Abstract

The growing demand on efficient and distributed optimization algorithms for large-scale data stimulates the popularity of Alternative Direction Methods of Multipliers (ADMM) in numerous areas, such as compressive sensing, matrix completion, and sparse feature learning. While linear equality constrained problems have been extensively explored to be solved by ADMM, there lacks a generic framework for ADMM to solve problems with nonlinear equality constraints, which are common in practical application (e.g., orthogonality constraints). To address this problem, in this paper, we proposed a new generic ADMM framework for handling nonlinear equality constraints, called neADMM. First, we propose the generalized problem formulation and systematically provide the sufficient condition for the convergence of neADMM. Second, we prove a sublinear convergence rate based on variational inequality framework and also provide an novel accelerated strategy on the update of the penalty parameter. In addition, several practical applications under the generic framework of neADMM are provided. Experimental results on several applications demonstrate the usefulness of our neADMM.

1 Introduction

In recent years, there is a growing demand on efficient computational methods for analyzing high-dimensional large-scale data in a variety of applications such as bioinformatics, medical imaging, social networks, and astronomy [25]. The Alternating Direction Method of Multiplier (ADMM) has received significant amount of attention in the last few years, because ADMM turned out to be a natural fit in the field of large-scale data-distributed machine learning and big-data related optimization [4]. For example, Zhang et al. discussed the nonnegative matrix factorization problem [28]; Wahlberg et al. presented an ADMM algorithm to deal with the total variance estimation problem [24]; Yang et al. dealt with the L1-regularization problem [27]. The classic ADMM seeks to solve the following convex problem with respect to $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$:

$$\begin{align*}
\text{minimize} & \quad f(x) + g(z) \\
\text{subject to} & \quad Ax + Bz = c
\end{align*}$$

where $f(x)$ and $g(z)$ are convex functions, $A \in \mathbb{R}^{p \times n}$, $B \in \mathbb{R}^{p \times m}$ and $c \in \mathbb{R}^p$. $f(x)$ and $g(z)$ can be either smooth or non-smooth. For example, $f(x)$ can be any loss functions such as hinge loss, log loss and square loss while $g(z)$ represents regularization terms like L1 penalty, L2 penalty and kernel penalty. The augmented Lagrangian methods (ALM) [14] can be utilized to alternately optimize $x$ and $z$. As a result, the advantage of the ADMM lies in the splitting scheme of two subproblems which are relatively easy to solve. More importantly, in numerous practical applications, the functions $f(x)$ and $g(x)$ are separable such that $f(x) = \sum^N_{i=1} f_i(x_i)$ and $g(z) = \sum^M_{i=1} g_i(z)$. Thus the $x$ and $z$-optimization can split into $N$ and $M$ separate problems respectively, each of which can be solved in parallel.

During the recent few years, many variants of ADMM have been developed and analyzed [11], [12], [9], [10], [1], [23], [29], [26] and [4]. For example, Wang et al. replaced the quadratic penalty in augmented Lagrangian with the Bregman divergence to achieve good convergence rate [25]; Ouyang et al. proposed an accelerated version of ADMM [22] while Ouyang et al. discussed the stochastic ADMM [21]; Chen et al. extended the ADMM to the multi-block problem [7]. The convergence rate of the ADMM has also been discussed extensively. For example, [13] developed an elegant variational inequality framework and proved the sublinear...
convergence rate $O(1/n)$ for convex problems. Recently, some other works proves the linear convergence rate on multi-block ADMM problem [15] [14]. However, most of these variants including the classic one only focus on linear constraints, while in reality many practical problems requires nonlinear constraints, such as collaborative filtering [5], nonnegative matrix factorization [19], 1-bit compressive sensing [3], and mesh processing [20]. There are only few works which focus specifically on several particular nonlinear equality constraints [16], [6] and [18]. But it still lacks a generic framework and a comprehensive analysis of the nonlinear equality constrained ADMM. This is because in this situation, the convexity and the existence of the minimum point of the augmented Lagrangian of Problem [4] cannot be guaranteed any more. Therefore, the commonly-used assumptions made by existing ADMM framework cannot be satisfied.

To address this challenge, we propose an novel generic ADMM framework that deals with the general nonlinear equality-constrained problems:

$$\begin{align*}
\text{minimize} & \sum_{i=1}^{n+1} F_i(x_i) \\
& f_1(x_1) = f_2(x_2) \\
& f_3(x_2) = f_4(x_3) \\
& \vdots \\
& f_{2n-1}(x_n) = f_{2n}(x_{n+1})
\end{align*}$$

subject to

with variables $x_i \in R^{m_i}$ ($i = 1, 2, \cdots, n + 1$), where $F_i(x)(i = 1, 2, \cdots, n + 1)$ and $f_{2i-1}(x_i)(i = 1, 2, \cdots, n)$ $f_{2i}(x_{i+1})(i = 1, 2, \cdots, n) \in R^{d_i}$ are convex functions, $f_{2i}(x_{i+1})(i = 1, 2, \cdots, n) \in R^{d_i}$ are concave functions. Therefore, the proposed Problem (2) is a generalization of the Problem [1] solved by classic ADMM. To the best of our knowledge, there is no existing work on ADMM that focuses on solving the Problem (2) and provides comprehensive theoretical analysis and guarantees. In this paper, we propose a new generic Nonlinear Equality-constrained ADMM (neADMM) to address this problem. It enjoys the advantages from ADMM in solving large-scale problems, and is capable of solving nonconvex problems with nonlinear equality constrained, which is a new promising domain of ADMM research. The major contributions of this paper can be summarized as follows:

- **Proposed a novel generic framework of nonlinear-equality constrained ADMM.** A generic neADMM framework is proposed to handle nonlinear equality constraints. Several classic ADMM-based methods are shown to be the special cases of our framework.
- **Analyzed the convergence and convergence rate for the generic problem.** We prove the sufficient and necessary conditions of the convergence to the global optima. Additionally, a sublinear ($O(1/n)$) convergence rate has been proved based on the variational inequality technique.

- **Developed new strategy for penalty parameter updates.** To further enhance the power of neADMM framework, we design a novel penalty parameter updating strategy that further weaken the condition for convergence. Theoretical analysis and experimental results show its effectiveness.

- **Conducted extensive experiments on several important applications.** Several important and challenging nonconvex optimization problems with nonlinear equality-constraints have been shown to fit our framework and enjoy our guarantees. Experiments on different datasets demonstrate the theoretical properties and practical effectiveness of our neADMM.

The rest of this paper is organized as follows. Section 2 presents the algorithms and theoretical analyses. Section 3 provides several representative nonconvex problems solvable to our neADMM. Section 4 verifies the theoretical properties and demonstrates the effectiveness of neADMM. Finally, the paper is concluded in Section 5.

## 2 Nonlinear Equality-constrained ADMM

In this section, we first present the problem formulation for the generic framework of neADMM. Then, a comprehensive convergence analysis is provided. Furthermore, a novel strategy for penalty parameter update is proposed and discussed. Finally, the novel neADMM is proved to enjoy a sublinear convergence rate.

