CLUSTERING DISJOINT SUBSPACES VIA SPARSE REPRESENTATION

Ehsan Elhamifar *René Vidal*

Center for Imaging Science, Johns Hopkins University, Baltimore MD 21218, USA

ABSTRACT

Given a set of data points drawn from multiple low-dimensional linear subspaces of a high-dimensional space, we consider the problem of clustering these points according to the subspaces they belong to. Our approach exploits the fact that each data point can be written as a *sparse* linear combination of all the other points. When the subspaces are *independent*, the sparse coefficients can be found by solving a linear program. However, when the subspaces are *disjoint*, but not independent, the problem becomes more challenging. In this paper, we derive theoretical bounds relating the principal angles between the subspaces and the distribution of the data points across all the subspaces under which the coefficients are guaranteed to be sparse. The clustering of the data is then easily obtained from the sparse coefficients. We illustrate the validity of our results through simulation experiments.

Index Terms— Subspace clustering, sparsity, subspace angles, disjoint subspaces.

1. INTRODUCTION

Given a set of data points drawn from a union of subspaces, subspace clustering refers to the problem of finding the number of subspaces, their dimensions, a basis for each subspace, and the segmentation of data. This is an important problem, which finds numerous applications in signal and image processing, *e.g.*, [1, 2], and computer vision, *e.g.*, [3, 4]. In most of these applications the data are embedded in high-dimensional spaces, while the underlying subspaces are low-dimensional. Under this assumption, a number of methods have been developed, including algebraic methods [5], spectral clustering-based methods [3] and statistical methods [6].

Recently, [7] proposed a subspace clustering method based on sparse representation. The key observation is that each data point in a union of linear subspaces can be written as a *sparse* linear combination of all other points. The segmentation of data can be obtained by applying spectral clustering to a similarity matrix built from the sparse coefficients. The question is then: how to efficiently find such a sparse representation? [7] proves that for *independent subspaces*, *i.e.*, subspaces such that the dimension of their union is the sum of their individual dimensions, a sparse representation can be found exactly by using convex ℓ_1 optimization. However, requiring the subspaces to be independent is a strong assumption in practice. For instance, when segmenting multiple rigid-body motions in a video sequence, the subspaces become partially dependent for articulated objects, or for objects moving in a common plane, as shown in [4].

In this paper, we address the more general problem of clustering *disjoint subspaces*. That is, every pair of subspaces intersects only at the origin, but the dimension of the union of all subspaces need not be the sum of their dimensions. We show that under certain conditions relating the principal angles between the subspaces and the distribution of the data points across all the subspaces, the sparse coefficients can still be found by using convex ℓ_1 optimization. This result represents a significant generalization with respect to the sparse recovery literature, which addresses the sparse recovery of signals in a *single* subspace, see *e.g.*, [8]. The subspace clustering problem addressed in this paper is also much more challenging than recent work on block-sparse signal recovery, whose goal is to write a signal as a linear combination of a small number of known subspace bases [9, 10, 11]. First, we do not know the basis for any of the subspaces nor do we know which data points belong to which subspace. Second, we do not have any restriction on the number of subspaces, while existing methods require this number to be large.

2. PROBLEM STATEMENT

Let $\{S_i\}_{i=1}^n$ be an arrangement of n linear subspaces of $\mathbb{R}^D$ of dimensions $\{d_i\}_{i=1}^n$. We will distinguish between the following two types of arrangements.

Definition 1 *n subspaces $\{S_i\}_{i=1}^n$ are called* independent *if $dim(\bigoplus_{i=1}^n S_i) = \sum_{i=1}^n dim(S_i)$, where $\bigoplus$ is the direct sum.*

Definition 2 *Two subspaces are said to be* disjoint *if they intersect only at the origin. n subspaces $\{S_i\}_{i=1}^n$ are said to be* disjoint *if every two subspaces are disjoint.*

Notice that the notion of independence is stronger than the notion of disjointedness. For example, $n \geq 3$ distinct lines in $\mathbb{R}^2$ are disjoint, but not independent.

