

A SMOOTHING PROXIMAL GRADIENT ALGORITHM FOR MATRIX RANK MINIMIZATION PROBLEM

QUAN YU, XINZHEN ZHANG

ABSTRACT. In this paper, we study the low-rank matrix minimization problem, where the loss function is convex but nonsmooth and the penalty term is defined by the cardinality function. We first introduce an exact continuous relaxation, that is, both problems have the same minimizers and the same optimal value. In particular, we introduce a class of lifted stationary point of the relaxed problem and show that any local minimizer of the relaxed problem must be a lifted stationary point. In addition, we derive lower bound property for the nonzero singular values of the lifted stationary point and hence also of the local minimizers of the relaxed problem. Then the smoothing proximal gradient (SPG) algorithm is proposed to find a lifted stationary point of the continuous relaxation model. Moreover, it is shown that the whole sequence generated by SPG algorithm converges to a lifted stationary point. At last, numerical examples show the efficiency of the SPG algorithm.

1. INTRODUCTION

Over the last decade, finding a low-rank matrix solution to a system or low-rank matrix optimization problem have received more and more attention. Numerous optimization models and methods have been proposed in [9, 10, 11, 18, 20, 25]. In this paper, we consider the following matrix rank minimization problem with cardinality penalty, that is,

$$(1.1) \quad \min \mathcal{F}_0(X) := f(X) + \lambda \cdot \text{rank}(X) = f(X) + \lambda \|\sigma(X)\|_0,$$

where $X \in \mathbb{R}^{m \times n}$ ($n \leq m$) and $\sigma(X) := (\sigma_1(X), \dots, \sigma_n(X))^T$ is a vector composed of X 's singular values with $\sigma_1(X) \geq \dots \geq \sigma_n(X) \geq 0$. Furthermore, $f : \mathbb{R}^{m \times n} \rightarrow [0, \infty)$ is convex (not necessarily smooth) and λ is a positive parameter.

One application of problem (1.1) is the low-rank matrix recovery problem [14, 17, 19, 24]. To solve such problem, traditional algorithms are always based on l_2 (or Frobenius)-nuclear model, that is, the loss function is a l_2 -norm for vector case or Frobenius norm for matrix case, and $\text{rank}(X)$ is relaxed as a matrix nuclear norm. However, these models are sensitive to non-Gaussian noise with outliers [12, 30, 29, 28]. To overcome this drawback, the l_1 model is considered in the problem with the outlier-resistant loss function. For example, the following loss

2020 *Mathematics Subject Classification.* 15A03, 15A83, 90C30, 65K05.

Key words and phrases. low-rank approximation, nonsmooth convex loss function, smoothing method.

function is considered in the low-rank matrix recovery problem

$$(1.2) \quad f(X) = \|\mathcal{A}(X) - b\|_1,$$

where the linear map $\mathcal{A} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^p$ and vector $b \in \mathbb{R}^p$ are given. Obviously, f is convex but not smooth. Considering low-rank matrix completion problem, a special case of low-rank matrix recovery problem, the corresponding loss function $f(X)$ can be written as

$$f(X) = \|P_\Omega(X - M)\|_1,$$

where $M \in \mathbb{R}^{m \times n}$ is a known matrix, Ω is an index set which locates the observed data, P_Ω is a linear operator that extracts the entries in Ω and fills the entries not in Ω with zeros. In the robust principal component analysis (RPCA) problem [2, 6, 23, 27], the loss function $f(X)$ is adopted as

$$(1.3) \quad f(X) = \|L - X\|_1,$$

where $L \in \mathbb{R}^{m \times n}$ denotes the observed data. The RPCA problem aims to decompose the matrix L as the sum of a low-rank matrix X and a sparse matrix $E = L - X \in \mathbb{R}^{m \times n}$.

It is known that matrix rank function is nonconvex and nonsmooth. In the matrix rank minimization problem, one of common used convex relaxations of rank function is matrix nuclear norm. Although the methods based nuclear norm relaxation have strong theoretical guarantees, the obtained approximation solutions under certain incoherence assumptions are usually hard to satisfy in real applications [4, 5]. In other words, the nuclear norm is not a perfect approximation to the rank function.

In [1], the capped l_1 function, a continuous relaxation of l_0 function, was adopted in penalized sparse regression problem with some advantages. Furthermore, a smoothing proximal gradient (SPG) algorithm with global convergence was proposed there. More recently, such technique was applied to group sparse optimization for images recovery in [22]. It is well-known that the matrix norm can be expressed as a vector norm of the singular value vector. Motivated by these, we consider whether such SPG algorithm can be generalized from sparse regression problem to low-rank matrix minimization or not.

For this aim, let $\Phi(X) = \sum_{i=1}^n \phi(\sigma_i(X))$ be a continuous relaxation of the rank function with the capped- l_1 function ϕ given by

$$(1.4) \quad \phi(t) = \min\{1, t/\nu\}, \quad t \geq 0,$$

where $\nu > 0$ is a parameter. Based on $\Phi(X)$, we consider the following continuous optimization problem for solving (1.1):

$$(1.5) \quad \min \mathcal{F}(X) := f(X) + \lambda \Phi(X).$$

Our contributions are as follows. We first present the continuous relaxation problem (1.5) of problem (1.1), which are shown to have the same global optimizers. Furthermore, the local minimizer of (1.5) is a lifted stationary point of (1.5) with an expected lower bound property of singular values. Then an SPG algorithm is proposed to get a lifted stationary point of (1.5) with global convergence.

Notations. We denote $[n] = \{1, 2, \dots, n\}$ and $\mathbb{D}^n = \{d \in \mathbb{R}^n : d_i \in \{1, 2\}, i \in [n]\}$. The space of $m \times n$ matrices is denoted by $\mathbb{R}^{m \times n}$. For a given matrix $X \in \mathbb{R}^{m \times n}$,

$\mathbb{B}_\delta(X)$ denotes the open ball centered at X with radius δ . In addition, $\mathcal{D}(x)$ denotes a diagonal matrix generated by vector x , whose dimension shall be clear from the context. Denote \mathbb{Q}^m the set of $m \times m$ -dimension unitary orthogonal matrix. Let $E_i = \mathcal{D}(e_i)$, where e_i is a unit vector whose i th entry is 1.

For any given $X, Y \in \mathbb{R}^{m \times n}$, the standard inner product of X and Y is denoted by $\langle X, Y \rangle$, that is, $\langle X, Y \rangle = \text{tr}(XY^T)$, where $\text{tr}(\cdot)$ denotes the trace of a matrix. The Frobenius norm of X is denoted by $\|X\|_F$, namely, $\|X\|_F = \sqrt{\text{tr}(XX^T)}$. Denote $\sigma(X) = (\sigma_1(X), \dots, \sigma_n(X))^T$ and

$$\mathcal{M}(X) = \{(U, V) \in \mathbb{Q}^m \times \mathbb{Q}^n : X = U\mathcal{D}(\sigma(X))V^T\}.$$

2. AN EXACT CONTINUOUS RELAXATION FOR (1.1)

In this section, we present the relationships between (1.1) and (1.5). Without specific explanation, Assumptions 1 and 2 are assumed throughout the paper.

Assumption 1. f is Lipschitz continuous with Lipschitz constant L_f .

Assumption 2. Positive parameter ν in (1.4) satisfies $\nu < \bar{\nu} := \lambda/L_f$.

2.1. Lifted stationary points of (1.5). Clearly, ϕ in (1.4) can be rewritten as a DC function, i.e.,

$$\phi(t) = \frac{t}{\nu} - \max\{\theta_1(t), \theta_2(t)\}$$

with $\theta_1(t) = 0$ and $\theta_2(t) = t/\nu - 1$. Denote

$$(2.1) \quad \mathcal{D}(t) = \{i \in \{1, 2\} : \theta_i(t) = \max\{\theta_1(t), \theta_2(t)\}\}.$$

Following Theorem 3.7 in [15], the Clarke subdifferential of Φ at X is given by

$$\partial\Phi(X) = \left\{ U\mathcal{D}(x)V^T : x \in \partial \sum_{i=1}^n \phi(\sigma_i(X)), (U, V) \in \mathcal{M}(X) \right\},$$

where $\partial\phi(x)$ is the Clarke subdifferential [8] of $\phi(x)$. Then we have the following definition.

Definition 2.1. *We say that X is a lifted stationary point of (1.5) if there exist $d_i \in \mathcal{D}(\sigma_i(X))$ for all $i \in [n]$ such that*

$$(2.2) \quad \lambda \sum_{i=1}^n \theta'_{d_i}(\sigma_i(X)) E_i \in \left\{ U^T \partial f(X) V + \frac{\lambda}{\nu} \mathcal{D}(\partial \|\sigma(X)\|_1) : (U, V) \in \mathcal{M}(X) \right\},$$

where $\sigma_i(X)$ is the i th largest singular value of X .