### 2.1 Algorithms

In the proposed Problem (2), $x_1$ and $x_{n+1}$ are called primal variables while the others are called intermediate variables because they build the connection between $x_1$ and $x_{n+1}$. Every equality constraint is split into two parts. This is a nonconvex problem because equality constraints are nonlinear. We form the augmented Lagrangian as follows:

$$L_\rho(x_1, x_2, \cdots, x_n, y_1, y_2, \cdots, y_n) =$$

$$\sum_{i=1}^{n} (F_i(x_i) + y_i^T (f_{2i-1}(x_i) - f_{2i}(x_{i+1})) +$$

$$\rho/2 \|f_{2i-1}(x_i) - f_{2i}(x_{i+1})\|_2^2 + F_{n+1}(x_{n+1})$$

where $\rho$ is the penalty parameter and $y_i^T$ a the dual variable. The neADMM algorithm consists of the following
steps:
\[
x_{1}^{k+1} = \arg \min_{x_1} L_{\rho}(x_1, x_2, \ldots, x_n+1, \ldots, y_{k+1})
\]
\[
x_{n+1}^{k+1} = \arg \min_{x_{n+1}} L_{\rho}(x_1^{k+1}, \ldots, x_n, x_{n+1}, \ldots, y_{k+1})
\]
\[
x_{2}^{k+1} = \arg \min_{x_2} L_{\rho}(x_1^{k+1}, x_2, \ldots, y_{k+1})
\]
\[
\vdots
\]
\[
x_{n}^{k+1} = \arg \min_{x_n} L_{\rho}(x_1^{k+1}, \ldots, x_n, \bar{y}_{n+1})
\]
\[
y_{1}^{k+1} = y_1 + \rho(f_1(x_1^{k+1}) - f_2(x_2^{k+1}))
\]
\[
\vdots
\]
\[
y_{n}^{k+1} = y_n + \rho(f_{n-1}(x_n^{k+1}) - f_{n}(x_{n+1}^{k+1})).
\]
The primal residuals \( r_i^{k+1} = f_{2i-1}(x_i^{k+1}) - f_{2i}(x_i^{k+1}) \) and the dual residuals \( s_i^{k+1} = \rho(\partial f_{2i-1}(x_i^{k+1}))^T(f_{2i}(x_i^{k+1}) - f_{2i}(x_i^{k+1}))(i \neq n) \), \( s_n^{k+1} = \rho(\partial f_{2n-1}(x_n^{k+1}))^T(f_{2n-1}(x_n^{k+1}) - f_{2n}(x_n^{k+1}))(\bar{x}_i, x_i) \) where \( \partial f(\bullet) \) is a subgradient operator.

It can be seen that the augmented Lagrangian is nonconvex, so each subproblem is not necessarily solvable. The following lemma gives sufficient conditions to confirm each subproblem is well defined.

**Lemma 1 (Well defined)** All subproblems in the ADMM algorithm are well defined if one of the following requirements is met:

1. An unique minimum point and maximum exist for \( f_1(x_1) \) and \( f_{2n}(x_{n+1}) \), respectively, either a unique maximum point exists for \( f_{2i-1}(x_i) \) or an unique minimum point exists for \( f_{2i-2}(x_i) \) where \( i = 2, \ldots, n \).
2. There exists a scalar \( L_i > 0 \) such that \( \forall \bar{x}_i, z \in \partial F_1(x_i) \),
\[
F(x_i) \geq F(\bar{x}_i) + z^T(x_i - \bar{x}_i) + 1/2L_i\|x_i - \bar{x}_i\|^2_2
\]
3. \( f_{2i-2}(x_i) \) and \( f_{2i-2}(x_i) \) are linear.

**Proof:** We first show that the \( x_1 \)-subproblem is well defined and other subproblems are similar. The \( x_1 \)-subproblem is shown as follows:
\[
x^{k+1} = \arg \min_x F_1(x_1) + \rho/2\|f_1(x_1) - f_2(x_2) + y_1^k/\rho\|^2_2
\]
(1) There exists a minimum point \( a \) and two points \( b \) and \( c \) such that \( b < a < c \), \( f(a) < f(b) \) and \( f(a) < f(c) \). So we have
\[
F(x_1) + \rho/2\|f_1(x_1) - f_2(x_2) + y_1^k/\rho\|^2_2 \geq F(a) + \rho/2\|d_1^T(x_1 - a) + f_1(b) - f_2(x_2) + y_1^k/\rho\|^2_2
\]
where \( d_1 \) and \( d_2 \) are in \( \partial f_1(b) \) and \( \partial f_1(c) \), respectively, and \( g \in \partial F_1(a) \).

As \( x_1 \to \infty \), the lower bound goes to \( \infty \), so the \( x_1 \)-subproblem is well defined.

2. This condition can be simply proved by
\[
F(x_1) + \rho/2\|f_1(x_1) - f_2(x_2) + y_1^k/\rho\|^2_2 \geq F(x_1) + z^T(x_i - x_1) + 1/2L_i\|x_i - \bar{x}_i\|^2_2
\]
The lower bound goes to \( \infty \), as \( x_1 \to \infty \).
(3) In this case, the problem is reduced into the original formulation of ADMM. □

### 2.2 Convergence Analysis

In this subsection, we present and prove the sufficient condition of the convergence of the proposed algorithm neADMM. Due to the space limitation, some theorems only the proof sketch are provided, of which the full versions are in the supplementary materials. We start with the following assumption:

**Assumption 1 (Saddle point assumption)** The optimal solution point \( (x_1^*, x_2^*, \ldots, x_{n+1}^*) \) is a saddle point for the unaugmented Lagrangian \( L_0 \).

The above assumption implies that the solution point is the minimum point for the unaugmented Lagrangian \( L_0 \) and therefore strong duality holds. Now we analyze the convergence by the following theorem.

**Theorem 1 (Convergence properties)** The neADMM satisfies the following convergence properties:

1. Residual convergence. The primal residuals \( r_i \to 0 \) and the dual residuals \( s_i \to 0 \) as \( k \to \infty \).
2. Objective convergence. Suppose \( p^* = \sum_{i=1}^{n+1} F_i(x_i^*) \) and \( p^k = \sum_{i=1}^{n+1} F_i(x_i^k) \), then \( p^k \to p^* \) as \( k \to \infty \).
3. Dual variable convergence. \( y_i^k \to y_i^* \) as \( k \to \infty \) where \( y_i^* \) is a dual optimal point and \( i = 1, \ldots, n \).

