

Article

A Fast Proximal Alternating Method for Robust Matrix Factorization of Matrix Recovery with Outliers

Ting Tao ¹, Lianghai Xiao ^{2,*} and Jiayuan Zhong ^{1,*}¹ School of Mathematics, Foshan University, Foshan 528011, China; taoting@fosu.edu.cn² College of Information Science and Technology, Jinan University, Guangzhou 510632, China

* Correspondence: xiaolh@jnu.edu.cn (L.X.); zjiayuan@foshan.edu.cn (J.Z.)

Abstract: This paper concerns a class of robust factorization models of low-rank matrix recovery, which have been widely applied in various fields such as machine learning and imaging sciences. An ℓ_1 -loss robust factorized model incorporating the $\ell_{2,0}$ -norm regularization term is proposed to address the presence of outliers. Since the resulting problem is nonconvex, nonsmooth, and discontinuous, an approximation problem that shares the same set of stationary points as the original formulation is constructed. Subsequently, a proximal alternating minimization method is proposed to solve the approximation problem. The global convergence of its iterate sequence is also established. Numerical experiments on matrix completion with outliers and image restoration tasks demonstrate that the proposed algorithm achieves low relative errors in shorter computational time, especially for large-scale datasets.

Keywords: matrix recovery with outliers; global convergence; column $\ell_{2,0}$ -norm; alternating method

MSC: 49J52; 49M27; 90C26



Academic Editor: Ioannis Tsoulos

Received: 20 March 2025

Revised: 21 April 2025

Accepted: 26 April 2025

Published: 29 April 2025

Citation: Tao, T.; Xiao, L.; Zhong, J. A Fast Proximal Alternating Method for Robust Matrix Factorization of Matrix Recovery with Outliers. *Mathematics* **2025**, *13*, 1466. <https://doi.org/10.3390/math13091466>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The low-rank matrix recovery problem seeks to recover a true yet unknown low-rank matrix M^* of rank r^* from the minimal feasible observations. This problem encompasses various applications in numerous fields, including signal and image processing, quantum state tomography, control and system identification, statistics, machine learning, etc. (see [1–5]). This paper considers a robust factorized model for the low-rank matrix recovery problem, in which the observations contain outliers. The model is expressed as

$$b = \mathcal{A}(M^*) + \omega,$$

where $\mathcal{A}: \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ is a sampling operator, $b \in \mathbb{R}^m$ is the observation vector, and ω is a sparse noise vector with arbitrarily large nonzero entries, while the rest of the elements are zero. In scenarios where outliers exist, the traditional smooth least-squared loss function becomes too sensitive to outliers and is probably biased. In view of this, studies [6,7] have proposed a non-smooth ℓ_p -loss with balance regularization model:

$$\min_{U \in \mathbb{R}^{n_1 \times \kappa}, V \in \mathbb{R}^{n_2 \times \kappa}} \Psi(U, V) := \|\mathcal{A}(UV^T) - b\|_p + \gamma \|UU^T - VV^T\|_F, \quad (1)$$

where $\gamma \geq 0$. When $\gamma > 0$, for $p = 1$ and $\kappa = r^*$, Li et al. [7] established exact matrix recovery based on the ℓ_1/ℓ_2 -restricted isometry property of the sampling operator $\mathcal{A}$ and

proved that the proposed subgradient method with geometrically diminishing step sizes converges linearly to the ground-truth matrix. For $p = 1$ or 2 and $\kappa = r^*$, Charisopoulos et al. [6] proved that the subgradient method and prox-linear method converge at a fast dimension-independent rate under the same assumptions as in [7]. When $\gamma = 0$, for $p = 1$, Ma and Fattahi [8] showed that the subgradient method converges to the true matrix under the Sign-RIP condition [[8], Definition 7]. Additionally, Wang et al. [9] developed a robust and fast rank-one matrix completion algorithm by minimizing the Welsch cost function.

Although the subgradient method exhibits favorable convergence properties in solving the problem (1), it is limited by some strict assumptions. Practically, the rank of the ground-truth matrix is generally unknown, and consequently, the regularization term in (1) fails to induce low-rank structures. Inspired by [10], this paper utilizes the $\ell_{2,0}$ -norm in the regularization term and studies the following ℓ_1 -loss factorized model of low-rank matrix recovery:

$$\min_{U \in \mathbb{R}^{n_1 \times \kappa}, V \in \mathbb{R}^{n_2 \times \kappa}} \Phi(U, V) := \|\mathcal{A}(UV^\top) - b\|_1 + \lambda(\|U\|_{2,0} + \|V\|_{2,0}), \quad (2)$$

where κ is an upper estimator of r^* , and $\|\cdot\|_{2,0}$ denotes the $\ell_{2,0}$ -norm of matrices, which measures the number of the nonzero column. The regularization term $\lambda(\|U\|_{2,0} + \|V\|_{2,0})$ is used to reduce rank through column sparsity. The ℓ_1 -loss function, on the other hand, is more robust against outliers and has been widely utilized for outlier detection studies [11–13]. Both terms of the objective function (2) are nonsmooth. On top of that, the function suffers from both nonconvexity and discontinuity due to the existence of the $\ell_{2,0}$ -norm regularization term. The discontinuity causes the subgradient algorithm to fail when it is applied to problem (2). While the alternating direction method of multipliers is commonly applied to address problem (2), there is still no theoretical guarantee of its convergence in the context of such nonconvex and nonsmooth optimization problems. For methods with convergence guarantees, the majority of them impose, within their objective functions, either the assumption of smoothness for at least one term or the continuity assumption. To the best of our knowledge, no state-of-the-art algorithm is capable of handling this type of nonconvex and discontinuous problem.

The major contributions of this work are threefold. Firstly, to handle the nonconvex and discontinuous problem (2), a novel potential function is constructed:

$$\min_{U \in \mathbb{R}^{n_1 \times \kappa}, V \in \mathbb{R}^{n_2 \times \kappa}, z \in \mathbb{R}^m} \Theta(U, V, z) := \frac{\alpha}{2} \|\mathcal{A}(UV^\top) - b - z\|^2 + \|z\|_1 + \lambda(\|U\|_{2,0} + \|V\|_{2,0}), \quad (3)$$

where $\alpha > 0$ is a given constant. It has been proven that, under mild assumptions, the sets of stationary points of (2) and (3) are equivalent. Moreover, it has been demonstrated that any global optimal solution to problem (3) is an approximate optimal solution to problem (2). This equivalence enables us to find solutions to problem (2) by solving problem (3), and the latter is more computationally efficient. Secondly, a proximal alternating linearized minimization (PALM) method is proposed to solve (3). The global convergence of PALM method is also established by exploiting the Kurdyka–Łojasiewicz (KL) property of the objective function. Finally, numerical experiments conducted on both synthetic and real-world datasets have validated that the PALM method for problem (3) outperforms SubGM [6,7] for problem (1) with $p = 1$.

The remainder of this paper is organized as follows. Section 2 introduces preliminaries and foundational concepts. In Section 3, the relationship between problems (2) and (3) is established. Section 4 presents the PALM method to solve the problem (3), with its global convergence established. In Section 5, the numerical performance of the PALM method is

evaluated, and its performance is compared with that of SubGM to solve the problem (1) through experiments on various datasets. Section 6 concludes the paper.

2. Preliminaries

This section introduces the notation employed and then presents the definitions of stationary points and the KL property.

