

Article

Not peer-reviewed version

---

# A Novel Symmetrical Inertial Alternating Direction Method of Multipliers with Proximal Term for Nonconvex Optimization with Applications

---

[Ji-hong Li](#) , [Heng-you Lan](#) <sup>\*</sup> , [Si-yuan Lin](#)

Posted Date: 15 May 2025

doi: 10.20944/preprints202505.1063.v1

Keywords: convergence analysis; nonconvex optimization; alternating direction method of multipliers; inertial term; Kurdyka-Lojasiewicz inequality



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

# A Novel Symmetrical Inertial Alternating Direction Method of Multipliers with Proximal Term for Nonconvex Optimization with Applications

Ji-hong Li <sup>1</sup>, Heng-you Lan <sup>1,2,\*</sup> and Si-yuan Lin <sup>1</sup>

<sup>1</sup> College of Mathematics and Statistics, Sichuan University of Science and Engineering, Zigong 643000, China

<sup>2</sup> Sichuan Province University Key Laboratory of Bridge Non-destruction Detecting and Engineering Computing, Zigong 643000, China

\* Correspondence: hengyoulan@163.com

**Abstract:** In this paper, we propose a novel alternating direction method of multipliers based on acceleration technique involving two symmetrical inertial terms for a class of nonconvex optimization problems with a two-block structure. To address the nonconvex subproblem, we introduce a proximal term to reduce the difficulty of solving this subproblem. For smooth subproblem, we employ a gradient descent method on the augmented Lagrangian function, which significantly reduces the computational complexity. Under the assumptions that the generated sequence is bounded and the auxiliary function satisfies Kurdyka-Łojasiewicz property, we establish the global convergence of the proposed algorithm. Finally, the effectiveness and superior performance of the proposed algorithm are validated through numerical experiments in signal processing and SCAD problems.

**Keywords:** convergence analysis; nonconvex optimization; alternating direction method of multipliers; inertial term; Kurdyka-Łojasiewicz inequality

**MSC:** 90C26; 65K05; 49K35; 41A25

## 1. Introduction

In recent years, nonconvex optimization problems have found widespread applications in science and engineering. For instance, Mohammadreza et al. [1] investigated the optimization of local nonconvex objective functions in time-varying networks based on the gradient tracking algorithm. And then, the researchers had explored optimizing nonconvex objective functions in multi-node networks under imperfect data exchange links. Zhang et al. [2] pointed out that traditional optimization methods often lead to target feature compression and information loss in motor imagery decoding, thereby reducing classification performance. Further, to address the high dimensionality and small sample size characteristics of motor imagery signals, Zhang et al. [2] proposed a nonconvex sparse regularization model constructed using the Cauchy function. This approach enables more accurate extraction of target features across multiple datasets while effectively suppressing noise interference. In addition, Tiddeman and Ghahremani [3] combined wavelet transforms with principal component analysis to propose a class of principal component waveform networks for solving linear inverse problems, they fully utilized the symmetry in wavelet transforms during the wavelet decomposition, ensuring the effectiveness of image reconstruction. For more related works, one can see [4–6] and the references therein.

It is well known that recovering sparse signals from incomplete observations is an important research direction in practical applications. The core objective is to find the optimal sparse solution to a system of linear equations, which can be formulated as the following model [7]:

$$\min c\|x\|_0 + \frac{1}{2}\|Ax - b\|,$$

where  $A$  is the measurement matrix,  $b$  is the observed data,  $x$  is a sparse signal,  $c > 0$  is a regularization parameter, and  $\|\cdot\|_0$  denotes the  $\ell_0$ -norm. However, Chartrand and Staneva [8] pointed out that the above problem represents a class of problems that are fundamentally difficult to solve. To overcome this challenge, Zeng et al. [9] proposed a relaxed objective function by replacing the  $l_0$  regularization with the  $l_{\frac{1}{2}}$  regularization, the problem is transformed into a more tractable nonconvex optimization problem. Therefore, adopting this modification becomes more reasonable in signal recovery problems, which leads to the following two-block nonconvex optimization problem:

$$\min c\|x\|_{\frac{1}{2}} + \frac{1}{2}\|Ax - b\|^2, \quad (1)$$

where  $\|x\|_{\frac{1}{2}} = \sum_{i=1}^n |x_i|^{\frac{1}{2}}$ . In general, we would consider introducing an auxiliary variable  $y$  to reformulate the problem (1) as follows

$$\begin{aligned} \min c\|x\|_{\frac{1}{2}} + \frac{1}{2}\|y\|^2 \\ \text{s.t. } Ax - y = b. \end{aligned} \quad (2)$$

Zeng et al. [9] pointed out that the iterative soft-thresholding algorithm can be used to solve the regularization problem, which was validated in context of problem (2). Meanwhile, Chen and Selesnick [10] performed a performance validation of model (2) using an improved overlapping shrinkage algorithm. In addition, further related works can be found in [11,12].

In statistical optimization, certain penalty methods exhibit limitations, such as vulnerability to data circumvention and biased estimation of significant variables [14]. To address these issues, Fan and Li [14] proposed the smoothly clipped absolute deviation (SCAD) penalty function. They developed optimization algorithms to solve non-concave penalized likelihood problems and demonstrated that this method possesses asymptotic oracle properties. Remarkably, with appropriate regularization parameter selection, the results can achieve nearly identical performance to the known true model. SCAD penalty problem proposed by scholars can be conceptually understood as

$$\begin{aligned} \min \sum_{i=1}^n h_k(|x_i|) + \frac{1}{2}\|y\|^2 \\ \text{s.t. } Ax - y = b, \end{aligned} \quad (3)$$

with  $A \in \mathbb{R}^{m \times n}$ ,  $y, b \in \mathbb{R}^m$ ,  $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$  and the penalty function  $h_k$  in the objective, we refer readers to (26) later. As shown above, problems of the forms (2) and (3) can be generalized to the following nonconvex optimization problem:

$$\begin{aligned} \min f(x) + g(y) \\ \text{s.t. } Ax + y = b, \end{aligned} \quad (4)$$

where  $f$  is a lower semicontinuous function from  $\mathbb{R}^n$  to  $\mathbb{R}$ , and  $g : \mathbb{R}^m \rightarrow \mathbb{R}$  is a differentiable function whose gradient is  $L$ -Lipschitz continuous with  $L > 0$ . Here,  $A$  denotes a matrix in  $\mathbb{R}^{m \times n}$ ,  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^m$ . Variants of model (4) have found applications in various fields, such as statistical learning [15–17], penalized zero-variance discriminant analysis [18] and image reconstruction [19,20].

It is well known that the alternating direction method of multipliers (ADMM) has gained widespread attention due to its balance between performance and efficiency. When the subproblems are independent, ADMM exhibits a unique symmetry. In fact, with appropriately designed update steps, this symmetry ensures that the convergence of ADMM is independent of the order in which the subproblems are updated [21]. In recent years, as nonconvex optimization problems have gained increasing attention, the convergence analysis of ADMM in nonconvex settings has become a research hotspot. Hong et al. [22], recognizing the strong empirical performance of ADMM on

nonconvex problems but the lack of theoretical guarantees, not only established the convergence theory for nonconvex ADMM but also overcame the limitation on the number of variable blocks. Wang et al. [23] demonstrated that incorporating the Bregman distance into ADMM can effectively simplify the computation of subproblems, emphasizing the feasibility of ADMM in nonconvex settings. Ding et al. [24] proposed a class of Semi-Proximal ADMM for solving low-rank matrix recovery problems. In the presence of noisy matrix data, by minimizing the nuclear norm, they effectively addressed the issues of Gaussian noise and related mixed noise. Guo et al. [25] provided insights into solving large-scale nonconvex optimization problems using ADMM. For more related work, readers may refer to [26–29].

The inertial acceleration technique, derived from the heavy-ball method, utilizes information from previous iterations to construct affine combinations [30]. Additionally, we observe that inertial techniques can employ different extrapolation strategies during the optimization process to enhance convergence speed. In their study of the general inertial proximal gradient method, Wu and Li [31] proposed two distinct extrapolation strategies to flexibly adjust the algorithm's convergence rate. Chen et al. [32] investigated an inertial proximal ADMM and established the global convergence of iterates under appropriate assumptions. In fact, Chao et al. [33] also discovered that embedding the inertial term into the  $y$ -subproblem can significantly improve the algorithm's convergence speed. Moreover, compared to the standard inertial update step  $\bar{x}^k = x^k + \eta(x^k - x^{k-1})$ , Wang et al. [34] considered a different update scheme,  $\bar{x}^k = x^k + \eta(x^k - \bar{x}^k)$ . This update step not only preserves the acceleration effect of inertia but also reduces computational errors introduced by the inertial term updates.

Unfortunately, the work of Wang et al. only considered the inertial update step for  $x$ . Inspired by the work in [34], we propose a novel symmetrical inertial alternating direction method of multipliers with proximal term (NIP-ADMM). Building upon Wang et al.'s inertial update step, we introduce an additional inertial update for  $y$  and incorporate  $\bar{y}^k$  into the  $x$ -subproblem update, this form of inertial update ensures that the primal variables are treated equally, thereby achieving faster acceleration. To simplify the computation of the subproblems, we introduce an approximation term in the  $x$ -subproblem, which under certain conditions, allows the nonconvex subproblem to be transformed into an approximate projection-type problem. Furthermore, since  $g(y)$  is convex, this property ensures that  $g(y)$  is well-defined, this enables us to abandon the traditional ADMM update scheme and instead adopt a gradient descent approach. This method requires only the computation of gradients at each iteration, significantly reducing computational complexity. Consequently, it offers substantial advantages when handling high-dimensional or large-scale datasets.

The structure of this paper is as follows. In Section 2, we review essential results required for further analysis. We present NIP-ADMM and analyze its convergence in Section 3. Numerical experiment and application to signal recovery in Section 4 highlight the benefits of the majorization and inertial techniques. Lastly, in Section 5, we sent a conclusion.

## 2. Preliminaries

In this section, we introduce key notations and definitions that are essential for the results to be developed and are utilized in the subsequent sections.

Assume  $\langle x, y \rangle = x^T y$ ,  $\|x\| = \sqrt{\langle x, x \rangle}$ . If matrix  $Q$  is a positive definite (semi-definite positive) matrix, then we have  $Q \succ 0$  ( $Q \succeq 0$ ). Given any  $n \times n$  matrix  $Q \succeq 0$  and a vector  $x \in \mathbb{R}^n$ , let  $\|x\|_Q := \sqrt{x^T Q x}$  be  $Q$ -norm of  $x$ . For the matrix  $G$ , we define  $\lambda_{\min(G)}$  and  $\lambda_{\max(G)}$  as the smallest and largest eigenvalues of  $G^T G$ , respectively. If we denote  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ , then the domain of the function is defined as  $\text{dom } f = \{x \in \mathbb{R}^n | f(x) < +\infty\}$ .

**Definition 1.** Let  $S \subseteq \mathbb{R}^n$ . Then the distance from point  $x \in \mathbb{R}^n$  to  $S$  is defined as  $d(x, S) = \inf_{y \in S} \|y - x\|$ . In particular, if  $S = \emptyset$ , then  $d(x, S) = +\infty$ .

