

## Article

# A New Proximal Iteratively Reweighted Nuclear Norm Method for Nonconvex Nonsmooth Optimization Problems

Zhili Ge <sup>1</sup>, Siyu Zhang <sup>2</sup>, Xin Zhang <sup>3,4,\*</sup> and Yan Cui <sup>5</sup><sup>1</sup> School of Mathematical Sciences, Nanjing Normal University of Special Education, Nanjing 210038, China; 250053@njts.edu.cn<sup>2</sup> School of Microelectronics and Data Science, Anhui University of Technology, Ma'anshan 243032, China; zhangsiyu@ahut.edu.cn<sup>3</sup> School of Mathematics and Physics, Suqian University, Suqian 223800, China<sup>4</sup> Key Laboratory of Numerical Simulation for Large Scale Complex Systems, Ministry of Education, Nanjing 210023, China<sup>5</sup> School of Artificial Intelligence, Nanjing Normal University of Special Education, Nanjing 210038, China; csyanncui@njts.edu.cn

\* Correspondence: 23106@squ.edu.cn

## Abstract

This paper proposes a new proximal iteratively reweighted nuclear norm method for a class of nonconvex and nonsmooth optimization problems. The primary contribution of this work is the incorporation of line search technique based on dimensionality reduction and extrapolation. This strategy overcomes parameter constraints by enabling adaptive dynamic adjustment of the extrapolation/proximal parameters  $(\alpha_k, \beta_k, \mu_k)$ . Under the Kurdyka–Łojasiewicz framework for nonconvex and nonsmooth optimization, we prove the global convergence and linear convergence rate of the proposed algorithm. Additionally, through numerical experiments using synthetic and real data in matrix completion problems, we validate the superior performance of the proposed method over well-known methods.

**Keywords:** nonconvex and nonsmooth optimization; proximal iteratively reweighted method; line search; convergence analysis

**MSC:** 49M37; 65K05; 90C26



Academic Editor: Andrea Scozzari

Received: 9 July 2025

Revised: 13 August 2025

Accepted: 13 August 2025

Published: 16 August 2025

**Citation:** Ge, Z.; Zhang, S.; Zhang, X.; Cui, Y. A New Proximal Iteratively Reweighted Nuclear Norm Method for Nonconvex Nonsmooth Optimization Problems. *Mathematics* **2025**, *13*, 2630. <https://doi.org/10.3390/math13162630>

**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

### 1.1. Problem Description

This work addresses a nonconvex and nonsmooth optimization problem within the real matrix space  $\mathbb{R}^{m \times n}$  ( $m \leq n$ )

$$\min_X \Psi(X) := f(X) + \sum_{i=1}^m g(\sigma_i(X)), \quad (1)$$

where  $\sigma_i(X)$  denotes the  $i$ -th singular value of  $X$ ,  $f$  is differentiable and the gradient is Lipschitz continuous with constant  $L_f$ ,  $g$  is differentiable concave and the gradient is Lipschitz continuous with constant  $L_g$  and  $g'(t) > 0$  for any  $t \in [0, +\infty)$ .

It is easy to see that  $\sum_{i=1}^m g(\sigma_i(X))$  is nonconvex and nonsmooth due to the nonsmoothness of  $\sigma_i(X)$  and the concavity of  $g$ . Thus, the overall function  $\Psi$  is nonconvex and nonsmooth even though  $f$  is differentiable (may be nonconvex). Note the generality of

problem (1), it has a wide applications, such as image processing [1], machine learning [2] and multiple category classification [3]. To illustrate this point, consider the well-known image recovery problem. In such a scenario,  $f(X) = \frac{1}{2} \|A(X) - b\|^2$  ( $A$  is a linear operator and  $b$  is a vector or matrix) generally represents the quadratic loss function, which is used to measure recovery performance. Consequently,  $f$  is always differentiable. On the other hand,  $\sum_{i=1}^m g(\sigma_i(X))$  is a nonconvex regularized term that is employed to obtain a low rank solution. Some common nonconvex regularized terms, including  $L_p$ , Log, ETP, Geman and Laplace can be found in [1,4]. The validity of the assumption of  $g$  can be verified through the nonnegativity of its second order derivatives and the median theorem.

### 1.2. Related Work

It is precisely because of the popularity and scope of problem (1) that there is a lot of related work, as can be seen in [5–13]. One of the more competitive methods is the well-known General Iterative Shrinkage and Thresholding (GIST) algorithm [14,15]. Applying the GIST algorithm to solve (1) needs to compute the proximal operator of a DC function  $\sum_{i=1}^m g(\sigma_i(X))$ . Unfortunately, this assumption of  $\sum_{i=1}^m g(\sigma_i(X))$  is less likely to be satisfied, since the DC decomposition of  $\sum_{i=1}^m g(\sigma_i(X))$  is not known in general. Thus, based on the key fact that nonnegativity and monotone decrease of  $\nabla g$ , Lu et al. [4] proposed Proximal Iteratively Reweighed Nuclear Norm algorithm (PIRNN). Sun et al. [16] refined the related convergence conclusions. Later, Ge et al. [1] gave the PIRNN with a more general Extrapolation (PIRNNE) and proved the convergence under the same assumptions. The concrete iterative scheme can be read as

$$\begin{cases} Y^k := X^k + \alpha_k(X^k - X^{k-1}), \\ Z^k := X^k + \beta_k(X^k - X^{k-1}), \\ X^{k+1} := \text{prox}_{\sum_{i=1}^m \mu_k w_i^k \sigma_i(X)}(Y^k - \mu_k \nabla f(Z^k)), \end{cases} \quad (2)$$

$$Z^k := X^k + \beta_k(X^k - X^{k-1}), \quad (3)$$

$$X^{k+1} := \text{prox}_{\sum_{i=1}^m \mu_k w_i^k \sigma_i(X)}(Y^k - \mu_k \nabla f(Z^k)), \quad (4)$$

where  $\alpha_k \in [0, 1)$ ,  $\beta_k \in [0, 1]$  are the extrapolation stepsizes,  $\{\mu_k\}$  is a nondecreasing parameter sequence,  $w_i^k := g'(\sigma_i(X^k))$  and for any  $Y \in \mathbb{R}^{m \times n}$ ,

$$\text{prox}_{\sum_{i=1}^m \mu_k w_i^k \sigma_i(X)}(Y) := \arg \min_X \left\{ \sum_{i=1}^m w_i^k \sigma_i(X) + \frac{1}{2\mu_k} \|X - Y\|_F^2 \right\}. \quad (5)$$

(4) has a closed-form solution if  $0 \leq w_1^k \leq w_2^k \leq \dots \leq w_m^k$ . In other words, for any  $Y \in \mathbb{R}^{m \times n}$ , one has  $US(\Lambda)V^\top \in \text{prox}_{\sum_{i=1}^m \mu_k w_i^k \sigma_i(X)}(Y)$ , where  $U\Lambda V^\top$  is the SVD of  $Y$ ,  $S(\Lambda) = \text{diag}\{(\Lambda_{i,i} - \mu_k w_i^k)_+\}_{1 \leq i \leq m}$  with  $(a)_+ = \max\{a, 0\}$  for any  $a \in \mathbb{R}$ .

Meanwhile, Phan et al. [17] devised an acceleration framework utilizing the partial singular value decomposition of reduced-dimensional matrices rather than full matrices, conditioned upon parameter specifications with  $\alpha_k = 0$ ,  $\beta_k = \frac{k}{k+3}$  and  $\mu_k = \mu$ . Xu et al. [18] integrated rank estimation via enhanced Gerschgorin disk analysis with learnable sub-matrix recovery, demonstrating state-of-the-art performance. Separately, Wen et al. [19] formulated an alternative accelerated matrix completion methodology employing continuation protocols and randomized truncated SVD, parameterized by  $\alpha_k = \beta_k = 0$  and  $\mu_{k+1} = \max\{\eta\mu_k, \mu_{\min}\}$ ,  $\eta < 1$ . Generalized framework [20] leveraged ADMM for nonconvex nonsmooth low-rank recovery with rigorous convergence guarantees. Some recent methods also combine other regularization techniques. Alternative regularization strategies included an image reconstruction factorization model using a total variation regularizer [21], a truncated error model using the difference between the nuclear norm and

Frobenius norm for impulse noise processing [22], and an accelerated iterative reweighted nuclear norm method combined with active manifold identification [23]. This paper focuses on the efficient computation of (5) based on PIRNNE. To the best of our knowledge, suitable parameter selection makes these algorithms have good numerical performance. The most famous optimal parameter choice is the Nesterov's acceleration, such as FISTA [24] and APG [25]. The optimal choice involving the inertial and proximal parameters of PIRNNE for the nonconvex nonsmooth problems considered in this paper are not explicit, which makes the algorithm unable to maintain its advantage. Whether there is an adaptive parameter selection is our concern.

### 1.3. Our Contribution

Fortunately, the line search strategy is widely used for nonconvex vector optimization to overcome restrictions on the involved parameters [26–31]. This strategy allows the parameters to be chosen initially with some aggressive values that are not below a specific threshold, and then updates adaptively parameters at each iteration according to the line search criterion, which can improve the numerical performance in implementation. A natural approach is to incorporate the line search strategy to the PIRNNE by updating the parameters  $\alpha_k$ ,  $\beta_k$  and  $\mu_k$  adaptively. Therefore, the main contributions of this paper are as follows:

- We propose a Proximal Iteratively Reweighted Nuclear Norm algorithm with Extrapolation and Line Search, denoted by PIRNNE-LS. This framework integrates line search with extrapolation and dimensionality reduction, circumventing parametric limitations. Parameters within the proposed method initialize aggressively above defined thresholds then undergo criterion-driven adaptive recalibration per iteration.
- We prove the subsequential convergence that each generated sequence converges to a stationary point of the considered problem. Especially, when the line search is monotone, we further establish its global convergence and linear convergence rate under the Kurdyka–Łojasiewicz framework.
- We conduct some experiments to evaluate the performance of the proposed method for solving the matrix completion problem. Some numerical results are reported the effectiveness and superiority of our proposed method.

