Subspace Clustering with Priors via Sparse Quadratically Constrained Quadratic Programming*

Yongfang Cheng, Yin Wang, Mario Sznaier, Octavia Camps
Electrical and Computer Engineering
Northeastern University, Boston, MA 02115, US
{cheng.yong,wang.yin@husky.neu.edu}, {msznaier,camps}@coe.neu.edu

Abstract

This paper considers the problem of recovering a subspace arrangement from noisy samples, potentially corrupted with outliers. Our main result shows that this problem can be formulated as a convex semi-definite optimization problem subject to an additional rank constrain that involves only a very small number of variables. This is established by first reducing the problem to a quadratically constrained quadratic problem and then using its special structure to find conditions guaranteeing that a suitably built convex relaxation is indeed exact. When combined with the standard nuclear norm relaxation for rank, the results above lead to computationally efficient algorithms with optimality guarantees. A salient feature of the proposed approach is its ability to incorporate existing a-priori information about the noise, co-ocurrences, and percentage of outliers. These results are illustrated with several examples.

1. Introduction

Many practical problems involve fitting piecewise models to a given set of sample points. Examples of applications include image compression [11], face recognition [3], motion segmentation [21], video segmentation [19] and system identification [15]. Due to its importance, a substantial research effort has been devoted to this problem, leading to many algorithms, that can be roughly classified into statistical, algebraic and self-representation based.

RANdom SAmple Consensus (RANSAC) [8] is an iterative approach that proceeds by fitting one subspace at each iteration to as many points as possible, using a sampling based approach, removing these inliers from the dataset and repeating the process, until a given threshold on the percentage of inliers has been exceeded. While in principle the algorithm provides robust estimates of the parameters of the subspaces, it may require a large number of iterations to do so. On the other hand, due to its random nature, limiting the number of iterations may lead to arbitrarily bad solutions.

Algebraic methods such as GPCA [16], exploit the properties of subspace arrangements by reducing the problem to estimating the coefficients of a multivariate polynomial from noisy measurements of its zeros. Once this polynomial has been found, the parameters of each subspace can be recovered via polynomial differentiation. While GPCA works well with clean data, its performance degrades quickly with the noise level. This drawback has motivated the approach in [18], where the original data is "cleaned" via rank minimization. Although this approach is shown to be capable of handling substantial noise level, its main drawback is its computational complexity. In addition, in the presence of noise there are no guarantees that the resulting polynomial can be factored as a product of linear forms (and hence the parameters of the subspaces are recovered).

Due to these drawbacks, several methods have been recently proposed to handle noisy samples by exploiting the geometric properties of subspaces to reduce the problem to that of looking for sparse or low rank solutions to a set of linear equations that encode the fact that subspaces are self-expressive (e.g. a point in a subspace can be expressed as a linear combination of other points in it). These methods include Sparse Subspace Clustering (SSC) [7], Robust PCA (RPCA) [5], Low Rank Representation (LRR) [13], Fixed Rank Representation (FRR) [14] and Robust Subspace Clustering (RSC) [19]. All of these methods typically involve using relaxations (such as nuclear norm for rank and the ℓ_1 norm for sparsity), in order to obtain tractable convex problems[1]. While in the noiseless case these relaxations are exact under suitable conditions on the distribution of the data, in the presence of noise such guarantees are usually lost. Further, finding the parameters of the subspaces re-

*This work was supported in part by NSF grants IIS–1318145 and ECCS–1404163; AFOSR grant FA9550-15-1-0392; and the Alert DHS Center of Excellence under Award Number 2013-ST-061-ED0001.

[1]FRR uses directly a non-convex formulation, but it is shown that, for noiseless data, the global optimum has a closed form solution.

quires performing first a spectral clustering to cluster the data. Thus, there is no direct control on the fitting error.

Paper contributions:

Motivated by these difficulties, in this paper we propose an alternative method for recovering a subspace arrangement from noisy samples. Its main idea is to recast the problem as a rank constrained semi-definite program, which in turn is relaxed to a sequence of convex optimization problems by using a reweighted nuclear norm as a surrogate for rank. We also provide easily testable conditions certifying that the relaxation is exact. Specifically, the contributions of the paper are:

- Establishing that the problem of subspace clustering can be recast into a quadratically constrained quadratic program (QCQP).
- Exploiting the sparse structure underlying this QCQP to show that it is equivalent to a convex semi-definite program subject to an additional (non-convex) rank constraint that involves only a very small number of variables (roughly the number of parameters needed to characterize the subspaces). Notably, the size of this constraint is independent of the number of data points to be clustered.
- Using the results above, together with the special sparse structure of the problem, to obtain convex relaxations whose computational complexity scales linearly with the number of data points, along with conditions certifying optimality of these relaxations.
- Developing a clustering algorithm that, contrary to most existing techniques, directly identifies a set of subspace parameters that guarantees a fitting error lower than a given bound. Further, this algorithm can easily accommodate existing co-ocurrence information (points known to be in the same or different subspaces), bounds on the number of outliers, and priors on the relative frequency of each subspace, to improve clustering performance. To the best of our knowledge, this ability is not shared by any other existing method.

The above contributions are illustrated with several examples, including both synthetic and real data, where the ability to incorporate priors is key to obtaining the correct segmentation.

2. Preliminaries

2.1. Notation

$\mathbb{R}, \mathbb{N}$	set of real number and non-negative integers
$\mathbf{I}$	Identity matrix
$\mathbf{M} \succeq \mathbf{N}$	the matrix $\mathbf{M} - \mathbf{N}$ is positive semidefinite
$\sigma_i(\mathbf{A})$	the i-th largest singular value of matrix $\mathbf{A}$
$\mathrm{Tr}(\mathbf{A})$	trace of the square matrix $\mathbf{A}$

2.2. Some properties of QCQP

In this paper we will reduce the subspace clustering problem to a QCQP of the form:

$$\begin{aligned} p^* = \quad & \min_{\mathbf{v}\in\mathbb{R}^n} \quad f_0(\mathbf{v}) \\ & \text{s.t.} \quad f_i(\mathbf{v}) \leq 0, \forall_{i=1}^q \\ & \qquad \mathbf{A}\mathbf{v} \leq \mathbf{b}, \end{aligned} \tag{1}$$

where $f_i(\mathbf{v}) = \mathbf{v}^T\mathbf{Q}_i\mathbf{v} + \mathbf{c}_i^T\mathbf{v} + d_i$ and $\mathbf{Q}_i \in \mathbb{R}^{n\times n}$ is a symmetric matrix, for $\forall i = 0, 1, \ldots, q$, $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. In the case where $\mathbf{Q}_i \succeq \mathbf{0}$ for each i, the QCQP (1) is a convex programming problem that can be solved in polynomial time [2]. On the other hand, it is well known that, in the general case the problem is NP-hard. Nevertheless, since these problems are ubiquitous in a wide range of areas, extensive efforts have been devoted to developing tractable relaxations and associated optimality certificates (see for instance [6, 4, 2] and references therein). In particular, a relaxation of interest in this paper can be obtained by introducing a new variable $\mathbf{V} \in \mathbb{R}^{n\times n}$ and rewriting (1) as:

$$\begin{aligned} p^* = \quad & \min_{\mathbf{v},\mathbf{V}} \quad \mathrm{Tr}(\mathbf{Q}_0\mathbf{V}) + \mathbf{c}_0^T\mathbf{v} + d_0 \\ & \text{s.t.} \quad \mathrm{Tr}(\mathbf{Q}_i\mathbf{V}) + \mathbf{c}_i^T\mathbf{v} + d_i \leq 0, \forall_{i=1}^q \\ & \qquad \mathbf{A}\mathbf{v} \leq \mathbf{b} \\ & \qquad \mathbf{V} = \mathbf{v}\mathbf{v}^T, \end{aligned} \tag{2}$$

where the non-convexity appears now only in the last equality constraint. A convex relaxation of this problem that provides a lower bound of the cost can now be obtained by replacing $\mathbf{V} = \mathbf{v}\mathbf{v}^T$ with the convex positive semidefiniteness constraint $\mathbf{V} - \mathbf{v}\mathbf{v}^T \succeq \mathbf{0}$, leading to the semidefinite program (SDP) [22]:

$$\begin{aligned} \tilde{p}^* = \quad & \min_{\mathbf{v},\mathbf{V}} \quad \mathrm{Tr}(\mathbf{Q}_0\mathbf{V}) + \mathbf{c}_0^T\mathbf{v} + d_0 \\ & \text{s.t.} \quad \mathrm{Tr}(\mathbf{Q}_i\mathbf{V}) + \mathbf{c}_i^T\mathbf{v} + d_i \leq 0, \forall_{i=1}^q \\ & \qquad \mathbf{A}\mathbf{v} \leq \mathbf{b} \\ & \qquad \begin{bmatrix} 1 & \mathbf{v}^T \\ \mathbf{v} & \mathbf{V} \end{bmatrix} \succeq \mathbf{0} \end{aligned} \tag{3}$$

We will refer to this as the SDP relaxation of (1).

Remark 1. *Since (3) is a relaxation of (1), then $\tilde{p}^* \leq p^*$. A trivial sufficient condition for $\tilde{p}^* = p^*$ is $\mathbf{V}^* = \mathbf{v}^*\mathbf{v}^{*T}$, that is, rank of $\begin{bmatrix} 1 & \mathbf{v}^{*T} \\ \mathbf{v}^* & \mathbf{V}^* \end{bmatrix}$ is 1.*

2.3. A reduced QCQP relaxation

While the relaxation (3) is convex, it has relatively poor scaling properties, due to the semi-definite constraint (recall that for $n \times n$ matrices, the computational complexity of these constraints scales as n^6). However, the problem often (as in this paper) has an underlying sparse structure that can be exploited to mitigate this growth. For notational

simplicity and without loss of generality (by absorbing linear terms into quadratic ones), rewrite the problem as

$$p^* = \min_{\mathbf{v}\in\mathbb{R}^n} f_0(\mathbf{v}) \text{ s.t. } f_i(\mathbf{v}) \leq 0, \forall_{i=1}^{q} \tag{4}$$

where $f_i(\mathbf{v}) = \mathbf{v}^T\mathbf{Q}_i\mathbf{v} + \mathbf{c}_i^T\mathbf{v} + d_i$.

Assume that $\{\mathbf{v}_k\}_{k=1}^{l}$ are subsets of $\mathbf{v}$, satisfying $\cup_{k=1}^{l}\mathbf{v}_k = \mathbf{v}$, each constraint $f_i(\mathbf{v})$ for $i = 1,\ldots,q$, depends only on a subsets of variables $\mathbf{v}_k$, and that the objective function $f_0(\mathbf{v})$ can be partitioned into a sum of the form $f_0(\mathbf{v}) = \sum_{k=1}^{l} p_k(\mathbf{v}_k)$. Problem (1) is said to satisfy the running intersection property [12] if there exists a reordering $\mathbf{v}_{k'}$ of $\mathbf{v}_k$ such that for every $k' = 1,\ldots,l-1$:

$$\mathbf{v}_{k'+1} \cap (\cup_{j=1}^{k'}\mathbf{v}_j) \subseteq \mathbf{v}_s \text{ for some } s \leq k'. \tag{5}$$

Then a convex relaxation of (1) can be obtained by replacing the condition $\mathbf{M} \doteq \begin{bmatrix} 1 & \mathbf{v}^T \\ \mathbf{v} & \mathbf{V} \end{bmatrix} \succeq \mathbf{0}$ with positive semi-definiteness of a collection of smaller matrices as follows:

$$\begin{aligned} \tilde{p}^*_{sparse} = \min_{\mathbf{M}_k} \quad & \textstyle\sum_{k=1}^{l} \mathrm{Tr}(\bar{\mathbf{Q}}_{k,0}\mathbf{M}_k) \\ \text{s.t.} \quad & \mathrm{Tr}(\bar{\mathbf{Q}}_i\mathbf{M}_i) \leq 0, \forall_{i=1}^{q} \\ & \mathbf{M}_k \succeq \mathbf{0}, \forall_{k=1}^{l} \\ & \mathbf{M}_i(I_{ij}, I_{ij}) = \mathbf{M}_j(I_{ij}, I_{ij}), \forall_{i,j=1}^{l}, i \neq j \end{aligned} \tag{6}$$

where $\bar{\mathbf{Q}}_{k,0}$ and $\bar{\mathbf{Q}}_i$ are symmetric matrices built from $\{\mathbf{Q}_i, \mathbf{c}_i, d_i\}$. $\mathbf{M}_i(I_{ij}, I_{ij})$ denotes the block of $\mathbf{M}_i$ with rows and columns corresponding to $\begin{bmatrix} 1 & \mathbf{v}_i^T \cap \mathbf{v}_j^T \end{bmatrix}^T$.

In the sequel, we will refer to the relaxation above as the "reduced SDP" relaxation. Clearly, $\tilde{p}^*_{sparse} \leq \tilde{p}^* \leq p^*$, since the relaxation (6) is looser than (3). However, under certain conditions, the equality holds.

Theorem 1. *If (4) satisfies the running intersection, then a sufficient condition for $\tilde{p}^*_{sparse} = \tilde{p}^* = p^*$ is that rank$(\mathbf{M}_k) = 1,\ k = 1,\ldots,l$.*

Proof. This is a special case of Theorem 3.7 in [12]. □

3. Problem Statement

The goal of this paper is to estimate a set of subspaces from noisy samples such that certain priors are satisfied, or show that none exists. Formally, this can be stated as:

Problem 1. *Given:*

- *A set of noisy samples $\mathbf{X} = \{\mathbf{x}_j \in \mathbb{R}^n : \mathbf{x}_j = \hat{\mathbf{x}}_j + \boldsymbol{\eta}_j\}_{j=1}^{N_p}$, drawn from N_s distinct linear subspaces $\{\mathcal{S}_i \subset \mathbb{R}^n\}_{i=1}^{N_s}$ of dimension $n-1$ of the form $\mathcal{S}_i = \{\hat{\mathbf{x}} \in \mathbb{R}^n : \mathbf{r}_i^T\hat{\mathbf{x}} = 0, \mathbf{r}_i \in \mathbb{R}^n, \|\mathbf{r}_i\|_2 = 1\}$.*
- *A-priori information consisting of (i) a bound ϵ on the distance from the noisy sample to the subspace it is drawn from, (ii) a bound N_o on the number of outliers, (iii) N_{f_i}, the relative frequency of each subspace, and (iv) point wise co-occurrence information.*

