= \bar{0} + \bar{0}^T \bar{x} + \frac{1}{2} \bar{x}^T V_l \Lambda_l V_l^T \bar{x} + R_l(\bar{\xi}_l),$$

$$= \frac{1}{2} \sum_{j=1}^{m} \left(<\bar{x}, \bar{v}_{l,j}>^2 \mathcal{K}_{l,j} \right) + R_l(\bar{\xi}_l),$$

$$= \frac{1}{2} \parallel \bar{x} \parallel_2^2 \mathcal{K}_l(\bar{x}) + R_l(\bar{\xi}_l) = f_{q,l}(\bar{x}) + R_l(\bar{\xi}_l),$$

where $\bar{\xi}_l \in (\bar{0}, \bar{x})$ depends on $\bar{x}$. Here, $f_{q,l}$ denotes the quadratic approximation of f_l, and $R_l(\bar{\xi}_l) = O(\parallel \bar{x} \parallel_2^3)$ represents the higher-order remainder terms in its Taylor series. The Hessian of f_l at the origin is represented by $\nabla^2 f_l(\bar{0})$, and

$$V_l = \begin{bmatrix} \bar{v}_{l,1} & \bar{v}_{l,2} & \cdots & \bar{v}_{l,m} \end{bmatrix}_{m \times m}, \qquad \Lambda_l = \text{diag}(\mathcal{K}_{l,1}, \mathcal{K}_{l,2}, \ldots, \mathcal{K}_{l,m})$$

denote respectively the eigenvector and eigenvalue matrices of $\nabla^2 f_l(\bar{0})$. Geometrically, $\mathcal{K}_l(\bar{x})$ represents the curvature at point P of the geodesic curve on $\mathcal{S}_l$ from P to $[\bar{x}^T \ f_1(\bar{x}) \ \ldots \ f_{n-m}(\bar{x})]^T$, where

$$\mathcal{K}_l(\bar{x}) = \sum_{j=1}^{m} \left(\frac{<\bar{x}, \bar{v}_{l,j}>^2}{\parallel \bar{x} \parallel_2^2} \mathcal{K}_{l,j} \right).$$

Given the above setting, recall from Section 2.3 that

$$\mathcal{K}_{max} = \mathcal{K}_{l',j'} \quad \text{where} \quad (l', j') := \underset{l,j}{\text{argmax}} |\mathcal{K}_{l,j}|.$$

Here, $|\mathcal{K}_{max}|$ is the largest absolute value of the principal curvatures among the hypersurfaces $\mathcal{S}_l$, for $l = 1, \ldots, n-m$. We remark that the sampling conditions derived throughout our analysis capture the second-order properties of the manifold at P in terms of the maximum curvature $\mathcal{K}_{max}$. Equipped with the above parametrization, we can now describe the estimation of the tangent space. Let us consider K points,

$$\{P_i\}_{i=1}^K = \left\{ [x_1^{(i)} \ x_2^{(i)} \ \ldots \ x_m^{(i)} \ f_1(\bar{x}_i) \ \ldots \ f_{n-m}(\bar{x}_i)]^T \right\}_{i=1}^K$$

in $\mathcal{N}_\varepsilon(P)$. Denoting the coordinates of the orthogonal projection of P_i on T_PS as $\bar{x}_i = [x_1^{(i)} \; x_2^{(i)} \; \ldots \; x_m^{(i)}]^T$, for $i = 1, \ldots, K$, we represent the points by the matrix $X^{(K)}$ as follows.

$$X^{(K)} = \begin{bmatrix} x_1^{(1)} & \cdots & x_1^{(K)} \\ \vdots & & \vdots \\ x_m^{(1)} & \cdots & x_m^{(K)} \\ f_1(\bar{x}_1) & \cdots & f_1(\bar{x}_K) \\ \vdots & & \vdots \\ f_{n-m}(\bar{x}_1) & \cdots & f_{n-m}(\bar{x}_K) \end{bmatrix}$$

where each f_l has the following form:

$$f_l(\bar{x}) \;=\; \frac{1}{2}\sum_{j=1}^m <\bar{x}, \bar{v}_{l,j}>^2 \mathcal{K}_{l,j} + O(\| \bar{x} \|_2^3) ; \quad l = 1, \ldots, n-m. \tag{4.1}$$

The optimal m-dimensional linear subspace, in the least squares sense, passing through P will be the one spanned by the eigenvectors corresponding to the m largest eigenvalues of

$$M^{(K)} = \frac{1}{K}XX^{T^{(K)}} = \begin{bmatrix} A^{(K)} & B^{(K)} \\ B^{(K)^T} & D^{(K)} \end{bmatrix} = U\Lambda U^T,$$

where the individual submatrices have the following form.

$$A^{(K)} = \begin{bmatrix} \frac{1}{K}\sum_i (x_1^{(i)})^2 & \cdots & \frac{1}{K}\sum_i x_1^{(i)}x_m^{(i)} \\ \vdots & & \vdots \\ \frac{1}{K}\sum_i x_m^{(i)}x_1^{(i)} & \cdots & \frac{1}{K}\sum_i (x_m^{(i)})^2 \end{bmatrix}, \quad B^{(K)} = \begin{bmatrix} \frac{1}{K}\sum_i x_1^{(i)}f_1(\bar{x}_i) & \cdots & \frac{1}{K}\sum_i x_1^{(i)}f_{n-m}(\bar{x}_i) \\ \vdots & & \vdots \\ \frac{1}{K}\sum_i x_m^{(i)}f_1(\bar{x}_i) & \cdots & \frac{1}{K}\sum_i x_m^{(i)}f_{n-m}(\bar{x}_i) \end{bmatrix}$$

and

$$D^{(K)} = \begin{bmatrix} \frac{1}{K}\sum_i f_1^2(\bar{x}_i) & \cdots & \frac{1}{K}\sum_i f_1(\bar{x}_i)f_{n-m}(\bar{x}_i) \\ \vdots & & \vdots \\ \frac{1}{K}\sum_i f_{n-m}(\bar{x}_i)f_1(\bar{x}_i) & \cdots & \frac{1}{K}\sum_i f_{n-m}^2(\bar{x}_i) \end{bmatrix}.$$

Furthermore,

$$U = \begin{bmatrix} \bar{u}_1 & \cdots & \bar{u}_m & \bar{u}_{m+1} & \cdots & \bar{u}_n \end{bmatrix} \text{ and } \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

are respectively the eigenvector and eigenvalue matrices of $\frac{1}{K}XX^{T^{(K)}}$ with $U^TU = UU^T = I_n$. Assume the ordering $\lambda_1 \geq \cdots \geq \lambda_m \geq \cdots \geq \lambda_n$. We then have

$$\widehat{T}_PS = \text{span}\{\bar{u}_1, \ldots, \bar{u}_m\} \text{ and } \quad T_PS = \text{span}\{\bar{e}_1, \ldots, \bar{e}_m\},$$

where $\{\bar{e}_j\}_{j=1}^m$ denote the first m of the n canonical basis vectors in $\mathbb{R}^n$. Now the angle between $\widehat{T}_PS$ and T_PS as per Definition 1 is given by

$$\cos^2(\angle \widehat{T}_PS, T_PS) \;:=\; \det(W^TW),$$

where $[W^T]_{i,j} = <\bar{u}_i, \bar{e}_j>$ for $1 \leq i, j \leq m$. Let us denote the first m columns of U by $U^{(m)}$ where

$$U^{(m)} = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}; \quad U_1 \in \mathbb{R}^{m\times m}, U_2 \in \mathbb{R}^{(n-m)\times m}.$$

Lemma 1 states the condition on $\| U_2 \|_F$ that guarantees a bound on $|\angle \widehat{T}_PS, T_PS|$.

Lemma 1. *Consider $K \geq m$ points in $\mathcal{N}_\varepsilon(P)$ sampled such that $\| U_2 \|_F < \tau < 1$ for some $0 < \tau < 1$. Then,*

$$|\angle \widehat{T}_PS, T_PS| < \cos^{-1}\sqrt{(1-\tau^2)^m}.$$

Proof. Clearly $U_1^TU_1 + U_2^TU_2 = I_{m\times m}$. Let $E = \begin{bmatrix} \bar{e}_1 & \ldots & \bar{e}_m \end{bmatrix}$. We have

$$W^TW = (U^{(m)^T}E)(E^TU^{(m)}) = U_1^TU_1 = I_{m\times m} - U_2^TU_2.$$

Denoting the eigenvalues of $U_2^TU_2$ as $\mu_1, \ldots, \mu_m$, we observe that

$$Tr(U_2^TU_2) \;=\; \| U_2 \|_F^2 \geq \mu_{max},$$

where $\mu_{max} = \max_{i=1\ldots m} \mu_i$. Using this result in conjunction with Definition 1, we arrive at the following inequality.

$$\cos^2(\angle \widehat{T}_P S, T_P S) := \det\left(I_{m \times m} - U_2^T U_2\right) = \prod_{i=1}^{m} (1 - \mu_i) \geq (1 - \mu_{max})^m.$$

Hence, the following bound on $|\angle \widehat{T}_P S, T_P S|$ clearly holds if $\| U_2 \|_F < \tau < 1$.

$$\cos^2(\angle \widehat{T}_P S, T_P S) > (1 - \tau^2)^m \Leftrightarrow |\angle \widehat{T}_P S, T_P S| < \cos^{-1}\sqrt{(1 - \tau^2)^m}.$$

$\square$

Remark 2. *We remark here that one can compare the column spaces of E and $U^{(m)}$ by also computing the difference between their projection matrices, i.e., $\| EE^T - U^{(m)}U^{(m)^T} \|_F^2$. It is easily verifiable that*

$$\| EE^T - U^{(m)}U^{(m)^T} \|_F^2 = 2 \| U_2 \|_F^2. \tag{4.2}$$

Hence when $\| U_2 \|_F < \tau < 1$, we have $\| EE^T - U^{(m)}U^{(m)^T} \|_F < \sqrt{2}\tau$. The core of our analysis involves deriving sampling conditions which guarantee that $\| U_2 \|_F$ is suitably upper bounded. Hence one can interchangeably use Eq. (4.2) or the notion of angle in Definition 1 to compare $T_P S$ and $\widehat{T}_P S$ with no change in the analysis and the sampling conditions derived later on. The only change would be in the expression for the error bound where instead of $\cos^{-1}\sqrt{(1 - \tau^2)^m}$ one would have the error term $\sqrt{2}\tau$. Our choice of using Definition 1 is purely motivated by our objective of measuring the deviation of $\widehat{T}_P S$ from $T_P S$ in a geometric way.

Finally, we note that, if the manifold S is a linear subspace of $\mathbb{R}^n$, then we have $B^{(K)} = 0$ and $D^{(K)} = 0$ implying U_2 to be trivially equal to zero. In other words, we have $\angle \widehat{T}_P S, T_P S = 0$ for any $K \geq m$. However, when S is a more general manifold, then its nonlinearity manifests itself in the form of error arising due to the local mappings $\{f_l\}_{l=1}^{n-m}$. Hence, in order to obtain a good locally linear approximation of $T_P S$, one intuitively expects that the points in $\mathcal{N}_\varepsilon(P)$ are sampled sufficiently close to P. In particular, one might wonder how far from P points can be sampled and also how many points need to be sampled in order to achieve a good approximation guarantee on $|\angle \widehat{T}_P S, T_P S|$. We now proceed to rigorously analyze these two questions in the following sections.

4.2. **Accuracy of tangent space estimation for quadratic embeddings.** We assume first that $\mathcal{N}_\varepsilon(P)$ is representable in terms of quadratic forms at the reference point P. In other words, for any point $[x_1 \ldots x_m \ f_1(\bar{x}) \ldots f_{n-m}(\bar{x})]$ in $\mathcal{N}_\varepsilon(P)$, we have

$$f_l(\bar{x}) = f_{q,l}(\bar{x}), \quad l = 1, \ldots, n - m,$$

where $f_{q,l}(\cdot)$ denotes the second order approximation of $f_l(\cdot)$.

We consider the points $\{P_i\}_{i=1}^{K}$ to be formed by sampling independently and uniformly at random in $T_P S$ such that

$$x_j^{(i)} \sim \mathcal{U}[-\nu, \nu] \text{ i.i.d.}, \qquad \forall i = 1, \ldots, K \quad \text{and} \quad j = 1, \ldots, m.$$

To begin with, Lemma 2 states precisely the condition on ν which guarantees that $\widehat{T}_P S = T_P S$ in the limit where $K \to \infty$.

Lemma 2. *As $K \to \infty$, $[M^{(K)}]_{i,j} \to [M]_{i,j}$ a.s. for every $1 \leq i, j \leq n$, where*

$$M = \begin{bmatrix} \frac{\nu^2}{3} I_{m \times m} & 0_{m \times (n-m)} \\ 0_{(n-m) \times m} & D_{(n-m) \times (n-m)} \end{bmatrix}, \quad [D]_{l,k} = \mathbb{E}[f_l(\bar{x}) f_k(\bar{x})] = \mathbb{E}[f_{q,l}(\bar{x}) f_{q,k}(\bar{x})],$$

and $l, k = 1, \ldots, n - m$. Furthermore, the following holds.
(1) Let D be dense, i.e., let $\{f_{q,l}\}_{l=1}^{n-m}$ be correlated. Then, if the sampling width satisfies

$$\nu < \sqrt{\frac{60}{m(n-m)(5m+4)|\mathcal{K}_{max}|^2}},$$

it holds that $\mathbb{P}(|\angle \widehat{T}_P S, T_P S| > 0) \to 0$ as $K \to \infty$.

(2) Let D be diagonal, i.e., let $\{f_{q,l}\}_{l=1}^{n-m}$ be uncorrelated. Then, if the sampling width satisfies

$$\nu < \sqrt{\frac{60}{m(5m+4)|\mathcal{K}_{max}|^2}},$$

it holds that $\mathbb{P}(|\angle \widehat{T}_P S, T_P S| > 0) \to 0$ as $K \to \infty$.

The proof of Lemma 2 is presented in Appendix A.1. The main idea here is to observe that the eigenspace corresponding to the eigenvalue $\frac{\nu^2}{3}$ is equal to the span of $\{\bar{e}_1, \ldots, \bar{e}_m\}$, which is the same as $T_P S$. Hence, the condition on the sampling width follows from the requirement that the noiseless spectra associated with $\frac{\nu^2}{3} I_m$ is separated from the noisy spectra associated with D arising on account of the manifold's curvature at P. In other words,

$$\frac{\nu^2}{3} > \rho(D)$$

where $\rho(D)$ denotes the spectral radius of D. For the sake of brevity, let us denote the bound on ν by

$$\nu_{\text{bound,quad}} = 1/\sqrt{3RL}$$

where

$$L = \frac{m(5m+4)|\mathcal{K}_{max}|^2}{180} \quad \text{and} \quad R = \begin{cases} (n-m), & \text{if D is dense} \\ 1, & \text{if D is diagonal.} \end{cases}$$

Therefore, $\nu_{\text{bound,quad}}$ depends on the structure of D.

Remark 3. *The case where D is dense is general. Therefore, the derived condition on ν can be used even if D is actually diagonal. Moreover, if D is diagonal, then we see that the condition on ν is considerably less restrictive. We note here that $\{f_{q,l}\}_{l=1}^{n-m}$ will typically be correlated if m is fixed and n is allowed to increase to a large value. This arises due to the requirement*

$$\mathbb{E}[f_{q,l}(\bar{x}) f_{q,k}(\bar{x})] = 0, \text{with } l, k = 1, \ldots, n-m, \quad l \neq k,$$

where a large value of n and a small value of m result in more equations than degrees of freedom. Hence, in order to have D diagonal, the manifold dimension m needs to be sufficiently large. In the case that the correlation matrix D is sparse, the sufficient condition on ν lies in between the two bounds stated in Lemma 2.

