Tangent space estimation bounds for smooth manifolds

Hemant Tyagi*, Elif Vural† and Pascal Frossard†
*Institute of Theoretical Computer Science
ETH Zurich, Switzerland
Email: htyagi@inf.ethz.ch
†School of Electrical Enginnering, EPFL, Switzerland
Emails: {elif.vural, pascal.frossard}@epfl.ch

***Abstract*—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 and (ii) the number of samples in the neighborhood affect the performance of learning algorithms. In this paper, we propose a theoretical analysis of local sampling conditions for the estimation of the tangent space at a point P lying on an 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. Considering a random sampling framework, we leverage recent results from random matrix theory to derive local sampling conditions for an accurate estimation of tangent subspace. Our main results state that the width of the sampling region in the tangent space guaranteeing an accurate estimation is inversely proportional to the manifold dimension, curvature, and the square root of the ambient space dimension. At the same time, we show that the number of samples increases quadratically with the manifold dimension and logarithmically with the ambient space dimension.**

I. Introduction

A data set that resides in a high-dimensional Euclidean 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 forms 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 [1], [2], [3]. 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 for a given estimation accuracy. Furthermore, the number of necessary samples is likely to vary according to the local characteristics of the manifold. In this work, we present an analysis of the sampling problem and derive conditions on the size of the sampling region and the number of samples for an accurate estimation of the tangent space.

There are many examples of dimensionality reduction algorithms such as [3], [4], [5], [6], which apply a local Principal Component Analysis (PCA) for the computation of the tangent space of the manifold. The performance of Singular Value Decomposition (SVD) or PCA under noise is a well-studied topic (see [7], [8], [9] and references within). However these studies do not involve the geometric structure of the data. Only a few recent works have studied the relation between PCA performance and data geometry. The work in [10] generalizes the idea of diffusion maps in dimensionality reduction [11] to vector diffusion maps. As part of their analysis, the authors have shown in particular that when the size ε of the local area for tangent space estimation is set to $\varepsilon = O(K^{-\frac{1}{m+2}})$ with K being the number of samples on the *whole* manifold and m being the dimension of the manifold, then the deviation between the estimated and the true tangent space is typically $O(\varepsilon^{3/2})$. Their work however considers a global sampling from a compact manifold while we focus here on the local manifold geometry. Finally, the accuracy of tangent space estimation from noisy manifold samples is analyzed in a work parallel to ours [12]. This study focuses on manifolds that are embedded with exactly quadratic forms and poses the sampling problem as the selection of a subset of samples from a set of noisy samples given a priori. On the contrary, we analyze 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.

Our contribution in this paper can be summarized as follows. Firstly, we determine a suitable upper bound on the size of the neighborhood in the tangent space within which the manifold can be sampled randomly. In the derivation of this bound, we consider the asymptotic case $K \to \infty$ with arbitrarily many manifold samples 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. Secondly, 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 the aforementioned bound. To this end, we utilize recent results from random matrix theory [13], [14]. Combining the two above results, we give a complete characterization of the local sampling conditions in Theorem 1.

The rest of the paper is organized as follows. Section II contains the formal outline of the problem. In Section III we present the main results along with a discussion. In Section IV we provide concluding remarks and possible directions for future work.

II. Problem setup

We consider an m-dimensional submanifold S of $\mathbb{R}^n$ with a smooth embedding in $\mathbb{R}^n$, $n \geq m+1$. Let $P \in S$ be a manifold point and $\mathcal{N}_\varepsilon(P)$ denote an ε-neighbourhood of P on S for some

This work has been performed while the first author was with the Signal Processing Laboratory LTS4 at EPFL.

The work has been partly funded by the Swiss National Science Foundation under Grant 200020-132772.

$\varepsilon > 0$

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

where $\| \ . \ \|_2$ stands for the ℓ_2-norm in $\mathbb{R}^n$. Let $T_P S$ denote the tangent space at P.