**Sketch of Proof:** According to the saddle point assumption, we could obtain
\[
p^* - p^{k+1} \leq \sum_{i=1}^{n} (y_i^*)^T x_i^{k+1}. \tag{3}
\]
The optimality conditions of all subproblems result in
\[
p^{k+1} - p^k \leq \sum_{i=1}^{n} \rho(f_{2i}(x_i^{k+1}) - f_{2i}(x_i^{k+1}))^T \frac{1}{\rho} \sum_{i=1}^{n} \rho(f_{2i-1}(x_i^{k+1}) - f_{2i-1}(x_i^{k+1})) + \rho(f_{2n-1}(x_n^{k+1}) - f_{2n-1}(x_n^{k+1}))^T \frac{1}{\rho} \sum_{i=1}^{n} \rho(f_{2n-1}(x_n^{k+1}) - f_{2n-1}(x_n^{k+1})). \tag{4}
\]
By adding Equation (3) and (4), regrouping terms, Let Lyapunov function
\[
V_k = \frac{1}{\rho} \sum_{i=1}^{n} \rho(f_{2i}(x_i^{k+1}) - f_{2i}(x_i^{k+1}))^T \frac{1}{\rho} \sum_{i=1}^{n} \rho(f_{2n-1}(x_n^{k+1}) - f_{2n-1}(x_n^{k+1})). \tag{5}
\]
\( V_k \) strictly decreases and \( V_k \geq 0 \), so \( \lim_{k \to \infty} V_k = 0 \). Therefore \( \lim_{k \to \infty} r_i = 0 \) and \( \lim_{k \to \infty} s_i = 0 \) (i = 1, 2, \ldots, n). Recall back Equation (3) and (4), the right side of inequalities approach zero. As a result, \( p^k \to p^* \). Overall, the neADMM converges. □
Corollary 1 If the proposed neADMM converges to the optimal solution point, the Assumption 7 holds.

Proof: This follows directly from the optimality conditions of each subproblem. □

2.3 A novel strategy for updating \( \rho \)

We propose a novel strategy to update the penalty parameter \( \rho \). The motivation behind this strategy is that as \( \rho \) increases, the neADMM focuses more on the nonlinear equality constraints and ensure the feasibility better. In addition, with this strategy, our convergence properties can be ensured in mild conditions.

We find that Theorem 1 still holds if there exists an increasing and unbounded sequence \( \{\eta_k^i\} \) such that it yields a bounded sequence \( \{y_k^i\} \) such that

\[
\eta_k^i = \sup \left( \frac{\eta_k^i}{\rho_k} \right) \quad \forall i.
\]

We substitute \( \rho_k \) for constant \( \rho \) in the previous proof, \( V_k \to 0 \) because \( \frac{1}{\rho_k} = o(\|f_2(x_{k+1}^i) - f_2(x_{k+1}^i)\|_2) \) and \( \frac{1}{\rho_k} = o(\|f_2(x_{k+1}^i) - f_2(x_{k+1}^i)\|_2) \) and hence \( r_k \to 0 \). By rewriter the right side of Equation (4), \( p_k \to p^* \) is proved. Notice that it should be careful to choose \( \rho^k \) because if \( \frac{1}{\rho^k} \) converges faster than a threshold, the neADMM would be divergent and ill-conditioned. As mentioned before, Theorem 3 implies an important convergence property: if objective and constraints are continuous functions, the neADMM converges to a point \( (x_{k}^1, x_{k}^2, \cdots, x_{k}^{n+1}) \), then it must be the global minimum.

Theorem 3 Assume that \( F(x_i) = (1, 2, \cdots, n+1) \), \( f_{2i-1}(x_i) \) is a continuous function, the solution set is nonempty, \( x_{k+1}^i \) are the global minimums of \( x_i \), and \( \|x_i - x_{k+1}^i\|_2 \) is a increasing sequence, we have \( \rho_k \) is a positive and increasing sequence and \( \rho_k \to \infty \).

For convenience, we consider \( n = 1 \). We suppose the limit point is \( (x_1, x_2) \). By definition,

\[
L_{\rho_k+1}(x_{k+1}^1, x_{k+1}^2, y_k^1) \leq L_{\rho_k+1}(x_1^1, x_2^1, y_1^1)
\]

We add these two inequalities together and obtain

\[
F_1(x_{k+1}^1) + \eta_k^1 T f_1(x_{k+1}^1) + \rho_k^{k+1} / 2 \| f_1(x_{k+1}^1) - f_2(x_k^2) \|_2^2 + F_2(x_{k+1}^2) - \eta_k^1 T f_2(x_{k+1}^2) + \rho_k^{k+1} / 2 \| f_1(x_{k+1}^1) - f_2(x_k^2) \|_2^2 \leq p^* + \rho_k^{k+1} / 4 \| f_1(x_{k+1}^1) - f_2(x_k^2) \|_2^2
\]

On the other hand,

\[
p^* = \inf_{f_1(x_1) = f_2(x_2)} F_1(x_1) + F_2(x_2) = \inf_{f_1(x_1) = f_2(x_2)} F_1(x_1) + F_2(x_2) + (y_k^1) T (f_1(x_1) - f_2(x_2))
\]

We take the infimum on the right side of Equation (6) over \( f_1(x_1) = f_2(x_2) \) and obtain

\[
F_1(x_{k+1}^1) + \eta_k^1 T f_1(x_{k+1}^1) + \rho_k^{k+1} / 2 \| f_1(x_{k+1}^1) - f_2(x_k^2) \|_2^2 + F_2(x_{k+1}^2) - \eta_k^1 T f_2(x_{k+1}^2) + \rho_k^{k+1} / 2 \| f_1(x_{k+1}^1) - f_2(x_k^2) \|_2^2 \leq p^* + \rho_k^{k+1} / 4 \| f_1(x_{k+1}^1) - f_2(x_k^2) \|_2^2
\]

Since \( \{y_k^i\} \) is a bounded sequence and hence have a limit point \( \eta_1 \). We take the upper limit in the inequality above using the continuity of \( F_1(x_1), F_2(x_2), f_1(x_1) \) and \( f_2(x_2) \) and get

\[
F_1(\eta_1) + (\eta_1) T f_1(\eta_1) + \lim sup \rho_k^{k+1} / 4 \| f_1(\eta_1) - f_2(\eta_2) \|_2^2 + F_2(\eta_2) - (\eta_1) T f_2(\eta_2) + \lim sup \rho_k^{k+1} / 2 \| f_1(\eta_1) - f_2(\eta_2) \|_2^2 \leq p^*
\]

Because \( k \) is an increasing sequence, we have \( f_1(\eta_1) = f_2(\eta_2) \) and \( F_1(\eta_1) + F_2(\eta_2) \leq p^* \). As a result, \( (\eta_1, \eta_2) \) is a global minimum point. □

2.4 Analysis on Convergence Rate

In this section, the sublinear convergence rate of the neADMM will be proved using the variational inequality framework. To achieve this, our Problem 2 is reformulated equivalently to the formulation under variational inequality framework.