We want now to find a lower bound on the number of samples K which guarantees that the deviation $|\angle \widehat{T}_P S, T_P S|$ is suitably upper bounded with high probability. Hence, we first derive a bound on K that guarantees some tail bounds on the eigenvalues of the submatrices of $M^{(K)}$. This bound is precisely stated in the following Lemma.

Lemma 3. *Let the sampling width be chosen such that $\nu < \nu_{bound,quad} = 1/\sqrt{3RL}$. Let $s_1 \in (0,1)$, $s_2 > e$, $s_3 > 0$ and $0 < p_1, p_2, p_3 < 1$ denote fixed constants. We define*

$$K_{bound}^{(1)} = \frac{6R_M}{(1-s_1)^2} \log\left((n-m+1)/p_1\right),$$

$$K_{bound}^{(2)} = \frac{R_D}{s_2 RL} \frac{\log((n-m)/p_2)}{\log(s_2/e)},$$

$$K_{bound}^{(3)} = \frac{\nu^6 R_\sigma + \frac{R_B \nu^3 s_3}{3}}{s_3^2/2} \log(n/p_3),$$

where

$$R_M = m + \frac{1}{4}(n-m)m^2\nu^2|\mathcal{K}_{max}|^2, \quad R_D = \frac{1}{4}(n-m)m^2|\mathcal{K}_{max}|^2,$$

$$R_\sigma = \frac{m^2|\mathcal{K}_{max}|^2}{12} \max\left\{(n-m), \frac{R(5m+4)}{15}\right\}, \text{ and } R_B = \frac{1}{2}m^{3/2}\sqrt{n-m}|\mathcal{K}_{max}|.$$

Then, let $K_{bound} = \max\{K_{bound}^{(1)}, K_{bound}^{(2)}, K_{bound}^{(3)}\}$. If the number of samples K satisfies $K > K_{bound}$, then the following bounds hold true with probability at least $1 - p_1 - p_2 - p_3$:

- *(i) $\lambda_m(M^{(K)}) > s_1 \frac{\nu^2}{3}$,*

- *(ii) $\rho(D^{(K)}) < s_2 \rho(D)$,*
- *(iii) $\| B^{(K)} \| < s_3$.*

The proof of Lemma 3 is presented in Appendix A.2. The lemma defines a sufficient bound on the sampling density, which in turn guarantees probabilistic bounds on the spectral norms of the perturbation matrices $B^{(K)}$ and $D^{(K)}$. Our proof builds on the recent results [22], [23], which give tail bounds on the eigenvalues of sums of independent random matrices.

We have seen earlier that, if ν is chosen to satisfy the appropriate bound on the sampling width, then $\mathbb{P}(|\angle \widehat{T}_P S, T_P S| > 0) \to 0$ as $K \to \infty$. We now employ Lemma 3 to show that, if $\nu < c\, \nu_{\text{bound,quad}}$ for some $c < e^{-1/2}$, and if $s_3 > 0$ is suitably upper bounded, then for $K > K_{bound}$, we have that $|\angle \widehat{T}_P S, T_P S|$ is bounded from above with high probability. This is stated precisely in Theorem 3.

Theorem 3. *Consider K points randomly sampled in $\mathcal{N}_\varepsilon(P)$ such that their projections to $T_P S$ are independent and uniform in the region $[-\nu, \nu]^m$, i.e.,*

$$x_j^{(i)} \sim U[-\nu, \nu] \quad i.i.d., \quad i = 1, \ldots, K, \quad j = 1, \ldots, m.$$

Under the notation defined earlier, assume that, for some fixed $s_1 \in (0, 1)$ and $s_2 > e$,

$$\nu < \sqrt{\frac{s_1}{s_2}} \nu_{bound,quad} = \sqrt{\frac{s_1}{3 s_2 R L}}.$$

Then, consider that, for some $\tau \in (0, 1)$,

$$0 < s_3 < \frac{(s_1 \frac{\nu^2}{3} - s_2 R L \nu^4) \tau}{\sqrt{m}} = s_{3\,bound,quad}. \tag{4.3}$$

Finally, let $0 < p_1, p_2, p_3 < 1$. Then, if $K > K_{bound}$, we have that

$$\mathbb{P}(|\angle \widehat{T}_P S, T_P S| < \cos^{-1} \sqrt{(1 - \tau^2)^m}) > 1 - p_1 - p_2 - p_3.$$

The proof is presented in Appendix A.3. In the proof, we use the conditions derived in Lemma 3 in order to obtain eigenvalue separation conditions for the correlation matrix constructed with a finite sampling, which are then used to derive a bound on $\|U_2\|_F$. Note that the error term τ is the variance error arising due to finite sampling; it goes to zero as $K \to \infty$.

4.3. **Analysis of the bounds for quadratic embedding.** We now proceed to analyze the dependence of the sampling parameters ν and K on the manifold dimension m, the maximum curvature $|\mathcal{K}_{max}|$ and the ambient space dimension n, where we assume that n is high (i.e., $n \to \infty$). We analyze this by considering two separate cases based on the structure of the matrix D.

(1) D **is dense.** When no assumption is made on the structure of D, we have $\nu_{\text{bound,quad}} = O(n^{-1/2} m^{-1} |\mathcal{K}_{max}|^{-1})$ as $n \to \infty$. Using this, one obtains from the corresponding expression of s_3 that

$$s_3 = O\left(n^{-1} m^{-5/2} |\mathcal{K}_{max}|^{-2} \tau\right).$$

For a given probability of success, we derive the sampling bound complexity as follows.

$$K_{bound}^{(1)} = O(R_M \log n) = O(m \log n),$$

$$K_{bound}^{(2)} = O\left(\frac{R_D}{RL} \log n\right) = O(\log n),$$

$$K_{bound}^{(3)} = O\left(\frac{\nu_{\text{bound,quad}}^6 R_\sigma + \frac{R_B \nu_{\text{bound,quad}}^3 s_3}{3}}{s_3^2} \log n\right) = O(\tau^{-2} m^2 \log n).$$

Thus $K_{bound} = O(\tau^{-2} m^2 \log n)$ as $n \to \infty$. Here, the number of samples is seen to depend quadratically on the manifold dimension and logarithmically on the ambient space dimension. Note that the dependency on n is milder in this case in comparison to the case where D is diagonal, which is due to the fact that the condition on the sampling width ν is stricter when D is dense.

(2) D **is diagonal.** We first observe that $\nu_{\text{bound,quad}}$ is independent of the dimension n. In particular, we have $\nu_{\text{bound,quad}} = O\left(m^{-1}|\mathcal{K}_{max}|^{-1}\right)$. Using this, one obtains from the corresponding expression of s_3 that

$$s_3 = O\left(m^{-5/2}|\mathcal{K}_{max}|^{-2}\tau\right).$$

For a given probability of success, we derive the sampling bound complexity as follows.

$$K_{bound}^{(1)} = O(R_M \log n) = O(n \log n),$$

$$K_{bound}^{(2)} = O\left(\frac{R_D}{RL} \log n\right) = O(n \log n),$$

$$K_{bound}^{(3)} = O\left(\frac{\nu_{\text{bound,quad}}^6 R_\sigma + \frac{R_B \nu_{\text{bound,quad}}^3 s_3}{3}}{s_3^2/2} \log n\right) = O(mn\tau^{-2} \log n).$$

Thus $K_{bound} = O(mn\tau^{-2} \log n)$ as $n \to \infty$. Hence, the number of samples has a linear dependence on the intrinsic dimension of the manifold and a loglinear dependence on the ambient space dimension.

4.4. **Accuracy of tangent space estimation for smooth embeddings.** In the previous section, we have assumed that $\mathcal{N}_\varepsilon(P)$ can be represented with quadratic forms. We now consider the more general scenario where the manifold is smoothly embedded in $\mathbb{R}^n$. In particular, we assume the smoothness class C^r, where $r > 2$, in order to be able to study the influence of the local curvature of the manifold. Under this assumption we have that $f_l(\cdot) = f_{q,l}(\cdot) + R_l(\cdot)$ for $l = 1, \ldots n - m$, where $f_{q,l}(\cdot)$ denotes the second order approximation of $f_l(\cdot)$ and $R_l(\cdot)$ denotes the higher-order terms. As each $f_l(\cdot)$ is defined over a compact domain, $R_l(\cdot)$ is bounded, i.e., $|R_l(\cdot)| = O \|\cdot\|_2^3$ for all $l = 1, \ldots, n - m$. Hence,

$$|R_l(\cdot)| < C_{s,l} \|\cdot\|_2^3 \quad l = 1, \ldots, n - m,$$

where the constant $C_{s,l} > 0$ depends on the magnitude of the third order derivatives of f_l in $\mathcal{N}_\varepsilon(P)$. We denote

$$C_s = \max_l C_{s,l}, \quad l = 1, \ldots, n - m.$$

Let us again consider the points $\{P_i\}_{i=1}^K$ to be formed by sampling independently and uniformly at random in $T_P S$ such that

$$x_j^{(i)} \sim \mathcal{U}[-\nu, \nu] \text{ i.i.d.} \qquad \forall i = 1, \ldots, K \quad \text{and} \quad j = 1, \ldots, m.$$

Using the same notation as before, when $f_l(\cdot) = f_{q,l}(\cdot) + R_l(\cdot)$, $l = 1, \ldots n - m$, we arrive at the following form for the local covariance matrix $M^{(K)}$.

$$M^{(K)} = M_q^{(K)} + \Delta^{(K)}$$

where

$$M_q^{(K)} = \begin{bmatrix} A^{(K)} & B^{(K)} \\ B^{(K)^T} & D^{(K)} \end{bmatrix}$$

is the covariance matrix considered in the previous section, with the submatrices $B^{(K)}$ and $D^{(K)}$ representing the error on account of the manifold's curvature at P. Furthermore,

$$\Delta^{(K)} = \begin{bmatrix} 0 & B_1^{(K)} \\ B_1^{(K)^T} & D_1^{(K)} \end{bmatrix}$$

is an additional error term arising on account of the higher-order Taylor series terms of the mappings $\{f_l\}_{l=1}^{n-m}$ with

$$B_1^{(K)} = \begin{bmatrix} \frac{1}{K}\sum_i x_1^{(i)} R_1(\bar\xi_{1,i}) & \cdots & \frac{1}{K}\sum_i x_1^{(i)} R_{n-m}(\bar\xi_{n-m,i}) \\ \vdots & & \vdots \\ \frac{1}{K}\sum_i x_m^{(i)} R_1(\bar\xi_{1,i}) & \cdots & \frac{1}{K}\sum_i x_m^{(i)} R_{n-m}(\bar\xi_{n-m,i}) \end{bmatrix}$$

and

$$[D_1]_{l,k}^{(K)} = \begin{cases} \frac{1}{K}\sum_i (R_l(\bar{\xi}_{l,i})R_k(\bar{\xi}_{k,i}) + R_l(\bar{\xi}_{l,i})f_{q,k}(\bar{x}_i) + R_k(\bar{\xi}_{k,i})f_{q,l}(\bar{x}_i)) & \text{if} \quad l \neq k \\ \frac{1}{K}\sum_i (R_l(\bar{\xi}_{l,i})^2 + 2f_{q,l}(\bar{x}_i)R_l(\bar{\xi}_{l,i})) & \text{if} \quad l = k. \end{cases}$$

To begin with, let us define

$$\delta(\nu) = C_s m^{3/2}\nu^3$$

where the factor $m^{3/2}$ appears since $\| \bar{x}_i \|_2^3 < m^{3/2}\nu^3$ for $i = 1, \dots, K$. We then observe that each entry of $\Delta^{(K)}$ can be bounded as

(1) $|[B_1^{(K)}]_{j,l}| < \nu\delta(\nu)$ for $j = 1, \dots m; l = 1, \dots, n-m$
(2) $|[D_1^{(K)}]_{l,k}| < \delta(\nu)^2 + \delta(\nu)m\nu^2|\mathcal{K}_{max}|$ for $l, k = 1, \dots, n-m$

where we used the fact that $|x_j^{(i)}| < \nu$ and $|R_l(\cdot)| < \delta(\nu)$ for obtaining (1); and $|f_{q,l}(\cdot)| < \frac{1}{2}m\nu^2|\mathcal{K}_{max}|$ for obtaining (2). Using the bounds on the entries, we obtain the following bounds on $\| B_1^{(K)} \|_F$ and $\| D_1^{(K)} \|_F$ respectively:

$$\| B_1^{(K)} \|_F < \sqrt{m(n-m)}\nu\delta(\nu) = \sqrt{m(n-m)}C_s m^{3/2}\nu^4 = \| B_1 \|_{F,bound},$$

$$\| D_1^{(K)} \|_F < (n-m)(\delta(\nu)^2 + \delta(\nu)m\nu^2|\mathcal{K}_{max}|)$$
$$= (n-m)C_s m^{5/2}\nu^5(C_s m^{1/2}\nu + |\mathcal{K}_{max}|) = \| D_1 \|_{F,bound}.$$

Now let us denote

$$B_1 = \mathbb{E}[B_1^{(K)}], \quad D_1 = \mathbb{E}[D_1^{(K)}], \text{ and } \Delta = \mathbb{E}[\Delta^{(K)}].$$

Due to the ergodicity of the sampling process, we have $B_1 = \lim_{K\to\infty} B_1^{(K)}$, $D_1 = \lim_{K\to\infty} D_1^{(K)}$, and $\Delta = \lim_{K\to\infty} \Delta^{(K)}$. Since the bounds on the entries of the perturbation submatrices $B_1^{(K)}$ and $D_1^{(K)}$ hold for all K, they are also valid for the entries of B_1 and D_1. Therefore, we get $\| B_1 \|_F < \| B_1 \|_{F,bound}$ and $\| D_1 \|_F < \| D_1 \|_{F,bound}$.

We first consider the case $K = \infty$, where we obtain $M = \lim_{K\to\infty} M^{(K)} = M_q + \Delta$. It was shown in Lemma 2 that

$$M_q = \begin{bmatrix} \frac{\nu^2}{3}I_{m\times m} & 0_{m\times(n-m)} \\ 0_{(n-m)\times m} & D_{(n-m)\times(n-m)} \end{bmatrix},$$

where $[D]_{l,k} = \mathbb{E}[f_{q,l}(\bar{x})f_{q,k}(\bar{x})]$, for $l, k = 1, \dots, n-m$. Given that M_q is now 'perturbed' by Δ, Lemma 4 states the conditions on the sampling width ν that guarantee an upper bound on the angle between $\widehat{T}_P S$ and $T_P S$.

Lemma 4. *Let the sampling width satisfy*

$$\nu < \frac{1}{[3((\beta_2 + RL) + \beta_3\alpha + \beta_4\alpha^2)]^{1/2}}$$

where $\beta_2 = 4C_s m^2(n-m)^{1/2}$, $\beta_3 = 2(n-m)C_s m^{5/2}|\mathcal{K}_{max}|$, $\beta_4 = 2(n-m)m^3 C_s^2$ and

$$\alpha = \min\left\{ (3(\beta_2 + RL))^{-1/2}, (3\beta_3)^{-1/3}, (3\beta_4)^{-1/4} \right\}.$$

Then, as $K \to \infty$,

$$\mathbb{P}\left(|\angle\widehat{T}_P S, T_P S| > \cos^{-1}\sqrt{(1 - m\sigma_\infty^2)^m} \right) \to 0$$

where

$$\sigma_\infty = \frac{\| B_1 \|_{F,bound}}{\frac{\nu^2}{3} - RL\nu^4 - 2(\| B_1 \|_{F,bound} + \| D_1 \|_{F,bound})}.$$

The proof is presented in Appendix A.4. The main idea here is to ensure that the spectrum associated with $\frac{\nu^2}{3}I_m$ is separated from the spectrum of the error arising due to the following factors:

(1) The curvature components $\{f_{q,l}\}_{l=1}^{n-m}$ which give rise to the correlation matrix D.
(2) The higher-order Taylor series terms of the smooth mappings $\{f_l\}_{l=1}^{n-m}$ giving rise to the perturbation matrix Δ.

