Iterative Complexity and Applications of Linearized
Generalized Alternating Direction Method of
Multipliers with Multi-block Case

Jian He\textsuperscript{a,b}, Bangzhong Zhang\textsuperscript{c}, Jinlin Li\textsuperscript{a,*d}

\textsuperscript{a}School of Economics and Business Administration, Chongqing University, Chongqing 400030, People’s Republic of China.
\textsuperscript{b}Chongqing Chemical Industry Vocational College, Chongqing 401228, People’s Republic of China.
\textsuperscript{c}Chongqing Tsinghua High School, Chongqing 400054, People’s Republic of China.
\textsuperscript{d}Postdoctoral Workstation of Chongqing Rural Commercial Bank, Chongqing 400023, People’s Republic of China.

Abstract

We consider a multi-block separable convex optimization problem with the linear constraints, where the objective function is the sum of $m$ individual convex functions without overlapping variables. The linearized version of the generalized alternating direction method of multipliers (L-GADMM) is particularly efficient for the two-block separable convex programming problem and its convergence was proved when two blocks of variables are alternatively updated. However, the convergence and some practical applications of the extension ($m \geq 3$) of the L-GADMM is still in its infancy. In this paper, we extend this algorithm to the general case where the objective function consists of the sum of $m$-block convex functions. Theoretically, we prove global convergence of the new method and establish the worst-case convergence rate in the ergodic and nonergodic senses for the proposed algorithm. The efficiency of the new method is further demonstrated through numerical results on the calibration of the correlation matrices.

Keywords: Generalized alternating direction method of multipliers; Convex optimization; Global convergence; Worst case convergence rate; Calibrating

\textsuperscript{*}Corresponding author
Email address: 20102127@cqu.edu.cn (Jinlin Li\textsuperscript{a,*d})
1. Introduction

We consider the multiple-block separable convex minimization problem with linear constraints.

$$
\min \left\{ \sum_{i=1}^{m} \vartheta_i(x_i) \left| \sum_{i=1}^{m} A_i x_i = b, x_i \in X_i, i = 1, 2, \ldots, m \right. \right\},
$$

(1)

where $m$ is any positive integer, $\vartheta_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R}$ ($i = 1, 2, \cdots, m$) is convex function but not necessarily smooth. Let $\ell$ be also a positive integer, $A_i \in \mathbb{R}^{\ell \times n_i}$ ($i = 1, 2, \cdots, m$) be given matrices, $b \in \mathbb{R}^\ell$ be a given vector, and $X_i \subseteq \mathbb{R}^{n_i}$ ($i = 1, 2, \cdots, m$) be nonempty closed convex set. Throughout this paper, we let $n = \sum_{i=1}^{m} n_i$, $x := (x_1, x_2, \ldots, x_m)$ and $X := \prod_{i=1}^{m} X_i$ and the matrice $A_i$ ($i = 1, 2, \cdots, m$) is assumed to be full column rank. Moreover, the solution set of (1) is assumed to be nonempty. The problem (1) has many numerous applications, e.g., the robust principal component analysis (PCA) model with noisy and incomplete data[1], signal or image processing [2, 3, 4], the quadratic discriminant analysis model [5], statistical learning [1, 6, 7, 8], network optimization [9, 10, 11], DSL dynamic spectrum management problem [12, 13].

Though the model (1) has a simple form, the efficient solutions is usually very challenging in practice because of the nonsmoothness of the objective functions and the process of high dimensionality arising from some real-world applications. Thus, it is efficient by the structure-utilizing algorithms for (1) to overcome the difficulties caused by the aforementioned characteristics. Based on the special structure, some operator splitting methods have been developed for the problem (1) with $m=2$. These methods involve the Peaceman-Rachford splitting method (PRSM) [5, 16], the alternating direction method of multipliers (ADMM) [14, 15], and their variants [17-26]. In fact, to solve the problem (1) with $m = 2$, a benchmark is the ADMM proposed originally in [14, 15]. And the schemes
of the ADMM for the optimization problem (1) with two-block are written as follows

\[
\begin{aligned}
    x_1^{k+1} &= \arg\min \{ \vartheta_1(x_1) - x_1^\top A_1^\top y^k + \frac{\rho}{2} \| A_1 x_1 + A_2 x_2^k - b \|^2 | x_1 \in X_1 \}, \\
    x_2^{k+1} &= \arg\min \{ \vartheta_2(x_2) - x_2^\top A_2^\top y^k + \frac{\rho}{2} \| A_1 x_1^{k+1} + A_2 x_2 - b \|^2 | x_2 \in X_2 \}, \\
    y^{k+1} &= y^k - \rho (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b),
\end{aligned}
\]

(2)

where \( y \in \mathbb{R}^\ell \) is the lagrange multiplier and \( \rho > 0 \) is a penalty parameter.

Gabay [27] pointed out that the ADMM was an application of the well-known Douglas-Rachford splitting method (DRSM) in [16] to the dual of (2); and as an application of the proximal point algorithm (PPA) in [28], Eckstein and Bertsekas [23] proposed the following generalized ADMM (GADMM) scheme:

\[
\begin{aligned}
    x_1^{k+1} &= \arg\min \{ \vartheta_1(x_1) - x_1^\top A_1^\top y^k + \frac{\rho}{2} \| A_1 x_1 + A_2 x_2^k - b \|^2 | x_1 \in X_1 \}, \\
    x_2^{k+1} &= \arg\min \{ \vartheta_2(x_2) - x_2^\top A_2^\top y^k + \frac{\rho}{2} \| A_1 x_1^{k+1} + (1 - \gamma)(b - A_2 x_2^k) + A_2 x_2^{k+1} - b \|^2 | x_2 \in X_2 \}, \\
    y^{k+1} &= y^k - \rho (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b),
\end{aligned}
\]

