An inexact splitting method for the subspace segmentation from incomplete and noisy observations

Liang, Renli; Bai, Yanqin; Lin, Hai Xiang

DOI
10.1007/s10898-018-0684-4

Publication date
2018

Document Version
Final published version

Published in
Journal of Global Optimization

Journal of Global Optimization (2019) 73:411–429
https://doi.org/10.1007/s10898-018-0684-4

An inexact splitting method for the subspace segmentation from incomplete and noisy observations

Renli Liang[1] · Yanqin Bai[1] · Hai Xiang Lin[2]

Received: 25 October 2017 / Accepted: 26 June 2018 / Published online: 4 July 2018

Abstract
Subspace segmentation is a fundamental issue in computer vision and machine learning, which segments a collection of high-dimensional data points into their respective low-dimensional subspaces. In this paper, we first propose a model for segmenting the data points from incomplete and noisy observations. Then, we develop an inexact splitting method for solving the resulted model. Moreover, we prove the global convergence of the proposed method. Finally, the inexact splitting method is implemented on the clustering problems in synthetic and benchmark data, respectively. Numerical results demonstrate that the proposed method is computationally efficient, robust as well as more accurate compared with the state-of-the-art algorithms.

Keywords Subspace segmentation · Low rank representation · Inexact augmented Lagrange multiplier method

Mathematics Subject Classification 65K05 · 90C25 · 90C30 · 94A08

1 Introduction

Subspace segmentation is a fundamental issue in computer vision and machine learning, which has numerous applications, including motion segmentation [15], face clustering [21], image representation [14], and system identification [32]. In fact, the data points often reside in or lie close to a union of low dimensional subspaces [18], such as facial images, motion, texture, and biological networks [3]. Consequently, it is important to solve the well-known subspace segmentation (or clustering) problem, whose goal is to segment the data into their

This research was supported by a grant from the National Natural Science Foundation of China (No. 11771275)..

✉ Yanqin Bai
yqbai@t.shu.edu.cn

[1] Department of Mathematics, Shanghai University, Shanghai, China

[2] Department of Applied Mathematics, Delft University of Technology, Delft, The Netherlands

respective clusters, with each cluster being a subspace. The subspace segmentation problem is formally defined as follows.

Definition 1 (*Subspace Segmentation* [18]) Given a set of data vectors $X = [X_1, \ldots, X_s] = [x_1, \ldots, x_n] \in R^{m\times n}$ drawn from a union of s subspaces $\{S_i\}_{i=1}^s$. Let X_i be a collection of $\bar{n}_i$ samples drawn from the subspace S_i, $n = \sum_{i=1}^s \bar{n}_i$. The task of subspace segmentation is to segment the data according to the underlying subspaces they are drawn from.

A variety of methods have been proposed in the past decades. In general, existing methods can be roughly divided into four categories: algebraic methods [15,22,28], iterative methods [2,13,27,33], statistical methods [10,23,26] and spectral clustering based methods [6,7,9,30,34] according to the review in [8]. Among many approaches for the subspace segmentation, the spectral clustering based methods have shown excellent performance. In particular, spectral clustering based methods consist of two main steps. Firstly, an affinity matrix (i.e., an undirected graph) is learned from the given data. Secondly, the segmentation results are obtained by using the affinity matrix to perform spectral clustering algorithms, such as the Normalized Cuts (NCut) [24]. Building a "good" affinity matrix is the key to achieve a good clustering result. The main difference among various spectral clustering based methods is the first step which learns an affinity matrix.

Liu et al. [18] proposed a low-rank representation (LRR) approach for clustering data drawn from a union of multiple linear subspaces. And an inexact augmented Lagrange multiplier (IALM) algorithm [17] was introduced to solve the following model:

$$\text{(P1)} \qquad \min_{Z,E} \|Z\|_* + \lambda\|E\|_\ell \quad \text{s.t. } X = AZ + E,$$

where $X \in R^{m\times n}$ is the given data matrix. $A \in R^{m\times d}$ is a dictionary that linearly spans the data space. $Z \in R^{d\times n}$ is the low-rank representation of data X with respect to the dictionary A. $E \in R^{m\times n}$ is the observation noise. $\lambda > 0$ is a positive weighting parameter and $\|\cdot\|_\ell$ indicates a certain regularization strategy, such as the l_1 norm. The nuclear norm $\|Z\|_*$ is defined as the sum of all the singular values of Z. After solving (P1), the optimal solution Z^* was used to define an affinity matrix Y for spectral clustering algorithms [24] to provide the final segmentation results. However, the convergence property of LRR was ambiguous without more assumptions. Hence, Xiao et al. [29] proposed a primal splitting and linearizing augmented Lagrangian (PSLAL) method for solving (P1) and established the global convergence. Recently, He et al. [12] proposed a splitting method for solving a general separable convex minimization problem. And they also established the global convergence and a worst-case convergence rate for the splitting method.

Inspired by the above works, we devote to consider more practical circumstances for the subspace segmentation problem. It has been pointed out in [25], the observed data X may be corrupted by both impulsive noise E (sparse but large) and Gaussian noise F (small but dense), e.g., $X = AZ + E + F$. We assume the Gaussian noise of the observed entries is small in the sense that $\|F\|_F \le \delta$, where $\|\cdot\|_F$ is the Frobenius norm and $\delta > 0$ is the Gaussian noise level. Besides, we also consider the case that only a fraction of entries of X can be observed. In particular, let $\Omega \subset \{1, \ldots, m\} \times \{1, \ldots, n\}$ be the index set of entries X that are observable. The same symbol $P_\Omega : R^{m\times n} \to R^{m\times n}$ in [5,11] is used to summarize the incomplete observation information, where $P_\Omega(X)_{ij} = X_{ij}$ if $(i, j) \in \Omega$ and $P_\Omega(X)_{ij} = 0$ otherwise.

Therefore, in this paper, we first propose the following new model for segmenting the data points from incomplete and noisy observations:

$$\text{(P2)} \qquad \min_{Z,E} \|Z\|_* + \lambda\|E\|_1 \quad \text{s.t. } \|P_\Omega(X - AZ - E)\|_F \le \delta,$$

where $\|E\|_1 := \sum_{i=1}^{m}\sum_{j=1}^{n}|E_{ij}|$ and $\lambda > 0$ is a positive weighting parameter. Obviously, (P2) is convex and the objective function of (P2) is nonsmooth. We then develop an inexact splitting method for solving a problem equivalent to (P2). Furthermore, we also prove that the inexact splitting method is globally convergent. Finally, numerical results on synthetic and real data sets demonstrate the effectiveness of our proposed method.

The paper is organized as follows. In Sect. 2, we first provide some preliminaries that are used in the latter analysis. In Sect. 3, we describe the inexact splitting method for solving the convex reformulation of (P2). In Sect. 4, the global convergence of the proposed method is established. Section 5 presents experiments that evaluate our method using the synthetic data and the real data. Lastly, we end with some concluding remarks in Sect. 6.

2 Preliminaries

In this section, we first summarize notations used in this paper. The l_1 norm, Frobenius norm, and $l_{2,1}$ norm of the matrix $X \in R^{m\times n}$ are respectively defined as

$$\|X\|_1 = \sum_{i=1}^{m}\sum_{j=1}^{n}|X_{ij}|, \quad \|X\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}X_{ij}^2}, \quad \|X\|_{2,1} = \sum_{j=1}^{n}\sqrt{\sum_{i=1}^{m}X_{ij}^2},$$

where X_{ij} is the (i, j)th component of X. For any two matrices $X, Y \in R^{m\times n}$, we define $\langle X, Y\rangle = trace(X^{\mathrm{T}}Y)$ (the standard trace inner product). Let $sign(X) : R^{m\times n} \to R^{m\times n}$ denote the sign function of X, e.g.,

$$[sign(X)]_{ij} = \begin{cases} 1, & \text{if } X_{ij} > 0, \\ 0, & \text{if } X_{ij} = 0, \\ -1, & \text{if } X_{ij} < 0. \end{cases}$$

We denote $abs(X) \in R^{m\times n}$ as the absolute value function, e.g., $[abs(X)]_{ij} = |X_{ij}|$. And $diag(x)$ denotes a square diagonal matrix with the elements of the vector x on the main diagonal. In the following, we briefly review some well-known results that are used in the latter analysis.

Lemma 1 [31] *For $\mu > 0$ and $t \in R$, the minimizer of*

$$\min_{s\in R}\ \mu|s| + \frac{1}{2}(s-t)^2$$

is given by

$$\max\{|t| - \mu, 0\} \cdot sign(t).$$

Lemma 2 [4] *Given $T \in R^{m\times n}$ of rank r, let $T = U_T\Sigma_T V_T^{\mathrm{T}}$, $\Sigma_T = diag(\sigma_1, \ldots, \sigma_r)$ be the singular value decomposition of T, where $U_T \in R^{m\times r}$, $\Sigma_T \in R^{r\times r}$ and $V_T \in R^{n\times r}$. For each $\mu > 0$, the solution of the following problem*

$$\min_{X\in R^{m\times n}}\ \mu\|X\|_* + \frac{1}{2}\|X - T\|_F^2$$

is given by $D_\mu(T) \in R^{m\times n}$, which is defined by

$$D_\mu(T) := U_T\Sigma_T^\mu V_T^{\mathrm{T}},$$

where $\Sigma_T^\mu = diag(\{\sigma_i - \mu\}^+) \in R^{r\times r}$ and $\{\cdot\}^+ = \max(0, \cdot)$.