**Definition 2.** For a differentiable convex function  $F : \mathbb{R}^n \rightarrow \mathbb{R}$ , the Bregman distance is defined by

$$\Delta_F(p, q) = F(p) - F(q) - \langle \nabla F(q), p - q \rangle,$$

where  $p, q \in \mathbb{R}^n$ .

**Definition 3.** Assume  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper lower semicontinuous function.

(i) The Frechet sub-differential of  $f$  at  $x \in \text{dom} f$  is denoted by  $\hat{\partial}f(x)$  and defined as:

$$\hat{\partial}f(x) = \left\{ \bar{x} \in \mathbb{R}^n : \liminf_{y \neq x, y \rightarrow x} \frac{f(y) - f(x) - \langle \bar{x}, y - x \rangle}{\|y - x\|} \geq 0 \right\}.$$

Among others, we set  $\hat{\partial}f(x) = \emptyset$  when  $x \notin \text{dom} f$ .

(ii) The limiting sub-differential of  $f$  at  $x \in \text{dom} f$  is written as  $\partial f(x)$  and defined by

$$\partial f(x) = \left\{ \bar{x} \in \mathbb{R}^n : \exists x_k \rightarrow x, f(x_k) \rightarrow f(x), \hat{x}_k \in \hat{\partial}f(x_k), \hat{x}_k \rightarrow \bar{x} \right\}.$$

**Proposition 1.** The sub-differential of a lower semicontinuous function  $f$  possesses several fundamental and significant properties as follows:

- (i) From Definition 3, which implies that  $\hat{\partial}f(x) \subseteq \partial f(x)$  holds for all  $x \in \mathbb{R}^n$ , and given that  $\partial f(x)$  is a closed set,  $\hat{\partial}f(x)$  is also a closed set.
- (ii) Suppose that  $(x^k, y^k)$  is a sequence that converges to  $(x^*, y^*)$ , and  $f(x^k)$  converges to  $f(x^*)$  with  $y^k \in \partial f(x^k)$ . Then, by the definition of the sub-differential, we have  $y^* \in \partial f(x^*)$ .
- (iii) If  $x^*$  is a local minimum of  $f$ , then it follows that  $0 \in \partial f(x^*)$ .
- (vi) Assuming that  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  is a continuously differentiable function, we can derive:

$$\partial(f + g)(x) = \partial f(x) + \nabla g(x).$$

**Definition 4.** We consider the point  $(x^*, y^*, \lambda^*)$  to be a critical point of the augmented Lagrangian function  $\hat{L}_\beta(x, y, \lambda)$  if it satisfies the following conditions:

$$\begin{cases} A^T \lambda^* \in \partial f(x^*), \\ B^T \lambda^* = \nabla g(y^*), \\ Ax^* + By^* = b. \end{cases}$$

**Definition 5 ([36]).** (Kurdyka-Łojasiewicz property (KLP)) Let  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower semicontinuous function. For  $\hat{p} \in \partial f$  ( $\partial f \neq \emptyset$ ), if there exists  $\varsigma \in (0, +\infty)$ , a neighborhood  $U$  of  $\hat{p}$ , and a function  $\varphi \in Q_\varphi$ , here  $Q_\varphi$  is the set of the concave function  $\varphi : [0, \varsigma) \rightarrow [0, +\infty)$ , then for any  $p \in \{p \in U \mid f(\hat{p}) < f(p) < f(\hat{p}) + \varsigma\}$ , the following inequality holds

$$\varphi'(f(p) - f(\hat{p}))d(0, \partial f(p)) \geq 1,$$

and we call that  $f$  satisfies KLP.

**Lemma 1 ([35]).** Suppose that the matrix  $B \in \mathbb{R}^{r \times p}$  is a non-zero matrix, and let  $\mu_B$  denote the smallest positive eigenvalue of the matrix  $BB^T$ . Then for each  $u \in \mathbb{R}^p$ , the following holds:

$$\|P_{Q^T} u\| \leq \frac{1}{\mu_B} \|Bu\|.$$

**Lemma 2 ([36]).** Assume  $B(x, y) = f(x) + g(y)$ , where  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  and  $g : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  are both proper lower semicontinuous functions. Then, for any  $(x, y) \in \text{dom} B = \text{dom} f \times \text{dom} g$ , we can obtain

$$\partial B(x, y) = \partial_x f(x, y) \times \partial_y g(x, y).$$

**Lemma 3 ([37]).** (Uniformized KLP) Let  $\Omega$  be a compact set and  $Q_\varphi$  be the same as in Definition 5. If a proper lower semicontinuous function  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is fixed at a point in  $\Omega$  and satisfies the



KLP at every point on  $\Omega$ , and there exist  $\varrho > 0$ ,  $\varsigma > 0$ , and  $\varphi \in Q_\varphi$  such that for any  $\hat{x} \in \Omega$  and  $x \in \{x \in \mathbb{R}^n : d(x, \Omega) < \varrho\} \cap [f(\hat{x}) < f(x) < f(\hat{x}) + \varsigma]$ , then the following inequality is satisfied:

$$\varphi'(f(x) - f(\hat{x}))d(0, \partial f(x)) \geq 1.$$

**Lemma 4 ([38]).** If the function  $c : \mathbb{R}^n \rightarrow \mathbb{R}$  is continuously differentiable, and  $\nabla c$  is Lipschitz continuous with constant  $L \geq 0$ , then for any  $x, y \in \mathbb{R}^n$ , the following result holds:

$$|c(y) - c(x) - \langle \nabla c(x), y - x \rangle| \leq \frac{L}{2} \|y - x\|^2.$$

### 3. Algorithm and Convergence Analysis

In this section, we first present the definition of the augmented Lagrangian function associated with problem (4) as follows:

$$L_\beta(x, y, \lambda) = f(x) + g(y) + \langle \lambda, Ax + y - b \rangle + \frac{\beta}{2} \|Ax + y - b\|^2, \quad (5)$$

where  $\lambda$  denotes the augmented Lagrange multiplier, and  $\beta > 0$  is a penalty parameter. Following this, we propose the NIP-ADMM for solving the problem (4), the proposed algorithm is outlined below:

- Remark 1.** (i) In NIP-ADMM, the inertial parameters  $\eta$  and  $\theta$  are both in  $(0, 1]$ , and  $S$  is a positive semi-definite matrix.
- (ii) The update scheme for the  $y$ -subproblem adopts the gradient descent method, where  $\nabla_y L_\beta$  is the gradient of the function  $L$  with respect to  $y$ , and  $\gamma$  is called the learning rate.
- (iii) The inertial structure we adopted employs a structurally balanced acceleration strategy. This update strategy is mathematically symmetric, with the only distinction being the values of the parameters  $\eta$  and  $\theta$ .

According to Algorithm 1, the optimality conditions for NIP-ADMM are obtained as

$$\begin{cases} 0 \in \partial f(x^{k+1}) - A^T \lambda^k + \beta A^T (Ax^{k+1} + \bar{y}^k - b) + S(x^{k+1} - \bar{x}^k), \\ 0 = \nabla g(y^k) + \lambda^k + \beta(Ax^{k+1} + y^k - b) - \frac{1}{\gamma}(y^{k+1} - y^k). \end{cases} \quad (6)$$

Before concluding this section, we present the following fundamental assumptions, which are essential for the convergence analysis.

---

#### Algorithm 1 NIP-ADMM

---

**Initialization:** Input  $x^1, y^1$ , and  $\lambda^1$ , let  $\bar{x}^0 = x^1$ , and  $\bar{y}^0 = y^1$ . Given constants  $\eta, \theta, \gamma, \beta$ . Set  $k = 1$ .

**While** "not converge" **Do**

1° Compute  $(\bar{x}^k, \bar{y}^k) = (x^k, y^k) + \theta(x^k - \bar{x}^{k-1}, 0) + \eta(0, y^k - \bar{y}^{k-1})$ .

2° Execute  $\arg \min \{L_\beta(x, \bar{y}^k, \lambda^k) + \frac{1}{2} \|x - \bar{x}^k\|_S^2\}$  to determine  $x^{k+1}$ .

3° Calculate  $y^{k+1} = y^k - \gamma \nabla_y L_\beta(x^{k+1}, y^k, \lambda^k)$ .

4° Update dual variable  $\lambda^{k+1} = \lambda^k + \beta(Ax^{k+1} + y^{k+1} - b)$ .

5° Let  $k = k + 1$ .

**End While**

**Output:** output  $(x^{k+1}, y^{k+1}, \lambda^{k+1})$  of the problem (4).

---

**Assumption 1.** (i)  $f : \mathbb{R}^{n_1} \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper lower semicontinuous function.  $g : \mathbb{R}^{n_2} \rightarrow \mathbb{R}$  is continuously differentiable, and  $\nabla g$  is Lipschitz continuous with a Lipschitz constant  $l_g > 0$ .

(ii)  $S$  is a positive semidefinite matrix.

(iii) For convenience, we introduce the following symbols:

$$\zeta = (x, y, \lambda), \quad \zeta^k = (x^k, y^k, \lambda^k), \quad \zeta^* = (x^*, y^*, \lambda^*), \quad \xi = \frac{1}{\gamma} - \beta,$$

$$\begin{aligned}\hat{\zeta}^k &= (x^k, y^k, \lambda^k, \bar{x}^k, x^{k-1}, y^{k-1}), \quad \sigma_0 = \frac{1}{\gamma} - \frac{L + \beta}{2} - \frac{2\xi^2}{\beta} - \frac{2(\xi + L)^2}{\beta}, \\ \hat{L}_\beta(\hat{\zeta}^k) &= L_\beta(\zeta^k) + \frac{(\xi + L)^2}{\beta} \|y^{k+1} - y^k\|^2 + \frac{1}{2} \|\bar{x}^{k+1} - x^k\|_S^2.\end{aligned}$$

(iv) To analyze the monotonicity of  $\{\hat{L}_\beta(\hat{\zeta}^k)\}$ , we set  $\sigma_0 > 0$ .

**Lemma 5.** *If Assumption 1 holds, for any  $k \geq 1$ , then*

$$\begin{aligned}\hat{L}_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}, \bar{x}^{k+1}, x^k, y^k) - \hat{L}_\beta(x^k, y^k, \lambda^k, \bar{x}^{k-1}, x^{k-1}, y^{k-1}) \\ \leq -\frac{1-\eta^2}{2} \|x^k - \bar{x}^{k-1}\|_S^2 - \sigma_0 \|y^{k+1} - y^k\|^2,\end{aligned}\tag{7}$$

where  $\eta \in (0, 1]$  is the inertial parameter in Algorithm 1.