Definition 1 (Variational inequality reformulation) Find

\[
w^* = (f_1(x_1^2), f_2(x_2^1), \cdots, f_{2n}(x_{n+1}^2), y_1^2, y_2^2, \cdots, y_n^2)^T \in \Omega = \mathbb{R}^{d_1} \times \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n} \times \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}
\]

such that

\[
VI(\Omega, J, H) : H(w) - H(w^*) + (w - w^*) T J(w^*) \geq 0, \quad \forall w \in \Omega
\]

where

\[
J(w) = \sum_{i=1}^{n+1} F_i(x_i) = (y_i - y_i, \cdots, y_n - y_n, f_2(x_2) - f_1(x_1), \cdots, f_{2n}(x_{n+1}) - f_{2n-1}(x_1))^T.
\]

So the solution set of the above problem \( \Omega^* \) is nonempty under the saddle point assumption. It is easy to find that

\[
(w_1 - w_2) T (J(w_1) - J(w_2)) = 0, \quad \forall w_1, w_2 \in \Omega
\]

which is used frequently. Now we describe the property of \( \Omega^* \). The below theorem extends from Theorem 2.1 in [13] and Theorem 2.3.5 in [8].
Theorem 4 The solution set of $VI(\Omega, J, H)$ is convex and could be characterized as

$$
\Omega^* = \cap_{w \in \Omega} \{ \overline{w} \in \Omega : H(u) - H(\overline{w}) + (w - \overline{w})^T J(w) \geq 0 \}
$$

(9)

Sketch of Proof: Firstly, we prove that $\Omega^* \subseteq \cap_{w \in \Omega} \{ \overline{w} \in \Omega : H(u) - H(\overline{w}) + (w - \overline{w})^T J(w) \geq 0 \}$ and $\cap_{w \in \Omega} \{ \overline{w} \in \Omega : H(u) - H(\overline{w}) + (w - \overline{w})^T J(w) \geq 0 \} \subseteq \Omega^*$. Therefore Equation (9) holds.

Then we need to prove the convexity of $\Omega^*$. For fixed $w \in \Omega$, the set $\{ \overline{w} \in \Omega : H(u) - H(\overline{w}) + (w - \overline{w})^T J(w) \geq 0 \}$ is convex, so their intersection is convex. □

The above theorem implies that we can find an $\varepsilon$-optimal solution $\overline{w}$ if it satisfies $H(u) - H(\overline{w}) + (w - \overline{w})^T J(w) \leq \varepsilon \forall w \in \Omega$ after $t$ iterations. Hence we focus on finding such a new sequence, thus establishing the worst-case sublinear convergence rate $O(1/t)$. Such a new sequence is introduced here.

$$
\hat{w}_i^k = (f_1(\tilde{y}_i^k), \ldots, f_n(\tilde{y}_n^k)), \tilde{y}_1^k, \tilde{y}_2^k, \ldots, \tilde{y}_n^k)
$$

To make the following proofs more concise, we define some important matrices, let $d = \sum_{i=1}^n d_i$

$$
C = \left[ \begin{array}{cc}
\rho A_{2d} & O_{2d \times d} \\
-B_{d \times 2d} & \frac{1}{\rho} I_d
\end{array} \right], \quad D = \left[ \begin{array}{cc}
\rho A_{2d} & O_{2d \times d} \\
-B_{d \times 2d} & \frac{1}{\rho} I_d
\end{array} \right],
$$

$$
E = \left[ \begin{array}{cc}
-I_{2d} & O_{2d \times d} \\
-\rho B_{d \times 2d} & I_d
\end{array} \right], \quad G = C^T + C - E^T DE
$$

(10)

where

$$
A = \text{diag}(O_{d_1}, I_{d_2}, \ldots, I_{d_n}, O_{d_m}),
$$

$$
B = \left[ \begin{array}{cccc}
O_{d_1} & O_{d_1 \times d_2} & \cdots & O_{d_1 \times d_n} \\
O_{d_2} & O_{d_2 \times d_2} & \cdots & O_{d_2 \times d_n} \\
\vdots & \vdots & \ddots & \vdots \\
O_{d_m} & O_{d_m \times d_2} & \cdots & O_{d_m \times d_n}
\end{array} \right]
$$

(11)

We find that $C = DE$ and

$$
w^{k+1} = w^k - E(w^k - \hat{w}^k).
$$

(13)

Besides, the notation of $G$ implies that it is symmetric and semi-definite because

$$
G = C^T + C - E^T DE =
\left[ \begin{array}{cc}
2\rho A_{2d} & -B_{d \times 2d} \\
-B_{d \times 2d}^T & \frac{1}{\rho} I_d
\end{array} \right] + \left[ \begin{array}{cc}
\rho A_{2d} + \rho B_{d \times 2d}^T & -B_{d \times 2d} \\
-B_{d \times 2d}^T & \frac{1}{\rho} I_d
\end{array} \right]
$$

(12)

using the fact that $B_{d \times 2d}^T B_{d \times 2d} = O_{2d}$. In the next two subsections, we study the worst-case convergence rate in two senses. In an ergodic sense, we discuss the convergence rate of the average of $\hat{w}^k$ over iterations. In a nonergodic sense, we explore the convergence rate of $\hat{w}^k$ directly.

2.4.1 In an Ergodic Sense

In this section, we prove the convergence rate in an ergodic sense. The below theorem describes the relationship between $\hat{w}^k$ and $\overline{w}$.

Theorem 5 Let $\tilde{w}^k$ be defined in Equation (10). Then

$$
H(u) - H(\tilde{w}^k) + (w - \tilde{w}^k)^T J(u) \geq 0, \forall w \in \Omega
$$

(14)

Sketch of Proof: By combining optimality conditions of all subproblems and the notation of $\tilde{w}^k$, we get

$$
H(u) - H(\tilde{w}^k) + \sum_{i=1}^{n-1} (f_{2i-1}(x_i) - f_{2i-1}(\hat{x}_i^k))^T y_i^k
$$

$$
= - (f_{2n}(x_{n+1}) - f_{2n}(\hat{x}_{n+1}^k))^T y_n^k
$$

$$
- \sum_{i=1}^{n-1} (f_{2i}(x_i+1) - f_{2i}(\hat{x}_{i+1}^k))^T y_{i+1}^k
$$

$$
+ (f_{2n-1}(x_n) - f_{2n-1}(\hat{x}_n^k))^T y_{n+1}^k \geq 0
$$

(15)

By substituting $y_i^k$ for $y_i^{k+1}$ and rewriting groups, we get

$$
H(u) - H(\tilde{w}^k) + (w - \tilde{w}^k)^T J(u)
$$

$$
= - \rho \sum_{i=1}^{n-1} (f_{2i}(x_i+1) - f_{2i}(\hat{x}_{i+1}^k))^T (f_{2i}(x_i+1) - f_{2i}(\hat{x}_{i+1}^k))
$$

$$
+ \rho (f_{2n-1}(x_n) - f_{2n-1}(\hat{x}_n^k))^T (f_{2n-1}(x_n) - f_{2n-1}(\hat{x}_n^k))
$$

$$
+ 1/\rho \sum_{i=1}^{n-1} (y_i - \hat{w})^T \hat{w}
$$

$$
- \sum_{i=1}^{n-1} (y_i - \hat{w})^T (f_{2i}(x_i+1) - f_{2i}(\hat{x}_{i+1}^k))
$$

$$
+ (y_n - \hat{w})^T (f_{2n-1}(x_n) - f_{2n-1}(\hat{x}_n^k)) \geq 0, \forall w \in \Omega
$$

(16)

Recall the definition of $C$, the desired inequality is proved. □

From the above theorem, we find that if $C(w^k - \tilde{w}^k) = 0$, then $\hat{w}$ is the optimal solution to the Formulation (6). Applying Equation (8) and $C = DE$, the above theorem could be rewritten as

$$
H(u) - H(\tilde{w}^k) + (w - \tilde{w}^k)^T J(u) \geq 0, \forall w \in \Omega
$$

(17)

To deal with the right side of inequality, we aim at changing it in term of the norm form $||w - \tilde{w}^k||_D = (((w - \tilde{w}^k)^T DE(w - \tilde{w}^k))/2).$ The following theorem provides the desired result we want.