2.1. Notations

Throughout this paper, $\mathbb{R}^{n_1 \times n_2}$ represents the vector space of all $n_1 \times n_2$ real matrices, equipped with the trace inner product $\langle X, Y \rangle = \text{trace}(X^\top Y)$ for $X, Y \in \mathbb{R}^{n_1 \times n_2}$ and its induced Frobenius norm. For a matrix $X \in \mathbb{R}^{n_1 \times n_2}$, Z_j denotes the j th column of Z , J_Z represents its index set of nonzero columns, and $\|X\|_F$ and $\|X\|_{2,0}$ denote the Frobenius norm and the column $\ell_{2,0}$ -norm of X , which means the number of nonzero columns. For a matrix $X \in \mathbb{R}^{n_1 \times n_2}$, let $\sigma(X) := (\sigma_1(X), \dots, \sigma_n(X))^\top$ with $\sigma_1(X) \geq \dots \geq \sigma_n(X)$, where $\sigma_i(X)$ denotes the singular value of X , $\Sigma_\kappa(X) := \text{Diag}(\sigma_1(X), \dots, \sigma_\kappa(X))$. For a vector $x \in \mathbb{R}^n$, $\|x\|_1$ denotes the ℓ_1 -norm and $\|x\|$ denotes the ℓ_2 -norm. Let $\nabla_U F(U, V)$ and $\nabla_V F(U, V)$ denote the partial gradient of F at (U, V) w.r.t. variable U and V , respectively. For any $x \in \mathbb{R}^m$, $\text{sign}(x) = [\text{sign}(x_1), \dots, \text{sign}(x_m)]$ with

$$\text{sign}(t) := \begin{cases} t/|t|, & t \neq 0; \\ 0, & t = 0. \end{cases} \quad \text{for any } t \in \mathbb{R}.$$

For convenience, define $G(U, V, z) := \frac{\alpha}{2} \|A(UV^\top) - b - z\|^2$,

$$F(U, V) := \|A(UV^\top) - b\|_1 \quad \text{and} \quad g(U) := \|U\|_{2,0}.$$

2.2. Stationary Points and ϵ -Global Optimal Solutions

From [14], the generalized subdifferentials of an extended real-valued function $h: \mathbb{R}^{n \times m} \rightarrow \overline{\mathbb{R}} := (-\infty, \infty]$ at a point where h attains a finite value are recalled.

Definition 1. Given a function $h: \mathbb{R}^{n \times m} \rightarrow \overline{\mathbb{R}}$ with a point x such that $h(x)$ is finite. The regular subdifferential of h at x is defined as follows:

$$\widehat{\partial}h(x) := \left\{ v \in \mathbb{R}^{n \times m} \mid \liminf_{x' \rightarrow x} \frac{h(x') - h(x) - \langle v, x' - x \rangle}{\|x' - x\|_F} \geq 0 \right\},$$

and the basic (or, say, limiting or Mordukhovich) subdifferential of h at x is defined as

$$\partial h(x) := \left\{ v \in \mathbb{R}^{n \times m} \mid \exists x^k \rightarrow x \text{ with } h(x^k) \rightarrow h(x) \text{ and } v^k \rightarrow v \text{ with } v^k \in \widehat{\partial}h(x^k) \right\}.$$

Consider a proper lower semicontinuous (lsc) function $h: \mathbb{R}^{n \times m} \rightarrow \overline{\mathbb{R}}$, the stationary point and ϵ -global optimal solution to problem $\min_{x \in \mathbb{R}^{n \times m}} h(x)$ are defined as follows.

Definition 2. For a point $x \in \mathbb{R}^{n \times m}$ such that $0 \in \partial h(x)$, it is called a stationary point of $\min_{x \in \mathbb{R}^{n \times m}} h(x)$.

Definition 3. For a point $x^* \in \mathbb{R}^{n \times m}$ such that

$$h(x^*) - \epsilon \leq h(x) \quad \forall x \in \mathbb{R}^{n \times m},$$

it is called an ϵ -global optimal solution of $\min_{x \in \mathbb{R}^{n \times m}} h(x)$.

2.3. Kurdyka–Łojasiewicz Property

From [15], the KL property of an extended real-valued function is stated as follows.

Definition 4. Given $h: \mathbb{X} \rightarrow \overline{\mathbb{R}}$ a proper lsc function. It is said to have the Kurdyka–Łojasiewicz (KL) property at $\bar{x} \in \text{dom } \partial h$ if there exists $\eta \in (0, \infty]$; a continuous concave function $\varphi: [0, \eta) \rightarrow \mathbb{R}_+$ satisfies

(i) $\varphi(0) = 0$ and φ is continuously differentiable on $(0, \eta)$, and

(ii) $\forall s \in (0, \eta), \varphi'(s) > 0$;

and a neighborhood $\mathcal{U}$ of $\bar{x}$ such that $\forall x \in \mathcal{U} \cap [h(\bar{x}) < h(x) < h(\bar{x}) + \eta]$,

$$\varphi'(h(x) - h(\bar{x})) \text{dist}(0, \partial h(x)) \geq 1.$$

If the function h satisfies the KL property at any points of $\text{dom } \partial h$, then h is called a KL function.

Remark 1. According to Lemma 2.1 of [15], a proper lsc function has the KL property at any noncritical point. Thus, to prove that a proper lsc function $h: \mathbb{X} \rightarrow \overline{\mathbb{R}}$ is a KL function, it suffices to check whether h has the KL property at any critical point.

3. Relationship Between Problems (2) and (3)

This section demonstrates the relationship between the stationary points and global optimal solutions of (2) and (3). On top of that, the stationary point sets of (2) and (3) are characterized. The subdifferential of Φ at an arbitrary point $(U, V) \in \mathbb{R}^{n_1 \times \kappa} \times \mathbb{R}^{n_2 \times \kappa}$ is given in the following lemma.

Lemma 1. Fix any $\lambda > 0$. Consider any $(U, V) \in \mathbb{R}^{n_1 \times \kappa} \times \mathbb{R}^{n_2 \times \kappa}$. Then, it holds that

$$\widehat{\partial}\Phi(U, V) = \partial\Phi(U, V) = \left\{ \begin{pmatrix} \mathcal{A}^*(\xi)V + \lambda\partial g(U) \\ [\mathcal{A}^*(\xi)]^\top U + \lambda\partial g(V) \end{pmatrix} \middle| \xi \in \partial\|\cdot\|_1(\mathcal{A}(UV^\top) - b) \right\},$$

where for any $X \in \mathbb{R}^{n \times m}$

$$\widehat{\partial}g(X) = \partial g(X) = S_1 \times \cdots \times S_m \text{ with } S_j = \begin{cases} \{0\}^n & \text{if } j \in J_X; \\ \mathbb{R}^n & \text{if } j \notin J_X. \end{cases}$$

Proof. Recall that $\Phi(U, V) = F(U, V) + \lambda(g(U) + g(V))$. From [[14], Corollary 10.9 and Exercise 10.10], it immediately follows that

$$\begin{aligned} \widehat{\partial}F(U, V) + \lambda\widehat{\partial}g(U) \times \lambda\widehat{\partial}g(V) &\subset \widehat{\partial}\Phi(U, V) \subset \partial\Phi(U, V) \\ &\subset \partial F(U, V) + \lambda\partial g(U) \times \lambda\partial g(V). \end{aligned}$$

Next, $\partial F(U, V)$ and $\widehat{\partial}F(U, V)$ are calculated. According to the chain rule [[14], Corollary 8.11 and Theorem 10.6], we obtain

$$\partial F(U, V) = \widehat{\partial}F(U, V) = \left\{ \begin{pmatrix} \mathcal{A}^*(\xi)V \\ [\mathcal{A}^*(\xi)]^\top U \end{pmatrix} \middle| \xi \in \partial\|\cdot\|_1(\mathcal{A}(UV^\top) - b) \right\}.$$

For any $U \in \mathbb{R}^{n_1 \times \kappa}$, by invoking [[14], Proposition 10.5], it immediately follows that

$$\widehat{\partial}g(U) = \partial g(U) = S_1 \times \cdots \times S_\kappa \text{ with } S_j = \begin{cases} \{0\}^{n_1} & \text{if } j \in J_U; \\ \mathbb{R}^{n_1} & \text{if } j \notin J_U. \end{cases} \quad (4)$$

This, along with the inclusion above, implies that the desired result holds. $\square$

The subdifferential of Θ is characterized by the following lemma.

Lemma 2. Fix any $\lambda > 0$. Consider any $(U, V, z) \in \mathbb{R}^{n_1 \times \kappa} \times \mathbb{R}^{n_2 \times \kappa} \times \mathbb{R}^m$. Then, it holds that

$$\widehat{\partial}\Theta(U, V, z) = \partial\Theta(U, V) = \begin{pmatrix} \alpha \mathcal{A}^*(\mathcal{A}(UV^\top) - b - z)V + \lambda \partial g(U) \\ \alpha [\mathcal{A}^*(\mathcal{A}(UV^\top) - b - z)]^\top U + \lambda \partial g(V) \\ \alpha(z - \mathcal{A}(UV^\top) + b) + \partial \|z\|_1 \end{pmatrix},$$

where ∂g is defined in Lemma 1.