(3)

where the parameter \( \gamma \in (0, 2) \) is a relaxation factor. If the parameter \( \gamma = 1 \), the generalized scheme (3) will reduce to the original ADMM scheme (2). Based on the above iterative scheme (3), Fang et al. [29] recently proposed the following linearized versions of GADMM (“L-GADMM” for short) for (3) by adding a proximal term to its subproblem:

\[
\begin{aligned}
    x_1^{k+1} &= \arg\min \{ \vartheta_1(x_1) - x_1^\top A_1^\top y^k + \frac{\rho}{2} \| A_1 x_1 + A_2 x_2^k - b \|^2 \\
    &\quad + \frac{1}{2} \| x_1 - x_1^k \|^2_2 | x_1 \in X_1 \}, \\
    x_2^{k+1} &= \arg\min \{ \vartheta_2(x_2) - x_2^\top A_2^\top y^k + \frac{\rho}{2} \| A_1 x_1^{k+1} + (1 - \gamma)(b - A_2 x_2^k) + A_2 x_2^{k+1} - b \|^2 | x_2 \in X_2 \}, \\
    y^{k+1} &= y^k - \rho (A_1 x_1^{k+1} + (1 - \gamma)(b - A_2 x_2^k) + A_2 x_2^{k+1} - b),
\end{aligned}
\]

(4)
and

\[
\begin{align*}
  x_{1}^{k+1} &= \arg\min \{ \vartheta_1(x_1) - x_1^\top A_1^\top y^k + \frac{\rho}{2} \| A_1 x_1 + A_2 x_2^k - b \|^2 \\
  &\quad + \frac{1}{2} \| x_1 - x_1^k \|^2_{P_1} | x_1 \in X_1 \}, \\
  x_{2}^{k+1} &= \arg\min \{ \vartheta_2(x_2) - x_2^\top A_2^\top y^k + \frac{\rho}{2} \| \gamma A_1 x_1^{k+1} + (1 - \gamma)(b - A_2 x_2^k) \|^2 \\
  &\quad + A_2 x_2 - b \|^2 + \frac{1}{2} \| x_2 - x_2^k \|^2_{P_2} | x_2 \in X_2 \}, \\
  y^{k+1} &= y^k - \rho(\gamma A_1 x_1^{k+1} + (1 - \gamma)(b - A_2 x_2^k) + A_2 x_2^{k+1} - b),
\end{align*}
\]

where $P_1 \in \mathbb{R}^{n_1 \times n_1}$ and $P_2 \in \mathbb{R}^{n_2 \times n_2}$ are symmetric positive definite matrices.

In this paper, we focus on the above linearized version of the GADMM, which is an efficient and simple acceleration algorithm. Note that we use the notation $\| x_1 \|_{P_1}$ to denote the quantity $\sqrt{x_1^\top P_1 x_1}$. If $X_1 = \mathbb{R}^{n_1}$ and $P_1 = \tau I_{n_1} - \rho A_1^\top A_1$ with the requirement $\tau > \rho \| A_1^\top A_1 \|_2$, where $\| \cdot \|_2$ denotes the spectral norm of a matrix, then the $x_1$-subproblem in (4) or (5) reduces to estimating the resolvent operator of $\vartheta_1$:

\[
x_1^{k+1} = (I + \frac{1}{\tau} \partial \vartheta_1)^{-1}(t) = \arg\min \{ \vartheta_1(x_1) + \frac{\tau}{2} \| x_1 - t \|^2 \},
\]

where $\partial(\cdot)$ denotes the subdifferential of a convex function and $t = \frac{1}{\tau}(\tau I_{n_1} - \rho A_1^\top A_1)x_1 - A_1^\top A_2 x_2^k + A_1^\top y^k + \rho A_1^\top b)$. Therefore, the scheme (4) includes the linearized version of the ADMM (see [26, 30, 31]) as a special case with $P_1 = \tau I_{n_1} - \rho A_1^\top A_1$ and $\gamma = 1$.

Theoretically, Fang et al. [29] showed the worst $\mathcal{O}(1/t)$ convergence rate in both ergodic and nonergodic senses for the L-GADMM (4) and (5). Numerically, Fang et al. [29] verified that both the L-GADMM (4) and the scheme (5) were quite efficient for solving some rather new and core applications arising in statistical learning. However, both the iterative scheme (4) and (5) are only suit for the minimization problem (1) with $m = 2$. In fact, there are some efficient algorithms to solve the problem (1) with $m \geq 3$, we thus refer to [20, 32, 33]. In this paper, we regroup the $m$ variables $x_i$ ($i = 1, 2, \cdots, m$) as two blocks $(x_1, x_2, \cdots, x_{m-1})$ and $x_m$ and further separate the subproblems related to the first block $(x_1, x_2, \cdots, x_{m-1})$ into $m - 1$ parallel smaller subproblems.
Furthermore, our algorithm is written as follows

$$
\begin{align*}
\begin{cases}
    x_j^{k+1} &= \text{argmin} \{ \vartheta_j(x_j) + \frac{\rho}{2} \| A_j x_j + \sum_{i=1, i \neq j}^m A_i x_i^k - b - \frac{y_j^k}{\rho} \|^2 + \frac{1}{2} \| x_j - x_j^k \|^2 \mid x_j \in \mathcal{A}_j \}, \\
    j &= 1, \cdots, m - 1, \\
    x_m^{k+1} &= \text{argmin} \{ \vartheta_m(x_m) + \frac{\rho}{2} \gamma \sum_{i=1}^{m-1} A_i x_i^{k+1} + (1 - \gamma)(b - A_m x_m^k) + A_m x_m - b - \frac{y_m^k}{\rho} \|^2 \\
    &+ \frac{1}{2} \| x_m - x_m^k \|^2 \mid x_m \in \mathcal{X}_m \}, \\
    y^{k+1} &= y^k - \rho \left( \gamma \sum_{i=1}^{m-1} A_i x_i^{k+1} + (1 - \gamma)(b - A_m x_m^k) + A_m x_m^{k+1} - b \right), 
\end{cases}
\end{align*}
$$

(6)

where $P_i \in \mathbb{R}^{n_i \times n_i} (i = 1, 2, \cdots, m - 1)$ are symmetric positive definite matrices, $\gamma \in (0, 2)$ is a relaxation factor and $P_m \in \mathbb{R}^{n_m \times n_m}$ is a symmetric matrix with $P_m + \frac{\rho}{\gamma} A_m^T A_m \succ 0$. Here for a matrix $P \succ 0$ means that $P$ is positive definite.