If (2.2) holds for all $d_i \in \mathcal{D}(\sigma_i(X))$, $i \in [n]$, then we call X a d -stationary point.

2.2. Characterizations of lifted stationary points of (1.5). We first show that for any element in $\mathcal{D}(\sigma_i(X))$ satisfying (2.2) is unique and well defined.

Proposition 2.2. *If \bar{X} is a lifted stationary point of (1.5), then the vector $d^{\bar{X}} = (d_1^{\bar{X}}, \dots, d_n^{\bar{X}})^T \in \prod_{i=1}^n \mathcal{D}(\sigma_i(\bar{X}))$ satisfying (2.2) is unique. In particular, for $i \in [n]$,*

$$(2.3) \quad d_i^{\bar{X}} = \begin{cases} 1 & \text{if } \sigma_i(\bar{X}) < \nu, \\ 2 & \text{if } \sigma_i(\bar{X}) \geq \nu. \end{cases}$$

Proof. For case of $\sigma_i(\bar{X}) \neq \nu$, the statement in this proposition follows. Hence, it suffices to consider the index i satisfying $\sigma_i(\bar{X}) = \nu$.

Now we assume that $d_i^{\bar{X}} = 1$ by contradiction when $\sigma_i(\bar{X}) = \nu$. By (2.2), there exists $\xi(\bar{X}) \in \partial f(\bar{X})$ such that $0 = (U^T \xi(\bar{X}) V)_{ii} + \lambda/\nu$, where $(U, V) \in \mathcal{M}(\bar{X})$. Then, $\lambda/\nu = |(U^T \xi(\bar{X}) V)_{ii}| \leq \|U^T \xi(\bar{X}) V\|_F = \|\xi(\bar{X})\|_F \leq L_f$. This leads to a contradiction to $\nu < \lambda/L_f$. Then, we can assert that $d_i^{\bar{X}} = 2$, and hence (2.3) holds for $\sigma_i(\bar{X}) = \nu$. \square

For a given $d = (d_1, \dots, d_n)^T \in \mathbb{D}^n$, we define

$$(2.4) \quad \Phi^d(X) := \sum_{i=1}^n \sigma_i(X)/\nu - \sum_{i=1}^n \theta_{d_i}(\sigma_i(X)).$$

It is easy to see that

$$\Phi(X) = \min_{d \in \mathbb{D}^n} \Phi^d(X), \quad \forall X \in \mathbb{R}^{m \times n}.$$

Furthermore, for a fixed \bar{X} , $\Phi(\bar{X}) = \Phi^{d^{\bar{X}}}(\bar{X})$ with $d^{\bar{X}}$ defined in (2.3). We next show that any local minimizer of (1.5) is a lifted stationary point of the problem.

Theorem 2.3. *Suppose that \bar{X} is a local minimizer of problem (1.5). Then \bar{X} is a lifted stationary point of (1.5), that is, (2.2) holds at \bar{X} .*

Proof. Since \bar{X} is a local minimizer of (1.5) satisfying (1.5), it gives

$$(2.5) \quad \begin{aligned} f(\bar{X}) + \lambda \Phi^{d^{\bar{X}}}(\bar{X}) &= f(\bar{X}) + \lambda \Phi(\bar{X}) \\ &\leq f(X) + \lambda \Phi(X) \leq f(X) + \lambda \Phi^{d^{\bar{X}}}(X), \quad \forall X \in \mathbb{B}_\rho(\bar{X}), \end{aligned}$$

where the first equality comes from $\Phi^{d^{\bar{X}}}(\bar{X}) = \Phi(\bar{X})$, $d^{\bar{X}}$ is defined as in (2.3) and the last inequality is due to $\Phi^{d^{\bar{X}}}(X) \geq \Phi(X)$, $\forall X \in \mathbb{R}^{m \times n}$. Then \bar{X} is a local minimizer of the problem

$$(2.6) \quad \min_X f(X) + \lambda \Phi^{d^{\bar{X}}}(X).$$

Hence, there exists some $(\bar{U}, \bar{V}) \in \mathcal{M}(\bar{X})$ such that

$$(2.7) \quad 0 = \partial f(\bar{X}) + \lambda \bar{U} \left(\frac{1}{\nu} \mathcal{D}(\partial \|\sigma(\bar{X})\|_1) - \sum_{i=1}^n \theta'_{d_i^{\bar{X}}}(\sigma_i(\bar{X})) E_i \right) \bar{V}^T,$$

which implies (2.2) at \bar{X} . \square

Now we show a lower bound property of the lifted stationary points of (1.5), which is similar to Lemma 2.3 in [1]. For the ease of the reader, we present the proof here.

Lemma 2.4. *If \bar{X} is a lifted stationary point of (1.5), then it holds that*

$$(2.8) \quad \sigma_i(\bar{X}) \in [0, \nu) \Rightarrow \sigma_i(\bar{X}) = 0, \quad i \in [n].$$

Proof. Suppose \bar{X} is a lifted stationary point of (1.5). Assume that $\sigma_i(\bar{X}) \in (0, \nu)$ for some $i \in [n]$. Then, $d_i^{\bar{X}} = 1$. By Definition 2.1, there exists $\xi(\bar{X}) \in \partial f(\bar{X})$ such that $0 = (U^T \xi(\bar{X}) V)_{ii} + \lambda/\nu$, where $(U, V) \in \mathcal{M}(\bar{X})$. Then, $\lambda/\nu = |(U^T \xi(\bar{X}) V)_{ii}| \leq \|U^T \xi(\bar{X}) V\|_F = \|\xi(\bar{X})\|_F \leq L_f$, which leads to a contradiction to $\nu < \lambda/L_f$. Thus, for any $i \in [n]$, $\sigma_i(\bar{X}) \in [0, \nu)$ implies that $\sigma_i(\bar{X}) = 0$. \square

2.3. Relationship between (1.1) and (1.5). This subsection presents the relationship between problem (1.1) and its continuous relaxation (1.5). According to the lower bound property of the lifted stationary points of (1.5) in Lemma 2.4, we are ready to link (1.1) and (1.5) by the following two results.

Theorem 2.5. *\bar{X} is a global minimizer of (1.1) if and only if it is a global minimizer of (1.5). Moreover, problems (1.1) and (1.5) have the same optimal value.*

Proof. Let \bar{X} be a global minimizer of (1.5), then \bar{X} is a lifted stationary point of (1.5) from Theorem 2.3. By (2.8), it follows $\Phi(\bar{X}) = \|\sigma(\bar{X})\|_0$. Then,

$$\begin{aligned} f(\bar{X}) + \lambda \|\sigma(\bar{X})\|_0 &= f(\bar{X}) + \lambda \Phi(\bar{X}) \\ &\leq f(X) + \lambda \Phi(X) \\ &\leq f(X) + \lambda \|\sigma(X)\|_0, \end{aligned}$$

where the last inequality comes from $\Phi(X) \leq \|\sigma(X)\|_0$, $\forall X \in \mathbb{R}^{m \times n}$. Thus, \bar{X} is a global minimizer of (1.1).

Next, suppose \bar{X} is a global minimizer of (1.1) but not a global minimizer of (1.5). Assume that \hat{X} is a global minimizer of (1.5) satisfying

$$f(\hat{X}) + \lambda \Phi(\hat{X}) < f(\bar{X}) + \lambda \Phi(\bar{X}).$$

As shown earlier, $\Phi(\hat{X}) = \|\sigma(\hat{X})\|_0$. Together with $\Phi(\bar{X}) \leq \|\sigma(\bar{X})\|_0$, we have $f(\hat{X}) + \lambda \|\sigma(\hat{X})\|_0 < f(\bar{X}) + \lambda \|\sigma(\bar{X})\|_0$, which leads to a contradiction. Thus, any global minimizer of (1.1) must be a global minimizer of (1.5).

Lemma 2.4 ensures that problems (1.1) and (1.5) have the same optimal value. \square

Proposition 2.6. *If \bar{X} is a local minimizer of (1.5), then it is a local minimizer of (1.1), and the objective functions of (1.1) and (1.5) have the same value at \bar{X} , i.e., $\mathcal{F}_{\ell_0}(\bar{X}) = \mathcal{F}(\bar{X})$.*

The proof is similar to the first part of Theorem 2.5 and hence we omit it here.

To end this subsection, we present Figure 1 to summarize the relationship between problems (1.1) and (1.5).