**Proof.** According to the definition of the Lagrangian function, one gets

$$\begin{aligned}L_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}) - L_\beta(x^{k+1}, y^{k+1}, \lambda^k) \\ = \langle \lambda^{k+1} - \lambda^k, Ap^{k+1} + q^{k+1} - b \rangle = \frac{1}{\beta} \|\lambda^{k+1} - \lambda^k\|^2,\end{aligned}\tag{8}$$

and we also can know that

$$\begin{aligned}L_\beta(x^{k+1}, y^{k+1}, \lambda^k) - L_\beta(x^{k+1}, y^k, \lambda^k) \\ = g(y^{k+1}) - g(y^k) - \langle \lambda^k, y^{k+1} - y^k \rangle + \frac{\beta}{2} \|Ax^{k+1} + y^{k+1} - b\|^2 \\ - \frac{\beta}{2} \|Ax^{k+1} + y^k - b\|^2.\end{aligned}\tag{9}$$

It follows from (6), (9) and Lemma 4 that

$$\begin{aligned}g(y^{k+1}) - g(y^k) - \langle \lambda^k, y^{k+1} - y^k \rangle \\ \leq \langle \nabla g(y^k), y^{k+1} - y^k \rangle + \frac{L}{2} \|y^{k+1} - y^k\|^2 - \langle \lambda^k, y^{k+1} - y^k \rangle \\ = \langle \lambda^{k+1} + (\frac{1}{\gamma} - \beta)(y^k - y^{k+1}), y^{k+1} - y^k \rangle + \frac{L}{2} \|y^{k+1} - y^k\|^2 - \langle \lambda^k, y^{k+1} - y^k \rangle \\ = (\frac{L}{2} + \beta - \frac{1}{\gamma}) \|y^{k+1} - y^k\|^2 + \langle \lambda^{k+1} - \lambda^k, y^{k+1} - y^k \rangle,\end{aligned}$$

and we get

$$\begin{aligned}\frac{\beta}{2} \|Ax^{k+1} + y^{k+1} - b\|^2 - \frac{\beta}{2} \|Ax^{k+1} + y^k - b\|^2 \\ = \frac{\beta}{2} \|Ax^{k+1} + y^{k+1} - b\|^2 - \frac{\beta}{2} \|(Ax^{k+1} + y^{k+1} - b) + (y^k - y^{k+1})\|^2 \\ = \frac{\beta}{2} \|Ax^{k+1} + y^{k+1} - b\|^2 - \frac{\beta}{2} (\|Ax^{k+1} + y^{k+1} - b\|^2 + \|y^k - y^{k+1}\|^2) \\ + \langle \beta(Ax^{k+1} + y^{k+1} - b), y^k - y^{k+1} \rangle \\ = -\frac{\beta}{2} \|y^k - y^{k+1}\|^2 + \langle \lambda^k - \lambda^{k+1}, y^k - y^{k+1} \rangle.\end{aligned}$$

Combining the above two formulas, one can declare

$$L_\beta(x^{k+1}, y^{k+1}, \lambda^k) - L_\beta(x^{k+1}, y^k, \lambda^k) \leq (\frac{L + \beta}{2} - \frac{1}{\gamma}) \|y^{k+1} - y^k\|^2.\tag{10}$$

Since  $x^{k+1}$  is the optimal solution to the subproblem with respect to  $x$ , one knows that

$$\begin{aligned} L_{\beta}(x^{k+1}, y^k, \lambda^k) - L_{\beta}(x^k, y^k, \lambda^k) &\leq \frac{1}{2} \|x^k - \bar{x}^k\|_S^2 - \frac{1}{2} \|x^k - \bar{x}^{k+1}\|_S^2 \\ &\leq \frac{1}{2} \|x^k - \bar{x}^{k-1}\|_S^2 - \frac{1}{2} \|x^k - \bar{x}^{k+1}\|_S^2 - \frac{1-\eta^2}{2} \|x^k - \bar{x}^{k-1}\|_S^2. \end{aligned} \quad (11)$$

Noticing Algorithm 1 and (6), one can see

$$\left(\frac{1}{\gamma} - \beta\right)(y^k - y^{k+1}) = \nabla g(y^k) - \lambda^{k+1}.$$

Thus, it is natural to derive the following process:

$$\begin{aligned} \|\lambda^{k+1} - \lambda^k\|^2 &= \|\nabla g(y^k) - \nabla g(y^{k-1}) + \left(\frac{1}{\gamma} - \beta\right)(y^{k+1} - y^k) - \left(\frac{1}{\gamma} - \beta\right)(y^k - y^{k-1})\|^2 \\ &\leq 2\left(L + \frac{1}{\gamma} - \beta\right)^2 \|y^k - y^{k-1}\|^2 + 2\left(\frac{1}{\gamma} - \beta\right)^2 \|y^{k+1} - y^k\|^2. \end{aligned} \quad (12)$$

Combining (8) and Equations (10)–(12), one can draw the following conclusions:

$$\begin{aligned} L_{\beta}(x^{k+1}, y^{k+1}, \lambda^{k+1}) &+ \frac{(\xi + L)^2}{\beta} \|y^{k+1} - y^k\|^2 + \frac{1}{2} \|\bar{x}^{k+1} - x^k\|_S^2 \\ &\leq L_{\beta}(x^k, y^k, \lambda^k) + \frac{(\xi + L)^2}{\beta} \|y^k - y^{k-1}\|^2 + \frac{1}{2} \|\bar{x}^k - x^{k-1}\|_S^2 \\ &\quad - \frac{1-\eta^2}{2} \|x^k - \bar{x}^{k-1}\|_S^2 - \sigma_0 \|y^{k+1} - y^k\|^2, \end{aligned} \quad (13)$$

where  $\xi = \frac{1}{\gamma} - \beta$  and  $\sigma_0 = \frac{1}{\gamma} - \frac{L+\beta}{2} - \frac{2\xi^2}{\beta} - \frac{2(\xi+L)^2}{\beta}$ , and we obtain the desired conclusion.  $\square$

According to Assumption 1 with  $\sigma_0 > 0$  and  $\eta \in (0, 1]$ , the monotonic non-increasing property of the sequence  $\{\hat{L}_{\beta}(\hat{\zeta}^k)\}$  is guaranteed.

**Lemma 6.** *If the sequence  $\zeta^k := \{x^k, y^k, \lambda^k\}$  generated by the algorithm is bounded, then we have*

$$\|\zeta^{k+1} - \zeta^k\|^2 < +\infty.$$

**Proof.** Since  $\{\zeta^k\}$  is bounded, it is evident that  $\hat{\zeta}^k$  is also bounded. Moreover, there exists an accumulation point, let us assume it to be  $\zeta^*$ , and there exists a subsequence  $\{\zeta_j^k\}$  of  $\{\zeta^k\}$  such that

$$\liminf_{j \rightarrow \infty} \hat{L}_{\beta}(\zeta_j^k) \geq \hat{L}_{\beta}(\zeta^*), \quad (14)$$

which implies that  $\{\hat{L}_{\beta}(\zeta_j^k)\}$  is bounded from below. From Lemma 5 and the condition  $k \geq 2$ , it follows that

$$\begin{aligned} \sum_{k=2}^n \sigma_0 \|y^{k+1} - y^k\|^2 + \sum_{k=2}^n \frac{1-\theta^2}{2} \|x^k - \bar{x}^{k-1}\|^2 \\ \leq \hat{L}_{\beta}(\zeta^2) - \hat{L}_{\beta}(\zeta^*). \end{aligned} \quad (15)$$

Given  $\sigma_0 > 0$ ,  $\theta \in [0, 1)$ , and  $S$  is a positive semi-definite matrix, it is evident that one can derive that

$$\sum_{k=0}^n \sigma_0 \|y^{k+1} - y^k\|^2 < \infty, \quad \sum_{k=0}^n \frac{1-\theta^2}{2} \|x^k - \bar{x}^{k-1}\|^2 < \infty. \quad (16)$$



By the inertial relationship, the following conclusion can be obtained:

$$\begin{aligned}
 \|x^{k+1} - x^k\|^2 &= \|x^{k+1} - \bar{x}^k + \bar{x}^k - x^k\|^2 \\
 &= \|x^{k+1} - \bar{x}^k + \theta(x^k - \bar{x}^{k-1})\|^2 \\
 &\leq 2\|x^{k+1} - \bar{x}^k\|^2 + 2\theta\|x^k - \bar{x}^{k-1}\|^2, \\
 \|y^{k+1} - y^k\|^2 &= \|y^{k+1} - \bar{y}^k + \bar{y}^k - y^k\|^2 \\
 &= \|y^{k+1} - \bar{y}^k + \eta(y^k - \bar{y}^{k-1})\|^2 \\
 &\leq 2\|y^{k+1} - \bar{y}^k\|^2 + 2\eta\|y^k - \bar{y}^{k-1}\|^2.
 \end{aligned} \tag{17}$$

Combining (12), (16), and (17), we have

$$\sum_{k=0}^n \|x^{k+1} - x^k\|^2 < \infty, \sum_{k=0}^n \|y^{k+1} - y^k\|^2 < \infty, \sum_{k=0}^n \|\lambda^{k+1} - \lambda^k\|^2,$$

thus, we have  $\|\zeta^{k+1} - \zeta^k\|^2 < +\infty$ .  $\square$

Now we give subsequential convergence analysis of NIP-ADMM.

**Theorem 1.** (Subsequential Convergence) *The sequence  $\{\zeta^k\}$  generated by NIP-ADMM is bounded, and assume  $S$  and  $\hat{S}$  are the sets of cluster points of  $\{\zeta^k\}$  and  $\{\hat{\zeta}^k\}$ , respectively. Under the assumptions and the conditions of Lemma 5, we have the following conclusion:*

- (i)  $M$  and  $\hat{M}$  are two non-empty compact sets. As  $k \rightarrow 0$ , it follows that  $d(\zeta^k, M) \rightarrow 0$  and  $d(\hat{\zeta}^k, \hat{M}) \rightarrow 0$ .
- (ii)  $\zeta^* \in M \iff \hat{\zeta}^* \in \hat{M}$ .
- (iii)  $M \subseteq \text{crit}L_\beta$ .
- (iv) The sequence  $\{\hat{L}_\beta(\hat{\zeta}^k)\}$  converges, and  $\hat{L}_\beta(\hat{\zeta}^*) = \inf_{k \in \mathbb{N}} \hat{L}_\beta(\hat{\zeta}^k) = \lim_{k \rightarrow \infty} \hat{L}_\beta(\hat{\zeta}^k)$ .

**Proof.** (i) Based on the definitions of  $M$  and  $\hat{M}$ , the conclusion can be satisfied.

(ii) Combining Lemma 5 with the definitions of  $\zeta^*$  and  $\hat{\zeta}^*$ , we obtain the desired conclusion.