Observe that, unlike in the case where f_l's are quadratic forms, the deviation $|\angle \widehat{T}_P S, T_P S|$ now does not converge to zero but to a residual bound $\cos^{-1} \sqrt{(1 - m\sigma_\infty^2)^m}$. This is on account of the additional error associated with the matrix Δ, which now perturbs the covariance matrix M_q. The error term $m\sigma_\infty^2$ can be interpreted as the bias error arising due to the nonzero sampling width. In particular, it is easily verifiable that

$$\sigma_\infty \to 0 \quad \text{as} \quad \nu \to 0.$$

Also note that, had the f_l's been quadratic forms, we would have $C_s = 0$ resulting in $\sigma_\infty = 0$. This gives us the result obtained in Lemma 2.

Remark 4. *We remark here that the choice of the sampling width ν satisfying the condition in Lemma 4 actually ensures the following bound:*

$$\frac{\nu^2}{3} - RL\nu^4 > 4 \parallel B_1 \parallel_{F,bound} + 2 \parallel D_1 \parallel_{F,bound}.$$

Using this implication in the expression for σ_∞, we obtain the trivial bound $\sigma_\infty < 1/2$. Furthermore, the residual angle bound term σ_∞ can be made arbitrarily small by choosing a sufficiently small ν.

We now proceed to the case $K < \infty$. Theorem 4, which is the main sampling theorem of this section, states the sufficient conditions on the sampling width ν and the number of samples K, such that the deviation $|\angle \widehat{T}_P S, T_P S|$ is suitably upper bounded with high probability.

Theorem 4. *Consider K points randomly sampled in $\mathcal{N}_\varepsilon(P)$ such that their projections to $T_P S$ are independent and uniform in the region $[-\nu, \nu]^m$, i.e.,*

$$x_j^{(i)} \sim \mathcal{U}[-\nu, \nu] \quad i.i.d., \quad i = 1, \dots, K, \text{ and } j = 1, \dots, m.$$

Under the notation defined earlier, assume that for some fixed $s_1 \in (0, 1)$ and $s_2 > e$, the following holds:

$$\nu < \left(\frac{s_1}{3[(\beta_2 + s_2 RL) + \beta_3 \alpha + \beta_4 \alpha^2]} \right)^{1/2} = \nu_{bound,smooth}.$$

Then, for some $\tau \in (0, 1)$, let $s_3 > 0$ be chosen such that

$$s_3 < [(s_1 \frac{\nu^2}{3} - s_2 RL\nu^4) - 2(\parallel B_1 \parallel_{F,bound} + \parallel D_1 \parallel_{F,bound})] \left(\frac{\tau^2}{m} + \sigma_f^2 \right)^{1/2} - \parallel B_1 \parallel_{F,bound} = s_{3\,bound,smooth}$$

(4.4)

where

$$\sigma_f = \frac{\parallel B_1 \parallel_{F,bound}}{(s_1 \frac{\nu^2}{3} - s_2 RL\nu^4) - 2(\parallel B_1 \parallel_{F,bound} + \parallel D_1 \parallel_{F,bound})}.$$

Finally, let $0 < p_1, p_2, p_3 < 1$. Then, if the number of samples satisfies $K > K_{bound}$, where K_{bound} is as derived in Lemma 3, the following holds true

$$\mathbb{P}(|\angle \widehat{T}_P S, T_P S| < \cos^{-1} \sqrt{(1 - \tau^2 - m\sigma_f^2)^m}) > 1 - p_1 - p_2 - p_3.$$

The proof is presented in Appendix A.5 and is built on the results of Lemma 4. It uses similar ideas to those in the proof of Theorem 3; however, the additional perturbation matrix Δ also plays a role in the derived bounds. Note that the approximation error consists of two terms - the variance term τ due to finite sampling and the bias term σ_f arising due to the nonzero sampling width ν.

Remark 5. *We again remark here that the choice of the sampling width ν in Theorem 4 ensures the following bound*

$$(s_1 \frac{\nu^2}{3} - s_2 RL\nu^4) > 4 \parallel B_1 \parallel_{F,bound} + 2 \parallel D_1 \parallel_{F,bound}.$$

Hence, it follows trivially that $\sigma_f < 1/2$. Furthermore, σ_f can be reduced appropriately by choosing a suitably downscaled sampling width. In particular, as shown in Section 4.5, in the worst case σ_f is $O(n^{-1/6})$ for large n. This implies that the effect of the bias error $m\sigma_f^2$ on the overall performance is typically mild.

4.5. **Analysis of the bounds for smooth embedding.** We now analyze the complexity of the parameters involved in the sampling analysis for large n (i.e., $n \to \infty$). We first observe the following for the perturbation terms β_2, β_3 and β_4 in Lemma 4:

$$\beta_2 = O(n^{1/2}m^2), \ \ \beta_3 = O(nm^{5/2}|\mathcal{K}_{max}|), \ \ \beta_4 = O(nm^3). \tag{4.5}$$

We now proceed by analyzing two different scenarios depending on the structure of the matrix D.

(1) **D is dense.** When the positive semidefinite matrix D is dense, the sampling width has complexity

$$\nu_{\text{bound,smooth}} = O(n^{-1/2}m^{-1}|\mathcal{K}_{max}|^{-1}).$$

This is similar to the bound in the quadratic embedding case. Next, we have the following complexities for the perturbation bounds $\| B_1 \|_{F,bound}$ and $\| D_1 \|_{F,bound}$.

$$\| B_1 \|_{F,bound} = O(n^{1/2}m^2\nu_{\text{bound,smooth}}^4) = O(n^{-3/2}m^{-2}|\mathcal{K}_{max}|^{-4})$$
$$\| D_1 \|_{F,bound} = O(nm^{5/2}\nu_{\text{bound,smooth}}^5|\mathcal{K}_{max}|) = O(n^{-3/2}m^{-5/2}|\mathcal{K}_{max}|^{-4})$$

Finally, we obtain the following complexity for the 'residual' angle bound term σ_f.

$$\sigma_f = O\left(\frac{\| B_1 \|_{F,bound}}{\nu_{\text{bound,smooth}}^2}\right) = O(n^{1/2}m^2\nu_{\text{bound,smooth}}^2) = O(n^{-1/2}|\mathcal{K}_{max}|^{-2}).$$

We observe that σ_f decays at a faster rate compared to the case where D is diagonal. The order of the dependency of the sampling width bound on n is higher in this case, which in turn implies a stricter bound on the high-order terms in the Taylor expansion. Finally, since the dependency of $\nu_{\text{bound,smooth}}$ on n, m, and $|\mathcal{K}_{max}|$ is the same as that of $\nu_{\text{bound,quad}}$, we obtain the same sampling complexity as in the quadratic embedding case:

$$K_{bound} = O(\tau^{-2}m^2 \log n) \quad \text{as} \quad n \to \infty.$$

Therefore, the number of samples has a quadratic dependence on the manifold dimension and a logarithmic dependence on the ambient space dimension.

(2) **D is diagonal.** It can be verified that the bound on the sampling width has the complexity

$$\nu_{\text{bound,smooth}} = O((\beta_3\alpha)^{-1/2}) = O(n^{-1/3}m^{-5/6}|\mathcal{K}_{max}|^{-1/3}).$$

This is in contrast to the quadratic embedding case, where we have seen that $\nu_{\text{bound,quad}}$ is independent of n. Moving on, we have the following complexities for the perturbation bounds $\| B_1 \|_{F,bound}$ and $\| D_1 \|_{F,bound}$:

$$\| B_1 \|_{F,bound} = O(n^{1/2}m^2\nu_{\text{bound,smooth}}^4) = O(n^{-5/6}m^{-4/3}|\mathcal{K}_{max}|^{-4/3}),$$
$$\| D_1 \|_{F,bound} = O(nm^{5/2}\nu_{\text{bound,smooth}}^5|\mathcal{K}_{max}|) = O(n^{-2/3}m^{-5/3}|\mathcal{K}_{max}|^{-2/3}).$$

From these orders of dependency, we arrive at the following complexity for the 'residual' angle bound term σ_f.

$$\sigma_f = O\left(\frac{\| B_1 \|_{F,bound}}{\nu_{\text{bound,smooth}}^2}\right) = O(n^{1/2}m^2\nu_{\text{bound,smooth}}^2) = O(n^{-1/6}m^{1/3}|\mathcal{K}_{max}|^{-2/3}).$$

Observe that σ_f decays with the increase in n, which is due to the decrease in $\nu_{\text{bound,smooth}}$. Notice also that σ_f gets smaller when the maximum curvature $|\mathcal{K}_{max}|$ increases. This can be intuitively explained as follows. It has been discussed in Section 3 that, as $f_{q,l}(\bar{x}) = f_{q,l}(-\bar{x})$, the normal components $f_{q,l}$ of the second-order terms constitute a symmetry around the origin. Meanwhile, the higher-order terms do not have such a symmetry in general, causing a bias on the tangent space estimation with PCA. The residual angle bound σ_f is associated with this bias resulting from the asymmetry of the normal components. As $|\mathcal{K}_{max}|$ increases, the second-order terms get more significant compared to the higher-order terms,

which strengthens the symmetry of the manifold and reduces the bias term. We now study the sampling complexity by analyzing $K_{bound}^{(1)}, K_{bound}^{(2)}$ and $K_{bound}^{(3)}$ separately. It can be easily verified that

$$K_{bound}^{(1)} = O(R_M \log n) = O(n^{1/3} m^{1/3} |\mathcal{K}_{max}|^{4/3} \log n),$$

$$\text{and} \quad K_{bound}^{(2)} = O\left(\frac{R_D}{RL} \log n\right) = O(n \log n).$$

Furthermore, by observing that

$$s_3 = O(\nu_{\text{bound,smooth}}^2 m^{-1/2} \tau) = O(n^{-2/3} m^{-13/6} |\mathcal{K}_{max}|^{-2/3} \tau), \tag{4.6}$$

we have

$$K_{bound}^{(3)} = O\left(\frac{\nu_{\text{bound,smooth}}^6 R_\sigma + \dfrac{R_B \nu_{\text{bound,smooth}}^3 s_3}{3}}{s_3^2} \log n\right)$$

$$= O(n^{1/3} m^{4/3} |\mathcal{K}_{max}|^{4/3} \log n).$$

Since $K_{bound} = \max\left\{K_{bound}^{(1)}, K_{bound}^{(2)}, K_{bound}^{(3)}\right\}$, we have

$$K_{bound} = O(K_{bound}^{(2)}) = O(n \log n) \quad \text{as} \quad n \to \infty.$$

Hence, the number of samples has a loglinear dependence on the ambient space dimension. In fact, although K_{bound} is chosen according to $K_{bound}^{(2)}$, this choice of K_{bound} implies that s_3 can be chosen up to the order $s_3 = O(n^{-1} m^{-3/2})$ by retaining the value of K_{bound}. Comparing this with the expression in (4.6), we see that the variance term is $\tau^2 = O(n^{-2/3} m^{4/3})$. Meanwhile, the bias term is $m\sigma_f^2 = O(n^{-1/3} m^{5/3} |\mathcal{K}_{max}|^{-4/3})$, which shows that the decay of the variance term with the increase in n is faster than the decay of the bias term. As we consider that n is large, we can neglect the variance term in comparison with the bias term. This gives

$$\tau^2 + m\sigma_f^2 = O(m\sigma_f^2) = O(n^{-1/3} m^{5/3} |\mathcal{K}_{max}|^{-4/3}).$$

The fact that the error resulting from finite sampling is negligible compared to the error due to the high-order Taylor terms when D is diagonal can be interpreted as follows. When D is diagonal, the samples can be chosen from a relatively wide region. Then, since the sampling width is large, the error in the estimation of the tangent space caused by the asymmetry in the geometric structure of the manifold dominates the error caused by finite sampling.

5. EXPERIMENTAL RESULTS

In this section we present experimental results for the empirical validation of the sampling conditions derived in the preceding sections. For the sake of brevity, we use the notation $\theta = \angle \widehat{T}_P S, T_P S$ to describe the angle between $T_P S$ and $\widehat{T}_P S$. Recall from Section 4 that, for any point P lying on a smooth m-dimensional manifold S in $\mathbb{R}^n$, the points lying in the neighborhood $\mathcal{N}_\varepsilon(P)$ of P have the following representation:

$$[\bar{x}^T \ f_1(\bar{x}) \ \dots \ f_{n-m}(\bar{x})]; \quad f_l : T_P S \to \mathbb{R}.$$

In the experiments, we study different manifold embeddings, where the functions f_l have the following form:

(1) *Quadratic form:* $f_l = \frac{1}{2} \sum_{j=1}^m \mathcal{K}_{l,j} x_j^2$

(2) *Smooth mapping 1:* $f_l = 1 - \exp\left(\frac{1}{2} \sum_{j=1}^m \mathcal{K}_{l,j} x_j^2\right)$

(3) *Smooth mapping 2:* $f_l = \sin\left(\frac{1}{2} \sum_{j=1}^m \mathcal{K}_{l,j} x_j^2\right)$

(4) *Smooth mapping 3:* $f_l = \sum_{j=1}^m \left(\frac{1}{2} \mathcal{K}_{l,j} x_j^2 + a_{l,j} x_j^3 + b_{l,j} x_j^4 + c_{l,j} x_j^5\right)$

In particular, we consider the mappings $\{f_l\}_{l=1}^{n-m}$ to be all of the same form. Furthermore, we focus on the general case where $\{f_{q,l}\}_{l=1}^{n-m}$ are correlated, or equivalently D is dense, which is the most generic scenario. Then, for a given value of $\mathcal{K}_{max}$, we select the principal curvatures $(\mathcal{K}_{l,1}, \dots, \mathcal{K}_{l,m})$ randomly from the interval $[0, |\mathcal{K}_{max}|]^m$ and then randomly assign the same sign ($+$ or $-$) to the elements of $\{\mathcal{K}_{l,j}\}_{j=1}^m$.

Furthermore for Smooth mapping 3, we select the coefficients $a_{l,j}, b_{l,j}$ and $c_{l,j}$ randomly from the interval $[0, 10]$ for $l = 1, \ldots, n - m$ and $j = 1, \ldots, m$.

We sample the points as explained in Section 4. We compute the tangent space with these samples points and compare the resulting estimation with the true tangent space by measuring the angle between both subspaces. Then we analyze the results from the perspective of the theoretical bounds on the width of the sampling regions and on the number of samples, which have been derived earlier in the paper. In particular, we consider the sampling width to have the value $\nu = \gamma \nu_{\text{bound,quad}}$, where

$$\nu_{\text{bound,quad}} := \sqrt{\frac{60}{m(n-m)(5m+4)|\mathcal{K}_{max}|^2}}.$$

The choice $\nu_{\text{bound,quad}}$ for the reference sampling width is due to the fact that it can be easily computed and it also provides a basis for comparing smooth embeddings with quadratic embeddings, made possible by varying the scale parameter γ.

In the first set of experiments, we examine the relation between the estimation error, i.e., the deviation $|\theta|$, and the sampling density K. We fix $m = 5$, $\mathcal{K}_{max} = 10$ and consider different values for the dimension of the ambient space, namely $n = 100, 500, 1000$. For each value of $(m, n, \mathcal{K}_{max})$, we choose the sampling width as a scaled version of theoretical bound, i.e., $\nu = \gamma \nu_{\text{bound,quad}}$. We estimate the tangent space with K samples and compute the approximation error with respect to the true tangent subspace. The results, shown in Fig. 3 have been averaged over 25 random trials for each value of K, where K is varied from 100 to 2000 in steps of 100.