Consider now a given collection of $N = \sum_{i=1}^n N_i$ noise-free data points drawn from the n subspaces $\{S_i\}_{i=1}^n$. We denote the matrix whose columns are the N_i points drawn from subspace S_i as $\boldsymbol{Y}_i = \big[\boldsymbol{y}_{i1}, \ \cdots, \ \boldsymbol{y}_{iN_i}\big] \in \mathbb{R}^{D \times N_i}$ and the matrix containing all the data points as

$\boldsymbol{Y} = \begin{bmatrix}\boldsymbol{Y}_1, & \ldots, & \boldsymbol{Y}_n\end{bmatrix}\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma} \in \mathbb{R}^{N\times N}$ is an unknown permutation matrix. We assume that we do not know *a priori* the bases for each one of the subspaces nor do we know which data points belong to which subspace. The *subspace clustering* problem refers to the problem of finding the number of subspaces, their dimensions, a basis for each subspace, and the segmentation of the data from $\boldsymbol{Y}$.

The sparse subspace clustering (SSC) algorithm (see [7]) addresses the subspace clustering problem using techniques from sparse representation theory. This algorithm is based on the observation that each data point $\boldsymbol{y} \in S_i$ can always be written as a linear combination of all the other data points in $\{S_i\}_{i=1}^n$. However, generically, the *sparsest* representation is obtained when the point $\boldsymbol{y}$ is written as a linear combination of points in its own subspace. In this case, the number of nonzero coefficients corresponds to the dimension of the subspace. It is shown in [7] that when the subspaces are independent and low-dimensional, *i.e.*, $d_i \ll D$, this sparse representation can be obtained by using ℓ_1 minimization. The segmentation of the data is found by applying spectral clustering to a similarity graph formed using the sparse coefficients. More specifically, the SSC algorithm proceeds as follows.

Algorithm 1 : Sparse Subspace Clustering (SSC)

Input: A set of points $\{\boldsymbol{y}_i\}_{i=1}^N$ lying in n subspaces $\{S_i\}_{i=1}^n$.

1: For every data point $\boldsymbol{y}_i$, solve the following optimization problem:
$$\min \|\boldsymbol{c}_i\|_1 \quad \text{subject to} \quad \boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{c}_i \tag{1}$$
where $\boldsymbol{X}_i = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_N]$.

2: Form a similarity graph with N nodes representing the N data points. Connect node i, representing $\boldsymbol{y}_i$, to the other $N-1$ nodes by edge weights $w_{ij} = |c_{ij}| + |c_{ji}|$.

3: Form the graph Laplacian matrix $\boldsymbol{L} \in \mathbb{R}^{N\times N}$. Infer the segmentation of the data from the n eigenvectors of $\boldsymbol{L}$ corresponding to the n smallest eigenvalues using the K-means algorithm [12].

Output: Segmentation of the data: $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$.

In this paper, we extend the SSC algorithm to the more general class of *disjoint subspaces*. We show that under certain conditions on the principal angles between the subspaces and the distribution of the data points across all the subspaces, a sparse representation can still be recovered by ℓ_1 minimization. The subspace angles are defined as follows.

Definition 3 *The* first (smallest) principal angle *between two subspaces S_i and S_j, denoted by θ_{ij}, is defined as:*
$$\cos(\theta_{ij}) = \max_{\mathbf{x}\in S_i, \boldsymbol{z}\in S_j} \frac{\mathbf{x}^\top \boldsymbol{z}}{\|\mathbf{x}\|_2\|\boldsymbol{z}\|_2} \tag{2}$$

From the definition it is clear that for two disjoint subspaces, the first principal angle is always greater than zero.

3. CLUSTERING OF DISJOINT SUBSPACES

In this section, we derive conditions that generalize SSC from independent to disjoint subspaces. Specifically, we show that when the smallest angle between any two subspaces is greater than a bound determined by the distribution of the data points across all the subspaces, a *sparse subspace representation* can be found by ℓ_1 minimization. By sparse subspace representation we mean writing a point in a subspace as a sparse linear combination of data points in the same subspace. Notice that, by contrast with the standard sparse recovery problem [8] or the block-sparse recovery problem [9, 10, 11], we are not concerned with the uniqueness of such a sparse representation, as long as it comes from the data points in the true subspace.