In our analysis, we represent the points in $\mathcal{N}_\varepsilon(P)$ via tangent space parameterization using local functions $f_l : T_P S \rightarrow \mathbb{R}$. There exists an ε such that all points $M \in \mathcal{N}_\varepsilon(P)$ can be uniquely represented in the form

$$[\bar{x}^T \ f_1(\bar{x}) \dots f_{n-m}(\bar{x})]^T. \quad \text{(II.1)}$$

Here $\bar{x} = [x_1 \dots x_m]^T$ denotes the coordinates of the orthogonal projection of manifold points onto $T_P S$. Note that, in (II.1), the coordinates are given with respect to the reference manifold point P, which is taken as the local origin. Furthermore, aligning the coordinate system with the tangent space at P, $T_P S$ can be represented as

$$T_P S = \text{span}\{\bar{e}_1, \dots, \bar{e}_m\},$$

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

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

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

$$f_l(\bar{x}) = f_{q,l}(\bar{x}) + R_l(\bar{x}); \quad l = 1, \dots, n-m \quad \text{(II.2)}$$

where $f_{q,l}$ is a quadratic form and $R_l(\bar{x})$ is the remainder term of $O(\| \bar{x} \|_2^3)$. The Hessian of f_l at the local origin P can be represented as

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

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

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

defined by f_l. We then define the maximum principal curvature at P as $\mathcal{K}_{max} := \mathcal{K}_{l',j'}$ where $(l', j') = \underset{l,j}{\text{argmax}} |\mathcal{K}_{l,j}|$. We consider that the tangent space is 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

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

The matrices U and $\Lambda \in \mathbb{R}^n$ represent respectively the eigenvector and eigenvalue matrices of $M^{(K)}$ where

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

with the ordering $\lambda_1 \geq \dots \geq \lambda_m \geq \dots \geq \lambda_n$. 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.,

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

Hence, $\widehat{T}_P S$ is the estimation of the true tangent space $T_P S$ at P with PCA. This is illustrated in Fig. 1. Finally, we characterize the accuracy of the tangent space estimation with the angle between $\widehat{T}_P S$ and $T_P S$, where we use the angle definition given in [15].

We can now state our problem formally. Given the above setting, we want to describe the conditions on the manifold samples $\{P_i\}_{i=1}^K$

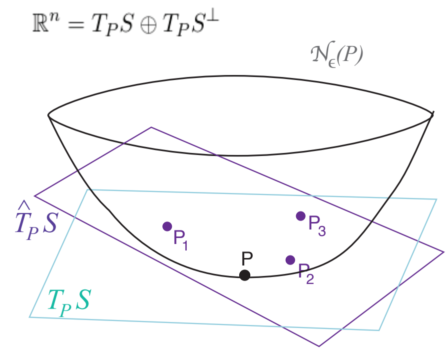

Fig. 1. The true tangent space $T_P S$ and the estimated tangent space $\widehat{T}_P S$ at point P.

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 from P_i to P?
- *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, we would like to determine the the dependency of the above bounds on n, m and $\mathcal{K}_{max}$. In order to address these 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, we assume that

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

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.

III. Main results

We now present our main results regarding the sampling of a smooth manifold. First, since we consider the sampling of $f_l(\bar{x})$ over the compact region $[-\nu, \nu]^m$, $R_l(\bar{x})$ is bounded over this region. Therefore, for each l there exists a constant $C_{s,l} > 0$ such that

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

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

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

The empirical covariance matrix $M^{(K)}$ corresponding to the samples $\{P_i\}_{i=1}^K$ is in the form $M^{(K)} = M_q^{(K)} + \Delta^{(K)}$

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

Here $M_q^{(K)}$ is the covariance matrix associated with the quadratic components $f_{q,l}(\bar{x})$ of the embeddings. The $m \times m$ matrix $A^{(K)}$ gives the covariance of the tangential components $\bar{x}_i$ of data points

P_i. As $K \to \infty$, the matrix $A^{(K)} \to \frac{\nu^2}{3} I_{m\times m}$ approaches a scaled version of the identity matrix; and therefore, the column space of $A^{(K)}$ approaches the true tangent space $T_P S$. The submatrices $B^{(K)}$ and $D^{(K)}$ represent the error on account of the nonzero manifold curvature at P, which stems from the second-order terms $f_{q,l}$. Meanwhile, $\Delta^{(K)}$ is an additional error term resulting from the higher-order Taylor terms R_l of the mappings f_l. We give the explicit formulation of $B_1^{(K)}$ and $D_1^{(K)}$ in [16, Section 4.4] and show that their Frobenius norms $\|B_1^{(K)}\|_F$ and $\|D_1^{(K)}\|_F$ can be bounded as

$$\begin{aligned}
\| B_1^{(K)} \|_F &< \| B_1 \|_{F,bound} := \sqrt{m(n-m)} C_s m^{3/2} \nu^4, \\
\| D_1^{(K)} \|_F &< \| D_1 \|_{F,bound} \\
&:= (n-m) C_s m^{5/2} \nu^5 (C_s m^{1/2} \nu + |\mathcal{K}_{max}|).
\end{aligned}$$