We first show in Figures 3a-3c the results obtained for a quadratic embedding. We observe that for the choice $\nu \approx 1.2 \nu_{\text{bound,quad}}$, $|\theta|$ decreases sharply towards $0°$ with the increasing values of K. On the other hand, the choice $\nu \approx 4 \nu_{\text{bound,quad}}$ causes $|\theta|$ to increase towards $90°$. The results are similar across different values of n. Furthermore, since D is dense and $m = 5$ in this experiment, Theorem 1 states that $K = O(\tau^{-2})$ and $|\theta| = O(\cos^{-1}(1 - \tau^2)^{5/2})$. Therefore, the order of the dependence of $|\theta|$ on K is expected as $|\theta| = O(\cos^{-1}(1 - K^{-1})^{5/2})$. We can see that the plotted curves are in accordance with this theoretical result. We remark that, in these experiments, the true upper bound on ν appears to be within a factor γ of $\nu_{\text{bound,quad}}$, where γ takes a value between 1.2 and 4.

Figures 3d-3f, 3g-3i and 3j-3l then show the experimental results for non-quadratic embeddings, in particular, for Smooth mappings 1, 2 and 3 respectively. Interestingly, the variation of $|\theta|$ with respect to K for non-quadratic mappings is almost identical to those for quadratic forms.

The theoretical bounds stated in Theorems 3 and 4 are directly implementable for this experiment where the variation of $|\theta|$ with K is examined. Therefore, we now provide a comparison of the theoretical bounds and the empirical results for this setup. We obtain the theoretical variation of $|\theta|$ with K as follows. We first choose $s_1 = 0.5$ (as $0 < s_1 < 1$) and $s_2 = 2e$ (as $s_2 > e$). Then, in the quadratic embedding, we compute $\nu = \gamma \nu_{bound,quad}$, where $\gamma = c \sqrt{s_1/s_2}$. Here, $c < 1$ is a scale parameter guaranteeing that ν is strictly less than $\sqrt{s_1/s_2} \nu_{bound,quad}$. In the smooth embedding, we compute $\nu = \gamma \nu_{bound,quad}$ where $\gamma = c \nu_{bound,smooth}/\nu_{bound,quad}$ with $c < 1$ so that ν is strictly less than $\nu_{bound,smooth}$. In both quadratic and smooth embeddings, we fix the values of the parameters p_1, p_2, p_3 to 0.01 and vary τ from 0.01 to 0.2 in steps of 0.01. For each value of τ, we compute s_3 such that it is slightly smaller than $s_{3\text{bound,quad}}$ in the quadratic embedding and $s_{3\text{bound,smooth}}$ in the smooth embeddings, which are respectively given in (4.3) and (4.4). This gives a value of $K_{bound}^{(3)}$ and hence K_{bound}. The angle bound $|\theta| = \cos^{-1} \sqrt{(1 - \tau^2)^m}$ for the quadratic embedding is computed using Theorem 3 , and the angle bound $\cos^{-1} \sqrt{(1 - \tau^2 - m\sigma_f^2)^m}$ for the smooth embeddings is given by Theorem 4. Evaluating the bounds at four different values of the c parameter (0.2, 0.4, 0.6, 0.8 for the quadratic embedding, and 0.1, 0.2, 0.3, 0.4 for the smooth embeddings), we obtain four subplots showing the variation of K_{bound} with $|\theta|$, each for a different value of ν. The results are given in Fig. 4 for the quadratic embedding and Figures 5, 6 and 7 for the smooth embeddings. The theoretical plots obtained for $n = 100, 500, 1000$ are displayed respectively in (a)-(c) in these figures. The experimental curves corresponding to these theoretical plots are then obtained by sampling the manifolds in the region $\nu = \gamma \nu_{bound,quad}$ for the same value of γ as in the theoretical plots, which are shown in the plots (d)-(f) of Figures 4-7.

The comparison of the theoretical bounds and the experimental plots given in Figures 4-7 shows the following. While the numerical values of the theoretical bounds obtained for the angle error are pessimistic

in comparison with the experimental values, we see that the theoretical variation of $|\theta|$ with K matches well the experimental one in both quadratic and smooth embeddings. Therefore, the theoretical results provide a good prediction of the dependence of the angle error on the number of samples.
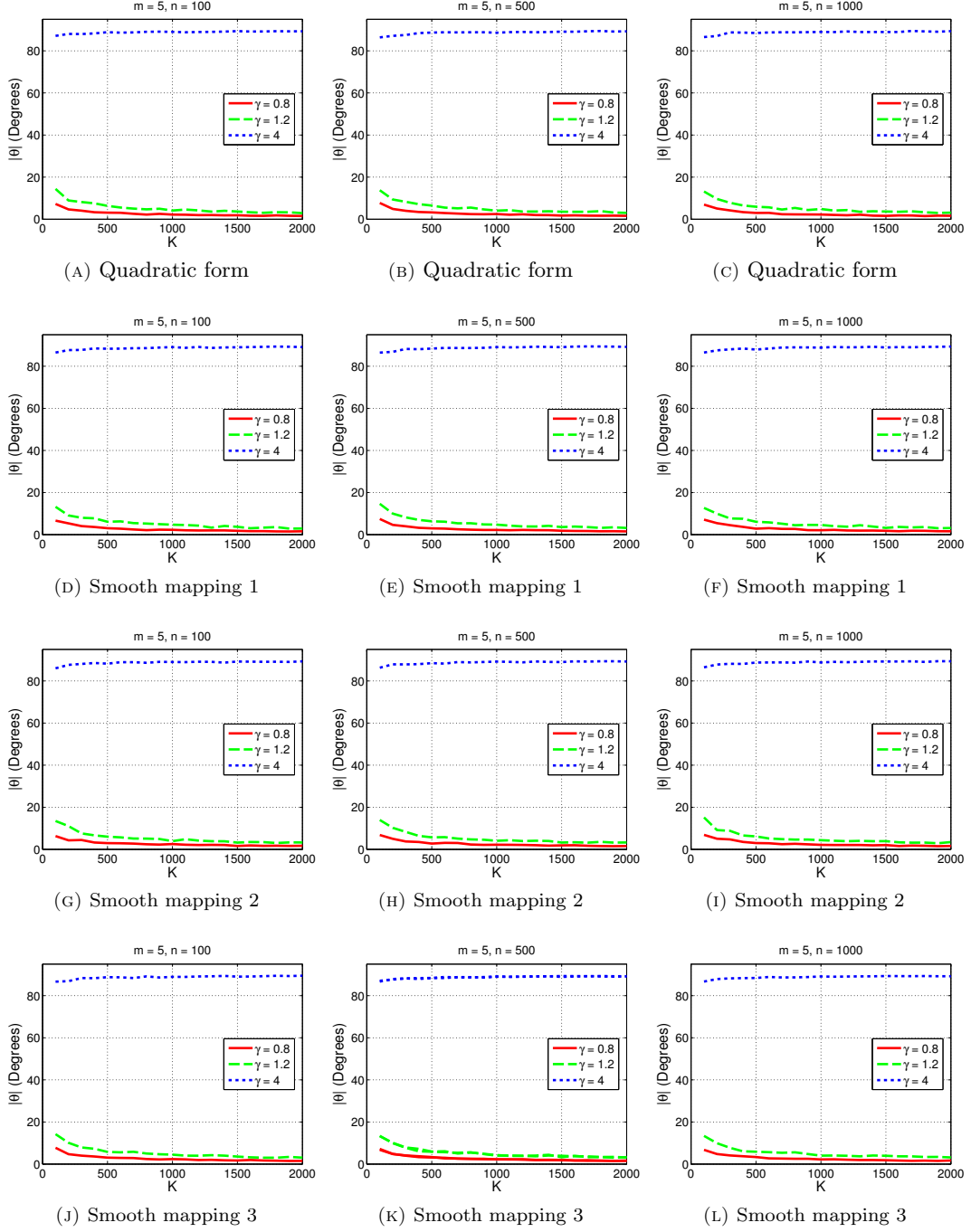

FIGURE 3. Variation of the deviation $|\theta|$ with respect to K for different sampling widths ν. For each type of mapping, $\nu = \gamma \nu_{\text{bound,quad}}$.

In the second set of experiments, we study the scaling of the true bound on the sampling width ν with the ambient space dimension n. To this end, we fix $m = 5$, $\mathcal{K}_{max} = 10$ and the number of samples is fixed at a sufficiently large value, i.e., $K = 2000$. Then, we vary n from 100 to 1000 in steps of 50. For each value of n,
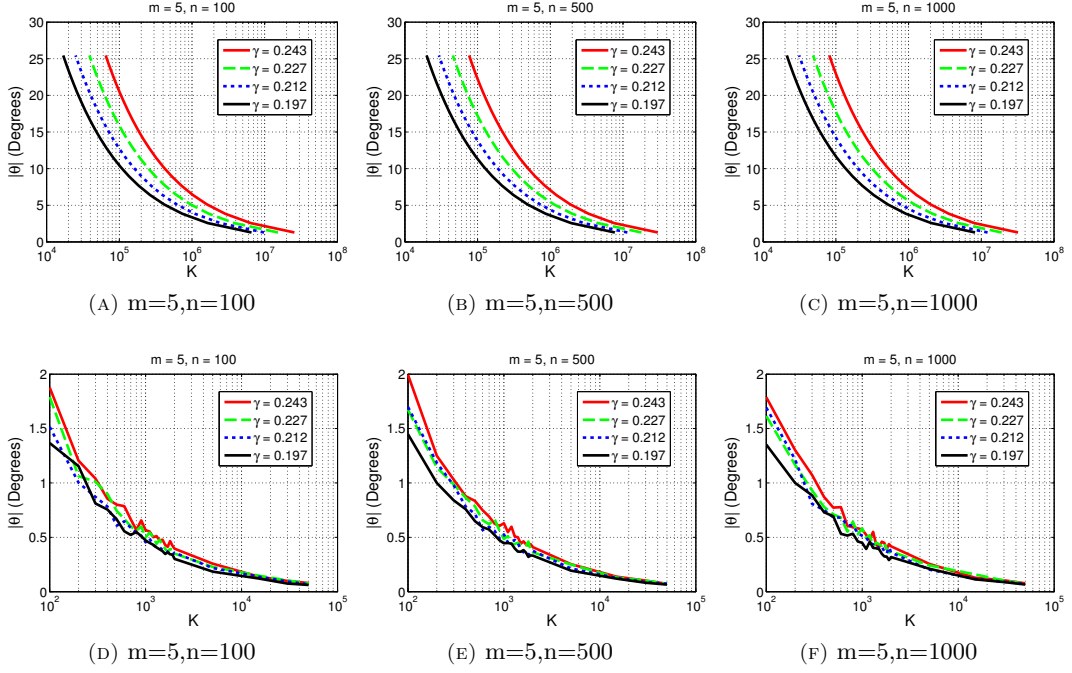
FIGURE 4. Variation of the deviation $|\theta|$ with respect to K for different sampling widths $\nu = \gamma\nu_{\text{bound,quad}}$ for quadratic embedding. Figures 4a-4c show theoretical plots while Figures 4d-4f show empirical plots.
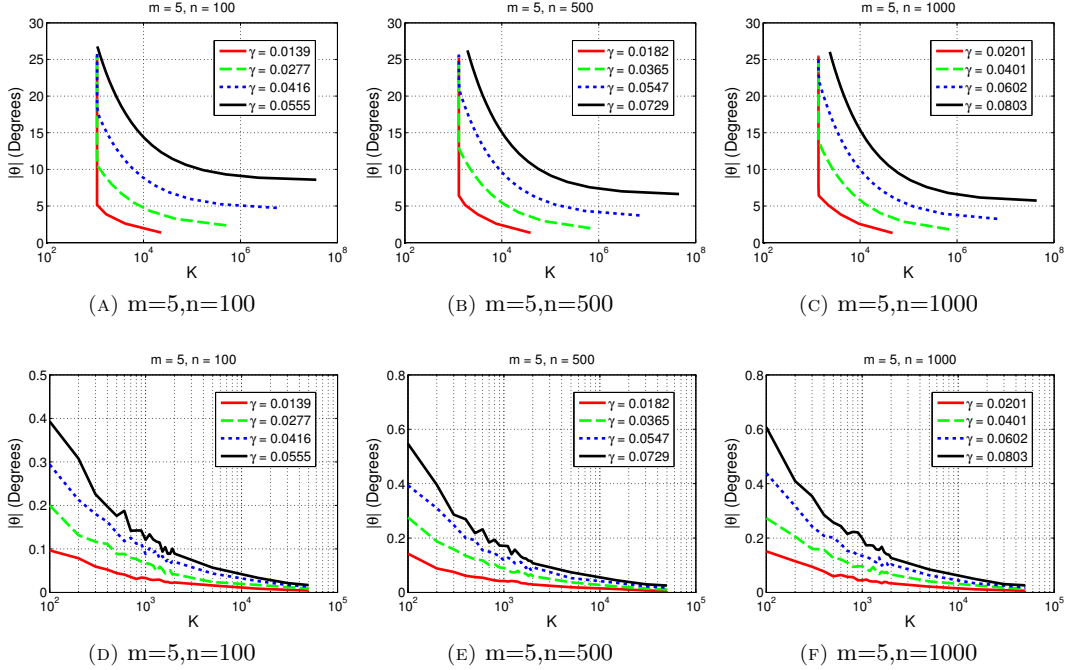

FIGURE 5. Variation of the deviation $|\theta|$ with respect to K for different sampling widths $\nu = \gamma\nu_{\text{bound,quad}}$ for smooth mapping 1. Figures 5a-5c show theoretical plots while Figures 5d-5f show empirical plots.

(A) m=5,n=100   (B) m=5,n=500   (C) m=5,n=1000

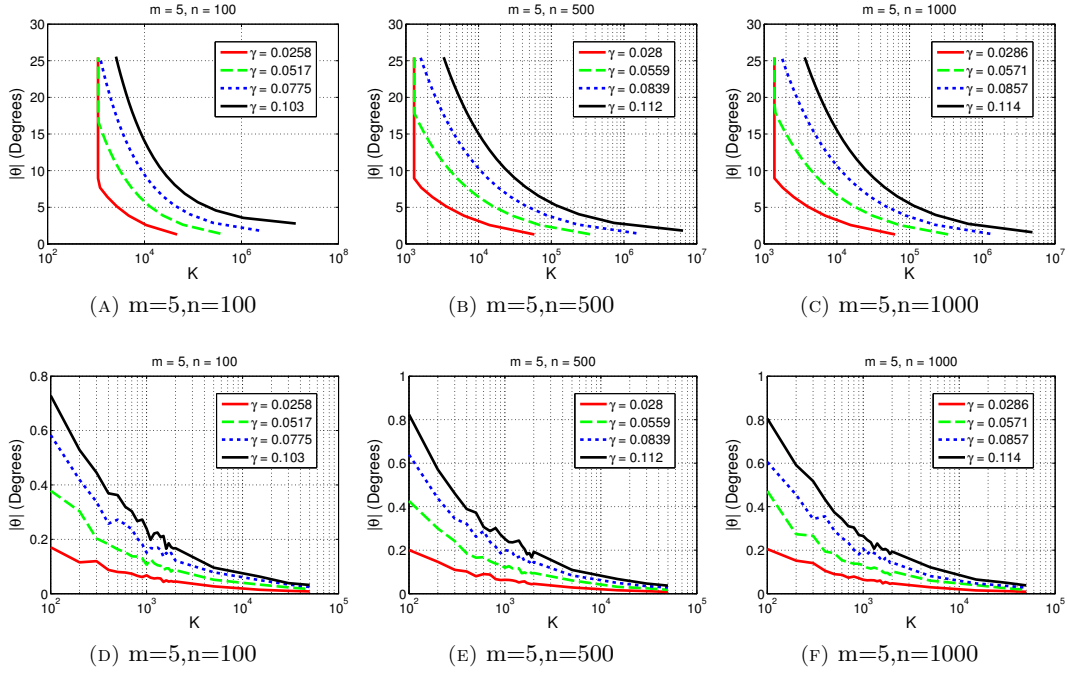(D) m=5,n=100   (E) m=5,n=500   (F) m=5,n=1000

FIGURE 6. Variation of the deviation $|\theta|$ with respect to K for different sampling widths $\nu = \gamma \nu_{\text{bound,quad}}$ for smooth mapping 2. Figures 6a-6c show theoretical plots while Figures 6d-6f show empirical plots.