Our approach for tackling the sparse subspace recovery problem is the following. We will show that the ℓ_1 norm of the sparse vector of coefficients for an arbitrary point in each subspace S_i with respect to all data points from the same subspace, $\boldsymbol{Y}_i$, is bounded from above. That is, for each $\boldsymbol{Y}_i$, there is a $\beta_i > 0$ such that for all $\boldsymbol{y} \in S_i$ we have $\|\boldsymbol{c}_i^*\|_1 \leq \beta_i$, where
$$\boldsymbol{c}_i^* = \text{argmin}\|\boldsymbol{c}\|_1 \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{Y}_i\,\boldsymbol{c}. \tag{3}$$
We will also show that the ℓ_1 norm of the sparse vector of coefficients of an arbitrary point in each subspace S_i with respect to the data points from all other subspaces, $\widehat{\boldsymbol{Y}}_i = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{i-1}, \boldsymbol{Y}_{i+1}, \ldots, \boldsymbol{Y}_n]$, is bounded from below. That is, for each i there exists a $\gamma_i > 0$ such that for all $\boldsymbol{y} \in S_i$, we have $\gamma_i \leq \|\widehat{\boldsymbol{c}}_i^*\|_1$, where
$$\widehat{\boldsymbol{c}}_i^* = \text{argmin}\|\boldsymbol{c}\|_1 \quad \text{subject to} \quad \boldsymbol{y} = \widehat{\boldsymbol{Y}}_i\,\boldsymbol{c}. \tag{4}$$
Finally, we will show in Theorem 1 that the sufficient condition $\beta_i < \gamma_i$ for all $i \in \{1, 2, \cdots, n\}$ guarantees the correctness of the ℓ_1 minimization, because it implies that
$$\|\boldsymbol{c}_i^*\|_1 \leq \beta_i < \gamma_i \leq \|\widehat{\boldsymbol{c}}_i^*\|_1. \tag{5}$$
Loosely speaking, this condition ensures that each point prefers to write itself as a linear combination of data points in its own subspace, hence the SSC algorithm is applicable.

3.1. Upper bound on the norm of the solution

We will first establish the upper bound on the ℓ_1 norm of the solution of the minimization problem when we choose points from the same subspace.

Let $\mathbb{W}_i$ be the set of all submatrices $\breve{\boldsymbol{Y}}_i \in \mathbb{R}^{D\times d_i}$ of $\boldsymbol{Y}_i$ that are full column rank. We can write any point $\boldsymbol{y} \in S_i$ as:
$$\boldsymbol{y} = \breve{\boldsymbol{Y}}_i\breve{\boldsymbol{c}}_i \Rightarrow \breve{\boldsymbol{c}}_i = (\breve{\boldsymbol{Y}}_i^\top\breve{\boldsymbol{Y}}_i)^{-1}\breve{\boldsymbol{Y}}_i^\top\boldsymbol{y}. \tag{6}$$
Using vector and matrix norm properties, we have:
$$\begin{aligned}\|\breve{\boldsymbol{c}}_i\|_1 &\leq \sqrt{d_i}\|\breve{\boldsymbol{c}}_i\|_2 = \sqrt{d_i}\,\|(\breve{\boldsymbol{Y}}_i^\top\breve{\boldsymbol{Y}}_i)^{-1}\breve{\boldsymbol{Y}}_i^\top\boldsymbol{y}\|_2 \\ &\leq \sqrt{d_i}\,\|(\breve{\boldsymbol{Y}}_i^\top\breve{\boldsymbol{Y}}_i)^{-1}\breve{\boldsymbol{Y}}_i^\top\|_{2,2}\|\boldsymbol{y}\|_2 = \frac{\sqrt{d_i}}{\sigma_{d_i}(\breve{\boldsymbol{Y}}_i)}\,\|\boldsymbol{y}\|_2,\end{aligned} \tag{7}$$

where $\sigma_l(\breve{\boldsymbol{Y}}_i)$ denotes the l-th largest singular value of $\breve{\boldsymbol{Y}}_i$. Thus, for the optimization problem in (3), we get:

$$\|\boldsymbol{c}_i^*\|_1 \leq \min_{\breve{\boldsymbol{Y}}_i \in \mathbb{W}_i} \|\breve{\boldsymbol{c}}_i\|_1 \leq \min_{\breve{\boldsymbol{Y}}_i \in \mathbb{W}_i} \frac{\sqrt{d_i}}{\sigma_{d_i}(\breve{\boldsymbol{Y}}_i)} \|\boldsymbol{y}\|_2 \triangleq \beta_i. \quad (8)$$