Proof. Recall that $\Theta(U, V, z) = G(U, V, z) + \|z\|_1 + \lambda(g(U) + g(V))$. Then, from [[14], Exercise 8.8(c) and Proposition 10.5], it immediately follows that

$$\begin{aligned} \nabla G(U, V, z) + \lambda \widehat{\partial}g(U) \times \lambda \widehat{\partial}g(V) \times \partial \|z\|_1 &\subset \widehat{\partial}\Theta(U, V, z) \subset \partial\Theta(U, V, z) \\ &\subset \nabla G(U, V, z) + \lambda \partial g(U) \times \lambda \partial g(V) \times \partial \|z\|_1. \end{aligned}$$

By the smoothness and the expression of $G(U, V, z)$, one has

$$\nabla G(U, V, z) = \begin{pmatrix} \alpha \mathcal{A}^*(\mathcal{A}(UV^\top) - b - z)V \\ \alpha [\mathcal{A}^*(\mathcal{A}(UV^\top) - b - z)]^\top U \\ \alpha(z - \mathcal{A}(UV^\top) + b) \end{pmatrix}.$$

Along with the above inclusion and (4), the desired results are obtained. $\square$

The following proposition states the equivalence between the stationary points of problems (2) and (3) under some mild conditions.

Proposition 1. If $(U, V, z) \in \mathbb{R}^{n_1 \times \kappa} \times \mathbb{R}^{n_2 \times \kappa} \times \mathbb{R}^m$ is an arbitrary stationary point of (3) satisfying the nonzero elements of $|\mathcal{A}(UV^\top) - b|$ not smaller than $\frac{1}{\alpha}$, then (U, V) is a stationary point of (2). Conversely, if $(U, V) \in \mathbb{R}^{n_1 \times \kappa} \times \mathbb{R}^{n_2 \times \kappa}$ is an arbitrary stationary point of (2) satisfying all elements of $|\mathcal{A}(UV^\top) - b|$ not smaller than $\frac{1}{\alpha}$, then there exists $z \in \mathcal{A}(UV^\top) - b - \text{sign}(\mathcal{A}(UV^\top) - b)$ such that (U, V, z) is a stationary point of (3).

Proof. Pick any stationary point (U, V, z) of (3). From Lemma 2, it holds that

$$\begin{cases} 0 \in \alpha \mathcal{A}^*(\mathcal{A}(UV^\top) - b - z)V + \lambda \partial g(U); & (5a) \\ 0 \in \alpha [\mathcal{A}^*(\mathcal{A}(UV^\top) - b - z)]^\top U + \lambda \partial g(V); & (5b) \\ 0 \in -\alpha(\mathcal{A}(UV^\top) - b - z) + \partial \|z\|_1. & (5c) \end{cases}$$

Write $y := \mathcal{A}(UV^\top) - b$. From (5c), it can be concluded that $z = \text{sign}(y) \circ \max(0, |y| - \frac{1}{\alpha})$. Then,

$$y - z = \text{sign}(y) \circ \min(|y|, \frac{1}{\alpha}), \quad (6)$$

Along with $|y_i| \geq \frac{1}{\alpha}$ or $y_i = 0$ for all $i = 1, \dots, m$, one has $y - z = \frac{1}{\alpha} \text{sign}(y)$. Thus, substituting $y - z = \frac{1}{\alpha} \text{sign}(y)$ into (5a) and (5b), one obtains

$$\begin{cases} 0 \in \mathcal{A}^*(\text{sign}(\mathcal{A}(UV^\top) - b))V + \lambda \partial g(U); \\ 0 \in [\mathcal{A}^*(\text{sign}(\mathcal{A}(UV^\top) - b))]^\top U + \lambda \partial g(V). \end{cases}$$

This together with the definition of $\text{sign}(\cdot)$ implies that (U, V) is a stationary point of (2).

Conversely, any (U, V) is a stationary point of (2). According to Lemma 1, there exists $\xi \in \partial \|\cdot\|_1(\mathcal{A}(UV^\top) - b)$ such that

$$\begin{cases} 0 \in \mathcal{A}^*(\xi)V + \lambda \partial g(U) \\ 0 \in [\mathcal{A}^*(\xi)]^\top U + \lambda \partial g(V) \end{cases}$$

Write $y := \mathcal{A}(UV^\top) - b$ and $z := y - \frac{1}{\alpha}\xi$. Then, (U, V, z) satisfies (5a) and (5b). By noting that $|\xi_i| \leq 1$ for all $i = 1, \dots, m$, and $|y_i| \geq \frac{1}{\alpha}$ for all $i = 1, \dots, m$, one has $\alpha(y - z) = \xi \in \partial \|\cdot\|_1(y - \frac{1}{\alpha}\xi) = \partial \|z\|_1$ which means that (5c) holds. This implies that (U, V, z) is a stationary point of (3). The proof is completed. $\square$

Remark 2. The converse of the above property generally does not hold, as the obtained vector y is expected to be sparse. This indicates that the set of stationary points for problem (3) is typically smaller than that of problem (2). Through the relationship between the global optimal solutions of (2) and (3) in next proposition, the stationary points of (3) exclude certain undesirable points contained in the stationary point set of (2).

Proposition 2. If (U^*, V^*, z^*) is a global optimal solution of (3), then (U^*, V^*) is a $\frac{m}{2\alpha}$ -global optimal solution of (2).

Proof. If (U^*, V^*, z^*) is a global optimal solution of (3), then (U^*, V^*, z^*) is a stationary point of (3). Write $y^* := \mathcal{A}(U^*(V^*)^\top) - b$. From (6), it holds that

$$y^* - z^* = \text{sign}(y^*) \circ \min(|y^*|, \frac{1}{\alpha}).$$

Let $I := \{i \in [m] \mid |y_i^*| < \frac{1}{\alpha}\}$ and $\bar{I} := \{i \in [m] \mid |y_i^*| \geq \frac{1}{\alpha}\}$. Then, one has

$$\begin{aligned} G(U^*, V^*, z^*) &= \frac{\alpha}{2} \|y^* - z^*\|^2 + \|z^*\|_1 \\ &= \frac{\alpha}{2} \|\text{sign}(y^*) \circ \min(|y^*|, \frac{1}{\alpha})\|^2 + \|\text{sign}(y^*) \circ \max(0, |y^*| - \frac{1}{\alpha})\|_1 \\ &= \frac{\alpha}{2} \sum_{i \in I} |y_i^*|^2 + \sum_{i \in \bar{I}} (|y_i^*| - \frac{1}{2\alpha}) \\ &= \|y^*\|_1 - \frac{m}{2\alpha} + \sum_{i \in \bar{I}} (\frac{\alpha}{2} |y_i^*|^2 - |y_i^*| + \frac{1}{2\alpha}) \geq \|y^*\|_1 - \frac{m}{2\alpha}, \end{aligned}$$

where the inequality is due to $|y_i^*| < \frac{1}{\alpha}$ for $i \in I$. Since (U^*, V^*, z^*) is a global optimal solution of (3), for any (U, V, z) with $z = \mathcal{A}(UV^\top) - b$

$$\Phi(U^*, V^*) - \frac{m}{2\alpha} \leq \Theta(U^*, V^*, z^*) \leq \Theta(U, V, z) = \Phi(U, V).$$

This means that (U^*, V^*, z^*) is a $\frac{m}{2\alpha}$ -global optimal solution of (2). $\square$

Remark 3. Proposition 2 demonstrates that global optimal solution (U^*, V^*, z^*) of problem (3) is an approximate optimal solution to problem (2), and as the parameter α increases, the value $\Phi(U^*, V^*)$ progressively approaches the optimal value of problem (2).

