Robust Principal Component Analysis with Missing Data

Fanhua Shang[†], Yuanyuan Liu[‡], James Cheng[†], Hong Cheng[‡]
[†]Dept. of Computer Science and Engineering, The Chinese University of Hong Kong
{fhshang, jcheng}@cse.cuhk.edu.hk
[‡]Dept. of Systems Engineering and Engineering Management, The Chinese University of Hong Kong
{yyliu, hcheng}@se.cuhk.edu.hk

ABSTRACT

Recovering matrices from incomplete and corrupted observations is a fundamental problem with many applications in various areas of science and engineering. In theory, under certain conditions, this problem can be solved via a natural convex relaxation. However, all current provable algorithms suffer from superlinear per-iteration cost, which severely limits their applicability to large scale problems. In this paper, we propose a robust principal component analysis (RPCA) plus matrix completion framework to recover low-rank and sparse matrices from missing and grossly corrupted observations. Under the unified framework, we first present a convex robust matrix completion model to replace the linear projection operator constraint by a simple equality one. To further improve the efficiency of our convex model, we also develop a scalable structured factorization model, which can yield an orthogonal dictionary and a robust data representation simultaneously. Then, we develop two alternating direction augmented Lagrangian (ADAL) algorithms to efficiently solve the proposed problems. Finally, we discuss the convergence analysis of our algorithms. Experimental results verified both the efficiency and effectiveness of our methods compared with the state-of-the-art algorithms.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database applications-Data Mining

Keywords

RPCA, robust matrix completion, low-rank, ADAL

1. INTRODUCTION

In recent years, high dimensional data are becoming increasingly ubiquitous such as digital photographs, surveillance videos and social network data. Such high dimensional data are becoming more commonly available due to the advance in data collection technologies. However, the "curse of dimensionality" has rendered many tasks such as inference, learning, and recognition, impractical. Principal component analysis (PCA) is arguably the most widely used technique for dimensionality reduction in statistical data analysis, mainly because it is simple to implement, can be solved efficiently, and is often effective in real-world applications such as latent semantic indexing, face recognition and text clustering. However, one of the main challenges faced by PCA is that the observed data is often contaminated by outliers or missing values.

This problem has drawn much attention from researchers in various communities such as data mining, machine learning, signal/image processing, and computer vision [21, 30, 5, 27, 34, 7, 16]. Based on compressive sensing and rank minimization, many methods for recovering low-rank and sparse matrices (also called robust principal component analysis or RPCA [30]) with incomplete or grossly corrupted observations have been proposed, such as principal component pursuit (PCP) [5] and outlier pursuit [31]. In principle, those methods aim to minimize a hybrid optimization problem involving both the l_1-norm and trace norm minimization. It is recognized that the l_1-norm and the trace norm as the convex surrogates for the l_0-norm and the rank function are powerfully capable of inducing sparse and low-rank, respectively [24, 5]. In addition, Xu et al. [31] used the $l_{1,2}$-norm to model corrupted columns.

Although RPCA has been well studied in recent years, there is little work focusing on RPCA plus matrix completion (also called robust matrix completion or RMC [7]). In this paper, we are particularly interested in the trace norm regularized problem for RMC:

$$\min_X \|X\|_* + \lambda f(X), \tag{1}$$

where $\|X\|_*$ is the trace norm of the desired matrix X, i.e., the sum of its singular values, $\lambda \geq 0$ is a regularization parameter, and $f(\cdot)$ denotes the loss function such as the l_2-norm loss or the l_1-norm (or $l_{1,2}$-norm) loss functions. In the following, we give a few examples of applications where the RMC is useful.

- **Robust principal component analysis** (**RPCA**). When the loss function is the l_1-norm loss, the model (1) is a RPCA problem, which is adopted by a great number of emerging approaches such as PCP [5], and has been successfully applied in many important problems such as latent semantic indexing [21], video surveillance [30, 5], and low-rank textures.
- **Matrix completion** (**MC**). When the loss function is the l_2-norm loss, and only a relatively small number

CIKM'14, November 3–7, 2014, Shanghai, China.

http://dx.doi.org/10.1145/2661829.2662083.

of entries are observed, the goal of this problem is to complete a low-rank matrix from incomplete samples of its entries. This problem, also called matrix completion, is fundamental for collaborative filtering [6], link prediction, and global positioning.

In this paper, we aim to recover both low-rank and sparse matrices from missing and grossly corrupted observations. Our solution provides a good approximation to the original data contaminated with both outliers and missing values. Unlike existing RPCA methods, our approach not only takes into account the fact that the observations are contaminated by additive outliers or missing values, but can also identify both low-rank and sparse components from missing and grossly corrupted observations. We develop a provable and scalable solution framework for RPCA and RMC problems, which is particularly useful in this "big data" era where many real-world applications need to deal with large, high dimensional data with missing and corrupted values. We conduct extensive experiments that verify both the efficiency and effectiveness of our methods.

We summarize the main contributions of our work as follows:

- We propose a unified RMC framework to recover both low-rank and sparse matrices from missing and grossly corrupted observations, where the loss function can be selected as the l_2-norm or the l_1-norm.
- We present both convex and non-convex scalable models to replace the linear projection operator constraint by a simple equality one. With the orthogonality constraint applied to the dictionary component, we convert the non-convex model into a smaller-scale matrix trace norm regularized problem.
- We develop two efficient alternating direction augmented Lagrangian (ADAL) algorithms to solve the proposed methods, which can be accelerated by adaptively changing the penalty parameter.
- Finally, we also provide the convergence analysis of our convex and non-convex algorithms.

This paper is organized as follows. We review background and related work in Section 2. In Section 3, we present a convex and a non-convex scalable trace norm regularized models. We develop two efficient ADAL algorithms in Section 4. We provide the theoretical analysis of our algorithms in Section 5. We report empirical results in Section 6, and conclude this paper in Section 7.

2. BACKGROUND AND PROBLEM FORMULATION

Given a matrix, some of its entries may not be observed due to problems in the acquisition process, e.g., loss of information or high cost of experiments to obtain complete data [1]. For the matrix completion problem, many iterative thresholding algorithms [25, 19] have been proposed to solve the trace norm regularized linear least squares problem

$$\min_X \|X\|_* + \frac{\lambda}{2}\|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(Z)\|_F^2, \tag{2}$$

where $\mathcal{P}_\Omega(Z)$ is defined as the projection of the matrix Z on the observed entries Ω: $\mathcal{P}_\Omega(Z)_{ij} = Z_{ij}$ if $(i,j) \in \Omega$ and $\mathcal{P}_\Omega(Z)_{ij} = 0$ otherwise. In other words, $f(\cdot)$ is the l_2-norm loss function, i.e., $f(X) = \frac{1}{2}\|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(Z)\|_F^2$.

A low-rank matrix X can be recovered from highly corrupted matrix $Z = X + E$ via the following trace norm and l_1-norm minimization problem

$$\min_X \|X\|_* + \lambda\|Z - X\|_1, \tag{3}$$

where $\|\cdot\|_1$ indicates the element-wise l_1-norm, i.e., $\|E\|_1 = \Sigma_{ij}|e_{ij}|$ and $E = Z - X$. The model (3) is called the RPCA problem [30], where $f(\cdot)$ is the l_1-norm loss function, i.e., $f(X) = \|Z - X\|_1$. Several algorithms have been developed to solve the convex optimization problem (3), such as PCP [5] and IALM [17].