3.2. Lower bound on the norm of the solution

In this part, we consider the problem of writing a point $\boldsymbol{y} \in S_i$ as a linear combination of data points in all subspaces except S_i. Our goal is to lower bound the ℓ_1 norm of $\widehat{\boldsymbol{c}}_i^*$ given by (4).

Let $\widehat{\boldsymbol{c}}_i^* = [\widehat{\boldsymbol{c}}_1^\top, \ldots, \widehat{\boldsymbol{c}}_{i-1}^\top, \widehat{\boldsymbol{c}}_{i+1}^\top, \ldots, \widehat{\boldsymbol{c}}_n^\top]^\top$ denote the solution of (4). We have:

$$\boldsymbol{y} = \widehat{\boldsymbol{Y}}_i \widehat{\boldsymbol{c}}_i^* \quad (9)$$

If we multiply both sides of (9) from left by $\boldsymbol{y}^\top$, we get:

$$\|\boldsymbol{y}\|_2^2 = \boldsymbol{y}^\top \boldsymbol{y} = \boldsymbol{y}^\top \widehat{\boldsymbol{Y}}_i \widehat{\boldsymbol{c}}_i^* \quad (10)$$

Applying the Holder's inequality ($|\boldsymbol{z}^* \mathbf{x}| \leq \|\boldsymbol{z}\|_\infty \|\mathbf{x}\|_1$) to the above equation, we obtain:

$$\|\boldsymbol{y}\|_2^2 \leq \|\widehat{\boldsymbol{Y}}_i^\top \boldsymbol{y}\|_\infty \|\widehat{\boldsymbol{c}}_i^*\|_1. \quad (11)$$

By recalling the definition of the smallest principal angle between two subspaces from Definition 3, we can write:

$$\begin{aligned} \|\boldsymbol{y}\|_2^2 &\leq \max_{j \neq i} \max_{1 \leq k \leq N_j} |\boldsymbol{y}^\top \boldsymbol{y}_{jk}| \, \|\widehat{\boldsymbol{c}}_i^*\|_1 \\ &\leq \max_{j \neq i} \cos(\theta_{ij}) \, \Delta_i \, \|\boldsymbol{y}\|_2 \, \|\widehat{\boldsymbol{c}}_i^*\|_1, \quad (12) \end{aligned}$$

where θ_{ij} is the first principal angle between S_i and S_j and

$$\Delta_i \triangleq \max_{j \neq i} \max_{1 \leq k \leq N_j} \|\boldsymbol{y}_{jk}\|_2. \quad (13)$$

In other words, Δ_i is the maximum ℓ_2 norm of all data points in all subspaces except subspace i. We can rewrite (12) as:

$$\gamma_i \triangleq \frac{\|\boldsymbol{y}\|_2}{\max_{j \neq i} \cos(\theta_{ij}) \, \Delta_i} \leq \|\widehat{\boldsymbol{c}}_i^*\|_1 \quad (14)$$

which establishes the lower bound on the ℓ_1 norm of the solution.

3.3. Sufficient condition for sparse subspace recovery

By combining the results in (8) and (14), we obtain our main result. Theorem 1 gives a sufficient condition under which the solution to the ℓ_1 optimization problem in (1) has nonzero elements corresponding only to the points in S_i.