4. A PALM Method for Solving Problem (3)

Recall that $G(U, V, z) = \frac{\alpha}{2} \|\mathcal{A}(UV^\top) - b - z\|^2$ is a smooth function. The gradients $\nabla_U G(\cdot, V, z)$ and $\nabla_V G(U, \cdot, z)$ are Lipschitz continuous with moduli defined as τ_V and τ_U , respectively. Fix any $(U', V', z') \in \mathbb{R}^{n_1 \times \kappa} \times \mathbb{R}^{n_2 \times \kappa} \times \mathbb{R}^m$; according to decent Lemma [[16], Proposition A.24], it holds that

$$G(U, V', z') \leq G(U', V', z') + \langle \nabla_U G(U', V', z'), U - U' \rangle + \frac{\tau_{V'}}{2} \|U - U'\|_F^2, \quad \forall U \in \mathbb{R}^{n_1 \times \kappa},$$

$$G(U', V, z') \leq G(U', V', z') + \langle \nabla_V G(U', V', z'), V - V' \rangle + \frac{\tau_{U'}}{2} \|V - V'\|_F^2, \quad \forall V \in \mathbb{R}^{n_2 \times \kappa}.$$

Let (U^k, V^k, z^k) be the current iteration. From the expression of Θ and the above two inequalities, one obtains

$$\begin{cases} \Theta(U, V^k, z^k) \leq \tilde{\Theta}_U(U, V^k, z^k) := G(U^k, V^k, z^k) + \langle \nabla_U G(U^k, V^k, z^k), U - U^k \rangle \\ \quad + \lambda \|U\|_{2,0} + \|z^k\|_1 + \frac{\tau_{V^k}}{2} \|U - U^k\|_F^2, \\ \Theta(U^k, V, z^k) \leq \tilde{\Theta}_V(U^k, V, z^k) := G(U^k, V^k, z^k) + \langle \nabla_V G(U^k, V^k, z^k), V - V^k \rangle \\ \quad + \lambda \|V\|_{2,0} + \|z^k\|_1 + \frac{\tau_{U^k}}{2} \|V - V^k\|_F^2, \end{cases}$$

which become equalities when $U = U^k$ and $V = V^k$. Consequently, $\tilde{\Theta}_U(\cdot, V^k, z^k)$ and $\tilde{\Theta}_V(U^k, \cdot, z^k)$ serve as majorizations for $\Theta(\cdot, V^k, z^k)$ at U^k and $\Theta(U^k, \cdot, z^k)$ at V^k , respectively. An algorithm for problem (3) is developed by alternately minimizing the following subproblem. The iteration steps are outlined as follows.

Remark 4. (i) For any $\alpha > 0$, the optimal solution of problem (7) has the following form $z^{k+1} = \text{sign}(y^k) \circ \max(0, |y^k| - \frac{1}{\alpha})$, where “ $\circ$ ” is the Hadamard product.

(i) Let $H^k = U^k - \frac{1}{\gamma_{1,k}} \nabla_U G(U^k, V^k, z^{k+1})$ and $S^k = V^k - \frac{1}{\gamma_{2,k}} \nabla_V G(U^{k+1}, V^k, z^{k+1})$. Then, the columns of U^{k+1} and V^{k+1} have the following closed form:

$$\begin{aligned} U_i^{k+1} &= \text{sign}[\max(0, \|H_i^k\| - \sqrt{2\lambda\gamma_{1,k}^{-1}})] H_i^k \quad \text{for } i = 1, \dots, r, \\ V_i^{k+1} &= \text{sign}[\max(0, \|S_i^k\| - \sqrt{2\lambda\gamma_{2,k}^{-1}})] S_i^k \quad \text{for } i = 1, \dots, r. \end{aligned}$$

For each $k \in \mathbb{N}$, write $w^k = (U^k, V^k, z^k)$. To establish the convergence of the sequence $\{w^k\}_{k \in \mathbb{N}}$, the following proposition is required.

Proposition 3. Let $\{w^k\}_{k \in \mathbb{N}}$ be the sequence generated by Algorithm 1. Then, the following statements hold.

Algorithm 1 (PALM Method for Solving (3))

- 1: **Input:** Sampling operator $\mathcal{A}$ and observation vector $b \in \mathbb{R}^m$
- 2: **Initialization:** Choose an initial $(U^0, V^0) \in \mathbb{R}^{n_1 \times \kappa} \times \mathbb{R}^{n_2 \times \kappa}$, set $\lambda, \mu > 0$
- 3: **while** stopping conditions are not satisfied, **do**
- 4: Compute $y^k = \mathcal{A}(U^k(V^k)^\top) - b$
- 5: Solve for z^{k+1} :

$$z^{k+1} = \arg \min_{z \in \mathbb{R}^m} \frac{\alpha}{2} \|y^k - z\|^2 + \|z\|_1. \quad (7)$$

- 6: Set $\gamma_{1,k} \geq \tau_{V^k} + \mu$, solve for U^{k+1} :

$$U^{k+1} \in \arg \min_{U \in \mathbb{R}^{n_1 \times \kappa}} \left\{ \langle \nabla_U G(U^k, V^k, z^{k+1}), U \rangle + \lambda \|U\|_{2,0} + \frac{\gamma_{1,k}}{2} \|U - U^k\|_F^2 \right\}$$

- 7: Set $\gamma_{2,k} \geq \tau_{U^{k+1}} + \mu$, solve for V^{k+1} :

$$V^{k+1} \in \arg \min_{V \in \mathbb{R}^{n_2 \times \kappa}} \left\{ \langle \nabla_V G(U^{k+1}, V^k, z^{k+1}), V \rangle + \lambda \|V\|_{2,0} + \frac{\gamma_{2,k}}{2} \|V - V^k\|_F^2 \right\}$$

- 8: **end while**

- 9: **Output:** $X = U^{k+1}(V^{k+1})^\top$
-

- (i) For each $k \in \mathbb{N}$, it holds that $\Theta(w^{k+1}) < \Theta(w^k) - \frac{\min\{\alpha, \mu\}}{2} \|w^k - w^{k+1}\|_F^2$. Hence, the sequence $\{\Theta(w^k)\}_{k \in \mathbb{N}}$ is convergent.
- (ii) For each $k \in \mathbb{N}$, $(A_1^{k+1}, A_2^{k+1}, A_3^{k+1}) \in \partial\Theta(w^{k+1})$ with

$$\begin{cases} A_1^{k+1} := \nabla_U G(w^{k+1}) - \nabla_U G(U^k, V^k, z^{k+1}) - \gamma_{1,k}(U^{k+1} - U^k); \\ A_2^{k+1} := \nabla_V G(w^{k+1}) - \nabla_V G(U^{k+1}, V^k, z^{k+1}) - \gamma_{2,k}(V^{k+1} - V^k); \\ A_3^{k+1} := \alpha(y^k - y^{k+1}). \end{cases}$$

Assume that $\{w^k\}_{k \in \mathbb{N}}$ is bounded. Then, there exists constant $c > 0$, such that

$$\text{dist}(0, \partial\Theta(w^{k+1})) \leq c \|w^k - w^{k+1}\|_F.$$

Proof. (i) By the optimality of z^{k+1} and the strong convexity of the subproblem (7) in Algorithm 1, one obtains

$$\|z^{k+1}\|_1 + \frac{\alpha}{2} \|y^k - z^{k+1}\|^2 + \frac{\alpha}{2} \|z^k - z^{k+1}\|^2 \leq \|z^k\|_1 + \frac{\alpha}{2} \|y^k - z^k\|^2. \quad (8)$$

Similar to U^{k+1} and V^{k+1} , one has

$$\begin{aligned} \langle \nabla_U G(U^k, V^k, z^{k+1}), U^{k+1} \rangle + \lambda \|U^{k+1}\|_{2,0} + \frac{\gamma_{1,k}}{2} \|U^{k+1} - U^k\|_F^2 \\ \leq \langle \nabla_U G(U^k, V^k, z^{k+1}), U^k \rangle + \lambda \|U^k\|_{2,0}; \\ \langle \nabla_V G(U^{k+1}, V^k, z^{k+1}), V^{k+1} \rangle + \lambda \|V^{k+1}\|_{2,0} + \frac{\gamma_{2,k}}{2} \|V^{k+1} - V^k\|_F^2 \\ \leq \langle \nabla_V G(U^{k+1}, V^k, z^{k+1}), V^k \rangle + \lambda \|V^k\|_{2,0}. \end{aligned}$$

The above two inequalities imply that

$$\begin{aligned} \lambda \|U^k\|_{2,0} + \lambda \|V^k\|_{2,0} - \frac{\gamma_{1,k}}{2} \|U^{k+1} - U^k\|_F^2 - \frac{\gamma_{2,k}}{2} \|V^{k+1} - V^k\|_F^2 \\ \geq \langle \nabla_U G(U^k, V^k, z^{k+1}), U^{k+1} - U^k \rangle + \lambda \|U^{k+1}\|_{2,0} + \lambda \|V^{k+1}\|_{2,0} \\ + \langle \nabla_V G(U^{k+1}, V^k, z^{k+1}), V^{k+1} - V^k \rangle. \end{aligned} \quad (9)$$

By the Lipschitz continuity of $\nabla_U G(\cdot, V, z)$ and $\nabla_V G(U, \cdot, z)$, one has

$$\begin{cases} G(U^{k+1}, V^k, z^{k+1}) \leq G(U^k, V^k, z^{k+1}) + \langle \nabla_U G(U^k, V^k, z^{k+1}), U^{k+1} - U^k \rangle \\ \quad + \frac{\tau_{V^k}}{2} \|U^{k+1} - U^k\|_F^2; \\ G(U^{k+1}, V^{k+1}, z^{k+1}) \leq G(U^{k+1}, V^k, z^{k+1}) + \langle \nabla_V G(U^{k+1}, V^k, z^{k+1}), V^{k+1} - V^k \rangle \\ \quad + \frac{\tau_{U^{k+1}}}{2} \|V^{k+1} - V^k\|_F^2. \end{cases}$$

Along with inequality (9), it immediately holds that

$$\begin{aligned} G(w^{k+1}) + \lambda \|U^{k+1}\|_{2,0} + \lambda \|V^{k+1}\|_{2,0} \leq G(U^k, V^k, z^{k+1}) + \lambda \|U^k\|_{2,0} + \lambda \|V^k\|_{2,0} \\ - \frac{\mu}{2} \|U^{k+1} - U^k\|_F^2 - \frac{\mu}{2} \|V^{k+1} - V^k\|_F^2. \end{aligned}$$

From the last inequality, the definition of Θ , the lower boundedness of Θ , and inequality (8), it follows that part (i) holds.