(A) m=5,n=100   (B) m=5,n=500   (C) m=5,n=1000

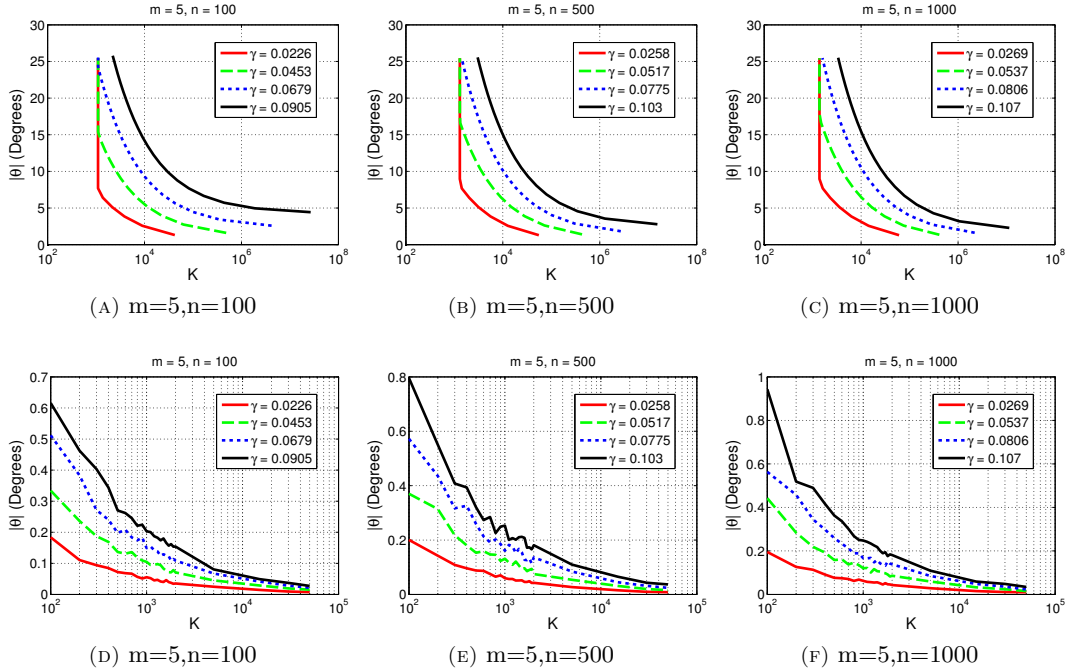(D) m=5,n=100   (E) m=5,n=500   (F) m=5,n=1000

FIGURE 7. Variation of the deviation $|\theta|$ with respect to K for different sampling widths $\nu = \gamma \nu_{\text{bound,quad}}$ for smooth mapping 3. Figures 7a-7c show theoretical plots while Figures 7d-7f show empirical plots.

we first initialize $\nu = 3\,\nu_{\text{bound,quad}}$ and then compute the bound on the sampling width by gradually reducing ν until $|\theta| < \theta_{\text{bound}}$. The value of $|\theta|$ is averaged over 25 random trials. We obtain four plots corresponding to the angle bounds $\theta_{\text{bound}} = 5^\circ, 10^\circ, 15^\circ$ and 20°, for the quadratic form and smooth mappings 1 and 2.

Fig. 8 shows the variation of ν with n. Importantly, we have observed that, for quadratic forms (Fig. 8a), the true bound ν on the sampling width is in line with its theoretical estimation $\nu_{\text{bound,quad}}$ as $\nu = \gamma\,\nu_{\text{bound,quad}}$, where γ lies approximately between 1.38 and 1.46. This is also true for smooth mappings (Figures 8b, 8c) indicating that the true bound on ν is approximately $O(n^{-1/2})$. It can also be observed that, at a fixed value of n, the bound on ν is larger when the angle bound θ_{bound} is greater, as expected.
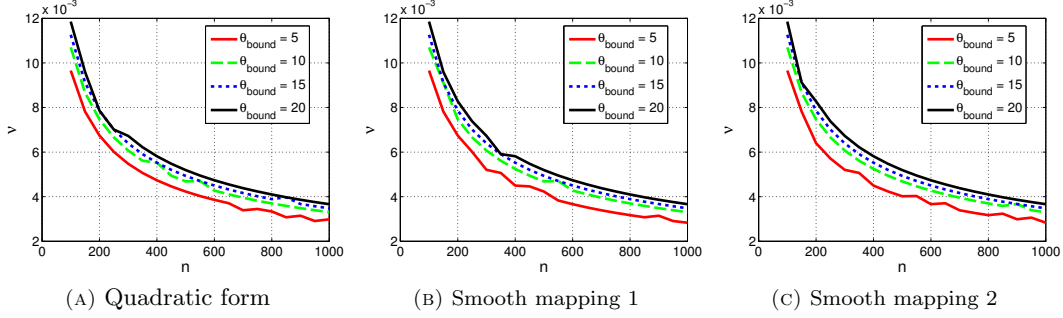

FIGURE 8. Maximum sampling width ν for which the deviation $|\theta| < \theta_{\text{bound}}$ is achieved for different values of n. Plots are shown for $\theta_{\text{bound}} = 5^\circ, 10^\circ, 15^\circ, 20^\circ$.

In the next experiment, we are interested in observing the dependency of the true sampling width bound ν on the maximum local curvature $\mathcal{K}_{max}$. We fix $m = 5$, set $K = 2000$ and choose a fixed $n \in \{100, 500, 1000\}$. We vary $\mathcal{K}_{max}$ from 0.5 to 10 in steps of 0.5. For each value of $\mathcal{K}_{max}$, we first initialize $\nu = 3\,\nu_{\text{bound,quad}}$ and then compute the bound on ν by gradually reducing ν until $|\theta| < |\theta_{\text{bound}}| = 5^\circ$. The value of $|\theta|$ is averaged over 25 random trials.

Fig. 9 shows the dependency of ν on $\mathcal{K}_{max}$ for different values of n. Similarly to the previous experiments, we observe that for quadratic forms, $\nu = \gamma\,\nu_{\text{bound,quad}}$, where γ is approximately between 1.32 and 1.46 (see Fig. 9a). We note the same behavior for the case of smooth mappings shown in Figures 9b and 9c. The results indicate that for fixed values of m and n, the true bound on ν matches the theoretical result $\nu \approx O(|\mathcal{K}_{max}|^{-1})$ derived in Section 4.5.
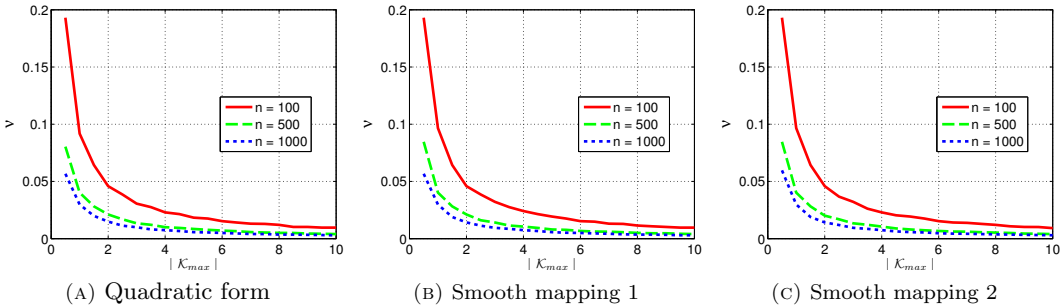

FIGURE 9. Maximum sampling width ν for which $|\theta| < |\theta_{\text{bound}}| = 5^\circ$ is achieved for different values of $\mathcal{K}_{max}$. The results are given for different dimensions of the ambient space n.

Then, we investigate the relation between the sampling density K and the embedding dimension n for a fixed sampling width ν. We choose $\mathcal{K}_{max} = 10$ and select several values for the dimension of the manifold $m \in \{5, 10, 15\}$. We vary n from 100 to 2000 in steps of 100. We set $\nu = \nu_{\text{bound,quad}}$, where $\nu_{\text{bound,quad}}$ is evaluated at the largest value of n such that the fixed sampling width ν is sufficiently small for the range of n under consideration. We denote the largest value of n by n_{large} in this experiment. For each value of

$(m, n, \mathcal{K}_{max})$, we compute the minimum number of samples needed in order to have $|\theta| < |\theta_{\text{bound}}| = 5°$. The value of $|\theta|$ is the average of 25 random trials.
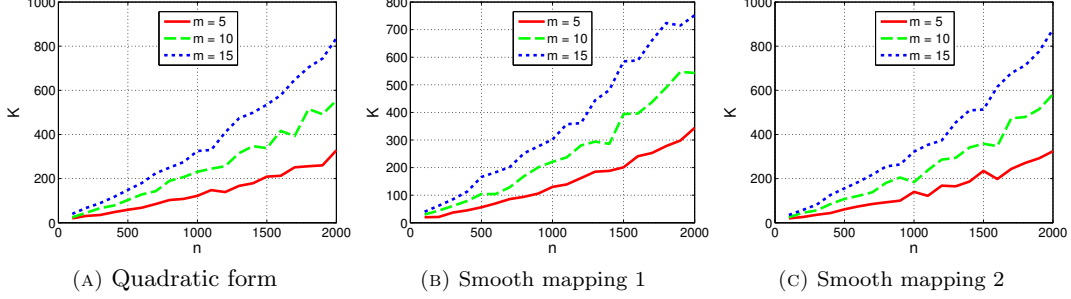

(A) Quadratic form   (B) Smooth mapping 1   (C) Smooth mapping 2

FIGURE 10. Minimum number of samples K for which $|\theta| < |\theta_{\text{bound}}| = 5°$ is achieved as n is varied. The results obtained for different dimensions of the manifold m.

Fig. 10 shows the variation of K with respect to n for the different mappings. We see that K increases with the ambient dimension n as expected. Furthermore, for a given n, we observe that increasing the dimension of the manifold increases K. We now show that this behaviour is well explained by our theoretical results in Section 4. In order to see this, we first note that $\nu = O(n_{\text{large}}^{-1/2} m^{-1} \mathcal{K}_{max}^{-1})$, which is due to the relation $\nu = O(n^{-1/2} m^{-1} \mathcal{K}_{max}^{-1})$ derived in Section 4.3 and the fact that we evaluate ν at $n = n_{\text{large}}$. Using this value of ν in the bounds on the sampling density stated in Lemma 3, one can easily verify that

$$K_{\text{bound}}^{(1)} = O\left(\left(m + \frac{n}{n_{\text{large}}}\right) \log n\right), \quad K_{\text{bound}}^{(2)} = O(\log n) \quad \text{and}$$

$$K_{\text{bound}}^{(3)} = O\left(\left(\frac{nm}{n_{\text{large}}} + \sqrt{\frac{nm}{n_{\text{large}}}}\right) \log n\right) \approx O\left(m\sqrt{n/n_{\text{large}}} \log n\right).$$

Since $n < n_{\text{large}}$, we obtain $K_{\text{bound}} = O\left(\left(m + \frac{n}{n_{\text{large}}}\right) \log n\right)$. This closely matches the behavior shown in Figures 10a-10c.

Finally, in the last experiment, we would like to look into the dependency of the sampling density K on the curvature term $\mathcal{K}_{max}$ for a fixed sampling width ν. We set $m = 5$ and pick several values of the ambient space dimension, i.e., $n \in \{100, 500, 1000\}$. We vary $\mathcal{K}_{max}$ from 0.5 to 10 in steps of 0.5. We fix the value of the sampling width as $\nu = \nu_{\text{bound,quad}}$, where $\nu_{\text{bound,quad}}$ is evaluated at the largest value of $\mathcal{K}_{max}$ (denoted as $\mathcal{K}_{\text{max,large}}$ in this experiment). For each value of $(m, n, \mathcal{K}_{max})$, we compute the minimum number of samples required in order to have $|\theta| < |\theta_{\text{bound}}| = 5°$. The value of $|\theta|$ is averaged over 25 random trials.

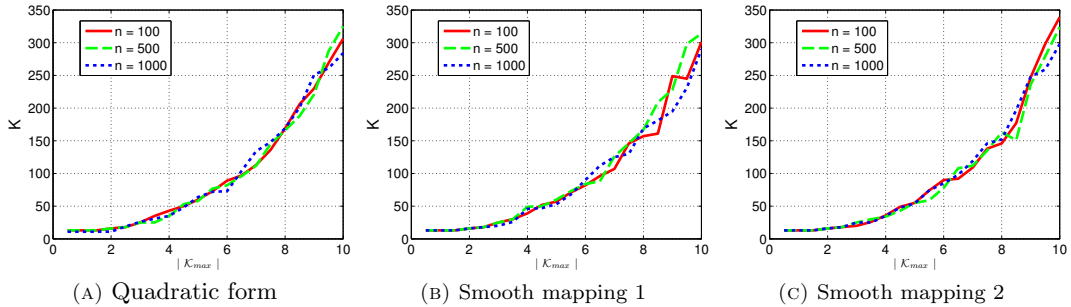(A) Quadratic form   (B) Smooth mapping 1   (C) Smooth mapping 2

FIGURE 11. Minimum number of samples K for which $|\theta| < |\theta_{\text{bound}}| = 5°$ is achieved as $\mathcal{K}_{max}$ is varied. The results are given for different values of n.

Fig. 11 shows the relation between K and $|\mathcal{K}_{max}|$ for the different mappings. We see that K increases with $|\mathcal{K}_{max}|$ as expected. Interestingly, we see that, for a fixed value of $\mathcal{K}_{max}$, a change in the embedding dimension n does not significantly affect K. We now show that such a variation of K with $|\mathcal{K}_{max}|$ for a fixed sampling width is explained by the theoretical results in Section 4. We first note that the sampling width is $\nu = O(n^{-1/2}m^{-1}|\mathcal{K}_{\mathrm{max,large}}|^{-1})$, which can be obtained from the results of Section 4.3 by evaluating the bounds on the sampling density at $\mathcal{K}_{max} = \mathcal{K}_{\mathrm{max,large}}$. From Lemma 3, one can then easily verify that

$$K_{\mathrm{bound}}^{(1)} = O\left(\left(m + \frac{\mathcal{K}_{max}^2}{\mathcal{K}_{\mathrm{max,large}}^2}\right)\log n\right), \quad K_{\mathrm{bound}}^{(2)} = O(\log n) \quad \text{and}$$

$$K_{\mathrm{bound}}^{(3)} = O\left((m\mathcal{K}_{max}^2\mathcal{K}_{\mathrm{max,large}}^{-2} + m^{1/2}|\mathcal{K}_{max}||\mathcal{K}_{\mathrm{max,large}}|^{-1})\log n\right) \approx O\left(m|\mathcal{K}_{max}|/|\mathcal{K}_{\mathrm{max,large}}|\log n\right).$$

As $|\mathcal{K}_{max}| < |\mathcal{K}_{\mathrm{max,large}}|$, we have $K_{\mathrm{bound}} = O\left(\left(m + \frac{\mathcal{K}_{max}^2}{\mathcal{K}_{\mathrm{max,large}}^2}\right)\log n\right)$. Thus, for a fixed n, the bound on K increases quadratically with $|\mathcal{K}_{max}|$, which is consistent with the curves presented in Figures 11a-11c. Furthermore, K_{bound} depends only logarithmically on n, suggesting that a change in n would affect the sampling density only *mildly*: this also matches the experimental results.

6. Discussion

In this section, we first discuss our results in view of the recent works from the literature. Then we show how our results could be used in practical applications.

We first position our study relatively to the works presented in [19] and [21], which are, to the best of our knowledge, the closest to our paper. In [19] the authors consider a global sampling from a compact manifold and relate the size of the neighborhood ε to the number of samples K through the condition $\varepsilon = O(K^{-\frac{1}{m+2}})$. From this aspect, our approach is significantly different. Our bound on ε is derived in the asymptotic limit where $K \to \infty$, so that it depends completely on the local manifold geometry. Furthermore, the analysis in [19] gives soft bounds that do not reflect the effect of the curvature, nor of the ambient space and manifold dimensions on the sampling conditions. Meanwhile, we derive worst-case bounds on both ε and K by explicitly taking into account the effect of curvature and dimensions.