Theorem 6 Let $\tilde{w}^k$ be defined in Equation (10) and matrices $D$ and $E$ be defined in Equation (11), then we have

$$
(w - \tilde{w}^k)^T DE(w - \tilde{w}^k) = 1/2(\|w - w^{k+1}\|^2_D - \|w - w^k\|^2_D - \|w^k - \tilde{w}^k\|^2_D)
$$

(18)

Proof: Firstly, $(w - \tilde{w}^k)^T DE(w - \tilde{w}^k) = (w - \tilde{w}^k)^T D(w - w^{k+1})$ by Equation (13), by applying identity

$$
(a - b)^T D(c - d) = 1/2(||a - d||^2_D - ||a - c||^2_D)
$$

$$
+ 1/2(||c - b||^2_D - ||d - b||^2_D),
$$

we have

$$
(w - \tilde{w}^k)^T D(w - w^{k+1}) = 1/2(\|w - w^{k+1}\|^2_D - \|w - w^k\|^2_D + \|w^k - \tilde{w}^k\|^2_D)
$$

$$
+ 1/2(\|w^k - \tilde{w}^k\|^2_D - \|w^{k+1} - \tilde{w}^k\|^2_D).
$$

(19)
We notice that
\[
\|w^k - \tilde{w}^k\|_D^2 - \|w^{k+1} - w^k\|_D^2
= \|w^k - \tilde{w}^k\|_D^2 - \|w^k - \tilde{w}^k - (w^k - \bar{w}^{k+1})\|_D^2
= \|w^k - \tilde{w}^k\|_D^2 - \|w^k - \bar{w}^{k+1} - E(w^k - \bar{w}^{k+1})\|_D^2
= (w^k - \tilde{w}^k)^T(2DE - E^TDE)(w^k - \tilde{w}^k)
= (w^k - \tilde{w}^k)^T((2C - E^TDE)(w^k - \tilde{w}^k)
= (w^k - \tilde{w}^k)^T(C^T + C - E^TDE)(w^k - \tilde{w}^k)
= (w^k - \tilde{w}^k)^T G(w^k - \tilde{w}^k)
= \|w^k - \tilde{w}^k\|_G^2
\]
using the notation of $G$. The equality holds. \hfill \Box

The above theorem introduces the following corollary, which is a fundamental result for proving the worse-case sublinear rate in an ergodic sense.

**Corollary 2** Let $\tilde{w}^k$ be defined in Equation (10) and matrices $D$ and $E$ be defined in Equation (11), then we have
\[
(w - \tilde{w}^k)^T DE(w^k - \tilde{w}^k) + 1/2\|w - w^k\|_D^2 \geq 0, \forall w, w \in \Omega
\]
(17)

**Proof:** It follows directly from Equation (10) by $\|w^k - \tilde{w}^k\|_G^2 \geq 0$. \hfill \Box

Now, the sublinear convergence rate could be proved based on the former two theorems.

**Theorem 7** Let $C$ be defined in Equation (11). For any integer number $t > 0$, we define $\bar{w}_t$, the average of $w^k$ which is given in Equation (10) as
\[
\bar{w}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{w}^k.
\]
(18)

Then $\bar{w}_t \in \Omega$ and
\[
H(\bar{w}_t) - H(u) + (\bar{w}_t - w)^T J(w) \leq \frac{1}{2(t+1)} \|w - w^0\|_D^2,
\forall w, w \in \Omega
\]
(19)

**Proof:** It is easy to prove that $\bar{w}_t \in \Omega$ directly from Equation (13). Equation (15) and (17) imply that
\[
H(u) - H(\bar{u}^k) + (w - \bar{w}^k)^T J(w)
+ 1/2\|w - w^k\|_D^2 \geq 0, \forall w, w \in \Omega
\]
Summing it over $k = 0, 1, \cdots, t$, we have
\[
(1+t)H(u) - \sum_{i=0}^{t} H(\tilde{u}^i) + ((1+t)w - \sum_{i=0}^{t} \tilde{w}^i)^T J(w)
+ 1/2\|w - w^0\|_D^2 \geq 0, \forall w, w \in \Omega
\]
Hence
\[
\frac{1}{t+1} \sum_{i=0}^{t} H(\tilde{u}^i) - H(u) + \frac{1}{t+1} \sum_{i=0}^{t} (\tilde{w}^i - w)^T J(w)
\leq \frac{1}{2(t+1)} \|w - w^0\|_D^2, \forall w, w \in \Omega
\]
Convexity of $H(\tilde{u}^i)$ leads to $H(\bar{u}^i) \leq \frac{1}{t+1} \sum_{i=0}^{t} H(\tilde{u}^i)$, and substitute it and the definition of $\tilde{w}^k$ into the above inequality, the theorem follows directly. \hfill \Box

From the above theorem, we find that the sequence $\tilde{w}^k$ is an approximation solution of the Formulation (7) after $t$ iterations with convergence rate $O(\frac{1}{t})$ in an ergodic sense.

### 2.4.2 In a Nonergodic Sense

In this section, we explore the worst-case convergence rate of the $w^k$ defined in Equation (10). Firstly, we give a fundamental result that is similar to Equation (17).

**Theorem 8** Let $\bar{w}$ be defined in Equation (10) and matrices $C$, $D$ and $E$ be defined in Equation (11), then we have
\[
H(\bar{w}^{k+1}) - H(\bar{w}^k) + (\bar{w}^{k+1} - \bar{w}^k)^T (J(\bar{w}^k) - C(\bar{w}^k - \bar{w}^0)) \geq 0.
\]
(20)

**Proof:** Firstly, we set $w = \bar{w}^{k+1}$ in Equation (14) and we obtain
\[
H(\bar{w}^{k+1}) - H(\bar{w}^k) + (\bar{w}^{k+1} - \bar{w}^k)^T (J(\bar{w}^k) - C(\bar{w}^k - \bar{w}^0)) \geq 0.
\]
(21)

Next, we let $k := k + 1$ in Equation (14) and we have
\[
H(u) - H(\bar{w}^{k+1}) + (w - \bar{w}^{k+1})^T (J(\bar{w}^{k+1}) - C(\bar{w}^{k+1} - \bar{w}^0)) \geq 0, \forall w, w \in \Omega
\]
and we set $w = \bar{w}^k$ and get
\[
H(\bar{w}^{k}) - H(\bar{w}^{k+1}) + (\bar{w}^k - \bar{w}^{k+1})^T (J(\bar{w}^{k+1}) - C(\bar{w}^{k+1} - \bar{w}^0)) \geq 0.
\]
(22)

We sum up Equation (21) and (22) to obtain
\[
(\bar{w}^{k+1} - \bar{w}^k)^T C((\bar{w}^{k+1} - \bar{w}^k) - (w - \bar{w}^k)) \geq 0
\]
using Equation (8). Adding the term
\[
((\bar{w}^{k+1} - \bar{w}^k) - (w - \bar{w}^k))^T C((\bar{w}^{k+1} - \bar{w}^k) - (w - \bar{w}^k)) \]
in the both side of the above inequality and using the fact that $w^T C w = 1/2 w^T (C^T + C) w$, we have
\[
\|w^{k+1} - w^k\|_2^2 \leq (\|w^{k+1} - w^k\|_2^2 - \|w - w^k\|_2^2) \geq 0, \forall w, w \in \Omega
\]
and using Equation (13) and $C = DE$, the theorem is proved. \hfill \Box

The next two theorems are the keys to prove the sublinear convergence rate in a nonergodic sense.