It is noteworthy that the above iterative scheme (6) has the following features:

(I) It minimizes $m$ subproblems with respect to $x_i (i = 1, 2, \cdots, m)$ separately and solves the first $m - 1$ subproblems in parallel; (II) The variable $x_m$ is updated by using all the latest information ($x_1^{k+1}, x_2^{k+1}, \cdots, x_{m-1}^{k+1}$). In fact, the scheme (6) is a partially parallel splitting method. (III) All the subproblems for the iterative scheme (6) are linearized and the corresponding matrices for the quadratic proximal terms are positive definite but the matrix $P_m$. Specially, the matrix $P_m$ is not necessarily positive definite, we only need to ensure that the matrix satisfies the condition: $P_m + \frac{\rho}{\gamma} A_m^T A_m \succ 0$.

**Our contribution:**

- We extend L-GADMM (6) to the general case where the objective function consists of the sum of $m$-block ($m \geq 2$) convex functions.
- We show the global convergence of the new method and establish the worst-case convergence rate in the ergodic and nonergodic senses for the proposed algorithm (6).
- The efficiency of the new method is further demonstrated through the calibration of the correlation matrices.
The rest of this paper is organized as follows. In section 2, we give some notations and characterize problem (1) by a variational inequality problem. Then, we propose L-GADMM iterative scheme and prove its global convergence in section 3. In section 4, we establish a worst-case $O(1/t)$ convergence rate in both ergodic and nonergodic senses for the proposed method. In Section 5, we apply the new algorithm to solve the matrix optimization problem and report some experiment results. Finally, we draw some conclusions in Section 6.

2. Preliminaries

In this section, we give some notations and summarize some properties that will be used throughout this paper.

We use $\mathbb{R}^n$ to denote an $n$-dimensional Euclidean space. The superscript $^\top$ represents the transpose operation for matrix/vector variables. For any a vector $x \in \mathbb{R}^n$, let $\|x\|_p = \sqrt{x^\top x}$ be the $p$-norm. Moreover, $I$ represents the identity matrix and $0$ represents a zero vector or matrix. It is well known that solving (1) is equivalent to solving a variational inequality problem (denoted by $VI(\mathcal{W}, \mathcal{F}, \vartheta)$): Find a vector $w^* \in \mathcal{W}$ such that

$$\vartheta(u) - \vartheta(u^*) + (w - w^*)^\top \mathcal{F}(w^*) \geq 0, \quad \forall w \in \mathcal{W},$$

where

$$u = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \quad w = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ y \end{pmatrix}, \quad \vartheta(u) = \sum_{i=1}^m \vartheta_i(x_i)$$

and

$$\mathcal{F}(w) = \begin{pmatrix} -A_1^\top y \\ -A_2^\top y \\ \vdots \\ -A_m^\top y \\ \sum_{i=1}^m A_ix_i - b \end{pmatrix}, \quad \mathcal{W} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_m \times \mathbb{R}^\ell,$$
The solution set of $VI(W, F, \vartheta)$ denoted $W^*$. Obviously, $W^*$ is nonempty since the solution set of (1) is nonempty. It is easy to verify that the mapping $F(\cdot)$ defined by (9) is monotone and satisfies the following a property:

$$(w_1 - w_2)^\top (F(w_1) - F(w_2)) \geq 0, \quad \forall \ w_1, w_2 \in W.$$

In the following, we will give some matrices which will be used later. Set

$$Q = \begin{pmatrix} G_1 & 0 & 0 \\ 0 & \rho A_m^\top A_m + P_m & (1 - \gamma)A_m^\top \\ 0 & -A_m & \frac{1}{\rho} I_\ell \end{pmatrix}, \quad (10)$$

$$M = \begin{pmatrix} I_1 & 0 & 0 \\ 0 & I_{n_m} & 0 \\ 0 & -\rho A_m & \gamma I_\ell \end{pmatrix}, \quad (11)$$

and

$$H = \begin{pmatrix} G_1 & 0 & 0 \\ 0 & P_m + \frac{\rho}{\gamma} A_m^\top A_m & \frac{1 - \gamma}{\gamma} A_m^\top \\ 0 & \frac{1 - \gamma}{\gamma} A_m & \frac{1}{\gamma \rho} I_\ell \end{pmatrix}, \quad (12)$$

where

$$I_1 = \begin{pmatrix} I_{n_1} & 0 & \cdots & 0 \\ 0 & I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_{n_{m-1}} \end{pmatrix}, \quad G_1 = \begin{pmatrix} P_1 & -\rho A_1^\top A_2 & \cdots & -\rho A_1^\top A_{m-1} \\ -\rho A_2^\top A_1 & P_2 & \cdots & -\rho A_2^\top A_{m-1} \\ \cdots & \cdots & \ddots & \cdots \\ -\rho A_{m-1}^\top A_1 & -\rho A_{m-1}^\top A_2 & \cdots & P_{m-1} \end{pmatrix}.$$
Lemma 2.1 Let the matrices $Q, M, H$ be defined in (10), (11) and (12), respectively. If $\alpha \in (0, 2)$ and $G_1 \succ 0$, then
(I). $Q = HM$.
(II). The matrix $H$ is positive definite.
(III). $N := Q^\top + Q - M^\top H M \succ 0$.

3. The global convergence of the L-GADMM scheme

In this section, we further develop the iterative scheme (6) for solving (1) and establish its global convergence.

Algorithm 1 A linearized GADMM for $VI(W, F, \vartheta)$

Input $\gamma \in (0, 2), \rho > 0, P_i \in \mathbb{R}^{n_i \times n_i}, (i = 1, 2, \cdots, m - 1)$ such that the matrix $G_1$ is positive definite and $P_m \in \mathbb{R}^{n_m \times n_m}$ is a symmetric matrix with $P_m + \frac{\rho}{\gamma} A_m^\top A_m \succ 0$.