(iii) Let  $\zeta \in M$ , then we obtain that a subsequence  $\{\zeta^{k_j}\}$  of  $\{\zeta^k\}$  can converge to  $\zeta^*$ . By combining Lemma 5, as  $k \rightarrow +\infty$ , one has  $\|\zeta^{k+1} - \zeta^k\| = 0$ , which implies  $\lim_{j \rightarrow +\infty} \zeta^{k_j} = \zeta^*$ . On one hand, noting that  $x^{k+1}$  is the optimal solution to the  $x$ -subproblem, we have

$$\begin{aligned}
 &f(x^{k+1}) - \langle \lambda^k, Ax^{k+1} \rangle + \frac{\beta}{2} \|Ax^{k+1} + \bar{y}^k - b\|^2 + \frac{1}{2} \|x^{k+1} - \bar{x}^k\|_S^2 \\
 &\leq f(x^*) - \langle \lambda^k, Ax^* \rangle + \frac{\beta}{2} \|Ax^* + \bar{y}^k - b\|^2 + \frac{1}{2} \|x^* - \bar{x}^k\|_S^2.
 \end{aligned} \tag{18}$$

From Lemma 6, we know that  $\lim_{k \rightarrow +\infty} \|x^{k+1} - \bar{x}^k\| = 0$ , and combining this with  $\lim_{k_j \rightarrow +\infty} \zeta^{k_j} = \zeta^*$ , we conclude that the equality  $\lim_{k_j \rightarrow +\infty} \sup f(x^{k_j+1}) \leq f(x^*)$  holds. On the other hand, since  $f(\cdot)$  is a lower semi-continuous function, we deduce that  $\lim_{k_j \rightarrow +\infty} f(x^{k_j+1}) \geq f(x^*)$ , so one gets

$$\lim_{k_j \rightarrow +\infty} f(x^{k_j+1}) = f(x^*). \tag{19}$$

Moreover, given the closedness of  $\partial f$  and the continuity of  $\nabla g$ , along with  $k = k_j \rightarrow +\infty$  and the optimality condition of NIP-ADMM (6), we assert that

$$\begin{cases} A^T \lambda^* \in \partial f(x^*), \\ B^T \lambda^* = \nabla g(y^*), \\ Ax^* + By^* = b, \end{cases}$$

and  $\zeta^* \in \text{crit} L_\beta$  as established in Definition 4.

- (v) Let  $\hat{\zeta}^* \in \hat{M}$ , and assume that there exists a subsequence  $\{\zeta^{k_j}\}$  of  $\{\zeta^k\}$  that converges to  $\zeta^*$ . Combining the relations (14), (19), and the continuity of  $g$ , we have

$$\lim_{k_j \rightarrow +\infty} \hat{L}_\beta(\zeta^{k_j}) = \hat{L}_\beta(\zeta^*).$$

Considering that  $\hat{L}(\hat{\zeta}^k)$  is monotonically non-increasing, it follows that  $\hat{L}(\hat{\zeta}^k)$  is convergent. Consequently, for any  $\hat{\zeta}^* \in \hat{S}$ , the relationship can be established as

$$\hat{L}_\beta(\hat{\zeta}^*) = \inf_{k \in \mathbb{N}} \hat{L}_\beta(\hat{\zeta}^k) = \lim_{k \rightarrow \infty} \hat{L}_\beta(\hat{\zeta}^k). \quad \square$$

Based on the definition of the augmented Lagrangian function and the semidefiniteness of the matrix  $S$ , the following can be defined with  $\zeta^k = (x^k, y^k, \lambda^k)$ :

$$\begin{cases} \epsilon_1^{k+1} = A^T(\lambda^k - \lambda^{k+1}) + \beta(y^{k+1} - y^k) + S(x^k - x^{k+1}), \\ \epsilon_2^{k+1} = \nabla_S(y^{k+1}) - \nabla_S(y^k) + (\lambda^k - \lambda^{k+1}) + \beta(y^{k+1} - y^k) \\ \quad + \frac{1}{\gamma}(y^k - y^{k+1}) + \frac{2(\epsilon + L)^2}{\beta}(y^{k+1} - y^k), \end{cases}$$

$$\begin{cases} \epsilon_3^{k+1} = -(Ax^{k+1} + y^{k+1} - b), \\ \epsilon_4^{k+1} = S(x^{k+1} - x^k), \\ \epsilon_5^{k+1} = -S(x^{k+1} - x^k), \\ \epsilon_6^{k+1} = -\frac{2(\epsilon + L)^3}{\beta}(y^{k+1} - y^k). \end{cases}$$

Then, the following result can be obtained.

**Lemma 7.** Let  $(\epsilon_1^{k+1}, \epsilon_2^{k+1}, \epsilon_3^{k+1}, \epsilon_4^{k+1}, \epsilon_5^{k+1}, \epsilon_6^{k+1})$  be contained in  $\partial \hat{L}_\beta(\hat{\zeta}^{k+1})$ . Then, there exists  $\psi > 0$  and  $k \geq 1$  such that

$$d(0, \partial \hat{L}_\beta(\hat{\zeta}^{k+1})) \leq \psi(\|y^{k+1} - y^k\| + \|x^{k+1} - x^k\| + \|y^k - y^{k-1}\| + \|y^{k+1} - \bar{y}^k\|).$$

**Proof.** By the definition of  $\hat{L}_\beta(\cdot)$  and  $\hat{\zeta}^k = (x^k, y^k, \lambda^k, \bar{x}^k, x^{k-1}, y^{k-1})$ , we can derive that

$$\begin{cases} \partial_x \hat{L}_\beta(\zeta^{k+1}) = \partial f(x^{k+1}) - A^T \lambda^{k+1} + \beta(Ax^{k+1} - y^{k+1} - b), \\ \partial_y \hat{L}_\beta(\zeta^{k+1}) = \nabla g(y^{k+1}) - \lambda^{k+1} + \beta(Ax^{k+1} - y^{k+1} - b) + \frac{2(\epsilon + L)^2}{\beta}(y^{k+1} - y^k), \\ \partial_\lambda \hat{L}_\beta(\zeta^{k+1}) = -(Ax^{k+1} + y^{k+1} - b), \quad \partial_{\bar{x}} \hat{L}_\beta(\zeta^{k+1}) = S(\bar{x}^{k+1} - x^k), \\ \partial_{\bar{x}} \hat{L}_\beta(\zeta^{k+1}) = -S(\bar{x}^{k+1} - x^k), \quad \partial_{\bar{y}} \hat{L}_\beta(\zeta^{k+1}) = -\frac{2(\epsilon + L)^3}{\beta}(y^{k+1} - y^k). \end{cases}$$

Combining the above expression with the optimality conditions of NIP-ADMM (6), which means

$$\begin{cases} \epsilon_1^{k+1} = A^\top(\lambda^k - \lambda^{k+1}) + \beta(y^{k+1} - \bar{y}^k) + S(\bar{x}^k - x^{k+1}), \\ \epsilon_2^{k+1} = \nabla g(y^{k+1}) - \nabla g(y^k) + (\lambda^k - \lambda^{k+1}) + \beta(y^{k+1} - y^k) + \frac{1}{\gamma}(y^k - y^{k+1}) \\ \quad + \frac{2(\varepsilon + L)^2}{\beta}(y^{k+1} - y^k), \\ \epsilon_3^{k+1} = -(Ax^{k+1} + y^{k+1} - b), \quad \epsilon_4^{k+1} = S(\bar{x}^{k+1} - x^k), \\ \epsilon_5^{k+1} = -S(\bar{x}^{k+1} - x^k), \quad \epsilon_6^{k+1} = -\frac{2(\varepsilon + L)^3}{\beta}(y^{k+1} - y^k). \end{cases} \quad (20)$$

It is easy to see from Lemma 2 that  $(\epsilon_1^{k+1}, \epsilon_2^{k+1}, \epsilon_3^{k+1}, \epsilon_4^{k+1}, \epsilon_5^{k+1}, \epsilon_6^{k+1}) \in \partial \hat{L}_\beta(\hat{\zeta}^{k+1})$ . Moreover, since  $g(\cdot)$  has a Lipschitz continuous gradient with respect to  $L$ , we get

$$\|\nabla g(y^{k+1}) - \nabla g(y^k)\| \leq L\|y^{k+1} - y^k\|,$$

therefore, according to (20), there exists a positive real number  $\psi_1$  such that

$$\begin{aligned} \|(\epsilon_1^{k+1}, \epsilon_2^{k+1}, \epsilon_3^{k+1}, \epsilon_4^{k+1}, \epsilon_5^{k+1}, \epsilon_6^{k+1})\| &\leq \psi_1(\|\lambda^{k+1} - \lambda^k\| + \|y^{k+1} - \bar{y}^k\| \\ &\quad + \|x^{k+1} - \bar{x}^k\| + \|y^{k+1} - y^k\|). \end{aligned}$$

Furthermore, combining this with (12), we know that there exists  $\psi_2 > 0$  and  $k > 1$  such that:

$$\|\lambda^{k+1} - \lambda^k\| \leq \psi_2(\|y^k - y^{k-1}\| + \|y^{k+1} - y^k\|),$$

thus, by selecting  $\psi = \max\{\psi_1, \psi_2\}$  and  $k > 1$ , we can further conclude that

$$\begin{aligned} d(0, \partial \hat{L}_\beta(\hat{\zeta}^{k+1})) &\leq \|(\epsilon_1^{k+1}, \epsilon_2^{k+1}, \epsilon_3^{k+1}, \epsilon_4^{k+1}, \epsilon_5^{k+1}, \epsilon_6^{k+1})\| \\ &\leq \psi(\|y^{k+1} - y^k\| + \|x^{k+1} - \bar{x}^k\| + \|y^k - y^{k-1}\| + \|y^{k+1} - \bar{y}^k\|), \end{aligned}$$

this concludes the proof.  $\square$

**Theorem 2.** (Global convergence) Suppose the sequence  $\{\zeta^k\}$  generated by NIP-ADMM is bounded, and the assumptions hold. If  $\hat{L}_\beta$  is a KL function, then

$$\|\zeta^{k+1} - \zeta^k\| < +\infty.$$

Moreover, the sequence  $\{\zeta^k\}$  converges to a critical point of  $L_\beta$ .

**Proof.** From Theorem 1, we know that  $\lim_{k \rightarrow \infty} \hat{L}_\beta(\hat{\zeta}^k) = \hat{L}_\beta(\hat{\zeta}^*)$ . For any  $\hat{\zeta}^* \in \hat{M}$ , the proof process needs to consider the following two cases:

- (i) For any  $k_0 > 1$  and given that  $\hat{L}_\beta(\hat{\zeta}^{(k_0)}) = \hat{L}_\beta(\hat{\zeta}^*)$ , it follows from Lemma 1 and Lemma 5 that there exists a constant  $K > 0$  such that

$$K(\|x^k - \bar{x}^{k-1}\|^2 + \|y^{k+1} - y^k\|^2) \leq \bar{L}_\beta(\hat{\zeta}^k) - \bar{L}_\beta(\hat{\zeta}^{k+1}) \leq \bar{L}_\beta(\hat{\zeta}_0^k) - \bar{L}_\beta(\hat{\zeta}^*).$$

It is clear that  $K = \min\{\sigma_0, \frac{1-\eta^2}{2}\lambda_{\min(s)}\}$ . As a result, we have  $\|y^{k+1} - y^k\| < +\infty$  and  $\|x^k - \bar{x}^{k-1}\| < +\infty$ . Combining (12) and (17), it follows that  $\|x^{k+1} - x^k\| < +\infty$  and  $\|\lambda^{k+1} - \lambda^k\| < +\infty$ . Finally, for any  $k > k_0$ , we conclude that  $\|\zeta^{k+1} - \zeta^k\| < +\infty$ , and the result holds.

