Subspace clustering through parametric representation and sparse optimization

Laurent Bako

Subspace clustering through parametric representation and sparse optimization

Laurent Bako

***Abstract*—We consider the problem of recovering a finite number of linear subspaces from a collection of unlabeled data points that lie in the union of the subspaces. The data are such that it is not known which data point originates from which subspace. To address this challenge, we show that the clustering problem is amenable to a sparse optimization problem. Considering a candidate subspace and the distances of the data points to that subspace, the foundation of the proposed method lies in the maximization of the number of zero distances. This can be relaxed into a convex optimization. Efficiency of the relaxation can be significantly increased by solving a sequence of reweighted convex optimization problems.**

***Index Terms*—Subspace Clustering, Sparse Optimization, Subspace Arrangement.**

I. Introduction

We consider the problem of estimating a finite number s of linear subspaces $\{\mathcal{S}_k\}_{k=1}^s$ of the euclidean vector space $\mathbb{R}^n$, with equal dimension d. More precisely, assume we are given a finite set $\{y_i\}_{i=1}^N$ of noise-free data points drawn from the subspace arrangement $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_s$. Then the problem of interest in this paper is to infer the individual subspaces $\mathcal{S}_1, \ldots, \mathcal{S}_s$ from the collected data. This can indeed be viewed as a clustering problem because the available data are mixed in the sense that we do not know which data point originates from which subspace.

The literature on subspace clustering/segmentation contains a variety of methods that can be roughly divided into four categories: algebraic methods [1], [2], [3], statistical methods [4], matrix factorization-based methods [5], [6], sparse representation-based methods [7], [8], [9]. For a detailed exposition of some of those methods we refer to the recent survey [10]. The more recent trend in the treatment of the subspace segmentation problem builds on some ideas borrowed from the field of compressed sensing [11], [12]. The subspace segmentation methods reported in [7] and [8] allow for an elegant retrieval of the subspaces, their dimensions and their number by solving a convex optimization problem and then applying spectral clustering. However, these methods assume that the subspaces to be estimated are independent, a requirement that imposes a constraint on the number and the dimensions of the subspaces. Another relevant method is the one described in [13] which relies on sparse representation over a fixed but optimized dictionary in a probabilistic framework.

In this note, we describe a technique for subspace clustering that uses sparse optimization. The method makes it possible to compute one subspace from the entire mixed dataset by minimizing the number of nonzero vectors obtained when the data are projected onto the orthogonal complement of a single subspace. In contrast to the methods mentioned above, the method presented here can estimate subspaces that are not independent, for example an arrangement of hyperplanes. The principle of the method is as follows. Considering a candidate subspace and the distances of the data points to that subspace, we propose to maximize the number of zero distances. Because this is a hard sparse optimization problem, we later consider a convex relaxation of it. While the initial sparse formulation requires only very mild conditions, the convex surrogate might necessitate a higher level of sparsity to yield the desired solution. This means that there must exist a subspace in the arrangement which contains the majority of the data points. To overcome this limitation, a sequence of reweighted convex optimization problems can be solved [14].

L. Bako is with Laboratoire Ampère – Ecole Centrale de Lyon – Université de Lyon, 69130 France. E-mail: laurent.bako@ec-lyon.fr

II. Description of the proposed method

We will assume, for the sake of exposition clarity, that all subspaces have equal and known dimension d. Also, the subspaces $\mathcal{S}_1, \ldots, \mathcal{S}_s$ are assumed to be discernible in the sense that for any pair (k, ℓ) of indices, $\mathcal{S}_k \not\subseteq \mathcal{S}_\ell$. But the method to be presented can be extended to arrangements of subspaces with different dimensions.

For any $i = 1, \ldots, s$, let $B_i \in \mathbb{R}^{n\times d}$, with $\mathrm{rank}(B_i) = d$, be a basis of the subspace $\mathcal{S}_i$. Denote with $\mathcal{S}_i^\perp$ the orthogonal complement of $\mathcal{S}_i$ in $\mathbb{R}^n$. If we assume noiseless data, then for any $y_j \in \mathcal{S}_i$, the orthogonal projection of y_j onto $\mathcal{S}_i^\perp$ is equal to zero, that is