A more general RMC model in [7] and [16] aims to simultaneously recover both low-rank and sparse components from incomplete and grossly corrupted observations via the convex optimization problem,

$$\begin{aligned} &\min_{X,E} \|X\|_* + \lambda\|E\|_1, \\ &\text{s.t., } \mathcal{P}_\Omega(X + E) = \mathcal{P}_\Omega(Z). \end{aligned} \tag{4}$$

Chen et al. [7] and Li [16] provided theoretical performance guarantees when minimizing trace norm plus l_1-norm succeeds in exact recovery. Although the RMC model (4) is a convex optimization problem, and can be solved by some convex algorithms, some additional variables need to be introduced. In addition, all current provable algorithms for RMC involve the singular vector decomposition (SVD), and thus they suffer from high computational cost of full or partial SVDs, which severely limits their applicability to large scale problems. To address the issues, we propose a scalable RMC method to recover matrices with missing and grossly corrupted observations.

3. CONVEX AND NON-CONVEX RMC MODELS

From the optimization problem (4), we can find that the optimal solution E_{Ω^C}=$\mathbf{0}$, where Ω^C is the complement of Ω, i.e., the index set of unobserved entries. Consequently, we have the following lemma.

LEMMA 1. *The RMC problem (4) is equivalent to the following convex optimization problem,*

$$\begin{aligned} &\min_{X,E} \|X\|_* + \lambda\|\mathcal{P}_\Omega(E)\|_1, \\ &\text{s.t., } \mathcal{P}_\Omega(X + E) = \mathcal{P}_\Omega(Z),\ E_{\Omega^C} = \mathbf{0}. \end{aligned} \tag{5}$$

3.1 Convex RMC Model

To efficiently solve the RMC problem (4) and avoid introducing some auxiliary variables, we can assume without loss of generality that the unobserved data Z_{Ω^C}=$\mathbf{0}$, and E_{Ω^C} may be any values such that $Z_{\Omega^C} = X_{\Omega^C} + E_{\Omega^C}$. Therefore, the constraint with a linear projection operator $\mathcal{P}_\Omega$ in (4) is simplified into $Z = X + E$. It is worth noting that at last E_{Ω^C} will be set to $\mathbf{0}$ for the expected output E. Hence, we replace the constraint $\mathcal{P}_\Omega(Z) = \mathcal{P}_\Omega(X+E)$ with $Z = X+E$, and achieve the following equivalent form:

$$\begin{aligned} &\min_{X,E} \|X\|_* + \lambda\|\mathcal{P}_\Omega(E)\|_1, \\ &\text{s.t., } X + E = Z. \end{aligned} \tag{6}$$

To further improve the efficiency of our convex model (6) and the scalability of handling large data sets, we also propose a scalable non-convex model.

3.2 Non-Convex RMC Model

The desired low-rank matrix X is factorized into two much smaller matrices $G \in \mathbb{R}^{m\times d}$ ($G^TG = I$) and $H \in \mathbb{R}^{n\times d}$ whose product is equal to X, i.e., $X = GH^T$, where d is an upper bound for the rank of the matrix X, i.e., $d \geq r = \text{rank}(X)$. We have the following lemma.

LEMMA 2. *Let G and H be two matrices of compatible dimensions, where G has orthogonal columns, i.e., $G^TG = I$, then we have $\|GH^T\|_* = \|H\|_*$.*

By substituting $GH^T = X$ and $\|H\|_* = \|X\|_*$ into (6), we obtain a much smaller-scale matrix trace norm minimization problem,

$$\begin{aligned} &\min_{G,H,E} \|H\|_* + \lambda\|\mathcal{P}_\Omega(E)\|_1, \\ &\text{s.t., } GH^T + E = Z,\ G^TG = I. \end{aligned} \tag{7}$$

THEOREM 1. *Suppose (X^*, E^*) is a solution of the convex problem (6) with rank$(X^*) = r$, then there exists the solution $G \in \mathbb{R}^{m\times d}$, $H \in \mathbb{R}^{n\times d}$ and $E \in \mathbb{R}^{m\times n}$ to the problem (7) with $d \geq r$ and $E_{\Omega^C} = \mathbf{0}$ such that $GH^T = X^*$, and (GH^T, E) is also a solution to the problem (6).*

PROOF. If we know that (X^*, E^*) is a solution to the convex optimization problem (6), it is also a solution to

$$\begin{aligned} &\min_{X,E,\text{rank}(X)=r} \|X\|_* + \lambda\|\mathcal{P}_\Omega(E)\|_1, \\ &\text{s.t., } \mathcal{P}_\Omega(Z) = \mathcal{P}_\Omega(X+E),\ \mathcal{P}_{\Omega^C}(E) = \mathbf{0}. \end{aligned}$$

Since for any (X^*, E^*) with rank$(X^*) = r$, we can find $G \in \mathbb{R}^{m\times d}$ and $H \in \mathbb{R}^{n\times d}$ satisfying $GH^T = X^*$ and $\mathcal{P}_\Omega(Z - GH^T) = \mathcal{P}_\Omega(E) = \mathcal{P}_\Omega(E^*)$, where $d \geq r$. Moreover, according to Lemma 2, we have

$$\begin{aligned} &\min_{G,H,E} \|H\|_* + \lambda\|\mathcal{P}_\Omega(E)\|_1, \\ &\quad \text{s.t., } \mathcal{P}_\Omega(Z) = \mathcal{P}_\Omega(GH^T + E),\ G^TG = I, \\ &= \min_{G,H,E} \|GH^T\|_* + \lambda\|\mathcal{P}_\Omega(E)\|_1, \\ &\quad \text{s.t., } \mathcal{P}_\Omega(Z) = \mathcal{P}_\Omega(GH^T + E), \\ &= \min_{X,E,\text{rank}(X)=r} \|X\|_* + \lambda\|\mathcal{P}_\Omega(E)\|_1, \\ &\quad \text{s.t., } \mathcal{P}_\Omega(Z) = \mathcal{P}_\Omega(X+E), \end{aligned}$$

where $\mathcal{P}_{\Omega^C}(E)$=$\mathbf{0}$. □

In the following, we will discuss how to solve our convex and non-convex models (6) and (7). It is worth noting that the RPCA problem (3) can be viewed as a special case of both (6) and (7), where all entries of the corrupted matrix are directly observed. We will develop two efficient alternating direction augmented Lagrangian (ADAL) solvers for solving our convex model (6) and non-convex model (7), respectively. It is also worth noting that although our non-convex algorithm will produce different estimations of G and H, the estimation of GH^T is stable as guaranteed by Theorem 1 and the convexity of the problem (6).

4. OPTIMIZATION ALGORITHMS

In this section, we will develop two efficient alternating direction augmented Lagrangian (ADAL) algorithms for solving both problems (6) and (7). First, we design a general convex ADAL scheme for solving the convex problem (6). Then we propose a similar procedure for solving the non-convex problem (7). The convergence analysis of our algorithms is provided in the next section.

Algorithm 1 Solving the problem (8) via ADAL

Input: γ_0.
Initialize: $x_0 = \mathbf{0}$, $z_0 = \mathbf{0}$ and $\gamma_0 = \mathbf{0}$.
  for $k = 0,\ 1,\ \cdots,\ T$ **do**
    $x_{k+1} = \arg\min_x\ \mathcal{L}_\mu(x, z_k, \gamma_k)$.
    $z_{k+1} = \arg\min_z\ \mathcal{L}_\mu(x_{k+1}, z, \gamma_k)$.
    $\gamma_{k+1} = \gamma_k + \mu(Ax_{k+1} + Bz_{k+1} - c)$.
  end for
Output: x_k and z_k.