**Theorem 9** Let $\bar{w}$ be defined in Equation (10) and matrices $C$, $D$ and $E$ be defined in Equation (11), then we have
\[
\|E(\bar{w}^{k+1} - \bar{w}^k)\|_D \leq \|E(w^{k+1} - \bar{w}^k)\|_D
\]
(23)

**Proof:** Setting $a = E(w^k - \bar{w}^k)$ and $b = E(w^{k+1} - \bar{w}^k)$ in the identity
\[
\|a\|_D^2 - \|b\|_D^2 = 2a^T D(a - b) - \|a - b\|_D^2
\]
We obtain
\[
\|E(w^k - \bar{w}^k)\|_D^2 - \|E(w^{k+1} - \bar{w}^k)\|_D^2
\]
\[
= 2(w^k - \bar{w}^k)^T E^T DE((w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^k))
\]
\[
- \|E(w^k - \bar{w}^k) - E(w^{k+1} - \bar{w}^k)\|_D^2
\]
Applying Equation (8) to the right side of the above equality, we have
\[
\|E(w^k - \bar{w}^k)\|_D^2 - \|E(w^{k+1} - \bar{w}^k)\|_D^2
\]
\[
\geq \|E(w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^k)\|_2^2 \geq 0
\]
using the notation of $G$. The theorem follows directly. \hfill \Box
We add these two and get

$$\|w^{k+1} - w^*\|_D^2 \leq \|w^k - w^*\|_D^2 - \|w^k - w^k\|_G^2$$

(24)

**Proof:** We let $w = w^*$ in Equation (15) and (16) to get

$$H(u^*) - H(u^k) + (w^* - w^k)^T J(w^*) \geq (w^* - w^k)^T DE(w^k - w^k)

= 1/2(\|w^* - w^{k+1}\|_D^2 - \|w^* - w^k\|_D^2 + \|w^k - w^k\|_G^2)

$$

We add these two and get

$$\|w^{k+1} - w^*\|_D^2 \leq \|w^k - w^*\|_D^2 - \|w^k - w^k\|_G^2 + 2(H(u^*) - H(u^k) + (w^* - w^k)^T J(w^*))

\leq \|w^k - w^*\|_D^2 - \|w^k - w^k\|_G^2$$

where the second inequality uses Equation (7) with $w = w^k$. □

It follows from Equation (24) and $G > 0$ that there is a constant $c_0 > 0$ such that

$$\|w^{k+1} - w^*\|_D^2 \leq \|w^0 - w^*\|_D^2 - c_0 \|E(w^k - w^k)\|_D^2$$

(25)

Now we are ready to prove the sublinear convergence rate $O(1/k)$ in the nonergodic sense.

**Theorem 11** Let $\tilde{w}$ be defined in Equation (19) and matrices $D$ and $E$ be defined in Equation (11), for any integer $t > 0$, we have

$$\|E(w^t - \tilde{w}^t)\|_D^2 \leq \frac{1}{(t+1)c_0} \|w^0 - \tilde{w}^0\|_D^2$$

(26)

**Proof:** It follows from Equation Equation (25) that

$$c_0 \sum_{t=0}^\infty \|E(w^t - \tilde{w}^t)\|_D^2 \leq \|w^0 - \tilde{w}^0\|_D^2.$$

Equation Equation (25) implies that $\|E(w^k - w^k)\|_D^2$ is non-increasing and hence

$$(t+1)\|E(w^t - \tilde{w}^t)\|_D^2 \leq \sum_{t=0}^t \|E(w^t - \tilde{w}^t)\|_D^2.$$

As a result, the theorem follows directly. □

2.5 Relationship to Classic Optimization Algorithms

When $n = 2$, $f_1(x) = Ax$ and $f_2(z) = -Bz + c$ are linear functions, the algorithm is reduced to Equation (1). The augmented Lagrangian is formulated as $L_{\rho}(x, z, y) = f(x) + g(z) + y^T (Ax + Bz - c) + \rho/2 \|Ax + Bz - c\|_2^2$ . The ADMM consists of three steps.

$$x^{k+1} = \arg \min_x L_{\rho}(x, z^k, y^k)

z^{k+1} = \arg \min_z L_{\rho}(x^{k+1}, z, y^k)

y^{k+1} = y^k + \rho (Ax^{k+1} + Bz^{k+1} - c)$$

When $n = 1$ and $f_1(x)$ is a linear function, the algorithm is reduced to the Augmented Lagrangian Method (ALM) [2].

$$\text{minimize } F_1(x)
\text{ subject to } f_1(x) = 0$$

The Augmented Lagrangian is formed as $L_{\rho}(x, y) = F_1(x) + y^T f_1(x) + \rho/2 \|f_1(x)\|_2^2$. The ALM consists of two steps.

$$x^{k+1} = \arg \min_x L_{\rho}(x, y^k)

y^{k+1} = y^k + \rho f_1(x).$$

3 Applications

In this section, several representative applications of our neADMM framework are presented. Analysis and discussions show these applications enjoy the good properties of our neADMM.

3.1 Two Numerical Examples

First, two concrete optimization problems with nonlinear equality constraints are provided and the solution based on our neADMM are discussed.

**Example 1**

$$\text{minimize } x + z
\text{ subject to } \sqrt{x} + \sqrt{z} = 1$$

The Augmented Lagrangian is

$$L_{\rho}(x, z, y) = x + z + y(\sqrt{x} + \sqrt{z} - 1) + \rho/2 (\sqrt{x} + \sqrt{z} - 1)^2$$

The ADMM consists of several steps

$$x^{k+1} = \arg \min_x (x + y^k / \sqrt{x} + \rho/2 (\sqrt{x} + \sqrt{z} - 1)^2)

z^{k+1} = \arg \min_z (z + y^k / \sqrt{x} + \rho/2 (\sqrt{x} + \sqrt{z} - 1)^2)

y^{k+1} = y^k + \rho (\sqrt{x^{k+1}} + \sqrt{z^{k+1}} - 1)$$

**Example 2**

$$\text{minimize } x + z
\text{ subject to } x^2 + z^2 = 1$$

The Augmented Lagrangian is $L_{\rho}(x, z, y) = x + z + y(x^2 + z^2 - 1) + \rho/2 (x^2 + z^2 - 1)^2$. The ADMM consists of several steps

$$x^{k+1} = \arg \min_x L_{\rho}(x, z^k, y^k)

z^{k+1} = \arg \min_z L_{\rho}(x^{k+1}, z, y^k)

y^{k+1} = y^k + (z^{k+1})^2 + (z^{k+1})^2 - 1$$

3.2 Optimization Problems with Orthogonality Constraints

The orthogonality constrained optimization problem is widely applied in various research areas [17], including 1-bit compressive sensing [3] as well as mesh processing [20]. The matrix version is shown below.