$$(I_n - P_i)\, y_j = 0 \tag{1}$$

where P_i is the projection matrix defined by $P_i = B_i(B_i^\top B_i)^{-1}B_i^\top$ and I_n stands for the identity matrix of order n. Note in passing that P_i is uniquely determined by the subspace $\mathcal{S}_i$; it does not depend on any specific basis B_i of $\mathcal{S}_i$. For, if the matrix B_i forms a basis for $\mathcal{S}_i$, any other basis has necessarily the form $\bar{B}_i = B_i T$ with $T \in \mathbb{R}^{n\times n}$ being a nonsingular matrix. As a consequence, it can be checked that the corresponding projection matrix $\bar{P}_i$ is equal to P_i. We can therefore describe the subspaces $\mathcal{S}_i$ with the matrices P_i, $i = 1, \ldots, s$. Each P_i and $I_n - P_i$ are orthogonal projections matrices that project the ambiant space onto $\mathcal{S}_i$ and $\mathcal{S}_i^\perp$ respectively. Let us formalize.

Definition 1. *A square matrix $P \in \mathbb{R}^{n\times n}$ is called an orthogonal projection matrix (or an orthogonal projector) if P is idempotent and symmetric, that is, $P^2 = P$ and $P^\top = P$.*

At this point, it is worth recalling some useful properties of orthogonal projectors.

Lemma 1. 1) *If P_i is an orthogonal projection matrix onto $\mathcal{S}_i$, then* $\mathrm{im}(P_i) = \ker(I_n - P_i) = \mathcal{S}_i$ *and* $\ker(P_i) = \mathrm{im}(I_n - P_i) = \mathcal{S}_i^{\perp}$,
2) *The eigenvalues of an orthogonal projector are* 0 *or* 1. *Hence, any orthogonal projector P_i satisfies $I_n \succeq P_i \succeq 0$, where $A \succeq B$ means that $A - B$ is a positive semi-definite matrix.*
3) *The trace of an orthogonal projector equals its rank.*

Let now $Y = \begin{bmatrix} y_1 & \cdots & y_N \end{bmatrix} \in \mathbb{R}^{n \times N}$ be a matrix formed with all the available data samples and define $E(P)$ to be the projection of Y onto $\mathcal{S}_i^{\perp}$, i.e.,

$$E(P) = (I_n - P)\, Y. \tag{2}$$

The method to be presented is based on the fundamental observation that the matrix $E(P)$ contains a significant number of zero column vectors. We shall then say that $E(P)$ is sparse (to some degree) as a sequence of vectors. More precisely, all the columns of the form $(I_n - P_i)y_j$ with $y_j \in \mathcal{S}_i$, are equal to zero. Let $E_2(P)$ be the vector obtained by applying the euclidean norm $\|\cdot\|_2$ to the columns of $E(P)$, i.e.,

$$E_2(P) = \begin{bmatrix} \|(I_n - P)\, y_1\|_2 & \dots & \|(I_n - P)\, y_N\|_2 \end{bmatrix} \in \mathbb{R}^N.$$

We can then search for one matrix P_i by solving the sparse optimization problem

$$\begin{aligned} &\min_P \|E_2(P)\|_0 \\ &\quad \text{s.t. } P = P^\top,\ I_n \succeq P \succeq 0,\ \mathrm{rank}(P) = d. \end{aligned} \tag{3}$$

Here, $\|\cdot\|_0$ refers to the ℓ_0 pseudo-norm which counts the number of nonzero entries in a vector.

Assumption 1. *The data $\{y_j\}_{j=1}^N$ are in general position in each subspace $\mathcal{S}_i$, that is, any group $\{y_{j_1}, \dots, y_{j_d}\} \subset \mathcal{S}_i$ of d data samples is formed of linearly independent vectors.*

Theorem 1. *Let $P^\star$ denote any solution to* (3)*. For any $P \in \mathbb{R}^{n \times n}$, define $\mathscr{I}(P) = \{y_j : (I - P)\, y_j = 0\}$. Under Assumption 1, if $|\mathscr{I}(P_i)| \geq sd$ for all $i = 1, \dots, s$, then*

$$P^\star \in \{P_1, \dots, P_s\}.$$

Proof. Because the matrices P_i are feasible points of the optimization problem (3), it holds that $|\mathscr{I}(P^\star)| \geq |\mathscr{I}(P_i)|$ for $i = 1, \dots, s$. We claim that there is an index $i^* \in \{1, \dots, s\}$ such that $|\mathscr{I}(P^\star) \cap \mathscr{I}(P_{i^*})| \geq d$. Assume, for contradiction, that this is not the case. Then by letting n_i denote the cardinality of the set $\mathscr{I}(P^\star) \cap \mathscr{I}(P_i)$, we have $n_i < d$, for $i = 1, \dots, s$. Since all the data points lie in $\mathscr{I}(P_1) \cup \dots \cup \mathscr{I}(P_s)$, it follows that