4.1 Generic Formulation

We describe our optimization algorithm based on the ADAL method (also known as the alternating direction method of multipliers) for solving (6). The ADAL method was introduced for optimization in the 1970's, and its origins can be traced back to techniques for solving partial difference equations in the 1950's. It has received renewed interest due to the fact that it is efficient to tackle large scale problems and solve optimization problems with multiple non-smooth terms in the objective function [17]. The ADAL can be considered as an approximation of the method of multipliers. It decomposes a large global problem into a series of smaller subproblems, and coordinates the solutions of subproblems to compute the globally optimal solution. The problem solved by ADAL takes the following generic form

$$\begin{aligned} &\min_{x\in\mathbb{R}^n, z\in\mathbb{R}^m}\ f(x) + g(z), \\ &\text{s.t., } Ax + Bz = c, \end{aligned} \tag{8}$$

where both $f(\cdot)$ and $g(\cdot)$ are convex functions. ADAL reformulates the problem using a variant of the augmented Lagrangian method as follows:

$$\mathcal{L}_\mu(x, z, \gamma) = f(x)+g(z)+\gamma^T(Ax+Bz-c)+\frac{\mu}{2}\|Ax+Bz-c\|_2^2,$$

where γ is the Lagrangian multiplier and μ is a penalty parameter. ADAL solves the problem (8) by iteratively minimizing $\mathcal{L}_\mu(x, z, \gamma)$ over $x,\ z$, and then updating γ, as outlined in **Algorithm 1** [2].

4.2 Convex ADAL Scheme

Our RMC optimization problem (6) can be solved by ADAL. The augmented Lagrangian of (6) is given by

$$\begin{aligned} \mathcal{L}_\mu(X, E, Y) &= \|X\|_* + \lambda\|\mathcal{P}_\Omega(E)\|_1 \\ &+ \langle Y,\ Z - X - E\rangle + \frac{\mu}{2}\|Z - X - E\|_F^2. \end{aligned} \tag{9}$$

Applying ADAL, we carry out the following updating steps in each iteration.

4.2.1 Updating X

With all other variables fixed, the optimal X is the solution to the following problem:

$$\min_X \|X\|_* + \frac{\mu}{2}\|X - Z + E - Y/\mu\|_F^2. \tag{10}$$

To solve the problem (10), the spectral soft-thresholding operation [6, 4] is considered as a shrinkage operation on the singular values and is defined as follows:

$$X = \text{prox}_{1/\mu}(T) := U\text{diag}(\max\{\sigma - \frac{1}{\mu}, 0\})V^T, \tag{11}$$

Algorithm 2 Solving RMC problem (6) via ADAL

Input: Given data $\mathcal{P}_{\Omega^C}(Z)$ and λ.
Initialize: $X_0 = E_0 = Y_0 = \mathbf{0}$, $\mu_0 = 10^{-4}$, $\mu_{max} = 10^{10}$, $\rho = 1.10$, and tol.
  while not converged **do**
    Update X_{k+1} by (11).
    Update E_{k+1} by (13) and (14).
    Update the multiplier Y_{k+1} by
      $Y_{k+1} = Y_k + \mu_k(Z - X_{k+1} - E_{k+1})$.
    Update the parameter μ_{k+1} by
      $\mu_{k+1} = \min(\rho\mu_k, \mu_{max})$.
    Check the convergence condition,
      $\|Z - X_{k+1} - E_{k+1}\|_F < \text{tol}$.
  end while
Output: X_k and E_k, where $(E_k)_{\Omega^C}$ is set to $\mathbf{0}$.

where $T := Z - E + Y/\mu$, $\max\{\cdot,\cdot\}$ should be understood element-wise, $U \in \mathbb{R}^{m\times r}$, $V \in \mathbb{R}^{n\times r}$, and $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_r)^T \in \mathbb{R}^{r\times 1}$ are obtained by SVD of T, i.e., $T = U\text{diag}(\sigma)V^T$.

4.2.2 Updating E

The optimal E with all other variables fixed is the solution to the following problem,

$$\min_E \lambda\|\mathcal{P}_\Omega(E)\|_1 + \frac{\mu}{2}\|E + X - Z - Y/\mu\|_F^2. \tag{12}$$

To solve the problem (12), we introduce the following well-known shrinkage (soft-thresholding) operator [8]:

$$\mathcal{S}_\tau(M_{ij}) := \begin{cases} M_{ij} - \tau, & M_{ij} > \tau, \\ M_{ij} + \tau, & M_{ij} < -\tau, \\ 0, & \text{otherwise.} \end{cases}$$

According to the soft-thresholding operator $\mathcal{S}_\tau$, the closed-form solution E_Ω to the problem (12) is given by

$$E_\Omega = \mathcal{S}_{\lambda/\mu}(Z - X + Y/\mu)_\Omega. \tag{13}$$

We can easily obtain the closed-form solution by zeroing the gradient of (12) with respect to E_{Ω^C}, i.e.,

$$E_{\Omega^C} = (Z - X + Y/\mu)_{\Omega^C}. \tag{14}$$

We can replace the l_1-norm loss function in the sparse component learning problem (12) with the $l_{2,1}$-norm loss function for a sparse solution, such as outlier pursuit [31] or low-rank representation [18] problems. The optimal solution to the problem with $l_{2,1}$ regularization can be obtained by the soft-thresholding operator in [32].

Based on the description above, we develop an ADAL iterative algorithm for solving the RMC problem (6), as outlined in **Algorithm 2**. In addition, E_{Ω^C} should be set to $\mathbf{0}$ for the expected output E. Moreover, an $O(1/k)$ convergence rate of Algorithm 2 can be established following the conclusion in [11]. A fixed μ is commonly used. But there are some schemes of varying the penalty parameter to achieve better convergence. This algorithm can also be accelerated by adaptively changing μ. An efficient strategy [17] is to let $\mu = \mu_0$ (the initialization in Algorithm 2) and increase μ_k iteratively by $\mu_{k+1} = \rho\mu_k$, where $\rho \in (1.0, 1.1]$ in general and μ_0 is a small constant.

Algorithm 2 can be easily applied to solve the RPCA problem (3), where all entries of the corrupted matrix are directly observed. Although we also use the PROPACK package [14] to compute a partial SVD as in [25, 4, 17], Algorithm 2 employs the SVD for the spectral soft-thresholding operation, and becomes slow or even impractical for large-scale problems. Therefore, we further propose an efficient ADAL algorithm for solving the non-convex problem (7) in Section 4.3. In addition, several researchers [13, 28] have provided some matrix rank estimation strategies to compute a good value r for the rank of the involved matrix. Thus, we only set a relatively large integer d such that $d \geq r$.

4.3 Non-Convex ADAL Scheme

Our non-convex RMC problem (7) can also be solved by ADAL. The augmented Lagrangian of (7) is given by

$$\begin{aligned}\mathcal{L}_\mu(G, H, E, Y) &= \|H\|_* + \lambda\|\mathcal{P}_\Omega(E)\|_1 \\ &+\langle Y, Z - GH^T - E\rangle + \frac{\mu}{2}\|Z - GH^T - E\|_F^2.\end{aligned} \tag{15}$$

We will derive our scheme for solving the following subproblems with respect to G, H and E, respectively,

$$\begin{aligned} G^* = &\arg\min_{G\in\mathbb{R}^{m\times d}} \mathcal{L}_\mu(G, H, E, Y), \\ &\text{s.t., } G^TG = I, \end{aligned} \tag{16}$$

$$H^* = \arg\min_{H\in\mathbb{R}^{n\times d}} \mathcal{L}_\mu(G^*, H, E, Y), \tag{17}$$

$$E^* = \arg\min_{E\in\mathbb{R}^{m\times n}} \mathcal{L}_\mu(G^*, H^*, E, Y). \tag{18}$$

4.3.1 Updating G

By fixing H and E at their latest values, and removing the terms that do not depend on G and adding some proper terms that do not depend on G, the optimization problem (16) with respect to G is reformulated as follows:

$$\begin{aligned} &\min_G \|GH^T - P\|_F^2, \\ &\text{s.t., } G^TG = I, \end{aligned} \tag{19}$$

where $P := Z - E + Y/\mu$. This is actually the well-known orthogonal procrustes problem [22], the optimal solution can be given by the SVD of the matrix PH, i.e.,

$$G^* = \hat{U}\hat{V}^T, \tag{20}$$

where $\hat{U}$ and $\hat{V}$ are given by the SVD of PH, i.e., $PH = \hat{U}\hat{S}\hat{V}^T$.