The work in [21] is parallel to ours and addresses a similar problem. The analysis is however clearly different in two main aspects. Firstly, the analysis in [21] assumes that the manifold is embedded with exactly quadratic forms and that the data consists of samples from the quadratic manifold corrupted with Gaussian noise. On the contrary, the type of the manifolds that we consider is more generic as we assume an embedding of the manifold with arbitrary smooth functions. In particular, we explicitly examine the effect of the deviation of the manifold from its second-order approximation on the accuracy of the tangent space estimation. Secondly, an important difference between both studies is that the data is already sampled in [21], where the problem consists of choosing the size of the subset of samples used in the tangent space estimation, while we assume that we have a rather direct control on the parameters of the local random sampling (sampling width and number of samples). Therefore, in [21], the number of samples N (which is K in our notation) and the sampling radius r (which is comparable to the sampling width ν in our notation) are directly dependent on each other. As the sampling is formulated as a subset selection problem, increasing the number of samples necessarily leads to choosing samples from a larger radius. The analysis is based on the assumption $r = c\,N^{1/d}$, where c is a constant and d is the dimension of the manifold (m with our notation); therefore, r and N can be represented in terms of a single parameter. Meanwhile, in our analysis, we consider a setting where we treat the sampling width ν and number of samples K as two different parameters.

Even if the frameworks in [21] and in this paper are quite different, we can try to compare results. It is assumed in [21] that the subset of samples selected for tangent space estimation corresponds to a sampling radius smaller than a threshold r_{max}, where r_{max} is the largest radius within which the manifold can be accurately represented with quadratic forms. We give a characterization of such a bound on the sampling width in Lemma 4 for arbitrary smooth manifolds, which is very relevant to the parameter r_{max} in their work. In [21], the parameter r_{max} is used as a predetermined constant and the study does not go into the analysis of r_{max} for non-quadratic manifolds. A direct comparison of the main results in both papers is difficult. However, we can compare the noiseless version of the Interpretable Main Result 1 in [21] and our results on quadratic manifolds in the following way. The denominator of the angle bound in Interpretable Main Result 1 quantifies the separation between the tangential and normal components of the computed eigenspace.

Furthermore, the sampling radius must be small enough to guarantee that the eigenvalues corresponding to the tangential components must be larger than those corresponding to the normal components. Then, an admissible sampling radius must be below the value of r that equates the denominator of the expression in Interpretable Main Result 1 to zero. Taking the noise variance as zero and observing the relation $K = O(n^{1/2} m |\mathcal{K}_{max}|)$, where K is the curvature parameter in [21], their result translates into the fact that the admissible sampling radius must be smaller than $O(n^{-1/2} m^{-1/2} |\mathcal{K}_{max}|^{-1})$ with our notation, where m, n and $|\mathcal{K}_{max}|$ are the parameters corresponding respectively to the intrinsic manifold dimension, the ambient space dimension and the curvature. This is in agreement with our result for quadratic embeddings (see Table 1), where we have calculated the admissible sampling width as $O(n^{-1/2} m^{-1} |\mathcal{K}_{max}|^{-1})$.

Now that our work has been properly positioned with respect to the related work, we discuss the usage of our results in practical applications. We can interpret our results in two important application areas, namely (i) the discretization of a manifold with a known parametric model - *manifold sampling* and (ii) the recovery of the tangent space of a manifold from a given set of data samples - *manifold learning*.

First, in order to use our results in a real application, the intrinsic dimension m of the manifold, the curvature parameter $\mathcal{K}_{max}$, and the higher-order deviation term C_s have to be known or estimated. In a manifold sampling application, m is already known and it is possible to estimate $\mathcal{K}_{max}$ in the following ways. If the manifold conforms to a known analytic model, it is easy to compute the values of the principal curvatures and the higher-order terms from the Taylor expansion of the model. If an analytic model is not known for the manifold, the curvature of a manifold of known parameterization can be estimated using results from Riemannian geometry such as [27] (Section V) and [28] (Proposition 2). The results in Section V of [27] are especially compatible with our definition of curvature, where we define $\mathcal{K}_{max}$ as the largest of the maximum principal curvatures of the hypersurfaces $\mathcal{S}_l$, $l = 1, \ldots, n - m$, each of which have a single normal direction. Although the work in [27] addresses an image registration problem, the analysis in Section V of [27] is generic and it describes a procedure to compute the maximum principal curvature of a manifold corresponding to a single normal direction, which is equal to the norm of the second fundamental form corresponding to the normal direction. Applying this procedure for all $n - m$ normal directions and taking the largest one of the maximum principal curvatures, one can compute the exact value of $\mathcal{K}_{max}$. Then, the deviation term C_s is the maximum of the constants $C_{s,l}$. Once the maximum principal curvature of $\mathcal{S}_l$ is computed as above, one can find a suitable bound for $C_{s,l}$ by looking at the deviation of $\mathcal{S}_l$ from its second order approximation.

Second, in a manifold learning application where only data samples are available, m, $\mathcal{K}_{max}$ and C_s are unknown and need to be estimated. The estimation of the intrinsic dimension of a data set has been studied in several works such as [29], [30] and [31]. It is also possible to obtain an estimate of the curvature from data samples using results such as in [32]. In [32], a method is proposed to estimate the intrinsic dimension of the manifold by examining the variation of the singular values of the data covariance matrix with respect to the radius of the neighborhood of samples used. It is observed that the singular values corresponding to the curvatures can be distinguished from the singular values corresponding to the tangential components by using the fact that the tangential and curvature singular values conform respectively to linear and quadratic fits as a function of the radius. In such a setting, the deviation of the curvature singular values from their quadratic fits for large values of the radius can possibly be related to the deviation term C_s.

Finally, in our results, we characterize the admissible sampling width for accurate tangent space estimation in terms of the tangent space distances, i.e., the distances between the projections of points on the tangent space and P. In a manifold sampling application, our analysis can be easily adapted to the parametric data model at hand since it assumes that the true tangent space of the manifold is aligned with the subspace generated by the first m canonical basis vectors. This can be achieved by applying a Gram-Schmidt orthonormalization to the tangent vectors of the data manifold and then performing a change of coordinates in $\mathbb{R}^n$ such that the subspace spanned by the original tangent vectors is mapped to the subspace generated by the first m canonical basis vectors. Meanwhile, in a manifold learning application where only data samples are available, one needs to adapt the bounds on the tangent space distance to bounds on the distance between actual data samples in the ambient space. This can be done in different ways. Based on our results, one can easily obtain some worst-case bounds on the ambient space distance by making use of the fact that the tangent space distance is upper bounded by the ambient space distance. This approach is expected to be effective if the ambient space dimension n is comparable to the intrinsic dimension m, or if the manifold has small curvature. Alternatively, if $n \gg m$ and the manifold has significant nonlinearity, the current results involving

the tangent space distance can be translated into approximate conditions on the ambient space distance with the help of the estimation $\|.\|_{ambient\ space} \approx O(\|.\|_{tangent\ space}\sqrt{n/m})$. Note that, using this estimation, the decay of the sampling width ν in the tangent space at the rate $O(n^{-1/2}m^{-1}|\mathcal{K}_{max}|^{-1})$ implies that the same width measured in the ambient space must change at the rate $O(\nu\sqrt{n/m}) = O(m^{-3/2}|\mathcal{K}_{max}|^{-1})$. Therefore, the sampling width in the ambient space does not decrease with the ambient space dimension. It is of $O(1)$ with respect to n; meanwhile, it decreases with m and $|\mathcal{K}_{max}|$. This means that, when applying PCA, the size of the neighborhood around a reference point in the ambient space must get smaller as the intrinsic dimension or the curvature of the manifold increases.

In this work, we have focused on a noiseless data model that is perfectly representable with smooth functions. However, in real applications, one may need to work with noisy data samples that exhibit a deviation from the manifold. One can possibly extend the study presented here to include the effect of noise in the analysis. This can be achieved by first identifying the sampling region for an accurate estimation of the tangent space and then determining a sufficient sampling density in that region. The admissible sampling region highly depends on the type of noise. One would expect to have no bias in the estimation for a random noise model with spherical symmetry, while a structured noise model may bias the estimation and necessitate stricter constraints on the sampling width. Then, the sampling density is expected to be affected by the variance of the noise. These effects can be characterized by studying the additional perturbation on the correlation matrices due to the noise.

7. Concluding Remarks

We have presented a theoretical analysis of the tangent space estimation at a point on a submanifold from a set of manifold samples that are selected locally at random. We have considered a setting where the manifold is embedded smoothly in $\mathbb{R}^n$ and the tangent space is estimated with local PCA. We have derived relations between the accuracy of the tangent space estimation and the sampling conditions. In particular, we have examined the effect of the local curvature of the manifold in tangent space estimation and shown that the size of the sampling neighborhood shall be inversely proportional to the manifold curvature. We have also seen that sampling conditions are affected by the correlation between the components of the second-order approximation of the embedding. The sampling width can be chosen larger when the components of the manifold in different dimensions are less correlated. The presented study can be used for obtaining performance guarantees in the discretization of parametrizable data and in manifold learning applications. Finally, our analysis assumes that the data samples are noiseless, i.e., the data lies exactly on the manifold. A future research direction resides therefore in the extension of the current results to a scenario where data samples are corrupted with noise.

8. Acknowledgments

References

[1] H. Tyagi. Local Sampling Analysis for Quadratic Embeddings of Riemannian Manifolds. Master's thesis, Ecole Polytechnique Fédérale de Lausanne, July 2011. Available: http://infoscience.epfl.ch/record/179897.

[2] E. Vural and P. Frossard. Discretization of Parametrizable Signal Manifolds. *IEEE Transactions on Image Processing*, 20(12):3621–3633, 2011.

[3] J.B. Tenenbaum, V.D. Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

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

[5] D. L. Donoho and C. E. Grimes. Hessian eigenmaps: Locally linear embedding techniques for highdimensional data. *Proc. Natl. Acad. Sci. USA*, 100:5591–5596, 2003.

[6] T. Lin, H. Zha, and S.U. Lee. Riemannian manifold learning for nonlinear dimensionality reduction. In *Proc. of Eur. Conf. Computer Vision*, 2006.

[7] Z. Zhang, J. Wang, and H. Zha. Adaptive Manifold Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(2):253–265, February 2012.

[8] H. Zha and Z. Zhang. Spectral properties of the alignment matrices in manifold learning. *SIAM Review*, 51:545–566, 2009.

[9] Z. Zhang and H. Zha. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM Journal of Scientific Computing*, 26:313–338, 2005.

[10] Y. Yang, F. Nie, S. Xiang, Y. Zhuang, and W. Wang. Local and global regressive mapping for manifold learning with out-of-sample extrapolation. In *Proc. of the 24th AIII Conf. on Artificial Intelligence*, 2010.

[11] Y. Zhan, J. Yin, G. Zhang, and E. Zhu. Incremental manifold learning algorithm using PCA on overlapping local neighborhoods for dimensionality reduction. In *Advances in Computation and Intelligence*, volume 5370 of *Lecture Notes in Computer Science*, pages 406–415. Springer Berlin/Heidelberg, 2008.

[12] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation, III. *SIAM J. Numer. Anal.*, 7, March 1970.

[13] P.A. Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12:99–111, 1972. 10.1007/BF01932678.

[14] V. Vu. Singular vectors under random perturbation. *Random Struct. Algorithms*, 39(4):526–538, December 2011.

[15] N. M. Faber, M. J. Meinders, P. Geladi, M. Sjöström, L. M. C. Buydens, and G. Kateman. Random error bias in principal component analysis. Part I. derivation of theoretical predictions. *Analytica Chimica Acta*, 304(3):257–271, 1995.

[16] T. W. Anderson. Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics*, 34(1):122–148, 1963.

[17] D. N. Lawley. Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, 43(1-2):128–136, June 1956.

[18] M. A. Girshick. On the sampling theory of roots of determinantal equations. *The Annals of Mathematical Statistics*, 10(3):203–224, 1939.

[19] A. Singer and H. Wu. Vector Diffusion Maps and the Connection Laplacian. *Comm. on Pure and App. Math.*, 2012.

[20] R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni, F. Warner, and S. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. In *Proceedings of the National Academy of Sciences*, pages 7426–7431, 2005.

[21] D. Kaslovsky and F.G. Meyer. Optimal tangent plane recovery from noisy manifold samples. Submitted to the Annals of Statistics, available at http://arxiv.org/abs/1111.4601v2.

[22] A. Gittens and J. Tropp. Tail bounds for all eigenvalues of a sum of random matrices. *Preprint*, 2011.

[23] J. Tropp. User-friendly tail bounds for sums of random matrices. *Preprint*, 2011.

[24] G.H. Golub and van Loan C.F. Matrix computations. The Johns Hopkins University Press, Baltimore, 1996.

[25] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with confidence from random samples. *Discrete and Computational Geometry*, 2006.

[26] H. Gunawan, O. Neswan, and W. Setya-Budhi. A formula for angles between subspaces of inner product spaces. *Contributions to Algebra and Geometry*, 46:311–320, 2005.

[27] E. Kokiopoulou, D. Kressner, and P. Frossard. Optimal image alignment with random projections of manifolds: algorithm and geometric analysis. *IEEE Transactions on Image Processing*, 20(6):1543–1557, 2011.

[28] L. Jacques and C. De Vleeschouwer. A geometrical study of matching pursuit parametrization. *IEEE Transactions on Signal Processing*, 56(7):2835–2848, July 2008.

[29] M. Hein. Intrinsic dimensionality estimation of submanifolds in Euclidean space. In *Proceedings of the 22nd International Conference on Machine Learning*, pages 289–296, 2005.

[30] E. Levina and P.J. Bickel. Maximum likelihood estimation of intrinsic dimension. In *Advances in Neural Information Processing Systems*, 2005.

[31] G. Chen, A.V. Little, M. Maggioni, and L. Rosasco. Some recent advances in multiscale geometric analysis of point clouds. *Wavelets and Multiscale Analysis: Theory and Applications*, March 2011.

[32] A.V. Little, J. Lee, Y.M. Jung, and M. Maggioni. Estimation of intrinsic dimensionality of samples from noisy low-dimensional manifolds in high dimensions with multiscale SVD. In *Proc. of S.S.P.*, 2009.

[33] H. Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen*, 71:441–479, 1912.