Establish whether the data is consistent with the a-priori assumptions and, if so, find a set of subspaces compatible with the a-priori information and assign the (inlier) samples to each subspace. That is, find $\{\mathbf{r}_i \in \mathbb{R}^n\}_{i=1}^{N_s}$ and a partition of the samples in $N_s + 1$ sets $\{\mathbf{X}_i\}_{i=1}^{N_s}, \mathbf{X}_o$ such that card$(\mathbf{X}_o) \leq N_o$ and card$(\mathbf{X}_i) = N_{f_i}N_p$ for each $i = 1,\ldots,N_s$,and

$$|\mathbf{r}_i^T\mathbf{x}| \leq \epsilon \text{ holds for } \forall \mathbf{x} \in \mathbf{X}_i. \tag{7}$$

4. A Convex Approach to Clustering

In this section we present the main result of this paper, a convex optimization approach to solving Problem 1. The main idea is to first recast the problem into an equivalent non-convex QCQP, which in turn can be reduced to an SDP subject to a non-convex rank constraint by exploiting the results outlined in section 2.2. Next, by exploiting the structure of the problem, we show that this non-convex rank constraint needs to be enforced only on a single matrix of a small size (substantially smaller than those involved in the reduced SDP relaxation). Finally, combining these results with standard nuclear norm surrogates for rank leads to the desired algorithm. For simplicity we will consider first the case with no outliers ($N_o = 0$), and without constraints on the relative frequency and co-ocurrences. The handling of these cases will be covered in sections 4.4 and 4.5 after presenting the basic algorithm and supporting theory.

4.1. Clustering as a Nonconvex QCQP

It can be easily shown that by introducing a set of binary variables $\{s_{i,j}\}$ that indicate whether $\mathbf{x}_j$ is drawn from $\mathcal{S}_i$ or not, Problem 1 is equivalent to:

Problem 2. *Determine the feasibility of the following set of quadratic inequalities:*

$$\begin{cases} |s_{i,j}\mathbf{r}_i^T\mathbf{x}_j| \leq \epsilon s_{i,j}, \forall_{i=1}^{N_s}\forall_{j=1}^{N_p} & \text{(8a)} \\ s_{i,j}^2 = s_{i,j}, s_{i,j} \geq 0, \forall_{i=1}^{N_s}\forall_{j=1}^{N_p} & \text{(8b)} \\ \Sigma_{i=1}^{N_s} s_{i,j} = 1, \forall_{j=1}^{N_p} & \text{(8c)} \\ \mathbf{r}_i^T\mathbf{r}_i = 1, \forall_{i=1}^{N_s} & \text{(8d)} \\ \mathbf{r}_1(1) \geq \mathbf{r}_2(1) \geq \cdots \geq \mathbf{r}_{N_s}(1) \geq 0 & \text{(8e)} \end{cases}$$

Here, (8a) is equivalent to $|\mathbf{r}_i^T\mathbf{x}_j| \leq \epsilon$ if $s_{i,j} \neq 0$ (hence $\mathbf{x}_j \in \mathbf{X}_j$) and trivially satisfied otherwise; (8b) imposes that $s_{i,j} \in \{0,1\}$; (8c) forces each sample $\mathbf{x}_i$ to be assigned to exactly one subspace; and (8e) eliminates the symmetry of the solutions. Thus, if (8) is feasible, then the subspaces recovered are characterized by their normals $\{\mathbf{r}_i\}$. On the other hand, infeasibility of (8) invalidates the a-priori information given in Problem 1.

Clearly, Problem 2 is a QCQP of the form (1), albeit non-convex due to the constraints (8a), (8b) and (8d). Applying

the SDP relaxation outlined in Section 2.2 leads to the following result.

Theorem 2. *Problem 2 is equivalent to establishing feasibility of*

$$\begin{cases} Tr(\mathbf{Q}_k\mathbf{M}) \leq 0, \forall_{k=1}^{K} & \text{(9a)} \\ \mathbf{M} \succeq \mathbf{0},\ \mathbf{M}(1,1) = 1 & \text{(9b)} \\ rank(\mathbf{M}) = 1 & \text{(9c)} \end{cases}$$

where (9a) denotes the linear (in)equalities,and $\mathbf{M}(i,j)$ *denotes the (i,j) entry of* $\mathbf{M}$.

Proof. It follows from Remark 1 by collecting all variables in (8) in a vector $\mathbf{v} \in \mathbb{R}^{N_s(n+N_p)}$, that is,

$$\mathbf{v} \doteq [\mathbf{r}_1^T, \cdots, \mathbf{r}_{N_s}^T, s_{1,1}, \cdots, s_{N_s,1}, \cdots, s_{1,j}, \cdots, s_{N_s,j}, \cdots, s_{1,N_p}, \cdots, s_{N_s,N_p}]^T,$$

and defining the rank 1 matrix $\mathbf{M} = \begin{bmatrix} 1 & \mathbf{v}^T \\ \mathbf{v} & \mathbf{v}\mathbf{v}^T \end{bmatrix}$. □

Corollary 1. *Consider the convex SDP problem*

$$\begin{cases} Tr(\mathbf{Q}_k\mathbf{M}) \leq 0, \forall_{k=1}^{K} \\ \mathbf{M} \succeq \mathbf{0},\ \mathbf{M}(1,1) = 1 \end{cases} \quad (10)$$

If (10) is infeasible, then Problem 2 is also infeasible. On the other hand, if this problem admits a rank 1 solution $\mathbf{M}^*$*, then* $\mathbf{M}^*(2:(n+N_p)N_s+1, 1)$ *is also a feasible solution to Problem 2.*

In principle, from the results above, one could attempt to solve Problem 2 by solving (10) or (9) where the rank constraint is relaxed to one involving minimizing the reweighted nuclear norm. Note however, that both (10) and (9) require solving SDPs whose computational complexity scales as $[N_s(n+N_p)+1]^6$, limiting the use of the algorithm to relatively few points.As shown next, these difficulties can be circumvented by exploiting the sparse structure of the problem.

4.2. Exploiting the Sparse Structure

To exploit the sparse structure of the problem, partition the constraints in (8) into the N_p+1 sets P_j, $j = 0, 1, \ldots, N_p$:

$$P_0 : \begin{cases} \mathbf{r}_i^T\mathbf{r}_i = 1, \forall_{i=1}^{N_s} \\ \mathbf{r}_1(1) \geq \mathbf{r}_2(1) \geq \cdots \geq \mathbf{r}_{N_s}(1) \geq 0, \end{cases}$$

$$\forall_{j=1}^{N_p}, P_j : \begin{cases} |s_{i,j}\mathbf{r}_i^T\mathbf{x}_j| \leq \epsilon s_{i,j}, \forall_{i=1}^{N_s} \\ s_{i,j}^2 = s_{i,j}, s_{i,j} \geq 0,\ \forall_{i=1}^{N_s} \\ \sum_{i=1}^{N_s} s_{i,j} = 1. \end{cases}$$

It is easily seen that P_0 is only associated with variables $\mathbf{v}_0 = [\mathbf{r}_1^T, \ldots, \mathbf{r}_{N_s}^T]^T \in \mathbb{R}^{nN_s}$, for $j = 1, \ldots, N_p$, P_j is only associated with variables $\mathbf{v}_j = [\mathbf{v}_0^T, s_{1,j}, \ldots, s_{N_s,j}]^T \in \mathbb{R}^{(n+1)N_s}$, and that each P_j can be reformulated as a set of quadratic constraints of the form

$$\mathbf{v}_j^T\mathbf{Q}_{k,j}\mathbf{v}_j + \mathbf{c}_{k,j}^T\mathbf{v}_j + d_{k,j} \leq 0, \forall_{k=1}^{K_j}, \quad (11)$$

where K_j is the number of constraints in P_j.