Initialize $(x_1, x_2, \cdots, x_m, y) = (x_0^1, x_0^2, \cdots, x_0^m, y^0), k = 0.$

while a stopping criterion is not satisfied do
(1) Compute $w^{k+1} = (x_1^{k+1}, x_2^{k+1}, \cdots, x_m^{k+1}, y^{k+1})$ by (6).
(2) Set $k = k + 1$.
end while

Output $x_1^{k+1}, x_2^{k+1}, \cdots, x_m^{k+1}$.

Throughtout this paper, we definite the following an auxiliary sequence $\{\tilde{w}^k\}$, which is used to prove the global convergence of Algorithm 1.

\[ \tilde{w}^k = \begin{pmatrix} x_1^k \\ x_2^k \\ \vdots \\ x_m^k \\ y^k \end{pmatrix} = \begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ \vdots \\ x_m^{k+1} \\ y^k - \rho(\sum_{i=1}^{m-1} A_i x_i^{k+1} + A_m x_m^k - b) \end{pmatrix}, \quad (13) \]
where \((x_1^{k+1}, x_2^{k+1}, \cdots, x_m^{k+1})\) is generated by the scheme (6) from \((x_1^{k}, x_2^{k}, \cdots, x_m^{k})\).

Furthermore, according to (10) and (13), we get

\[
w^{k+1} = w^k - \mathcal{M}(w^k - \bar{w}^k).
\]

(14)

Now, we give the convergence analysis of Algorithm 1.

**Lemma 3.1** Let the sequence \(\{w^k\}\) be generated by L-GACMM scheme (6).

Then we have

\[
\vartheta(w) - \vartheta(\bar{w}^k) + (w - \bar{w}^k)^\top \mathcal{F}(\bar{w}^k) \geq (w - \bar{w}^k)^\top \mathcal{Q}(w^k - \bar{w}^k), \quad \forall w \in \mathcal{W},
\]

(15)

where \(\mathcal{F}, \mathcal{Q}\) is defined in (9) and (10), respectively.

**Proof** By the first-order optimality condition of the \(x_j\)-subproblem \((j = 1, \ldots, m-1)\) in (6), for any \(x_j \in \mathcal{X}_j\), we have

\[
\vartheta_j(x_j) - \vartheta_j(x_j^{k+1}) + (x_j - x_j^{k+1})^\top \left\{-A_j^\top (y^k - \rho(\gamma \sum_{i=1,i \neq j} A_i x_i^k + (1 - \gamma)(b - A_m x_m^k) + A_m x_m^{k+1} - b)) + P_j(x_j^{k+1} - x_j^k)\right\} \geq 0.
\]

(16)

It follows from (13) and (16) that

\[
\vartheta_j(x_j) - \vartheta_j(x_j^k) + (x_j - x_j^k)^\top \left\{-A_j^\top y^k + \rho A_j^\top \left(\sum_{i=1,i \neq j} A_i(x_i^k - \bar{x}_i^k)\right) + P_j(x_j^k - \bar{x}_j^k)\right\} \geq 0, \quad \forall x_j \in \mathcal{X}_j.
\]

(17)

Furthermore, by the optimality of \(x_m\)-subproblem in (6), for any \(x_m \in \mathcal{X}_m\), we obtain

\[
\vartheta_m(x_m) - \vartheta_m(x_m^{k+1}) + (x_m - x_m^{k+1})^\top \left\{-A_m^\top \left(y^k - \rho \left(\gamma \sum_{j=1}^{m-1} A_j x_j^{k+1} + (1 - \gamma)(b - A_m x_m^k) + A_m x_m^{k+1} - b)\right)\right) + P_m(x_m^{k+1} - x_m^k)\right\} \geq 0.
\]

(18)

From (13) and (18), we have

\[
\vartheta_m(x_m) - \vartheta_m(x_m^k) + (x_m - x_m^k)^\top \left(-A_m^\top \bar{y}^k - (1 - \gamma)A_m^\top (y^k - \bar{y}^k) + (P_m + \rho A_m A_m)(\bar{x}_m^k - x_m^k)\right) \geq 0, \quad \forall x_m \in \mathcal{X}_m.
\]

(19)
Moreover, by (13), we get
\[(y - \bar{y}^k)^\top \left\{ \sum_{j=1}^{m} A_j \bar{x}_j^k - b \right\} - \mathcal{A}_m (\bar{x}_m^k - x_m^k) + \frac{1}{\rho} (\bar{y}^k - y^k) \geq 0, \forall y \in \mathcal{R}^\ell. \]

(20)

It follows from (8), (9), (10), (17), (19) and (20) that
\[\vartheta(u) - \vartheta(\bar{u}^k) + (w - \bar{w}^k)^\top \{ \mathcal{F}(\bar{w}^k) + \mathcal{Q}(\bar{w}^k - w^k) \} \geq 0, \forall w \in \mathcal{W}, \]

(21)

which implies (15).

From (21), if \(w^k = \bar{w}^k\), we have
\[\vartheta(u) - \vartheta(\bar{u}^k) + (w - \bar{w}^k)^\top \mathcal{F}(\bar{w}^k) \geq 0, \forall w \in \mathcal{W}. \]

(22)

It follows from (22) that \(\bar{w}^k = (\bar{x}_1^k, \bar{x}_2^k, ..., \bar{x}_m^k, \bar{y}^k)\) is a solution of \(VI(\mathcal{W}, \mathcal{F}, \vartheta)\).

Moreover, by (13), we can obtain the following a result which implies the Algorithm 1 to produce two infinite sequences \(\{ w^k \} \) and \(\{ \bar{w}^k \} \):

Remark 3.1 If \(A_i x_i^k = A_i \bar{x}_i^k (i = 1, 2, ..., m)\) and \(y^k = \bar{y}^k\), then \(w^{k+1}\) produced by Algorithm 1 is a solution of \(VI(\mathcal{W}, \mathcal{F}, \vartheta)\).