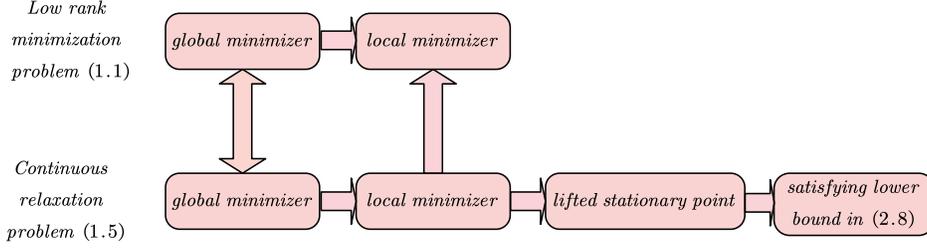


FIGURE 1. Links between problems (1.1) and (1.5).

3. NUMERICAL ALGORITHM AND ITS CONVERGENCE ANALYSIS

In this section, we establish a numerical algorithm to find a lifted stationary point of (1.5). We first introduce some useful preliminary results on smoothing methods and the proximal gradient algorithm, then we propose a proximal gradient algorithm based on the smoothing method. After that we present the convergence of the proposed algorithm.

3.1. Smoothing approximation method and proximal gradient method.

Smoothing approximation method is a common-used numerical method for solving nonsmooth optimization problems. For more details, see [21] and references therein. For the sake of completeness, we recall a class of smoothing functions for $f(X)$ in (1.5).

Definition 3.1. We call $\tilde{f} : \mathbb{R}^{m \times n} \times [0, \bar{\mu}] \rightarrow \mathbb{R}$ with $\bar{\mu} > 0$ a smoothing function of the convex function f in (1.5), if $\tilde{f}(\cdot, \mu)$ is continuously differentiable in $\mathbb{R}^{m \times n}$ for any fixed $\mu > 0$ and satisfies the following conditions:

- (i) $\lim_{X \rightarrow \bar{X}, \mu \downarrow 0} \tilde{f}(X, \mu) = f(\bar{X}), \forall \bar{X} \in \mathbb{R}^{m \times n};$
- (ii) (convexity) $\tilde{f}(X, \mu)$ is convex with respect to X for any fixed $\mu > 0;$
- (iii) (gradient consistency) $\left\{ \lim_{Z \rightarrow X, \mu \downarrow 0} \nabla_Z \tilde{f}(Z, \mu) \right\} \subseteq \partial f(X), \forall X \in \mathbb{R}^{m \times n};$
- (iv) ($\tilde{f}(X, \cdot)$ Lipschitz continuity with respect to μ) there exists a positive constant κ such that

$$\left| \tilde{f}(X, \mu_2) - \tilde{f}(X, \mu_1) \right| \leq \kappa |\mu_1 - \mu_2|, \forall X \in \mathbb{R}^{m \times n}, \mu_1, \mu_2 \in [0, \bar{\mu}];$$
- (v) ($\nabla_X \tilde{f}(\cdot, \mu)$ Lipschitz continuity with respect to X) there exists a constant $L > 0$ such that for any $\mu \in (0, \bar{\mu}], \nabla_X \tilde{f}(\cdot, \mu)$ is Lipschitz continuous with Lipschitz constant $L\mu^{-1}$.

Throughout this paper, we denote $\tilde{f}(X, \mu)$ a smoothing function of $f(X)$ in (1.5). For convenience of notation, the gradient of $\tilde{f}(X, \mu)$ with respect to X is denoted as $\nabla \tilde{f}(X, \mu)$. Furthermore, Definition 3.1-(iv) indicates that

$$(3.9) \quad \left| \tilde{f}(X, \mu) - f(X) \right| \leq \kappa \mu, \quad \forall X \in \mathbb{R}^{m \times n}, 0 < \mu \leq \bar{\mu}.$$

Some smoothing functions of the l_1 loss function in (1.2) can be found in Example 3.1 of [1] and we omit it here.

Some notations are listed here.

$$\tilde{\mathcal{F}}^d(X, \mu) \triangleq \tilde{f}(X, \mu) + \lambda\Phi^d(X) \quad \text{and} \quad \tilde{\mathcal{F}}(X, \mu) \triangleq \tilde{f}(X, \mu) + \lambda\Phi(X),$$

where \tilde{f} is a smoothing function of f , $\mu > 0$ and $d \in \mathbb{D}^n$. For any fixed $\mu > 0$ and $d \in \mathbb{D}^n$, both $\tilde{\mathcal{F}}^d(X, \mu)$ and $\tilde{\mathcal{F}}(X, \mu)$ are nonconvex. Moreover, $\tilde{\mathcal{F}}^d(X, \mu)$ is smooth, but $\tilde{\mathcal{F}}(X, \mu)$ is nonsmooth. Moreover,

$$(3.10) \quad \tilde{\mathcal{F}}^d(X, \mu) \geq \tilde{\mathcal{F}}(X, \mu), \quad \forall d \in \mathbb{D}^n, X \in \mathbb{R}^{m \times n}, \mu \in (0, \bar{\mu}].$$

Now we are ready to recall some preliminaries on proximal gradient method.

Similar to the analysis in Subsection 3.2 of [1], we have a closed-form solution to proximal operator of $\tau\Phi^d$ as follows.

Lemma 3.2. *For any given vectors $d \in \mathbb{D}^n$, $w \in \mathbb{R}_+^n$, and a positive number $\tau > 0$, the proximal operator of $\mathbf{prox}_{\tau\Phi^d}(w)$ has a closed-form solution; i.e.,*

$$(3.11) \quad \hat{x} = \mathbf{prox}_{\tau\Phi^d}(w) := \arg \min_{x \in \mathbb{R}_+^n} \left\{ \tau\Phi^d(\mathcal{D}(x)) + \frac{1}{2}\|x - w\|_F^2 \right\}$$

can be calculated by

$$(3.12) \quad \hat{x}_i = \max\{\bar{w}_i - \tau/\nu, 0\}, \quad i \in [n],$$

where

$$\bar{w}_i = \begin{cases} w_i & \text{if } d_i = 1, \\ w_i + \frac{\tau}{\nu} & \text{if } d_i = 2. \end{cases}$$

Theorem 3.3. *For given $W \in \mathbb{R}^{m \times n}$ and $\tau > 0$, let $U\mathcal{D}(w)V^T$ be the singular value decomposition of W and $\hat{x} = \mathbf{prox}_{\tau\Phi^d}(w)$. Then $\hat{x}_1 \geq \hat{x}_2 \geq \dots \geq \hat{x}_n \geq 0$ and $\hat{X} = U\mathcal{D}(\hat{x})V^T$ is an optimal solution of the problem*

$$(3.13) \quad \min_X \left\{ \tau\Phi^d(X) + \frac{1}{2}\|X - W\|_F^2 \right\}.$$

Proof. From $w_1 \geq w_2 \geq \dots \geq w_n \geq 0$, it is clear that $d_1 \geq d_2 \geq \dots \geq d_n$. Next, we will prove that $\hat{x}_1 \geq \hat{x}_2 \geq \dots \geq \hat{x}_n \geq 0$. We split the proof into three cases.

Case 1. $d_i = d_{i+1} = 2$. By (3.12), it holds

$$\hat{x}_i = w_i \geq w_{i+1} = \hat{x}_{i+1} \geq 0.$$

Case 2. $d_i = 2$ and $d_{i+1} = 1$. By (3.12), it holds

$$\hat{x}_i = w_i \geq \max\{w_{i+1} - \tau/\nu, 0\} = \hat{x}_{i+1} \geq 0.$$

Case 3. $d_i = d_{i+1} = 1$. By (3.12), it holds

$$\hat{x}_i = \max\{w_i - \tau/\nu, 0\} \geq \max\{w_{i+1} - \tau/\nu, 0\} = \hat{x}_{i+1} \geq 0.$$

Combining with all cases, the non-increasing of \hat{x}_i is asserted.

Invoking by [16, Proposition 2.1] with $F(X) = \tau\Phi^d(X)$, $\phi(t) = t^2/2$, $\|\cdot\| = \|\cdot\|_F$ and using the fact that \hat{x} is an optimal solution of $\mathbf{prox}_{\tau\Phi^d}(w)$, it is concluded that $\hat{X} = U\mathcal{D}(\hat{x})V^T$ is an optimal solution of (3.13). \square

Before we end this subsection, we consider the following approximation of $\tilde{\mathcal{F}}^d(\cdot, \mu)$ on a given matrix Z

$$(3.14) \quad Q_{d,\gamma}(X, Z, \mu) = \tilde{f}(Z, \mu) + \langle X - Z, \nabla \tilde{f}(Z, \mu) \rangle + \frac{1}{2}\gamma\mu^{-1}\|X - Z\|_F^2 + \lambda\Phi^d(X)$$

with a constant $\gamma > 0$. Then, minimization problem $\min_X Q_{d,\gamma}(X, Z, \mu)$ has a closed form, denoted by \hat{X} , which can be calculated by Theorem 3.3 with $\tau = \lambda\gamma^{-1}\mu$ and $W = Z - \gamma^{-1}\mu\nabla\tilde{f}(Z, \mu)$.