4.3.2 Updating H

By fixing G and E, the optimization problem (17) with respect to H can be rewritten as:

$$\min_H \frac{\mu}{2}\|G^*H^T - P\|_F^2 + \|H\|_*. \tag{21}$$

According to Theorem 2.1 in [4], the closed-form solution to the problem (21) is given by the following theorem.

THEOREM 2. *The trace norm minimization problem (21) has a closed-form solution given by:*

$$H^* = \text{prox}_{1/\mu}(P^TG^*). \tag{22}$$

PROOF. The first-order necessary and sufficient optimality condition for the convex problem (21) is given by

$$0 \in \partial\|H\|_* + \mu(G^*H^T - P)^TG^*,$$

Algorithm 3 Solving RMCMF problem (7) via ADAL

Input: Given data $\mathcal{P}_{\Omega^C}(Z)$ and λ.

Initialize: $G_0 = \text{eye}(m,d)$, $H_0 = \mathbf{0}$, $E_0 = Y_0 = \mathbf{0}$, $\mu_0 = 10^{-4}$, $\mu_{max} = 10^{10}$, $\rho = 1.10$, and tol.

while not converged **do**

Update G_{k+1} by (20).

Update H_{k+1} by (22).

Update E_{k+1} by (26) and (27).

Update the multiplier Y_{k+1} by $Y_{k+1} = Y_k + \mu_k(Z - G_{k+1}H_{k+1}^T - E_{k+1})$.

Update the parameter μ_{k+1} by $\mu_{k+1} = \min(\rho\mu_k, \mu_{max})$.

Check the convergence condition, $\|Z - G_{k+1}H_{k+1}^T - E_{k+1}\|_F < \text{tol}$.

end while

Output: G_k, H_k and E_k, where $(E_k)_{\Omega^C}$ is set to $\mathbf{0}$.

where $\partial\|H\|_*$ denotes the set of subgradients of the trace norm (optimality conditions for trace norm are given in [4]). Since $(G^*)^T G^* = I$, the optimality condition for the problem (21) is rewritten as follows:

$$0 \in \partial\|H\|_* + \mu(H - P^T G^*). \quad (23)$$

(23) is also the first-order optimality condition for the following problem,

$$\min_H \frac{\mu}{2}\|H - P^T G^*\|_F^2 + \|H\|_*. \quad (24)$$

According to Theorem 2.1 in [4], the optimal solution of the problem (24) is given by (22). □

4.3.3 Updating E

By fixing all other variables, the optimal E is the solution to the following problem:

$$\min_E \lambda\|\mathcal{P}_\Omega(E)\|_1 + \frac{\mu}{2}\|E + G^*(H^*)^T - Z - Y/\mu\|_F^2. \quad (25)$$

The updating steps for E are very similar to (13) and (14), where X is replaced by $G^*(H^*)^T$ as follows:

$$E_\Omega^* = \mathcal{S}_{\lambda/\mu}(Z - G^*(H^*)^T + Y/\mu)_\Omega, \quad (26)$$

and

$$E_{\Omega^C}^* = (Z - G^*(H^*)^T + Y/\mu)_{\Omega^C}. \quad (27)$$

Following the above analysis, we obtain an ADAL algorithm to solve the matrix factorization based RMC (RMCMF) problem (7), as outlined in **Algorithm 3**. In addition, E_{Ω^C} should be set to $\mathbf{0}$ for the output E. Algorithm 3 can also be easily applied to solve the RPCA problem (3).

5. ALGORITHM ANALYSIS

We now provide convergence analysis and complexity analysis for our algorithms.

5.1 Convergence Analysis

The convergence of ADAL to solve the standard form (8) was studied in [9, 2]. We establish the convergence of Algorithm 2 by transforming the RMC problem (6) into a standard form (8), and show that the transformed problem satisfies the condition needed to establish the convergence. In Algorithm 2, we state that our algorithm alternates between two blocks of variables, X and E. Let x denote the vectorization of X, i.e., $\text{x} = \text{vec}(X) \in \mathbb{R}^{mn\times 1}$, $\text{e} = \text{vec}(E) \in \mathbb{R}^{mn\times 1}$ and $\text{z} = \text{vec}(Z) \in \mathbb{R}^{mn\times 1}$, and $f(X) := \|X\|_*$ and $g(E) := \lambda\|\mathcal{P}_\Omega(E)\|_1$. We can write the equivalence constraint in (6) as the following form:

$$A\text{x} - B\text{e} = \text{z},$$

where both $A \in \mathbb{R}^{mn\times mn}$ and $B \in \mathbb{R}^{mn\times mn}$ are the identity matrices. By the definition $f(X)$ and $g(E)$, it is easy to verify that the problem (6) and Algorithm 2 satisfy the conditions in Algorithm 1. Hence, the convergence of Algorithm 2 is given as follows:

THEOREM 3. *Consider the RMC problem (6), where both $f(X)$ and $g(E)$ are convex functions, and A and B are both identity matrices and have full column rank. The sequence $\{X_k, E_k\}$ generated by Algorithm 2 converges to an optimal solution $\{X^*, E^*\}$ of the problem (6).*

Hence, the sequence $\{X_k, E_k\}$ converges to an optimal solution to the RMC problem (4), where $(E_k)_{\Omega^C}=\mathbf{0}$. Moreover, the convergence of our derived Algorithm 3 for the non-convex problem (7) is guaranteed, as shown in the following theorem.

THEOREM 4. *Let (G_k, H_k, E_k) be a sequence generated by Algorithm 3, then we have the following conclusions:*

1. *(G_k, H_k, E_k) approaches to a feasible solution, i.e., $lim_{k\to\infty}\|Z - G_k H_k^T - E_k\|_F = 0$.*
2. *Both sequences $G_k H_k^T$ and E_k are Cauchy sequences.*
3. *(G_k, H_k, E_k) converges to a KKT point of the problem (7).*

The proof of this theorem can be found in APPENDIX.

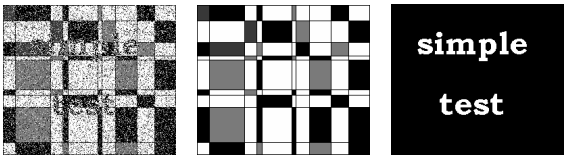
5.2 Complexity Analysis

For the convex problem (6), the running time of Algorithm 2 is dominated by that of performing SVD on the matrix of size $m \times n$. For the non-convex problem (7), Algorithm 3 performs SVD on much smaller matrices of sizes $m \times d$ and $d \times n$, and some matrix multiplications in (22). Hence, the total time complexity of Algorithm 2 and Algorithm 3 are $O(tmn^2)$ and $O(t(d^2m + mnd))$ $(d \ll n \leq m)$, respectively, where t is the number of iterations.

5.3 Connections to Existing Approaches

Our non-convex method is the scalable version of our convex method for both RPCA and RMC problems. In addition, the computational complexity of existing convex algorithms is at least $O(mn^2)$. It means that common RPCA (e.g., PCP [5]) and RMC (e.g., SpaRCS [27]) methods cannot handle large-scale problems, while our non-convex method has a complexity practically linear to the input matrix size and scales well to handle large-scale problems.