$$|\mathscr{I}(P^\star)| \leq n_1 + \dots + n_s < sd \leq |\mathscr{I}(P_i)|,$$

$i = 1, \dots, s$. However this contradicts the definition of $P^\star$, hence proving the existence of an index i^* such that $\mathscr{I}(P^\star) \cap \mathscr{I}(P_{i^*})$ contains at least d different vectors. Note that

$$\mathrm{span}\left[\mathscr{I}(P^\star) \cap \mathscr{I}(P_{i^*})\right] \subset \ker(I_n - P_{i^*}) = \mathcal{S}_{i^*} \tag{4}$$

By invoking Assumption 1 with the fact that $|\mathscr{I}(P^\star) \cap \mathscr{I}(P_{i^*})| \geq d$, the dimension of the span of $\mathscr{I}(P^\star) \cap \mathscr{I}(P_{i^*})$ is equal to d. Since $\dim(\mathcal{S}_{i^*}) = d$, it can be concluded from (4) that

$$\mathrm{span}\left[\mathscr{I}(P^\star) \cap \mathscr{I}(P_{i^*})\right] = \mathcal{S}_{i^*} \subset \ker(I_n - P^\star).$$

It follows that $P^\star$ possesses 1 as an eigenvalue with geometric multiplicity larger or equal to $d = \dim(\mathcal{S}_{i^*})$. Since $P^\star$ is symmetric, it is diagonalizable. Hence, there exists an orthonormal matrix $U \in \mathbb{R}^{n \times n}$ such that $P^\star$ can be factorized in the form $P^\star = U \Sigma U^\top$, with $\Sigma = \mathrm{diag}(\lambda_1, \dots, \lambda_d, \lambda_{d+1}, \dots, \lambda_n)$, $\lambda_1 = \lambda_2 = \dots = \lambda_d = 1$. Recalling that $\mathrm{rank}(P^\star) = d$, we have necessarily, $\lambda_{d+1} = \dots = \lambda_n = 0$ so that $P^\star$ obeys $(P^\star)^2 = P^\star$. It can also be noticed that $\mathrm{rank}(I_n - P^\star) = \mathrm{rank}(U(I_n - \Sigma)U^\top) = n - d$, which, by the rank theorem, implies that $\dim\left(\ker(I_n - P^\star)\right) = d$ and $\mathcal{S}_{i^*} = \ker(I_n - P^\star)$. From this, we deduce that $P^\star$ is in fact the orthogonal projection matrix onto $\mathcal{S}_{i^*}$. Hence $P^\star = P_{i^*}$ concluding the proof. □

Problem (3) is in general very hard to solve exactly because it involves a combinatorial search. The difficulty arises essentially from the presence of the ℓ_0-norm objective and the rank constraint. To get a more tractable optimization problem, a classical idea intensively used in the field of compressed sensing [12], [11], is to approximate the ℓ_0 pseudo-norm with the ℓ_1-norm[1], the main advantage being that the last one is convex and therefore efficiently solvable. This leads to the following formulation

$$\begin{aligned} &\min_P \|W E_\ell(P)\|_1 \\ &\quad \text{s.t. } P = P^\top,\ I_n \succeq P \succeq 0,\ \mathrm{rank}(P) = d \end{aligned} \tag{5}$$

where $W = \mathrm{diag}(w_1, \cdots, w_N)$ is a user-defined weighting diagonal matrix with positive entries. Prior knowledge, when available, can be incorporated in W; by default W is set to the identity matrix I_N. The next step is to relax the rank constraint. Knowing from Lemma 1 that the rank of a projection matrix equals its trace, we can try to enforce the rank constraint by replacing it with a constraint on the trace. As a result, we obtain a completely convex relaxation of (3) in the following form

$$\begin{aligned} &\min_P \sum_{i=1}^N w_i \|(I_n - P)\, y_i\|_2 \\ &\quad \text{s.t. } P = P^\top,\ I_n \succeq P \succeq 0,\ \mathrm{tr}(P) = d. \end{aligned} \tag{6}$$

We first look at well-posedness by showing that if a single subspace was to be estimated, then (6) is a convenient formulation for recovering the orthogonal projection matrix.