(ii) Under the optimal conditions of U^{k+1} , V^{k+1} and z^{k+1} , one has

$$\begin{cases} 0 \in \alpha(z^{k+1} - y^k) + \partial \|\cdot\|_1(z^{k+1}); \\ 0 \in \nabla_U G(U^k, V^k, z^{k+1}) + \gamma_{1,k}(U^{k+1} - U^k) + \lambda \partial \|\cdot\|_{2,0}(U^{k+1}); \\ 0 \in \nabla_V G(U^{k+1}, V^k, z^{k+1}) + \gamma_{2,k}(V^{k+1} - V^k) + \lambda \partial \|\cdot\|_{2,0}(V^{k+1}). \end{cases}$$

Then, by Lemma 2, it is not hard to obtain

$$(A_1^{k+1}, A_2^{k+1}, A_3^{k+1}) \in \partial \Theta(w^{k+1}).$$

Hence, $\text{dist}(0, \partial \Theta(w^{k+1})) \leq \|(A_1^{k+1}, A_2^{k+1}, A_3^{k+1})\|_F$. By the boundedness of $\{w^k\}_{k \in \mathbb{N}}$ and the Lipschitz continuity of $\nabla_U G(\cdot, V, z)$ and $\nabla_V G(U, \cdot, z)$, the desired result is obtained. $\square$

From [[15], Section 4], Θ is a KL function. By Proposition 3 and using the same reasoning as in [[17], Theorem 1], the main convergence results are established.

Theorem 1. Suppose that the sequence $\{w^k\}_{k \in \mathbb{N}}$ generated by Algorithm 1 is bounded. Then, the sequence $\{w^k\}_{k \in \mathbb{N}}$ is convergent and its limit $(\tilde{U}, \tilde{V}, \tilde{z})$ is a critical point of (3). If the nonzero elements of $|\mathcal{A}(\tilde{U}\tilde{V}^\top) - b|$ are not smaller than $\frac{1}{\alpha}$, then $(\tilde{U}, \tilde{V})$ is a critical point of Φ .

5. Numerical Experiments

The efficiency of Algorithm 1 is validated by solving matrix completion problems with outliers under uniform sampling, and its performance is compared with that of SubGM from [7] for solving problem (1). All numerical tests are conducted in MATLAB 2024b on a laptop computer running on a 64-bit Windows Operating System with an Intel(R) Core(TM) i9-13905H CPU 2.60GHz and 32 GB RAM (Intel, Santa Clara, CA, USA).

5.1. Implementation Details of Algorithms

Matrix completion problems involving outliers are considered. To generate the sampling operator $\mathcal{A}$, a random index set $\Omega = \{(i_t, j_t) \mid t = 1, \dots, m\}$ is assumed, with indices sampled independently from a uniform distribution. The mapping $\mathcal{A}$ is then defined by $\mathcal{A}(X) := (X_{i_1, j_1}, X_{i_2, j_2}, \dots, X_{i_m, j_m})^\top$ for $X \in \mathbb{R}^{n_1 \times n_2}$, and $b = \mathcal{A}(M_\Omega)$ where $M_\Omega \in \mathbb{R}^{n_1 \times n_2}$, with

$$[M_\Omega]_{i_t, j_t} = \begin{cases} 0 & \text{if } (i_t, j_t) \notin \Omega, \\ M_{i_t, j_t}^* + \omega_t & \text{if } (i_t, j_t) \in \Omega \end{cases} \quad \text{for } t = 1, 2, \dots, m. \quad (10)$$

The true matrix $M^* \in \mathbb{R}^{n_1 \times n_2}$ and the sparse noisy vector $\omega = (\omega_1, \dots, \omega_m)^\top$ are considered. The true matrix $M^* = M_L^*(M_R^*)^\top \in \mathbb{R}^{n_1 \times n_2}$ of rank r^* is generated by independently sampling entries of $M_L^* \in \mathbb{R}^{n_1 \times r^*}$ and $M_R^* \in \mathbb{R}^{n_2 \times r^*}$ from the standard normal distribution $\mathcal{N}(0, 1)$. The nonzero entries of the sparse noisy vector $\omega = (\omega_1, \dots, \omega_m)^\top$ follow one of two distributions: (i) Student's t-distribution with four degrees of freedom, scaled by $\sqrt{2}$, or (ii) Laplace distribution with density $d(u) = 0.5 \exp(-|u|)$. Unless otherwise specified, the number of nonzero entries in ω is set to $\lfloor 0.3m \rfloor$ for all tests. The parameters are set as $\mu = 10^{-8}$ and $\lambda = c_\lambda \|b\|$, where $c_\lambda > 0$ is specified in the experiments. Unless otherwise stated, it takes $\kappa = \min(n_1, n_2, 150)$. The initial point (U^0, V^0) of Algorithm 1 is defined as $(P_1 \Sigma_\kappa(M_\Omega)^{1/2}, Q_1 \Sigma_\kappa(M_\Omega)^{1/2})$, with P_1 and Q_1 denoting the matrices consisting of the first κ left and right singular vectors of M_Ω , respectively. For SubGM, as in [7], the Polyak step size rule [18] is employed. Since the optimal value of (1) is typically unknown in practice, the step size $\frac{\Psi(U^k, V^k) - \min_{(U, V)} \Psi(U, V)}{\|\zeta^k\|_F^2}$ is replaced with $\frac{0.05 \Psi(U^k, V^k)}{\|\zeta^k\|_F^2}$, where $\zeta^k \in \partial \Psi(U^k, V^k)$ for $k \in \mathbb{N}$.

Algorithm 1 is terminated at iteration $w^k = (U^k, V^k, z^k)$, when $k \geq k_{\max}$ or one of the following conditions is satisfied:

$$\frac{\max_{j \in \{1, \dots, 19\}} |\Theta(w^k) - \Theta(w^{k-j})|}{\max\{1, \Theta(w^k)\}} \leq \epsilon_1 \quad \text{or} \quad \frac{R_1^k + R_2^k + R_3^k}{1 + \|b\|} \leq \epsilon_2 \quad \text{for } k \geq 30,$$

where

$$\begin{cases} R_1^k = \alpha \|y^{k-1} - y^k\|; \\ R_2^k = \|\nabla_U G(U^{k-1}, V^{k-1}, z^k) - \nabla_U G(w^k) + \gamma_{1,k-1}(U^{k-1} - U^k)\|_F; \\ R_3^k = \|\nabla_V G(U^k, V^{k-1}, z^k) - \nabla_V G(w^k) + \gamma_{2,k-1}(V^{k-1} - V^k)\|_F. \end{cases}$$

For a fair comparison, the initial point of SubGM is set to be identical to that of Algorithm 1. SubGM is terminated at iteration w^k , whenever $k > k_{\max}$ or either of the following conditions is satisfied: $\frac{\max_{j \in \{1, \dots, 9\}} |\Psi(U^k, V^k) - \Psi(U^{k-j}, V^{k-j})|}{\max\{1, \Psi(U^k, V^k)\}} \leq \epsilon_3$ or

$$\frac{\|U^{k+1}(V^{k+1})^\top - U^k(V^k)^\top\|_F}{1 + \|U^{k+1}(V^{k+1})^\top\|_F} \leq \epsilon_4 \quad \text{with } k \geq 200. \quad (11)$$

Unless otherwise stated, $\epsilon_1 = 10^{-4}$, $\epsilon_2 = 10^{-5}$ and $k_{\max} = 1000$ are set for Algorithm 1, while $\epsilon_3 = 5 \times 10^{-4}$, $\epsilon_4 = 5 \times 10^{-4}$, and $k_{\max} = 500$ are set for SubGM.