In the following, we show that the accuracy of \(\bar{w}^k\) to a solution of \(VI(\mathcal{W}, \mathcal{F}, \vartheta)\) is measured by the quantity \((w - \bar{w}^k)^\top \mathcal{Q}(u^k - \bar{w}^k)\).

Lemma 3.2 Let the sequence \(\{ w^k \} \) be generated by Algorithm. Then we have
\[\(w - \bar{w}^k\)^\top \mathcal{Q}(w^k - \bar{w}^k) = \frac{1}{2} (\|w - w^{k+1}\|_{\mathcal{H}}^2 - \|w - w^k\|_{\mathcal{H}}^2 + \|w^k - \bar{w}^k\|_{\mathcal{H}}^2), \forall w \in \mathcal{W}, \]

(23)

where \(\mathcal{N} = \mathcal{Q} + \mathcal{Q}^\top - \mathcal{M}^\top \mathcal{H} \mathcal{M}\).

Proof It follows from the positive definiteness of \(\mathcal{H}\) that
\[\(w - \bar{w}^k\)^\top \mathcal{H}(w^k - w^{k+1}) = \frac{1}{2} (\|w - w^{k+1}\|_{\mathcal{H}}^2 - \|w - w^k\|_{\mathcal{H}}^2)
+ \frac{1}{2} (\|w^k - \bar{w}^k\|_{\mathcal{H}}^2 - \|w^{k+1} - w^k\|_{\mathcal{H}}^2). \]

Furthermore, by the identity (14) and \(\mathcal{Q} = \mathcal{H} \mathcal{M}\), we have
\[\(w - \bar{w}^k\)^\top \mathcal{Q}(w^k - \bar{w}^k) = \frac{1}{2} (\|w - w^{k+1}\|_{\mathcal{H}}^2 - \|w - w^k\|_{\mathcal{H}}^2)
+ \frac{1}{2} (\|w^k - \bar{w}^k\|_{\mathcal{H}}^2 - \|w^{k+1} - w^k\|_{\mathcal{H}}^2). \]

(24)
And
\[ \|w^k - \bar{w}^k\|_H^2 - \|w^{k+1} - \bar{w}^k\|_H^2 = \|w^k - \bar{w}^k\|_H^2 - \|(w^k - \bar{w}^k) - (w^k - w^{k+1})\|_H^2 
= \|w^k - \bar{w}^k\|_H^2 - \|(w^k - \bar{w}^k) - M(w^k - \bar{w}^k)\|_H^2 
= (w^k - \bar{w}^k)^\top (Q^\top + Q - M^\top H M) (w^k - \bar{w}^k) 
= \|w^k - \bar{w}^k\|_N^2, \quad (25) \]

where \( N = Q + Q^\top - M^\top H M \). Thus, the assertion (23) is proved directly by (24) and (25).

In the following, we give a crucial inequality for the global convergence of Algorithm 1 that indicates the sequence \( \{w^k\} \) generated by Algorithm 1 is Fejér monotone with respect to the solution \( W^* \).

**Lemma 3.3** Let \( \{w^k\} \) be the sequence generated by Algorithm 1. Then we have
\[ \|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - \bar{w}^k\|_N^2, \quad \forall w^* \in W^*. \quad (26) \]

**Proof** It follows from (15) and (23) that we obtain
\[ \vartheta(u) - \vartheta(\bar{u}^k) + (w - \bar{w}^k)^\top F(\bar{w}^k) \geq \frac{1}{2} \left( \|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2 + \|w^k - w^*\|_N^2 \right), \quad \forall w \in W. \quad (27) \]

Furthermore, by setting \( w = w^* \in W^* \) in (27) and the monotonicity of \( F(\cdot) \), we get
\[ 0 \geq \vartheta(w^*) - \vartheta(u^k) + (w^* - \bar{w}^k)^\top F(\bar{w}^k) 
\geq \frac{1}{2} \left( \|w^* - w^{k+1}\|_H^2 - \|w^* - w^k\|_H^2 + \|w^k - w^*\|_N^2 \right), \]
which implies the assertion (26). \( \Box \)

Now, we establish the global convergence of Algorithm 1.

**Theorem 3.1** The sequence \( \{w^k\} \) generated by Algorithm 1 converges to an optimal solution which belongs to \( W^* \) for problem (1) from any starting point.

**Proof** It follows from (26) that
\[ \|w^k - \bar{w}^k\|_N^2 \leq \|w^k - w^*\|_H^2 - \|w^{k+1} - w^*\|_H^2, \quad \forall w^* \in W^*. \quad (28) \]
Summing the above inequality over $k = 0, 1, \ldots, \infty$, we have
\[ \sum_{k=0}^{\infty} \|w^k - \bar{w}^k\|_N^2 \leq \|w^0 - w^*\|_H^2, \]
which implies that
\[ \lim_{k \to \infty} \|w^k - \bar{w}^k\|_N^2 = 0. \] (29)
It follows from (29) that $\bar{w}^k \to w^k$ ($k \to \infty$). Moreover, by (15) and (29), we have
\[ \lim_{k \to \infty} \{\vartheta(u) - \vartheta(\bar{u}^k) + (w - \bar{w}^k)^\top F(\bar{w}^k)\} \geq 0, \quad \forall w \in \mathcal{W}, \] (30)
From (26), we know that the sequence $\{w^k\}$ is bounded. Then, it follows from (29) that the sequence $\{\bar{w}^k\}$ is also bounded. Thus, the sequence $\{\bar{w}^k\}$ has at least one cluster point. We assume that $w^\infty$ is a cluster point of $\{w^k\}$ and the subsequence $\{\bar{w}^{k_j}\}$ converges to $w^\infty$ (That is, $\bar{w}^{k_j} \to w^\infty$ ($j \to \infty$) ). From (30), we get
\[ \vartheta(u) - \vartheta(u^\infty) + (w - w^\infty)^\top F(w^\infty) \geq 0, \quad \forall w \in \mathcal{W}, \]
which indicates that $w^\infty \in \mathcal{W}^\ast$. It follows from (29) that $\lim_{k \to \infty} \|w^k - \bar{w}^k\|_H^2 = 0$. Therefore, for any $l > 0$, there exists an integer $N$ such that
\[ \|w^{k_N} - \bar{w}^{k_N}\|_H < \frac{l}{2}, \quad \|\bar{w}^{k_N} - w^\infty\|_H < \frac{l}{2}. \] (31)
Furthermore, It follows from (26) and (31) that for any $k \geq k_N$
\[ \|w^k - w^\infty\|_H \leq \|w^{k_N} - w^\infty\| \leq \|w^{k_N} - \bar{w}^{k_N}\|_H + \|\bar{w}^{k_N} - w^\infty\|_H < l, \]
which shows that $\lim_{k \to \infty} w^k = w^\infty$. It follows from $w^\infty \in \mathcal{W}^\ast$ that we thus complete the proof.