Lemma 2. *Assume $s = 1$, i.e., the data are drawn from a single subspace $\mathcal{S}_1$. If* $\mathrm{rank}(Y) = \dim(\mathcal{S}_1) = d$*, then problem* (6) *has a unique solution $P^\star$ which corresponds exactly to the orthogonal projection matrix onto $\mathcal{S}_1$.*

Proof. Let P_1 be the orthogonal projection matrix onto $\mathcal{S}_1$. It is clear that P_1 is a feasible point for problem (6) and

[1]The ℓ_1-norm of a vector $z = [z_1 \ \dots \ z_N]^\top \in \mathbb{R}^N$ is given by $\|z\|_1 = |z_1| + \dots + |z_N|$.

minimizes the associated cost function (whose optimal value is actually equal to zero). Therefore P_1 is obviously one solution to problem (6). We just need to prove uniqueness. We do so by showing that any solution $P^\star$ to (6) is necessarily equal to P_1. In effect, if $P^\star$ solves (6), then $(I_n - P^\star)\,Y = 0$. Since $\mathrm{rank}(Y) = d$, we have $(I_n - P^\star)\,y = 0$ for any $y \in \mathcal{S}_1$, the span of Y. In other words, 1 is an eigenvalue of $P^\star$ with geometric multiplicity at least d. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $P^\star$ enumerated in descending order. We have $\lambda_1 = \cdots = \lambda_d = 1$, so that $\lambda_{d+1} + \cdots + \lambda_n = \mathrm{tr}(P^\star) - d = 0$. With $P^\star \succeq 0$, this leads to $\lambda_{d+1} = \cdots = \lambda_n = 0$. $P^\star$ is the orthogonal projection matrix onto $\mathcal{S}_1$ and therefore $P^\star = P_1$.

Consider now the general case where $s \geq 2$. Next we state a condition under which a subspace can be exactly recovered by solving (6). Let $\mathcal{H} = \big\{\Delta \in \mathbb{R}^{n\times n} : \Delta = \Delta^\top, \mathrm{tr}(\Delta) = 0, \|\Delta\|_2 \leq 1\big\}$.

Theorem 2. *Assume that $w_1 + \cdots + w_N = 1$. Let*

$$\begin{aligned} r(Y) = &\max \|\Delta\|_2 \\ &\text{s.t. } \Delta \in \mathcal{H},\ \|\Delta Y\|_{2,\mathrm{col}} = 1 \end{aligned} \tag{7}$$

where $\|A\|_{2,\mathrm{col}}$ is, for any matrix A, the sum of the 2-norms of the column vectors of A. If there is a P_i satisfying

$$\sum_{j\in\mathscr{I}(P_i)} w_j > 1 - \frac{1}{2\eta r(YW)} \tag{8}$$

with $\eta = \max_j \|y_j\|_2$, then P_i is the unique solution to problem (6)*.*

Proof. P_i uniquely solves problem (6) if, for any nonzero matrix Δ in $\mathbb{R}^{n\times n}$ such that $P_i + \Delta$ is feasible, it holds that

$$\sum_{j=1}^N w_j \|(I_n - P_i)\,y_j\|_2 < \sum_{j=1}^N w_j \|(I_n - P_i - \Delta)\,y_j\|_2 .$$

Feasibility of $P_i + \Delta$ translates into $\mathrm{tr}(\Delta) = 0$, $\Delta^\top = \Delta$ and $I_n - P_i \succeq \Delta \succeq -P_i$. These, together with the fact that P_i is a projection matrix implies that Δ lies $\mathcal{H}$. An equivalent form of the inequality above is

$$\begin{aligned} \sum_{j\notin\mathscr{I}(P_i)} w_j \Big[\, \|(I_n - P_i)\,y_j\|_2 - \|(I_n - P_i - \Delta)\,y_j\|_2 \,\Big] \\ < \sum_{j\in\mathscr{I}(P_i)} w_j \|\Delta y_j\|_2 . \end{aligned}$$