Theorem 1 *Consider a collection of data points drawn from n subspaces $\{S_i\}_{i=1}^n$ of dimensions $\{d_i\}_{i=1}^n$. Let $\boldsymbol{Y}_i$ denote the data points on S_i and $\widehat{\boldsymbol{Y}}_i$ denote the data points on the other subspaces. Let $\mathbb{W}_i$ be the set of all full rank submatrices $\breve{\boldsymbol{Y}}_i \in \mathbb{R}^{D \times d_i}$ of $\boldsymbol{Y}_i$. If the sufficient condition*

$$\max_{\breve{\boldsymbol{Y}}_i \in \mathbb{W}_i} \sigma_{d_i}(\breve{\boldsymbol{Y}}_i) > \sqrt{d_i} \, \Delta_i \max_{j \neq i} \cos(\theta_{ij}) \quad (15)$$

is satisfied for all $i \in \{1, \ldots, n\}$, then for every nonzero $\boldsymbol{y} \in S_i$, the solution to the following optimization problem

$$\begin{bmatrix} \boldsymbol{c}_i^* \\ \widehat{\boldsymbol{c}}_i^* \end{bmatrix} = \text{argmin} \left\| \begin{bmatrix} \boldsymbol{c}_i \\ \widehat{\boldsymbol{c}}_i \end{bmatrix} \right\|_1 \text{ subject to } \boldsymbol{y} = [\boldsymbol{Y}_i, \ \widehat{\boldsymbol{Y}}_i] \begin{bmatrix} \boldsymbol{c}_i \\ \widehat{\boldsymbol{c}}_i \end{bmatrix} \quad (16)$$

gives the sparse subspace solution with $\boldsymbol{c}_i^* \neq 0$ and $\widehat{\boldsymbol{c}}_i^* = 0$.

Proof. We prove the theorem by contradiction. Assume $\widehat{\boldsymbol{c}}_i^* \neq 0$ and define

$$\bar{\boldsymbol{y}} \triangleq \boldsymbol{y} - \boldsymbol{Y}_i \boldsymbol{c}_i^* = \widehat{\boldsymbol{Y}}_i \widehat{\boldsymbol{c}}_i^* \in S_i.$$

Since $\bar{\boldsymbol{y}}$ lives in S_i, we can write it as a linear combination of points in S_i. Let

$$\bar{\boldsymbol{c}}_i^* = \text{argmin} \, \|\boldsymbol{c}\|_1 \quad \text{subject to} \quad \bar{\boldsymbol{y}} = \boldsymbol{Y}_i \, \boldsymbol{c}.$$

Also, from $\bar{\boldsymbol{y}} = \widehat{\boldsymbol{Y}}_i \widehat{\boldsymbol{c}}_i^*$, we have that $\bar{\boldsymbol{y}}$ is in the range space of $\widehat{\boldsymbol{Y}}_i$. Let

$$\widetilde{\boldsymbol{c}}_i^* = \text{argmin} \, \|\boldsymbol{c}\|_1 \quad \text{subject to} \quad \bar{\boldsymbol{y}} = \widehat{\boldsymbol{Y}}_i \, \boldsymbol{c}.$$

Since $\bar{\boldsymbol{y}} = \widehat{\boldsymbol{Y}}_i \widehat{\boldsymbol{c}}_i^* = \widehat{\boldsymbol{Y}}_i \widetilde{\boldsymbol{c}}_i^*$, from the optimality of $\widetilde{\boldsymbol{c}}_i^*$ in the above optimization problem, we have $\|\widetilde{\boldsymbol{c}}_i^*\|_1 \leq \|\widehat{\boldsymbol{c}}_i^*\|_1$. Using (8) and (14), the sufficient condition in (15) guarantees that (5) is satisfied for $\bar{\boldsymbol{c}}_i^*$ and $\widetilde{\boldsymbol{c}}_i^*$, so we have $\|\bar{\boldsymbol{c}}_i^*\|_1 < \|\widetilde{\boldsymbol{c}}_i^*\|_1 \leq \|\widehat{\boldsymbol{c}}_i^*\|_1$. Finally, from $\boldsymbol{y} = \boldsymbol{Y}_i(\boldsymbol{c}_i^* + \bar{\boldsymbol{c}}_i^*) = [\boldsymbol{Y}_i \ \widehat{\boldsymbol{Y}}_i] \begin{bmatrix} \boldsymbol{c}_i^* \\ \widehat{\boldsymbol{c}}_i^* \end{bmatrix}$ we get:

$$\|\boldsymbol{c}_i^* + \bar{\boldsymbol{c}}_i^*\|_1 \leq \|\boldsymbol{c}_i^*\|_1 + \|\bar{\boldsymbol{c}}_i^*\|_1 < \|\boldsymbol{c}_i^*\|_1 + \|\widehat{\boldsymbol{c}}_i^*\|_1 = \left\| \begin{bmatrix} \boldsymbol{c}_i^* \\ \widehat{\boldsymbol{c}}_i^* \end{bmatrix} \right\|_1$$

which contradicts optimality of $\begin{bmatrix} \boldsymbol{c}_i^{*\top} & \widehat{\boldsymbol{c}}_i^{*\top} \end{bmatrix}^\top$. ∎

Loosely speaking, the sufficient condition in Theorem 1 states that the first principal angle between any pair of subspaces needs to be above a certain bound that depends on the distribution of the data points across all the subspaces. Notice that this bound can be rather high when the norms of the data points are oddly distributed, *e.g.*, when the maximum norm of the data points in one subspace is much smaller than the minimum norm of the data points in all other subspaces. Since the segmentation of the data does not change when the data points are scaled, a weaker sufficient condition can be obtained when all the data points are scaled so that they have the same ℓ_2 norm. In this case, the sufficient condition in (15) reduces to

$$\max_{\breve{\boldsymbol{Y}}_i \in \mathbb{W}_i} \sigma_{d_i}(\breve{\boldsymbol{Y}}_i) > \sqrt{d_i} \max_{j \neq i} \cos(\theta_{ij}) \quad (17)$$

where the columns of all matrices $\breve{\boldsymbol{Y}}_i$ in $\mathbb{W}_i$ are normalized so that their ℓ_2 norm is equal to one. This suggest applying the SSC algorithm after normalizing the data points so that $\|\boldsymbol{y}_{jk}\|_2 = 1$, for all j and k.

4. EXPERIMENTAL RESULTS

In [7], SSC was applied to the segmentation of a large database of 155 video sequences. The results showed that SSC outperforms state-of-the-art motion segmentation algorithms by a factor of 3. In this section, our goal is not to demonstrate the performance of SSC on a large database. Instead, we aim to evaluate the validity of the proposed theoretical results through simulation experiments.

We consider $n = 3$ subspaces of the same dimension d embedded in a $D = 100$ dimensional space. We generate bases $\{\boldsymbol{U}_i\}_{i=1}^3$ for the subspaces in such a way that $\text{rank}([\boldsymbol{U}_1, \boldsymbol{U}_3]) = 2d$ and $\boldsymbol{U}_2$ is in the range space of $[\boldsymbol{U}_1, \boldsymbol{U}_3]$, *i.e.*, $\text{rank}([\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3]) = 2d$. For a fixed value of d, we change the minimum angle between subspaces, θ, as well as the number of points in each subspace N_g. For each pair of (θ, N_g), we compute two different errors. First, we measure the error in the sparse recovery of data points $\{\boldsymbol{y}_i\}_{i=1}^{3N_g}$, where $\|\boldsymbol{y}_i\|_2 = 1$ for all i, by

$$SpErr = \frac{1}{3N_g} \sum_{i=1}^{3N_g} (1 - \frac{\|\boldsymbol{c}^*_{iq_i}\|_1}{\|\boldsymbol{c}^*_i\|_1}) \in [0, 1], \qquad (18)$$

where $\boldsymbol{c}^*_i = [\boldsymbol{c}^{*\top}_{i1}, \boldsymbol{c}^{*\top}_{i2}, \boldsymbol{c}^{*\top}_{i3}]^\top$ is the solution of the ℓ_1 minimization and $q_i \in \{1, 2, 3\}$ denotes the index of the subspace to which $\boldsymbol{y}_i$ belongs. Second, we measure the percentage of data points which have been assigned to the wrong subspace (misclassification rate), after applying spectral clustering to the similarity graph obtained from the sparse coefficients. We generate subspaces with a minimum subspace angle θ in the interval $\cos(\theta) \in [0.50, 0.99]$. We also change the number of points in each subspace in the interval of $N_g \in [d + 1, 20d]$. For each (θ, N_g), we perform 20 trials and compute the average of each one of the two errors. The results for two different values of $d = 2$ and 5 are shown in Figures 1 and 2. Blue denotes zero error and red indicates high value of the error.