The matrix recovery performance is evaluated using the relative error (RE), defined by

$$\text{RE} := \frac{\|X^{\text{out}} - M^*\|_F}{\|M^*\|_F},$$

where $X^{\text{out}} = U^{\text{out}}(V^{\text{out}})^\top$ represents the output of a solver. All test results are averaged over 5 instances of each experiment.

5.2. Parameter Sensitivity Analysis

In this section, the effect of the parameters on the performance of Algorithm 1 is evaluated. Figure 1 illustrates the curves of RE, recovered rank, and running time (in seconds) for Algorithm 1 when solving problem (3) with $n_1 = n_2 = 1000$, $r^* = 5$, under Case I and Case II. As shown in Figure 1, for $2 \leq \alpha \leq 8.5$, Algorithm 1 yields both the exact rank recovery and the lowest relative error in both cases. Values of α exceeding 8.5 degrade the algorithm's performance. The RE increases significantly, accompanied by an increase in computational time. In view of this, α is set to 5 for all subsequent experiments.

Under the settings $n_1 = n_2 = 1000$, $r^* = 5$, $\kappa = 10r^*$, and $\text{SR} = 0.2$, Figure 2 illustrates the RE, recovered rank, and running time curves of Algorithm 1 for varying λ . The results show that a specific interval of λ values (e.g., $8 \leq c_\lambda \leq 24$) achieves low relative errors while accurately recovering the true rank of matrix M^* .

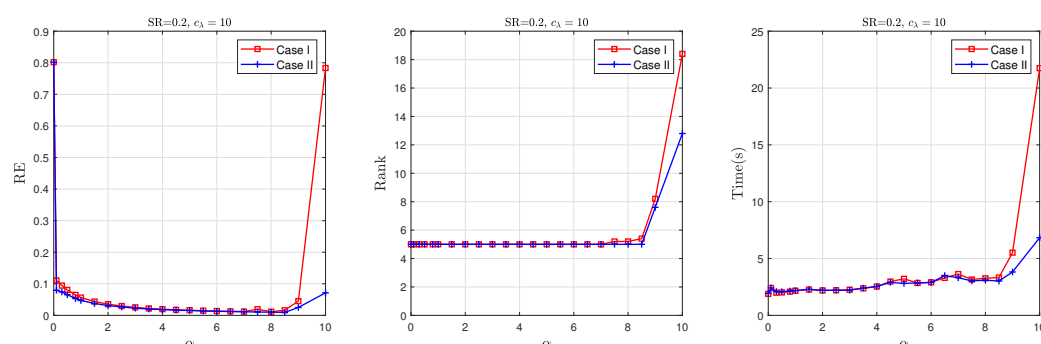


Figure 1. Curves of RE, rank, and running time of Algorithm 1 with different α values.

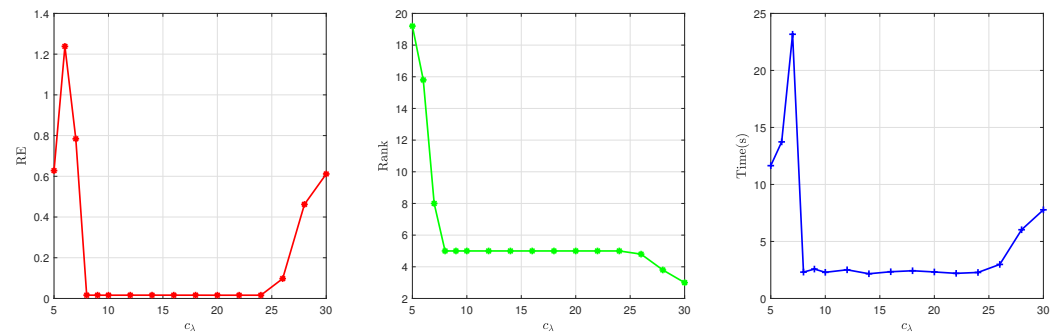


Figure 2. Curves of relative error, rank, and running time of Algorithm 1 in Case I.

For $n := n_1 = n_2$ and $r^* = 10$, Figure 3 shows the variation in running time and RE as n increases. These results indicate that the computational time remains below 160 s when $n \leq 7000$ and does not exceed 400 s even when $n = 10^4$. With increasing n , the RE decreases rapidly, remaining below 0.015 for $n \geq 3000$.

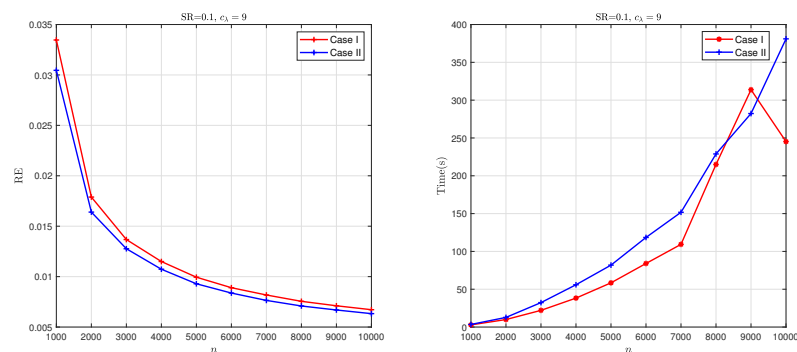


Figure 3. RE and running time curves of Algorithm 1 in Case I-II with different values of n .

For Figures 4 and 5, the settings are $n_1 = n_2 = 1000$ and $r^* = 5$. Figure 4 displays the average RE across five repetitions for sampling ratios or SRs $= \{0.1, 0.15, 0.2, \dots, 0.5\}$. Under uniform sampling, the RE decreases monotonically with increasing SR, while the running time consistently stays within 5 s.

In Figure 5, the variation in RE with the ratio of nonzero entries in the noise vector (NZR) is examined. NZR is set to $\{0.1, 0.15, 0.18, \dots, 0.5\}$. As shown in Figure 5, the RE increases accordingly with NZR, while the running time remains consistently lower than 4 s. Although the difficulty of the problem escalates as NZR increases, the RE remains close to 0.02, even when NZR reaches 0.5.

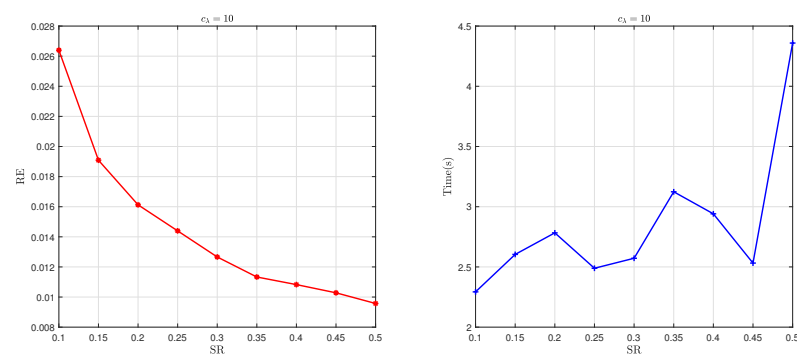


Figure 4. RE and running time curves of Algorithm 1 in Case I with different SRs.

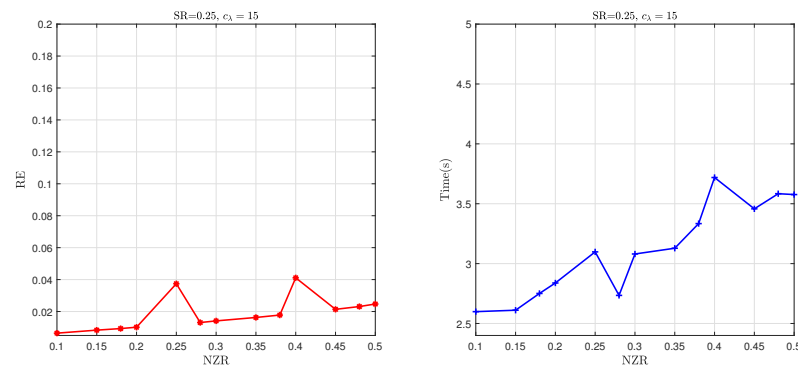


Figure 5. RE and running time curves of Algorithm 1 in Case I with different NZRs.