From the triangle inequality property of vector 2-norm, the following identity $\|a\|_2 - \|a+b\|_2 \leq \|b\|_2$ holds for any two vectors a and b of compatible dimensions. By applying this, it can be observed that for P_i to be a solution to problem (6), it is sufficient that

$$\sum_{j\notin\mathscr{I}(P_i)} w_j \|\Delta y_j\|_2 < \sum_{j\in\mathscr{I}(P_i)} w_j \|\Delta y_j\|_2$$

for any $\Delta \in \mathbb{R}^{n\times n}$ obeying the conditions above. Adding the term $\sum_{j\notin\mathscr{I}(P_i)} w_j \|\Delta y_j\|_2$ on each side of the previous inequality yields, after simple algebraic manipulations,

$$\max_{\substack{\|\Delta YW\|_{2,\mathrm{col}}=1 \\ \Delta\in\mathcal{H}}} \Big[\sum_{j\notin\mathscr{I}(P_i)} w_j \|\Delta y_j\|_2 \Big] < 1/2.$$

Note that the maximum here could have been taken over the set

$$\mathcal{H} \cap \Big\{\Delta : \|\Delta YW\|_{2,\mathrm{col}} = 1\Big\} \cap \Big\{\Delta : I_n \succeq P_i + \Delta \succeq 0\Big\}.$$

Thanks to the identity $\max_{\Delta\in A} \|\Delta\|_2 \leq \max_{\Delta\in B} \|\Delta\|_2$ whenever $A \subset B$, we can restrict the maximum to the first two members of the above chain of intersections and get a more conservative condition. Finally, from the inequality $\|\Delta y_j\|_2 \leq \eta \|\Delta\|_2$, it can be seen that the maximum above is bounded from above by $\eta\big(1 - \sum_{j\in\mathscr{I}(P_i)} w_j\big) r(YW)$. Therefore P_i is the unique solution to (6) if

$$\eta\Big(1 - \sum_{j\in\mathscr{I}(P_i)} w_j\Big) r(YW) < 1/2$$

as claimed.

A. Enhancing sparsity

As shown by Theorem 1, the formulation (3) can obtain a solution to the subspace clustering problem under mild conditions. However, for the convex relaxation (6) to yield the desired solution, a relatively high level of sparsity might be needed. That is, a large majority of the data must pertain to the same subspace. This may not be the case in general. A possible solution is to solve a sequence of problems of the type (6) with different weights computed iteratively [14], [15]. The iterative scheme can be defined for a fixed number $r_{\max}$ of iterations as follows. At iteration $r = 0, \ldots, r_{\max}$, compute

$$\begin{aligned} P^{(r)} = &\arg\min_P \sum_{j=1}^N w_j^{(r)} \|(I_n - P)y_j\|_2 , \\ &\text{s.t. } P = P^\top,\ I_n \succeq P \succeq 0,\ \mathrm{tr}(P) = d \end{aligned} \tag{9}$$

with weights defined, for all j by $w_j^{(0)} = 1/N$, and $w_j^{(r)} = v_j^{(r)} / \sum_{j=1}^N v_j^{(r)}$ for $r \geq 1$, where

$$v_j^{(r)} = \frac{1}{\big\|(I_n - P^{(r-1)})y_j\big\|_2^\nu + \zeta},$$

$\zeta > 0$ is a small number preventing division by zero, r is the iteration number and $\nu \geq 1$. Since we are dealing here with a sequence of convex optimization problems, they can be numerically implemented using any convex solver, see e.g., [16]. Many simulations results show that the iterative scheme increases significantly the recoverability of the P_i's in situations where the degree of sparsity is low.

After a first parameter matrix P_1 is identified, we still need to identify the other subspaces. For this purpose, we can remove the data samples pertaining to the already identified subspace and repeat the iterative algorithm over the remaining set of data. The different steps are described in Algorithm II-A.

III. Special case of independent subspaces

In this section we analyze the particular case where the subspaces $\mathcal{S}_1, \ldots, \mathcal{S}_s$ are independent, that is when they satisfy

$$\mathcal{S}_i \cap (\mathcal{S}_1 + \cdots + \mathcal{S}_{i-1} + \mathcal{S}_{i+1} + \cdots + \mathcal{S}_s) = \{0\}$$

for any i. If this is the case, then for all i, $\bar{\mathcal{S}}_i \triangleq \oplus_{j\neq i}\mathcal{S}_j \subset \ker(P_i)$ while $\mathrm{im}(P_i) = \mathcal{S}_i$. Here $\oplus$ refers to direct sum of