(ii) Assume that for any  $k > 0$ , the inequality  $L_\beta(\hat{\zeta}^k) > L_\beta(\hat{\zeta}^*)$  holds. Since

$$\lim_{k \rightarrow \infty} d(\hat{\zeta}^k, \hat{M}) = 0,$$

it follows that for any  $\epsilon_1 > 0$ , there exists  $k_1 > 0$  such that for all  $k \geq k_1$ , we have:

$$d(\hat{\zeta}^k, \hat{M}) < \epsilon_1.$$

Moreover, noting that

$$\lim_{k \rightarrow \infty} \hat{L}_\beta(\hat{\zeta}^k) = \hat{L}_\beta(\hat{\zeta}^*),$$

it implies that for any  $\epsilon_2 > 0$ , there exists  $k_2 > 0$  such that for all  $k > k_2$ , the inequality holds:

$$L_\beta(\hat{\zeta}^k) < L_\beta(\hat{\zeta}^*) + \epsilon_2.$$

Hence, given  $\epsilon_1$  and  $\epsilon_2$ , when  $k = \max\{k_1, k_2\}$ , we have

$$d(\hat{\zeta}^k, \hat{M}) < \epsilon_1, \quad L_\beta(\hat{\zeta}^*) < L_\beta(\hat{\zeta}^k) < L_\beta(\hat{\zeta}^*) + \epsilon_2.$$

And, based on Lemma 3, the following holds for all  $k > \bar{k}$ , it can be deduced that

$$\varphi'(\hat{L}_\beta(\hat{\zeta}^k) - \hat{L}_\beta(\hat{\zeta}^*))d(0, \partial \hat{L}_\beta(\hat{\zeta}^k)) \geq 1.$$

Furthermore, using the concavity of  $\varphi$ , we derive the following:

$$\begin{aligned} \varphi(L_\beta(\hat{\zeta}^k) - L_\beta(\hat{\zeta}^*)) - \varphi(L_\beta(\hat{\zeta}^{k+1}) - L_\beta(\hat{\zeta}^*)) &\geq \varphi'(L_\beta(\hat{\zeta}^k) - L_\beta(\hat{\zeta}^*)) \\ &\times (L_\beta(\hat{\zeta}^k) - L_\beta(\hat{\zeta}^{k+1})). \end{aligned}$$

Noting the fact that  $\varphi'(\hat{L}_\beta(w^k) - \hat{L}_\beta(w^*)) > 0$ , together with the conclusion obtained in Lemma 7, we can infer that

$$\begin{aligned} L_\beta(\hat{\zeta}^k) - L_\beta(\hat{\zeta}^{k+1}) &\leq \frac{\varphi(L_\beta(\hat{\zeta}^k) - L_\beta(\hat{\zeta}^*)) - \varphi(L_\beta(\hat{\zeta}^{k+1}) - L_\beta(\hat{\zeta}^*))}{\varphi'(L_\beta(\hat{\zeta}^k) - L_\beta(\hat{\zeta}^*))} \\ &\leq \Pi_{[\varphi], [k+1, k]} \psi T_{[k, k+1]}, \end{aligned} \quad (21)$$

where  $\Pi_{[\varphi], [k, k+1]}$  represents  $\varphi(L_\beta(\hat{\zeta}^k) - L_\beta(\hat{\zeta}^*)) - \varphi(L_\beta(\hat{\zeta}^{k+1}) - L_\beta(\hat{\zeta}^*))$ , and  $T_{[k, k+1]}$  represents  $\|y^{k+1} - y^k\| + \|x^{k+1} - x^k\| + \|y^k - y^{k-1}\| + \|y^{k+1} - \bar{y}^k\|$ . Combining Lemma 5, we can rewrite (21) as follows

$$\varphi(\|x^k - \bar{x}^{k-1}\|^2 + \|y^{k+1} - \bar{y}^k\|^2) \leq \Pi_{[\varphi], [k+1, k]} \psi T_{[k, k+1]}, \quad (22)$$

where (22) can be equivalently expressed as

$$\sqrt{2\|x^k - \bar{x}^{k-1}\|^2 + 2\|y^{k+1} - \bar{y}^k\|^2} \leq 2\sqrt{\frac{2\psi}{\phi}} T_{[k, k+1]}^{\frac{1}{2}} \Pi_{[\varphi], [k+1, k]}^{\frac{1}{2}}.$$

By applying the Cauchy-Schwarz inequality and multiplying both sides by 6, we obtain

$$6\left(\|x^k - \bar{x}^{k-1}\|^2 + \|y^{k+1} - \bar{y}^k\|^2\right)^{\frac{1}{2}} \leq \sqrt{\frac{2\psi}{\phi}} T_{[k, k+1]}^{\frac{1}{2}} \Pi_{[\varphi], [k+1, k]}^{\frac{1}{2}}.$$

Then, by further applying the fundamental inequality, we can deduce that

$$\begin{aligned}
 6\left(\|x^k - \bar{x}^{k-1}\| + \|y^{k+1} - \bar{y}^k\|\right) &\leq 2T_{[k,k+1]}^{\frac{1}{2}} \sqrt{\frac{18\psi}{\phi} \Pi_{[\varphi],[k+1,k]}^{\frac{1}{2}}} \\
 &\leq T_{[k,k+1]} + \frac{18\psi}{\phi} \Pi_{[\varphi],[k+1,k]} \\
 &= \|y^{k+1} - y^k\| + \|x^{k+1} - \bar{x}^k\| + \|y^k - y^{k-1}\| \\
 &\quad + \|y^{k+1} - \bar{y}^k\| + \frac{18\psi}{\phi} \Pi_{[\varphi],[k+1,k]} \\
 &\leq 2\|y^{k+1} - \bar{y}^k\| + 2\|y^k - \bar{y}^{k-1}\| + \|x^{k+1} - \bar{x}^k\| \\
 &\quad + \|y^{k-1} - \bar{y}^{k-2}\| + \frac{18\psi}{\phi} \Pi_{[\varphi],[k+1,k]}. \tag{23}
 \end{aligned}$$

Next, summing up (23) from  $k = p + 1$  to  $k = z$  and rearranging the terms, one gets

$$\begin{aligned}
 5 \sum_{k=p+3}^z \|x^k - \bar{x}^{k-1}\| + \sum_{k=p+3}^z \|y^{k+1} - \bar{y}^k\| &\leq 3\|y^{p+1} - \bar{y}^p\| + \|y^{p+2} - \bar{y}^{p+1}\| \\
 &\quad + \|x^{p+1} - \bar{x}^p\| - 3\|y^{z+1} - \bar{y}^z\| \\
 &\quad - \|y^{z+2} - \bar{y}^{z+1}\| - \|x^{z+1} - \bar{x}^z\| \\
 &\quad + \frac{18\psi}{\phi} \Pi_{[\varphi],[p+1,z+1]}.
 \end{aligned}$$

Furthermore, as  $0 \leq \varphi(L_\beta(\hat{\zeta}^k) - L_\beta(\hat{\zeta}^*))$  and  $m$  approaches positive infinity, we can conclude that

$$\sum_{k=p+1}^{+\infty} (5\|x^k - \bar{x}^{k-1}\| + \|y^{k+1} - \bar{y}^k\|) < +\infty,$$

which implies

$$\sum_{k=0}^{+\infty} \|x^k - \bar{x}^{k-1}\| < +\infty, \quad \sum_{k=0}^{+\infty} \|y^k - \bar{y}^{k-1}\| < +\infty.$$

Based on the relationship between (12) and (17), we can assert that

$$\sum_{k=0}^{+\infty} \|x^k - x^{k-1}\| < +\infty, \quad \sum_{k=0}^{+\infty} \|y^k - y^{k-1}\| < +\infty, \quad \sum_{k=0}^{+\infty} \|\lambda^k - \lambda^{k-1}\| < +\infty.$$

This demonstrates that  $\{\zeta^k\}$  forms a Cauchy sequence, which ensures its convergence. By applying Theorem 1, it follows that  $\{\zeta^k\}$  converges to a critical point of  $L_\beta$ .

□

#### 4. Numerical Simulations

In this section, we demonstrate the application of NIP-ADMM in signal recovery and SCAD penalty problems. To verify the effectiveness of the algorithm, we compare it with Bregman modification of ADMM (BADMM) proposed by Wang et al. [23] and inertial proximal ADMM (IPADMM) proposed by Pi et al. [32]. All codes were implemented in MATLAB 2024b and executed on a Windows 11 system equipped with an AMD Ryzen R9-9900X CPU.

#### 4.1. Signal Recovery

In this subsection on signal recovery, we consider the previously mentioned model (2).

$$\begin{aligned} \min c \|x\|_{\frac{1}{2}}^{\frac{1}{2}} + \frac{1}{2} \|y\|^2 \\ \text{s.t. } Ax - y = b, \end{aligned} \quad (24)$$

where  $\|x\|_{\frac{1}{2}}^{\frac{1}{2}} = \sum_{i=1}^n |x_i|^{\frac{1}{2}}$ , and  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $y \in \mathbb{R}^m$ ,  $x \in \mathbb{R}^n$ . To evaluate the effectiveness of NIP-ADMM, we compare it with BADMM [23] and IPADMM [32]. We construct the following framework to solve problem (24). We set  $S = e_0 I - \beta A^\top A$ , where  $I$  denotes the identity matrix.

$$\begin{cases} (\bar{x}^k, \bar{y}^k) = (x^k, y^k) + \theta(x^k - \bar{x}^{k-1}, 0) + \eta(0, y^k - \bar{y}^{k-1}), \\ x^{k+1} = H\left(\frac{1}{e_0} \left(-A^\top \lambda^k + \beta A^\top (\bar{y}^k + b) + (e_0 I - \beta A^\top A) \bar{x}^k\right), \frac{2c}{e_0}\right), \\ y^{k+1} = y^k - \gamma(y^k + \lambda^k - \beta(Ax^{k+1} - y^k - b)), \\ \lambda^{k+1} = \lambda^k + \beta(Ax^{k+1} - y^{k+1} - b). \end{cases}$$

Here,  $H(\cdot, \cdot)$  represents the half-shrinkage operator proposed by Xu et al. [39], which is defined as

$$H(x, e) = \{h_e(x_1), h_e(x_2), \dots, h_e(x_n)\}^T,$$

where the function  $h_e(x_i)$  for  $i = 1, 2, \dots, n$  is defined by

$$\begin{cases} h_e(x_i) = \frac{2x_i}{3} \left(1 + \cos \frac{2}{3} \left(\pi - \arccos \left(\frac{e}{8} \left(\frac{x_i}{3}\right)^{-\frac{3}{2}}\right)\right)\right), & |x_i| > \frac{\sqrt[3]{54}}{4} e^{\frac{2}{3}}, \\ 0, & \text{otherwise.} \end{cases}$$