$$\text{minimize } L(G) + \Omega(G)
\text{ subject to } G^T G = I$$
where \( G \) is a symmetric matrix, \( L(G) \) is a loss function, \( \Omega(G) \) is a regularization term and \( I \) is an identity matrix. The vector version with spherical constraints is shown below.

\[
\begin{align*}
\text{minimize} & \quad L(g) + \lambda \Omega(g) \\
\text{subject to} & \quad \|g\|_2^2 = 1
\end{align*}
\]

The notation is the same as the above. We take the matrix version as an example. We introduce the variable \( W \), the problem is formulated as:

\[
\begin{align*}
\text{minimize} & \quad L(G) + \lambda \Omega(W) \\
\text{subject to} & \quad G^T G = I, W = G
\end{align*}
\]

We form the augmented Lagrangian \( L\rho(G, W, Y_1, Y_2) = L(G) + \lambda \Omega(W) + \langle Y_1, G^T G - I \rangle + \langle Y_2, W - G \rangle - \rho/2 \|G^T G - I\|_2^2 - \rho/2 \|W - G\|_2^2 \). The ADMM consists of three parts:

\[
\begin{align*}
G^{k+1} &= \arg \min_G L\rho(G, W^k, Y_1^k, Y_2^k) \\
W^{k+1} &= \arg \min_W L\rho(G^k, W, Y_1^k, Y_2^k) \\
Y_1^{k+1} &= Y_1^k + \rho((G^{k+1})^T G^{k+1} - I) \\
Y_2^{k+1} &= Y_2^k + \rho(W^{k+1} - G^{k+1})
\end{align*}
\]

### 3.3 1-Bit Compressive Sensing

In recent years, 1-bit compressive sensing became an important research topic of compressive sensing [3]. A common formulation of it could be:

\[
\begin{align*}
\text{minimize} & \quad \|x\|_1 + \lambda/2 \sum \min(Y \phi x, 0)^2 \\
\text{subject to} & \quad \|x\|_2 = 1
\end{align*}
\]

Now we apply our novel ADMM model to this problem. The problem is reformulated as:

\[
\begin{align*}
\text{minimize} & \quad \|w\|_1 + \lambda/2 \sum \min(z, 0)^2 \\
\text{subject to} & \quad \|x\|_2 = 1, Y \phi w = z, w = x
\end{align*}
\]

The augmented Lagrangian is formed as

\[
\begin{align*}
L\rho(x, z, w, y_1, y_2) &= \|w\|_1 + \lambda/2 \sum \min(z, 0)^2 + \\
&\rho/2 \|\|x\|_2 - 1 + y_1 / \rho\|_2^2 + \rho/2 \|Y \phi w - z + y_2 / \rho\|_2^2 + \\
&\rho/2 \|w - x + y_1 / \rho\|_2^2
\end{align*}
\]

The ADMM consists of several steps:

\[
\begin{align*}
x^{k+1} &= \arg \min \|\|x\|_2 - 1 + y_1 / \rho\|_2^2 + \|w - x + y_2 / \rho\|_2^2 \\
y^{k+1} &= \arg \min \lambda/2 \sum \min(z, 0)^2 + \rho/2 \|Y \phi w - z + y_2 / \rho\|_2^2 \\
w^{k+1} &= \arg \min \rho/2 \|w - x^{k+1} + y_2 / \rho\|_2^2 \\
y_1^{k+1} &= y_1^k + \rho((x^{k+1})^2 - 1) \\
y_2^{k+1} &= y_2^k + \rho(Y \phi w^{k+1} - z^{k+1}) \\
y_3^{k+1} &= y_3^k + \rho(w^{k+1} - x^{k+1})
\end{align*}
\]

The primal residuals and the dual residuals are defined as:

\[
\begin{align*}
r_1^{k+1} &= \|x^{k+1}\|_2^2 - 1 \\
r_2^{k+1} &= Y \phi w^{k+1} - r_2^{k+1} \\
r_3^{k+1} &= w^{k+1} - x^{k+1} \\
s_1^{k+1} &= \rho(w^{k} - w^{k+1}) \\
s_2^{k+1} &= \rho Y \phi (w^{k} - w^{k+1})
\end{align*}
\]

### 4 Experiments

In this section, we evaluate our neADMM with several practical applications, including two numerical examples and 1-bit compressive sensing application. The two numerical examples illustrate convergence properties when the saddle point assumption holds. The 1-bit signal reconstruction problem is considered as an application of our model. The effectiveness and the efficiency of our novel ADMM is assessed against a popular existing method. All the experiments were conducted on a 64-bit machine with Intel(R) core(TM) quad-core processor (i3-3217U CPU@ 1.80GHZ) and 4.0GB memory.

#### 4.1 Two Numerical Examples

The results of two numerical examples mentioned in the previous section are presented in this subsection. The parameter \( \rho \) was set to 1. Figure 1 shows residual plots of two examples. The ADMM converged to the optimal points within tens of iterations in both examples. \( r \) and \( s \) decreased drastically in less than 10 iterations, then dropped slowly as they approached to 0. Figure 2 shows objective values during iterations for two examples. We find that objective values changed when \( r \) and \( s \) decreased drastically, and then remained stable and indicated that the neADMM reached global convergence as expected. These two examples illustrate the point that the neADMM can lead to convergence even if equality constraints are nonlinear as long as the saddle point assumption is satisfied.

#### 4.2 1-bit Compressive Sensing

In this section, we compare our neADMM in the 1-bit signal reconstruction problem with the Randomized Fix Point Continuation algorithm (RFPC) [3]. RFPC is a variant of Fix Point Continuation method.

##### 4.2.1 Parameter Settings and Metrics

Here are the definitions of key parameters: \( N \) represents the number of signals, \( M \) denotes the number of measurements and...
$K$ denotes the number of nonzero signals. In this experiment, we set $N = 512$. The generation process of simulated data is as follows: $K$ is randomly chosen from $1$ to $N$. Each nonzero signal is drawn from a standard normal distribution. The measurement matrix $\phi$ whose coefficients are i.i.d also drawn from a standard normal distribution.

The neADMM is compared with the RFPC. The common tuning variable $\lambda$ was set to $10$. $\rho$ was set to $\rho^{k+1} = \rho^k + 1$ with $\rho^0 = 100$.

In this experiment, the 1-bit signal reconstruction task was to reconstruct sparse signals based on signal measurements. The Mean Square Error (MSE) was utilized to evaluate model performance. MSE is the expected bias between original signals and reconstructed signals.