Algorithm 1 Subspace clustering algorithm

Inputs: $\{y_j\}_{j=1}^N$

1) *Initialization:* $\mathscr{P} \leftarrow \emptyset$, $J \leftarrow \{1, \ldots, N\}$
2) **Repeat**
   - Compute, up to convergence, the sequence $\left\{P^{(r)}\right\}$ defined in (9) based on the data whose indices are contained in J.
   - Record the identified matrix $\hat{P}$: $\mathscr{P} \leftarrow \mathscr{P} \cup \left\{\hat{P}\right\}$.
   - Remove from J the indices of data y_j satisfying $\left\|(I_n - \hat{P})y_j\right\|_2 \leq \delta$, where δ is a threshold.

   Until $|J| = 0$
3) *Return* $\mathscr{P}$ and set $s = |\mathscr{P}|$.

subspaces. For $s > 2$, the number of samples pertaining to subspace $\bar{\mathcal{S}}_i$ is generally larger than those pertaining to $\mathcal{S}_i$. As a result, solving problem (6) is likely to yield the orthogonal projector onto $\bar{\mathcal{S}}_i$. More precisely, the following corollary of Theorem 2 holds.

Corollary 1. *Assume that the subspaces $\mathcal{S}_1, \ldots, \mathcal{S}_s$ are independent. Define $P^\star$ as*

$$P^\star = \arg\min_P \sum_{j=1}^N \|(I_n - P)y_j\|_2 \quad \text{s.t. } P = P^\top,\ I_n \succeq P \succeq 0,\ \mathrm{tr}(P) \leq d(s-1). \tag{10}$$

If $\min_{i=1,\ldots,s} |\mathscr{I}(P_i)| < \dfrac{1}{2\eta r(Y)}$, then $\{y_j : j \in \mathscr{I}^c(P^\star)\} \subset \mathcal{S}_{i^o}$, where $i^o = \arg\min_i |\mathscr{I}(P_i)|$ and $\mathscr{I}^c(P^\star) = \{1, \ldots, N\} \setminus \mathscr{I}(P^\star)$.

Proof. Denote with $\hat{P}$ the projection matrix onto $\bar{\mathcal{S}}_{i^o} \triangleq \oplus_{j \neq i^o} \mathcal{S}_i$. Since the subspaces are independent, $\mathrm{tr}(\hat{P}) = \mathrm{rank}(\hat{P}) = \dim(\bar{\mathcal{S}}_{i^o}) = d(s-1)$. Hence $\hat{P}$ is feasible for the underlying optimization problem in (10). The condition $\min_{i=1,\ldots,s} |\mathscr{I}(P_i)| < \frac{1}{2\eta r(Y)}$ is equivalent to $\left|\mathscr{I}(\hat{P})\right| > N - \frac{1}{2\eta r(Y)}$, which, by Theorem 2, implies that $\hat{P}$ is the unique solution to (10). It follows that $P^\star = \hat{P}$. Data which are not in $\mathscr{I}(P^\star)$ are necessarily in $\mathcal{S}_{i^o} = \mathrm{span}(\mathscr{I}(P_{i^o}))$. □

The corollary suggests that it is presumably easier to segment the subspaces when they are independent.

IV. Robust formulation

The method derived above can still work in the presence of a moderate amount of noise. However, when the noise level is high, we might need to resort to a more robust implementation. The convex problem (6) can, similarly as in [17], be reformulated in the form

$$\begin{aligned} \min_{P,\xi} \ & \sum_{j=1}^N w_j \xi_j \\ \text{s.t. } & \|(I_n - P)y_j\|_2 \leq \varepsilon + \xi_j,\ j = 1, \ldots, N \\ & P = P^\top,\ I_n \succeq P \succeq 0,\ \mathrm{tr}(P) = d \\ & \xi_j \geq 0,\ j = 1, \ldots, N \end{aligned} \tag{11}$$

where ε is a user-defined parameter to be tuned in function of the noise level. The rationale behind the formulation (11) is that, for an appropriately chosen ε, the distances $\|(I_n - P_i)y_j\|_2$ should be less than ε for all the y_j close to the subspace spanned by P_i.