APPENDIX A. m-DIMENSIONAL SMOOTH MANIFOLDS IN $\mathbb{R}^n$

A.1. **Proof of Lemma 2.**

Proof. Observe that each entry of $M^{(K)}$ is the sum of K i.i.d. random variables. Therefore, by the Strong Law of Large Numbers as $K \to \infty$, $[M^{(K)}]_{i,j}$ converges a.s. to $[M]_{i,j}$ for all $1 \leq i, j \leq n$, where each entry of M is the expected value of the random variable involved in the summation of the corresponding entry of $M^{(K)}$. Let

$$M = \begin{bmatrix} A & B \\ B^T & D \end{bmatrix}.$$

Consider the entries of A. We have for $j, k = 1, \ldots, m$,

$$[A]_{j,k} = \mathbb{E}[x_j x_k] = \begin{cases} 0 & \text{if } j \neq k \\ \frac{\nu^2}{3} & \text{if } j = k \end{cases}$$

Consider the entries of B. We have for $j = 1, \ldots, m$ and $l = 1, \ldots, n - m$,

$$
[B]_{j,l} \;=\; \mathbb{E}[x_j f_l(\bar{x})] = \mathbb{E}[x_j \frac{1}{2} \sum_{k=1}^{m} < \bar{x}, \bar{v}_{l,k} >^2 \mathcal{K}_{l,k}]
$$

$$
= \frac{1}{2} \mathbb{E}[x_j \sum_{k=1}^{m} (x_1 v_{l,k,1} + \cdots + x_1 v_{l,k,m})^2 \mathcal{K}_{l,k}] = 0.
$$

The above result follows as each term in the expansion of $x_j f_l(\bar{x})$ has at least one odd power of x_j, and the expected value of each term is thus 0. Now, consider the diagonal entries of D. We have

$$
[D]_{l,l} \;=\; \mathbb{E}[f_l^2(\bar{x})] = \frac{1}{4} \mathbb{E}\left[\left(\sum_{j=1}^{m} < \bar{x}, \bar{v}_{l,j} >^2 \mathcal{K}_{l,j} \right)^2 \right] \le \frac{1}{4} |\mathcal{K}_{max}|^2 (\mathbb{E}[\| \bar{x} \|_2^4]).
$$

Furthermore,

$$
\mathbb{E}[\| \bar{x} \|_2^4] = \mathbb{E}[\sum_{j=1}^{m} x_j^4 + 2 \sum_{k<j} x_k^2 x_j^2] = \frac{m\nu^4}{5} + 2\frac{m(m-1)}{2} \left(\frac{\nu^2}{3} \right)^2 = \frac{m(5m+4)\nu^4}{45}.
$$

Hence

$$
0 \;\le\; [D]_{l,l} \;\le\; \frac{m(5m+4)\nu^4}{180} |\mathcal{K}_{max}|^2 \quad (l = 1, \ldots, n - m).
$$

We have the following bounds for $l, k = 1, \ldots, n - m, \; l \neq k$ on the off-diagonal entries of D

$$
[D]_{l,k} = \mathbb{E}[f_l(\bar{x}) f_k(\bar{x})]
$$

$$
= \frac{1}{4} \mathbb{E}[(\sum_{j=1}^{m} < \bar{x}, \bar{v}_{l,j} >^2 \mathcal{K}_{l,j})(\sum_{j=1}^{m} < \bar{x}, \bar{v}_{k,j} >^2 \mathcal{K}_{k,j})]
$$

$$
\le \frac{1}{4} |\mathcal{K}_{max}|^2 \mathbb{E}[\| \bar{x} \|_2^4] = \frac{m(5m+4)\nu^4}{180} |\mathcal{K}_{max}|^2.
$$

Similarly, it holds that

$$
[D]_{l,k} \;\ge\; -\frac{m(5m+4)\nu^4}{180} |\mathcal{K}_{max}|^2.
$$

Hence, M has the form

$$
M = \begin{bmatrix} \frac{\nu^2}{3} I_{m \times m} & 0_{m \times (n-m)} \\ 0_{(n-m) \times m} & D_{(n-m) \times (n-m)} \end{bmatrix},
$$

where

$$
0 \;\le\; [D]_{l,l} \;\le\; \frac{m(5m+4)\nu^4}{180} |\mathcal{K}_{max}|^2,
$$

$$
-\frac{m(5m+4)\nu^4}{180} |\mathcal{K}_{max}|^2 \;\le\; [D]_{l,k} \;\le\; \frac{m(5m+4)\nu^4}{180} |\mathcal{K}_{max}|^2 \quad (l \neq k).
$$

Therefore, for $l, k = 1, \ldots, n - m$,

$$
|[D]_{l,k}| < \frac{m(5m+4)\nu^4}{180} |\mathcal{K}_{max}|^2 = [D]_{bound}.
$$

Observe that the eigenspace of M corresponding to the eigenvalue $\frac{\nu^2}{3}$ is equal to the span of $\{\bar{e}_1, \ldots, \bar{e}_m\}$, which is the same as $T_P S$. Hence, as $K \to \infty$, we obtain the implication

$$
\frac{\nu^2}{3} \;>\; \rho(D) \;\Rightarrow\; |\angle \widehat{T}_P S, T_P S| \to 0,
$$

where $\rho(D)$ denotes the spectral radius of D, which is positive definite. In the case where D is diagonal, we have

$$
\rho(D) < [D]_{bound} = \frac{m(5m+4)\nu^4}{180} |\mathcal{K}_{max}|^2.
$$

Therefore, for this case, any value of ν satisfying

$$\frac{\nu^2}{3} > \frac{m(5m+4)\nu^4}{180}|\mathcal{K}_{max}|^2 \text{ or equivalently } \nu < \sqrt{\frac{60}{m(5m+4)|\mathcal{K}_{max}|^2}}$$

ensures that $|\angle \widehat{T}_P S, T_P S| \to 0$ as $K \to \infty$. In the scenario where D is dense, we have the stricter condition

$$\rho(D) < \| D \|_F \le (n-m)\frac{\nu^4(5m+4)m}{180}|\mathcal{K}_{max}|^2.$$

Thus, for this case, any value of ν satisfying

$$\frac{\nu^2}{3} > (n-m)\frac{m(5m+4)\nu^4}{180}|\mathcal{K}_{max}|^2 \text{ or equivalently } \nu < \sqrt{\frac{60}{m(n-m)(5m+4)|\mathcal{K}_{max}|^2}}$$

ensures that $|\angle \widehat{T}_P S, T_P S| \to 0$ as $K \to \infty$.   □

A.2. **Proof of Lemma 3.** We first recall two recent results on the tail bounds for the eigenvalues of sums of independent random matrices. The first result concerns upper and lower tail bounds on all eigenvalues of a sum of independent positive semidefinite matrices as stated in Theorem 4.1 in [22].

Theorem 5 (Eigenvalue Chernoff Bounds). *Consider a finite sequence $\{X_i\}$ of independent random positive semidefinite matrices where $X_i \in \mathbb{R}^{n \times n}$ with $\| X_i \| \le R$ a.s. Given an integer $k \le n$ define*

$$\mu_k = \lambda_k(\sum_j \mathbb{E}[X_j])$$

Then

$$\mathbb{P}(\lambda_k(\sum_j X_j) \ge t\mu_k) \le (n-k+1)\left[\frac{e}{t}\right]^{t\mu_k/R} \text{ where } t > e \quad and$$

$$\mathbb{P}(\lambda_k(\sum_j X_j) \le s\mu_k) \le ke^{\frac{-(1-s)^2\mu_k}{2R}}, \quad s \in (0,1).$$

The second result concerns an upper tail bound on the operator norm of a sum of zero-mean independent random matrices which can moreover be rectangular. This result is stated in the form of Theorem 1.3 in [23].

Theorem 6 (Matrix Bernstein: Rectangular Case). *Consider a finite sequence $\{Z_j\}$ of independent random matrices, $Z_j \in \mathbb{R}^{d_1 \times d_2}$. Assume that each random matrix satisfies*

$$\mathbb{E}[Z_j] = 0 \quad and \quad \| Z_j \| \le R \quad a.s.$$

Define

$$\sigma^2 := \max \left\{ \| \sum_k \mathbb{E}[Z_k Z_k^*] \|, \| \sum_k \mathbb{E}[Z_k^* Z_k] \| \right\}.$$

Then for all $t \ge 0$,

$$\mathbb{P}(\| \sum_k Z_k \| \ge t) \le (d_1 + d_2) \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right).$$

We now proceed to prove the Lemma A.2.

Proof. We have $M^{(K)} = \frac{1}{K}\sum_{i=1}^K \bar{p}_i \bar{p}_i^T$, where $\bar{p}_i = [\bar{x}_i^T \ f_1(\bar{x}_i) \ldots f_{n-m}(\bar{x}_i)]^T \in \mathbb{R}^n$. Now,

$$\| \frac{1}{K}\bar{p}_i\bar{p}_i^T \| \le \frac{1}{K} \| \bar{p}_i \|_2^2 \le \frac{1}{K}(m\nu^2 + \frac{1}{4}(n-m)m^2\nu^4|\mathcal{K}_{max}|^2)$$

$$= \frac{1}{K}\nu^2 R_M \quad \text{a.s.},$$

where $R_M = m + \frac{1}{4}(n-m)m^2\nu^2|\mathcal{K}_{max}|^2$. Here we used the fact that

$$|f_l(\bar{x})| \le \frac{1}{2}m\nu^2|\mathcal{K}_{max}| \text{ for } \bar{x} \in [-\nu, \nu]^m \text{and } l = 1, \ldots, n-m.$$

Furthermore, since $\nu < \nu_{\text{bound,quad}}$,

$$\mu_j = \lambda_j \left(\mathbb{E}\left[\frac{1}{K} \sum_{i=1}^{K} \bar{p}_i \bar{p}_i^T \right] \right) = \frac{\nu^2}{3}, \quad j = 1, \ldots m.$$

Hence, by applying Theorem 5, we have the following for $s_1 \in (0,1)$:

$$\mathbb{P}(\lambda_m(M^{(K)}) \leq s_1 \nu^2 / 3) \leq (n - m + 1) \exp \left(\frac{-(1-s_1)^2 \frac{\nu^2}{3}}{2\nu^2 R_M / K} \right)$$

$$= (n - m + 1) \exp \left(\frac{-(1-s_1)^2 K}{6 R_M} \right). \tag{A.1}$$

Then, we have, $D^{(K)} = \frac{1}{K} \sum_{i=1}^{K} \bar{q}_i \bar{q}_i^T$ where $\bar{q}_i = [f_1(\bar{x}_i) \ldots f_{n-m}(\bar{x}_i)]^T \in \mathbb{R}^{n-m}$. Furthermore,

$$\| \frac{1}{K} \bar{q}_i \bar{q}_i^T \| \leq \frac{1}{K} \| \bar{q}_i \|_2^2 \leq \frac{R_D \nu^4}{K},$$

where $R_D = \frac{1}{4}(n-m)m^2 |\mathcal{K}_{max}|^2$. Applying Theorem 5 for $\rho(D^{(K)}) = \lambda_1(D^{(K)})$, we can write

$$\mathbb{P}(\rho(D^{(K)}) \geq s_2 \rho(D)) \leq (n - m) \left[\frac{e}{s_2} \right]^{\frac{s_2 \rho(D) K}{R_D \nu^4}}, \quad s_2 > e.$$

We have seen in Section A.1 that $\rho(D) < RL\nu^4$. Using this, we obtain the following tail bound:

$$\mathbb{P}(\rho(D^{(K)}) \geq s_2 RL\nu^4) \leq (n - m) \left[\frac{e}{s_2} \right]^{\frac{s_2 RLK}{R_D}}, \quad s_2 > e. \tag{A.2}$$

We proceed now to derive an upper bound on $\| B^{(K)} \|$ by applying Theorem 6. First, observe that

$$B^{(K)} = \frac{1}{K} \sum_{i=1}^{K} \bar{x}_i \bar{q}_i^T.$$

By using the bounds

$$\| \bar{x}_i \|_2 \leq \nu \sqrt{m} \quad \text{and} \quad \| \bar{q}_i \|_2 \leq \frac{1}{2} m \nu^2 \sqrt{n-m} |\mathcal{K}_{max}|,$$

we obtain

$$\| \frac{1}{K} \bar{x}_i \bar{q}_i^T \| \leq \frac{1}{K} \| \bar{x}_i \|_2 \| \bar{q}_i \|_2 \leq \frac{R_B \nu^3}{K},$$

where $R_B = \frac{1}{2} m^{3/2} (n-m)^{1/2} |\mathcal{K}_{max}|$. The parameter σ^2 defined in Theorem A.2 has the following form

$$\sigma^2 = \max \left\{ \frac{1}{K^2} \| \sum_{i=1}^{K} \mathbb{E}[\bar{x}_i \bar{q}_i^T \bar{q}_i \bar{x}_i^T] \|, \frac{1}{K^2} \| \sum_{i=1}^{K} \mathbb{E}[\bar{q}_i \bar{x}_i^T \bar{x}_i \bar{q}_i^T] \| \right\}$$

$$\leq \max \{ \frac{1}{K^2} \sum_{i=1}^{K} \| \mathbb{E}\left[\| \bar{q}_i \|_2^2 \bar{x}_i \bar{x}_i^T \right] \|, \frac{1}{K^2} \sum_{i=1}^{K} \| \mathbb{E}\left[\| \bar{x}_i \|_2^2 \bar{q}_i \bar{q}_i^T \right] \| \}.$$

Now the terms $\| \mathbb{E}\left[\| \bar{q}_i \|_2^2 \bar{x}_i \bar{x}_i^T \right] \|$ and $\| \mathbb{E}\left[\| \bar{x}_i \|_2^2 \bar{q}_i \bar{q}_i^T \right] \|$ can be bounded from above as follows.

$$\| \mathbb{E}\left[\| \bar{q}_i \|_2^2 \bar{x}_i \bar{x}_i^T \right] \| = \sup_{\|\bar{y}\|_2 = 1} \bar{y}^T \mathbb{E}\left[\| \bar{q}_i \|_2^2 \bar{x}_i \bar{x}_i^T \right] \bar{y} \leq (\| \bar{q}_i \|_2^2)_{max} \| \mathbb{E}\left[\bar{x}_i \bar{x}_i^T \right] \|,$$

$$\| \mathbb{E}\left[\| \bar{x}_i \|_2^2 \bar{q}_i \bar{q}_i^T \right] \| = \sup_{\|\bar{y}\|_2 = 1} \bar{y}^T \mathbb{E}\left[\| \bar{x}_i \|_2^2 \bar{q}_i \bar{q}_i^T \right] \bar{y} \leq (\| \bar{x}_i \|_2^2)_{max} \| \mathbb{E}\left[\bar{q}_i \bar{q}_i^T \right] \|.$$

Observe that, for $i = 1, \ldots, K$, we have

$$\| \mathbb{E}[\bar{x}_i \bar{x}_i^T] \| = \| \frac{\nu^2}{3} I_m \| = \frac{\nu^2}{3} \quad \text{and} \quad \| \mathbb{E}[\bar{q}_i \bar{q}_i^T] \| = \rho(D) < RL\nu^4.$$

Furthermore, using the aforementioned upper bounds on $\| \bar{q}_i \|_2$ and $\| \bar{x}_i \|_2$, we arrive at the following:

$$\sigma^2 \leq \max \left\{ \frac{(n-m)m^2 \nu^6 |\mathcal{K}_{max}|^2}{12 K}, \frac{m\nu^2}{K} \rho(D) \right\}$$

$$\leq \max \left\{ \frac{(n-m)m^2 \nu^6 |\mathcal{K}_{max}|^2}{12 K}, \frac{mRL\nu^6}{K} \right\} = \frac{\nu^6 R_\sigma}{K},$$

where

$$R_\sigma := \frac{m^2 |\mathcal{K}_{max}|^2}{12} \max \left\{ n - m, \frac{R(5m + 4)}{15} \right\}.$$

Employing the bounds on $\| \frac{1}{K} \bar{x}_i \bar{q}_i^T \|$ and σ^2 in Theorem 6, we obtain the following tail bound.

$$\mathbb{P}(\| B^{(K)} \| > s_3) \leq n \exp \left(\frac{-(s_3^2/2)K}{\nu^6 R_\sigma + \frac{R_B \nu^3 s_3}{3}} \right), \quad s_3 > 0 \tag{A.3}$$

Lastly, let $0 < p_1, p_2, p_3 < 1$ denote the upper bounds on the probabilities of the events

$$\left\{ \lambda_m(M^{(K)}) \leq s_1 \nu^2/3 \right\}, \left\{ \rho(D^{(K)}) \geq s_2 RL\nu^4 \right\}, \left\{ \| B^{(K)} \| > s_3 \right\},$$

respectively. This is clearly achieved by choosing

$$K > \max \left\{ K_{bound}^{(1)}, K_{bound}^{(2)}, K_{bound}^{(3)} \right\} = K_{bound},$$

where $K_{bound}^{(1)}, K_{bound}^{(2)}, K_{bound}^{(3)}$ are as defined in the statement of Lemma 3. Applying the union bound, we arrive at the stated result. $\qquad \square$

A.3. Proof of Theorem 3.