### 4.2.2 Performance

In this subsection, experimental results for both the neADMM and the RFPC are analyzed. Two graphs plotted on prime residuals $r_1, r_2, r_3$ and dual residuals $s_1, s_2$ for the neADMM are shown in Figure 3 when $K = 16$ and $M = 50$. It indicated that all residuals were reduced to very small numbers near 0. However, as they approached to 0, they vibrated irregularly and therefore the neADMM failed to converge. This is because the saddle point assumption doesn’t hold. However, the neADMM was still applicable to this problem as long as these residuals were small enough.

Figure 4 shows the relationship between objective value and measurements for different choices of $K$. Overall, the objective of the neADMM was far lower than that of the RFPC, and the trends for different $K$ were similar. When $K = 16$, the objective value of the neADMM remained stable, except there was a slight raise when $M = 450$, while the objective value of the RFPC fluctuated drastically, it reached a local peak at around 6000 when $M = 200$, hit bottom at 2500 when $M = 400$, and raised quickly to 8000. When $K = 32$, the objective value of the neADMM raised to about 2000 and 8000 when $M = 200$ and $M = 400$, respectively; the objective value of the RFPC increased to 20000 when $M = 500$; when $K = 48$ and $K = 64$, the objective value of the neADMM raised when fluctuation.

Table 1 summarizes the MSE of the neADMM compared with the RFPC on the simulated data for four different choices of $K$. Overall, most MSEs of the neADMM are smaller than those of the RFPC. When $K = 16$ and $K = 32$, all MSEs of the neADMM were smaller than those of the RFPC. When $K = 48$, the neADMM was superior to the RFPC expect when $M = 400$, while the trend reversed when $K = 64$.

| Method | $K = 16$ | $K = 32$ | $K = 48$ | $K = 64$ |
|--------|----------|----------|----------|----------|
| RFPC   | 0.0138   | 0.0238   | 0.0316   | 0.0323   |
| neADMM | 0.0337   | 0.0318   | 0.0311   | 0.0305   |

Table 1: MSE of the RPFC and the neADMM for five choices of $M$ and four choices of $K$.

5 Conclusions

This paper proposes neADMM, a new generic nonconvex ADMM framework that focuses on the nonlinear equality-constrained problems. The theoretical properties including convergence and convergence rate are comprehensively analyzed. New strategy on penalty-parameter updating has also been proposed to further strengthen the applicability of the framework. Several important, challenging, and practical problems such as 1-bit compressive sensing are selected to demonstrate the usefulness of our neADMM. Finally, extensive experiments were conducted to demonstrate the effectiveness of our neADMM.

References

[1] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681–695, 2011.

[2] D. P. Bertsekas. *Nonlinear programming*. Athena scientific Belmont, 1999.

[3] P. T. Boufounos and R. G. Baraniuk. 1-bit compressive sensing. In *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*, pages 16–21. IEEE, 2008.
[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[5] E. Candes and B. Recht. Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119, 2012.

[6] M. Celis, J. Dennis, and R. Tapia. A trust region strategy for nonlinear equality constrained optimization. *Numerical optimization*, 1984:71–82, 1985.

[7] C. Chen, B. He, Y. Ye, and X. Yuan. The direct extension of admm for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming*, 155(1-2):57–79, 2016.

[8] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.

[9] M. A. Figueiredo and J. M. Bioucas-Dias. Deconvolution of poissonian images using variable splitting and augmented lagrangian optimization. In *Statistical Signal Processing, 2009. SSP’09. IEEE/SP 15th Workshop on*, pages 733–736. IEEE, 2009.

[10] M. A. Figueiredo and J. M. Bioucas-Dias. Restoration of poissonian images using alternating direction optimization. *IEEE transactions on Image Processing*, 19(12):3133–3145, 2010.

[11] J.-F. Giovannelli and A. Coulais. Positive deconvolution for superimposed extended source and point sources. *Astronomy & Astrophysics*, 439(1):401–412, 2005.

[12] T. Goldstein and S. Osher. The split bregman method for l1-regularized problems. *SIAM journal on imaging sciences*, 2(2):323–343, 2009.

[13] B. He and X. Yuan. On the o(1/n) convergence rate of the douglas–rachford alternating direction method. *SIAM Journal on Numerical Analysis*, 50(2):700–709, 2012.

[14] M. Hong and Z.-Q. Luo. On the linear convergence of the alternating direction method of multipliers. *Mathematical Programming*, pages 1–35, 2012.

[15] F. Iutzeler, P. Bianchi, P. Ciblat, and W. Hachem. Linear convergence rate for distributed optimization with the alternating direction method of multipliers. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 5046–5051. IEEE, 2014.

[16] S. J. Julier and J. J. LaViola. On kalman filtering with nonlinear equality constraints. *IEEE Transactions on Signal Processing*, 55(6):2774–2784, 2007.

[17] R. Lai and S. Osher. A splitting method for orthogonality constrained problems. *Journal of Scientific Computing*, 58(2):431–449, 2014.

[18] C. T. Lawrence and A. L. Tits. Nonlinear equality constraints in feasible sequential quadratic programming. *Optimization Methods and Software*, 6(4):265–282, 1996.

[19] C.-J. Lin. Projected gradient methods for nonnegative matrix factorization. *Neural computation*, 19(10):2756–2779, 2007.

[20] T. Neumann, K. Varanasi, C. Theobalt, M. Magnor, and M. Wacker. Compressed manifold modes for mesh processing. In *Computer Graphics Forum*, volume 33, pages 35–44. Wiley Online Library, 2014.

[21] H. Ouyang, N. He, L. Tran, and A. G. Gray. Stochastic alternating direction method of multipliers. *ICML (1)*, 28:80–88, 2013.

[22] Y. Ouyang, Y. Chen, G. Lan, and E. Pasiliao Jr. An accelerated linearized alternating direction method of multipliers. *SIAM Journal on Imaging Sciences*, 8(1):644–681, 2015.

[23] Q. Tran-Dinh and V. Cevher. A primal-dual algorithmic framework for constrained convex minimization. *arXiv preprint arXiv:1406.5403*, 2014.

[24] B. Wahlberg, S. Boyd, M. Annergren, and Y. Wang. An admm algorithm for a class of total variation regularized estimation problems. *IFAC Proceedings Volumes*, 45(16):83–88, 2012.

[25] H. Wang and A. Banerjee. Bregman alternating direction method of multipliers. In *Advances in Neural Information Processing Systems*, pages 2816–2824, 2014.

[26] X. Wang, Z. Zhang, Y. Ma, X. Bai, W. Liu, and Z. Tu. Robust subspace discovery via relaxed rank minimization. *Neural computation*, 26(3):611–635, 2014.

[27] J. Yang and Y. Zhang. Alternating direction algorithms for \ell_1-problems in compressive sensing. *SIAM journal on scientific computing*, 33(1):250–278, 2011.

[28] Y. Zhang. An alternating direction algorithm for nonnegative matrix factorization. *preprint*, 2010.

[29] L. Zhao, Q. Sun, J. Ye, F. Chen, C.-T. Lu, and N. Ramakrishnan. Multi-task learning for spatio-temporal event forecasting. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1503–1512. ACM, 2015.