Notice that $\mathbf{v}_j \cap \left(\cup_{k=0}^{j-1}\mathbf{v}_k\right) = \mathbf{v}_0$ holds for $\forall j = 1, \ldots, N_p$ and that $\cup_{j=0}^{N_p}\mathbf{v}_j = \mathbf{v}$. Thus, (8) exhibits the running intersection property and the results in Section 2.3 allow to reformulate (9) involving positive semi-definiteness of matrices of substantially smaller size than that of $\mathbf{M}$ in (9). To this effect, introduce positive semi-definite matrices of the form $\mathbf{M}_j = \begin{bmatrix} 1 & m_j(\mathbf{v}_j^T) \\ m_j(\mathbf{v}_j) & m_j(\mathbf{v}_j\mathbf{v}_j^T) \end{bmatrix}$ for $j = 0, 1, \ldots, N_p$, where $m_j(\bullet)$ is a variable locating in the same position as $\bullet$ in $\begin{bmatrix} 1 & \mathbf{v}_j^T \\ \mathbf{v}_j & \mathbf{v}_j\mathbf{v}_j^T \end{bmatrix}$, and consider the following rank constrained SDP problem:

Problem 3. *Determine the feasibility of*

$$\begin{cases} Tr(\bar{\mathbf{Q}}_{k,j}\mathbf{M}_j) \leq 0, \forall_{k=1}^{K_j}, \forall_{j=0}^{N_p} & \text{(12a)} \\ \mathbf{M}_j \succeq \mathbf{0}, \mathbf{M}_j(1,1) = 1, \forall_{j=0}^{N_p} & \text{(12b)} \\ \mathbf{M}_j(1:nN_s+1, 1:nN_s+1) = \mathbf{M}_0, \forall_{j=1}^{N_p} & \text{(12c)} \\ rank(\mathbf{M}_j) = 1, \forall_{j=0}^{N_p} & \text{(12d)} \end{cases}$$

where $\bar{\mathbf{Q}}_{k,j} \doteq \begin{bmatrix} d_{k,j} & 0.5\mathbf{c}_{k,j}^T \\ 0.5\mathbf{c}_{k,j} & \mathbf{Q}_{k,j} \end{bmatrix}$ *and* $\mathbf{M}$*(1:j,1:j) denotes the sub matrix formed by the first* j *rows and columns of* $\mathbf{M}$.

From Theorem 2 it follows that the problem above is equivalent to Problem 2. However, contrary to (10), it involves N_p+1 matrices of dimension around $[(n+1)N_s+1] \times [(n+1)N_s+1]$. Hence its computational complexity grows as $N_p[(n+1)N_s]^6$, that is, linearly with the number of data points. However, this formulation requires enforcing (N_p+1) rank constraints, a very challenging task. Surprisingly, as shown next, the special structure of the problem makes N_p of these constraints redundant, allowing for developing an equivalent of Problem 2 with a single rank constraint imposed on a $(nN_s+1) \times (nN_s+1)$ matrix.

Theorem 3. *Problem 2 is equivalent to checking the feasibility of*

$$\begin{cases} \text{(12a)} - \text{(12c)} \\ rank(\mathbf{M}_0) = 1. \end{cases} \quad (13)$$

Proof. Given in the Appendix. □

4.3. A Convex Optimization Based Algorithm

Theorem 3 suggests that a convex algorithm whose complexity scales linearly with the number of data points can

be obtained by seeking rank-1 solutions to (13) via iterative minimization of a re-weighted nuclear norm of $\mathbf{M}_0$ [17] subject to (12a)-(12c). This idea leads to Algorithm 1.

Algorithm 1 Subspace Clustering via QCQP

1: **Initialize:** $k=0$, $\mathbf{W}^{(0)}=\mathbf{I}$, $0<\delta\ll 1$, $k_{\max}$
2: **repeat**
3: solve

$$\{\mathbf{M}_j^{(k)}\} = \arg\min \quad \mathrm{Tr}(\mathbf{W}^{(k)}\mathbf{M}_0) \quad \text{s.t.} \quad (12a)-(12c)$$

4: update $\mathbf{W}^{(k+1)}=[\mathbf{M}_0^{(k)}+\sigma_2(\mathbf{M}_0^{(k)})]^{-1}$, $k=k+1$;
5: **until** $\sigma_2(\mathbf{M}_0^{(k)})<\delta\sigma_1(\mathbf{M}_0^{(k)})$ or $k>k_{\max}$.

4.4. Handling Outliers

Outliers, defined as points $\mathbf{x}$ that lie beyond a given distance ϵ from every subspace in the arrangement, (e.g. such that $\min_i |\mathbf{r}_i^T\mathbf{x}|>\epsilon$) can be handled by simply relaxing the requirement that each point must be assigned to a certain subspace, leading to the following QCQP:

$$\begin{aligned} p^* = \max \quad & \textstyle\sum_{j=1}^{N_p}\sum_{i=1}^{N_s} s_{i,j} \\ \text{s.t.} \quad & |s_{i,j}\mathbf{r}_i^T\mathbf{x}_j| \le \epsilon s_{i,j},\ s_{i,j}^2 = s_{i,j}\ \forall_{i=1}^{N_s}\forall_{j=1}^{N_p} \\ & \textstyle\sum_{i=1}^{N_s} s_{i,j} \le 1, \forall_{j=1}^{N_p} \\ & \mathbf{r}_i^T\mathbf{r}_i = 1, \forall_{i=1}^{N_s} \\ & \mathbf{r}_1(1)\ge\mathbf{r}_2(1)\ge\cdots\ge\mathbf{r}_{N_s}(1)\ge 0 \end{aligned} \tag{14}$$

which seeks to maximize the number of inliers.

Since (14) exhibits a sparsity pattern similar to that in (8), it can be solved by a modified version of Algorithm 1, where the equality constraint $\sum_{i=1}^{N_s} s_{i,j}=1$ is replaced by $\sum_{i=1}^{N_s} s_{i,j}\le 1$ (to accommodate the case where $\mathbf{x}_j$ is an outlier), and the objective function is replaced by

$$\tilde{p} = \Sigma_{i=1}^{N_s}\Sigma_{j=1}^{N_p}(1-m_j(s_{i,j})) + \lambda\mathrm{Tr}(\mathbf{W}^{(k)}\mathbf{M}_0), \tag{15}$$

where $\lambda>0$ is a parameter. Thus, this objective function penalizes both the number of outliers and the rank of $\mathbf{M}_0$.