Proof. We start with the following identity for $i = 1, \dots, m$

$$M^{(K)} \bar{u}_i \; = \; \lambda_i(M^{(K)}) \bar{u}_i, \tag{A.4}$$

where

$$M^{(K)} = \begin{bmatrix} A^{(K)} & B^{(K)} \\ B^{(K)^T} & D^{(K)} \end{bmatrix}.$$

Here $\lambda_1(M^{(K)}) \geq \lambda_2(M^{(K)}) \geq \cdots \geq \lambda_n(M^{(K)})$ denote the eigenvalues of $M^{(K)}$ and $\bar{u}_i = [\bar{u}_{i,1}^T \; \bar{u}_{i,2}^T]^T$ denote its corresponding eigenvectors. Using Eq. (A.4), we obtain the following inequality.

$$B^{(K)^T} \bar{u}_{i,1} + D^{(K)} \bar{u}_{i,2} = \lambda_i(M^{(K)}) \bar{u}_{i,2}$$
$$\Rightarrow (\lambda_m(M^{(K)}) - \rho(D^{(K)})) \| \bar{u}_{i,2} \|_2 < \| B^{(K)} \| . \tag{A.5}$$

Now, provided that K is chosen such that $K > K_{bound}$, the following events hold with high probability.

$$\left\{ \lambda_m(M^{(K)}) > s_1 \frac{\nu^2}{3} \right\}, \quad \left\{ \rho(D^{(K)}) < s_2 RL\nu^4 \right\}, \quad \left\{ \| B^{(K)} \| < s_3 \right\}, \tag{A.6}$$

where $s_1 \in (0, 1), s_2 > e$ and $s_3 > 0$. From (A.6) and (A.5), we conclude that the following inequality holds with high probability.

$$(s_1 \frac{\nu^2}{3} - s_2 RL\nu^4) \| \bar{u}_{i,2} \|_2 < s_3.$$

The L.H.S. of the above inequality is positive if $\nu < \sqrt{s_1/(3 s_2 RL)}$. Assuming that this is satisfied, we obtain

$$\| \bar{u}_{i,2} \| \; < \; \frac{s_3}{s_1 \frac{\nu^2}{3} - s_2 RL\nu^4} \; = \; \sigma_s$$
$$\Rightarrow \; \| U_2 \|_F \; < \; \sqrt{m} \sigma_s.$$

Furthermore, we have from Lemma 1 that

$$\| U_2 \|_F < \tau < 1 \Rightarrow |\angle \widehat{T}_P S, T_P S| < \cos^{-1}(\sqrt{(1 - \tau^2)^m}). \tag{A.7}$$

Lastly, we see that Eq. (A.7) is ensured if the following holds

$$\sqrt{m} \sigma_s < \tau \Leftrightarrow \frac{\sqrt{m} s_3}{s_1 \frac{\nu^2}{3} - s_2 RL\nu^4} < \tau \Leftrightarrow s_3 < \frac{(s_1 \frac{\nu^2}{3} - s_2 RL\nu^4) \tau}{\sqrt{m}}.$$

Therefore, for these choices of ν and the constants s_1, s_2, s_3, we get the bound on $|\angle \widehat{T}_P S, T_P S|$ stated in the theorem.

$\qquad \square$

A.4. **Proof of Lemma 4.**

Proof. For $K = \infty$, we have $M = M_q + \Delta$, where

$$M_q = \begin{bmatrix} \frac{\nu^2}{3} I_m & 0 \\ 0 & D \end{bmatrix}$$

and

$$
\begin{aligned}
\| \Delta \|_F &< 2 \| B_1 \|_F + \| D_1 \|_F, \\
&< 2\sqrt{m(n-m)}\nu\delta(\nu) + (n-m)(\delta(\nu)^2 + \delta(\nu)m\nu^2|\mathcal{K}_{max}|), \\
&= 2C_s m^{3/2}\sqrt{m(n-m)}\nu^4 + (n-m)(C_s^2 m^3 \nu^6 + C_s m^{5/2}\nu^5|\mathcal{K}_{max}|), \quad (A.8) \\
&= 2 \| B_1 \|_{F,bound} + \| D_1 \|_{F,bound} = \| \Delta \|_{F,bound}.
\end{aligned}
$$

Now, if there is no perturbation on M_q, the eigenvectors $\{\bar{e}_1, \ldots, \bar{e}_m\}$ corresponding to $\frac{\nu^2}{3}$ span $T_P S$. As M_q is actually perturbed by Δ, we analyze the perturbation of the space formed by the span of $\{\bar{e}_1, \ldots, \bar{e}_m\}$. We first observe from Weyl's inequality [33] the following bounds on the eigenvalues $\{\lambda_i(M)\}_{i=1}^n$ of M:

$$\lambda_i(M) \in [\lambda_i(M_q) - \| \Delta \|_{F,bound}, \ \lambda_i(M_q) + \| \Delta \|_{F,bound}], \quad i = 1, \ldots, n.$$

Here, $\lambda_i(M_q) = \frac{\nu^2}{3}$, for $i = 1, \ldots, m$. Furthermore, $\{\lambda_i(M_q)\}_{i=m+1}^n$ are the eigenvalues of D. In order to analyze the perturbation on $\text{span}\{\bar{e}_1, \ldots, \bar{e}_m\}$, we would like to guarantee the 'separation' of $\{\lambda_i(M)\}_{i=1}^m$ from $\{\lambda_i(M)\}_{i=m+1}^n$. Denoting $\rho(D)$ to be the spectral radius of D, we have the following sufficient condition to guarantee this separation.

$$
\begin{aligned}
\frac{\nu^2}{3} - \| \Delta \|_{F,bound} &> \rho(D) + \| \Delta \|_{F,bound}, \\
\Leftrightarrow \quad \frac{\nu^2}{3} - \rho(D) &> 2 \| \Delta \|_{F,bound}. \quad (A.9)
\end{aligned}
$$

Now, as shown in Section A.1, $\rho(D) < RL\nu^4$, where $L = \frac{m(5m+4)|\mathcal{K}_{max}|^2}{180}$, and

$$
R = \begin{cases} 1 & ; \text{ if D is diagonal} \\ (n-m) & ; \text{ if D is dense.} \end{cases}
$$

Using this fact along with Eq. (A.8) in Eq. (A.9), we arrive at the following sufficient condition that guarantees the separation of eigenvalues:

$$\frac{\nu^2}{3} - RL\nu^4 > \beta_2 \nu^4 + \beta_3 \nu^5 + \beta_4 \nu^6, \quad (A.10)$$

$$\Leftrightarrow (\beta_2 + RL)\nu^2 + \beta_3 \nu^3 + \beta_4 \nu^4 < \frac{1}{3} \quad (A.11)$$

where $\beta_2 = 4C_s m^{3/2}\sqrt{m(n-m)}$, $\beta_3 = 2(n-m)C_s m^{5/2}|\mathcal{K}_{max}|$ and $\beta_4 = 2(n-m)C_s^2 m^3$. Now, clearly the solution to Eq. (A.10) needs to satisfy the following conditions.

$$
\begin{aligned}
(\beta_2 + RL)\nu^2 < 1/3 &\Leftrightarrow \nu < (3(\beta_2 + RL))^{-1/2}, \\
\beta_3 \nu^3 < 1/3 &\Leftrightarrow \nu < (3\beta_3)^{-1/3}, \\
\beta_4 \nu^4 < \frac{1}{3} &\Leftrightarrow \nu < (3\beta_4)^{-1/4}.
\end{aligned}
$$

Equivalently, the solution to Eq. (A.10) satisfies $\nu < \alpha$, where

$$\alpha = \min\left\{ (3(\beta_2 + RL))^{-1/2}, (3\beta_3)^{-1/3}, (3\beta_4)^{-1/4} \right\}.$$

We thus arrive at the following sufficient condition on ν in order to guarantee Eq. (A.11) and consequently Eq. (A.9):

$$
\begin{aligned}
\nu^2((\beta_2 + RL) + \beta_3 \alpha + \beta_4 \alpha^2) &< \frac{1}{3}, \\
\Leftrightarrow \nu^2 &< \frac{1}{3[(\beta_2 + RL) + \beta_3 \alpha + \beta_4 \alpha^2]}. \quad (A.12)
\end{aligned}
$$

We now proceed to bound the angle between $T_P S$ and $\widehat{T}_P S$ by using the identity $M\bar{u}_i = \lambda_i(M)\bar{u}_i$, $\forall i = 1, \ldots, m$. We obtain

$$B_1^T \bar{u}_{i,1} + (D + D_1)\bar{u}_{i,2} = \lambda_i(M)\bar{u}_{i,2}.$$

By taking the l_2-norm of both sides and using the fact that $\| D\bar{u}_{i,2} \|_2 \leq \rho(D) \| \bar{u}_{i,2} \|_2$, we obtain

$$(\lambda_i(M) - \rho(D) - \| D_1 \|_F) \| \bar{u}_{i,2} \|_2 < \| B_1 \|_F .$$

Now, $\lambda_i(M) > \frac{\nu^2}{3} - \| \Delta \|_F$ for $i = 1, \ldots, m$. Therefore, if ν is chosen to satisfy Eq. (A.12), then the following holds true.

$$
\begin{aligned}
\lambda_i(M) - \rho(D) - \| D_1 \|_F &> \frac{\nu^2}{3} - \| \Delta \|_F - \rho(D) - \| D_1 \|_F, \\
&> \frac{\nu^2}{3} - RL\nu^4 - 2(\| B_1 \|_{F,bound} + \| D_1 \|_{F,bound}), \\
&> 0.
\end{aligned}
$$

Using the above facts, we obtain the following upper bound on $\| \bar{u}_{i,2} \|_2$.

$$\| \bar{u}_{i,2} \|_2 < \frac{\| B_1 \|_{F,bound}}{\frac{\nu^2}{3} - RL\nu^4 - 2(\| B_1 \|_{F,bound} + \| D_1 \|_{F,bound})} = \sigma_\infty. \tag{A.13}$$

Finally, to conclude the proof, we obtain the bound on $|\angle \widehat{T}_P S, T_P S|$ by using Lemma 1.

$$\| U_2 \|_F^2 = \sum_{i=1}^m \| \bar{u}_{i,2} \|_2^2 < m\sigma_\infty^2,$$

$$\Rightarrow \cos^2(\angle \widehat{T}_P S, T_P S) > (1 - \| U_2 \|_F^2)^m > (1 - m\sigma_\infty^2)^m.$$

$$\square$$

A.5. Proof of Theorem 4.

Proof. The proof follows along the lines of the proof of Theorem 3. We start with the following identity for $i = 1, \ldots, m$:

$$M^{(K)}\bar{u}_i = \lambda_i(M^{(K)})\bar{u}_i. \tag{A.14}$$

We have, $M^{(K)} = M_q^{(K)} + \Delta^{(K)}$ where

$$M_q^{(K)} = \begin{bmatrix} A^{(K)} & B^{(K)} \\ B^{(K)^T} & D^{(K)} \end{bmatrix} \text{ and } \Delta^{(K)} = \begin{bmatrix} 0 & B_1^{(K)} \\ B_1^{(K)^T} & D_1^{(K)} \end{bmatrix}.$$

Let $\lambda_1(M^{(K)}) \geq \lambda_2(M^{(K)}) \cdots \geq \lambda_n(M^{(K)})$ denote the eigenvalues of $M^{(K)}$ and $\bar{u}_i = [\bar{u}_{i,1}^T \ \bar{u}_{i,2}^T]^T$ denote its corresponding eigenvectors. Using Eq. (A.14), we obtain the following:

$$(B^{(K)^T} + B_1^{(K)^T})\bar{u}_{i,1} + (D^{(K)} + D_1^{(K)})\bar{u}_{i,2} = \lambda_i(M^{(K)})\bar{u}_{i,2}$$

$$\Rightarrow (\lambda_m(M^{(K)}) - \rho(D^{(K)}) - \| D_1 \|_{F,bound}) \| \bar{u}_{i,2} \|_2 \tag{A.15}$$

$$< \| B^{(K)} \| + \| B_1 \|_{F,bound} .$$

We observe by Weyl's inequality [33] that the following holds true:

$$\lambda_m(M^{(K)}) \geq \lambda_m(M_q^{(K)}) - \| \Delta \|_{F,bound} = \lambda_m(M_q^{(K)}) - 2 \| B_1 \|_{F,bound} - \| D_1 \|_{F,bound} . \tag{A.16}$$

If K is chosen such that $K > K_{bound}$, the following events hold with high probability:

$$\left\{ \lambda_m(M_q^{(K)}) > s_1 \frac{\nu^2}{3} \right\}, \quad \left\{ \rho(D^{(K)}) < s_2 RL\nu^4 \right\}, \quad \left\{ \| B^{(K)} \| < s_3 \right\}, \tag{A.17}$$

where $s_1 \in (0,1)$, $s_2 > e$ and $s_3 > 0$. Thus, using Eq. (A.16) and Eq. (A.17) in Eq. (A.15), we obtain the following:

$$(s_1 \frac{\nu^2}{3} - s_2 RL\nu^4 - 2(\| B_1 \|_{F,bound} + \| D_1 \|_{F,bound})) \| \bar{u}_{i,2} \|_2 < s_3 + \| B_1 \|_{F,bound} .$$

Similarly to the proof of Lemma 4, one can show that the following condition is sufficient to ensure that the L.H.S. of the above inequality is strictly positive

$$\nu^2 < \frac{s_1}{3[(\beta_2 + s_2 RL) + \beta_3 \alpha + \beta_4 \alpha^2]}. \tag{A.18}$$

In particular, the above condition ensures the following:

$$s_1 \frac{\nu^2}{3} - s_2 RL\nu^4 > 2 \parallel \Delta \parallel_{F,bound} = 4 \parallel B_1 \parallel_{F,bound} + 2 \parallel D_1 \parallel_{F,bound}.$$

Now, assuming that ν satisfies Eq. (A.18), we arrive at the following bound on $\parallel \bar{u}_{i,2} \parallel_2$ for $i = 1, \ldots, m$:

$$\parallel \bar{u}_{i,2} \parallel_2 < \frac{s_3 + \parallel B_1 \parallel_{F,bound}}{(s_1 \frac{\nu^2}{3} - s_2 RL\nu^4) - 2(\parallel B_1 \parallel_{F,bound} + \parallel D_1 \parallel_{F,bound})}$$

$$= \sigma_s.$$

The above bound on $\parallel \bar{u}_{i,2} \parallel_2$ implies that $\parallel U_2 \parallel_F^2 < m\sigma_s^2$. Let

$$\sigma_f := \frac{\parallel B_1 \parallel_{F,bound}}{(s_1 \frac{\nu^2}{3} - s_2 RL\nu^4) - 2(\parallel B_1 \parallel_{F,bound} + \parallel D_1 \parallel_{F,bound})}.$$

If for some $\tau \in (0, 1)$

$$m\sigma_s^2 < \tau^2 + m\sigma_f^2 \Leftrightarrow \sigma_s < (\tau^2/m + \sigma_f^2)^{1/2}, \tag{A.19}$$

then from Lemma 1 we obtain

$$\cos^2(\angle \widehat{T}_P S, T_P S) > 1 - \tau^2 - m\sigma_f^2.$$

Finally, we see that Eq. (A.19) is ensured if the following holds.

$$\frac{s_3 + \parallel B_1 \parallel_{F,bound}}{(s_1 \frac{\nu^2}{3} - s_2 RL\nu^4) - 2(\parallel B_1 \parallel_{F,bound} + \parallel D_1 \parallel_{F,bound})} < \left(\frac{\tau^2}{m} + \sigma_f^2\right)^{1/2}$$

$$\Leftrightarrow s_3 < [(s_1 \frac{\nu^2}{3} - s_2 RL\nu^4) - 2(\parallel B_1 \parallel_{F,bound} + \parallel D_1 \parallel_{F,bound})] \left(\frac{\tau^2}{m} + \sigma_f^2\right)^{1/2}$$

$$- \parallel B_1 \parallel_{F,bound}.$$

This completes the proof. □