3 The inexact splitting method

In this section, we first reformulate the model (P2) to a convex separable model. We then derive an inexact splitting method for solving this convex separable model.

3.1 Reformulation and optimality

In the following theorem, we construct the alternative formulation and establish its equivalence to (P2).

Theorem 1 *Let* (Z^*, E^*, F^*) *be an optimal solution to*

$$\text{(P3)} \quad \begin{aligned} \min_{Z,E,F} \quad & \|Z\|_* + \lambda \|P_\Omega(E)\|_1 \\ \text{s.t.} \quad & P_\Omega(X) = AZ + E + F, \\ & F \in \mathscr{F} := \{F, \|F\|_F \le \delta\}. \end{aligned}$$

Then $(Z^*, P_\Omega(E^*))$ *is an optimal solution to (P2).*

Proof Suppose that (Z^*, E^*, F^*) is an optimal solution to (P3). Then, we have $\|P_\Omega(X) - AZ^* - E^*\|_F \le \delta$. Furthermore, $\|P_\Omega(X - AZ^* - P_\Omega(E^*))\|_F = \|P_\Omega(X - AZ^* - E^*)\|_F \le \|P_\Omega(X) - AZ^* - E^*\|_F \le \delta$. Hence, the feasibility of $(Z^*, P_\Omega(E^*))$ is verified. Now suppose that $(Z^*, P_\Omega(E^*))$ is not optimal to (P2). Then there exists an optimal solution $(\bar{Z}, \bar{E})$ to (P2), such that

$$\|\bar{Z}\|_* + \lambda \|\bar{E}\|_1 < \|Z^*\|_* + \lambda \|P_\Omega(E^*)\|_1, \tag{1}$$

$$\|P_\Omega(X - A\bar{Z} - \bar{E})\|_F \le \delta. \tag{2}$$

We claim that $\bar{E}_{ij} = 0, \forall (i, j) \notin \Omega$. Otherwise, $(\bar{Z}, P_\Omega(\bar{E}))$ is feasible to (P2) and has a strictly smaller objective function value than $(\bar{Z}, \bar{E})$, which contradicts the optimality of $(\bar{Z}, \bar{E})$. Hence, we have

$$\|\bar{Z}\|_* + \lambda \|P_\Omega(\bar{E})\|_1 = \|\bar{Z}\|_* + \lambda \|\bar{E}\|_1. \tag{3}$$

By defining a new matrix $\widetilde{E}$ as

$$\widetilde{E}_{ij} = \begin{cases} \bar{E}_{ij}, & \text{if } (i, j) \in \Omega, \\ -(A\bar{Z})_{ij}, & \text{otherwise,} \end{cases}$$

we have $\|P_\Omega(\widetilde{E})\|_1 = \|P_\Omega(\bar{E})\|_1$ and

$$P_\Omega(X) - A\bar{Z} - \widetilde{E} = P_\Omega(X - A\bar{Z} - \bar{E}). \tag{4}$$

Therefore, from (2) and (4), we have that $(\bar{Z}, \widetilde{E}, \widetilde{F})$ is feasible to (P3), where $\widetilde{F} = P_\Omega(X) - A\bar{Z} - \widetilde{E}$. Combination this with (1) and (3), we obtain

$$\|\bar{Z}\|_* + \lambda \|P_\Omega(\widetilde{E})\|_1 = \|\bar{Z}\|_* + \lambda \|P_\Omega(\bar{E})\|_1 < \|Z^*\|_* + \lambda \|P_\Omega(E^*)\|_1,$$

which contradicts the optimality of (Z^*, E^*, F^*). Therefore, $(Z^*, P_\Omega(E^*))$ is an optimal solution to (P2). □

Let $P_\Omega(X) = M$. The Lagrangian function of (P3) is defined as

$$L(Z, E, F, \Lambda) = \|Z\|_* + \lambda \|P_\Omega(E)\|_1 + \langle \Lambda, M - AZ - E - F \rangle,$$

where $\Lambda \in R^{m\times n}$ is the Lagrange multiplier associated with the equality constraint in (P3). Obviously, $(Z^*, E^*, F^*) \in R^{d\times n} \times R^{m\times n} \times \mathscr{F}$ is a solution of (P3) if and only if there exists $\Lambda^* \in R^{m\times n}$ such that:

$$\begin{aligned} &\langle F' - F^*, -\Lambda^*\rangle \geq 0, \quad \forall F' \in \mathscr{F}, \\ &0 \in \lambda\partial(\|P_\Omega(E^*)\|_1) - \Lambda^*, \\ &0 \in \partial(\|Z^*\|_*) - A^{\mathrm{T}}\Lambda^*, \\ &M^* = AZ^* + E^* + F^*, \end{aligned} \tag{5}$$

where $\partial(\cdot)$ denotes the subgradient operator of a convex function.

3.2 The inexact splitting method for solving (P3)

Next, we propose an inexact splitting method for solving (P3), which is an extension of the method in [12,25]. Recall that the method in [12] was proposed for solving a general separable convex problem with linear constraints. And all the coefficient matrices of linear constraints were assumed to be full column rank. However, for the subspace segmentation problem, the matrix A is not full column rank. Furthermore, because there exists the linear operator A, we are no longer able to obtain the exact solution of one subproblem. For these reasons, we need to further modify the splitting method in [12] for solving (P3).

The augmented Lagrangian function of (P3) is

$$L_\rho(Z, E, F, \Lambda) = \|Z\|_* + \lambda\|P_\Omega(E)\|_1 + \langle\Lambda, M - AZ - E - F\rangle + \frac{\rho}{2}\|M-AZ - E - F\|_F^2,$$

where $\Lambda \in R^{m\times n}$ is the Lagrange multiplier and $\rho > 0$ is the penalty parameter. Following the same iteration scheme in [12], the new iterate $(Z^{k+1}, E^{k+1}, F^{k+1}, \Lambda^{k+1})$ is generated via the following scheme:

$$F^{k+1} = \arg\min_{\|F\|_F\leq\delta} L_\rho(Z^k, E^k, F, \Lambda^k), \tag{6a}$$

$$\widetilde{\Lambda}^k = \Lambda^k - \rho(AZ^k + E^k + F^{k+1} - M), \tag{6b}$$

$$E^{k+1} = \arg\min_{E\in R^{m\times n}} \lambda\|P_\Omega(E)\|_1 + \frac{\rho\eta}{2}\left\|E - \left(E^k + \frac{\widetilde{\Lambda}^k}{\rho\eta}\right)\right\|_F^2, \tag{6c}$$

$$Z^{k+1} = \arg\min_{Z\in R^{d\times n}} \|Z\|_* + \frac{\rho\eta}{2}\left\|AZ - \left(AZ^k + \frac{\widetilde{\Lambda}^k}{\rho\eta}\right)\right\|_F^2, \tag{6d}$$

$$\Lambda^{k+1} = \widetilde{\Lambda}^k - \rho(E^{k+1} - E^k) - \rho(AZ^{k+1} - AZ^k). \tag{6e}$$

Remark 1 The convergence is valid no matter which alternating order among the variables $(F^{k+1}, E^{k+1}, Z^{k+1})$ is used. We decide to perform the alternating tasks in the order of $F^{k+1} \to E^{k+1} \to Z^{k+1}$.

Then, we deduce the closed-form solutions of the problem (6a) and the problem (6c) in Theorems 2 and 3, respectively.

Theorem 2 *The optimal solution F^{k+1} of the problem (6a) is given by*

$$F^{k+1} = \frac{\min\{\|N^k\|_F, \delta\}}{\|N^k\|_F} \cdot N^k, \tag{7}$$

where $N^k = M + \Lambda^k/\rho - AZ^k - E^k$.