Remark 2. *A bound N_o on the number of outliers can be handled via a constraint of the form $\sum_{i=1}^{N_s}\sum_{j=1}^{N_p} s_{i,j}\ge N_p-N_o$. However, since this constraint subverts the sparsity of the problem, the formulation (14) is preferable.*

4.5. Handling Additional A-Priori Information

In many scenarios of practical interest, a-priori information on the labels of some sample points is available. For instance, in surveillance videos, it is easy to identify points lying on buildings, and hence background, and often points belonging to moving targets. Similarly, in many situations, information is available about the size of the target, and thus on the relative frequency of points in the corresponding subspace. As shown below, this additional information can be incorporated into our formulation by simply adding suitable quadratic constraints on the variables $s_{i,j}$. Specifically:
(i) $f\%$ of $\mathbf{X}$ belongs to $\mathcal{S}_i \iff \sum_{j=1}^{N_p} s_{i,j} = 0.01 f N_p$;
(ii) $\mathbf{x}_m, \mathbf{x}_n$ belong to the same subspace $\iff s_{i,m} = s_{i,n}, \forall i=1,\cdots,N_s$;
(iii) $\mathbf{x}_m, \mathbf{x}_n$ belong to different subspaces $\iff s_{i,m}s_{i,n}=0, \forall i=1,\cdots,N_s$.
The advantages of being able to exploit this information will be illustrated in Section 6.2.

4.6. Recovery Properties

A salient feature of the proposed approach is its ability to certify optimality of the solution provided by Algorithm 1. Specifically, from Theorem 3 it follows that, in the case of data corrupted by outliers, $rank(\mathbf{M}_0)=1$ certifies that the correct clustering has been found. Similarly, in the presence of noise, this condition guarantees that the recovered subspaces fit the inliers within the given noise bound ϵ, and are consistent with the given a-priori information.

5. Further Complexity Reduction

As discussed in section 4.2, exploiting the sparse structure of the problem leads to Algorithm 1 which only requires imposing positive semi-definiteness on N_p+1 matrices of dimension at most $[(n+1)N_s+1]\times[(n+1)N_s+1]$. Hence, its complexity scales as $(nN_s)^6$. Further computational complexity reduction can be achieved by proceeding in a greedy fashion where subspaces are determined step by step rather than simultaneously as in Section 4.3. At each step, samples drawn from a specific subspace are considered as inliers while all other points are considered outliers. Thus, instead of introducing N_s binary variables $\{s_{i,j}\}_{i=1}^{N_s}$ for each sample as in Section 4.2, here only one binary variable s_j is needed to indicate whether $\mathbf{x}_j$ is an inlier. At each step the resulting problem reduces to a QCQP of the form:

$$\begin{aligned} p^* = \max_{s_j,\mathbf{r}} \quad & \textstyle\sum_{j=1}^{N_p} s_j \\ \text{s.t.} \quad & |s_j\mathbf{r}^T\mathbf{x}_j| \le \epsilon s_j,\ s_j^2 = s_j,\ \forall_{j=1}^{N_p} \\ & \mathbf{r}^T\mathbf{r}=1,\ \mathbf{r}(1)\ge 0 \end{aligned} \tag{16}$$

Proceeding as in section 4.2 it can be shown that this problem also exhibits the running intersection property. Combining this observation with a reasoning similar to the one used in the proof of Theorem 3 leads to the following result:

Theorem 4. *The problem (16) is equivalent to*

$$\begin{aligned} \tilde{p}^* = \max_{\mathbf{M}_j} \quad & \textstyle\sum_{j=1}^{N_p} m_j(s_j) \\ \text{s.t.} \quad & Tr(\bar{\mathbf{Q}}_{k,j}\mathbf{M}_j)\le 0, \forall_{k=1}^{K_j}, \forall_{j=0}^{N_p} \\ & \mathbf{M}_j\succeq\mathbf{0}, \mathbf{M}_j(1,1)=1, \forall_{j=0}^{N_p} \\ & \mathbf{M}_j(1:n+1,1:n+1)=\mathbf{M}_0, \forall_{j=1}^{N_p} \\ & rank(\mathbf{M}_0)=1. \end{aligned} \tag{17}$$

Remark 3. *Compared to the nongreedy formulation (12), the number of variables and the size of the positive semi-definite matrix in (17) are reduced to $\mathcal{O}(n^2)$ and $\mathcal{O}(n)$ respectively, independent of the number of subspaces N_s.*

This result leads to Algorithm 2 for subspace clustering.

Algorithm 2 Greedy Subspace Clustering by QCQP

1: **Initialize:** $n_s = 0$, $n_o = N_p$, $\mathbf{X}_o = \mathbf{X}$, $N = \{1, \ldots, N_o\}$;
2: **while** $n_o > n - 1$ **do**
3: $n_s = n_s + 1$;
4: solve (17) by re-weighted nuclear norm relaxation of rank with samples $\mathbf{X}_o$;
5: $J \subset N$ is the union of j with $s_j = 1$, then $\mathbf{x}_j \in \mathcal{S}_{n_s}$, $\mathbf{X}_o = \mathbf{X}_o \setminus \{\mathbf{x}_j, j \in J\}$, $N = N \setminus J$, and $n_o = n_o - \text{cardinality}(J)$;
6: **end while**

6. Experiments

In this section, we demonstrate the advantage of the proposed method using both synthetic data and a non-trivial application: planar segmentation by homography learning.

6.1. Synthetic Data

We first investigate the performance of the QCQP-based subspace clustering algorithm on synthetic data as the number of subspaces, their dimensions and noise levels changed. For each set of experiments, we randomly generated 20 instances with sample points lying on a union of multiple linear subspaces corrupted by noise, normal to the corresponding subspace, and with uniform random magnitudes in $[0.8\epsilon, \epsilon]$. A comparison of the performance of Algorithm 1, implemented in Matlab using CVX [9], against existing state-of-the-art methods is summarized in Table 1[2]. As shown there, in all cases the proposed method outperformed the others in terms of the worst-case fitting error of the identified subspaces, given by:

$$\text{err}_\text{f} = \max_{j \in \{1,\cdots,N_p\}} \min_{i \in \{1,\cdots,N_s\}} |\mathbf{r}_i^T \mathbf{x}_j| \quad \text{s.t. } ||\mathbf{r}_i||_2 = 1, \forall_{i=1}^{N_s}$$

where $\mathbf{r}_i$'s are the normals to the subspaces found by each algorithm. For algorithms that cannot obtain $\mathbf{r}_i$ directly, like SSC and LRR, we calculated $\mathbf{r}_i$ by fitting a subspace to each cluster of points produced by the algorithm.

Next, the effect of outliers was investigated, by randomly corrupting 4 to 6 points with noise of magnitude 3ϵ (the distribution of the data is shown in Fig. 1). Using Algorithm 1 modified to solve (14) instead of (8) led to the results shown in the last row of Table 1. As shown there, the proposed algorithm could detect the outliers (black dots) precisely and fit the inliers to subspaces within the given error bound.

[2]In order to provide a meaningful comparison, for each set of experiments, the adjustable parameters of each of the algorithms listed in the table were experimentally tuned to minimize the misclassification rate.