3.2. SPG algorithm. In this subsection, a proximal gradient algorithm based on the smoothing method, denoted by SPG for simplicity, will be established for finding a lifted stationary point of (1.5).

The following assumptions are needed in the convergence analysis of the SPG algorithm:

- (A1) Assumption 1 and Assumption 2 hold;
- (A2) \tilde{f} is a smoothing function of f defined in Definition 3.1;
- (A3) The global minimum point of \mathcal{F} in (1.5) (or \mathcal{F}_{ℓ_0} in (1.1)) is bounded.

Borrowing from L_f in Assumption 1, ν can be defined such that problems (1.1) and (1.5) have the consistency in Theorem 2.5 and Proposition 2.6. Parameter κ in Definition 3.1 is used in the SPG algorithm, which can be calculated exactly for most smoothing functions [7]. The value of L in Definition 3.1 is not necessary, and we will use a simple line search method to find an acceptable value at each iteration of the SPG algorithm.

Based on the above assumptions, the SPG algorithm for solving (1.5) is outlined as Algorithm 1 here.

At each iteration, the proximal gradient algorithm is adopted for solving $\min_X Q_{d^k, \gamma_k}(X, X^k, \mu_k)$ with fixed μ_k , γ_k , and d^k . The values of γ_k are chosen independently in Step 1 of each iteration. Step 3 updates the smoothing parameter μ_k by (3.18), where $\tilde{\mathcal{F}}(X^{k+1}, \mu_k) + \kappa\mu_k$ can be seen as an energy function, with monotone nonincreasing property, which can be seen from Lemma 3.5. If the energy function decreases more than the given scale, then the smoothing parameter μ_k is still acceptable; otherwise, we reduce it by the updating rule (3.19). Let

$$\mathcal{N}^s = \{k \in \mathbb{N} : \mu_{k+1} \neq \mu_k\},$$

and denote n_r^s the r th smallest number in \mathcal{N}^s . Then, we can update $\{\mu_k\}$ by

$$(3.20) \quad \mu_k = \mu_{n_r^s+1} = \frac{\mu_0}{(n_r^s + 1)^\sigma}, \quad \forall n_r^s + 1 \leq k \leq n_{r+1}^s,$$

which will be used in the proof of Lemma 3.6.

Algorithm 1 SPG algorithm.

Input: Let $X^0 \in \mathbb{R}^{m \times n}$ and $\mu_{-1} = \mu_0 \in (0, \bar{\mu}]$. Choose $\rho, \sigma > 1, \alpha > 0$, and $0 < \underline{\gamma} \leq \bar{\gamma}$. Set $k := 0$.

while not converge **do**

Step 1. Choose $\gamma_k \in [\underline{\gamma}, \bar{\gamma}]$ and let $d^k := d^{X^k}$, where d^{X^k} is defined in (2.3).

Step 2. 2a) Compute

$$(3.15) \quad \hat{X}^{k+1} = \arg \min_X Q_{d^k, \gamma_k}(X, X^k, \mu_k).$$

2b) If \hat{X}^{k+1} satisfies

$$(3.16) \quad \tilde{\mathcal{F}}^{d^k}(\hat{X}^{k+1}, \mu_k) \leq Q_{d^k, \gamma_k}(\hat{X}^{k+1}, X^k, \mu_k),$$

let

$$(3.17) \quad X^{k+1} := \hat{X}^{k+1}$$

and go to **Step 3**. Otherwise, $\gamma_k := \rho\gamma_k$, and return to Step 2a).

Step 3. If

$$(3.18) \quad \tilde{\mathcal{F}}(X^{k+1}, \mu_k) + \kappa\mu_k - \tilde{\mathcal{F}}(X^k, \mu_{k-1}) - \kappa\mu_{k-1} \leq -\alpha\mu_k,$$

set $\mu_{k+1} = \mu_k$; otherwise, set

$$(3.19) \quad \mu_{k+1} := \frac{\mu_0}{(k+1)^\sigma}.$$

Let $k := k + 1$ and go to **Step 1**.

end while

Output: X^{k+1} .

3.3. Convergence analysis. In this subsection, we will present the convergence analysis for the SPG algorithm.

Let $\{X^k\}$, $\{\gamma_k\}$ and $\{\mu_k\}$ be the sequences generated by the SPG algorithm. We first show that the SPG algorithm is well-defined. Then we establish some basic properties of the iterates $\{X^k\}$, $\{\gamma_k\}$ and $\{\mu_k\}$ in Lemma 3.4-3.7. Next, the subsequential convergence of $\{X^k\}$ to a lifted stationary point of (1.5) is established in Proposition 3.8. Finally, we prove the global sequence convergence of iterates $\{X^k\}$ in Theorem 3.9.

Lemma 3.4. *The SPG algorithm is well-defined and $\{\gamma_k\} \subseteq [\underline{\gamma}, \max\{\bar{\gamma}, \rho L\}]$.*

Proof. Clearly, (3.16) holds if and only if

$$\tilde{f}(\hat{X}^{k+1}, \mu_k) \leq \tilde{f}(X^k, \mu_k) + \left\langle \nabla \tilde{f}(X^k, \mu_k), \hat{X}^{k+1} - X^k \right\rangle + \frac{1}{2} \gamma_k \mu_k^{-1} \left\| \hat{X}^{k+1} - X^k \right\|_F^2.$$

Invoking Definition 3.1-(v), (3.16) holds when $\gamma_k \geq L$. Thus, the updating of γ_k in Step 2 is at most $\log_\eta(L/\underline{\gamma}) + 1$ times at each iteration. Hence, the SPG algorithm is well-defined, and we have that $\gamma_k \leq \max\{\bar{\gamma}, \rho L\}$, $\forall k \in \mathbb{N}$. \square

Lemma 3.5. *For any $k \in \mathbb{N}$, we have*

$$(3.21) \quad \tilde{\mathcal{F}}(X^{k+1}, \mu_k) \leq \tilde{\mathcal{F}}(X^k, \mu_k),$$

which implies that $\{\tilde{\mathcal{F}}(X^{k+1}, \mu_k) + \kappa\mu_k\}$ is nonincreasing.

Proof. By (3.15), it follows

$$Q_{d^k, \gamma_k}(X^{k+1}, X^k, \mu_k) \leq Q_{d^k, \gamma_k}(X, X^k, \mu_k), \quad \forall X \in \mathbb{R}^{m \times n}.$$

From (3.14), upon rearranging the terms, we have

$$(3.22) \quad \begin{aligned} \lambda\Phi^{d^k}(X^{k+1}) &\leq \lambda\Phi^{d^k}(X) + \left\langle X - X^{k+1}, \nabla \tilde{f}(X^k, \mu_k) \right\rangle + \frac{1}{2}\gamma_k\mu_k^{-1} \|X - X^k\|_F^2 \\ &\quad - \frac{1}{2}\gamma_k\mu_k^{-1} \|X^{k+1} - X^k\|_F^2. \end{aligned}$$

Moreover, (3.16) can be written as

$$(3.23) \quad \begin{aligned} \tilde{\mathcal{F}}^{d^k}(X^{k+1}, \mu_k) &\leq \tilde{f}(X^k, \mu_k) + \left\langle X^{k+1} - X^k, \nabla \tilde{f}(X^k, \mu_k) \right\rangle \\ &\quad + \frac{1}{2}\gamma_k\mu_k^{-1} \|X^{k+1} - X^k\|_F^2 + \lambda\Phi^{d^k}(X^{k+1}). \end{aligned}$$

Summing up (3.22) and (3.23), there holds

$$(3.24) \quad \begin{aligned} \tilde{\mathcal{F}}^{d^k}(X^{k+1}, \mu_k) &\leq \tilde{f}(X^k, \mu_k) + \lambda\Phi^{d^k}(X) + \left\langle X - X^k, \nabla \tilde{f}(X^k, \mu_k) \right\rangle \\ &\quad + \frac{1}{2}\gamma_k\mu_k^{-1} \|X - X^k\|_F^2, \quad \forall X \in \mathbb{R}^{m \times n}. \end{aligned}$$

For a fixed $\mu > 0$, the convexity of $\tilde{f}(X, \mu)$ with respect to X leads to

$$(3.25) \quad \tilde{f}(X^k, \mu_k) + \left\langle X - X^k, \nabla \tilde{f}(X^k, \mu_k) \right\rangle \leq \tilde{f}(X, \mu_k), \quad \forall X \in \mathbb{R}^{m \times n}.$$

Combining (3.24) and (3.25) and recalling the definition of $\tilde{\mathcal{F}}^{d^k}$, it follows

$$(3.26) \quad \tilde{\mathcal{F}}^{d^k}(X^{k+1}, \mu_k) \leq \tilde{\mathcal{F}}^{d^k}(X, \mu_k) + \frac{1}{2}\gamma_k\mu_k^{-1} \|X - X^k\|_F^2, \quad \forall X \in \mathbb{R}^{m \times n}.$$

Letting $X = X^k$ in (3.26) and by $d^k = d^{X^k}$, we obtain $\Phi^{d^k}(X^k) = \Phi(X^k)$, and hence

$$(3.27) \quad \tilde{\mathcal{F}}^{d^k}(X^{k+1}, \mu_k) \leq \tilde{\mathcal{F}}(X^k, \mu_k).$$

Thanks to (3.10), $\tilde{\mathcal{F}}^{d^k}(X^{k+1}, \mu_k) \geq \tilde{\mathcal{F}}(X^{k+1}, \mu_k)$. Therefore, (3.27) leads to (3.21).

