Local Convex Representation with Pruning for Manifold Clustering

Tao Yang and Chun-Guang Li
School of Information and Communication Engineering
Beijing University of Posts and Telecommunications, Beijing 100876, P.R. China
{taoyang, lichunguang}@bupt.edu.cn

***Abstract*—High-dimensional data in many applications can be considered as samples drawn from a union of multiple low-dimensional manifolds. Assigning data points into their own manifolds is referred to *manifold clustering*. Inspired by recent advances in subspace clustering, in this paper, we present an efficient approach for manifold clustering, called Local Convex Representation (LCR), in which each data point is represented as a convex combination of other points in the local neighborhood and under some mild conditions the nonzero coefficients are guaranteed to correspond to the data points lying on the same manifold. Moreover, we incorporate the estimated intrinsic dimension of the manifold to prune the minor nonzero coefficients and validate that the pruning step helps LCR yield remarkable improvements. Experiments on synthetic data as well as real world data demonstrate promising performance.**

***Index Terms*—manifold clustering, convex combination, intrinsic dimension**

I. Introduction

In many real-world applications such as image and video processing, we need to deal with a large amount of high-dimensional data. Such data can often be well approximated by a union of multiple low-dimensional manifolds, where each manifold corresponds to a class or a category. In such cases, it is important to segment the data into multiple groups where each group contains data points from the same manifold. This problem is known as *manifold clustering* [1] and has been successfully applied to various computer vision tasks over the past few years [2]–[4]. Formally, manifold clustering[1] refers to the problem as follows: *Let $X \in \mathbb{R}^{D\times N}$ be a real-valued data matrix, consisting of column vectors drawn from a union of n smooth submanifolds of $\mathbb{R}^D$, $\bigcup_{\ell=1}^{n}\{\mathcal{M}^{(\ell)}\}$, of dimension d_ℓ, for $\ell = 1, ..., n$. The goal of manifold clustering is to segment the columns of X into their corresponding submanifolds.*

Subspace Clustering. When data lie in multiple flat manifolds, this is a problem called *subspace clustering* [5]. Over the past decade, a number of approaches to subspace clustering have been proposed [6]–[10]. Among them, methods based on spectral clustering [11] have become extremely popular. Such methods consist of two steps: a) learning an affinity matrix from the data and then b) applying spectral clustering to this affinity matrix. State-of-the-art subspace clustering methods construct an affinity matrix based on *self-expression* model [6], which states that each data point $\boldsymbol{x}_j$ is expressed as a linear combination of all other data points, i.e., $\boldsymbol{x}_j = \sum_{i\neq j} c_{ij}\boldsymbol{x}_i + \boldsymbol{e}_j$, where the coefficients c_{ij} are used to induce the affinity between points i and j via $a_{ij} = \frac{1}{2}(|c_{ij}| + |c_{ji}|)$ and $\boldsymbol{e}_j$ tolerates the errors. To find the solution $\boldsymbol{c}_j$ of subspace-preserving property [5], i.e., the nonzero coefficients of $\boldsymbol{c}_j$ correspond only to data points from the same subspace as $\boldsymbol{x}_j$, a proper regularizer is usually imposed on $\boldsymbol{c}_j$. For example, in [6], the ℓ_1 norm as a convex surrogate over the ℓ_0 norm is used to promote sparse solution of $\boldsymbol{c}_j$; in [12], a mixture of ℓ_1 and ℓ_2 norms are used to gain a balance between connectivity and correctness; in [9], a scalable algorithm for elastic net based subspace clustering with theoretical guarantees is provided; and in [7], self-expression model is conducted implicitly in feature space to cope with the nonlinear subspaces. These algorithms, however, take advantage of the global linear relations among data points which might not be suitable for clustering nonlinear manifolds.

Manifold Clustering. When data lie in multiple highly nonlinear manifolds, the aforementioned algorithms have no longer their advantages. The existing approaches to cluster nonlinear manifolds can be divided into two categories: the *global* methods and the *local* methods. The global methods assume that the manifolds have different intrinsic dimensions and thus the data can be clustered according to the dimensions rather than the manifolds themselves [13]–[16]. However, in many real-world problems this assumption is often violated because different manifolds might still share the same intrinsic dimension. The local methods usually construct a pairwise affinity to depict the local geometry and then apply spectral clustering. Different methods differ in the way to exploit the local geometry in the data. For example, local proximity is used in Laplacian Eigenmaps [17], [18], local linearity is used in [19]–[24], curvature information is used in [25], principal angle information is investigated in [26], sparse representation is explored in [27]. Among them, *local self-expression* model based methods have received the most attention owning to their elegance and effectiveness, e.g., the local linearity based methods [23] and the weighted sparse representation based method [27]. However, there is either tradeoff parameters to tune or a lack of theoretical justification.