As the results show, when either θ or N_g is small, which means that the sufficient condition is less likely to be satisfied, we get higher errors. By increasing the values of θ or N_g, we get lower errors and when values of both (θ, N_g) are sufficiently large, we get zero error. Also, the results verify that the misclassification rate is highly dependent on the sparse recovery error. Specifically, both errors follow the same pattern showing that the success of clustering relies on the success of sparse subspace recovery.

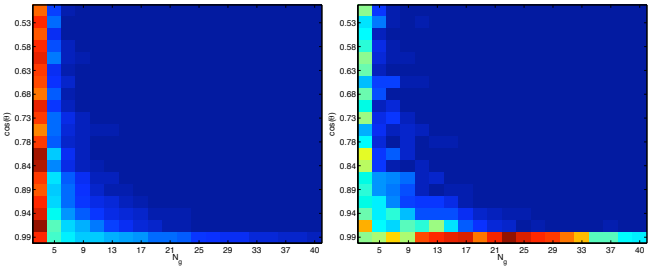

Fig. 1. Sparse recovery error (left) and misclassification rate (right) for $d = 2$ as a function of $(\cos(\theta), N_g)$.

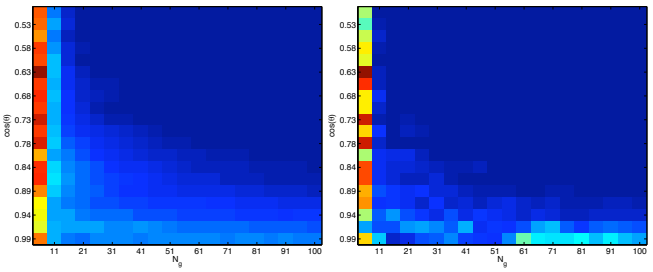

Fig. 2. Sparse recovery error (left) and misclassification rate (right) for $d = 5$ as a function of $(\cos(\theta), N_g)$.

5. ACKNOWLEDGMENT

This work was partially supported by startup funds from JHU, and by grants NSF CNS-0931805, NSF CNS-0834470.

6. REFERENCES

[1] W. Hong, J. Wright, K. Huang, and Y. Ma, "Multi-scale hybrid linear models for lossy image representation," *IEEE Trans. on Image Processing*, vol. 15, no. 12, pp. 3655–3671, 2006.

[2] Allen Y. Yang, John Wright, Yi Ma, and S. Shankar Sastry, "Unsupervised Segmentation of Natural Images Via Lossy Data Compression," *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 212–225, 2008.

[3] J. Yan and M. Pollefeys, "A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate," in *European Conf. on Computer Vision*, 2006, pp. 94–106.

[4] R. Vidal, R. Tron, and R. Hartley, "Multiframe motion segmentation with missing data using PowerFactorization and GPCA," *International Journal of Computer Vision*, vol. 79, no. 1, pp. 85–105, 2008.

[5] R. Vidal, Y. Ma, and S. Sastry, "Generalized Principal Component Analysis (GPCA)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1–15, 2005.

[6] M. Tipping and C. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Computation*, vol. 11, no. 2, pp. 443–482, 1999.

[7] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[8] David L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ^1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, Jun 2006.

[9] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inform. Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[10] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Compressed sensing of block-sparse signals: Uncertainty relations and efficient recovery," *submitted to in IEEE Trans. Signal Processing*, June 2009.

[11] A. Ganesh, Z. Zhou, and Y. Ma, "Separation of a subspace-sparse signal: Algorithms and conditions," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2009.

[12] Richard Duda, Peter Hart, and David Stork, *Pattern Classification*, Wiley-Interscience, October 2004.