4. The worst case $O(1/t)$ convergence rate

In this section, we establish the worst-case convergence rate for the Algorithm 1 in both ergodic and nonerodic senses.
4.1. A worst case $O(1/t)$ convergence rate in a ergodic sense

In this subsection, we show a worst $O(1/t)$ convergence rate of the Algorithm 1 in an ergodic sense.

**Theorem 4.1** Let $\{w^k\}$ be the sequence generated by the Algorithm 1. For any integer $t > 0$, let $w_t = \frac{1}{t+1} \sum_{k=0}^t \bar{w}^k$, then we have $w_t \in \mathcal{W}$ and

$$
\vartheta(u_t) - \vartheta(u) + (w_t - w)^\top \mathcal{F}(w) \leq \frac{1}{2(t+1)} \|w - w^0\|_H^2, \quad \forall w \in \mathcal{W}.
$$

(32)

**Proof** Obviously, we know that $\bar{w}^k \in \mathcal{W}$ for all $k \geq 0$. It follows from the convexity of $X_1, X_2, ..., X_m$ that $w_t \in \mathcal{W}$. And because of the monotonicity of $\mathcal{F}(\cdot)$, we have

$$
\vartheta(u) - \vartheta(\bar{w}^k) + (w - \bar{w}^k)^\top \mathcal{F}(w) + \frac{1}{2} \|w^k - w\|_H^2 \geq \frac{1}{2} \|w^{k+1} - w\|_H^2, \quad \forall w \in \mathcal{W}.
$$

(33)

Summing the above inequality over $k = 0, 1, ..., t$, we get

$$
(t+1)\vartheta(u) - \sum_{k=0}^t \vartheta(\bar{w}^k) + \left((t+1)w - \sum_{k=0}^t \bar{w}^k\right)^\top \mathcal{F}(w) + \frac{1}{2} \|w - w^0\|_H^2 \geq 0, \quad \forall w \in \mathcal{W}.
$$

(34)

According to the definition of $w_t$ and the convexity of $\vartheta(u)$, (34) can be rewritten as

$$
\vartheta(u_t) - \vartheta(u) + (w_t - w)^\top \mathcal{F}(w) \leq \frac{1}{2(t+1)} \|w - w^0\|_H^2, \quad \forall w \in \mathcal{W}.
$$

(35)

We thus complete the proof. 🅿️

Let an compact set $G \subset \mathcal{W}$ and define

$$
g = \sup \left\{ \|w - w^0\|_H | w \in G \right\},
$$

where $w^0 = (x_1^0, x_2^0, ..., x_m^0, y^0)$ is the initial point. After $t$ iterations of Algorithm 1, we can find a $w_t \in \mathcal{W}$ such that

$$
\sup_{w \in G} \left\{ \vartheta(u_t) - \vartheta(u) + (w_t - w)^\top \mathcal{F}(w) \right\} \leq \frac{g^2}{2t},
$$

which implies that the Algorithm 1 reaches

$$
\vartheta(u_t) - \vartheta(u) + (w_t - w)^\top \mathcal{F}(w) \leq \varepsilon, \quad \forall w \in G,
$$

at most $t = \frac{g^2}{2\varepsilon}$ iterations. We thus establish the worst-case $O(1/t)$ convergence rate for the Algorithm 1 in an ergodic sense.
4.2. a worst case $O(1/t)$ convergence rate in a nonergodic sense

In this subsection, we give a worst $O(1/t)$ convergence rate of the Algorithm 1 in a nonergodic sense. We first have to mention that the term $\|w^k - w^{k+1}\|_H^2$ can be used to measure the accuracy of an iteration.

**Lemma 4.1** let $\{w^{k+1}\}$ be generated by Algorithm 1, Then $\bar{w}^k$ defined in (13) is a solution to (7) if $\|w^k - w^{k+1}\|_H^2 = 0$.

**Proof** By Lemma 2.1, Lemma 3.1 and the equation (14), we have

$$\vartheta(u) - \vartheta(\bar{w}^k) + (w - \bar{w}^k)^\top \mathcal{F}(\bar{w}^k) \geq (w - \bar{w}^k)^\top \mathcal{H}(w^k - w^{k+1}), \ \forall w \in \mathcal{W}, \quad (36)$$

Clearly, $\mathcal{H}$ is positive definite since $G_1$ is positive definite in Algorithm 1 and $\gamma \in (0, 2)$. Then the right-hand side of (36) vanishes if $\|w^{k+1} - w^k\|_H^2 = 0$, since we can obtain $\mathcal{H}(w^{k+1} - w^k) = 0$ whenever $\|w^{k+1} - w^k\|_H^2 = 0$. The assertion is proved. \(\Box\)

Now, we are ready to establish a worst-case $O(1/t)$ convergence rate for Algorithm 1 in a nonergodic sense. First, we show the following two lemmas.