From the problems (3) and (6), we can see that in essence our convex method for RPCA problems is equivalent to IALM [17]. The models in [10] and [20] are the special cases of our model (7) when $\lambda \to \infty$. Moreover, the models in [33] and [3] focus only on the desired low-rank matrix. In this sense, them can be viewed as the special cases of our model (7). From the above complexity analysis, both schemes have the same theoretical computational complexity. However, from the experimental results in the next section, we can



(a) (b) (c)

Figure 1: Image used in the text removal experiment: (a) Input image; (b) Original image; (c) Outlier mask.

see that our non-convex method usually runs much faster than the methods in [33] and [3]. The following bilinear regularized matrix factorization formulation in [3] is one of the most similar model to our model (7),

$$\min_{G,H} \lambda f(Z - GH^T) + \frac{1}{2}(\|G\|_F^2 + \|H\|_F^2). \qquad (28)$$

6. EXPERIMENTAL RESULTS

In this section, we evaluate both the effectiveness and efficiency of our RMC and RMCMF algorithms for solving RMC problems such as text removal, face reconstruction and background modeling. All experiments were performed using Matlab 7.11 on an Intel(R) Core (TM) i5-4570 (3.20 GHz) PC running Windows 7 with 8GB main memory.

6.1 Text Removal

We first conduct an experiment by considering a simulated image processing task on artificially generated data, and the goal is to remove some generated text from an image. The ground-truth image is of size 256×222 with rank equal to 10 for the data matrix. We then add to the image a short phase in text form which plays the role of outliers. Fig. 1 shows the image together with the clean image and outliers mask. For fairness, we set the rank of all the algorithms to 20, which is two times the true rank of the underlying matrix. The input data are generated by setting 30% of the randomly selected pixels of the image as missing entries. We compare our two methods, RMC and RMCMF, with the state-of-the-art methods, PCP [5], SpaRCS[1] [27], RegL1[2] [33] and BF-ALM [3]. We set the regularization parameter $\lambda = 1/\sqrt{\max(m,n)}$ and the stopping tolerance tol $= 10^{-4}$ for all algorithms in this experiment.

The results obtained by different methods are visually shown in Fig. 2, where the outlier detection accuracy (the score Area Under the receiver operating characteristic Curve, AUC) and the error of low-rank component recovery (i.e., $\|X - Z\|_F / \|Z\|_F$, where Z and X denote the ground-truth image matrix and the recovered image matrix, respectively) are also presented. As far as low-rank matrix recovery is concerned, the five RMC methods including SpaRCS, RegL1, BF-ALM, RMC and RMCMF, outperform PCP, not only visually but also quantitatively. For outlier detection, it can be seen that our methods RMC and RMCMF perform better than the other methods. In short, RMC and RMCMF significantly outperform PCP, SpaRCS, RegL1 and BF-ALM in terms of both low-rank matrix recovery and spare outlier identification. Moreover, the running time of PCP, SpaRCS, RegL1, BF-ALM, RMC and RMCMF is 15.39sec, 5.74sec, 3.86sec, 2.62sec, 1.10sec and 0.23sec, respectively.

[1] http://www.ece.rice.edu/~aew2/sparcs.html
[2] https://sites.google.com/site/yinqiangzheng/

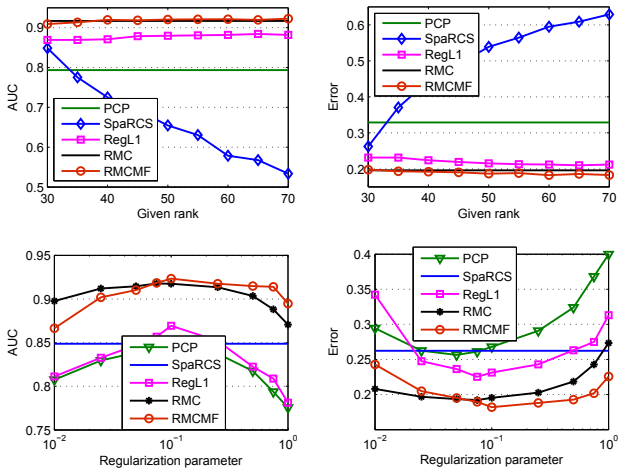


Figure 3: Performance of PCP, SpaRCS, RegL1, RMC and RMCMF in terms of AUC (left) and Error (right) with varying ranks (the first row) or regularization parameters (the second row).

Moreover, we further evaluate the robustness of our methods, RMC and RMCMF, with respect to the regularization parameter λ and the given rank changes. We conduct some experiments on the artificially generated data, and illustrate the outlier detection accuracy (AUC) and the error (Error) of low-rank component recovery, where the rank parameter of our RMCMF method, SpaRCS and RegL1 is chosen from $\{30, 35, \cdots, 70\}$, and the regularization parameter λ of RMC, RMCMF, PCP and RegL1 is chosen from the grid $\{0.01, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75, 1\}$. Notice that we only report the results of RegL1 due to the similar performance of RegL1 and BF-ALM. The average AUC and Error results of 10 independent runs are shown in Fig. 3, from which we can see that our RMCMF method performs much more robust than SpaRCS and RegL1 in terms of AUC and Error with respect to the given rank. Furthermore, our RMC and RMCMF methods are much more robust than PCP and RegL1 in terms of AUC and Error against the regularization parameter.

6.2 Face Reconstruction

We also test our methods for the face reconstruction problems. The face database used in this experiment is a part of Extended Yale Face Database B [15] with large corruptions. The part of Extended Yale-B consists of 320 frontal face images of the first 5 classes, and each subset contains 64 images with varying illumination conditions and heavily "shadows" as outliers. The resolution of all images is 192×168 and the pixel values are normalized to [0, 1], then the pixel values are used to form data vectors of dimension 32,256. The input data are generated by setting 40% of the randomly selected pixels of each image as missing entries.

Fig. 4 shows some original and reconstructed images by RegL1 and CWM[3] [20], and our methods, RMC and RMCMF, where the average computational time (in seconds)

[3] http://www4.comp.polyu.edu.hk/~cslzhang/papers.htm

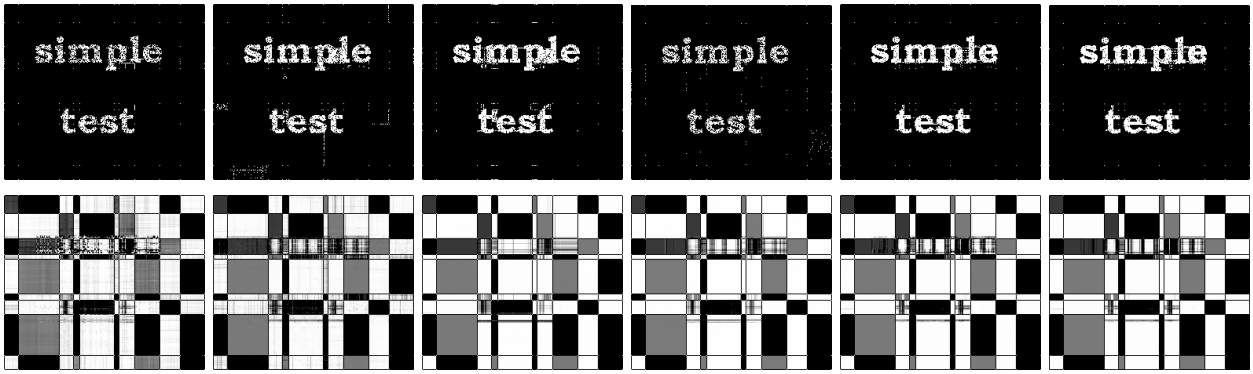


Figure 2: Text removal results of different methods, where the first row shows the recovered foreground masks and the second row shows the recovered background images: (a) PCP (AUC: 0.7934; Error: 0.3290); (b) SpaRCS (AUC: 0.8487; Error: 0.2623); (c) RegL1 (AUC: 0.8792; Error: 0.2291); (d) BF-ALM (AUC: 0.8568; Error: 0.2435); (e) RMC (AUC: 0.9206; Error: 0.1987); (f) RMCMF (AUC: 0.9197; Error: 0.1996).