V. Experiments

This section provides some numerical results to illustrate the previous theory. Two experiments are considered: one with noise-free data and the other with noisy data. Both experiments are conducted in the same way as follows. We generate 100 datasets in $\mathbb{R}^6$. Each dataset contains 400 data points normally sampled from $s = 4$ different subspaces of the same dimension d. To challenge the proposed algorithm, each dataset is designed such that the number of samples originating from all subspaces is the same (this is indeed the most difficult scenario according to the discussions above). The algorithm is then run on each of the 100 datasets. The experiment is repeated for different values of the subspaces' dimension $d = 1, \ldots, 5$. Results are reported in Tables I and II. For more comments on the effects of the user-defined parameters ζ and δ, see [15].

A number of observations can be made from these results: (a) The reweighted scheme (9) endows the proposed framework with a capability to accommodate subspace clustering problems where none of the subspaces contains the absolute majority of the data; (b) The proposed method can segment arrangements of subspaces with non trivial intersections. However, such arrangements seem more difficult to segment (more iterations are needed) than those formed of independent subspaces. Comparison with the SSC algorithm [7] reveals the benefit of the proposed approach.

dimension d	1	2	3	4	5
# iterations	2	5	5	10	10

TABLE I
NUMBER OF ITERATIONS OF THE REWEIGHTED SCHEME (9) (ON NOISE-FREE DATA) NECESSARY FOR ACHIEVING 100% SUCCESSES IN THE CLUSTERING. PARAMETERS ARE $\zeta = 1e-3$, $\delta = 0.1$, $\nu = 5$.

dimension d	1	2	3	4	5
this paper	0.022	0.027	0.030	0.029	0.035
SSC [7]	0.074	0.158	0.350	0.519	1.028

TABLE II
AVERAGE ESTIMATION ERROR FOR 10 ITERATIONS OF THE REWEIGHTED SCHEME (9) ON NOISY DATA. THE NOISE IS GENERATED FROM A GAUSSIAN DISTRIBUTION OF MEAN ZERO AND VARIANCE 0.05.

VI. Conclusion

We have described a technique for estimating multiple subspaces from unlabeled data which lie in the union of the subspaces. The main idea hinges on solving a sequence of sparsity-inducing optimization problems. In contrast to comparable methods using sparse representation, the subspaces to be segmented need not be independent here. Theoretical correctness of the method is analyzed under some sufficient condition related to the proportion of data drawn from each subspace.

ACKNOWLEDGMENT

The author thanks Prof. René Vidal for stimulating discussions.

REFERENCES

[1] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1–15, 2005.

[2] Y. Ma, A. Yang, H. Derksen, and R. Fossum, "Estimation of subspace arrangements with applications in modeling and segmenting mixed data," *SIAM Review*, vol. 50, pp. 413–458, 2008.

[3] K. Kanatani, "Motion segmentation by subspace separation and model selection," in *IEEE International Conference on Computer Vision*, 2001.

[4] M. A. Fischler and R. C. Bolles, "RANSAC random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 26, pp. 381–395, 1981.

[5] J. P. Costeira and T. Kanade, "A multibody factorization method for independently moving objects," *International Journal of Computer Vision*, vol. 29, pp. 159–179, 1998.

[6] A. Gruber and Y. Weiss, "Multibody factorization with uncertainty and missing data using the EM algorithm.," in *Computer Vision and Pattern Recognition*, 2004.

[7] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 2765–2781, 2013.

[8] G. Liu, H. Xu, and S. Yan, "Exact subspace segmentation and outlier detection by low-rank representation," *Journal of Machine Learning Research*, vol. http://arxiv.org/abs/1109.1646, 2011.

[9] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 171–184, 2013.

[10] R. Vidal, "Subspace clustering," *IEEE Signal Processing Magazine*, vol. 28, pp. 52–68, 2011.

[11] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 1289-1306, p. 52, 2006.

[12] E. J. Candès and T. Tao, "Reflections on compressed sensing," *IEEE Information Theory Society Newsletter*, vol. 58, pp. 20–23, 2008.

[13] A. Adler, M. Elad, and Y. Hel-Or, "Probabilistic subspace clustering via sparse representations," *IEEE Signal Processing Letters*, vol. 20, pp. 63–66, 2013.

[14] E. J. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008.

[15] L. Bako, "Identification of switched linear systems via sparse optimization," *Automatica*, vol. 47, pp. 668–677, 2011.

[16] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.2 (june 2009, build 711).".

[17] L. Bako, V. L. Le, F. Lauer, and G. Bloch, "Identification of MIMO switched state-space models," in *American Control Conference, Washington DC, USA*, 2013.