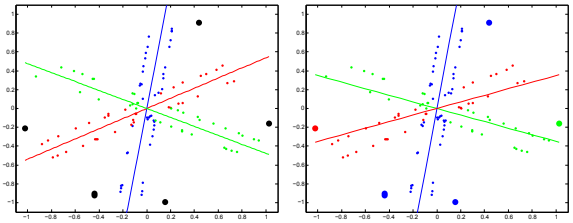

Figure 1: Fitting inliers vs. Fitting both Inliers and Outliers

6.2. Planar Segmentation

In this section we illustrate the advantages of taking into account prior information in subspace clustering by applying our algorithm to planar segmentation by homography estimation, which is an important problem in image registration and computation of camera motion [10]. Given the homogeneous coordinates of N_p correspondences from two images $\{(\mathbf{p}_j, \mathbf{p}'_j)\}_{j=1}^{N_p} \in \mathbb{R}^3$, assuming that these N_p points are on the same plane, let $\mathbf{p}_j = \begin{bmatrix} x_j & y_j & 1 \end{bmatrix}^T$ and $\mathbf{p}'_j = \begin{bmatrix} x'_j & y'_j & 1 \end{bmatrix}^T$, and let $\mathbf{H} \in \mathbb{R}^{3\times 3}$ denote the homography. Then $\mathbf{h} \doteq \text{vectorize}(\mathbf{H}^T)$ satisfies

$$\begin{bmatrix} \mathbf{x}_j^T \\ \mathbf{x}_{j+N_p}^T \end{bmatrix} \mathbf{h} = 0, \text{ with } \begin{cases} \mathbf{x}_j^T = \begin{bmatrix} \mathbf{p}_j^T & \mathbf{0}_{1\times 3} & -x'_j \mathbf{p}_j^T \end{bmatrix} \\ \mathbf{x}_{j+N_p}^T = \begin{bmatrix} \mathbf{0}_{1\times 3} & \mathbf{p}_j^T & -y'_j \mathbf{p}_j^T \end{bmatrix} \end{cases} \tag{18}$$

meaning $\mathbf{h}$ lies in the null space of the matrix $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{2N_p} \end{bmatrix}^T$.

Now assume that $\{(\mathbf{p}_j, \mathbf{p}'_j)\}_{j=1}^{N_p}$ are distributed on N_s different planes (shown as the purple dots in Figure 2). In this case (18) no longer holds for all $j = 1, \cdots, N_p$ with a single $\mathbf{h}$. Instead the planes can be segmented by clustering $\{\mathbf{x}_j\}_{j=1}^{2N_p}$ to N_s subspaces $\mathcal{S}_i$, characterized by different normal vectors $\mathbf{h}_i$, $i = 1, \cdots, N_s$ (shown as the blue dots and red dots in Fig. 2). Thus subspace clustering techniques can be applied to planar segmentation. Specifically, we can formulate this problem as seeking a feasible solution to:

$$\begin{cases} ||\mathbf{h}_i||_2^2 = 1, \forall_{i=1}^{N_s} & \text{(19a)} \\ |s_{i,j} \mathbf{h}_i^T \mathbf{x}_j| \le \epsilon s_{i,j}, \; s_{i,j}^2 = s_{i,j}, \forall_{i=1}^{N_s} \forall_{j=1}^{2N_p} & \text{(19b)} \\ \Sigma_{i=1}^{N_s} s_{i,j} = 1, \forall_{j=1}^{2N_p} & \text{(19c)} \\ s_{i,j} = s_{i,j+N_p}, \forall_{i=1}^{N_s} \forall_{j=1}^{N_p} & \text{(19d)} \end{cases}$$

where (19a)-(19c) define a subspace clustering problem similar to Problem 2, and (19d) represents the prior information that $\mathbf{x}_j$ and $\mathbf{x}_{j+N_p}$ should be in the same subspace since they are derived from a single correspondence $(\mathbf{p}_j, \mathbf{p}'_j)$ as in (18), which cannot be enforced by the existing subspace clustering methods.

Table 1: Performance comparison for different synthetic data scenarios, n: dimension of data, d_i: Dimension of each subspace, N_i: Number of samples on each subspace, ϵ: error bound, μ and σ: mean and standard deviation of err_r, (*): experiments with outliers.

n	d_i	N_i	ϵ	Algorithm 1	Denoised GPCA[18]	GPCA	SSC	LRR
3	[2 2]	[50 50]	0.10	$\mu = 0.0992$ $\sigma = 0.0003$	$\mu = 0.1015$ $\sigma = 0.0034$	$\mu = 0.1234$ $\sigma = 0.0230$	$\mu = 0.2465$ $\sigma = 0.0705$	$\mu = 0.3349$ $\sigma = 0.1565$
3	[2 2]	[50 50]	0.15	$\mu = 0.1485$ $\sigma = 0.0008$	$\mu = 0.1608$ $\sigma = 0.0140$	$\mu = 0.2980$ $\sigma = 0.1696$	$\mu = 0.3224$ $\sigma = 0.0831$	$\mu = 0.4054$ $\sigma = 0.1417$
3	[2 2]	[50 50]	0.20	$\mu = 0.1978$ $\sigma = 0.0010$	$\mu = 0.2020$ $\sigma = 0.0143$	$\mu = 0.4495$ $\sigma = 0.1459$	$\mu = 0.3793$ $\sigma = 0.1112$	$\mu = 0.4635$ $\sigma = 0.1083$
4	[3 3]	[50 50]	0.15	$\mu = 0.1481$ $\sigma = 0.0013$	$\mu = 0.2058$ $\sigma = 0.1029$	$\mu = 0.3229$ $\sigma = 0.1430$	$\mu = 0.3495$ $\sigma = 0.0588$	$\mu = 0.3622$ $\sigma = 0.0885$
2	[1 1 1]	[40 40 40]	0.10	$\mu = 0.0993$ $\sigma = 0.0003$	$\mu = 0.2190$ $\sigma = 0.0212$	$\mu = 0.4943$ $\sigma = 0.0173$	$\mu = 0.1649$ $\sigma = 0.0452$	$\mu = 0.1021$ $\sigma = 0.0071$
$2^{(*)}$	[1 1 1]	[40 40 40]	0.10	$\mu = 0.0996$ $\sigma = 0.0003$	$\mu = 0.4601$ $\sigma = 0.1355$	$\mu = 0.6362$ $\sigma = 0.1219$	$\mu = 0.2986$ $\sigma = 0.1086$	$\mu = 0.2223$ $\sigma = 0.0461$

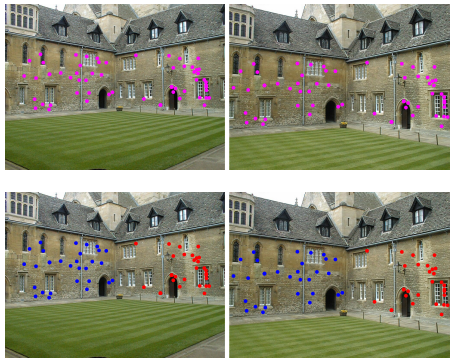

Figure 2: Images for Planar Segmentation: Merton I

We tested on three pairs of real images *Merton I*, *Merton II*, and *Wadham*, given by VGG, University of Oxford. Given each pair of images, firstly VLfeat toolbox [23] was used to obtain the SIFT features of two images, and correspondences were defined by those pairs of points whose ℓ_2 norm are less than 5. Among these correspondences, we randomly generated 20 instances with 30 correspondences on each plane and $N_s = 2$. ϵ was determined by calculating a ground truth homography matrix for each plane, $\mathbf{H}_1$, $\mathbf{H}_2$ and $\epsilon = \max\{\text{erf}_1, \text{erf}_2\}$, $\text{erf}_i = \max_{j \in \mathcal{S}_i} |\mathbf{x}_j^T \text{vec}(\mathbf{H}_i^T)|$, $i = 1, 2$. Performance was evaluated by the misclassification rate err_1 among $2N_p$ samples, with

$$\text{err}_\text{l} = \frac{\begin{pmatrix}\text{Number of Points with Different} \\ \text{Labels from the Ground Truth}\end{pmatrix}}{\text{Total Number of Sample Points } 2N_p} \times 100\%. \qquad (20)$$