Proof The subproblem (6a) with respect to F is equivalent to

$$\begin{aligned} F^{k+1} &= \arg\min_{\|F\|_F \le \delta} L_\rho(Z^k, E^k, F, \Lambda^k) \\ &= \arg\min_{\|F\|_F \le \delta} \left\| AZ^k + E^k + F - M - \frac{\Lambda^k}{\rho} \right\|_F^2 \\ &= \arg\min_{\|F\|_F \le \delta} \|F - N^k\|_F^2. \end{aligned} \tag{8}$$

Obviously, N^k is the optimal solution when $\|N^k\|_F \le \delta$. Next, we only consider the situation that $\|N^k\|_F > \delta$. The problem in (8) is equivalent to the following problem:

$$\text{(P4)} \qquad \min \|F - N^k\|_F^2, \quad \text{s.t. } \|F\|_F^2 \le \delta^2.$$

The Lagrangian function of (P4) is defined as

$$L(F, \bar{\lambda}) = \|F - N^k\|_F^2 + \bar{\lambda}\left(\|F\|_F^2 - \delta^2\right),$$

where $\bar{\lambda} \in R$ is the Lagrange multiplier. It is easy to derive the optimality condition of (P4). More specifically, $F^{k+1} \in R^{m\times n}$ is a solution of (P4) if and only if there exists $\lambda^{k+1} \in R$ that satisfies the following conditions:

$$F^{k+1} = \frac{1}{1+\lambda^{k+1}} N^k, \tag{9a}$$

$$\lambda^{k+1}\left(\|F^{k+1}\|_F^2 - \delta^2\right) = 0, \tag{9b}$$

$$\|F^{k+1}\|_F^2 \le \delta^2, \quad \lambda^{k+1} \ge 0. \tag{9c}$$

Clearly, (9a) and (9c) imply that $\lambda^{k+1} > 0$ for $\|N^k\|_F > \delta$. Furthermore, by substituting (9a) into (9b), it yields $\lambda^{k+1} = \frac{\|N^k\|_F}{\delta} - 1$ and $F^{k+1} = \frac{\delta}{\|N^k\|_F} N^k$. This completes the proof. □

Theorem 3 *The optimal solution E^{k+1} of the subproblem (6c) can be written in closed-form as follows:*

$$E^{k+1} = \max\left\{abs(P_\Omega(T)) - \frac{\lambda}{\rho\eta}\mathbf{1}, 0\right\} \odot sign(P_\Omega(T)) + P_{\Omega^C}(T), \tag{10}$$

where $T = E^k + \frac{\widetilde{\Lambda}^k}{\rho\eta}$, $\mathbf{1} \in R^{m\times n}$ is the matrix with all components equal to one. And $\odot$ denotes the componentwise multiplication operator. Ω^C is the complementary set of Ω in index set $\Gamma = \{1, \ldots, m\} \times \{1, \ldots, n\}$.

Proof Using the definitions of the l_1 norm and the Frobenius norm, we have

$$\begin{aligned} E^{k+1} &= \arg\min_{E\in R^{m\times n}} \lambda\|P_\Omega(E)\|_1 + \frac{\rho\eta}{2}\|E - T\|_F^2 \\ &= \arg\min_{E\in R^{m\times n}} \sum_{i=1}^{m}\sum_{j=1}^{n} \lambda|P_\Omega(E)_{ij}| + \frac{\rho\eta}{2}(E_{ij} - T_{ij})^2. \end{aligned}$$

Note that $P_\Omega(E)_{ij} = 0$ for $(i, j) \notin \Omega$. We therefore have $E_{ij}^{k+1} = T_{ij}$ for $(i, j) \notin \Omega$. Moreover, by Lemma 1, we obtain $E_{ij}^{k+1} = \max\{|T_{ij}| - \lambda/\rho\eta, 0\} \cdot sign(T_{ij})$ for $(i, j) \in \Omega$. Thus, the optimal solution of the subproblem (6c) is given by (10), and the proof is complete. □

Because there exists a linear operator A, Lemma 2 does not yield the closed-form solution for the subproblem (6d). And it is expensive to obtain the exact solution. Furthermore, our linear operator A does not satisfy the assumption in [12] that the linear operator is full column rank. This causes the main difficulty to apply the splitting method in [12] directly. However, it is unnecessary to solve this subproblem exactly to achieve the high precision solution in order to guarantee the convergence. We approximate the subproblem by linearizing the quadratic term of its objective function. Subsequently, the resulting approximate problem is simple enough to have a closed-form solution. Let

$$G_k = -\frac{1}{\rho\eta} A^{\mathrm{T}} \widetilde{\Lambda}^k$$

be the gradient of

$$\frac{1}{2}\left\| AZ - \left(AZ^k + \frac{\widetilde{\Lambda}^k}{\rho\eta} \right) \right\|_F^2$$

at current Z^k. And we have

$$\frac{1}{2}\left\| AZ - \left(AZ^k + \frac{\widetilde{\Lambda}^k}{\rho\eta} \right) \right\|_F^2 \approx \frac{1}{2}\left\| \frac{\widetilde{\Lambda}^k}{\rho\eta} \right\|_F^2 + < G_k, Z - Z_k > + \frac{1}{2\tau}\|Z - Z_k\|_F^2, \quad (11)$$

where $\tau > 0$ is a positive scalar, and the last term is the so-called proximal points term. Hence, instead of solving (6d), the next iteration is generated by

$$\begin{aligned} Z^{k+1} &= \arg\min_{Z\in R^{d\times n}} \|Z\|_* + \frac{\rho\eta}{2}\left\| AZ - \left(AZ^k + \frac{\widetilde{\Lambda}^k}{\rho\eta} \right) \right\|_F^2 \\ &\approx \arg\min_{Z\in R^{d\times n}} \|Z\|_* + \rho\eta < G_k, Z - Z_k > + \frac{\rho\eta}{2\tau}\|Z - Z_k\|_F^2 \\ &= \arg\min_{Z\in R^{d\times n}} \|Z\|_* + \frac{\rho\eta}{2\tau}\|Z - Z_k + \tau G_k\|_F^2 \\ &= D_{\tau/\rho\eta}(Z_k - \tau G_k). \end{aligned} \quad (12)$$

Now we are ready to describe our algorithm, named the Inexact Splitting Method or ISM, as in Algorithm 1.

Algorithm 1: ISM for solving the problem (P3)

Input Choose tolerance parameter $\epsilon \geq 0$, $\eta = 2.01$, $\tau = 0.99/\lambda_{\max}(A^{\mathrm{T}}A)$, multiplier vector $\Lambda^0 = 0$, penalty parameter $\rho^0 > 0$. Initial $Z^0 = 0$, $F^0 = 0$, $E^0 = 0$, $\mu = 1.1$, $\rho_{max} = 10^6$. Set the iteration counter $k = 0$.

Output An approximate optimal solution $(Z^{k+1}, E^{k+1}, F^{k+1})$ of problem (P3).

while $\|AZ^k + E^k + F^k - M\|_\infty > \epsilon$ or $\|Z^k - Z^{k-1}\|_\infty + \|E^k - E^{k-1}\|_\infty > \epsilon$ **do**

Step 1 update F^{k+1} via (7);
Step 2 update $\widetilde{\Lambda}^k$ via $\widetilde{\Lambda}^k = \Lambda^k - \rho(AZ^k + E^k + F^{k+1} - M)$;
Step 3 update E^{k+1} via (10);
Step 4 update Z^{k+1} via $Z^{k+1} = D_{\tau/\rho\eta}(Z_k - \tau G_k)$, where $G_k = -\frac{1}{\rho\eta} A^{\mathrm{T}} \widetilde{\Lambda}^k$;
Step 5 update the multiplier via $\Lambda^{k+1} = \widetilde{\Lambda}^k - \rho(E^{k+1} - E^k) - \rho(AZ^{k+1} - AZ^k)$;
Step 6 update the parameter ρ^{k+1} via $\rho^{k+1} = \min(\rho_{max}, \mu\rho^k)$, and $k = k + 1$.

return F^{k+1}, E^{k+1}, and Z^{k+1};

4 Convergence analysis

This section is devoted to prove the global convergence of Algorithm 1. We first define some notations which will simplify our further analysis. For the iterate $(F^{k+1}, E^{k+1}, Z^{k+1}, \Lambda^{k+1})$ generated by the ISM from the given iterate $(F^k, E^k, Z^k, \Lambda^k)$, we temporarily denote $\widetilde{F}^k = F^{k+1}$, $\widetilde{E}^k = E^{k+1}$, and $\widetilde{Z}^k = Z^{k+1}$. For convenience, we use the notations

$$\omega = \begin{pmatrix} F \\ E \\ Z \\ \Lambda \end{pmatrix}, \quad \omega^* = \begin{pmatrix} F^* \\ E^* \\ Z^* \\ \Lambda^* \end{pmatrix}, \quad \nu = \begin{pmatrix} E \\ Z \\ \Lambda \end{pmatrix}, \quad \textit{and} \quad \nu^* = \begin{pmatrix} E^* \\ Z^* \\ \Lambda^* \end{pmatrix}.$$

For any positive integer k, we also use the notations

$$\omega^k = \begin{pmatrix} F^k \\ E^k \\ Z^k \\ \Lambda^k \end{pmatrix}, \quad \widetilde{\omega}^k = \begin{pmatrix} \widetilde{F}^k \\ \widetilde{E}^k \\ \widetilde{Z}^k \\ \widetilde{\Lambda}^k \end{pmatrix}, \quad \nu^k = \begin{pmatrix} E^k \\ Z^k \\ \Lambda^k \end{pmatrix}, \quad \textit{and} \quad \widetilde{\nu}^k = \begin{pmatrix} \widetilde{E}^k \\ \widetilde{Z}^k \\ \widetilde{\Lambda}^k \end{pmatrix}.$$

Let I_m denote the identity matrix in $R^{m\times m}$. And two more matrices are defined as

$$G = \begin{bmatrix} \eta\rho I_m & 0 & 0 \\ 0 & \frac{\eta\rho}{\tau} I_d & 0 \\ 0 & 0 & \frac{1}{\rho} I_m \end{bmatrix}, \quad \textit{and} \quad d(\nu^k - \widetilde{\nu}^k) = \begin{bmatrix} E^k - \widetilde{E}^k \\ Z^k - \widetilde{Z}^k \\ \Lambda^k - \widetilde{\Lambda}^k - \rho(E^k - \widetilde{E}^k) - \rho A(Z^k - \widetilde{Z}^k) \end{bmatrix}. \tag{13}$$

With the notation of $d(\nu^k - \widetilde{\nu}^k)$, it is easy to see that the iterative scheme of ISM is equivalent to the form:

$$\nu^{k+1} = \nu^k - d(\nu^k - \widetilde{\nu}^k). \tag{14}$$

Moreover, we assume that the solution set of (5), denoted by $\mathscr{W}^*$, is nonempty. We thus have $\mathscr{V}^* = \{\nu^*, \omega^* \in \mathscr{W}^*\}$ is also nonempty. Before we are going to prove the convergence of the proposed method, we prove two useful lemmas.