Paper Contributions. Inspired by recent work [28], we propose to induce the affinity for manifold clustering by Local Convex Representation (LCR), which is easier to compute and can guarantee that the nonzero coefficients relate to the data points from the same manifold under mild conditions.

[1]Strictly speaking, we should term it as submanifold clustering.

Moreover, we incorporate the estimated intrinsic dimension to prune the minor coefficients in LCR and validate that the pruning step helps LCR to yield superior clustering accuracy.

II. Our Proposed Approach

Assume that a collection of N data points $\{\boldsymbol{x}_i | \boldsymbol{x}_i \in \mathbb{R}^D, i = 1, ..., N\}$ are uniformly and sufficiently sampled from n locally separable smooth submanifolds which are embedded in D-dimensional Euclidean space $\mathbb{R}^D$.

Local Convex Representation. Consider a data point $\boldsymbol{x}_j$ which lies in submanifold $\mathcal{M}^{(\ell)}$, and denote the indexes of the k nearest neighbors of $\boldsymbol{x}_j$ as $\mathcal{N}(\boldsymbol{x}_j) \subset \{1, ..., N\}$. Assume that the submanifold is locally flat. Then, finding the affine sparse representation of $\boldsymbol{x}_j$ with respect to the k nearest neighbors is solving the following problem:

$$\min_{\boldsymbol{c}_j} \|\boldsymbol{c}_j\|_1 \text{ s.t. } \boldsymbol{x}_j = \sum_{i \in \mathcal{N}(\boldsymbol{x}_j)} c_{ij} \boldsymbol{x}_i, \sum_{i \in \mathcal{N}(\boldsymbol{x}_j)} c_{ij} = 1, \quad (1)$$

where $c_{ij} = 0$ for $i \notin \mathcal{N}(\boldsymbol{x}_j)$. If $\boldsymbol{x}_j$ lies in the relative interior of the convex hull of data points $\{\boldsymbol{x}_i, i \in \mathcal{N}(\boldsymbol{x}_j)\}$, the optimal solution must be nonnegative [28]. Thus, for $\boldsymbol{x}_j$ lying in the relative interior, problem (1) is equivalent to finding a convex representation as follows:

$$\min_{\{c_{ij}\}} \|\boldsymbol{c}_j\|_1 \text{ s.t. } \boldsymbol{x}_j = \sum_{i \in \mathcal{N}(\boldsymbol{x}_j)} c_{ij} \boldsymbol{x}_i, \sum_{i \in \mathcal{N}(\boldsymbol{x}_j)} c_{ij} = 1, \; c_{ij} \geq 0. \quad (2)$$

For the k neighboring points, we arrange the data points from submanifold $\mathcal{M}^{(\ell)}$ as the columns of matrix $X_j^{(\ell)}$, and the data points from other submanifolds as the columns of matrix $X_j^{(-\ell)}$. If $\boldsymbol{x}_j$ lies in the relative interior of the convex hull of the columns in $X_j^{(\ell)}$ and the affine hull of the columns in $X_j^{(\ell)}$ does not intersect the convex hull of $X_j^{(-\ell)}$ then, as established in [28] (see Theorem IV.2), the nonzero coefficients in the optimal solution of problem (2) relate only to data points from $\mathcal{M}^{(\ell)}$ (i.e., the columns in $X_j^{(\ell)}$).