Figure 4: Face reconstruction results. From left column to right column: input corrupted images (black pixels denote missing entries), original images, reconstruction results by CWM (1554.91sec), RegL1 (2710.68sec), RMC (54.19sec) and RMCMF (21.18sec).

of all these algorithms on each people's faces is presented. It can be observed that RMC and RMCMF perform much better than the other methods visually, as they effectively eliminate the heavy noise and "shadows" and simultaneously complete the missing entries. In other words, our RMC and RMCMF methods can achieve the latent features underlying the original images even though the observed data is corrupted by both outliers and missing values. And impressively, both RMC and RMCMF are also significantly faster than RegL1 and CWM.

6.3 Background Modeling

In this experiment we test our methods on real surveillance videos for object detection and background subtraction as a robust matrix completion problem. Background modeling is a crucial task for motion segmentation in surveillance videos. A video sequence satisfies the low-rank and sparse structures, because the background of all the frames is controlled by few factors and hence exhibits low-rank property, and the foreground is detected by identifying spatially localized sparse residuals [30, 5]. We test our methods on real surveillance videos for object detection and background subtraction on four color surveillance videos: Bootstrap, Lobby, Hall and Mall databases[4]. The data matrix Z consists of the first 400 frames of size 120×160. Since all the original videos have colors, we first reshape every frame of the video into a long column vector and then collect all the columns into a data matrix Z with size of 57600×400. Moreover, the input data is generated by setting 10% of the randomly selected pixels of each frame as missing entries.

Figs. 5 and 6 illustrate the foreground and background separation results on the Bootstrap and Mall databases, where the first and fourth columns represent the incomplete input images, the second and fifth columns show the low-rank recoveries, and the third and sixth columns show the sparse components. It is clear that the background can be effectively extracted by RMC, RMCMF, BF-ALM, and GRASTA[5] [12]. Notice that SpaRCS [27] could not yield experimental results on these databases because it ran out of memory. We can see that the decomposition results of RMCMF and RMC are slightly better than that of GRASTA and BF-ALM. As pointed out in [23], the theoretical reason for the unsatisfactory performance of the l_1-penalty is that the irrepresentable condition is not met. Hence, RMCMF incorporating with matrix factorization is more accurate in recovering the low-rank matrix than RMC. Furthermore, we also provide the CPU time consumption of these algorithms on all four databases, as shown in Table 1, from which we can see that RMCMF is more than 7 times faster than RMC, more than 4 times faster than GRASTA, and more than 2 times faster than BF-ALM. This further shows that RMCMF has good scalability and can address large-scale problems.

7. CONCLUSIONS

We proposed a unified RMC framework for RPCA and RMC problems. We first presented two matrix trace norm regularized models that replace the linear projection operator constraint by a simple equality one. Then we de-

[4] `http://perception.i2r.a-star.edu.sg/bkmodel/bkindex`
[5] `https://sites.google.com/site/hejunzz/grasta`

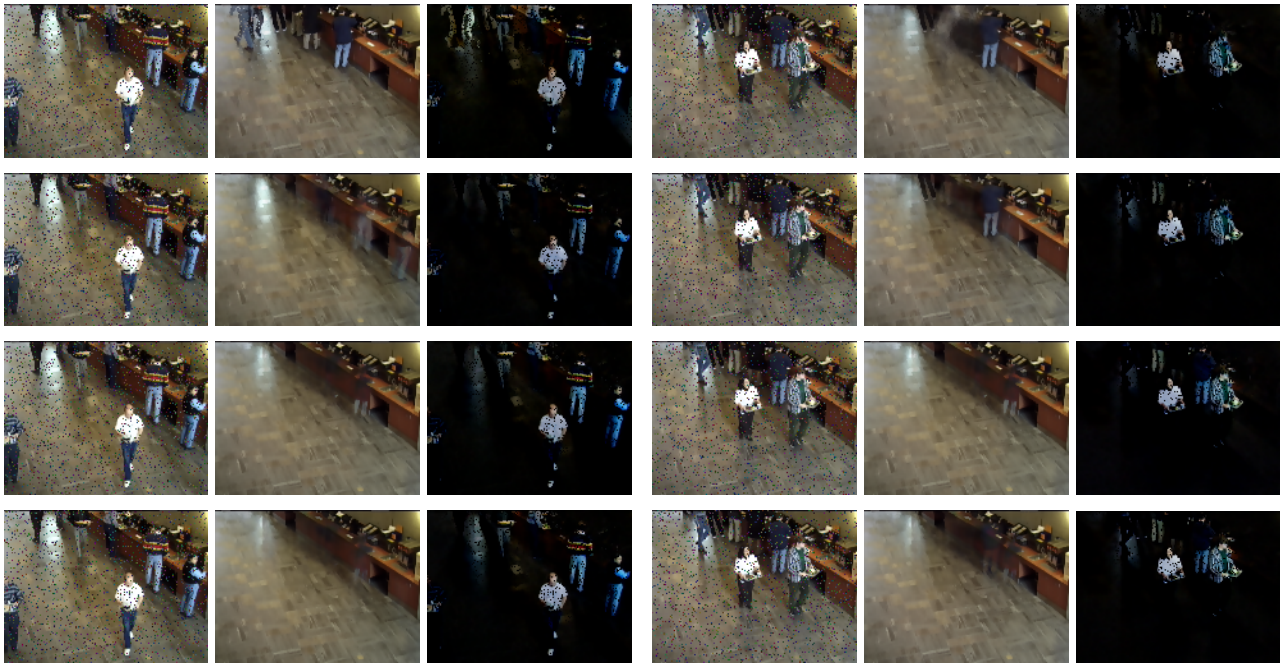


Figure 5: Foreground and background separation results of different algorithms on the Bootstrap data set, where the first, second, third and last rows show the recovered low-rank and sparse images by GRASTA, BF-ALM, RMC and RMCMF, respectively.

Table 1: Comparison of time costs in CPU seconds of GRASTA, BF-ALM, RMC and RMCMF on background modeling datasets.

Datasets	Sizes	GRASTA	BF-ALM	RMC	RMCMF
Bootstrap	$57,600 \times 400$	153.65	93.17	344.13	38.32
Lobby	$61,440 \times 400$	187.43	139.83	390.78	50.08
Hall	$76,032 \times 400$	315.11	153.45	461.48	67.73
Mall	$245,760 \times 200$	493.92	–	–	94.59

veloped two efficient ADAL algorithms to solve our convex and non-convex low-rank and sparse matrix decomposition problems. Finally, we analyzed the convergence of our algorithms. Experimental results on synthetic and real-world data sets demonstrated the superior performance of our methods compared with the state-of-the-art methods in terms of both efficiency and effectiveness.

Both our algorithms are essentially the Gauss-Seidel schemes of ADAL, and their Jacobi-type update schemes can be easily implemented in parallel. Hence, our algorithms are well suited for parallel and distributed computing and are particularly attractive for solving certain large-scale problems. Moreover, our methods can easily extended to the general nonconvex low-rank inducing penalty problem [26]. For future work, we will consider the compressing RMC (also called compressive principal component pursuit) problem with the general linear operator as in [29].