Since $\tilde{\mathcal{F}}(X, \mu) = \tilde{f}(X, \mu) + \lambda\Phi(X)$, it is clear that

$$\tilde{\mathcal{F}}(X, \mu_k) - \tilde{\mathcal{F}}(X, \mu_{k-1}) = \tilde{f}(X, \mu_k) - \tilde{f}(X, \mu_{k-1}) \leq \kappa(\mu_{k-1} - \mu_k),$$

where the last inequality comes from Definition 3.1 (iv). Together with (3.21), there holds

$$(3.28) \quad \tilde{\mathcal{F}}(X^{k+1}, \mu_k) + \kappa\mu_k \leq \tilde{\mathcal{F}}(X^k, \mu_k) + \kappa\mu_k \leq \tilde{\mathcal{F}}(X^k, \mu_{k-1}) + \kappa\mu_{k-1},$$

which implies the nonincreasing property of $\{\tilde{\mathcal{F}}(X^{k+1}, \mu_k) + \kappa\mu_k\}$. \square

Lemma 3.6. *The following statements hold:*

- (i) $\sum_{k=0}^{\infty} \mu_k \leq \Lambda$ with $\Lambda = \frac{1}{\alpha} \left(\tilde{\mathcal{F}}(X^0, \mu_{-1}) + \kappa \mu_{-1} - \min \mathcal{F}(X) \right) + \frac{\mu_0 \sigma}{\sigma - 1} < \infty$;
- (ii) $\lim_{k \rightarrow \infty} \mu_k = 0$.

Proof. (i) From (3.20), we have

$$(3.29) \quad \sum_{k \in \mathcal{N}^s} \mu_k = \sum_{r=1}^{\infty} \frac{\mu_0}{(n_r^s + 1)^\sigma} \leq \sum_{k=1}^{\infty} \frac{\mu_0}{k^\sigma} \leq \frac{\mu_0 \sigma}{\sigma - 1},$$

where n_r^s is the r th smallest element in \mathcal{N}^s . By (A3) and (3.9), we see that

$$(3.30) \quad \tilde{\mathcal{F}}(X^{k+1}, \mu_k) + \kappa \mu_k \geq \mathcal{F}(X^{k+1}) \geq \min \mathcal{F}(X) = \min \mathcal{F}_{\ell_0}(X) > -\infty,$$

where the equality follows from Theorem 2.5. When $k \notin \mathcal{N}^s$, (3.18) can be rewritten as

$$\alpha \mu_k \leq \tilde{\mathcal{F}}(X^k, \mu_{k-1}) + \kappa \mu_{k-1} - \tilde{\mathcal{F}}(X^{k+1}, \mu_k) - \kappa \mu_k,$$

which together with the nonincreasing property of $\left\{ \tilde{\mathcal{F}}(X^{k+1}, \mu_k) + \kappa \mu_k \right\}$ and (3.30) implies that

$$(3.31) \quad \sum_{k \notin \mathcal{N}^s} \mu_k \leq \frac{1}{\alpha} \left(\tilde{\mathcal{F}}(X^0, \mu_{-1}) + \kappa \mu_{-1} - \min \mathcal{F}(X) \right).$$

Combining (3.29) and (3.31), the proof for the estimation in item (i) is completed.

(ii) From (i), (ii) is obvious. \square

Lemma 3.7. *Suppose that the sequence $\{X^k\}$ generated by SPG algorithm is bounded. Then there exists $K \in \mathbb{N}$ such that for all $k \geq K$, it holds that*

- (i) $\left\| \nabla \tilde{f}(X^k, \mu_k) \right\|_F < \frac{1}{2} (\lambda/\nu + L_f)$;
- (ii) $\left\| X^{k+1} - X^k \right\|_F \leq (\sqrt{n} + 1) (\lambda/\nu) \gamma_k^{-1} \mu_k$;
- (iii) $\sum_{k=0}^{\infty} \left\| X^{k+1} - X^k \right\|_F < \infty$.

Proof. (i) We argue it by contradiction. Suppose that there exists a subsequence of $\{X^k\}$, denoted by $\{X^{k_i}\}$, such that

$$(3.32) \quad \left\| \nabla \tilde{f}(X^{k_i}, \mu_{k_i}) \right\|_F \geq \frac{1}{2} (\lambda/\nu + L_f) > L_f, \quad \forall i \in \mathbb{N}.$$

Since $\{X^{k_i}\}$ is bounded, there exists a subsequence of $\{X^{k_i}\}$ (also denoted by $\{X^{k_i}\}$ for simplicity) and \bar{X} such that $\lim_{i \rightarrow \infty} X^{k_i} = \bar{X}$. Due to $\lim_{i \rightarrow \infty} \mu_{k_i} = 0$, the property of \tilde{f} in Definition 3.1-(iii) and (3.32) imply the existence of $\bar{\xi} \in \partial f(\bar{X})$ such that $\|\bar{\xi}\|_F > L_f$, which leads to a contradiction to the definition of L_f given in Assumption 1. Hence, (i) is established.

(ii) Let $W^k = X^k - \gamma_k^{-1} \mu_k \nabla \tilde{f}(X^k, \mu_k)$ and $U^k \mathcal{D}(w^k) (V^k)^T$ be the singular value decomposition of W^k . By (3.12), we have

$$(3.33) \quad \left\| X^{k+1} - W^k \right\|_F = \left\| x^{k+1} - w^k \right\|_F \leq \sqrt{n} (\lambda/\nu) \gamma_k^{-1} \mu_k.$$

From (i), there exists $K \in \mathbb{N}$ such that for all $k > K$, it holds that

$$(3.34) \quad \|W^k - X^k\|_F = \left\| -\gamma_k^{-1} \mu_k \nabla \tilde{f}(X^k, \mu_k) \right\|_F \leq (\lambda/\nu) \gamma_k^{-1} \mu_k.$$

Combining (3.34) and (3.33), we have

$$\|X^{k+1} - X^k\|_F \leq \|X^{k+1} - W^k\|_F + \|W^k - X^k\|_F \leq (\sqrt{n} + 1) (\lambda/\nu) \gamma_k^{-1} \mu_k,$$

which completes the proof for item (ii).

(ii) From Lemma 3.6-(i) and (ii) of this lemma, we have

$$\begin{aligned} \sum_{k=0}^{\infty} \|X^{k+1} - X^k\|_F &\leq \sum_{k=0}^{K-1} \|X^{k+1} - X^k\|_F + \sum_{k=K}^{\infty} \|X^{k+1} - X^k\|_F \\ &\leq \sum_{k=0}^{K-1} \|X^{k+1} - X^k\|_F + (\sqrt{n} + 1) (\lambda/\nu) \underline{\gamma}^{-1} \sum_{k=K}^{\infty} \mu_k < \infty. \end{aligned}$$

□

Proposition 3.8. *Suppose that the sequence $\{X^k\}$ generated by SPG algorithm is bounded. Then any accumulation point of $\{X^k\}$ is a lifted stationary point of (1.5).*

Proof. Suppose that \bar{X} is an accumulation point of any convergence subsequent $\{X^{k_i}\}$. By Lemma 3.7-(iii), we have

$$(3.35) \quad \sum_{i=0}^{\infty} \|X^{k_{i+1}} - X^{k_i}\|_F \leq \sum_{k=0}^{\infty} \|X^{k+1} - X^k\|_F < \infty,$$

which implies that

$$(3.36) \quad \lim_{i \rightarrow \infty} \|X^{k_{i+1}} - X^{k_i}\|_F = 0 \text{ and } \lim_{i \rightarrow \infty} X^{k_{i+1}} = \bar{X}.$$