5.3. Numerical Comparisons with SubGM

The performance of Algorithm 1 in terms of solution quality and running time is compared with that of SubGM. Since the model (1) does not promote low-rank structures, $\kappa = 3r^*$ and $n_1 = n_2 = n$ are adopted for the following tests. The numerical results of the two methods are obtained under the same stopping criterion described in Section 5.1, with $\epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon_4 = 5 \times 10^{-4}$, and $k_{\max} = 500$. Tables 1 and 2 report the average results across five randomly generated instances for each setting. Compared with SubGM, Algorithm 1 consistently achieves the lowest RE in significantly less computational time.

Table 1. The results of the algorithm for synthetic data in Case I.

n	(r^*, SR)	c_λ	Algorithm 1			SubGM		
			RE	Rank	Time (s)	RE	Rank	Time (s)
1000	(5, 0.10)	10	2.61×10^{-2}	5	1.50	1.26×10^{-1}	15	4.04
	(5, 0.15)	10	1.94×10^{-2}	5	1.48	7.71×10^{-2}	15	4.71
	(5, 0.20)	10	1.59×10^{-2}	5	1.60	6.13×10^{-2}	15	5.28
	(5, 0.20)	10	1.40×10^{-2}	5	1.73	4.93×10^{-2}	15	5.56
	(10, 0.10)	6	3.34×10^{-2}	10	1.88	2.58×10^{-1}	30	4.35
	(10, 0.15)	6	2.26×10^{-2}	10	2.88	1.38×10^{-1}	30	5.26
	(10, 0.20)	6	2.15×10^{-2}	10	2.34	8.20×10^{-2}	30	5.92
	(10, 0.20)	6	1.54×10^{-2}	10	2.33	5.88×10^{-2}	30	6.26
3000	(10, 0.10)	6	1.36×10^{-2}	10	15.0	3.61×10^{-2}	30	40.6
	(10, 0.15)	6	1.05×10^{-2}	10	17.1	2.91×10^{-2}	30	50.6
	(10, 0.20)	6	8.95×10^{-3}	10	18.3	2.23×10^{-2}	30	51.4
	(10, 0.20)	6	1.24×10^{-2}	10	20.5	2.64×10^{-2}	30	56.3
	(20, 0.10)	6	1.57×10^{-2}	20	20.3	7.73×10^{-2}	60	44.1
	(20, 0.15)	6	1.15×10^{-2}	20	20.0	3.41×10^{-2}	60	49.6
	(20, 0.20)	6	9.52×10^{-3}	20	22.4	2.42×10^{-2}	60	55.5
	(20, 0.20)	6	8.33×10^{-3}	20	23.2	2.07×10^{-2}	60	61.6
5000	(10, 0.10)	10	9.92×10^{-3}	10	45.8	2.29×10^{-2}	30	140.9
	(10, 0.15)	10	7.87×10^{-3}	10	50.0	1.93×10^{-2}	30	138.8
	(10, 0.20)	10	6.73×10^{-3}	10	46.9	1.87×10^{-2}	30	141.4
	(10, 0.20)	10	6.01×10^{-3}	10	50.1	1.61×10^{-2}	30	157.7
8000	(10, 0.10)	10	7.63×10^{-3}	10	121.3	1.53×10^{-2}	30	312.1
	(10, 0.15)	10	8.21×10^{-3}	10	140.2	1.62×10^{-2}	30	387.8
	(10, 0.20)	10	5.21×10^{-3}	10	144.9	1.40×10^{-2}	30	432.9
	(10, 0.20)	10	4.71×10^{-3}	10	154.7	1.49×10^{-2}	30	475.7

Table 2. The results of the algorithm for synthetic data in Case II.

Algorithm 1						SubGM		
n	(r^*, SR)	c_λ	RE	Rank	Time (s)	RE	Rank	Time (s)
1000	(5, 0.10)	10	2.42×10^{-2}	5	1.37	8.76×10^{-2}	15	3.85
	(5, 0.15)	10	1.82×10^{-2}	5	1.43	6.14×10^{-2}	15	4.81
	(5, 0.20)	10	1.49×10^{-2}	5	1.39	5.03×10^{-2}	15	5.49
	(5, 0.20)	10	1.31×10^{-2}	5	1.28	4.61×10^{-2}	15	5.70
1000	(10, 0.10)	6	3.00×10^{-2}	10	1.86	1.85×10^{-1}	30	4.31
	(10, 0.15)	6	2.07×10^{-2}	10	1.88	9.23×10^{-2}	30	4.26
	(10, 0.20)	6	1.65×10^{-2}	10	2.24	6.02×10^{-2}	30	5.92
	(10, 0.20)	6	1.42×10^{-2}	10	2.28	4.71×10^{-2}	30	6.26
3000	(10, 0.10)	6	1.27×10^{-2}	10	15.5	3.11×10^{-1}	30	40.6
	(10, 0.15)	6	9.85×10^{-3}	10	16.1	2.71×10^{-2}	30	45.6
	(10, 0.20)	6	8.33×10^{-3}	10	17.3	2.27×10^{-2}	30	50.4
	(10, 0.20)	6	7.40×10^{-3}	10	18.5	2.11×10^{-2}	30	55.3
3000	(20, 0.10)	6	1.44×10^{-2}	20	20.3	5.30×10^{-2}	60	44.1
	(20, 0.15)	6	1.07×10^{-2}	20	22.0	2.79×10^{-2}	60	49.6
	(20, 0.20)	6	8.82×10^{-3}	20	22.4	2.11×10^{-2}	60	54.5
	(20, 0.20)	6	7.73×10^{-3}	20	23.2	1.85×10^{-2}	60	59.6
5000	(10, 0.10)	10	9.29×10^{-3}	10	40.8	2.07×10^{-2}	30	110.9
	(10, 0.15)	10	7.37×10^{-3}	10	48.0	1.84×10^{-2}	30	126.8
	(10, 0.20)	10	6.29×10^{-3}	10	46.9	1.75×10^{-2}	30	141.4
	(10, 0.20)	10	5.62×10^{-3}	10	50.1	1.61×10^{-2}	30	156.7
8000	(10, 0.10)	10	7.13×10^{-3}	10	99.3	1.51×10^{-2}	30	283.1
	(10, 0.15)	10	5.71×10^{-3}	10	114.2	1.56×10^{-2}	30	337.8
	(10, 0.20)	10	4.91×10^{-3}	10	114.9	1.50×10^{-2}	30	353.9
	(10, 0.20)	10	4.38×10^{-3}	10	124.7	1.48×10^{-2}	30	395.7

Reconstructing images from limited and noisy measurements is a key application of low-rank matrix completion. To further evaluate the effectiveness of the proposed method, experiments were conducted on images from the ZJU dataset [19]. Each image has dimensions of 300×300 . The initial points for both algorithms were obtained using the same procedure described in Section 5.1, with $\kappa = 40$.

The results are presented in Table 3 and Figure 6. The images were undersampled using various sampling rates and masks and then restored using a similar process reported in [9]. Table 3 lists the relative error and peak signal-to-noise ratio (PSNR) for each image restoration task. The results show that the proposed algorithm outperforms SubGM in terms of both computational time and restoration quality.

Table 3. Summary of experimental results of image recovery.

	Algorithm	RE	PSNR	Time (s)
Sampling Ratio = 50%	Algorithm 1	0.0708	30.57	1.72
	SubGM	0.0991	27.82	2.60
Sampling Ratio = 20%	Algorithm 1	0.1178	24.22	0.76
	SubGM	0.1361	23.07	1.84
Sampling Ratio = 10%	Algorithm 1	0.0369	33.61	0.59
	SubGM	0.1113	24.06	1.49
Image with text mask	Algorithm 1	0.1639	21.10	0.48
	SubGM	0.1700	20.82	4.84
Image with cross mask	Algorithm 1	0.0934	16.71	1.34
	SubGM	0.1811	7.40	5.67