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 $\Delta = \lim_{K\to\infty} \Delta^{(K)}$. Consequently, one can show that [16]

$$\|B_1\|_F < \|B_1\|_{F,bound}, \qquad \|D_1\|_F < \|D_1\|_{F,bound}.$$

Equipped with the above definitions and properties, we are now ready to state our main results about the sampling of smooth manifolds. We characterize the sampling conditions for accurate tangent space estimation by first defining a region of sampling in the tangent space and then determining the number of samples to be chosen from this region. We begin with the region of sampling and present in Lemma 1 the conditions on the sampling width ν that guarantee an upper bound on the angle between $\widehat{T}_P S$ and $T_P S$, provided that the number of samples is arbitrarily large.

Lemma 1: Let the sampling width satisfy

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

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

$$R = n - m, \qquad L = \frac{m(5m+4)|\mathcal{K}_{max}|^2}{180}$$

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 of Lemma 1 is presented in [16, Appendix A.4]. The stated result is derived from the condition that the spectrum associated with $\frac{\nu^2}{3} I_m$, whose corresponding eigenvectors give the true tangent space $T_P S$, is separated from the spectrum of the error. There are two sources of error here; namely, the curvature components $f_{q,l}$ which give rise to the correlation matrix $D = \lim_{K\to\infty} D^{(K)}$ and the higher-order terms R_l yielding the perturbation matrix Δ. The lemma states that the angle $|\angle \widehat{T}_P S, T_P S|$ between the estimated and true tangent spaces converges to the residual bound $\cos^{-1} \sqrt{(1 - m\sigma_\infty^2)^m}$ as the number of samples tends to infinity. The error term $m\sigma_\infty^2$ can be interpreted as the bias error resulting from the fact that a smooth embedding has a non-symmetric structure around the origin in general. In particular, it is easily verifiable that $\sigma_\infty \to 0 \quad \text{as} \quad \nu \to 0$; i.e., the bias approaches zero as the sampling width shrinks to 0. Also note that, when the f_l's are quadratic forms, this bias term vanishes to yield $\sigma_\infty = 0$, which is due to the symmetry of quadratic forms around the origin [16].

We now proceed to the finite sampling case $K < \infty$ and give our complete main result. In Theorem 1, we state 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 1: Let $s_1 \in (0,1)$ and $s_2 > e$ be fixed constants. Assume that the sampling width ν is such that

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

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

$$\begin{aligned}
s_3 <& [(s_1 \frac{\nu^2}{3} - s_2 RL\nu^4) - 2(\| B_1 \|_{F,bound} + \| D_1 \|_{F,bound})] \\
&\left(\frac{\tau^2}{m} + \sigma_f^2 \right)^{1/2} - \| B_1 \|_{F,bound}
\end{aligned}$$

where

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

Finally, let $0 < p_1, p_2, p_3 < 1$. Assume that the number of samples satisfies $K > K_{bound}$, where $K_{bound} = \max\{K_{bound}^{(1)}, K_{bound}^{(2)}, K_{bound}^{(3)}\}$

$$\begin{aligned}
K_{bound}^{(1)} &= \frac{6 R_M}{(1 - s_1)^2} \log\left((n - m + 1)/p_1\right), \\
K_{bound}^{(2)} &= \frac{R_D}{s_2 RL} \frac{\log((n-m)/p_2)}{\log(s_2/e)}, \\
K_{bound}^{(3)} &= \frac{\nu^6 R_\sigma + \frac{R_B \nu^3 s_3}{3}}{s_3^2/2} \log(n/p_3)
\end{aligned}$$

and

$$\begin{aligned}
R_M &= m + \frac{1}{4}(n-m) m^2 \nu^2 |\mathcal{K}_{max}|^2 \\
R_D &= \frac{1}{4}(n-m) m^2 |\mathcal{K}_{max}|^2, \quad R_B = \frac{1}{2} m^{3/2} \sqrt{n-m} |\mathcal{K}_{max}|
\end{aligned}$$