Recalling $X^{k_{i+1}} = \hat{X}^{k_{i+1}}$ defined in (3.15) and by first-order optimality condition, we have

$$(3.37) \quad \nabla \tilde{f}(X^{k_i}, \mu_{k_i}) + \gamma_{k_i} \mu_{k_i}^{-1} (X^{k_{i+1}} - X^{k_i}) + \lambda \zeta^{k_i} = 0, \forall \zeta^{k_i} \in \partial \Phi^{d^{k_i}}(X^{k_{i+1}}).$$

Since the elements in $\{d^{k_i} : i \in \mathbb{N}\}$ are finite and $\lim_{i \rightarrow \infty} X^{k_{i+1}} = \bar{X}$, there exists a subsequence of $\{k_i\}$, denoted as $\{k_{i_j}\}$, and $\bar{d} \in \mathcal{D}(\sigma(\bar{X}))$ such that $d^{k_{i_j}} = \bar{d}$, $\forall j \in \mathbb{N}$. By the definition of $\partial \Phi^{\bar{d}}$ and $\lim_{j \rightarrow \infty} X^{k_{i_j}+1} = \bar{X}$, it gives

$$(3.38) \quad \left\{ \lim_{j \rightarrow \infty} \zeta^{k_{i_j}} : \zeta^{k_{i_j}} \in \partial \Phi^{d^{k_{i_j}}}(X^{k_{i_j}+1}) \right\} \subseteq \partial \Phi^{\bar{d}}(\bar{X}).$$

Along with the subsequence $\{k_{i_j}\}$ and letting $j \rightarrow \infty$ in (3.37), from Definition 3.1-(iii), (3.36) and (3.38), we obtain that there exist $\bar{\xi} \in \partial f(\bar{X})$ and $\bar{\zeta}^{\bar{d}} \in \partial \Phi^{\bar{d}}(\bar{X})$ such that

$$(3.39) \quad \bar{\xi} + \lambda \bar{\zeta}^{\bar{d}} = 0.$$

By $\bar{d} \in \mathcal{D}(\sigma(\bar{X}))$ and the definition of $\Phi^{\bar{d}}$ in (2.4), (3.39) implies that \bar{X} is a lifted stationary point of (1.5). □

Theorem 3.9. *Suppose that the sequence $\{X^k\}$ generated by SPG algorithm is bounded. Then $\{X^k\}$ is globally convergent to a lifted stationary point of (1.5), i.e., there exists a lifted stationary point \bar{X} of (1.5) such that $\lim_{k \rightarrow \infty} X^k = \bar{X}$.*

Proof. Suppose that $\{X^{k_j}\}$ is a convergent subsequence of $\{X^k\}$ with $\lim_{j \rightarrow \infty} X^{k_j} = \bar{X}$. By Proposition 3.8, \bar{X} is a lifted stationary point of (1.5).

For any $t, s \in \mathbb{N}$, we have

$$(3.40) \quad \|X^{t+s+1} - \bar{X}\|_F \leq \|X^t - \bar{X}\|_F + \sum_{k=t}^{t+s} \|X^{k+1} - X^k\|_F.$$

For any given $\epsilon > 0$, there exists $K_1 > 0$ such that

$$(3.41) \quad \|X^{k_j} - \bar{X}\|_F \leq \epsilon/2, \quad \sum_{k=k_j}^{\infty} \|X^{k+1} - X^k\|_F \leq \epsilon/2, \quad \forall k_j \geq K_1.$$

Here the first inequality dues to $\lim_{j \rightarrow \infty} X^{k_j} = \bar{X}$ and the second inequality comes from Lemma 3.7-(iii).

By letting $t = \bar{k}_j \geq K_1$ in (3.40) and from (3.41), we obtain $\|X^k - \bar{X}\|_F \leq \epsilon, \forall k > K_1 + 1$. From the arbitrariness of $\epsilon > 0$, $\lim_{k \rightarrow \infty} X^k = \bar{X}$ follows. \square

4. NUMERICAL EXPERIMENTS

In this section we conduct numerical experiments to test the performance of the SPG method. In particular, we apply it to solve the problem (1.1) with $f(X) = \|P_\Omega(X - M)\|_1$, that is,

$$(4.42) \quad \min_{X \in \mathbb{R}^m \times n} \mathcal{F}_{l_0}(X) := \|P_\Omega(X - M)\|_1 + \lambda \|\sigma(X)\|_0.$$

We conduct extensive experiments to evaluate our method and then comparing it with some existing methods, including FPCA [18], SVT [3] and VBMFL1 [29]. The platform is Matlab R2014a under Windows 10 on a desktop of a 3.2GHz CPU and 8GB memory. We adopt the root-mean-square error (RMSE) as evaluation metrics

$$RMSE := \sqrt{\frac{\|X^* - M\|_F^2}{mn}},$$

and the final performance of each simulation is evaluated by obtaining an ensemble average of the relative error with T independent Monte Carlo runs.

In the simulation, a typical two-component Gaussian mixture model (GMM) is used as the non-Gaussian noise model. The probability density function (PDF) of GMM is defined as

$$p_v(i) = (1 - c)N(0, \sigma_A^2) + cN(0, \sigma_B^2),$$

where $N(0, \sigma_A^2)$ represents general the noise disturbance with variance σ_A^2 , and $N(0, \sigma_B^2)$ stands for outliers that occur occasionally with a large variance σ_B^2 . The variable c controls the occurrence probability of outliers.

4.1. Random Matrix Completion. In this subsection, we aim to recover a random matrix $M \in \mathbb{R}^{m \times n}$ with rank r based on a subset of entries $\{M_{ij}\}_{(i,j) \in \Omega}$. In detail, we first generate random matrices $M_L = \text{unifrnd}(-0.1, 0.3, m, r) \in \mathbb{R}^{m \times r}$ and $M_R = \text{unifrnd}(-0.1, 0.3, n, r) \in \mathbb{R}^{n \times r}$, then let $M = M_L M_R^T$. We then sample a subset with sampling ratio SR uniformly at random, where $SR = |\Omega|/(mn)$. In our experiment, we set $m = n$. The GMM noise are set at $\sigma_A^2 = 0.0001, \sigma_B^2 = 0.1, c = 0.1$. The rank r is set to 30 and the sampling ratio SR is set to 0.8. For each simulation, an average relative error is obtained via 100 Monte Carlo runs with different realizations of M, Ω and noise.

The performance is firstly compared for different μ_0 under different μ_k iterative methods in step 3 of SPG algorithm. We compare different μ_0 under $\alpha = 0.8$ and $\alpha = +\infty$ ¹. μ_0 increases from 10 to 100 with increment 10 and the size of the square matrix m is set to 150. From Figure 2, we can see that the larger μ_0 becomes, the smaller the value of RMSE, but the more time and iteration steps it costs. It can also be observed that $\alpha = 0.8$ has the better performance than $\alpha = \infty$. This is because when $\alpha = 0.8$, it is necessary to reduce μ_k when (3.18) is not satisfied. When $\alpha = \infty$, no matter how much $\tilde{\mathcal{F}}(X^{k+1}, \mu_k) + \kappa\mu_k$ decreases, μ_k will be reduced. It can be seen that the SPG algorithm can accelerate the convergence speed by adjusting the strategy of μ_k in step 3.

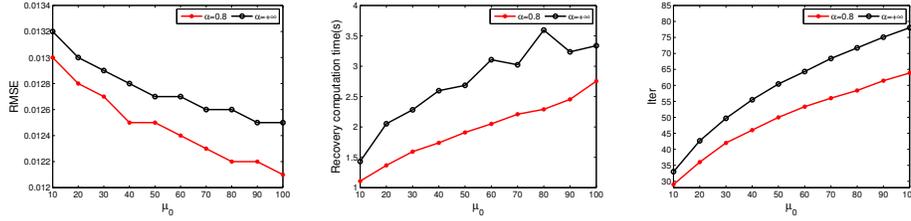


FIGURE 2. Curves of RMSE, average running times and average number of running iterations with different μ_0 .

Secondly, the performance of the algorithms for different sizes of completion problems. The size of the square matrix m increases from 100 to 200 with increment 10. Figure 3 shows the curves of the average RMSE and running times in terms of different matrix sizes m . As can be seen from Figure 3, the SPG algorithm achieves comparably lower average RMSE than the other algorithms, while FPCA and SVT algorithms based on l_2 norm have higher RMSE values. Moreover, as the size of the matrix increases, the average RMSE values decrease for all algorithms. With the increase of matrix size, the average running time of all algorithms increases gradually, but the average running time of SPG algorithm is the least. The average running time of VBMFL1 algorithm based on l_1 norm increases much faster than

¹ $\alpha = +\infty$ means $\mathcal{N}^s = \{1, 2, \dots\}$.

the other three algorithms. In summary, SPG algorithm performs best in four algorithms.