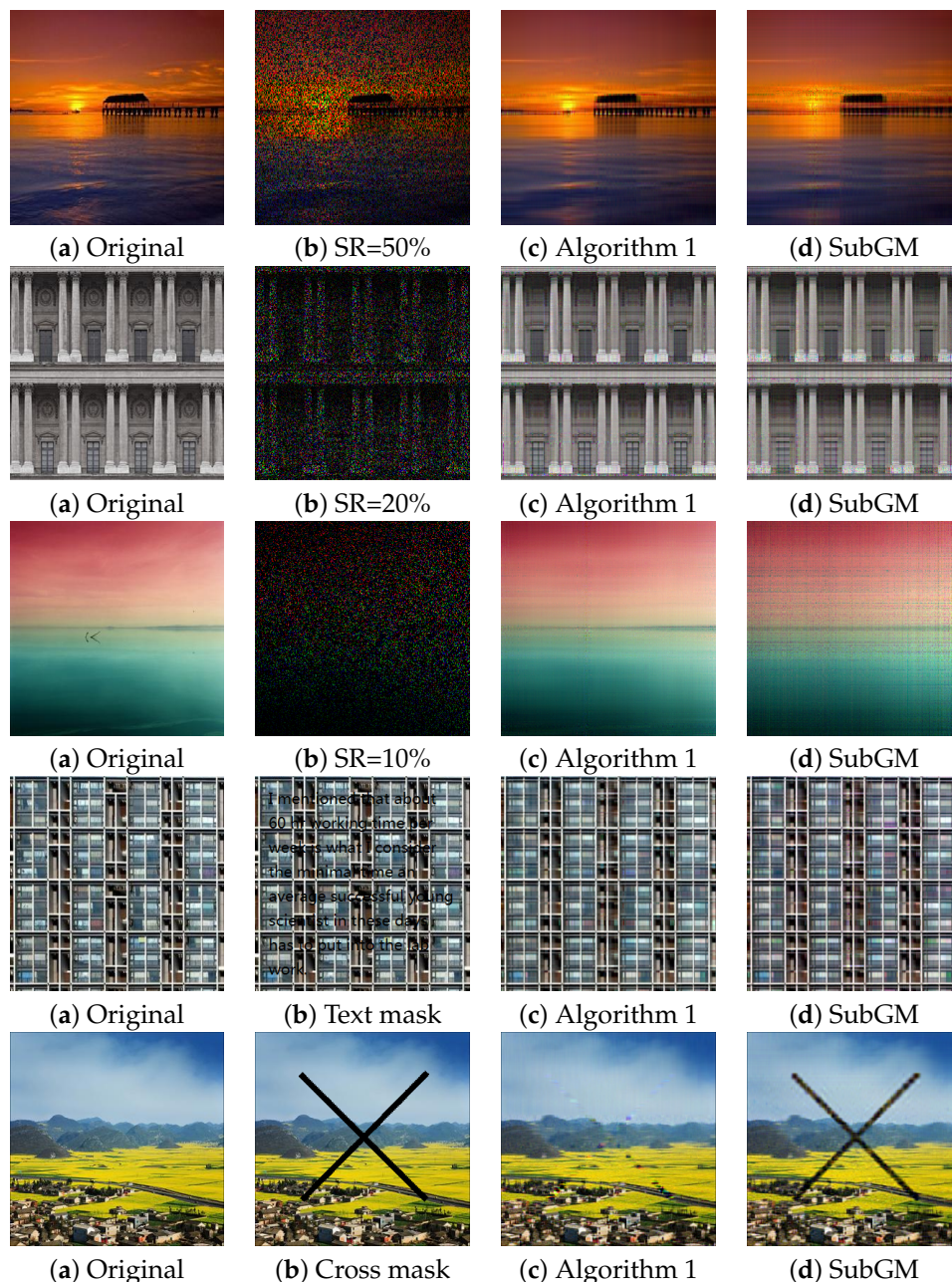


Figure 6. Experimental results of image recovery.

6. Conclusions

This paper has focused on robust factorization models for low-rank matrix recovery, which are of great significance in multiple fields such as machine learning and imaging sciences. To deal with the challenge of outliers, an ℓ_1 -loss robust factorized model with $\ell_{2,0}$ -norm regularization is proposed. Given the nonconvex and discontinuous nature of the problem, an approximation problem (3) is constructed. Under mild assumptions, the equivalence of stationary points between the original and the approximation problems is verified, and it is proven that the global optimum of the approximation problem can serve as an approximate optimal solution for the original problem in terms of objective value. On top of that, a fast PALM method is proposed to solve the approximation problem, and the global convergence of the iterate sequence is established. The numerical experiments on synthetic datasets with outliers and image restoration tasks demonstrate that the PALM method achieves low relative errors within a significantly shorter computational time, particularly when dealing with large-scale datasets. Despite the promising results of the

proposed approach, there are some limitations. The performance of the algorithm relies on the proper selection of parameters such as λ and α . To address these limitations, future research works will include the development of adaptive parameter-tuning strategies.

Author Contributions: Methodology, T.T. and J.Z.; Validation, L.X.; Formal analysis, T.T. and L.X.; Investigation, J.Z.; Writing—original draft, T.T., L.X. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: T.T. was supported by the Guangdong Basic and Applied Basic Research Foundation No. 2023A1515111167. L.X. was supported by the Guangdong Basic and Applied Basic Research Foundation No. 2022A1515110959. J.Z. was funded by the National Natural Science Foundation of China No. 12401630, the Educational Commission of Guangdong Province No. 2023KQNCX073, and the Natural Science Foundation of Guangdong Province No. 2023A1515110558.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Candès, E.J.; Recht, B. Exact matrix completion via convex optimization. *Found. Comput. Math.* **2009**, *9*, 717–772. [\[CrossRef\]](#)
2. Davenport, M.A.; Romberg, J. An overview of low-rank matrix recovery from incomplete observations. *IEEE J. Sel. Top. Signal Process.* **2016**, *10*, 608–622. [\[CrossRef\]](#)
3. Fazel, M.; Hindi, H.; Boyd, S. Rank Minimization and Applications in System Theory. *Proc. 2004 Am. Control Conf.* **2004**, *4*, 3273–3278.
4. Gross, D.; Liu, Y.K.; Flammia, S.T.; Becker, S.; Eisert, J. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **2010**, *105*, 150401. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Negahban, S.; Wainwright, M.J. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Stat.* **2011**, *39*, 1069–1097. [\[CrossRef\]](#)
6. Charisopoulos, V.; Chen, Y.; Davis, D.; Díaz, M.; Ding, L.; Drusvyatskiy, D. Low-Rank Matrix Recovery with Composite Optimization: Good Conditioning and Rapid Convergence. *Found. Comput. Math.* **2021**, *21*, 1505–1593. [\[CrossRef\]](#)
7. Li, X.; Zhu, Z.H.; So, A.M.; Vidal, R. Nonconvex Robust Low-Rank Matrix Recovery. *SIAM J. Optim.* **2020**, *30*, 660–686. [\[CrossRef\]](#)
8. Ma, J.H.; Fattahi, S. Global Convergence of Sub-gradient Method for Robust Matrix Recovery: Small Initialization, Noisy Measurements, and Over-parameterization. *J. Mach. Learn. Res.* **2023**, *24*, 1–84.
9. Wang, Z.Y.; So, H.C.; Liu, Z.F. Fast and robust rank-one matrix completion via maximum correntropy criterion and half-quadratic optimization. *Signal Process.* **2022**, *198*, 108580. [\[CrossRef\]](#)
10. Tao, T.; Qian, Y.T.; Pan, S.H. Column $\ell_{2,0}$ -norm regularized factorization model of low-rank matrix recovery and its computation. *SIAM J. Optim.* **2022**, *32*, 959–988. [\[CrossRef\]](#)
11. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis. *J. ACM* **2011**, *11*, 1–37. [\[CrossRef\]](#)
12. Josz, C.; Ouyang, Y.; Zhang, R.; Lavaei, J.; Sojoudi, S. A theory on the absence of spurious solutions for nonconvex and nonsmooth optimization. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 2441–2449.
13. Li, Y.; Sun, Y.; Chi, Y. Low-rank positive semidefinite matrix recovery from corrupted rank-one measurements. *IEEE Trans. Signal Process.* **2017**, *65*, 397–408. [\[CrossRef\]](#)
14. Rockafellar, R.T.; Wets, R.J.-B. *Variational Analysis*; Springer: Berlin/Heidelberg, Germany, 1998.
15. Attouch, H.; Bolte, J.; Redont, P.; Soubeyran, A. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res.* **2010**, *35*, 438–457. [\[CrossRef\]](#)
16. Nocedal, J.; Wright, S.J. *Numerical Optimization*; Springer: Berlin/Heidelberg, Germany, 2000.
17. Bolte, J.; Sabach, S.; Teboulle, M. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.* **2014**, *146*, 459–494. [\[CrossRef\]](#)
18. Polyak, B.T. Minimization of unsmooth functions. *USSR Comput. Math. Math. Phys.* **1969**, *9*, 14–29. [\[CrossRef\]](#)
19. Hu, Y.; Zhang, D.; Ye, J.; Li, X.; He, X. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 2117–2130. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.