8. ACKNOWLEDGMENTS

We thank the reviewers for giving us many constructive comments, with which we have significantly improved our paper. This research is supported in part by SHIAE Grant No. 8115048, MSRA Grant No. 6903555, GRF No. 411211, and CUHK direct grant Nos. 4055015 and 4055017.

9. REFERENCES

[1] E. Acar, D. Dunlavy, T. Kolda, and M. Mørup. Scalable tensor factorizations with missing data. In *SDM*, pages 701–711, 2010.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, 2011.

[3] R. Cabral, F. Torre, J. Costeira, and A. Bernardino. Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition. In *ICCV*, pages 2488–2495, 2013.

[4] J. Cai, E. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.*, 20(4):1956–1982, 2010.

[5] E. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *J. ACM*, 58(3):1–37, 2011.

[6] E. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.

[7] Y. Chen, A. Jalali, S. Sanghavi, and C. Caramanis. Low-rank matrix recovery from errors and erasures. *IEEE Trans. Inform. Theory*, 59(7):4324–4337, 2013.

[8] I. Daubechies, M. Defrise, and C. DeMol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pur. Appl. Math.*, 57(11):1413–1457, 2004.

[9] J. Eckstein and D. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Prog.*, 55(1):293–318, 1992.

Figure 6: Foreground and background separation results of different algorithms on the Mall data set, where the first and second rows show the recovered low-rank and sparse images by GRASTA and RMCMF, respectively.

[10] A. Eriksson and A. van den Hengel. Efficient computation of robust low-rank matrix approximations in the presence of missing data using the l1 norm. In *CVPR*, pages 771–778, 2010.

[11] B. He and X. Yuan. On the O(1/n) convergence rate of the Douglas-Rachford alternating direction method. *SIAM J. Numer. Anal.*, 50(2):700–709, 2012.

[12] J. He, L. Balzano, and A. Szlam. Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video. In *CVPR*, pages 1568–1575, 2012.

[13] R. Keshavan, A. Montanari, and S. Oh. Matrix completion from a few entries. *IEEE Trans. Inform. Theory*, 56(6):2980–2998, 2010.

[14] R. Larsen. Propack-software for large and sparse svd calculations. 2005.

[15] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(5):684–698, 2005.

[16] X. Li. Compressed sensing and matrix completion with constant proportion of corruptions. *Constr. Approx.*, 37:73–99, 2013.

[17] Z. Lin, R. Liu, and Z. Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *NIPS*, pages 612–620, 2011.

[18] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1):171–184, 2013.

[19] S. Ma, D. Goldfarb, and L. Chen. Fixed point and Bregman iterative methods for matrix rank minimization. *Math. Prog.*, 128(1):321–353, 2011.

[20] D. Meng, Z. Xu, L. Zhang, and J. Zhao. A cyclic weighted median method for L1 low-rank matrix factorization with missing entries. In *AAAI*, 2013.

[21] K. Min, Z. Zhang, J. Wright, and Y. Ma. Decomposition background topics from keywords by principal component pursuit. In *CIKM*, pages 269–278, 2010.

[22] H. Nick. Matrix procrustes problems. 1995.

[23] Y. She and A. Owen. Outlier detection using nonconvex penalized regression. *J. Am. Stat. Assoc.*, 106(494):626–639, 2011.

[24] M. Tao and X. Yuan. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM J. Optim.*, 21(1):57–81, 2011.

[25] K.-C. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *Pac. J. Optim.*, 6:615–640, 2010.

[26] S. Wang, D. Liu, and Z. Zhang. Nonconvex relaxation approaches to robust matrix recovery. In *IJCAI*, pages 1764–1770, 2013.

[27] A. Waters, A. Sankaranarayanan, and R. Baraniuk. SpaRCS: Recovering low-rank and sparse matrices from compressive measurements. In *NIPS*, pages 1089–1097, 2011.

[28] Z. Wen, W. Yin, and Y. Zhang. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Math. Prog. Comp.*, 4(4):333–361, 2012.

[29] J. Wright, A. Ganesh, K. Min, and Y. Ma. Compressive principal component pursuit. *Inform. Infer.*, 2:32–68, 2013.

[30] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization. In *NIPS*, pages 2080–2088, 2009.

[31] H. Xu, C. Caramanis, and S. Sanghavi. Robust PCA via outlier pursuit. In *NIPS*, pages 2496–2504, 2010.

[32] J. Yang, W. Yin, Y. Zhang, and Y. Wang. A fast algorithm for edge-preserving variational multichannel image restoration. *SIAM J. Imag. Sci.*, 2(2):569–592, 2009.

[33] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi. Practical low-rank matrix approximation under robust L1-norm. In *CVPR*, pages 1410–1417, 2012.

[34] T. Zhou and D. Tao. GoDec: Randomized low-rank & sparse matrix decomposition in noisy case. In *ICML*, pages 33–40, 2011.

APPENDIX

We first prove that the boundedness of the multiplier and some variables of Algorithm 3, and then analyze the conver-

gence of Algorithm 3. To prove the boundedness, we first give the following lemmas.

LEMMA 3. *Let $\mathcal{X}$ be a real Hilbert space endowed with an inner product $\langle\cdot\rangle$ and a corresponding norm $\|\cdot\|$ (the trace norm or the l_1 norm), and $y \in \partial\|x\|$, where $\partial\|\cdot\|$ denotes the subgradient. Then $\|y\|^* = 1$ if $x \neq 0$, and $\|y\|^* \leq 1$ if $x = 0$, where $\|\cdot\|^*$ is the dual norm of the norm $\|\cdot\|$.*

LEMMA 4. *Let $Y_{k+1} = Y_k + \mu_k(Z - G_{k+1}H_{k+1}^T - E_{k+1})$, $\hat{Y}_{k+1} = Y_k + \mu_k(Z - G_{k+1}H_{k+1}^T - E_k)$, and $\widetilde{Y}_{k+1} = Y_k + \mu_k(Z - G_{k+1}H_k^T - E_k)$, where G_{k+1} is the optimal solution of the problem (19). Then the sequences $\{Y_k\}$, $\{\hat{Y}_k\}$, $\{\widetilde{Y}_k\}$, $\{H_k\}$ and $\{E_k\}$ produced by Algorithm 3 are all bounded.*

PROOF. By the first-order optimality condition of the problem (18) with respect to E_{k+1}, we have

$$0 \in \partial_{(E_{k+1})_\Omega}\mathcal{L}_{\mu_k}(G_{k+1}, H_{k+1}, E_{k+1}, Y_k),$$

and $(Y_k + \mu_k(Z - G_{k+1}H_{k+1}^T - E_{k+1}))_\Omega \in \lambda\partial\|(E_{k+1})_\Omega\|_1$, i.e., $(Y_{k+1})_\Omega \in \lambda\partial\|(E_{k+1})_\Omega\|_1$.

Furthermore, by substituting $Y_{k+1} = Y_k + \mu_k(Z - G_{k+1}H_{k+1}^T - E_{k+1})$ into (18), we have $(Y_{k+1})_{\Omega^C} = 0$.

By Lemma 4, we have

$$\|Y_{k+1}\|_\infty = \|(Y_{k+1})_{\Omega^C} + (Y_{k+1})_\Omega\|_\infty \leq \lambda,$$

where $\|\cdot\|_\infty$ denotes the matrix l_∞-norm, i.e., $\|M\|_\infty = \max_{i,j}|M_{i,j}|$. Thus, the sequence $\{Y_k\}$ is bounded.