Lemma 3 *Let ν^k, $\widetilde{\nu}^k$, $d(\nu^k - \widetilde{\nu}^k)$, and G be defined as before. Let $\nu^* \in \mathscr{V}^*$. Then, we have*

$$\langle \nu^k - \nu^*, G \cdot d(\nu^k - \widetilde{\nu}^k)\rangle \geq \langle \nu^k - \widetilde{\nu}^k, G \cdot d(\nu^k - \widetilde{\nu}^k)\rangle. \tag{15}$$

Proof The optimality condition of (6a) implies that

$$\langle F^{'} - F^{k+1}, F^{k+1} - N^k\rangle \geq 0, \quad \forall F^{'} \in \mathscr{F}. \tag{16}$$

Using (6b), (16) is equivalent to

$$\langle F^{'} - F^{k+1}, -\widetilde{\Lambda}^k\rangle \geq 0, \quad \forall F^{'} \in \mathscr{F}. \tag{17}$$

Moreover, based on the optimality conditions of (6c) and (12) , we have

$$\begin{aligned} &\langle E^{'} - E^{k+1}, \lambda G_1^{k+1} - \widetilde{\Lambda}^k + \eta\rho(\widetilde{E}^k - E^k)\rangle \geq 0, \quad \forall E^{'} \in R^{m\times n}, \\ &\langle Z^{'} - Z^{k+1}, G_2^{k+1} - A^{\mathrm{T}}\widetilde{\Lambda}^k + \frac{\eta\rho}{\tau}(\widetilde{Z}^k - Z^k)\rangle \geq 0, \quad \forall Z^{'} \in R^{d\times n}, \end{aligned} \tag{18}$$

where $G_1^{k+1} \in \partial\|P_\Omega(E^{k+1})\|_1$, and $G_2^{k+1} \in \partial\|Z^{k+1}\|_*$.

On the other hand, based on the optimality condition of (P3), we have

$$\langle F^{k+1} - F^*, -\Lambda^*\rangle \geq 0,$$

$$\begin{aligned}
&\langle E^{k+1} - E^*, \lambda S_1 - \Lambda^* \rangle \geq 0, \quad S_1 \in \partial \|P_\Omega(E^*)\|_1, \\
&\langle Z^{k+1} - Z^*, S_2 - A^{\mathrm{T}}\Lambda^* \rangle \geq 0, \quad S_2 \in \partial \|Z^*\|_*, \\
&AZ^* + E^* + F^* - M = 0.
\end{aligned} \tag{19}$$

Let $F^{'} = F^*$ in (17), $E^{'} = E^*$ and $Z^{'} = Z^*$ in (18). By adding the resulted inequalities (17) and (18) to (19), we obtain

$$\begin{aligned}
&\langle F^* - F^{k+1}, -(\widetilde{\Lambda}^k - \Lambda^*) \rangle + \left\langle E^* - E^{k+1}, \lambda \left(G_1^{k+1} - S_1 \right) - (\widetilde{\Lambda}^k - \Lambda^*) + \eta\rho(\widetilde{E}^k - E^k) \right\rangle \\
&+ \left\langle Z^* - Z^{k+1}, \left(G_2^{k+1} - S_2 \right) - A^{\mathrm{T}}(\widetilde{\Lambda}^k - \Lambda^*) + \frac{\eta\rho}{\tau}(\widetilde{Z}^k - Z^k) \right\rangle \geq 0.
\end{aligned} \tag{20}$$

Note that the operator of the subgradient of a convex function is monotone. Hence, we have

$$\left\langle E^* - E^{k+1}, G_1^{k+1} - S_1 \right\rangle \leq 0, \quad \left\langle Z^* - Z^{k+1}, G_2^{k+1} - S_2 \right\rangle \leq 0. \tag{21}$$

In addition, recall that $F^* + E^* + AZ^* = M$, we have the following identity:

$$\begin{aligned}
&\langle F^* - F^{k+1}, -(\widetilde{\Lambda}^k - \Lambda^*) \rangle + \langle E^* - E^{k+1}, -(\widetilde{\Lambda}^k - \Lambda^*) \rangle + \langle A(Z^* - Z^{k+1}), -(\widetilde{\Lambda}^k - \Lambda^*) \rangle \\
&+ \langle F^{k+1} + E^{k+1} + AZ^{k+1} - M, -(\widetilde{\Lambda}^k - \Lambda^*) \rangle = 0.
\end{aligned} \tag{22}$$

Furthermore, it follows from (6b) and (6e) that

$$F^{k+1} + \widetilde{E}^k + A\widetilde{Z}^k - M = \frac{1}{\rho}(\Lambda^k - \widetilde{\Lambda}^k) - (E^k - \widetilde{E}^k) - A(Z^k - \widetilde{Z}^k). \tag{23}$$

According to (20), (21), (22), and (23), we thus obtain

$$\begin{aligned}
0 \leq\ & \langle E^* - E^{k+1}, \eta\rho(\widetilde{E}^k - E^k) \rangle + \left\langle Z^* - Z^{k+1}, \frac{\eta\rho}{\tau}(\widetilde{Z}^k - Z^k) \right\rangle \\
&+ \langle F^{k+1} + E^{k+1} + AZ^{k+1} - M, \widetilde{\Lambda}^k - \Lambda^* \rangle \\
=\ & \eta\rho \langle E^* - E^{k+1}, \widetilde{E}^k - E^k \rangle + \frac{\eta\rho}{\tau} \langle Z^* - Z^{k+1}, \widetilde{Z}^k - Z^k \rangle \\
&+ \left\langle \frac{1}{\rho}(\Lambda^k - \widetilde{\Lambda}^k) - (E^k - \widetilde{E}^k) - A(Z^k - \widetilde{Z}^k), \widetilde{\Lambda}^k - \Lambda^* \right\rangle.
\end{aligned}$$

Following the definition of G, we then obtain

$$\langle \widetilde{\nu}^k - \nu^*, G \cdot d(\nu^k - \widetilde{\nu}^k) \rangle \geq 0.$$

Therefore, by the fact of $\nu^k - \nu^* = \nu^k - \widetilde{\nu}^k + \widetilde{\nu}^k - \nu^*$, we have

$$\langle \nu^k - \nu^*, G \cdot d(\nu^k - \widetilde{\nu}^k) \rangle \geq \langle \nu^k - \widetilde{\nu}^k, G \cdot d(\nu^k - \widetilde{\nu}^k) \rangle,$$

which is the assertion of this lemma. □

Lemma 4 *Let $\nu^* \in \mathscr{V}^*$ and let the sequence $\{\nu^k\}$ be generated by the proposed ISM. Then, the sequence satisfies*

$$\begin{aligned}
&\|\nu^{k+1} - \nu^*\|_G^2 \leq \|\nu^k - \nu^*\|_G^2 \\
&- \rho(\eta - 2)\|E^k - \widetilde{E}^k\|_F^2 - \rho\left[\frac{\eta}{\tau} - 2\lambda_{\max}(A^{\mathrm{T}}A)\right]\|Z^k - \widetilde{Z}^k\|_F^2 - \frac{1}{\rho}\|\Lambda^k - \widetilde{\Lambda}^k\|_F^2,
\end{aligned} \tag{24}$$

where $\lambda_{\max}(A^{\mathrm{T}}A)$ denotes the largest eigenvalue of $A^{\mathrm{T}}A$ and

$$\|\nu^{k+1} - \nu^*\|_G^2 = \langle \nu^{k+1} - \nu^*, G \cdot (\nu^{k+1} - \nu^*) \rangle.$$

Proof Due to (14), we can easily derive that

$$\begin{aligned}\|\nu^{k+1}-\nu^*\|_G^2 &= \|\nu^k - d(\nu^k-\widetilde{\nu}^k)-\nu^*\|_G^2 \\ &= \|\nu^k-\nu^*\|_G^2 - 2\langle \nu^k-\nu^*, G\cdot d(\nu^k-\widetilde{\nu}^k)\rangle + \|d(\nu^k-\widetilde{\nu}^k)\|_G^2.\end{aligned} \tag{25}$$