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

Then, with probability larger than $1 - p_1 - p_2 - p_3$,

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

The proof of Theorem 1 is presented in [16, Appendix A.5]. The theorem builds on Lemma 1 that considers the case $K \to \infty$. In the proof of the theorem, in order to account for finite K, we use the results of [13], [14] in order to probabilistically bound how much the tangent space estimated with K samples deviates from the tangent space in Lemma 1 estimated with infinitely many samples. The parameters s_i are used to make the link between the estimation error and the sampling conditions (sampling width and sampling density), whereas the probability constants p_i establish the relation between the error probability and the sampling density K.

Note that the tangent space estimation error in this case consists of two terms - the variance term τ due to finite sampling and the bias term σ_f resulting from the asymmetric manifold geometry, which is the probabilistic counterpart of the bias term σ_∞ in Lemma 1. In particular, the number of samples K is related to the variance term τ through the parameter s_3, such that a larger number of samples decreases the variance, bringing thus the estimation error closer to its asymptotic value given in Lemma 1.

Remark: Let us now interpret our results with respect to the variation of the sampling conditions with the manifold parameters. As shown in [16], the results of our analysis translate into the fact that the choices $\nu = O(n^{-1/2}m^{-1}|\mathcal{K}_{max}|^{-1})$ and $K = O(\tau^{-2}m^2 \log n)$ ensure for large n 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. In this work, the sampling width ν is measured on the tangent space $T_P S$. However, using the estimation $\|.\|_{ambient\,space} \approx O(\|.\|_{tangent\,space}\sqrt{n/m})$, we see that the stated bound on ν implies that the sampling width measured in the ambient space must change at the rate $O(\nu\sqrt{n/m}) = O(m^{-3/2}|\mathcal{K}_{max}|^{-1})$. This practically 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 m or the curvature $\mathcal{K}_{max}$ of the manifold increases, whereas it is not affected by the ambient space dimension n. On the other hand, the number of samples K increases quadratically with m and logarithmically with n.

Let us now briefly discuss the usage of our results with regards to 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*. 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. First, 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 [17, Section V] and [18, Proposition 2]. On the other hand, 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 [19], [20] and [21]. It is also possible to obtain an estimate of the curvature and the deviation term C_s from data samples using results such as in [22].

IV. Conclusions

We have presented a theoretical analysis of the tangent space estimation at a point on a submanifold of $\mathbb{R}^n$ 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. 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.

Acknowledgment

The authors would like to thank Prof. Daniel Kressner and Dr. Bart Vandereycken for the helpful discussions and comments on the manuscripts.

References

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

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

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

[4] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimension reduction via local tangent space alignment," *SIAM Journal of Scientific Computing*, vol. 26, pp. 313–338, 2002.

[5] 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 Art. Intel.*, 2010.

[6] 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*, vol. 5370 of *Lecture Notes in Computer Science*, pp. 406–415. Springer Berlin/Heidelberg, 2008.

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

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

[9] V. Vu, "Singular vectors under random perturbation," *Random Struct. Algorithms*, vol. 39, no. 4, pp. 526–538, Dec. 2011.

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

[11] 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*, 2005, pp. 7426–7431.

[12] 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.

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

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

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

[16] H. Tyagi, E. Vural, and P. Frossard, "Tangent space estimation for smooth embeddings of Riemannian manifolds," [Online]. Available: http://arxiv.org/pdf/1208.1065.pdf.

[17] E. Kokiopoulou, D. Kressner, and P. Frossard, "Optimal image alignment with random projections of manifolds: algorithm and geometric analysis," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1543–1557, 2011.

[18] L. Jacques and C. De Vleeschouwer, "A geometrical study of matching pursuit parametrization," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2835–2848, July 2008.

[19] M. Hein, "Intrinsic dimensionality estimation of submanifolds in Euclidean space," in *Proceedings of the* 22^{nd} *International Conference on Machine Learning*, 2005, pp. 289–296.

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

[21] 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.

[22] 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.