**Lemma 4.2** Let the sequence $\{w^k\}$ be generated by Algorithm 1 with $\gamma \in (0, 2)$ and the associated $\{\bar{w}^k\}$ be defined in (13); the matrix $Q$ be defined in (10). Then, we have

$$(\bar{w}^k - \bar{w}^{k+1})^\top Q \{(w^k - w^{k+1}) - (\bar{w}^k - \bar{w}^{k+1})\} \geq 0. \quad (37)$$

**Proof** Setting $w = \bar{w}^{k+1}$ in (13), we obtain

$$\vartheta(\bar{u}^{k+1}) - \vartheta(\bar{w}^k) + (w^{k+1} - w^k)^\top \mathcal{F}(\bar{w}^k) \geq (w^{k+1} - w^k)^\top Q(w^k - \bar{w}^k), \ \forall w \in \mathcal{W}, \quad (38)$$

It is easy to know that (13) is also true for $k := k + 1$, that is

$$\vartheta(u) - \vartheta(\bar{u}^{k+1}) + (w - \bar{w}^{k+1})^\top \mathcal{F}(\bar{w}^{k+1}) \geq (w - \bar{w}^{k+1})^\top Q(w^{k+1} - \bar{w}^{k+1}), \ \forall w \in \mathcal{W}, \quad (39)$$

By setting $w = \bar{w}^k$ in the above inequality, we have

$$\vartheta(\bar{w}^k) - \vartheta(\bar{u}^{k+1}) + (\bar{w}^k - \bar{w}^{k+1})^\top \mathcal{F}(\bar{w}^{k+1}) \geq (\bar{w}^k - \bar{w}^{k+1})^\top Q(w^{k+1} - \bar{w}^{k+1}), \ \forall w \in \mathcal{W}, \quad (39)$$
Adding (38) and (39) and using the monotonicity of $F$, the assertion (37) is proved immediately.

\[ \Box \]

**Lemma 4.3** Let the sequence $\{w^k\}$ be generated by the Algorithm 1 and $\{\bar{w}^k\}$ be defined in (13); the matrices $Q$, $M$ and $H$ be defined in (10), (11) and (12), respectively. Then we have

\[ (w^k - \bar{w}^k) \top M \top H M \{ (w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1}) \} \geq \frac{1}{2} \| (w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1}) \|^2_{(Q\top + Q)}. \]  

(40)

**Proof** Adding the equation

\[ \{ (w^k - w^{k+1}) - (\bar{w}^k - \bar{w}^{k+1}) \} \top Q \{ (w^k - w^{k+1}) - (\bar{w}^k - \bar{w}^{k+1}) \} = \frac{1}{2} \| (w^k - w^{k+1}) - (\bar{w}^k - \bar{w}^{k+1}) \|^2_{(Q\top + Q)} \]

to both sides of (37), we have

\[ (w^k - w^{k+1}) \top Q \{ (w^k - w^{k+1}) - (\bar{w}^k - \bar{w}^{k+1}) \} \geq \frac{1}{2} \| (w^k - w^{k+1}) - (\bar{w}^k - \bar{w}^{k+1}) \|^2_{(Q\top + Q)}. \]  

(41)

It follows from $Q = H M$, (14) and (41) that the assertion (40) holds. \[ \Box \]

**Theorem 4.4** Let $\gamma \in (0, 2)$, the sequence $\{w^k\}$ be generated by Algorithm 1 and the matrix $H$ be defined in (12). Then, we have

\[ \| w^{k+1} - w^{k+2} \|^2_H \leq \| w^k - w^{k+1} \|^2_H. \]  

(42)

**Proof** Setting $c = M(w^k - \bar{w}^k)$ and $d = M(w^{k+1} - \bar{w}^{k+1})$, we get

\[ \| c \|^2_H - \| d \|^2_H = 2c \top H (c - d) - \| c - d \|^2_H. \]

We thus have

\[ \| M(w^k - \bar{w}^k) \|^2_H - \| M(w^{k+1} - \bar{w}^{k+1}) \|^2_H \]

\[ = 2(w^k - \bar{w}^k) M \top H M \{ (w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1}) \} \]

\[ - \| M \{ (w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1}) \} \|^2_H. \]  

(43)
By (40) and (43), we have
\[
\| \mathcal{M}(w_k - \bar{w}_k) \|^2_{H} - \| \mathcal{M}(w_{k+1} - \bar{w}_{k+1}) \|^2_{H} \\
\geq \| (w_k - \bar{w}_k) - (w_{k+1} - \bar{w}_{k+1}) \|^2_{\mathcal{H}} - \| \mathcal{M}((w_k - \bar{w}_k) - (w_{k+1} - \bar{w}_{k+1})) \|^2_{\mathcal{H}} \\
= \| (w_k - \bar{w}_k) - (w_{k+1} - \bar{w}_{k+1}) \|^2_{\mathcal{N}},
\]
(44)
where \( \mathcal{N} = \mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T \mathcal{H} \mathcal{M} \). It follows from \( \mathcal{Q} = \mathcal{H} \mathcal{M} \) that
\[
\mathcal{N} = (\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T \mathcal{Q}
\]
\[
= \begin{pmatrix}
2\mathcal{G}_1 & 0 & 0 \\
0 & 2(p_m + \rho A_m^T A_m) & -\gamma A_m^T \\
0 & -\gamma A_m & 2 \rho I_{\ell}
\end{pmatrix} - \begin{pmatrix}
\mathcal{G}_1 & 0 & 0 \\
0 & p_m + 2\rho A_m^T A_m & -\gamma A_m^T \\
0 & -\gamma A_m & 2 \rho I_{\ell}
\end{pmatrix}
\]
\[
= \begin{pmatrix}
\mathcal{G}_1 & 0 & 0 \\
0 & p_m & 0 \\
0 & 0 & 2 - \gamma \rho I_{\ell}
\end{pmatrix}.
\]
By the positive definiteness of \( P_m \) and \( \mathcal{G}_1 \), it is easy to see that \( \mathcal{N} \) is positive definite when \( \gamma \in (0, 2) \). Then, we have
\[
\| \mathcal{M}(w_k - \bar{w}_k) \|^2_{H} - \| \mathcal{M}(w_{k+1} - \bar{w}_{k+1}) \|^2_{H} \geq 0.
\]
By \( w_k - w_{k+1} = \mathcal{M}(w_k - \bar{w}_k) \), the assertion (42) follows immediately. \( \square \)

In order to further finish the proof of a worst-case \( \mathcal{O}(1/t) \) convergence rate for Algorithm 1 in a nonergodic sense, we give the following two lemmas.