By using (13) and (15), we obtain

$$\begin{aligned}\langle \nu^k-\nu^*, G\cdot d(\nu^k-\widetilde{\nu}^k)\rangle &\geq \langle \nu^k-\widetilde{\nu}^k, G\cdot d(\nu^k-\widetilde{\nu}^k)\rangle \\ &= \eta\rho\|E^k-\widetilde{E}^k\|_F^2 + \frac{\eta\rho}{\tau}\|Z^k-\widetilde{Z}^k\|_F^2 + \frac{1}{\rho}\|\Lambda^k-\widetilde{\Lambda}^k\|_F^2 \\ &\quad - \langle \Lambda^k-\widetilde{\Lambda}^k, E^k-\widetilde{E}^k + A(Z^k-\widetilde{Z}^k)\rangle.\end{aligned} \tag{26}$$

On the other hand, we have

$$\begin{aligned}&\|d(\nu^k-\widetilde{\nu}^k)\|_G^2 \\ &= \eta\rho\|E^k-\widetilde{E}^k\|_F^2 + \frac{\eta\rho}{\tau}\|Z^k-\widetilde{Z}^k\|_F^2 + \frac{1}{\rho}\|\Lambda^k-\widetilde{\Lambda}^k-\rho(E^k-\widetilde{E}^k)-\rho A(Z^k-\widetilde{Z}^k)\|_F^2 \\ &= \eta\rho\|E^k-\widetilde{E}^k\|_F^2 + \frac{\eta\rho}{\tau}\|Z^k-\widetilde{Z}^k\|_F^2 + \frac{1}{\rho}\|\Lambda^k-\widetilde{\Lambda}^k\|_F^2 \\ &\quad - 2\langle \Lambda^k-\widetilde{\Lambda}^k, E^k-\widetilde{E}^k + A(Z^k-\widetilde{Z}^k)\rangle + \rho\|E^k-\widetilde{E}^k + A(Z^k-\widetilde{Z}^k)\|_F^2.\end{aligned} \tag{27}$$

Substituting (26) and (27) into (25), and using the fact

$$\rho\|E^k-\widetilde{E}^k + A(Z^k-\widetilde{Z}^k)\|_F^2 \leq 2\rho\|E^k-\widetilde{E}^k\|_F^2 + 2\rho\|A(Z^k-\widetilde{Z}^k)\|_F^2,$$

it is easy to derive that

$$\begin{aligned}\|\nu^{k+1}-\nu^*\|_G^2 &\leq \|\nu^k-\nu^*\|_G^2 - \eta\rho\|E^k-\widetilde{E}^k\|_F^2 - \frac{\eta\rho}{\tau}\|Z^k-\widetilde{Z}^k\|_F^2 - \frac{1}{\rho}\|\Lambda^k-\widetilde{\Lambda}^k\|_F^2 \\ &\quad + 2\rho\|E^k-\widetilde{E}^k\|_F^2 + 2\rho\|A(Z^k-\widetilde{Z}^k)\|_F^2 \\ &\leq \|\nu^k-\nu^*\|_G^2 - \rho(\eta-2)\|E^k-\widetilde{E}^k\|_F^2 \\ &\quad - \rho\left[\frac{\eta}{\tau} - 2\lambda_{\max}(A^{\mathrm{T}}A)\right]\|Z^k-\widetilde{Z}^k\|_F^2 - \frac{1}{\rho}\|\Lambda^k-\widetilde{\Lambda}^k\|_F^2,\end{aligned}$$

where $\lambda_{\max}(A^{\mathrm{T}}A)$ denotes the largest eigenvalue of $A^{\mathrm{T}}A$. The inequality (24) thus holds, and the lemma is proved. □

Now, we are ready to prove the convergence of the proposed method.

Theorem 4 *Let $\{\nu^k\}$ and $\{\omega^k\}$ be the sequences of the proposed ISM. If $\eta > 2$ and $0 < \tau < 1/\lambda_{\max}(A^{\mathrm{T}}A)$, then $\{\omega^k\}$ converges to a solution point of (P3).*

Proof The proof consists of the following two claims.

1. Any clustering point of $\{\omega^k\}$ is a solution point of (P3).
2. The sequence $\{\omega^k\}$ converges to some ω^∞.

The boundedness of $\{\nu^k\}$ is obvious based on (24). Thus, the rest is to prove the boundedness of F^k. It follows from (24) that

$$\sum_{k=0}^{\infty}\rho(\eta-2)\|E^k-\widetilde{E}^k\|_F^2 + \rho\left[\frac{\eta}{\tau} - 2\lambda_{\max}(A^{\mathrm{T}}A)\right]\|Z^k-\widetilde{Z}^k\|_F^2 + \frac{1}{\rho}\|\Lambda^k-\widetilde{\Lambda}^k\|_F^2 < +\infty.$$

By assumption $\eta > 2$ and $0 < \tau < 1/\lambda_{\max}(A^{\mathrm{T}}A)$, it further implies that

$$\lim_{k\to\infty} \|E^k - \widetilde{E}^k\|_F = 0, \quad \lim_{k\to\infty} \|Z^k - \widetilde{Z}^k\|_F = 0, \quad \lim_{k\to\infty} \|\Lambda^k - \widetilde{\Lambda}^k\|_F = 0. \tag{28}$$

Recall that (6b) implies that $F^{k+1} = M - AZ^k - E^k - \frac{1}{\rho}(\widetilde{\Lambda}^k - \Lambda^k)$. Then the boundedness of F^k is ensured by the boundedness of $\{\Lambda^k - \widetilde{\Lambda}^k\}$ and $\{\nu^k\}$. We thus have that $\{\omega^k\}$ has at least one cluster point. Let

$$\omega^\infty = \begin{bmatrix} F^\infty \\ E^\infty \\ Z^\infty \\ \Lambda^\infty \end{bmatrix}$$

be a cluster point of the sequence $\{\omega^k\}$, and let $\{\omega^{k_j}\}$ be the subsequence converging to ω^∞. From (17), (18), and (28), we get

$$\begin{aligned}
&\lim_{k_j\to\infty} \langle F' - F^{k_j}, -\widetilde{\Lambda}^{k_j-1}\rangle \geq 0, \quad \forall F' \in \mathscr{F}, \\
&\lim_{k_j\to\infty} \left\langle E' - E^{k_j}, \lambda G_1^{k_j} - \widetilde{\Lambda}^{k_j-1}\right\rangle \geq 0, \quad G_1^{k_j} \in \partial \left\| P_\Omega(E^{k_j})\right\|_1, \quad \forall E' \in R^{m\times n}, \\
&\lim_{k_j\to\infty} \left\langle Z' - Z^{k_j}, G_2^{k_j} - A^{\mathrm{T}}\widetilde{\Lambda}^{k_j-1}\right\rangle \geq 0, \quad G_2^{k_j} \in \partial\|Z^{k_j}\|_*, \quad \forall Z' \in R^{d\times n}.
\end{aligned} \tag{29}$$

Note that (6e) and (28) indicate that $\lim_{k_j\to\infty} \|\Lambda^{k_j} - \widetilde{\Lambda}^{k_j-1}\|_F = 0$. Hence, we have

$$\begin{aligned}
&\lim_{k_j\to\infty} \langle F' - F^{k_j}, -\Lambda^{k_j}\rangle \geq 0, \quad \forall F' \in \mathscr{F}, \\
&\lim_{k_j\to\infty} \left\langle E' - E^{k_j}, \lambda G_1^{k_j} - \Lambda^{k_j}\right\rangle \geq 0, \quad G_1^{k_j} \in \partial \left\| P_\Omega(E^{k_j})\right\|_1, \quad \forall E' \in R^{m\times n}, \\
&\lim_{k_j\to\infty} \left\langle Z' - Z^{k_j 1}, G_2^{k_j} - A^{\mathrm{T}}\Lambda^{k_j}\right\rangle \geq 0, \quad G_2^{k_j} \in \partial\|Z^{k_j}\|_*, \quad \forall Z' \in R^{d\times n}.
\end{aligned} \tag{30}$$

Furthermore, (23) and (28) together imply that $\lim_{k_j\to\infty} F^{k_j} + \widetilde{E}^{k_j-1} + A\widetilde{Z}^{k_j-1} - M = 0$. Therefore, from notations: $\widetilde{E}^{k_j-1} = E^{k_j}$, $\widetilde{Z}^{k_j-1} = Z^{k_j}$, we have

$$\lim_{k_j\to\infty} F^{k_j} + E^{k_j} + AZ^{k_j} - M = 0. \tag{31}$$

From (30) and (31), it is obvious that the cluster point ω^∞ is a solution point of (P3). Thus, the first claim is proved.