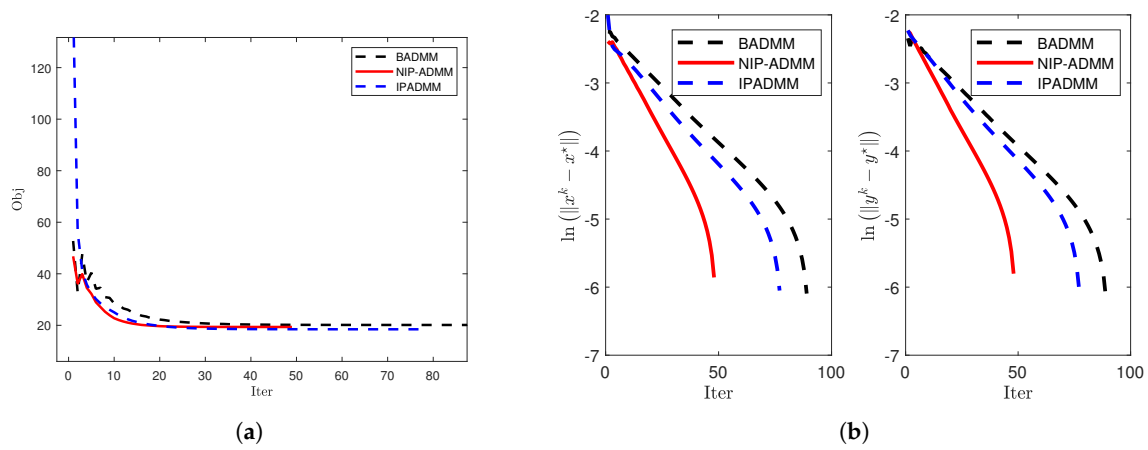
In this setup, the entries of matrix  $A$  are drawn from a standard normal distribution, with each column adjusted for normalization. The starting vector  $x_0$  is initialized as a sparse vector, containing at least 100 non-zero components. The values  $\bar{x}^0, \bar{y}^0, x^0, \lambda^0$  are set initially to zero. To simulate the observation vector with added noise, we generate  $b$  using  $b = Ax^0 + v$ , where  $v \sim N(0, 10^{-3}I)$ . For the regularization parameter  $c$ , we compute

$$c = 0.1 \|A^\top b\|_\infty.$$

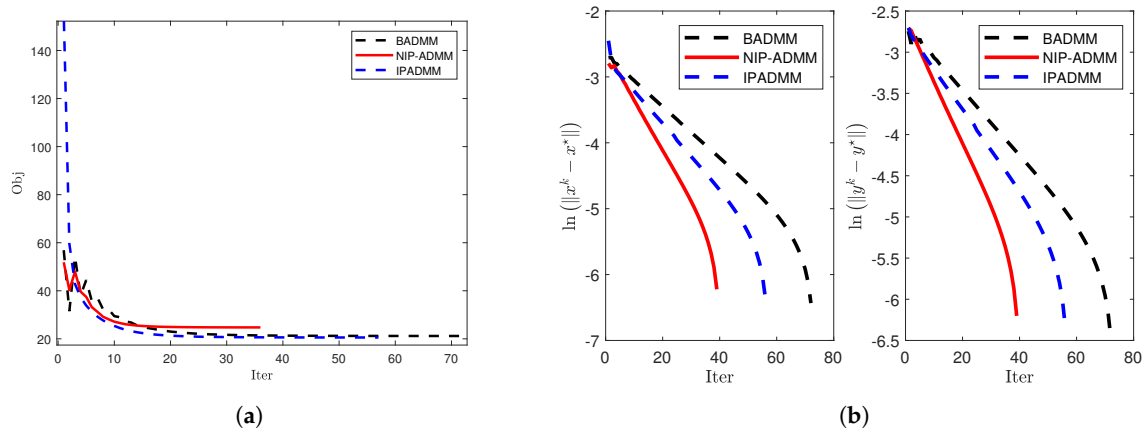
Based on Assumption 1, the parameters are chosen as  $\beta = 3$  and  $e_0 = 10$ . The error trend is depicted as  $\log_{10}(\|x^k - x^*\|)$  and  $\log_{10}(\|y^k - y^*\|)$ . At the  $(k+1)$ -th iteration, the primal residual is expressed as  $r^{k+1} = Ax^{k+1} + y^{k+1} - b$ , while the Dual residual is represented as  $s^{k+1} = \beta A^\top (x^{k+1} - \bar{x}^k)$ . Termination occurs when both conditions are met:  $\|r^{k+1}\|_2 \leq \sqrt{n} \times 10^{-4} + 10^{-3} \cdot \max(\|Ax^k\|_2, \|y^k\|_2)$  and  $\|s^{k+1}\|_2 \leq \sqrt{n} \times 10^{-4} + 10^{-3} \cdot \|A^\top \lambda^k\|_2$ . During the experiments, to satisfy Assumption 1, we set  $\gamma = 0.3$ . Table 1 shows that when  $m = n = 1000$ , selecting the inertial parameter values  $\theta = 0.8$  and  $\eta = 0.75$  produces satisfactory results. Therefore, in subsequent experiments, we also adopt  $\theta = 0.8$  and  $\eta = 0.75$ . The metrics include the number of iterations (Iter), CPU running time (CPUT) and the objective function value (Obj). To better present the experimental results, we retain two decimal places for Obj and four decimal places for CPUT.

The numerical results consistently demonstrate the superior performance of NIP-ADMM compared to BADMM and IPADMM (see Table 2): for  $m = n = 1000$ , NIP-ADMM shows faster convergence in terms of both objective value and error reduction (see Figure 1); for  $m = 3000$  and  $n = 4000$ , the inclusion of the inertial term further highlights its effectiveness (see Figure 2); and for larger-scale models with  $m = n = 6000$ , it is proven that NIP-ADMM is more suitable than BADMM and IPADMM for handling large-scale problems (see Figure 3).

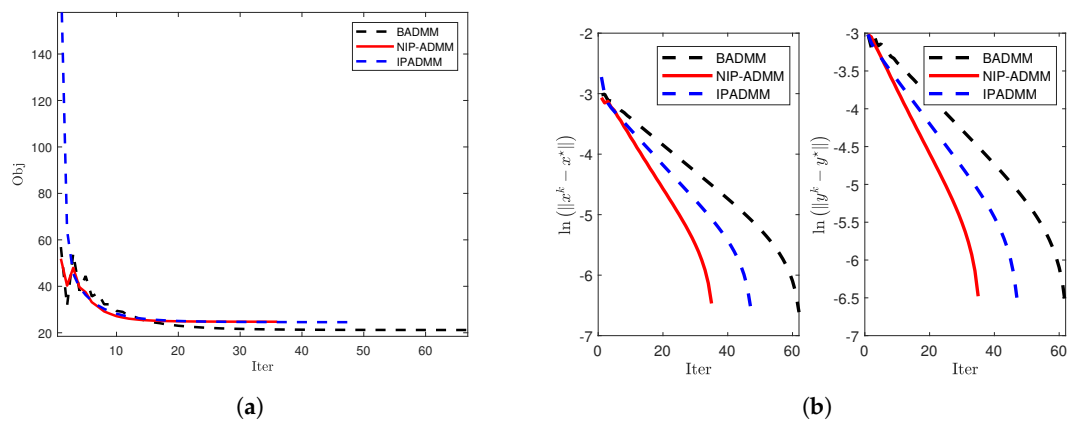




**Figure 1.** Comparison of convergence when  $m = 1000$  and  $n = 1000$ : (a) The objective value when  $m = 1000$  and  $n = 1000$ . (b)  $\log \|x^k - x^*\|$  and  $\log \|y^k - y^*\|$  under  $m = n = 1000$ .



**Figure 2.** Comparison of convergence when  $m = 3000$  and  $n = 3000$ : (a) The objective value when  $m = 3000$  and  $n = 3000$ . (b)  $\log \|x^k - x^*\|$  and  $\log \|y^k - y^*\|$  under  $m = n = 3000$ .



**Figure 3.** Comparison of convergence when  $m = 6000$  and  $n = 6000$ : (a) The objective value when  $m = 6000$  and  $n = 6000$ . (b)  $\log \|x^k - x^*\|$  and  $\log \|y^k - y^*\|$  under  $m = n = 6000$ .

**Table 1.** This is a table caption. Tables should be placed in the main text near to the first time they are cited.

$\theta$	$\eta$	Iter	CPUT(s)	$\theta$	$\eta$	Iter	CPUT(s)
0.2	0.2	75	2.2392	0.6	0.7	54	1.6014
0.3	0.2	78	2.3039	0.8	0.8	49	1.4583
0.3	0.3	69	1.9476	0.8	0.75	49	1.4309
0.5	0.5	60	1.7622	0.85	0.85	56	1.6445
0.6	0.6	56	1.6516	0.9	0.9	84	2.4614

**Table 2.** Comparison of iteration effect between NIP-ADMM, IPADMM and BADMM.

$m$	$n$	NIP-ADMM			IPADMM			BADMM		
		Iter	CPUT(s)	Obj	Iter	CPUT(s)	Obj	Iter	CPUT(s)	Obj
1000	1000	49	1.2698	19.36	78	2.1407	18.46	90	2.3407	20.14
1500	2000	44	4.9152	23.17	72	8.4978	22.12	76	8.5387	23.69
3000	3000	40	15.0464	21.02	57	22.3823	20.56	73	27.4115	21.18
3000	4000	55	34.0601	23.21	98	62.8206	23.11	76	48.3825	23.22
4000	5000	36	40.6110	24.02	53	61.7431	23.09	65	74.4521	24.03
4500	5500	40	61.7638	24.05	45	71.7627	23.79	67	102.6028	24.06
6000	6000	40	88.7702	24.99	48	108.3045	24.56	63	135.8133	25.00

#### 4.2. SCAD Penalty Problem

We note that the smoothly clipped absolute deviation (SCAD) penalty problem in statistics can be formulated as the following model [14,31]:

$$\begin{aligned} \min \quad & \sum_{i=1}^n h_{\kappa}(|x_i|) + \frac{1}{2} \|y\|^2 \\ \text{s.t.} \quad & Ax - y = b, \end{aligned} \quad (25)$$

with  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $y \in \mathbb{R}^m$ ,  $x \in \mathbb{R}^n$  and the penalty function  $h_{\kappa}(\cdot)$  in the objective is defined as:

$$h_{\kappa}(\theta) = \begin{cases} \kappa\theta, & \theta \leq \kappa, \\ \frac{-\theta^2 + 2c\kappa\theta - \kappa^2}{2(c+1)}, & \kappa < \theta \leq c\kappa, \\ \frac{(c+1)\kappa^2}{2}, & \theta > c\kappa, \end{cases} \quad (26)$$

where  $c > 2$  and  $\kappa > 0$ , being the knots of the quadratic spline function. In the reference signal recovery subsection, we similarly set  $\psi(x) = \frac{1}{2} \|x\|_S^2$  and  $\phi(y) = 0$ , where  $S = \mu I - \beta A^T A$ . For the problem (25), the  $x$ -subproblem can be expressed as

$$\begin{aligned} x^{k+1} &= \arg \min \left\{ \sum_{i=1}^n g_{\kappa}(|x_i|) - \langle p^k, Ax \rangle + \frac{\beta}{2} \|Ax - \bar{y}^k - b\|^2 + \frac{1}{2} \|x - \bar{x}^k\|_{\mu I - \beta A^T A}^2 \right\} \\ &= \arg \min \left\{ \sum_{i=1}^n g_{\kappa}(|x_i|) + \frac{\mu}{2} \left\| x - \frac{\beta A^T (\bar{y}^k + b + \frac{p^k}{\beta} - A \bar{x}^k)}{\mu} \right\|^2 \right\}. \end{aligned}$$