The remainder of this paper is organized as follows. Section 2 provides the preliminaries needed for the theoretical analysis in the subsequent sections. Section 3 details PIRNNE-LS for the specified problem. Section 4 analyzes the subsequential convergence. Specifically, we discuss the global convergence and linear convergence rate under the Kurdyka–Łojasiewicz framework in the case of monotone line search. Section 5 reports numerical results on synthetic and empirical datasets. Concluding conclusions appear in Section 6.

## 2. Preliminaries

In this section, we recall some definitions and properties which will be used in the analysis.

### 2.1. Basic Concepts in Variational and Convex Analysis

For an extended-real-valued function  $J := \mathbb{R}^n \rightarrow (-\infty, \infty]$ , its domain is defined by  $\text{dom}(J) := \{x \in \mathbb{R}^n : J(x) < +\infty\}$ . If  $\text{dom}(J) \neq \emptyset$  and  $J(x) > -\infty$  for any  $x \in \text{dom}(J)$ , we say the function  $J$  is proper. If it is lower semicontinuous, we say it is closed. For any subset  $T \subseteq \mathbb{R}^n$  and any point  $x \in \mathbb{R}^n$ , the distance from  $x$  to  $T$  is defined by  $\text{dist}(x, T) := \inf\{\|y - x\| : y \in T\}$ , and we have that  $\text{dist}(x, T) = \infty$  for all  $x$  when  $T = \emptyset$ .

Next, we give the definition of subdifferential which plays a central role in nonconvex optimization.

**Definition 1** ([32,33]). (Subdifferentials) Let  $J : \mathbb{R}^n \rightarrow (-\infty, \infty]$  be a proper and lower semicontinuous function.

- (i) For a given  $x \in \text{dom}(J)$ , the Fréchet subdifferential of  $J$  at  $x$ , written by  $\hat{\partial}J(x)$ , is the set of all vectors  $u \in \mathbb{R}^n$  that satisfy

$$\liminf_{y \neq x, y \rightarrow x} \frac{J(y) - J(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0.$$

When  $x \notin \text{dom}(J)$ , we set  $\hat{\partial}J(x) = \emptyset$ .

- (ii) The limiting subdifferential, or simply the subdifferential, of  $J$  at  $x$ , written by  $\partial J(x)$ , is defined by

$$\partial J(x) := \{u \in \mathbb{R}^n \mid \exists x^k \rightarrow x, \text{ s.t. } J(x^k) \rightarrow J(x) \text{ and } u^k \in \hat{\partial}J(x^k) \rightarrow u \text{ as } k \rightarrow \infty\}. \quad (6)$$

- (iii) A point  $x^*$  is called the (limiting) critical point or stationary point of  $J$  if it satisfies  $0 \in \partial J(x^*)$ , and the set of critical points of  $J$  is denoted by  $\text{crit}J$ .

**Assumption 1.**  $\Psi(X) \rightarrow +\infty$  iff  $\|X\|_F \rightarrow \infty$ .

## 2.2. Kurdyka–Łojasiewicz Property

Now we recall the Kurdyka–Łojasiewicz (KL) property [33–35], which would help us to establish the global convergence. Many functions have KL properties, like semi-algebraic functions defined in an  $\mathcal{o}$ -minimal structure, and others discussed in [32].

**Definition 2.** (KL property and KL function) Let  $J : \mathbb{R}^n \rightarrow (-\infty, \infty]$  be a proper and lower semicontinuous function.

- (i) The function  $J$  is said to have KL property at  $x^* \in \text{dom}(\partial J)$  if there exists  $\eta \in (0, +\infty]$ , a neighborhood  $U$  of  $x^*$ , and a continuous and concave function  $\varphi : [0, \eta) \rightarrow \mathbb{R}^+$  such that
- $\varphi(0) = 0$  and  $\varphi$  is continuously differentiable on  $(0, \eta)$  with  $\varphi' > 0$ ;
  - for all  $x \in U \cap \{z \in \mathbb{R}^n \mid J(x^*) < J(z) < J(x^*) + \eta\}$ , the following KL inequality holds:

$$\varphi'(J(x) - J(x^*)) \text{dist}(0, \partial J(x)) \geq 1.$$

- (ii) If  $J$  satisfies the KL property at each point of  $\text{dom}(\partial J)$ , then  $J$  is called a KL function.

Let  $\Phi_\eta$  be the set of function  $\varphi$  which satisfies the involved conditions in Definition 2 (i). In the following, we give a uniformized KL property which was established in [33].

**Lemma 1** ([33], Lemma 6). (Uniformized KL property) Let  $\Omega$  be a compact set and  $J : \mathbb{R}^n \rightarrow (-\infty, \infty]$  be a proper and lower semicontinuous function. Assume that  $J$  is a constant on  $\Omega$  and satisfies the KL property at each point of  $\Omega$ . Then, there exists  $\zeta, \eta > 0$  and  $\varphi \in \Phi_\eta$  such that for all  $\bar{x} \in \Omega$  and all  $x$  in the following intersection

$$\{z \in \mathbb{R}^n \mid \text{dist}(z, \Omega) < \zeta\} \cap \{z \in \mathbb{R}^n \mid J(\bar{x}) < J(z) < J(\bar{x}) + \eta\},$$

one has

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

### 3. The Proposed Method

This section advances a novel computational framework addressing the nonconvex and nonsmooth optimization problem (1). Under the functional decomposition of  $f$  in (1), we posit existence of convex functions  $f_1$  and  $f_2$  exhibiting Lipschitz continuous gradients such that  $f := f_1 - f_2$ . Consistent with established literature [1,36–39], the Lipschitz constant  $L_f$  governing  $\nabla f$  conforms to  $L_f \leq L$ , where  $\nabla f_1$  and  $\nabla f_2$  possess respective Lipschitz moduli  $L > 0$  and  $l \geq 0$  under the condition  $L \geq l$ . The formal computational architecture is instantiated in Algorithm 1.

---

**Algorithm 1** PIRNNE-LS for solving (1)

---

Choose  $\eta_1, \eta_2, \tau, p_{\min} \in (0, 1)$ ,  $\alpha_{\max}, \beta_{\max}, d > 0, \delta \in [0, 1)$ ,  $0 < \mu_{\min} \leq \frac{1-\delta}{L+2d} \leq \mu_{\max}$ . For given  $X^0 \in \mathbb{R}^{m \times n}$ ,  $X^{-1} = X^0$ , let  $\tilde{E}_0 := \Psi(X^0)$  and set  $k := 0$ .

**while** stopping criterion is not satisfied, **do**

**Step 1.** Choose  $\alpha_k^0 \in [0, \alpha_{\max}]$ ,  $\beta_k^0 \in [0, \beta_{\max}]$  and  $\mu_k^0 \in [\mu_{\min}, \mu_{\max}]$ , set  $\alpha_k := \alpha_k^0$ ,  $\beta_k := \beta_k^0$ ,  $\mu_k := \mu_k^0$ , then

**(1a)** Compute  $Y^k, Z^k$  by (2) and (3), respectively.

**(1b)** Compute the SVD of  $Y^k - \mu_k \nabla f(Z^k)$ , i.e.,  $Y^k - \mu_k \nabla f(Z^k) := \tilde{U}^k \tilde{\Lambda}^k (\tilde{V}^k)^\top$ ; Compute the singular value of  $X^k$ , and let  $w_i^k := g'(\sigma_i(X^k))$  for  $i = 1, \dots, m$ .

**(1c)** Compute

$$X^{k+1} := \tilde{U}^k S(\tilde{\Lambda}^k) (\tilde{V}^k)^\top, \quad (7)$$

where  $S(\tilde{\Lambda}^k) := \text{diag}\{(\tilde{\Lambda}_{i,i}^k - \mu_k w_i^k)_+\}_{1 \leq i \leq m}$ .

**(1d)** If

$$E_\delta(X^{k+1}, X^k, \mu_k) - \tilde{E}_k \leq -\frac{d}{2} \|X^{k+1} - X^k\|_F^2, \quad (8)$$

is satisfied, go to **Step 2**, where  $E_\delta$  is defined in (9). Otherwise, set  $\alpha_k = \eta_1 \alpha_k$ ,  $\beta_k = \eta_2 \beta_k$ ,  $\mu_k = \max\{\tau \mu_k, \mu_{\min}\}$  and go to **Step (1a)**.

**Step 2.**  $\tilde{E}_{k+1} = p_k E_\delta(X^{k+1}, X^k, \mu_k) + (1 - p_k) \tilde{E}_k$  for  $p_k \in [p_{\min}, 1]$ , then let  $k = k + 1$  and go to **Step 1**.

**end while**

---

Within Algorithm 1, analogous to references [28,31], we postulate the potential function:

$$E_\delta(U, V, \mu) := \Psi(V) + \frac{\delta}{4\mu} \|U - V\|_F^2, \quad (9)$$

where  $E_\delta : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \times \mathbb{R}^+ \rightarrow (-\infty, \infty]$  and  $\delta \in [0, 1)$  signifies an assigned nonnegative constant. Moreover, Algorithm 1 permits the selection of arbitrary initial values  $\alpha_k^0 \in [0, \alpha_{\max}]$ ,  $\beta_k^0 \in [0, \beta_{\max}]$  and  $\mu_k^0 \in (\mu_{\min}, \mu_{\max}]$  per iteration. Subsequent adaptive refinement occurs governed by the line search criterion (8). This methodology markedly enhances the procedure's adaptability and numerical efficiency. Furthermore, contingent upon specific conditions, users may initially and intuitively select  $\mu_{\min}$  and  $\mu_{\max}$ ; subsequently, determination of  $d$  ensues based upon their stipulated conditions.