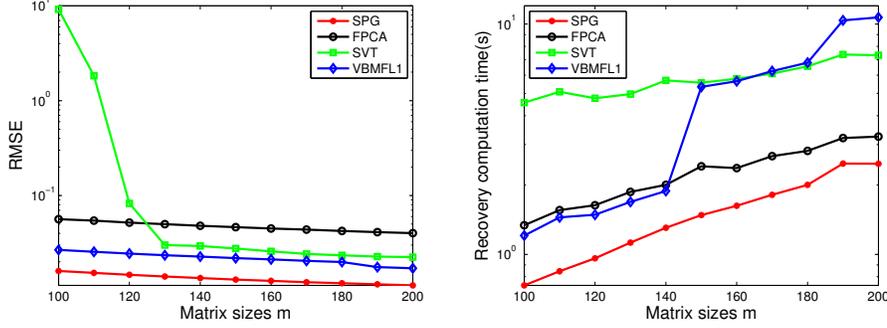


FIGURE 3. Curves of RMSE and average running times with different matrix size m .

4.2. Image Inpainting. In this subsection, the performance of the algorithms is compared for some image inpainting tasks with non-Gaussian noise. Note that grayscale images can be expressed as matrix. When the matrix data is of low-rank, or numerical low-rank, the image inpainting problem can be modeled as matrix completion problem. To evaluate the algorithm performances under non-Gaussian noise, a mixture of Gaussian is selected for the noise model. We adopt the peak signal-to-noise ratio (PSNR) as evaluation metrics, which is defined by

$$\text{PSNR} := 10 \log_{10} \left(\frac{mn}{\|X^* - M\|_F^2} \right).$$

A higher PSNR represents better recovery performance.

We use the USC-SIPI image database² to evaluate our method for image inpainting. In our test, we randomly select 6 images from this database for testing and the images are normalized in the range $[0, 1]$. In Figure 4, we consider the case where entries are missing at random by sampling ratio $SR = 0.9$. The GMM noise are set at $\sigma_A^2 = 0.001$, $\sigma_B^2 = 0.1$, $c = 0.1$. From Figure 4, we can see that the image restored by FPCA and SVT algorithm with l_2 norm is very blurred, while the image restored by SPG and VBMFL1 algorithm with l_1 norm is relatively clear, indicating that the recovery effect of l_1 norm for non-Gaussian noise is better than that of l_2 norm. In the image restored by VBMFL1 algorithm, the recovery effect is not good for those isolated small pixels, especially in ‘‘Chart’’. It may be that these abnormal small pixels are treated as outliers. The image restored by SPG algorithm performs well in all aspects. At the same time, in order to compare the recovery effect of the five algorithms more clearly, we give the PSNR and running time of the four algorithms in Table 1. It can be seen from the table that the VBMFL1 algorithm based on l_1 norm have higher PSNR values than the FPCA and SVT algorithms based on l_2 norm, but the running time is much longer. SPG algorithm has the highest PSNR value and short running time. We can assert that SPG is the best of the four algorithms.

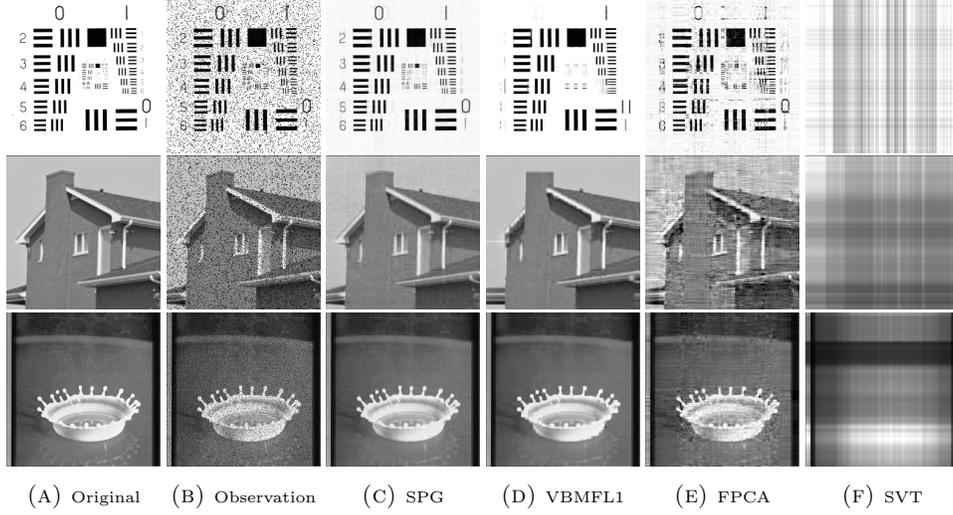


FIGURE 4. Image inpainting sample of image under a mixture of Gaussian noise.

TABLE 1. Image inpainting performance comparison under a mixture of Gaussian noise: PSNR and running times.

Method \ Image	Chart		House		Splash	
	PSNR	time	PSNR	time	PSNR	time
SPG	26.21	7.33	29.19	9.34	31.92	39.02
VBMFL1	21.43	65.67	28.08	53.84	30.88	183.73
FPCA	16.96	7.45	20.35	4.75	24.05	24.64
SVT	10.97	14.33	17.20	8.22	13.79	22.13

In Figure 5 and Table 2, we consider the case where entries are missing at random by sampling ratio $SR = 0.7$, Gaussian noise with variance 0.0001 is added to the observed pixels. It can be seen from Figure 5 and Table 2 that the recovery effect of the algorithm based on l_1 norm is similar to that of the algorithm based on l_2 norm in Gaussian noise, and the running time is relatively long. However, SPG algorithm still has the highest PSNR value and shorter running time. Therefore, whether Gaussian noise or non-Gaussian noise, SPG algorithm performs best for image restoration.

4.3. MRI Volume Dataset. The resolution of the MRI volume dataset³ is of size 217×181 with 181 slices and we selected the 38th slice and the 88th slice for the experiment. We consider the case where entries are missing at random by sampling ratio $SR = 0.9$. The GMM noise are set at $\sigma_A^2 = 0.0001, \sigma_B^2 = 0.1, c = 0.01$.

From Figure 6 and Histogram 7, we can see that the effect of FPCA and SVT to restore images is very poor. The effect of VBMFL1 algorithm to restore images is

²<http://sipi.usc.edu/database/>

³<http://graphics.stanford.edu/data/voldata/>

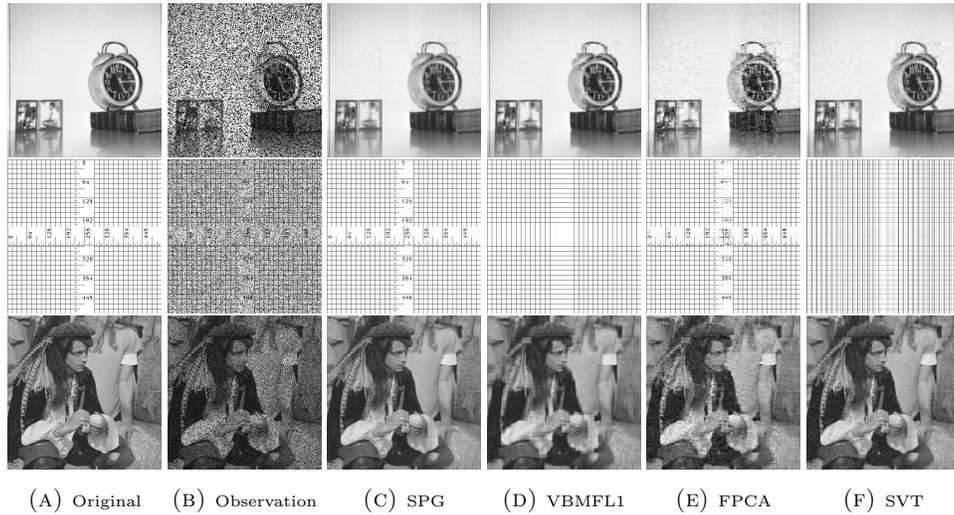


FIGURE 5. Image inpainting sample of image under Gaussian noise.

TABLE 2. Image inpainting performance comparison under Gaussian noise: PSNR and running times.

Method \ Image	Clock		Ruler		Man	
	PSNR	time	PSNR	time	PSNR	time
SPG	29.65	3.62	28.65	12.37	30.11	21.17
VBMFL1	28.23	25.48	17.88	15.73	23.82	114.15
FPCA	24.13	4.82	21.75	23.02	22.68	22.58
SVT	28.87	27.64	10.37	40.73	25.86	81.60

good, but the running time is relatively long. SPG algorithm to restore the image effect and good running time is short. In summary, the SPG algorithm has the best recovery effect.