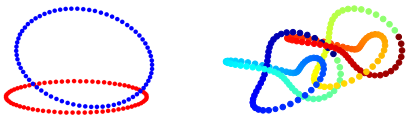
To tolerate some deviations in the local neighborhood, we relax (2) as follows:

$$\min_{\{c_{ij}\}} \|\boldsymbol{c}_j\|_1 + \frac{\lambda}{2} \|\boldsymbol{x}_j - \sum_{i \in \mathcal{N}(\boldsymbol{x}_j)} c_{ij} \boldsymbol{x}_i\|_2^2 \text{ s.t. } \sum_{i \in \mathcal{N}(\boldsymbol{x}_j)} c_{ij} = 1, \; c_{ij} \geq 0, \quad (3)$$

where $\lambda > 0$. Because of the non-negativity constraint, we reformulate problem (3) into an equivalent form without tradeoff parameter λ as follows:

$$\min_{\{c_{ij}\}} \|\boldsymbol{x}_j - \sum_{i \in \mathcal{N}(\boldsymbol{x}_j)} c_{ij} \boldsymbol{x}_i\|_2^2 \quad \text{s.t.} \sum_{i \in \mathcal{N}(\boldsymbol{x}_j)} c_{ij} = 1, \; c_{ij} \geq 0. \quad (4)$$

We call the optimal solution in problem (4) as Local Convex Representation (LCR). While problem (4) is the subproblem in convex LLE [20], we see from the derivation above, for those relative interior points, (4) is effectively an ℓ_1-minimization with nonnegative affine constraint. Thus, as an analogue to [28], we can also establish that, the nonzero coefficients in the optimal solution of (4) for the relative interior points are able to detect which points belong to the same submanifold under mild conditions.



(a) Two rings (b) Two trefoil-knots

Fig. 1. Toy data illustration.

Estimate Intrinsic Dimension to Prune Minor Coefficients. Since that problem (4) has no thresholding effect on the minor coefficients, we adopt the intrinsic dimension as the cue information to prune those minor coefficients. While the intrinsic dimension can be estimated by different methods [29]–[33], we use the most classical method— local PCA [29], which examines the eigenvalue spectra of the local covariance matrices, i.e., performing a local PCA in the neighborhood of each data point. To be specific, the first step in the process is to calculate the local PCA for each data point. Then, we take an average of the eigenvalues on the entire data set, and estimate the intrinsic dimension by the number of eigenvalues obtained when 95% energy is retained.

Once the intrinsic dimension $\hat{d}$ is estimated, as in [34], we keep the $\hat{d} + 1$ largest nonnegative coefficients and set those minor coefficients to zero. By doing so, we preserve only the connections from $\hat{d} + 1$ neighbors—which are in high probability from the same submanifold. The comparison between the performance with or without the pruning step is given in experiments.

For clarity, we summarize our proposed approach as following three steps:

- For data point $\boldsymbol{x}_j$, find its k nearest neighbors and compute the local convex representation for $\boldsymbol{x}_j$ via (2).
- Estimate the intrinsic dimension $\hat{d}$ via local PCA and keep the leading $\hat{d} + 1$ coefficients.
- Induce the affinity via $a_{ij} = \frac{1}{2}(c_{ij} + c_{ji})$ and apply spectral clustering.

III. Experiments

In this section, we evaluate the proposed approach on synthetic and real data to validate its effectiveness.

A. Experiments on Synthetic Data

We consider two rings embedded in $\mathbb{R}^3$ (see Fig. 1(a)), where the two manifolds are close to each other. From Fig. 1(a), we observe that the two rings are very close to each other on one side. We perform LLMC [23], SMCE [27], and LCR, respectively. Experiments show that the number of the misclassified points is 9, 0, and 0, respectively, when using $k = 5$. Furthermore, we find the data point (i.e., no. 55, at where the two rings are closet to each other) and its 5 nearest neighbors (i.e., no. 54, 56 on the red ring, and 126, 127, and 128 at the blue ring), and list the reconstruction coefficients computed by LLMC, SMCE, and LCR, respectively, in Table I. We can see that both SMCE and LCR yield the reasonably good results; whereas LLMC does not. Nevertheless, while the nonzero coefficients of LCR are not perfect compared to

SMCE, the first two dominating coefficients correctly relate to the submanifold to which the data point (i.e., no.55) belongs.