**Remark 1.** Observe that PIRNNE-LS still necessitates computing the singular value decomposition of a large-scale  $Y^k - \mu_k \nabla f(Z^k)$ , potentially exorbitant. The subsequent lemma ensures  $X^{k+1}$  is derivable through a reduced matrix's SVD.  $Y^k - \mu_k \nabla f(Z^k)$  has  $\hat{q}$  singular values  $\tilde{\Lambda}_{i_1, i_1}^k \geq \dots \geq \tilde{\Lambda}_{i_{\hat{q}}, i_{\hat{q}}}^k$  such that  $\tilde{\Lambda}_{i_j, i_j}^k > \mu_k w_{i_j}^k$ . Henceforth Algorithm 1 yields:  $\tilde{U}_{\hat{q}}^k \tilde{\Lambda}_{\hat{q}}^k (\tilde{V}_{\hat{q}}^k)^\top := (\tilde{u}_{i_1}^k, \dots, \tilde{u}_{i_{\hat{q}}}^k) \text{diag}(\tilde{\Lambda}_{i_1, i_1}^k, \dots, \tilde{\Lambda}_{i_{\hat{q}}, i_{\hat{q}}}^k) (\tilde{v}_{i_1}^k, \dots, \tilde{v}_{i_{\hat{q}}}^k)^\top$  as the rank- $\hat{q}$  SVD of  $Y^k - \mu_k \nabla f(Z^k)$ , where  $\tilde{u}_{i_j}^k$  and  $\tilde{v}_{i_j}^k$  denote left and right singular vectors for  $\tilde{\Lambda}_{i_j, i_j}^k$ , respectively.

**Lemma 2** ([17]). Let  $Q$  be a  $\mathbb{R}^{m \times q}$  matrix with orthogonal columns,  $\widetilde{U}_{\hat{q}}^k \widetilde{\Lambda}_{\hat{q}}^k (\widetilde{V}_{\hat{q}}^k)^\top$  be the rank- $\hat{q}$  SVD of  $Y^k - \mu_k \nabla f(Z^k)$ ,  $\widetilde{U}_Q^k \widetilde{\Lambda}_Q^k (\widetilde{V}_Q^k)^\top$  be the SVD of  $Q^\top (Y^k - \mu_k \nabla f(Z^k))$ , and  $\text{span}(\widetilde{U}_{\hat{q}}^k) \subset \text{span}(Q)$ , where  $q \geq \hat{q}$ . Thus,  $X^{k+1} := Q \widetilde{U}_Q^k S(\widetilde{\Lambda}_Q^k) (\widetilde{V}_Q^k)^\top$  is a solution to (5), where  $S(\widetilde{\Lambda}_Q^k) = \text{diag}([\widetilde{\Lambda}_Q^k]_{i_j, i_j} - \mu_k w_{i_j}^k)_{i_1 \leq i_j \leq i_{\hat{q}}}$ .

## 4. Convergence Analysis

This section mainly delineates subsequential convergence, global convergence and linear convergence rate. We start by analyzing the convergence of the subsequence.

### 4.1. Subsequential Convergence of Nonmontone Line Search

Initially, we establish monotone nonincreasing property of the sequence  $\{E_\delta(X^{k+1}, X^k, \mu_k)\}$ .

**Lemma 3.** Let  $\{X^k\}$  be the sequence generated by Algorithm 1. If for any  $k \geq 0$ , the parameters  $\alpha_k$ ,  $\beta_k$  and  $\mu_k$  satisfy

$$\mu_k < \frac{1}{L}, \quad \alpha_k \leq \sqrt{\frac{\mu_k \delta (1 - \mu_k L)}{16 \mu_{k-1}}}, \quad \text{and} \quad \beta_k \leq \sqrt{\frac{\delta (1 - \mu_k L)}{4 \mu_{k-1} (4 \mu_k L^2 + L + l)}}. \quad (10)$$

Then, we have

$$E_\delta(X^{k+1}, X^k, \mu_k) - E_\delta(X^k, X^{k-1}, \mu_{k-1}) \leq \frac{\mu_k L - 1 + \delta}{4 \mu_k} \|X^{k+1} - X^k\|_F^2. \quad (11)$$

**Proof.** Since  $X^{k+1}$  is a minimizer of the optimization problem in (4), we get

$$\begin{aligned} \sum_{i=1}^m w_i^k \sigma_i(X^{k+1}) &\leq \sum_{i=1}^m w_i^k \sigma_i(X^k) + \langle \nabla f(Z^k), X^k - X^{k+1} \rangle \\ &\quad + \frac{1}{2 \mu_k} \|X^k - Y^k\|_F^2 - \frac{1}{2 \mu_k} \|X^{k+1} - Y^k\|_F^2. \end{aligned} \quad (12)$$

From the Lipschitz continuity of  $\nabla f$  with the modulus  $L_f$  and  $L_f \leq L$ , it follows from ([40], Lemma 1.2.3) that

$$f(X^{k+1}) \leq f(Z^k) + \langle \nabla f(Z^k), X^{k+1} - Z^k \rangle + \frac{L}{2} \|X^{k+1} - Z^k\|_F^2. \quad (13)$$

Similarly to the technique of ([1], Lemma 4), we get

$$\begin{aligned} f(X^{k+1}) + \sum_{i=1}^m w_i^k \sigma_i(X^{k+1}) - f(X^k) - \sum_{i=1}^m w_i^k \sigma_i(X^k) \\ \leq \frac{1}{2 \mu_k} \|X^k - Y^k\|_F^2 - \frac{1}{2 \mu_k} \|X^{k+1} - Y^k\|_F^2 + \frac{L}{2} \|X^{k+1} - Z^k\|_F^2 + \frac{l}{2} \|X^k - Z^k\|_F^2. \end{aligned} \quad (14)$$

Next, it follows from (2) and (3) that

$$\begin{aligned} X^k - Y^k &= -\alpha_k (X^k - X^{k-1}), \quad X^{k+1} - Y^k = X^{k+1} - X^k - \alpha_k (X^k - X^{k-1}), \\ X^k - Z^k &= -\beta_k (X^k - X^{k-1}), \quad X^{k+1} - Z^k = X^{k+1} - X^k - \beta_k (X^k - X^{k-1}). \end{aligned} \quad (15)$$

Merging (14), (15), the concavity of  $g$  and the definition of  $E_\delta$  in (9), we have

$$\begin{aligned}
& E_\delta(X^{k+1}, X^k, \mu_k) - E_\delta(X^k, X^{k-1}, \mu_{k-1}) \\
&= f(X^{k+1}) + \sum_{i=1}^m g(\sigma_i(X^{k+1})) \\
&\quad + \frac{\delta}{4\mu_k} \|X^{k+1} - X^k\|_F^2 - f(X^k) - \sum_{i=1}^m g(\sigma_i(X^k)) - \frac{\delta}{4\mu_{k-1}} \|X^k - X^{k-1}\|_F^2 \\
&\leq f(X^{k+1}) - f(X^k) + \sum_{i=1}^m w_i^k (\sigma_i(X^{k+1}) - \sigma_i(X^k)) \\
&\quad + \frac{\delta}{4\mu_k} \|X^{k+1} - X^k\|_F^2 - \frac{\delta}{4\mu_{k-1}} \|X^k - X^{k-1}\|_F^2 \\
&\leq \frac{1}{2\mu_k} \alpha_k^2 \|X^k - X^{k-1}\|_F^2 - \frac{1}{2\mu_k} \|(X^{k+1} - X^k) - \alpha_k(X^k - X^{k-1})\|_F^2 \\
&\quad + \frac{L}{2} \|(X^{k+1} - X^k) - \beta_k(X^k - X^{k-1})\|_F^2 + \frac{l}{2} \beta_k^2 \|X^k - X^{k-1}\|_F^2 \\
&\quad + \frac{\delta}{4\mu_k} \|X^{k+1} - X^k\|_F^2 - \frac{\delta}{4\mu_{k-1}} \|X^k - X^{k-1}\|_F^2 \\
&= -\frac{1}{2\mu_k} \|X^{k+1} - X^k\|_F^2 + \frac{\alpha_k}{\mu_k} \langle X^{k+1} - X^k, X^k - X^{k-1} \rangle + \frac{L}{2} \|X^{k+1} - X^k\|_F^2 \\
&\quad + \frac{L}{2} \beta_k^2 \|X^k - X^{k-1}\|_F^2 - L\beta_k \langle X^{k+1} - X^k, X^k - X^{k-1} \rangle + \frac{l}{2} \beta_k^2 \|X^k - X^{k-1}\|_F^2 \\
&\quad + \frac{\delta}{4\mu_k} \|X^{k+1} - X^k\|_F^2 - \frac{\delta}{4\mu_{k-1}} \|X^k - X^{k-1}\|_F^2, \tag{16}
\end{aligned}$$

By using the Young inequality, we obtain

$$\frac{\alpha_k}{\mu_k} \|X^{k+1} - X^k\|_F \cdot \|X^k - X^{k-1}\|_F \leq \frac{1 - \mu_k L}{8\mu_k} \|X^{k+1} - X^k\|_F^2 + \frac{2\alpha_k^2}{\mu_k(1 - \mu_k L)} \|X^k - X^{k-1}\|_F^2,$$

and

$$-L\beta_k \|X^{k+1} - X^k\|_F \cdot \|X^k - X^{k-1}\|_F \leq \frac{1 - \mu_k L}{8\mu_k} \|X^{k+1} - X^k\|_F^2 + \frac{2\mu_k \beta_k^2 L^2}{1 - \mu_k L} \|X^k - X^{k-1}\|_F^2.$$

Substituting above two inequalities into (16), we have

$$\begin{aligned}
& E_\delta(X^{k+1}, X^k, \mu_k) - E_\delta(X^k, X^{k-1}, \mu_{k-1}) \\
&\leq \frac{\mu_k L - 1 + \delta}{4\mu_k} \|X^{k+1} - X^k\|_F^2 \\
&\quad + \left( \frac{2\alpha_k^2}{\mu_k(1 - \mu_k L)} + \frac{L}{2} \beta_k^2 + \frac{l}{2} \beta_k^2 + \frac{2\mu_k \beta_k^2 L^2}{1 - \mu_k L} - \frac{\delta}{4\mu_{k-1}} \right) \|X^k - X^{k-1}\|_F^2.
\end{aligned}$$

Furthermore, it follows from (10) that

$$\frac{(L + l)}{2} \beta_k^2 + \frac{2\mu_k \beta_k^2 L^2}{1 - \mu_k L} \leq \frac{\delta}{8\mu_{k-1}} \quad \text{and} \quad \frac{2\alpha_k^2}{\mu_k(1 - \mu_k L)} \leq \frac{\delta}{8\mu_{k-1}}.$$

Hence, the assertion (11) follows immediately. The proof is completed.  $\square$

**Lemma 4.** (Well-definedness of the line search criterion) Let  $\{X^k\}$  be the sequence generated by Algorithm 1. Then, for any  $k \geq 0$ , criterion (8) shall be satisfied within finite inner iterations.