Analysis. As reported in Table 2, our proposed approach outperformed GPCA and the denoised GPCA proposed in [18]. For LRR and SSC given by

$$\text{(LRR): } \min ||\mathbf{Z}||_* + \lambda ||\mathbf{E}||_{2,1}, \text{s.t. } \mathbf{X} = \mathbf{XZ} + \mathbf{E}$$
$$\text{(SSC): } \min ||\mathbf{C}||_1 + 0.5\tau ||\mathbf{E}||_F^2, \text{s.t. } \mathbf{X} = \mathbf{XC} + \mathbf{E}, \text{diag}(\mathbf{C}) = 0,$$

we plotted the performance for $\lambda \in [10^{-4}, 1]$ and $\tau \in [10^{-3}, 10]$ in Fig. 3, from which we can see that the range of λ (τ) for LRR (SSC) to be competitive with the proposed approach in terms of classification accuracy, is quite small, roughly $\lambda^* \in [0.003, 0.01]$, $\tau^* \in [0.4, 0.7]$. Thus, in this case, LRR and SSC were quite sensitive to the parameters, as shown in Fig. 3 and Fig. 4. In addition, such small values of λ^* (or τ^*) placed virtually no penalty on the noise

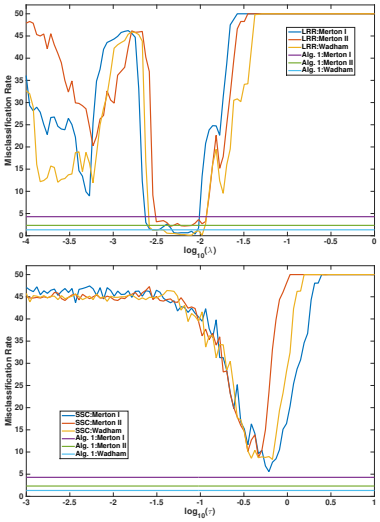

Figure 3: Average Performance of LRR and SSC over 20 Instances

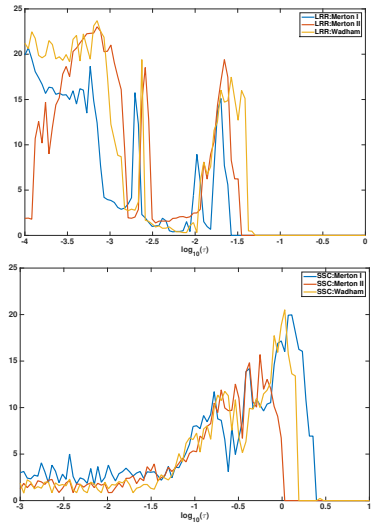

Figure 4: Variance of LRR and SSC over 20 Instances

terms. As a result, they yielded solutions where the denoised data $\mathbf{X} - \mathbf{E}$ fitted poorly the actual data. For example, for $\lambda = 0.01$, the misclassification rate of LRR for *Merton I* was 3.33%. However, as shown in Fig. 5, LRR produced a solution where where $\mathbf{X}$ and $\mathbf{E}$ had roughly the same scale. A similar situation (also shown in Fig. 5) arised with SSC when $\tau = 0.55$, the value yielding the lowest misclassification rate (5%). In contrast, the proposed method did not

have to be tuned and yielded an accurate estimate of the homography parameters.

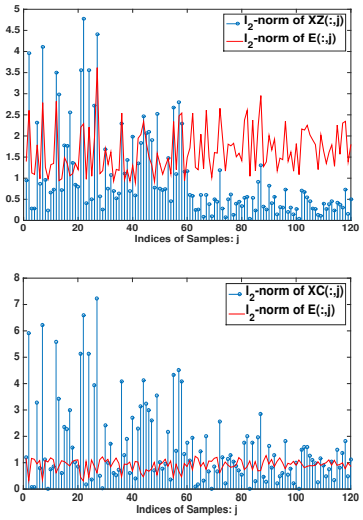

Figure 5: Denoised Data vs. Noise. (Left: LRR with $\lambda = 0.01$; Right: SSC with $\tau = 0.55$)

The specific reason why the existing methods performed worse is the structure of the samples for (19). It is easy to show that $\left[\begin{array}{c|c|c|c} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{N_p} \end{array}\right] = \left[\begin{array}{c} * \\ \hline \mathbf{0}_{3\times N_p} \\ \hline * \end{array}\right]$, and $\left[\begin{array}{c|c|c|c} \mathbf{x}_{N_p+1} & \mathbf{x}_{N_p+2} & \cdots & \mathbf{x}_{2N_p} \end{array}\right] = \left[\begin{array}{c} \mathbf{0}_{3\times N_p} \\ \hline * \\ \hline * \end{array}\right]$,where $*$ denotes the nonzero entries. Thus, without any prior information, the existing methods are likely to cluster $\{\mathbf{x}_j\}_{j=1}^{N_p}$ to a subspace, and $\{\mathbf{x}_j\}_{j=N_p+1}^{2N_p}$ to the second subspace. On the other hand, by enforcing the prior information that $\mathbf{x}_j$ and $\mathbf{x}_{N_p+j}$, for $j = 1, \ldots, N_p$, belong to the same subspace, the proposed approach has an extremely low misclassification rate. The running time of each method, averaged over 60 instances, is summarized in Table 3.

Table 2: Average Performance for Planar Segmentation: (%)

Dataset	Alg. 1	[18]	GPCA
Merton I	**4.29** ($\pm$3.45)	49.12 ($\pm$3.91)	50.00 ($\pm$0)
Merton II	**2.33** ($\pm$2.85)	49.58 ($\pm$0.66)	46.04 ($\pm$10.99)
Wadham	**1.33** ($\pm$2.19)	49.74 ($\pm$0.92)	45.69 ($\pm$7.02)

Table 3: Running Time (sec)

Alg. 1	[18]	GPCA	SSC	LRR
683.74	730.94	0.1421	0.1904	0.3553

7. Conclusions

In this paper we propose a new approach to the problem of identifying an arrangement of subspaces from noisy samples, potentially corrupted by outliers. The main idea is to recast the problem into a QCQP, which in turn can be solved by solving convex semi-definite programs. A salient feature of the proposed approach is its ability to exploit available a-priori information on the percentage of outliers, relative number of points in each subspace and partial labelings. These advantages were illustrated with several examples comparing the performance of the proposed method vis-à-vis existing ones. The main drawback of the proposed method stems from the need to solve semi-definite programs. However, exploiting the underlying sparse structure of the problem allows for imposing the semi-definite constraints only on a collection of smaller matrices, leading to an algorithm whose complexity scales linearly with the number of data points. Research is currently underway seeking to further reduce the computational burden.