On the one hand, the  $x$ -subproblem can be equivalently formulated as:

$$\min_{x \in \mathbb{R}^n} \left\{ \sum_{i=1}^n g_{\kappa}(|x_i|) + \frac{1}{2\nu} \|x - q\|^2 \right\}.$$

On the other hand, under the condition that  $1 + \nu \leq c$ , we can update  $x$  using the following rule [31]:

$$x_i := \begin{cases} \text{sign}(q_i)(|q_i| - \kappa\nu)_+, & |q_i| \leq (1 + \nu)\kappa, \\ \frac{(c-1)q_i - \text{sign}(q_i)\kappa\nu}{c-1-\nu}, & (1 + \nu)\kappa < |q_i| \leq c\kappa, \\ q_i, & |q_i| > c\kappa, \end{cases}$$

where  $(\cdot)_+$  represents the positive part operator, which is defined as  $(x)_+ = \max(0, x)$ , applying NIP-ADMM to solve the problem (25), one yields that

$$\begin{cases} \bar{x}^k = x^k + \theta(x^k - \bar{x}^{k-1}), \bar{y}^k = y^k + \eta(y^k - \bar{y}^{k-1}), \\ x^{k+1} = \arg \min \left\{ \sum_{i=1}^n g_\kappa(|x_i|) + \frac{\mu}{2} \left\| x - \frac{-A^\top \lambda^k + \beta A^\top (\bar{y}^k + b) + \mu_0 \bar{x}^k - A^\top A \bar{x}^k}{\mu} \right\|^2 \right\}, \\ y^{k+1} = y^k - \gamma(y^k - \lambda^k - \beta(Ax^{k+1} - y^k - b)), \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - y^{k+1} - b). \end{cases}$$

Similarly, the update scheme of BADMM can be represented by the following procedure:

$$\begin{cases} x^{k+1} = \arg \min \left\{ \sum_{i=1}^n g_\kappa(|x_i|) + \frac{\mu}{2} \left\| x - \frac{\beta A^\top (y^k + b + \frac{\lambda^k}{\beta} - Ax^k) + \mu x^k}{\mu} \right\|^2 \right\}, \\ y^{k+1} = \frac{1}{1+\beta} \left( \beta(Ax^{k+1} - b) - \lambda^k \right), \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - y^{k+1} - b). \end{cases}$$

Utilize the IPADMM to address model (25) and derive the following iterative scheme:

$$\begin{cases} (\bar{x}^k, \bar{y}^k, \bar{\lambda}^k) = (x^k, y^k, \lambda^k) + \theta(x^k - x^{k-1}, y^k - y^{k-1}, \lambda^k - \lambda^{k-1}), \\ x^{k+1} = \arg \min \left\{ \sum_{i=1}^n g_\kappa(|x_i|) + \frac{\mu}{2} \left\| x - \frac{-A^\top \bar{\lambda}^k + \beta A^\top (\bar{y}^k + b) + \mu_0 \bar{x}^k - A^\top A \bar{x}^k}{\mu} \right\|^2 \right\}, \\ \lambda^{k+1} = \bar{\lambda}^k - \beta(Ax^{k+1} - \bar{y}^k - b), \\ y^{k+1} = y^k - \gamma(y^k - \lambda^{k+1} - \beta(Ax^{k+1} - y^k - b)). \end{cases}$$

In this experiment, we generate a random  $m \times n$  matrix  $A$ , and perform row and column normalization. Here, we choose to generate a vector  $z$  of dimension  $n$  with a sparsity ratio of  $\frac{100}{m}$ . The vector  $b$  is represented as the sum of  $Az$  and a Gaussian noise vector with zero mean and variance 0.001. The initial variables  $x_0$  and  $y_0$  are set as zero vectors, serving as the starting point for optimization. To improve numerical efficiency, in this experiment, we set  $c = 5$  and  $\kappa = 0.1$ . Under the condition that Assumption 1 is satisfied, we configure  $\gamma = 0.1$ ,  $\beta = 12$ , and  $\mu = 100$  for NIP-ADMM and other algorithms. The stopping criterion for the updates is set as

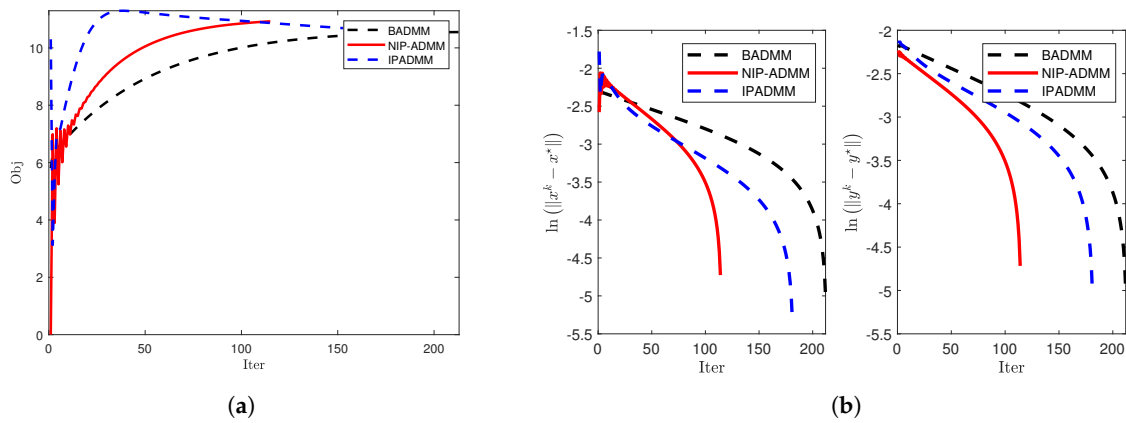
$$\max \left\{ \|x^{k+1} - x^k\|, \|y^{k+1} - y^k\| \right\} \leq 10^{-2}.$$

In Table 3, we set  $m = n = 1000$ . The results in the table support our choice of the inertial parameters  $\eta = 0.9$  and  $\theta = 0.9$ . Under these conditions, the NIP-ADMM requires the fewest iterations and the least running time. Therefore, we selected the inertial parameters  $\eta = 0.9$  and  $\theta = 0.9$  for the experiments.

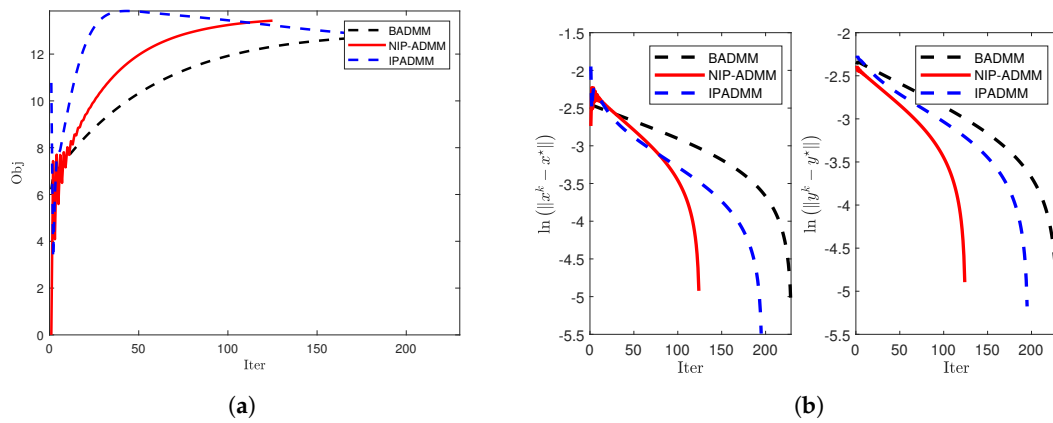
To further investigate the convergence behavior of the algorithms, we plot the update curves of the objective function (left figure) and the iteration error (right figure) against the number of iterations for each algorithm under three different dimensions (see Figures 4–6). The numerical results demonstrate that NIP-ADMM achieves nearly the same performance as IPADMM and BADMM but with significantly fewer iterations.

In Table 4, we present the test results of NIP-ADMM, BADMM and IPADMM under different dimensions. Although there are slight discrepancies in the Obj values between the two methods, by focusing on the metrics of Iter and CPUT, it is evident that NIP-ADMM demonstrates a significant

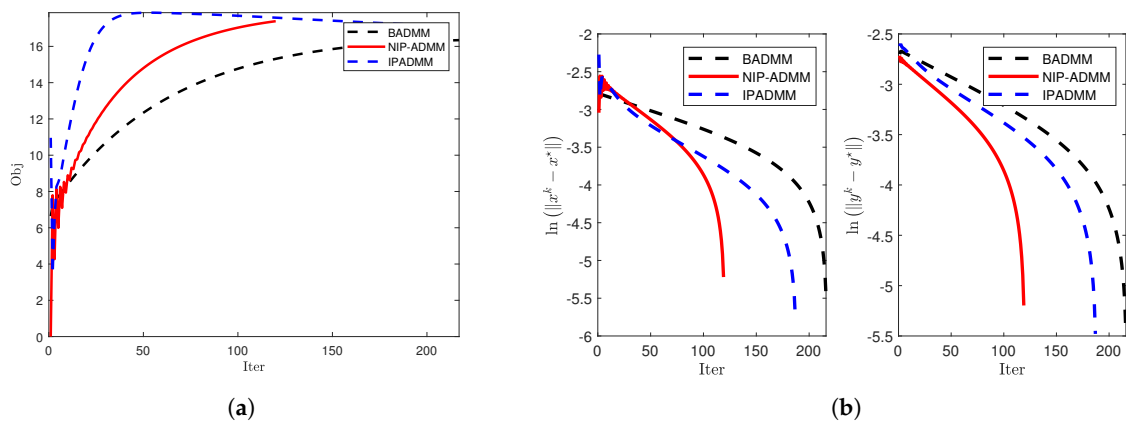
advantage over BADMM and IPADMM. This advantage becomes even more pronounced in higher-dimensional scenarios.



**Figure 4.** Comparison of convergence when  $m = 1000$  and  $n = 1000$ : (a) The objective value when  $m = 1000$  and  $n = 1000$ . (b)  $\log \|x^k - x^*\|$  and  $\log \|y^k - y^*\|$  under  $m = n = 1000$ .



**Figure 5.** Comparison of convergence when  $m = 1000$  and  $n = 1000$ : (a) The objective value when  $m = 1500$  and  $n = 1500$ . (b)  $\log \|x^k - x^*\|$  and  $\log \|y^k - y^*\|$  under  $m = n = 1500$ .



**Figure 6.** Comparison of convergence when  $m = 1000$  and  $n = 1000$ : (a) The objective value when  $m = 3000$  and  $n = 3000$ . (b)  $\log \|x^k - x^*\|$  and  $\log \|y^k - y^*\|$  under  $m = n = 3000$ .

Table 3. Numerical results of NIP-ADMM with different  $\theta, \eta$ .