**Proof.** This proof advances via contradiction. Initial focus rests upon  $k = 0$ . Observe that  $\mu_0 = \tilde{\mu}_0$  with  $\mu_{\min} \leq \tilde{\mu}_0 \leq \frac{1-\delta}{L+2d}$  holds incontrovertibly after finite inner iterations. Here,  $Y^0 = X^0$  and  $Z^0 = X^0$ . Then, from (4), we have



$$\sum_{i=1}^m w_i^0 \sigma_i(X^1) \leq \sum_{i=1}^m w_i^0 \sigma_i(X^0) + \langle \nabla f(X^0), X^0 - X^1 \rangle - \frac{1}{2\tilde{\mu}_0} \|X^1 - X^0\|_F^2,$$

where  $w_i^0 = g'(\sigma_i(X^0))$ . From ([40], Lemma 1.2.3), we obtain

$$f(X^1) - f(X^0) \leq \langle \nabla f(X^0), X^1 - X^0 \rangle + \frac{L}{2} \|X^1 - X^0\|_F^2.$$

Together with the concavity of  $g$ , we have

$$\begin{aligned} \Psi(X^1) - \Psi(X^0) &= f(X^1) + \sum_{i=1}^m g(\sigma_i(X^1)) - f(X^0) - \sum_{i=1}^m g(\sigma_i(X^0)) \\ &\leq f(X^1) - f(X^0) + \sum_{i=1}^m w_i^k (\sigma_i(X^1) - \sigma_i(X^0)) \\ &\leq \frac{L}{2} \|X^1 - X^0\|_F^2 - \frac{1}{2\tilde{\mu}_0} \|X^1 - X^0\|_F^2. \end{aligned}$$

This inequality connotes that (11) holds. Since  $\tilde{\mu}_0 \leq \frac{1-\delta}{L+2d} \leq \frac{2-\delta}{2(L+d)}$ , the criterion (8) shall invariably hold. Suppose that there exists a smallest  $k > 0$  such that the criterion (8) can not be satisfied. It means that the line search criterion (8) is satisfied for the former  $k - 1$  iterations. At the  $k - 1$ -th iteration, there exists  $\mu_{k-1}$  such that

$$E_\delta(X^k, X^{k-1}, \mu_{k-1}) - \tilde{E}_{k-1} \leq -\frac{d}{2} \|X^k - X^{k-1}\|_F^2.$$

Thus, we have  $\tilde{E}_{k-1} \leq E_\delta(X^k, X^{k-1}, \mu_{k-1})$ . Further, Step 2 of Algorithm 1 defines  $\tilde{E}_{k-1}$ , and we obtain

$$\begin{aligned} \tilde{E}_k &= p_{k-1} E_\delta(X^k, X^{k-1}, \mu_{k-1}) + (1 - p_{k-1}) \tilde{E}_{k-1} \\ &\leq p_{k-1} E_\delta(X^k, X^{k-1}, \mu_{k-1}) + (1 - p_{k-1}) E_\delta(X^k, X^{k-1}, \mu_{k-1}) \\ &= E_\delta(X^k, X^{k-1}, \mu_{k-1}). \end{aligned} \quad (17)$$

Since Algorithm 1's Step (1d) necessitates  $\mu_k \leq \mu_{\min}$ ,  $\mu_k = \mu_{\min}$  becomes admissible. Similarly, from Step (1d) in Algorithm 1, we know that

$$\alpha_k \leq \sqrt{\frac{\mu_k \delta (1 - \mu_k L)}{16\mu_{k-1}}}, \quad \text{and} \quad \beta_k \leq \sqrt{\frac{\delta (1 - \mu_k L)}{4\mu_{k-1} (4\mu_k L^2 + L + I)}}$$

must be satisfied. Consequently,  $\mu_k = \mu_{\min}$  and (10) are obtained. In addition, since  $d > 0$  and  $\delta \in [0, 1)$ , it holds that  $0 < \mu_{\min} \leq \frac{1-\delta}{L+2d}$  and, thus,  $\frac{\mu_{\min} L - 1 + \delta}{4\mu_{\min}} \leq -\frac{d}{2}$ . Together with Lemma 3 and (17), we have

$$\begin{aligned} E_\delta(X^{k+1}, X^k, \mu_{\min}) - \tilde{E}_k &\leq E_\delta(X^{k+1}, X^k, \mu_{\min}) - E_\delta(X^k, X^{k-1}, \mu_{k-1}) \\ &\leq \frac{\mu_{\min} L - 1 + \delta}{4\mu_{\min}} \|X^{k+1} - X^k\|_F^2 \\ &\leq -\frac{d}{2} \|X^{k+1} - X^k\|_F^2. \end{aligned} \quad (18)$$

This necessitates satisfaction of line search criterion (8) at the  $k$ -th iteration, inducing contradiction. The proof is completed.  $\square$



We obtain the subsequential convergence of Algorithm 1 in the following Theorem.

**Theorem 1.** Let  $\{X^k\}$  be the sequence generated by Algorithm 1. Then, we have

- (i) the sequence  $\{\tilde{E}_k\}$  is nonincreasing;
- (ii)

$$\lim_{k \rightarrow \infty} \|X^{k+1} - X^k\|_F = 0;$$

- (iii)  $\{X^k\}$  is bounded and any cluster point  $\tilde{X} = \lim_{k \rightarrow \infty} X^{k_i}$  of  $\{X^k\}$  is a critical point of  $\Psi$ .

**Proof.** (i) Invoking line search criterion (8) and  $\tilde{E}_{k+1}$ 's definition in Algorithm 1, we have

$$\begin{aligned} \tilde{E}_{k+1} &= p_k E_\delta(X^{k+1}, X^k, \mu_k) + (1 - p_k) \tilde{E}_k \\ &\leq p_k \left( \tilde{E}_k - \frac{d}{2} \|X^{k+1} - X^k\|_F^2 \right) + (1 - p_k) \tilde{E}_k \\ &\leq \tilde{E}_k - \frac{p_{\min} \cdot d}{2} \|X^{k+1} - X^k\|_F^2. \end{aligned} \quad (19)$$

It indicates the sequence  $\{\tilde{E}_k\}$  is nonincreasing.

- (ii) Summing up (19) from  $k = 1, \dots, N$ , we obtain

$$\begin{aligned} &\frac{p_{\min} \cdot d}{2} \|X^{k+1} - X^k\|_F^2 \\ &\leq \sum_{k=1}^N (\tilde{E}_k - \tilde{E}_{k+1}) \\ &= \tilde{E}_1 - \tilde{E}_{N+1} \\ &= E_\delta(X^1, X^0, \mu_0) - E_\delta(X^{N+1}, X^N, \mu_N) \\ &\leq \Psi(X^1) + \frac{\delta}{4\mu_0} \|X^1 - X^0\|_F^2 - \Psi(X^N) \\ &\leq \Psi(X^1) + \frac{\delta}{4\mu_0} \|X^1 - X^0\|_F^2 - \underline{\Psi} < \infty, \end{aligned}$$

where the validity of the second inequality is deducible from (17), while the third stems directly from the definition of  $E_\delta$  specified in (9). And the last inequality follows from  $X^1 \in \text{dom}\Psi$ ,  $\mu_0 > 0$  and  $\delta \geq 0$ . From the fact that  $p_{\min}, d > 0$  and  $N \rightarrow \infty$ , the assertion (ii) is consequently established.

- (iii) The sequence  $\{\tilde{E}_k\}$  exhibits a monotone decrease, an outcome established by the result (i). We have

$$\tilde{E}_k \leq \tilde{E}_{k-1} \leq \dots \leq \tilde{E}_1 = \Psi(X^1) + \frac{\delta}{4\mu_0} \|X^1 - X^0\|_F^2 < \infty.$$

Again, from the definition of  $\tilde{E}_k$  and (8), we have

$$\begin{aligned} &\Psi(X^1) + \frac{\delta}{4\mu_0} \|X^1 - X^0\|_F^2 \\ &\geq \tilde{E}_k = p_{k-1} E_\delta(X^k, X^{k-1}, \mu_{k-1}) + (1 - p_{k-1}) \tilde{E}_{k-1} \\ &\geq E_\delta(X^k, X^{k-1}, \mu_{k-1}) \geq \Psi(X^k). \end{aligned}$$

Consequently,  $\Psi$  is upper-bounded. Reiterating Assumption 1, the sequence  $\{X^k\}$  remains confined and contains at least one cluster point. Assign  $\tilde{X}$  as such a cluster point. Then, there exists a subsequence  $\{X^{k_j}\}$  of  $X^k$  such that  $\lim_{j \rightarrow \infty} X^{k_j} = \tilde{X}$ . Then, next proof is

similar to ([1], Proof of Theorem 1 (iii)); it is easy to derive that  $0 \in \partial\Psi(\tilde{X})$ , which implies that  $\tilde{X} \in \text{crit}\Psi$ . This completes the proof.  $\square$

#### 4.2. Global Convergence and Linear Convergence Rate of Monotone Line Search

This subsection primarily discusses the global convergence and linear convergence rate of the proposed monotone line search ( $p_k = 1$ ) Algorithm 1 under the KL framework. First, we introduce the following two frequently used lemmas, whose proofs are similar to ([1], Lemma 5) and ([1], Lemma 6), so we will not repeat them here. We proceed to use the notation  $\Delta_k := X^k - X^{k-1}$  in this subsection.