A. Proof of Theorem 3

(Necessity) Suppose $\mathbf{v}^*$ is a feasible solution to (8), then the rank-1 matrices $\mathbf{M}_j^* = \begin{bmatrix} 1 & \mathbf{v}_j^{*T} \\ \mathbf{v}_j^* & \mathbf{v}_j^*\mathbf{v}_j^{*T} \end{bmatrix}$, $j = 0, .., N_p$, are a feasible solution to (12).

(Sufficiency) Suppose $\mathbf{M}_j^*$, $j = 0, \ldots, N_p$, is a feasible solution to (12). Then from $\text{rank}(\mathbf{M}_0^*) = 1$, we know

$$m_0(\mathbf{r}_i(k)^2)^* = m_0(\mathbf{r}_i(k))^{*2}, \ \forall_{i=1}^{N_s}\forall_{k=1}^{n}. \tag{21}$$

Combining (21) and $m_j(s_{i,j}^2)^* = m_j(s_{i,j})^*$, from $\mathbf{M}_j^* \succeq \mathbf{0}$, for $j = 1, \ldots, N_p$, it follows that its principal minor $\mathbf{M}_{j,i,k} \succeq \mathbf{0}$, $k = 1, .., n$, $i = 1, \ldots, N_s$,

$$\mathbf{M}_{j,i,k} = \left[\begin{array}{c|cc} 1 & m_j(\mathbf{r}_i(k))^* & m_j(s_{i,j})^* \\ \hline m_j(\mathbf{r}_i(k))^* & m_j(\mathbf{r}_i(k))^{*2} & m_j(\mathbf{r}_i(k)s_{i,j})^* \\ m_j(s_{i,j})^* & m_j(\mathbf{r}_i(k)s_{i,j})^* & m_j(s_{i,j})^* \end{array}\right],$$

and its Schur complement of the first block is also positive semi-definite, that is,

$$\begin{aligned} &\begin{bmatrix} m_j(\mathbf{r}_i(k))^{*2} & m_j(\mathbf{r}_i(k)s_{i,j})^* \\ m_j(\mathbf{r}_i(k)s_{i,j})^* & m_j(s_{i,j})^* \end{bmatrix} \\ &- \begin{bmatrix} m_j(\mathbf{r}_i(k))^* \\ m_j(s_{i,j})^* \end{bmatrix} \begin{bmatrix} m_j(\mathbf{r}_i(k))^* & m_j(s_{i,j})^* \end{bmatrix} \\ &= \begin{bmatrix} 0 & \Delta \\ \Delta & m_j(s_{i,j})^* - m_j(s_{i,j})^{*2} \end{bmatrix} \succeq \mathbf{0} \end{aligned} \tag{22}$$

where $\Delta = m_j(\mathbf{r}_i(k)s_{i,j})^* - m_j(\mathbf{r}_i(k))^* m_j(s_{i,j})^*$, which implies

$$\Delta = 0, \text{ or } m_j(\mathbf{r}_i(k)s_{i,j})^* = m_j(\mathbf{r}_i(k))^* m_j(s_{i,j})^*. \tag{23}$$

Substituting (23) into the linear inequality constraints (12a) associated with (8a): $\begin{cases} m_j(s_{i,j}\mathbf{r}_i)^{*T}\mathbf{x}_j \le \epsilon m_j(s_{i,j})^* \\ m_j(s_{i,j}\mathbf{r}_i)^{*T}\mathbf{x}_j \ge -\epsilon m_j(s_{i,j})^* \end{cases}$, leads to: $\begin{cases} m_j(s_{i,j})m_j(\mathbf{r}_i)^{*T}\mathbf{x}_j \le \epsilon m_j(s_{i,j})^* \\ m_j(s_{i,j})m_j(\mathbf{r}_i)^{*T}\mathbf{x}_j \ge -\epsilon m_j(s_{i,j})^* \end{cases}$, which is equivalent to

$$|m_j(\mathbf{r}_i)^{*T}\mathbf{x}_j| \le \epsilon, \tag{24}$$

for any $m_j(s_{i,j})^* > 0$. Thus, $\mathbf{x}_j$ belongs to the subspace normal to $m_j(\mathbf{r}_i)^*$. Finally, the conditions $m_j(s_{i,j})^* \ge 0$ and $\sum_{i=1}^{N_s} m_j(s_{i,j})^* = 1$ guarantee that for each $j = 1, \ldots, N_p$, there exists at least one $i_0 \in \{1, \ldots, N_s\}$, such that $m_j(s_{i_0,j})^* > 0$. Hence each point $\mathbf{x}_j$ belongs to at least one of the subspaces characterized by the normals $m_0(\mathbf{r}_i)^*$, $i = 1, \ldots, N_s$, which establishes feasibility of (8).

References

[1] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical programming*, 95(1):3–51, 2003.

[2] K. M. Anstreicher. On convex relaxations for quadratically constrained quadratic programming. *Mathematical programming*, 136(2):233–251, 2012. 2

[3] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(2):218–233, 2003. 1

[4] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004. 2

[5] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011. 1

[6] A. dAspremont and S. Boyd. Relaxations and randomized methods for nonconvex qcqps. *EE392o Class Notes, Stanford University*, 2003. 2

[7] E. Elhamifar and R. Vidal. Sparse subspace clustering. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2790–2797. IEEE, 2009. 1

[8] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 1

[9] M. Grant, S. Boyd, and Y. Ye. Cvx: Matlab software for disciplined convex programming, 2008. 6

[10] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 6

[11] W. Hong, J. Wright, K. Huang, and Y. Ma. Multiscale hybrid linear models for lossy image representation. *Image Processing, IEEE Transactions on*, 15(12):3655–3671, 2006. 1

[12] J. B. Lasserre. Convergent sdp-relaxations in polynomial optimization with sparsity. *SIAM Journal on Optimization*, 17(3):822–843, 2006. 3

[13] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. 2012. 1

[14] R. Liu, Z. Lin, F. De la Torre, and Z. Su. Fixed-rank representation for unsupervised visual learning. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 598–605. IEEE, 2012. 1

[15] Y. Ma and R. Vidal. Identification of deterministic switched arx systems via identification of algebraic varieties. In *Hybrid Systems: Computation and Control*, pages 449–465. Springer, 2005. 1

[16] Y. Ma, A. Y. Yang, H. Derksen, and R. Fossum. Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM review*, 50(3):413–458, 2008. 1

[17] K. Mohan and M. Fazel. Iterative reweighted algorithms for matrix rank minimization. *J. of Machine Learning Research*, 13:3441–3473, 2012. 5

[18] N. Ozay, M. Sznaier, C. Lagoa, and O. Camps. Gpca with denoising: A moments-based convex approach. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3209–3216. IEEE, 2010. 1, 7, 8

[19] M. Soltanolkotabi, E. Elhamifar, and E. Candes. Robust subspace clustering. *arXiv preprint arXiv:1301.2603*, 2013. 1

[20] A. Tewari, P. Ravikumar, and I. S. Dhillon. Greedy algorithms for structurally constrained high dimensional problems. *Advances in Neural Information Processing Systems 24*, pages 882–890, 2011.

[21] R. Tron and R. Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007. 1

[22] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996. 2

[23] A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. In *Proceedings of the international conference on Multimedia*, pages 1469–1472. ACM, 2010. 7