$\theta$	$\eta$	Iter	CPUT(s)	$\theta$	$\eta$	Iter	CPUT(s)
0.2	0.2	196	2.0092	0.6	0.7	149	1.5017
0.3	0.2	187	1.9122	0.8	0.7	134	1.3693
0.3	0.3	181	1.8690	0.8	0.9	133	1.3503
0.4	0.5	170	1.7650	0.9	0.8	127	1.3467
0.5	0.5	159	1.6438	0.9	0.9	126	1.3100

Table 4. Comparison of iteration effect between NIP-ADMM, IPADMM and BADMM.

$m$	$n$	NIP-ADMM			IPADMM			BADMM		
		Iter	CPUT(s)	Obj	Iter	CPUT(s)	Obj	Iter	CPUT(s)	Obj
1000	1000	121	1.3366	10.91	213	2.3065	10.55	182	1.9632	10.55
1000	1300	115	1.9246	12.96	211	3.5724	12.48	174	2.9611	13.13
1500	1000	130	1.9270	8.92	228	3.2943	8.59	172	2.4709	7.49
1500	1300	140	3.0474	13.38	259	5.7832	12.90	215	4.7147	11.88
1500	1500	125	3.6104	13.43	230	6.6865	12.81	196	5.4584	12.71
1800	1500	146	4.6432	13.47	257	8.0396	12.94	209	6.1925	11.83
1800	2000	115	5.8341	15.00	210	10.7513	14.29	182	9.1033	14.69
2500	2000	142	8.9043	14.95	250	15.6397	14.29	201	12.2370	13.07
2900	2700	134	15.3647	17.70	245	28.4289	16.71	203	22.9945	16.50
3000	3000	125	17.1686	17.20	217	34.1575	16.34	188	25.0864	17.22
3500	3000	128	20.2808	16.87	234	37.1725	15.84	194	30.6876	15.94
3500	3500	123	24.5455	19.80	223	44.4163	18.69	200	39.2771	19.43

5. Conclusion

The Purpose of this is to propose a novel symmetrical inertial ADMM with proximal term for solving nonconvex two-block optimization problems. Under certain conditions, if the objective function satisfies Kurdyka-Łojasiewicz property, the sequence generated by the algorithm globally converges to a stationary point. In numerical experiments, we apply the algorithm to signal recovery and SCAD penalty problems, and its superiority is validated. Notably, by continuously adjusting the inertial parameters, we identify a set of parameters that enhances the convergence speed of the algorithm.

Furthermore, we believe that future work could explore whether the convergence of the algorithm can be guaranteed when the objective function is non-separable. Additionally, it would be worthwhile to investigate whether introducing inertial terms into the  $y$ -subproblem and the multiplier  $\lambda$  could further accelerate the convergence speed of the algorithm.

**Author Contributions:** Conceptualization, J.-h.L. and H.-y.L.; methodology, J.-h.L.; software, J.-h.L. and S.-y.L.; validation, J.-h.L. and H.-y.L.; writing—original draft preparation, J.-h.L. and H.-y.L.; writing—review and editing, J.-h.L. and H.-y.L.; visualization, J.-h.L, H.-y.L. and S.-y.L.; supervision, H.-y.L.; project administration, H.-y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Innovation Fund of Postgraduate, Sichuan University of Science & Engineering (Y2024340), the Scientific Research and Innovation Team Program of Sichuan University of Science and Engineering (SUSE652B002) and the Opening Project of Sichuan Province University Key Laboratory of Bridge Non-destruction Detecting and Engineering Computing (2023QZJ01).

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

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

## References

1. Zhang, S.R.; Wang, Q.H.; Zhang, B.X.; Liang, Z.; Zhang, L.; Li L.L.; Huang, G.; Zhang, Z.G.; Feng, B.; Yu, T.Y. Cauchy non-convex sparse feature selection method for the high-dimensional small-sample problem in motor imagery EEG decoding. *Front. Neurosci.* **2023**, *17*, 1292724. [CrossRef] [PubMed]
2. 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]
3. Tiddeman, B.; Ghahremani, M. Principal component wavelet networks for solving linear inverse problems. *Symmetry*. **2021**, *13*, 1083. [CrossRef]
4. Xia, Z.C.; Liu, Y.; Hu, C.; Jiang, H.J. Distributed nonconvex optimization subject to globally coupled constraints via collaborative neurodynamic optimization. *Neural Netw.* **2025**, *184*, 107027. [CrossRef] [PubMed]
5. Yu, G.; Fu, H.; Liu, Y.F. High-dimensional cost-constrained regression via nonconvex optimization. *Technometrics*. **2021**, *64*, 52-64. [CrossRef]
6. Merzbacher, C.; Mac Aodha, O.; Oyarzún, D.A. Bayesian optimization for design of multiscale biological circuits. *ACS Synth. Biol.* **2023**, *12*, 2073-2082. [CrossRef] [PubMed]
7. Kim, S.j.; Koh, K.; Lustig, M.; Boyd, S.; Gorinevsky, D. An interior-point method for large-scale  $\ell_1$ -regularized least Squares. *IEEE J. Sel. Top. Signal Process.* **2007**, *1*, 606-617. [CrossRef]
8. Chartrand, R.; Staneva, V. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*. **2008**, *24*, 035020. [CrossRef]
9. Zeng, J.S.; Lin, S.B.; Wang, Y.; Xu, Z.B.  $L_{1/2}$  regularization: convergence of iterative half thresholding algorithm. *IEEE Trans. Signal Process.* **2014**, *62*, 2317-2329. [CrossRef]
10. Chen, P.Y.; Selesnick, I.W. Group-sparse signal denoising: non-convex regularization, convex optimization. *IEEE Trans. Signal Process.* **2014**, *62*, 3464-3478. [CrossRef]
11. Bai, Z.L. Sparse Bayesian learning for sparse signal recovery using  $\ell_{1/2}$ -norm. *Appl. Acoust.* **2023**, *207*, 109340. [CrossRef]
12. Wang, C.; Yan, M.; Rahimi, Y.; Lou, Y.F. Accelerated schemes for the  $L_1 / L_2$  minimization. *IEEE Trans. Signal Process.* **2020**, *68*, 2660-2669. [CrossRef]
13. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183-202. [CrossRef]
14. Fan, J.Q.; Li, R.Z. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **2001**, *96*, 1348-1360. [CrossRef]
15. Bai, J.C.; Zhang, H.C.; Li, J.C. A parameterized proximal point algorithm for separable convex optimization. *Optim. Lett.* **2018**, *12*, 1589-1608. [CrossRef]
16. Wen, F.; Liu, P.L.; Liu, Y.P.; Qiu, R.C.; Yu, W.X. Robust sparse recovery in impulsive noise via  $\ell_p - \ell_1$  optimization. *IEEE Trans. Signal Process.* **2017**, *65*, 105-118. [CrossRef]
17. Zhang, H.M.; Gao, J.B.; Qian, J.J.; Yang, J.; Xu, C.Y.; Zhang, B. Linear regression problem relaxations solved by nonconvex ADMM with convergence analysis. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *34*, 828-838. [CrossRef]
18. Ames, B.P.W.; Hong, M.Y. Alternating direction method of multipliers for penalized zero-variance discriminant analysis. *Comput. Optim. Appl.* **2016**, *64*, 725-754. [CrossRef]
19. Zietlow, C.; Lindner, J.K.N. ADMM-TGV image restoration for scientific applications with unbiased parameter choice. *Numer. Algorithms*. **2024**, *97*, 1481-1512. [CrossRef]
20. Bian, F.M.; Liang, J.W.; Zhang, X.Q. A stochastic alternating direction method of multipliers for non-smooth and non-convex optimization. *Inverse Problems*. **2021**, *37*, 075009. [CrossRef]
21. Parikh, N.; Boyd, S. Proximal Algorithms; Now Publishers: Braintree, MA, USA, 2014.
22. Hong, M.Y.; Luo, Z.Q.; Razaviyayn, M. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM J. Optim.* **2016**, *26*, 337-364. [CrossRef]
23. Wang, F.H.; Xu, Z.B.; Xu, H.K. Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems. *arXiv* **2014**, arXiv:1410.8625.
24. Ding, W.; Shang, Y.; Jin, Z.; Fan, Y. Semi-proximal ADMM for primal and dual robust Low-Rank matrix restoration from corrupted observations. *Symmetry*. **2024**, *16*, 303. [CrossRef]
25. Guo, K.; Han, D.R.; Wu, T.T. Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints. *Int. J. Comput. Math.* **2016**, *94*, 1653-1669. [CrossRef]
26. Wang, Y.; Yin, W.T.; Zeng, J.S. Global convergence of ADMM in nonconvex nonsmooth optimization. *J. Sci. Comput.* **2019**, *78*, 29-63. [CrossRef]



27. Wang, F.H.; Cao, W.F.; Xu, Z.B. Convergence of multi-block Bregman ADMM for nonconvex composite problems. *Sci. China Inf. Sci.* **2018**, *61*, 122101. [[CrossRef](#)]
28. Barber, R.F.; Sidky, E.Y. Convergence for nonconvex ADMM, with applications to CT imaging. *J. Mach. Learn. Res.* **2024**, *25*, 1-46. [[CrossRef](#)] [[PubMed](#)]
29. Wang, X.F.; Yan, J.C.; Jin, B.; Li, W.H. Distributed and parallel ADMM for structured nonconvex optimization problem, *IEEE Trans. Cybern.* **2021**, *51*, 4540-4552. [[CrossRef](#)] [[PubMed](#)]
30. Alvarez, F.; Attouch, H. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Anal.* **2001**, *9*, 3-11. [[CrossRef](#)]
31. 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](#)]
32. Chen, C.H.; Chan, R.H.; Ma, S.Q.; Yang, J.F. Inertial proximal ADMM for linearly constrained separable convex optimization. *SIAM J. Imaging Sci.* **2015**, *8*, 2239-2267. [[CrossRef](#)]
33. Chao, M.T.; Zhang, Y.; Jian, J.B. An inertial proximal alternating direction method of multipliers for nonconvex optimization. *Int. J. Comput. Math.* **2020**, *98*, 1199-1217. [[CrossRef](#)]
34. Wang, X.Q.; Shao, H.; Liu, P.J.; Wu, T. An inertial proximal partially symmetric ADMM-based algorithm for linearly constrained multi-block nonconvex optimization problems with applications. *J. Comput. Appl. Math.* **2023**, *420*, 114821. [[CrossRef](#)]
35. Gonçalves, M.L.N.; Melo, J.G.; Monteiro, R.D.C. Convergence rate bounds for a proximal ADMM with over-relaxation stepsize parameter for solving nonconvex linearly constrained problems. *arXiv* **2017**, arXiv:1702.01850.
36. 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](#)]
37. Bolte, J.; Sabach, S.; Teboulle, M. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.* **2014**, *146*, 459-494. [[CrossRef](#)]
38. Nesterov, Y. *Introductory Lectures on Convex Optimization: A Basic Course*; Springer: NY, USA, 2004.
39. Xu, Z.B.; Chang, X.Y.; Xu, F.M.; Zhang, H.  $L_{1/2}$  regularization: A thresholding representation theory and a fast solver. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 1013-1027. [[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.