Finally, we prove the second claim. Based on the first claim, the convergence of $\{\nu^k\}$ is immediately implied by the fact that $\{\nu^k\}$ is Fejér monotone with respect to the set $\mathscr{V}^*$ under G-norm, e.g., see [1]. In other words, $(E^k, Z^k, \Lambda^k) \to (E^\infty, Z^\infty, \Lambda^\infty)$. Recall that $F^{k+1} = M - AZ^k - E^k - \frac{1}{\rho}(\widetilde{\Lambda}^k - \Lambda^k)$. It follows from (28) that $\lim_{k\to\infty} F^{k+1} = F^\infty$, where $F^\infty = M - AZ^\infty - E^\infty$. Overall, we have shown that the sequence $\{\omega^k\}$ converges to ω^∞, which is a solution point of (P3). This completes the proof. □

5 Numerical results

In this section, we test the performance of the proposed method for clustering problems on the synthetic data and on the Extended Yale B face database [16]. All experiments are performed with MATLAB 7.14 and run on a PC (2.70G Hz, 8GB RAM). For all test examples, the

data matrix itself is taken as the dictionary, i.e., $A = X$. After obtaining Z^*, we use a post-processing step as Algorithm 2 in [19] to perform the segmentation. In detail, let $U^*\Sigma^*V^*$ denote the skinny SVD of Z^*. Z^* is used to build the affinity matrix Y as $Y_{ij} = [\tilde{U}\tilde{U}^{\mathrm{T}}]_{ij}^2$, where $\tilde{U} = U^*(\Sigma^*)^{\frac{1}{2}}$. Segmentation results are obtained by using Y to perform a spectral clustering algorithm [24], which segments the data samples into s clusters, where s denotes the number of the subspaces.

To study the segmentation performances of ISM, we compare ISM to some previous subspace segmentation methods, including LRR_1 [18], $\mathrm{LRR}_{2,1}$ [18], and PSLAL [29]. Specifically, LRR_1 denotes the l_1 norm regularization strategy in (P1). And $\mathrm{LRR}_{2,1}$ denotes $l_{2,1}$ norm regularization strategy in (P1). The segmentation result is evaluated by the segmentation Error (Err.), which is defined as follows:

$$\text{Err.} = \frac{\text{number of misclassfied points}}{\text{total number of points}}.$$

5.1 Synthetic data

The synthetic data set is created by the following procedure (see also [20,29]). Five independent subspaces $\{S_i\}_{i=1}^5$ are constructed, whose bases $\{U_i\}_{i=1}^5$ are generated by $U_{i+1} = TU_i$ $(1 \le i \le 4)$, where T denotes a random rotation and U_i denotes a random orthogonal matrix of dimension 150×4. Hence, each subspace S_i has a rank of 4 and the data points have an ambient dimension of 150. $\bar{n}$ data points are sampled from each subspace by using $X_i^0 = U_i Q_i$ $(1 \le i \le 5)$, where Q_i being a $4 \times \bar{n}$ independent and identically distributed $\mathcal{N}(0, 1)$ matrix. In summary, the whole data matrix is formulated as $X = [X_1^0, \ldots, X_5^0] \in R^{150\times n}$ with rank $r = 20$ and $n = 5 \times \bar{n}$. The index of observed entries Ω is determined at random. Let the quantity sr represent the ratios of the observed entries, i.e., $\|\Omega\|/mn$. Furthermore, we add the sparse noise and Gaussian noise as follows. We randomly choose some data vectors to be corrupted with uniformly distributed noise between $[-1,1]$. The quantity spr represents the percentage of the corrupted data vectors. Besides, 20% of entries are contaminated with Gaussian noise $\mathcal{N}(0, 0.01)$.

5.1.1 Segmentation performances

In this experiment, we apply LRR_1, $\mathrm{LRR}_{2,1}$, PSLAL, and ISM to solve the problems with different values of sr and spr. The parameter λ for LRR_1, $\mathrm{LRR}_{2,1}$, and PSLAL is set to 10^{-3}, 10^{-1}, 10^{-1}, respectively. For all other parameters, we use the default values as provided by the authors. For ISM, we set $\delta = 10^{-3}$, $\rho^0 = 1$, $\lambda = 10^{-2}$, and the stopping tolerance $\varepsilon = 10^{-6}$. Ten realizations are computed for each (sr, spr) pair. Table 1 reports the average segmentation errors and the CPU time for the case $n = 500$.

Table 1 shows that the errors of all methods decrease as the number of sr increases, whereas segmentation errors increase as the number of spr increases. This may be caused by the fact that both more missing data and more noises increase the difficulty in the subspace segmentation. In addition, we can see that our proposed method almost consistently outperforms other methods on these test examples in terms of segmentation errors. These results confirm that model (P3) is quite robust and efficient at various levels of noise.

Moreover, we also demonstrate the segmentation performances of ISM under an increasing number of data points. Since it is costly to perform NCut [24] for computing the final segmentation results for data points $n > 1800$. For the inner memory limitation, we therefore

Table 1 Comparison of segmentation errors (%) and the computing time (in seconds) for various approaches with different pairs of (sr, spr)

spr	sr	LRR_1		$LRR_{2,1}$		PSLAL		ISM	
		Time	Err.	Time	Err.	Time	Err.	Time	Err.
0.05	0.5	7.71	77.92	8.99	11.16	4.25	11.20	1.78	10.16
	0.6	7.24	69.16	8.93	6.48	5.02	6.52	1.71	6.04
	0.7	6.99	9.12	7.70	3.96	4.22	3.96	1.53	3.96
	0.8	7.60	5.16	8.39	1.52	4.70	1.52	1.74	1.60
	0.9	7.43	6.08	7.98	1.04	4.69	1.04	1.69	1.00
0.1	0.5	7.33	34.32	8.01	68.20	4.48	68.44	1.72	26.64
	0.6	7.44	29.20	9.36	48.80	5.21	48.88	1.63	15.52
	0.7	7.17	34.32	7.82	21.24	4.29	21.64	1.58	10.08
	0.8	7.43	9.64	8.22	12.28	4.54	12.88	1.71	5.64
	0.9	7.26	11.72	7.97	7.40	4.67	7.80	1.63	3.48

Table 2 Comparison of segmentation errors (%) and the computing time (in seconds) for various approaches with different sizes of n

n	spr	sr	LRR_1		$LRR_{2,1}$		PSLAL		ISM	
			Time	Err.	Time	Err.	Time	Err.	Time	Err.
200	0.05	0.6	3.03	78.50	3.13	18.00	1.51	20.00	1.03	5.00
		0.8	3.08	22.50	3.16	2.00	1.38	2.00	0.84	2.00
	0.1	0.6	3.11	78.50	3.42	25.00	1.35	25.00	0.75	24.00
		0.8	3.08	61.00	3.33	8.50	1.52	8.50	0.76	8.00
600	0.05	0.6	7.31	38.00	7.97	8.67	4.35	8.83	1.61	6.50
		0.8	7.10	31.00	8.05	2.33	4.78	2.50	1.70	1.67
	0.1	0.6	7.30	16.50	8.59	75.33	5.14	76.83	1.59	15.83
		0.8	7.81	20.33	8.62	69.00	5.43	68.83	1.72	6.50
1000	0.05	0.6	11.57	5.20	8.52	78.30	9.76	76.30	2.56	4.60
		0.8	11.48	1.30	8.56	79.60	8.33	76.10	2.58	1.20
	0.1	0.6	11.38	23.80	5.55	79.70	10.34	79.70	2.50	12.50
		0.8	11.18	6.00	5.47	77.90	10.38	78.20	2.52	4.60
1400	0.05	0.6	15.09	13.86	7.65	78.21	13.96	78.14	3.24	4.29
		0.8	15.05	2.21	10.62	75.79	14.04	76.21	3.34	1.64
	0.1	0.6	15.21	13.64	7.45	78.50	14.29	78.50	3.28	12.14
		0.8	15.69	6.50	7.80	77.43	14.77	77.43	3.38	5.71
1800	0.05	0.6	19.84	6.44	12.99	77.22	18.61	77.22	4.46	4.83
		0.8	19.59	1.28	9.71	78.83	18.19	78.83	4.27	0.94
	0.1	0.6	20.01	16.78	9.87	78.22	17.86	77.72	4.13	14.72
		0.8	19.16	5.28	9.28	77.72	18.35	77.44	4.24	4.06

choose to test the examples with $n \leq 1800$. We vary the number of data points from 200 to 1800 with an increment of 400. The average results of each setting over five instances are reported in Table 2. All parameters are set in the same way as the previous example.