**Lemma 5.** Let  $\{X^k\}$  be the sequence generated by Algorithm 1. Then, there exist some  $K \in \mathbb{N}$  and  $b > 0$  such that for all  $k \geq K$ , there exists  $\omega^{k+1} \in \partial E_\delta(X^{k+1}, X^k, \mu_k)$  such that

$$\|\omega^{k+1}\|_F \leq b(\|\Delta_{k+1}\|_F + \|\Delta_k\|_F).$$

Denote the cluster point set of  $\{X^{k+1}, X^k, \mu_k\}$  by  $\Xi$ . Then, we summarize some properties of the cluster point set  $\Xi$ .

**Lemma 6.** Let  $\{X^k\}$  be the sequence generated by Algorithm 1 with  $p_k = 1$ . Then, we have

- (i)  $\Xi$  is nonempty and  $\Xi \subseteq \text{crit}E_\delta$ ;
- (ii)

$$\lim_{k \rightarrow \infty} \text{dist}((X^{k+1}, X^k, \mu_k), \Xi) = 0;$$

- (iii)  $E_\delta$  and  $\Psi$  are equal and constant on  $\Xi$ , i.e., there exists a constant  $\kappa$  such that for any  $(\tilde{X}, \tilde{X}, \tilde{\mu}) \in \Xi$ ,  $E_\delta(\tilde{X}, \tilde{X}, \tilde{\mu}) = \Psi(\tilde{X}) = \kappa$ .

**Theorem 2.** Let  $\{X^k\}$  be the sequence generated by Algorithm 1 with  $p_k = 1$  for  $k$ 's large enough and  $E_\delta$  is a KL function.

- (i) The whole sequence  $\{X^k\}$  manifests finite length  $\sum_{k=0}^{\infty} \|\Delta_{k+1}\|_F < +\infty$  and  $\{X^k\}$  globally converges to a point  $\tilde{X}$  in  $\text{crit}\Psi$ .
- (ii) Moreover, if the KL function can be taken in the form  $\varphi(s) = \rho s^{1-t}$  for some  $t \in (0, 1/2]$ , the whole sequences  $\{X^k\}$  and  $\{E_\delta(X^{k+1}, X^k, \mu_k)\}$  are linearly convergent.

**Proof.** (i) Assume that  $(\tilde{X}, \tilde{X}, \tilde{\mu}) \in \Xi \subseteq \text{crit}E_\delta$ . Then, there exists a subsequence  $\{(X^{k_i+1}, X^{k_i}, \mu_{k_i})\}$  of  $\{(X^{k+1}, X^k, \mu_k)\}$  converging to  $(\tilde{X}, \tilde{X}, \tilde{\mu})$ . Let  $k_1 \in \mathbb{N}$  be such that  $p_k = 1$  for all  $k \geq k_1$ , and we know that  $\tilde{E}_{k+1} = E_\delta(X^{k+1}, X^k, \mu_k)$ . It follows from Theorem 1 (iii) and the continuity of  $\Psi$  that  $\lim_{i \rightarrow \infty} \Psi(X^{k_i}) = \Psi(\tilde{X})$ . Again from Theorem 1 (i) and (ii), we have  $\lim_{k \rightarrow \infty} \|\Delta_k\|_F = 0$ , and  $\{E_\delta(X^{k+1}, X^k, \mu_k)\}$  is nonincreasing for all  $k \geq k_1$ . Thus, we get  $\lim_{k \rightarrow \infty} E_\delta(X^{k+1}, X^k, \mu_k) = \kappa$ , and  $E_\delta(X^{k+1}, X^k, \mu_k) \geq \kappa$  for all  $k \geq k_1$ .

If there exists an integer  $\bar{k}$  such that  $E_\delta(X^{\bar{k}}, X^{\bar{k}-1}, \mu_{\bar{k}-1}) = \kappa$ , then from (8),  $\forall k \geq \bar{k}$ , we have

$$\begin{aligned} \frac{d}{2} \|\Delta_{k+1}\|_F^2 &\leq E_\delta(X^k, X^{k-1}, \mu_{k-1}) - E_\delta(X^{k+1}, X^k, \mu_k) \\ &\leq E_\delta(X^{\bar{k}}, X^{\bar{k}-1}, \mu_{\bar{k}-1}) - E_\delta(X^{k+1}, X^k, \mu_k) \\ &\leq E_\delta(X^{\bar{k}}, X^{\bar{k}-1}, \mu_{\bar{k}-1}) - \kappa \\ &= 0. \end{aligned}$$

Thus, we have  $X^{k+1} = X^k$  for any  $k > \bar{k}$  and the assertion  $\sum_{k=0}^{\infty} \|\Delta_{k+1}\|_F < +\infty$  holds directly. Otherwise, since  $\{E_\delta(X^{k+1}, X^k, \mu_k)\}$  is nonincreasing for all  $k \geq k_1$ , we have  $\{E_\delta(X^{k+1}, X^k, \mu_k)\} > \kappa$  for all  $k \geq k_1$ . Now, we consider the sequence  $\{(X^{k+1}, X^k, \mu_k)\}_{k=0}^{\infty}$ . It follows from Lemma 6 that the cluster point set  $\Xi$  of  $\{(X^{k+1}, X^k, \mu_k)\}_{k=0}^{\infty}$  is nonempty and compact, and for any  $(\tilde{X}, \tilde{X}, \tilde{\mu}) \in \Xi$ , we have

$$E_\delta(\tilde{X}, \tilde{X}, \tilde{\mu}) = \Psi(\tilde{X}) = \kappa.$$

Thus, for any  $\eta > 0$ , there exists a nonnegative integer  $k_2 \geq k_1$  such that  $E_\delta(X^{k+1}, X^k, \mu_k) < \kappa + \eta$  for any  $k > k_2$ . In addition, for any  $\kappa > 0$ , there exists a positive integer  $k_3 \geq k_1$  such that  $\text{dist}((X^{k+1}, X^k, \mu_k), \Xi) < \kappa$  for all  $k > k_3$ . Consequently, for any  $\eta, \kappa > 0$ ,  $k > k_4 := \max\{k_2, k_3, K\}$ , where  $K$  is given by Lemma 5, we have

$$\text{dist}((X^{k+1}, X^k, \mu_k), \Xi) < \kappa, \quad \text{and} \quad \kappa < E_\delta((X^{k+1}, X^k, \mu_k)) < \kappa + \eta.$$

By using Lemma 1 with  $\Omega := \Xi$ , for any  $k \geq k_4$ , we have

$$\varphi'(E_\delta((X^{k+1}, X^k, \mu_k)) - \kappa) \text{dist}(0, \partial E_\delta(X^{k+1}, X^k, \mu_k)) \geq 1. \quad (20)$$

The remaining global convergence arguments are similar to ([1], Theorem 2);  $\{X^k\}$  is a Cauchy sequence and, hence, it is convergent. By using Lemma 6 (i), there exists  $(\tilde{X}, \tilde{X}, \tilde{\mu}) \in \text{crit}E_\delta$  with  $\tilde{X} \in \text{crit}\Psi$  such that  $\lim_{k \rightarrow \infty} X^k = \tilde{X}$ .

(ii) Denote  $\Theta_k := E_\delta(X^k, X^{k-1}, \mu_{k-1}) - \kappa$ . It follows from (20) that

$$\begin{aligned} 1 &\leq \varphi'(\Theta_{k+1}) \text{dist}\left(0, \partial E_\delta(X^{k+1}, X^k, \mu_k)\right) \\ &\leq (1-t)\rho b(\Theta_{k+1})^{-t} (\|\Delta_{k+1}\| + \|\Delta_k\|) \\ &\leq (1-t)\rho b(\Theta_{k+1})^{-t} \sqrt{2\|\Delta_{k+1}\|^2 + 2\|\Delta_k\|^2} \\ &\leq (1-t)\rho b(\Theta_{k+1})^{-t} \sqrt{4/d [E_\delta(X^{k-1}, X^{k-2}, \mu_{k-2}) - E_\delta(X^{k+1}, X^k, \mu_k)]} \\ &= c(\Theta_{k+1})^{-t} \sqrt{\Theta_{k-1} - \Theta_{k+1}}, \end{aligned}$$

where  $c = 2(1-t)\rho b/\sqrt{d}$ , the second inequality follows from Lemma 5 and the fourth one follows from (8), together with the  $p_k = 1$ . Since  $\Theta_k \rightarrow 0$ , there exists  $k_5$  such that  $\Theta_k \leq 1$ . Then, for all  $k \geq k_6 := \max\{k_4, k_5\}$ , it follows from (20) that for any  $k \geq k_6 + 1$ ,

$$\Theta_k \leq (\Theta_k)^{2t} \leq c^2(\Theta_{k-2} - \Theta_k),$$

which means that

$$\Theta_k \leq \frac{c^2}{1+c^2} \Theta_{k-2}.$$

So, the sequences  $\{E_\delta(X^{2k+1}, X^{2k}, \mu_{2k})\}$  and  $\{E_\delta(X^{2k}, X^{2k-1}, \mu_{2k-1})\}$  are both Q-linearly convergent. This indicates that the entire sequence  $\{E_\delta(X^{k+1}, X^k, \mu_k)\}$  is R-linearly convergent. By combining this with (8), we can infer that there exist  $N > 0$  and  $k_0$  and  $q \in (0, 1)$  so that for each  $k \geq k_0$ ,  $\|\Delta_{k+1}\| \leq Nq^k$ . Consequently,

$$\|X^k - X^*\| \leq \sum_{i=k}^{\infty} \|\Delta_{i+1}\| \leq \frac{N}{1-q} q^k,$$

which means that  $\{X^k\}$  is  $R$ -linearly convergent. This completes the proof.  $\square$