TABLE I
COMPARISON ON COEFFICIENTS COMPUTED BY DIFFERENT METHODS

Nonzero Coefficients	w_{54}	w_{56}	w_{126}	w_{127}	w_{128}
LLMC [23]	0.4949	0.4949	-0.0591	0.1284	-0.0591
SMCE [27]	0.4989	0.4989	0	0	0
LCR	0.4976	0.4976	4.63e-8	0.0047	4.63e-8

In Fig. 1(b), we show two trefoil-knots in $\mathbb{R}^3$. As in [27], we generate the synthetic data, which are sampled from two trefoil-knots embedded in $\mathbb{R}^{100}$ and corrupted with small Gaussian noise. Note that the data points are sampled such that among the 2 nearest neighbors of 1% of the data points there are points from the other submanifold. Also, among the 3 and 5 nearest neighbors of 9% and 18% of the data points, respectively, there are points from the other submanifold. For such points, the nearest neighbors-based methods [19], [20], [23] will connect the nearby points in the other submanifold and assign large coefficients to the connection. We compare LCR with SMCE [27] and LLMC [23], and report the experimental results in Table II. As can be read that, SMCE yields perfect clustering accuracy when λ is larger.[2] While LLMC and LCR cannot yield the perfect clustering, LLMC and LCR when combined with the pruning step lead to perfect clustering. This suggests the effectiveness of the pruning step.

TABLE II
COMPARISONS ON CLUSTERING ACCURACY (%) OF DIFFERENT METHODS AS A FUNCTION OF PARAMETER λ OR k.

λ	0.1	1	10	50	70	100	200
SMCE [27]	84.5	94.0	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
k	2	3	4	5	6	8	10
LLMC [23]	87.0	89.5	86.5	88.5	87.5	65.0	60.0
LLMC + $\hat{d}$	87.0	91.5	**100.0**	**100.0**	94.5	80.0	63.5
LCR	87.0	85.5	86.5	85.5	87.5	62.0	61.0
LCR + $\hat{d}$	87.0	88.5	**100.0**	**100.0**	**100.0**	93.0	90.5

B. Experiments on Real World Datasets

To verify the effectiveness of our approach, we conduct experiments on datasets COIL-20, COIL-100 and UMIST.

The Columbia Object Image Library (COIL). COIL-20 is a database of gray-scale images of 20 objects (see Fig. 2(a)). The objects were placed on a motorized turntable against a black background. The turntable was rotated through 360 degrees to vary object pose with respect to a fix camera. Images of the objects were taken at pose intervals of 5 degrees. This corresponds to 72 images per object. COIL-100 contains 100 objects with 72 images per object. In experiments, we resize each image into 32×32.

The Face Database UMIST. The UMIST Face Database consists of 1012 images of 20 individuals (mixed race/gender/apparence) (see Fig. 2(b)). The images are taken under a range of continuous pose changes. We resize each image into 32×32.

[2]Unfortunately, setting λ larger cannot lead to additional performance improvements on COIL20, COIL100 and UMIST.

(a) COIL-20 (b) UMIST

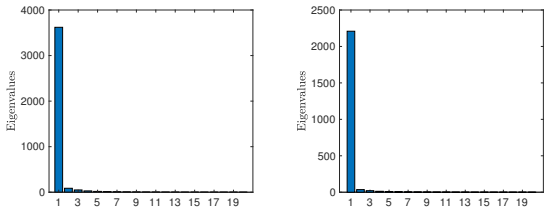
Fig. 2. Illustration of Samples in Datasets.

TABLE III
CLUSTERING ACCURACY (%) ON COIL-20, COIL-100 AND UMIST

Method	COIL-20	COIL-100	UMIST
k-means [35]	64.64	46.46	53.60
SSC [6]	85.14	55.00	58.50
EnSC-ORGEN [9]	74.25	69.24	68.30
SSC-OMP [10]	54.10	33.61	50.79
Kernel SSC [7]	75.35	52.82	57.20
LLMC [23]	92.57	77.78	72.43
SMCE [27]	92.29	80.75	68.77
LCR	85.00	76.36	72.13
LLMC + $\hat{d}$	80.56	80.06	51.68
LCR + $\hat{d}$	**100.00**	**91.51**	**76.98**

Experimental Results. To evaluate the performance of LCR, we select the following seven baseline algorithms, including k-means [35], SSC [6], EnSC-ORGEN [9], SSC-OMP [10], Kernel SSC [7], LLMC [23], and SMCE [27]. The hyper-parameters in each algorithm are tuned to gain the best performance. To prune the minor coefficients in LCR, we estimate the intrinsic dimension of data set using local PCA. To be specific, we find the eigenvalues of the local covariance matrix within a local neighborhood of 20 nearest neighbors, and take an average of the eigenvalues on the entire dataset. We show the averaged eigenvalue bars in Fig. 3. The big eigenvalues gap between the first leading eigenvalue and the second eigenvalue indicates the intrinsic dimension $\hat{d}$ as 1, which can also be calculated by preserving 95% energy of the eigenvalues. The estimated intrinsic dimension is consistent with the generation process of the dataset.