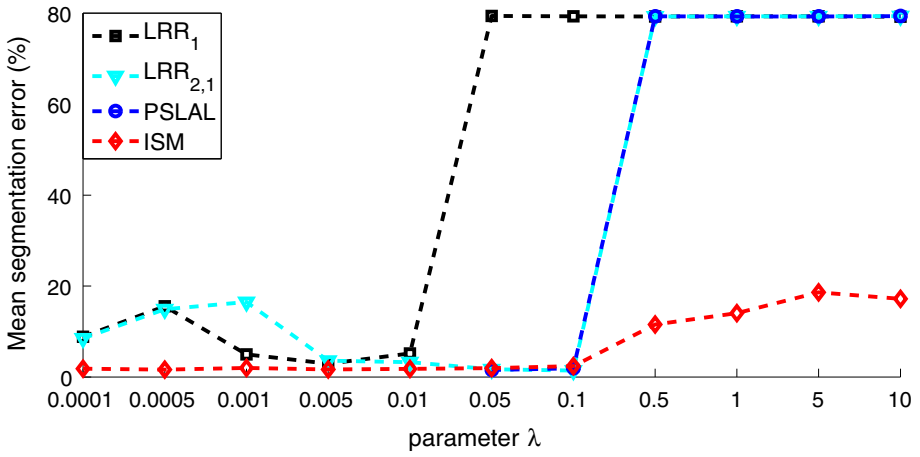

Fig. 1 The influences of the parameter λ on different methods

As shown in Table 2, ISM outperforms LRR_1, $\mathrm{LRR}_{2,1}$, and PSLAL in terms of clustering accuracy. What is more, our method stands out from other methods thanks to the consideration of the incomplete and noisy observations.

5.1.2 Effect of λ

The parameter λ is used to balance the effects between the low-rank part AZ and the noise part E. In general, the choice of λ depends on the prior knowledge of the data error level. For example, when the errors are small, we should use a relatively large λ. In contrast, when the errors are large, we should set λ to be rather small. Figure 1 shows the performance of ISM while the parameter λ varies from 10^{-4} to 10. For this experiment, we set $n = 500$, $sr = 0.8$, and $spr = 0.05$. For each λ, we create five instances randomly and report the average segmentation errors.

Figure 1 illustrates that the segmentation errors obtained by all methods increase when λ is large. Besides, there exists a range of parameters λ where ISM obtains segmentation errors less than 5%. We also notice that PSLAL fails to segment the data for $\lambda \leq 10^{-2}$. This may be caused by the small value of λ leading to a too small penalty on the noise. Generally, our method is less sensitive to λ than other methods.

5.2 Face clustering

We now turn to the real clustering tasks using the Extended Yale B face database [16]. In this database, there are frontal face images of 38 human subjects under 9 poses and 64 illumination conditions. The database partitions these images into 38 classes and each one contains 64 face images with size 192×168. We only consider the first eight subjects of them (see Fig. 2). In other words, there are 512 images used in experiments. To reduce the computational cost, we resize the test images into 48×42 and re-scale pixels into [0, 1]. The parameter δ of (P3) is set to $0.5 \max\{\min |X_{ij}|\}$.

Fig. 2 Example images from Extended Yale B

Table 3 Segmentation errors (%) and the computing time (in seconds) on the Extended Yale B database

No. subjects	LRR_1		$LRR_{2,1}$		PSLAL		ISM	
	Time	Err.	Time	Err.	Time	Err.	Time	Err.
2	3.49	2.34	3.27	2.34	3.50	1.56	5.21	1.56
3	6.50	9.38	6.09	5.21	6.43	5.73	9.01	4.17
4	10.93	8.98	10.26	6.64	10.93	3.52	16.11	3.13
5	16.07	6.56	14.55	4.38	14.83	3.75	20.20	2.50
6	23.93	10.68	20.49	3.39	21.19	6.25	27.77	2.34
7	31.53	12.05	26.48	8.93	28.29	10.71	35.88	7.81
8	47.06	12.30	41.16	9.77	41.95	12.11	51.53	9.18

5.2.1 Results on original Extended Yale B database

We first investigate the segmentation performances of LRR_1, $LRR_{2,1}$, PSLAL, and ISM on the Extended Yale B database. We use the first $N \in \{2, 3, 4, 5, 6, 7, 8\}$ subject classes for the face clustering. The parameter λ of LRR_1, $LRR_{2,1}$, PSLAL is set to 0.02, 0.2, 10^3, respectively. And we set $\lambda = 10^4$ and $\rho^0 = 10^{-2}$ for ISM. All other parameters are set to the default values. Table 3 shows the results of applying different methods to the original data.

From Table 3, we can see that our method almost achieves the lowest segmentation errors for all these test examples. In short, ISM outperforms other methods in terms of clustering accuracy. These results clearly show that our consideration for the incomplete and noisy observations positively contributes in improving clustering performance.

Moreover, our approach can also be applied to the error correction. Figure 3 shows some examples of the ISM in removing the shadows or specular lights. Unlike the experiments setting in the previous test, the parameter is chosen as $\lambda = 0.1$ for ISM in this test. As shown in Fig. 3, ISM removes the heavy noise well.

5.2.2 Results on contiguous occlusions corruptions

We now aim to demonstrate the ability of the ISM in dealing with the missing data. We randomly add contiguous occlusions to images with block size 5×5, 10×10, 15×15, 20×20, 25×25, and 30×30, as shown in Fig. 4. For each size, the percentage of data being corrupted varies from 10 to 50%. In this test, we only consider the first five subjects of the Extended Yale B face database. In other words, there are 320 images used for this experiment. The parameter λ of LRR_1, $LRR_{2,1}$, PSLAL, ISM is set to 10^{-3}, 10^{-1}, 10^3, 10^3, respectively. And all the other parameters are set in the same way as the previous test. And we report the segmentation errors of different methods on various levels of occlusions in Fig. 5.

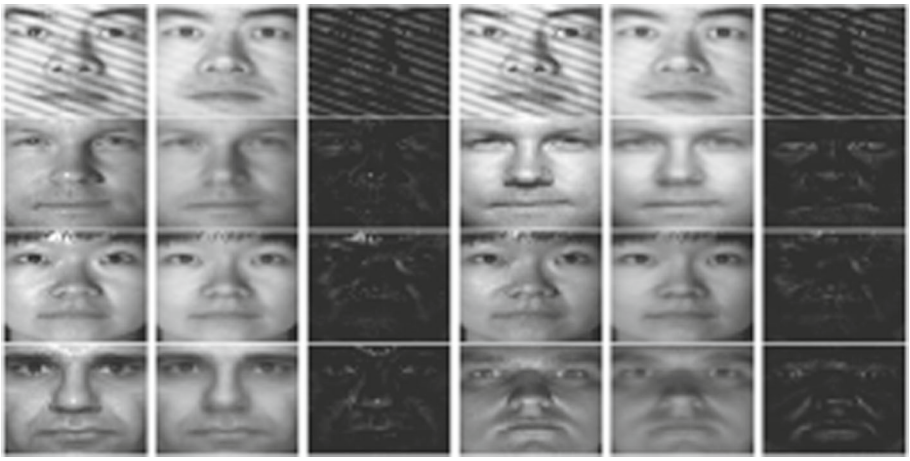

Fig. 3 Some examples using ISM to correct the corrupted images. The original data (first and the fourth column), the corrected data (the second and the fifth column) and the errors (the third and the last column)

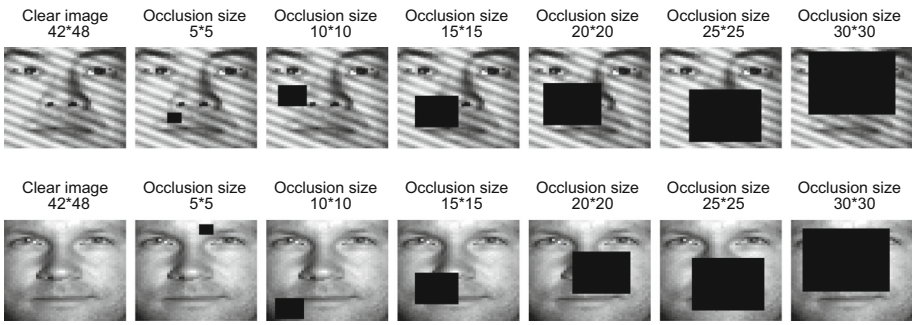

Fig. 4 Some examples of original and corrupted images with different levels of occlusions

From Fig. 5, one can see that the segmentation error curves of our method are always below those of the other methods under six different scenarios. It indicates that ISM outperforms other algorithms in terms of segmentation errors in dealing with occlusion corruptions. This also suggests that the model (P3) is robust to the contiguous occlusions corruptions.

We also visualize the effectiveness of the ISM in error correction. Figure 6 shows some examples that ISM recovers images from the 30×30 block noise with the corrupted percentage 50%. Clearly, ISM can remove block noises well.