## 5. Numerical Results

This section evaluates the algorithm's efficacy through resolution of the matrix completion problem:

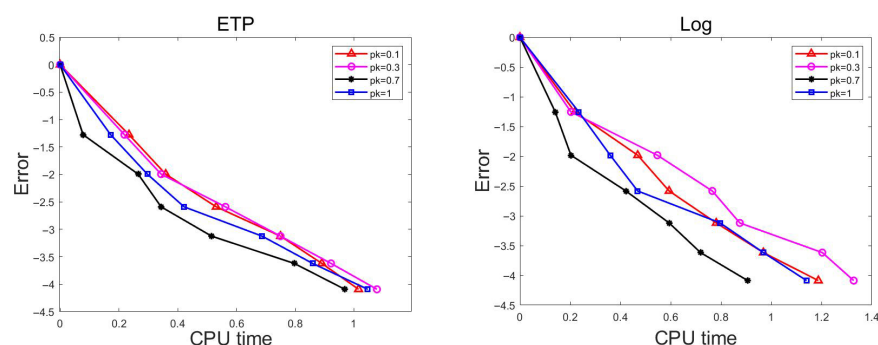
$$\min_X \sum_{i=1}^m g(\sigma_i(X)) + \frac{1}{2} \|\mathcal{P}_\Omega(X) - Y\|_F^2, \quad (21)$$

where  $\Omega$  constitutes the sample index set while  $\mathcal{P}_\Omega : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$  operates linearly, preserving  $\Omega$  entries intact and nullifying others. Define  $f := f_1 - f_2$ ,  $f_1 := \frac{1}{2} \|\mathcal{P}_\Omega(X) - Y\|_F^2$ ,  $f_2 := 0$ ;  $\nabla f_1$  and  $\nabla f_2$  exhibit Lipschitz constants  $L = 1$  and  $l = 0$ . Algorithm performance manifests using both synthetic and empirical datasets. Implementation employed MATLAB 2020a on a Windows 10 platform with Intel(R) Core(TM) i7-1165G7 processor (2.80 GHz) and 16 GB RAM. Testing prioritizes ETP and Log penalty functions, replicating parameter selections from [1,4] as predominantly optimal.

### 5.1. Synthetic Data

Within this synthetic trial, we fabricate a rank  $-r$  matrix  $X^*$  as  $M_L M_R$ , where  $M_L \in \mathbb{R}^{m \times r}$ , and  $M_R \in \mathbb{R}^{r \times n}$  originate from MATLAB's rand command. Half of the elements in  $X^*$  are randomly and uniformly missing. Here, the observed matrix  $Y = \mathcal{P}_\Omega(X^*)$ , with  $\lambda_0 = \|Y\|_\infty$ , and  $\lambda := \lambda_t = 10^{-3} \lambda_0$  in the model. Termination occurs when  $\|\mathcal{P}_\Omega(X) - Y\|_\infty \leq 10^{-3}$ .

PIRNNE-LS integrates a line search strategy to eliminate parametric restrictions. Validating this approach, we examine ETP and Log nonconvex penalties under four  $p_k$  scenarios:  $p_k = p \in \{0.1, 0.3, 0.7, 1\}$ , noting monotonicity when  $p = 1$  and non-monotonicity for  $p < 1$ . Tests employ  $m = n = 500$ ,  $r = 50$ ,  $\alpha_k^0 = 0.1$ ,  $\beta_k^0 = 0.1$ ,  $\mu_k^0 = 1$  for each  $k \in N$ , with Algorithm 1 (Step 1) parameters  $\eta_1 = 0.4$ ,  $\eta_2 = 0.35$ ,  $\tau = 0.45$ ,  $d = 0.1$  and  $\delta = 0.1$ . Maximum iterations cap at 1000. Figure 1 charts error metric evolution against CPU duration. Figure 1 indicates PIRNNE-LS with  $p = 0.7$  and  $p = 1$  markedly outperforms alternatives.



**Figure 1.** Evolution of the error value with respect to the CPU time.

### 5.2. Real Images

In this subsection, we primarily undertake a comparative analysis of our proposed algorithm with APIRNN in [17] and PIRNNE in [1]. For the APIRNN and PIRNNE algorithms, the choice of involved parameters  $\alpha_k, \beta_k$  is the same as in [1,17], respectively. To better demonstrate algorithmic enhancements, we implement (i) monotone line search ( $p_k \equiv 1$ ), designated PIRNNE-mLS, and (ii) nonmonotone line search ( $p_k \equiv 0.7$ ), designated PIRNNE-nLS.

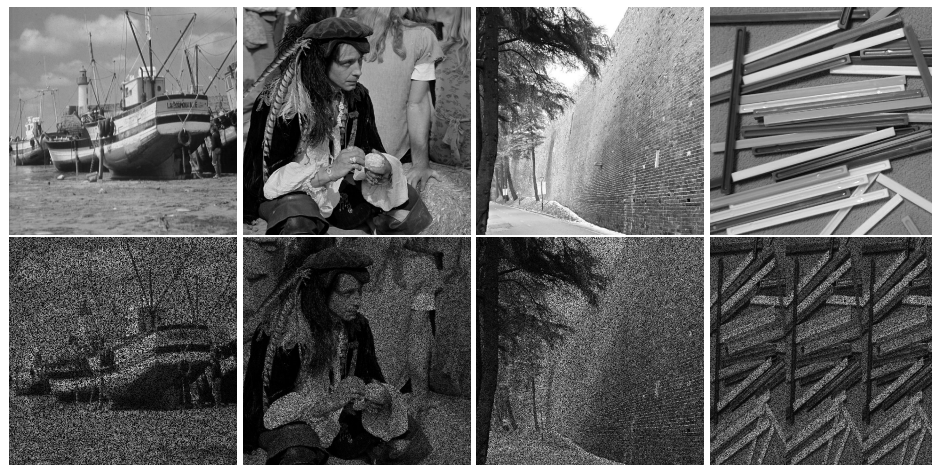
To more comprehensively demonstrate algorithmic efficacy, we constructed four 2D images, “Boat ( $512 \times 512$ )”, “Man ( $1024 \times 1024$ )”, “City Wall ( $512 \times 512$ )” and “Spillikins

$(512 \times 512)$ ", alongside four 3D counterparts, "Bottles  $(512 \times 512)$ ", "Texture  $(512 \times 512)$ ", "House  $(256 \times 256)$ ", "Clock  $(512 \times 512)$ ", visualized in Figures 2 and 3. It is universally acknowledged that although not all authentic images possess low-rank characteristics, the essential information is primarily determined by the higher singular values. Consequently, it is feasible to recover corrupted images through low-rank approximation. For 3D imagery containing three separate channels, matrix completion executes independently per channel. The approach for selecting parameters remains identical to that employed in the artificial example, and the termination criterion is  $\|\mathcal{P}_\Omega(X) - Y\|_\infty \leq 10^{-3}$ .

Due to space constraints, we concentrate upon the ETP penalty function to demonstrate recovery efficacy. The algorithm's recovery capability quantifies via Signal-to-Noise Ratio (SNR), defined as

$$\text{SNR}(u, \bar{u}) = 10 \log_{10} \frac{\|u - \bar{u}\|^2}{\|u - u^*\|^2},$$

where  $u$  and  $\bar{u}$  signify the original image and mean of the original image, and  $u^*$  the reconstructed image.



**Figure 2.** The list of pictures in order: Boat, Man, City Wall and Spillikins. First row: original images, second row: noisy images.



**Figure 3.** The list of pictures in order: Bottles, Texture, House and Clock. First row: original images, second row: noisy images.

For this evaluation, random values perturb 50% of image elements, with  $\Omega$  denoting a set of random values. These corrupted images appear in Figures 2 and 3's second row. SNR values