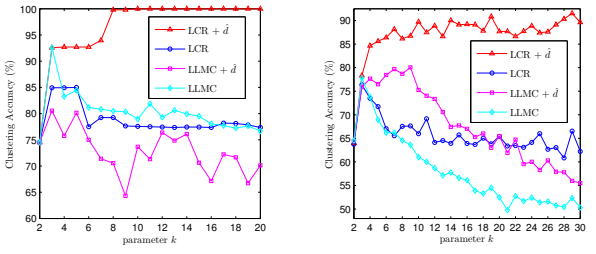
Experimental results are reported in Table III. As could be read, SMCE and LLMC are the two well-performed baselines, which even excel the performance of LCR. Nevertheless, if the pruning approach is used, LCR+$\hat{d}$ yields the best clustering accuracy on all the three datasets. Note that, while LLMC also yields remarkable performance on COIL-20 and UMIST, using the pruning strategy in LLMC degenerate the performance. In addition, while SMCE [27] also yields very competitive results, there is a need to tune the tradeoff parameter λ, which does not exist in LCR. Compare to tuning the parameter λ, estimating the intrinsic dimension with local PCA is more practical in a clustering task.



(a) COIL-20 (b) COIL-100

Fig. 3. The eigenvalues computed by local PCA with 20 nearest neighbors.

To evaluate the clustering accuracy with different local neighborhood parameter k, we conduct experiments on datasets COIL-20 and COIL-100 and display the clustering accuracy curves as a function of the parameter k in Fig. 4. As can be observed, while LLMC and LCR yield comparable clustering accuracy in small k, the performance degenerates dramatically when using a larger k. When the pruning strategy is used, the clustering accuracy of LCR is improved significantly if k is not too small; whereas LLMC does not.



(a) COIL-20 (b) COIL-100

Fig. 4. Comparison on clustering accuracy with/without pruning

IV. Conclusion

We have proposed an efficient method for manifold clustering, called Local Convex Representation. By building its connection to the ℓ_1 minimization based local affine sparse representation with an extra nonnegativity constraint, we have revealed that for some data points the nonzero coefficients of LCR are guaranteed to detect the data points lying on the same submanifold under mild conditions. Moreover, we have validated by experiments on synthetic data and real world datasets that the clustering accuracy of LCR can be significantly improved by keeping only the dominating $\hat{d}+1$ coefficients, where $\hat{d}$ is the intrinsic dimension of the submanifold estimated via local PCA.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant 61876022 and the Open Project Fund from the Key Laboratory of Machine Perception (MOE), Peking University.