6 Conclusion

In this paper, we have considered more practical circumstances for the subspace segmentation problem. First, we have extended the well-known model (P1) [18] to more practical circumstances: only a fraction of entries of data can be observed, and the observed data are corrupted by both impulsive and Gaussian noise. Then, an inexact splitting method has been developed for solving the resulted model. Furthermore, we have proved the global convergence of this method. Experimental results on synthetic and benchmark data illustrate that

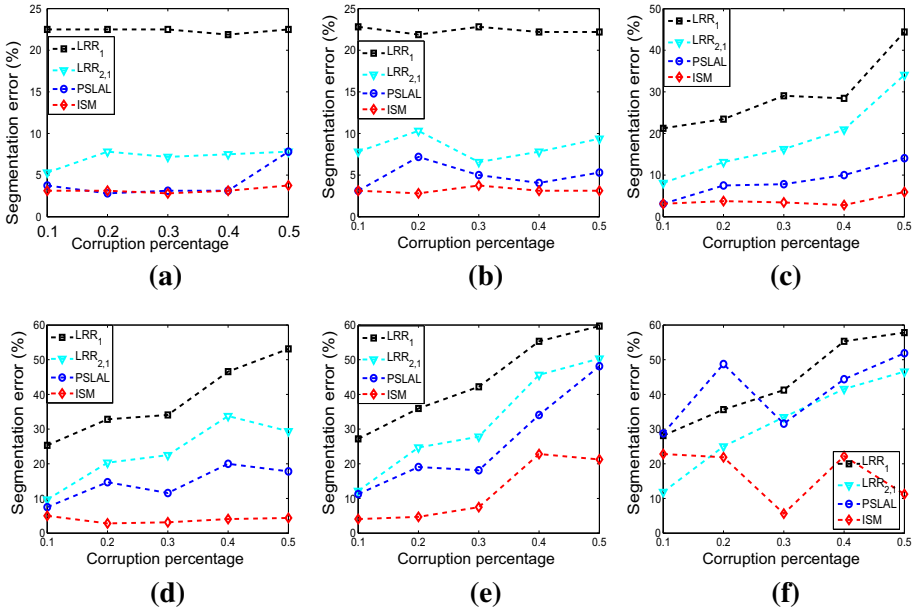

Fig. 5 Segmentation errors of different methods with various levels of occlusions. **a** Occlusion size 5×5. **b** Occlusion size 10×10. **c** Occlusion size 15×15. **d** Occlusion size 20×20. **e** Occlusion size 25×25. **f** Occlusion size 30×30

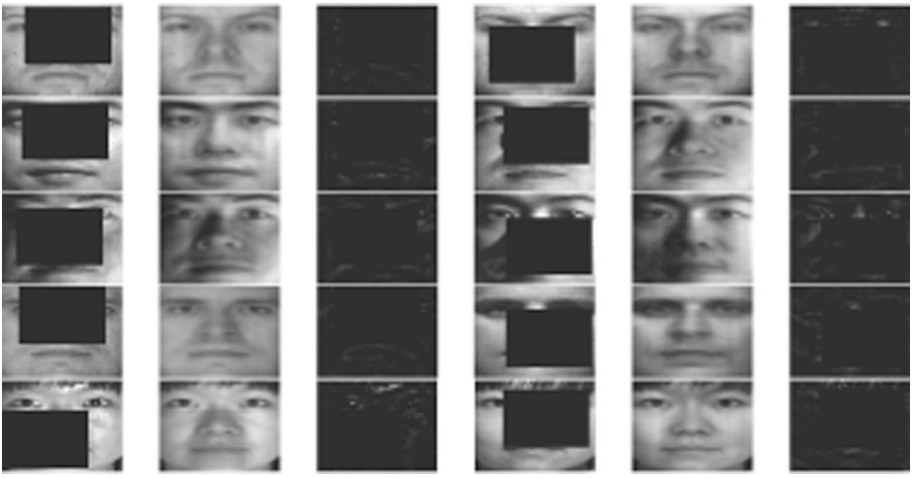

Fig. 6 Some examples using ISM to recover the corrupted images. The contaminated data (first and the fourth column), the corrected data (the second and the fifth column), and the errors (the third and the last column)

the proposed method is computationally efficient, robust as well as more accurate compared with the state-of-the-art algorithms.

Acknowledgements The authors would like to thank the financial support from the China Scholarship Council.

References

1. Bauschke, H., Combettes, P.: A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces. Math. Oper. Res. **26**(2), 248–264 (2001)
2. Bradley, P., Mangasarian, O.: k-Plane clustering. J. Glob. Optim. **16**(1), 23–32 (2000). https://doi.org/10.1023/A:1008324625522
3. Butenko, S., Chaovalitwongse, W., Pardalos, P.: Clustering Challenges in Biological Networks. World Scientific, Singapore (2009)
4. Cai, J., Candès, E., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim. **20**(4), 1956–1982 (2010)
5. Candès, E., Plan, Y.: Matrix completion with noise. Proc. IEEE **98**(6), 925–936 (2010)
6. Candès, E., Tao, T.: Decoding by linear programming. IEEE Trans. Inf. Theory **51**(12), 4203–4215 (2005)
7. Chen, G., Lerman, G.: Spectral curvature clustering (SCC). Int. J. Comput. Vis. **81**(3), 317–330 (2009)
8. Elhamifar, E., Vidal, R.: Sparse subspace clustering: algorithm, theory, and applications. IEEE Trans. Pattern Anal. Mach. Intell. **35**(11), 2765–2781 (2013). https://doi.org/10.1109/TPAMI.2013.57
9. Goh, A., Vidal, R.: Segmenting motions of different types by unsupervised manifold clustering. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–6 (2007)
10. Gruber, A., Weiss, Y.: Multibody factorization with uncertainty and missing data using the EM algorithm. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 707–714 (2004). https://doi.org/10.1109/CVPR.2004.1315101
11. Han, L., Bi, S.: Two-stage convex relaxation approach to low-rank and sparsity regularized least squares loss. J. Glob. Optim. (2017). https://doi.org/10.1007/s10898-017-0573-2
12. He, B., Tao, M., Yuan, X.: A splitting method for separable convex programming. IMA J. Numer. Anal. **35**(1), 394–426 (2015)
13. Ho, J., Yang, M., Lim, J., Lee, K., Kriegman, D.: Clustering appearances of objects under varying illumination conditions. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 11–18 (2003)
14. Hong, W., Wright, J., Huang, K., Ma, Y.: Multiscale hybrid linear models for lossy image representation. IEEE Trans. Image Process. **15**(12), 3655–3671 (2006)
15. Kanatani, K.: Motion segmentation by subspace separation: model selection and reliability evaluation. Int. J. Image Graph. **2**(2), 179–197 (2002)
16. Lee, K., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 684–698 (2005)
17. Lin, Z., Chen, M., Ma, Y.: The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices (2010). Eprint arXiv:1009.5055
18. Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y.: Robust recovery of subspace structures by low-rank representation. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 171–184 (2013)
19. Liu, G., Yan, S.: Latent low-rank representation for subspace segmentation and feature extraction. In: 2011 International Conference on Computer Vision, pp. 1615–1622 (2011). https://doi.org/10.1109/ICCV.2011.6126422
20. Liu, Y., Jiao, L., Shang, F.: A fast tri-factorization method for low-rank matrix recovery and completion. Pattern Recognit. **46**(1), 163–173 (2013)
21. Lu, C., Min, H., Zhao, Z., Zhu, L., Huang, D., Yan, S.: Robust and Efficient Subspace Segmentation via Least Squares Regression, pp. 347–360. Springer, Berlin (2012)
22. Ma, Y., Yang, A., Derksen, H., Fossum, R.: Estimation of subspace arrangements with applications in modeling and segmenting mixed data. SIAM Rev. **50**(3), 413–458 (2008)
23. Rao, S., Tron, R., Vidal, R., Ma, Y.: Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories. IEEE Trans. Pattern Anal. Mach. Intell. **32**(10), 1832–1845 (2010)
24. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 888–905 (2000)
25. Tao, M., Yuan, X.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM J. Optim. **21**(1), 57–81 (2011)
26. Tipping, M., Bishop, C.: Mixtures of probabilistic principal component analyzers. Neural Comput. **11**(2), 443–482 (1999)
27. Tseng, P.: Nearest q-flat to m points. J. Optim. Theory Appl. **105**(1), 249–252 (2000)
28. Vidal, R., Ma, Y., Sastry, S.: Generalized principal component analysis (GPCA). IEEE Trans. Pattern Anal. Mach. Intell. **27**(12), 1945–1959 (2005)
29. Xiao, Y., Wu, S., Li, D.: Splitting and linearizing augmented Lagrangian algorithm for subspace recovery from corrupted observations. Adv. Comput. Math. **38**(4), 837–858 (2013)

30. Yan, J., Pollefeys, M.: A General Framework for Motion Segmentation: Independent, Articulated, Rigid, Non-rigid, Degenerate and Non-degenerate, pp. 94–106. Springer, Berlin (2006)
31. Yang, J., Yin, W., Zhang, Y., Wang, Y.: A fast algorithm for edge-preserving variational multichannel image restoration. SIAM J. Imaging Sci. **2**(2), 569–592 (2009)
32. Zhang, C., Bitmead, R.: Subspace system identification for training-based MIMO channel estimation. Automatica **41**(9), 1623–1632 (2005)
33. Zhang, T., Szlam, A., Lerman, G.: Median k-flats for hybrid linear modeling with many outliers. In: 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, pp. 234–241 (2009)
34. Zhang, T., Szlam, A., Wang, Y., Lerman, G.: Hybrid linear modeling via local best-fit flats. Int. J. Comput. Vis. **100**(3), 217–240 (2012)