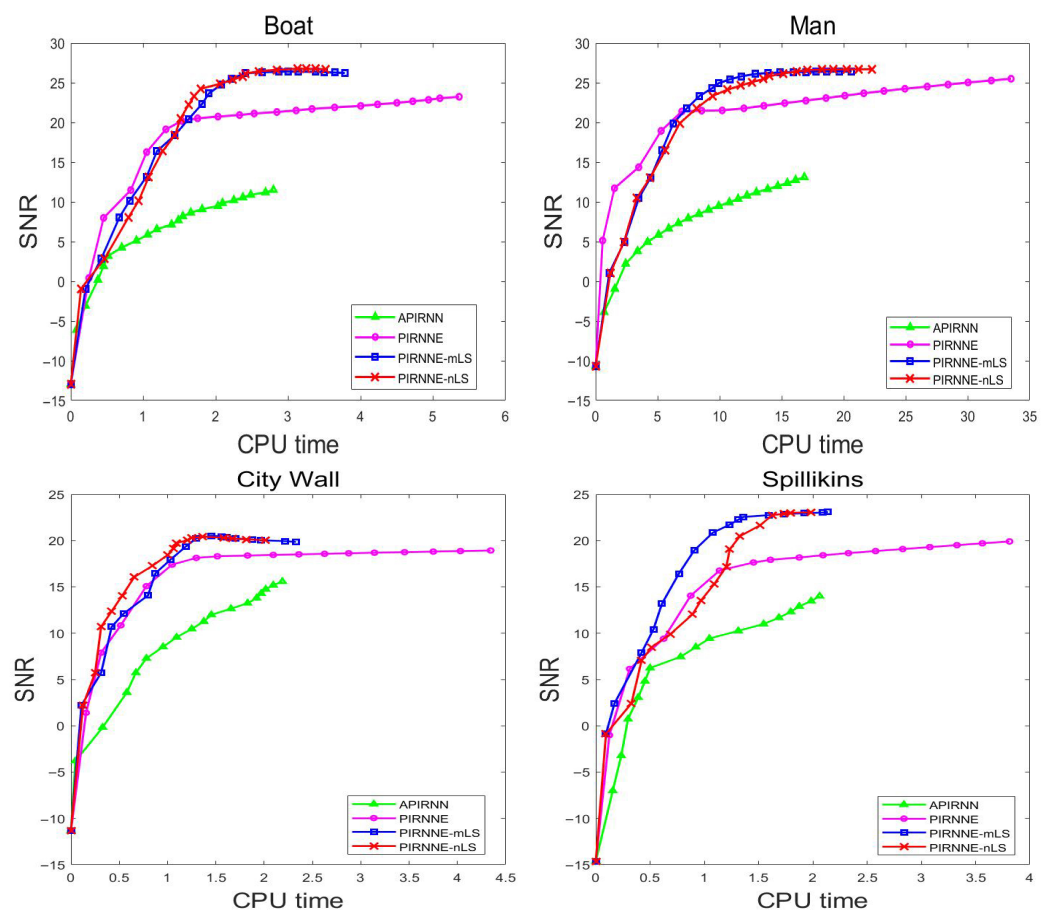


relative to CPU duration—required by four methods to attain minimal processing time—are plotted across Figures 4 and 5. The results in Figures 4 and 5 show that PIRNNE-LS (including PIRNNE-mLS and PIRNNE-nLS) outperforms traditional APIRNN and PIRNNE.

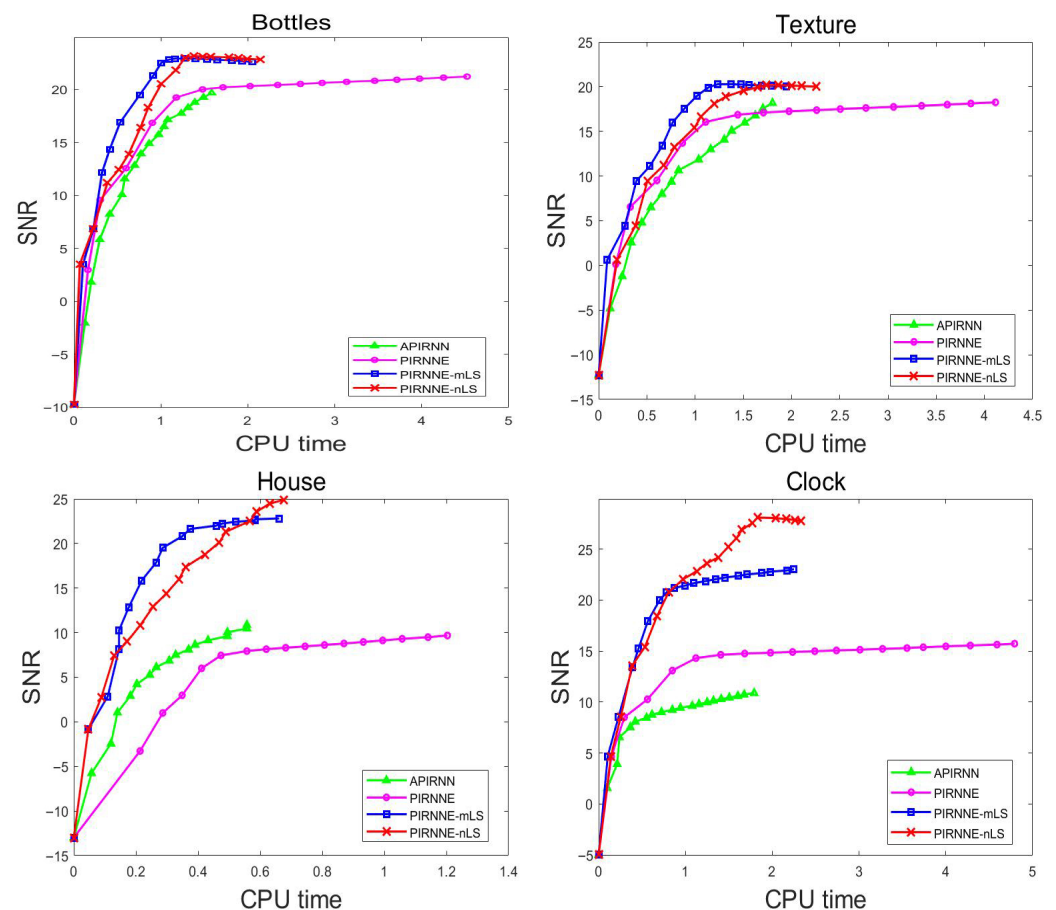
Furthermore, we report the number of iterations and CPU time in seconds and SNR values in Table 1. In the presented results, we use “Iter.”, “Time” and “SNR” to denote the number of iterations, CPU time in seconds and SNR value, respectively. In color images, Iter., Time and SNR represent the mean values of the three channels. From Table 1, we observe that our proposed PIRNNE-mLS and PIRNNE-nLS have better recovery performance.

**Table 1.** Numerical results of tested algorithms with Boat, Man, City Wall, Spillikins, Bottles, Texture, House and Clock.

	APIRNN			PIRNNE			PIRNNE-mLS			PIRNNE-nLS		
	Iter.	Time	SNR	Iter.	Time	SNR	Iter.	Time	SNR	Iter.	Time	SNR
Boat	121	3.94	19.49	51	2.87	23.96	43	1.90	26.23	43	1.90	26.69
Man	117	30.02	23.29	56	23.90	26.64	45	15.61	26.35	44	15.41	26.79
City Wall	56	1.43	17.88	40	1.61	19.08	38	1.05	19.74	35	0.96	20.05
Spillikins	68	1.57	20.05	59	1.96	22.39	34	1.03	23.08	33	1.01	23.09
Bottles	66	4.76	21.92	56	6.34	22.26	46	3.10	22.50	38	2.98	22.74
Texture	57	4.72	19.81	52	6.09	19.55	42	3.13	20.94	35	2.80	21.95
House	109	1.17	21.96	114	2.74	21.37	39	0.58	23.94	38	0.55	24.76
Clock	229	12.10	20.43	155	17.36	24.92	55	5.01	24.28	47	3.30	27.99



**Figure 4.** Evolution of SNR values of Boat, Man, City Wall and Spillikins with respect to the CPU time.



**Figure 5.** Evolution of SNR values of Bottles, Texture, House and Clock with respect to the CPU time.

### 5.3. Movie Recommendation System

In order to further evaluate the performance of the proposed algorithm, we test our algorithm on the MovieLens dataset [41]. MovieLens dataset contains anonymous ratings of movies by users. Three subsets of the dataset are employed: 100 K, 1 M and 10 M, with varying numbers of users, movies and ratings as described in Table 2.

**Table 2.** Dataset descriptions. The number of users, items and ratings used in each dataset.

Dataset		Users	Movies	Ratings
MovieLens	100 K	943	1682	100,000
	1 M	6040	3449	999,714
	10 M	69,878	10,677	10,000,054

The experiments were conducted on a workstation equipped with an Intel Xeon Gold 5218R processor (20 cores/40 threads), 64 GB of RAM and dual NVIDIA GeForce RTX 4090 GPU. The software environment is Ubuntu 22.04.4 LTS and MATLAB R2020a. The key measuring metrics are computational efficiency via GPU seconds (Time), recovery accuracy via RMSE and the objective value, which are adopted to determine the algorithm's superiority. The comparative results of different algorithms on the MovieLens dataset subsets are presented in Table 3. It should be emphasized that our advantages are not apparent with small data. However, our algorithms have a marked advantage with big data.



**Table 3.** Comparative results of tested algorithms on the MovieLens dataset subsets.

Dataset		Method	Time	RMSE	Objective Value
MovieLens	100 K	APIRNN	2.06	1.0410	$6.4081 \times 10^2$
		PIRNNE	1.11	1.0216	$7.8860 \times 10^2$
		PIRNNE-mLS	2.00	1.0468	$5.9634 \times 10^2$
		PIRNNE-nLS	1.98	1.0450	$6.1770 \times 10^2$
	1 M	APIRNN	6.24	0.8855	$1.1575 \times 10^5$
		PIRNNE	7.88	1.0343	$1.2641 \times 10^5$
		PIRNNE-mLS	5.23	0.8844	$1.0311 \times 10^5$
		PIRNNE-nLS	5.22	0.8844	$1.0311 \times 10^5$
	10 M	APIRNN	28.23	0.9483	$1.9444 \times 10^6$
		PIRNNE	238.86	1.0063	$2.1513 \times 10^6$
		PIRNNE-mLS	14.49	0.9483	$1.9435 \times 10^6$
		PIRNNE-nLS	13.78	0.9483	$1.9435 \times 10^6$

## 6. Conclusions

This paper addresses a class of nonconvex and nonsmooth optimization problems that are commonly encountered in various applications. Based on existing dimension reduction and extrapolation techniques, we propose a more generalized proximal iterative reweighted nuclear norm method. This method utilizes a line search mechanism to avoid parameter constraints, thereby providing greater flexibility in parameter selection. As a result, it is feasible to expand the application of this method in the future. In theory, we prove the subsequential convergence. Furthermore, for the case of monotone line search, we prove the global convergence and linear convergence rate of the algorithm under the KL framework. Finally, we validate the effectiveness of the algorithm through numerical results on synthetic and real data. We will construct a new nonconvex optimization model with distributed characteristics and design corresponding algorithms [42,43] based on the low rank of matrix. This will be our future research work.

**Author Contributions:** Conceptualization, Z.G.; methodology, Z.G.; formal analysis, Z.G.; writing—original draft preparation, Z.G.; software, X.Z.; validation, X.Z. and S.Z.; writing—review and editing, Z.G., S.Z., X.Z. and Y.C.; visualization, X.Z. and S.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (12471290), Natural Science Research of the Jiangsu Higher Education Institutions of China (20KJA520003), Six Talent Peaks Project of Jiangsu Province (JY-051), Open Fund of the Key Laboratory of NSLSCS, Ministry of Education, Suqian Sci&Tech Program (M202206) and Qing Lan Project.