References

[1] R. Souvenir and R. Pless, "Manifold clustering," in *ICCV*, vol. 1, 2005, pp. 648–653.
[2] H. E. Cetingül and R. Vidal, "Sparse riemannian manifold clustering for hardi segmentation," in *International Symposium on Biomedical Imaging: From Nano to Macro*, 2011, pp. 1750–1753.
[3] M. Hayat, M. Bennamoun, and A. A. El-Sallam, "Clustering of video-patches on grassmannian manifold for facial expression recognition from 3d videos," in *Workshop on Applications of Computer Vision*, 2013, pp. 83–88.
[4] H. E. Cetingül, M. J. Wright, P. M. Thompson, and R. Vidal, "Segmentation of high angular resolution diffusion mri using sparse riemannian manifold clustering," *IEEE Trans. on Medical Imaging*, vol. 33, no. 2, pp. 301–317, 2013.
[5] R. Vidal, "Subspace clustering," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 52–68, 2011.
[6] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *CVPR*, 2009, pp. 2790–2797.
[7] V. M. Patel and R. Vidal, "Kernel sparse subspace clustering," in *ICIP*, 2014, pp. 2849–2853.
[8] C.-G. Li and R. Vidal, "Structured sparse subspace clustering: A unified optimization framework," in *CVPR*, 2015, pp. 277–286.
[9] C. You, C.-G. Li, D. P. Robinson, and R. Vidal, "Oracle based active set algorithm for scalable elastic net subspace clustering," in *CVPR*, 2016, pp. 3928–3937.
[10] C. You, D. Robinson, and R. Vidal, "Scalable sparse subspace clustering by orthogonal matching pursuit," in *CVPR*, 2016, pp. 3918–3927.
[11] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *NIPS*, 2002, pp. 849–856.
[12] Y. Panagakis and C. Kotropoulos, "Elastic net subspace clustering applied to pop/rock music structure analysis," *Pattern Recognition Letters*, vol. 38, pp. 46–53, 2014.
[13] D. Barbará and P. Chen, "Using the fractal dimension to cluster datasets," in *ACM KDD*, 2000, pp. 260–264.
[14] E. Arias-Castro, "Clustering based on pairwise distances when the data is of mixed dimensions," *IEEE Trans. on Information Theory*, vol. 57, no. 3, pp. 1692–1706, 2011.
[15] A. Gionis, A. Hinneburg, S. Papadimitriou, and P. Tsaparas, "Dimension induced clustering," in *ACM KDD*, 2005, pp. 51–60.
[16] P. Mordohai and G. G. Medioni, "Unsupervised dimensionality estimation and manifold learning in high-dimensional spaces by tensor voting." in *IJCAI*, 2005, pp. 798–803.
[17] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *NIPS*, 2002, pp. 585–591.
[18] ——, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
[19] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
[20] L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds," *Journal of Machine Learning Research*, vol. 4, no. 6, pp. 119–155, 2003.
[21] M. Polito and P. Perona, "Grouping and dimensionality reduction by locally linear embedding," in *NIPS*, 2002, pp. 1255–1262.
[22] A. Tsai, C.-F. Westin, A. O. Hero, and A. S. Willsky, "Fiber tract clustering on manifolds with dual rooted-graphs," in *CVPR*, 2007, pp. 1–6.
[23] A. Goh and R. Vidal, "Segmenting motions of different types by unsupervised manifold clustering," in *CVPR*, 2007, pp. 1–6.
[24] E. Arias-Castro, G. Chen, G. Lerman *et al.*, "Spectral clustering based on local linear approximations," *Electronic Journal of Statistics*, vol. 5, pp. 1537–1587, 2011.
[25] G. Chen and G. Lerman, "Spectral curvature clustering (scc)," *International Journal of Computer Vision*, vol. 81, no. 3, pp. 317–330, 2009.
[26] Y. Wang, Y. Jiang, Y. Wu, and Z.-H. Zhou, "Spectral clustering on multiple manifolds," *IEEE Trans. on Neural Networks*, vol. 22, no. 7, pp. 1149–1161, 2011.
[27] E. Elhamifar and R. Vidal, "Sparse manifold clustering and embedding," in *NIPS*, 2011, pp. 55–63.
[28] C.-G. Li, C. You, and R. Vidal, "On geometric analysis of affine sparse subspace clustering," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1520–1533, 2018.
[29] K. W. Pettis, T. A. Bailey, A. K. Jain, and R. C. Dubes, "An intrinsic dimensionality estimator from near-neighbor information," *IEEE Trans. PAMI*, no. 1, pp. 25–37, 1979.
[30] E. Levina and P. J. Bickel, "Maximum likelihood estimation of intrinsic dimension," in *NIPS*, 2005, pp. 777–784.
[31] C. Szepesvári, J.-Y. Audibert *et al.*, "Manifold-adaptive dimension estimation," in *ICML*, 2007, pp. 265–272.
[32] K. M. Carter, R. Raich, and A. O. Hero III, "On local intrinsic dimension estimation and its applications," *IEEE Trans. on Signal Processing*, vol. 58, no. 2, pp. 650–663, 2009.
[33] M. D. Gupta and T. S. Huang, "Regularized maximum likelihood for intrinsic dimension estimation," *arXiv:1203.3483*, 2012.
[34] C.-G. Li, J. Guo, and H. Zhang, "Learning bundle manifold by double neighborhood graphs," in *ACCV*, 2009, pp. 321–330.
[35] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. PAMI*, no. 7, pp. 881–892, 2002.