Acknowledgement Xinzhen Zhang was partly supported by the National Natural Science Foundation of China (Grant No. 11871369). Quan Yu was partly supported by Tianjin Research Innovation Project for Postgraduate Students (Grant No. 2020YJSS140).

REFERENCES

[1] W. Bian and X. Chen, “A smoothing proximal gradient algorithm for nonsmooth convex regression with cardinality penalty”, *SIAM Journal on Numerical Analysis*, 58 (2020), 858-883.

[2] T. Bouwmans and E. H. Zahzah, “Robust PCA via principal component pursuit: a review for a comparative evaluation in video surveillance”, *Computer Vision and Image Understanding*, 122 (2014), 22-34.

[3] J. F. Cai, E. J. Candes and Z. Shen, “A singular value thresholding algorithm for matrix completion”, *SIAM Journal on Optimization*, 20 (2008), 1956-1982.

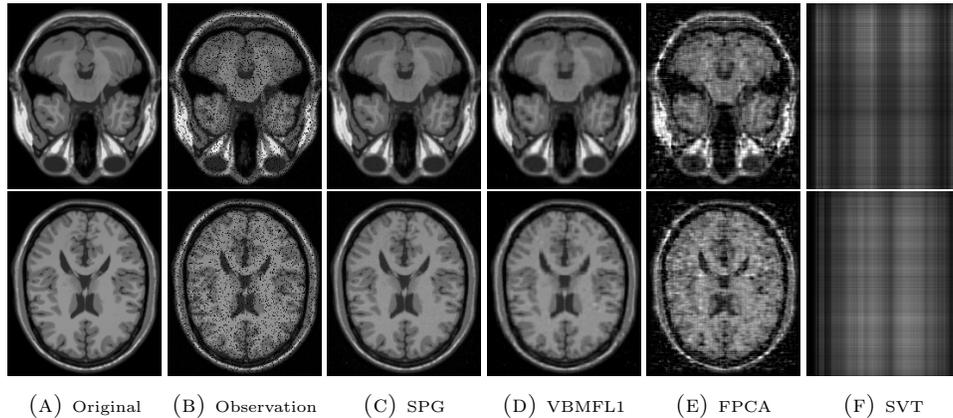


FIGURE 6. Completion results of the MRI Volume Dataset.

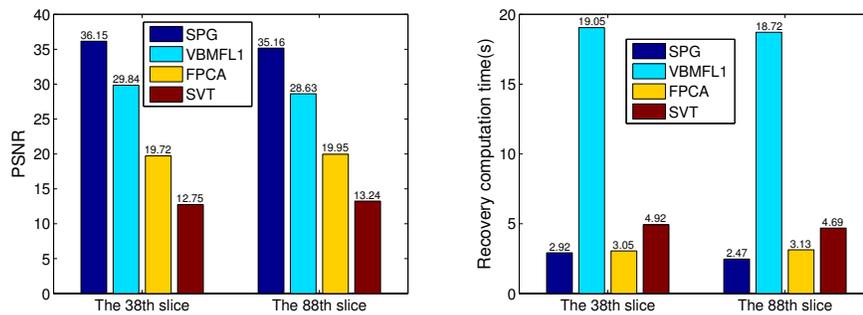


FIGURE 7. Histogram of representation results for the MRI Volume Dataset.

- [4] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization”, *Foundations of Computational Mathematics*, 9 (2009), 717-772.
- [5] E. J. Candès and T. Tao, “The power of convex relaxation: near-optimal matrix completion”, *IEEE Transactions on Information Theory*, 56 (2010), 2053-2080.
- [6] F. Cao, J. Chen, H. Ye, J. Zhao and Z. Zhou, “Recovering low-rank and sparse matrix based on the truncated nuclear norm”, *Neural Networks*, 85 (2017), 10-20.
- [7] X. Chen, “Smoothing methods for nonsmooth, nonconvex minimization”, *Mathematical Programming*, 134 (2012), 71-99.
- [8] F. H. Clarke, “Optimization and nonsmooth analysis”, *Society for Industrial and Applied Mathematics*, (1990).
- [9] M. Fazel, H. Hindi and S. P. Boyd, “A rank minimization heuristic with application to minimum order system approximation”, *Proceedings of the American Control Conference*, 6 (2001), 4734-4739.
- [10] M. Fornasier, H. Rauhut and R. Ward, “Low-rank matrix recovery via iteratively reweighted least squares minimization”, *SIAM Journal on Optimization*, 21 (2011), 1614-1640.
- [11] J. P. Haldar and D. Hernando, “Rank-constrained solutions to linear matrix equations using powerfactorization”, *IEEE Signal Processing Letters*, 16 (2009), 584-587.
- [12] Y. He, F. Wang, Y. Li, J. Qin and B. Chen, “Robust matrix completion via maximum correntropy criterion and half-quadratic optimization”, *IEEE Transactions on Signal Processing*, 68 (2020), 181-195.

- [13] M. J. Lai, Y. Xu and W. Yin, “Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization”, *SIAM Journal on Numerical Analysis*, 51 (2013), 927-957.
- [14] C. Lee and E. Y. Lam, “Computationally efficient truncated nuclear norm minimization for high dynamic range imaging”, *IEEE Transactions on Image Processing*, 25 (2016), 4145-4157.
- [15] A. S. Lewis and H. S. Sendov, “Nonsmooth analysis of singular values. Part II: applications”, *Set-Valued Analysis*, 13 (2005), 243-264.
- [16] Z. Lu, Y. Zhang and X. Li, “Penalty decomposition methods for rank minimization”, *Optimization Methods and Software*, 30 (2015), 531-558.
- [17] Z. Lu, Y. Zhang and J. Lu, “ l_p regularized low-rank approximation via iterative reweighted singular value minimization”, *Computational Optimization and Applications*, 68 (2017), 619-642.
- [18] S. Ma, D. Goldfarb and L. Chen, “Fixed point and Bregman iterative methods for matrix rank minimization”, *Mathematical Programming*, 128 (2011), 321-353.
- [19] T. H. Ma, Y. Lou and T. Z. Huang, “Truncated l_{1-2} models for sparse recovery and rank minimization”, *SIAM Journal on Imaging Sciences*, 10 (2017), 1346-1380.
- [20] D. Meng, Z. Xu, L. Zhang and J. Zhao, “A cyclic weighted median method for L1 low-rank matrix factorization with missing entries”, *Seventh AAAI Conference on Artificial Intelligence*, (2013), 704-710.
- [21] J. Nocedal and S. J. Wright, “Numerical optimization”, Springer, Berlin, 2006.
- [22] L. Pan and X. Chen, “Group sparse optimization for images recovery using capped folded concave functions”, *SIAM Journal on Imaging Sciences*, 14 (2021), 1-25.
- [23] C. Peng, Z. Kang and Q. Cheng, “A fast factorization-based approach to robust PCA”, *2016 IEEE 16th International Conference on Data Mining*, (2016), 1137-1142.
- [24] X. Su, Y. Wang, X. Kang and R. Tao, “Nonconvex truncated nuclear norm minimization based on adaptive bisection method”, *IEEE Transactions on Circuits and Systems for Video Technology*, 29 (2019), 3159-3172.
- [25] N. Wang, T. Yao, J. Wang and D. Y. Yeung, “A probabilistic approach to robust matrix factorization”, *Computer Vision - ECCV 2012*, (2012), 126-139.
- [26] Z. Wen, W. Yin and Y. Zhang, “Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm”, *Mathematical Programming Computation*, 4 (2012), 333-361.
- [27] H. Xu, C. Caramanis and S. Sanghavi, “Robust PCA via outlier pursuit”, *IEEE Transactions on Information Theory*, 58 (2012), 3047-3064.
- [28] W. J. Zeng and H. C. So, “Outlier-robust matrix completion via l_p -minimization”, *IEEE Transactions on Signal Processing*, 66 (2018), 1125-1140.
- [29] Q. Zhao, D. Meng, Z. Xu, W. Zuo and Y. Yan, “ L_1 -norm low-rank matrix factorization by variational bayesian method”, *IEEE Transactions on Neural Networks and Learning Systems*, 26 (2015), 825-839.
- [30] Y. Zheng, G. Liu, S. Sugimoto, S. Yan and M. Okutomi, “Practical low-rank matrix approximation under robust L1-norm”, *2012 IEEE Conference on Computer Vision and Pattern Recognition*, (2012), 1410-1417.

SCHOOL OF MATHEMATICS, TIANJIN UNIVERSITY, TIANJIN 300350, CHINA.

Email address: QuanYu527@163.com

SCHOOL OF MATHEMATICS, TIANJIN UNIVERSITY, TIANJIN 300350, CHINA.

Email address: xzzhang@tju.edu.cn