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

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Ge, Z.L.; Zhang, X.; Wu, Z.M. A fast proximal iteratively reweighted nuclear norm algorithm for nonconvex low-rank matrix minimization problems. *Appl. Numer. Math.* **2022**, *179*, 66–86. [\[CrossRef\]](#)
- Argyriou, A.; Evgeniou, T.; Pontil, M. Convex multi-task feature learning. *Mach. Learn.* **2008**, *73*, 243–272. [\[CrossRef\]](#)
- Amit, Y.; Fink, M.; Srebro, N.; Ullman, S. Uncovering shared structures in multiclass classification. In Proceedings of the the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007. [\[CrossRef\]](#)
- Lu, C.Y.; Tang, J.H.; Yan, S.C.; Lin, Z.C. Generalized nonconvex nonsmooth low-rank minimization. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; IEEE Computer Society: Washington, DC, USA, 2014; pp. 4130–4137.

5. Dong, W.S.; Shi, G.M.; Li, X.; Ma, Y.; Huang, F. Compressive sensing via nonlocal low-rank regularization. *IEEE Trans. Image Process.* **2014**, *23*, 3618–3632. [\[CrossRef\]](#)
6. Fazel, M.; Hindi, H.; Boyd, S.P. Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. In Proceedings of the 2003 American Control Conference, Denver, CO, USA, 4–6 June 2003; IEEE: New York, NY, USA, 2003; pp. 2156–2162.
7. Hu, Y.; Zhang, D.B.; Ye, J.P.; Li, X.L.; He, X.F. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Trans. Pattern Anal.* **2013**, *35*, 2117–2130. [\[CrossRef\]](#)
8. Lu, C.Y.; Zhu, C.B.; Xu, C.Y.; Yan, S.C.; Lin, Z.C. Generalized singular value thresholding. In Proceedings of the 29th AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; pp. 1215–1221. [\[CrossRef\]](#)
9. Todeschini, A.; Caron, F.; Chavent, M. Probabilistic low-rank matrix completion with adaptive spectral regularization algorithms. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 845–853.
10. Toh, K.C.; Yun, S.W. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pac. J. Optim.* **2010**, *6*, 615–640.
11. Zhang, X.; Peng, D.T.; Su, Y.Y. A singular value shrinkage thresholding algorithm for folded concave penalized low-rank matrix optimization problems. *J. Glob. Optim.* **2024**, *88*, 485–508. [\[CrossRef\]](#)
12. Tao, T.; Xiao, L.H.; Zhong, J.Y. A Fast Proximal Alternating Method for Robust Matrix Factorization of Matrix Recovery with Outliers. *Mathematics* **2025**, *13*, 1466. [\[CrossRef\]](#)
13. Cui, A.G.; He, H.Z.; Yuan, H. A Designed Thresholding Operator for Low-Rank Matrix Completion. *Mathematics* **2024**, *12*, 1065. [\[CrossRef\]](#)
14. Gong, P.H.; Zhang, C.S.; Lu, Z.S.; Huang, J.H.Z.; Ye, J.P. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; pp. 37–45.
15. Nakayama, S.; Narushima, Y.; Yabe, H.; Inexact proximal DC Newton-type method for nonconvex composite functions. *Comput. Optim. Appl.* **2024**, *87*, 611–640. [\[CrossRef\]](#)
16. Sun, T.; Jiang, H.; Cheng, L.Z. Convergence of proximal iteratively reweighted nuclear norm algorithm for image processing. *IEEE Trans. Image Process.* **2017**, *26*, 5632–5644. [\[CrossRef\]](#)
17. Phan, D.N.; Nguyen, T.N. An accelerated IRNN-Iteratively Reweighted Nuclear Norm algorithm for nonconvex nonsmooth low-rank minimization problems. *J. Comput. Appl. Math.* **2021**, *396*, 113602. [\[CrossRef\]](#)
18. Xu, Z.Q.; Zhang, Y.L.; Ma, C.; Yan, Y.C.; Peng, Z.L.; Xie, S.L.; Wu, S.Q.; Yang, X.K. LERE: Learning-Based Low-Rank Matrix Recovery with Rank Estimation. In Proceedings of the 38th AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; pp. 16228–16236. [\[CrossRef\]](#)
19. Wen, Y.W.; Li, K.X.; Chen, H.F. Accelerated matrix completion algorithm using continuation strategy and randomized SVD. *J. Comput. Appl. Math.* **2023**, *429*, 115215. [\[CrossRef\]](#)
20. Zhang, H.M.; Qian, F.; Shi, P.; Du, W.L.; Tang, Y.; Qian, J.J.; Gong, C.; Yang, J. Generalized Nonconvex Nonsmooth Low-Rank Matrix Recovery Framework With Feasible Algorithm Designs and Convergence Analysis. *IEEE Trans. Neur. Net. Lear.* **2023**, *34*, 5342–5353. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Li, B.J.; Pan, S.H.; Qian, Y.T. Factorization model with total variation regularizer for image reconstruction and subgradient algorithm. *Pattern Recogn.* **2026**, *170*, 112038. [\[CrossRef\]](#)
22. Guo, H.Y.; Huang, Z.H.; Zhang, X.Z. Low rank matrix recovery with impulsive noise. *Appl. Math. Lett.* **2022**, *134*, 108364. [\[CrossRef\]](#)
23. Wang, H.; Wang, Y.; Yang, X.Y. Efficient Active Manifold Identification via Accelerated Iteratively Reweighted Nuclear Norm Minimization. *J. Mach. Learn. Res.* **2024**, *25*, 1–44. <https://jmlr.org/papers/v25/23-0449.html>
24. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [\[CrossRef\]](#)
25. Li, H.; Lin, Z.C. Accelerated proximal gradient methods for nonconvex programming. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 377–387.
26. Grippo, L.; Lampariello, F.; Lucidi, S. A nonmonotone line search technique for Newton’s method. *SIAM J. Numer. Anal.* **1986**, *23*, 707–716. [\[CrossRef\]](#)
27. Wright, S.J.; Nowak, R.; Figueiredo, M.A.T. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.* **2008**, *57*, 2479–2493. [\[CrossRef\]](#)
28. Wu, Z.M.; Li, C.S.; Li, M.; Andrew, L. Inertial proximal gradient methods with Bregman regularization for a class of nonconvex optimization problems. *J. Glob. Optim.* **2021**, *79*, 617–644. [\[CrossRef\]](#)
29. Liu, J.Y.; Cui, Y.; Pang, J.S.; Sen, S. Two-stage stochastic programming with linearly bi-parameterized quadratic recourse. *SIAM J. Optimiz.* **2020**, *30*, 2530–2558. [\[CrossRef\]](#)

30. Wang, J.Y.; Petra, C.G. A sequential quadratic programming algorithm for nonsmooth problems with upper- $C^2$  Objective. *SIAM J. Optimiz.* **2023**, *33*, 2379–2405. [[CrossRef](#)]
31. Yang, L. Proximal gradient method with extrapolation and line search for a class of nonconvex and nonsmooth problems. *J. Optimiz. Theory App.* **2024**, *200*, 68–103. [[CrossRef](#)]
32. Attouch, H.; Bolte, J.; Svaiter, B.F. Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. *Math. Program.* **2013**, *137*, 91–129. [[CrossRef](#)]
33. Bolte, J.; Sabach, S.; Teboulle, M. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.* **2014**, *146*, 459–494. [[CrossRef](#)]
34. Kurdyka, K. On gradients of functions definable in o-minimal structures. *Ann. I. Fourier* **1998**, *48*, 769–783. [[CrossRef](#)]
35. Attouch, H.; Bolte, J.; Redont, P.; Soubeyran, A. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality. *Math. Oper. Res.* **2010**, *35*, 438–457. [[CrossRef](#)]
36. Ge, Z.L.; Wu, Z.M.; Zhang, X. An extrapolated proximal iteratively reweighted method for nonconvex composite optimization problems. *J. Glob. Optim.* **2023**, *86*, 821–844. [[CrossRef](#)]
37. Guo, K.; Han, D.R. A note on the Douglas-Rachford splitting method for optimization problems involving hypoconvex functions. *J. Glob. Optim.* **2018**, *72*, 431–441. [[CrossRef](#)]
38. Wen, B.; Chen, X.J.; Pong, T.K. Linear convergence of proximal gradient algorithm with extrapolation for a class of nonconvex nonsmooth minimization problems. *SIAM J. Optimiz.* **2017**, *27*, 124–145. [[CrossRef](#)]
39. Wu, Z.M.; Li, M. General inertial proximal gradient method for a class of nonconvex nonsmooth optimization problems. *Comput. Optim. Appl.* **2019**, *73*, 129–158. [[CrossRef](#)]
40. Nesterov, Y. *Introductory Lectures on Convex Optimization: A Basic Course*; Springer: New York, NY, USA, 2013. [[CrossRef](#)]
41. Harper, F.M.; Konstan, J.A. The MovieLens Datasets: History and Context. *ACM TiiS* **2015**, *5*, 1–19. [[CrossRef](#)]
42. Li, S.; Li, Q.W.; Zhu, Z.H.; Tang, G.G.; Wakin, M.B. The global geometry of centralized and distributed low-rank matrix recovery without regularization. *IEEE Signal Proc. Let.* **2020**, *27*, 1400–1404. [[CrossRef](#)]
43. Doostmohammadian, M.; Gabidullina, Z.R.; Rabiee, H.R. Nonlinear perturbation-based non-convex optimization over time-varying networks. *IEEE Trans. Netw. Sci. Eng.* **2024**, *11*, 6461–6469. [[CrossRef](#)]

**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.