By the iteration procedure of Algorithm 3, we have

$$\begin{aligned}&\mathcal{L}_{\mu_k}(G_{k+1}, H_{k+1}, E_{k+1}, Y_k) \leq \mathcal{L}_{\mu_k}(G_{k+1}, H_{k+1}, E_k, Y_k)\\ \leq&\mathcal{L}_{\mu_k}(G_{k+1}, H_k, E_k, Y_k) \leq \mathcal{L}_{\mu_k}(G_k, H_k, E_k, Y_k)\\ =&\mathcal{L}_{\mu_{k-1}}(G_k, H_k, E_k, Y_{k-1}) + \beta_k\|Y_k - Y_{k-1}\|_F^2,\end{aligned}$$

where $\beta_k = \frac{1}{2}\mu_{k-1}^{-2}(\mu_{k-1} + \mu_k)$ and $\mu_k = \rho\mu_{k-1}$. Since

$$\begin{aligned}&\sum_{k=1}^{\infty}\frac{1}{2}\mu_{k-1}^{-2}(\mu_{k-1} + \mu_k) = \frac{\rho(\rho+1)}{2\mu_0}\sum_{k=1}^{\infty}\frac{1}{\rho^k}\\ =&\frac{\rho(\rho+1)}{2\mu_0(\rho-1)} < \infty,\end{aligned}$$

we have that $\{\mathcal{L}_{\mu_k}(G_{k+1}, H_{k+1}, E_{k+1}, Y_k)\}$ is upper bounded due to the boundedness of $\{Y_k\}$. Then

$$\begin{aligned}&\|H_k\|_* + \lambda\|\mathcal{P}_\Omega(E_k)\|_1\\ =&\mathcal{L}_{\mu_{k-1}}(G_k, H_k, E_k, Y_{k-1}) - \langle Y_{k-1}, Z - G_kH_k^T - E_k\rangle\\ &- \frac{\mu_{k-1}}{2}\|Z - G_kH_k^T - E_k\|_F^2,\\ =&\mathcal{L}_{\mu_{k-1}}(G_k, H_k, E_k, Y_{k-1}) - \frac{1}{2\mu_{k-1}}(\|Y_k\|_F^2 - \|Y_{k-1}\|_F^2),\end{aligned}$$

is upper bounded, i.e., $\{H_k\}$ and $\{E_k\}$ are bounded. Since $\|G_kH_k^T\|_* = \|H_k\|_*$, $\{G_kH_k^T\}$ is also bounded.

We next prove that $\{\widetilde{Y}_k\}$ is bounded. Since G_{k+1} is the optimal solution of the problem (19), then we have

$$\begin{aligned}&\|Y_k + \mu_k(Z - G_{k+1}H_k^T - E_k)\|_F^2\\ \leq&\|Y_k + \mu_k(Z - G_kH_k^T - E_k)\|_F^2.\end{aligned}$$

By the definition of $\widetilde{Y}_{k+1}$, and $\mu_{k+1} = \rho\mu_k$, thus,

$$\|\widetilde{Y}_{k+1}\|_F^2 \leq \|(1+\rho)Y_k - \rho Y_{k-1}\|_F^2.$$

By the boundedness of H_k and Y_k, then the sequence $\{\widetilde{Y}_k\}$ is bounded.

The first-order optimality condition of the problem (21) with respect to H_{k+1} is rewritten as follows:

$$G_{k+1}^T\hat{Y}_{k+1} \in \partial\|H_{k+1}^T\|_*.$$

According to Lemma 4, we have $\|G_{k+1}^T\hat{Y}_{k+1}\|_2 \leq 1$, where $\|M\|_2$ denotes the spectral norm of M and is equal to the maximum singular value of M. Thus, $G_{k+1}^T\hat{Y}_{k+1}$ is bounded. Let $G_{k+1}^\perp$ denote the orthogonal complement of G_{k+1}, i.e., $G_{k+1}^\perp G_{k+1} = 0$, then we have

$$\begin{aligned}&(G_{k+1}^\perp)^T\hat{Y}_{k+1}\\ =&(G_{k+1}^\perp)^T(Y_k + \mu_k(Z - G_{k+1}H_{k+1}^T - E_k)),\\ =&(G_{k+1}^\perp)^T(Y_k + \mu_k(Z - E_k)),\\ =&(G_{k+1}^\perp)^T(Y_k + \mu_k(Z - G_{k+1}H_k^T - E_k)),\\ =&(G_{k+1}^\perp)^T\widetilde{Y}_{k+1}.\end{aligned}$$

Thus, $\{(G_{k+1}^\perp)^T\hat{Y}_{k+1}\}$ is bounded due to the boundedness of $\{\widetilde{Y}_k\}$. Then we have

$$\begin{aligned}&\|\hat{Y}_{k+1}\|_2 = \|G_{k+1}^T\hat{Y}_{k+1} + (G_{k+1}^\perp)^T\hat{Y}_{k+1}\|_2\\ \leq&\|G_{k+1}^T\hat{Y}_{k+1}\|_2 + \|(G_{k+1}^\perp)^T\hat{Y}_{k+1}\|_2.\end{aligned}$$

Since both $G_{k+1}^T\hat{Y}_{k+1}$ and $(G_{k+1}^\perp)^T\hat{Y}_{k+1}$ are bounded, the sequence $\{\hat{Y}_k\}$ is bounded. □

Proof of Theorem 4:

PROOF. **1.** By $Z - G_{k+1}H_{k+1}^T - E_{k+1} = \mu_k^{-1}(Y_{k+1} - Y_k)$, the boundedness of $\{Y_k\}$ and $\lim_{k\to\infty}\mu_k = \infty$, we have

$$\lim_{k\to\infty} Z - G_{k+1}H_{k+1}^T - E_{k+1} = 0.$$

Thus, (G_k, H_k, E_k) approaches to a feasible solution.

2. We prove that the sequences $\{E_k\}$ and $\{G_kH_k^T\}$ are Cauchy sequences.

By $\|E_{k+1} - E_k\| = \mu_k^{-1}\|Y_{k+1} - \hat{Y}_{k+1}\| = o(\mu_k^{-1})$ and

$$\sum_{k=1}^{\infty}\mu_{k-1}^{-1} = \frac{\rho}{\mu_0(\rho-1)} < \infty,$$

thus, $\{E_k\}$ is a Cauchy sequence, and it has a limit E^*.

Similarly, $\{G_kH_k^T\}$ is also a Cauchy sequence, therefore it has a limit $\{G^*(H^*)^T\}$.

3. According to Algorithm 3, the first-order optimality conditions of the problems (25) and (21) at the k-th iteration is formulated as follows:

$$G_k^T(Y_{k-1} + \mu_{k-1}(Z - G_kH_k^T - E_{k-1})) \in \partial\|H_k^T\|_*,$$

and

$$(Y_{k-1} + \mu_{k-1}(Z - G_kH_k^T - E_k))_\Omega \in \lambda\partial\|(E_k)_\Omega\|_1.$$

Since both $\{E_k\}$ and $\{G_kH_k^T\}$ are Cauchy sequences, let E^* and $G^*(H^*)^T$ be limit of $\{E_k\}$ and $\{G_kH_k^T\}$, respectively. By the definition of Y_k, we have

$$(G^*)^TY^* \in \partial\|(H^*)^T\|_*,\ (Y^*)_\Omega \in \lambda\partial\|(E^*)_\Omega\|_1,$$

and (G^*, H^*, E^*) is a feasible solution, i.e.,

$$Z = G^*(H^*)^T + E^*,$$

and $G^*(G^*)^T = I$.

Thus, (G^*, H^*, E^*) is a KKT point of the problem (7). □