**Lemma 4.4** Let \( \{x_m^k\} \) be the sequence generated by Algorithm 1 with \( \gamma \in (0, 2) \).

Then, we have
\[
(x_m^k - x_m^{k+1})^T A_m^T (y^k - y^{k+1}) \geq \frac{1}{2} \| x_m^k - x_m^{k+1} \|^2_{P_m} - \frac{1}{2} \| x_m^{k-1} - x_m^k \|^2_{P_m}.
\]
(45)

**Proof** By the optimality condition of the \( x_m \)-subproblem in lemma 3.1 and (6), there exists \( \xi \in \partial \vartheta_m(x_m^{k+1}) \) such that
\[
(x_m - x_m^{k+1})^T \{ \xi - A_m^T y^{k+1} + P_m(x_m^{k+1} - x_m^k) \} \geq 0, \quad \forall x_m \in \mathcal{X}_m,
\]
(46)
where $\partial \vartheta_m(x_m)$ is a subdifferential of $\vartheta_m(x_m)$. Setting $x_m = x_m^k$ in (46), then we have
\[
(x_m^k - x_m^{k+1})^\top \{ \xi - A_m^\top y^{k+1} + P_m(x_m^{k+1} - x_m^k) \} \leq 0.
\] (47)

Furthermore, setting $k := k + 1$ in (46), there exists $\zeta \in \partial \vartheta_m(x_m)$ such that
\[
(x_m - x_m^k)^\top \{ \zeta - A_m^\top y^k + P_m(x_m^k - x_m^{k-1}) \} \geq 0.
\] (48)

Similarly, by $x_m = x_m^{k+1}$ in (48), we have
\[
(x_m^{k+1} - x_m^k)^\top \{ \zeta - A_m^\top y^k + P_m(x_m^k - x_m^{k-1}) \} \geq 0.
\] (49)

Adding (47) to (49) and using the monotonicity of the $\partial \vartheta_m(\cdot)$, we get
\[
(x_m^k - x_m^{k+1})^\top A_m(y^k - y^{k+1}) \geq (x_m^{k+1} - x_m^k)^\top P_m(x_m^{k+1} - x_m^k + x_m^{k-1} - x_m^k)
\]
\[
= \|x_m^{k+1} - x_m^k\|^2_{P_m} + (x_m^{k+1} - x_m^k)^\top P_m(x_m^{k-1} - x_m^k).
\]

By the inequality
\[
(x_m^{k+1} - x_m^k)^\top P_m(x_m^{k-1} - x_m^k) \geq -\frac{1}{2\rho}\|x_m^{k+1} - x_m^k\|^2_{P_m} - \frac{1}{2\rho}\|x_m^{k-1} - x_m^k\|^2_{P_m},
\]
the assertion (45) is proved. \hfill \square

**Lemma 4.5** Let the sequence $\{w^k\}$ be generated by the Algorithm 1 with $\gamma \in (0, 2)$ and the associated $\{\bar{w}^k\}$ be defined in (13), then there exists $0 < \sigma_\gamma \leq 1$ such that
\[
\|w^k - \bar{w}^k\|^2_N \geq \sigma_\gamma \left( \|R^k - \bar{R}^k\|^2_{G_1} + \|x_m^k - \bar{x}_m^k\|^2_{\bar{P}_m} + \frac{\gamma}{\rho} \|y^k - \bar{y}^k\|^2 \right),
\] (50)

where
\[
N = Q^\top + Q - M^\top H M, \quad R^k = \begin{pmatrix} x_1^k \\ x_2^k \\ \vdots \\ x_m^{k-1} \end{pmatrix}
\] (51)

**Proof.** By the definition of $Q, M, H, G_1$ and $R^k$, we have
\[
\|w^k - \bar{w}^k\|^2_N
\]
\[
= \|R^k - \bar{R}^k\|^2_{G_1} + \|x_m^k - \bar{x}_m^k\|^2_{\bar{P}_m} + \frac{2 - \gamma}{\rho} \|y^k - \bar{y}^k\|^2
\]
\[
\geq \min \left\{ \frac{2 - \gamma}{\gamma}, 1 \right\} \left( \|R^k - \bar{R}^k\|^2_{G_1} + \|x_m^k - \bar{x}_m^k\|^2_{\bar{P}_m} + \frac{\gamma}{\rho} \|y^k - \bar{y}^k\|^2 \right).
\] (52)
Let \( \sigma_\gamma = \min \left\{ \frac{2-\gamma}{\gamma}, 1 \right\} \), then it follows from (52) that the assertion (50) holds. \( \Box \)

Finally, we establish a worst-case \( \mathcal{O}(1/t) \) convergence rate for the Algorithm 1 in a nonergodic sense.

**Theorem 4.5** Let the sequence \( \{w^k\} \) be generated by Algorithm 1 with \( \gamma \in (0,2) \). Then we have

\[
\|w^t - w^{t+1}\|_{\mathcal{H}}^2 \leq \frac{1}{t} \left( \frac{1}{\sigma_\gamma} \|w^0 - w^*\|_{\mathcal{H}}^2 + \|x_m^0 - x_m^1\|_{\mathcal{H}}^2 \right). \tag{53}
\]

**Proof** By \( w = w^* \) in (28), we have

\[
\vartheta(u^*) - \vartheta(\bar{u}^k) + (w^* - \bar{w}^k)^\top \mathcal{F}(w^*) \\
\geq \frac{1}{2} (\|w^* - w^{k+1}\|_{\mathcal{H}}^2 - \|w^* - w^k\|_{\mathcal{H}}^2) + \frac{1}{2} \|w^k - \bar{w}^k\|_{\mathcal{N}}^2,
\]

where \( \mathcal{N} = Q^\top + Q - \mathcal{M}^\top \mathcal{H} \mathcal{M} \). By (7) and (54), we can obtain

\[
\|w^k - \bar{w}^k\|_{\mathcal{N}}^2 \leq \|w^* - w^k\|_{\mathcal{H}}^2 - \|w^* - w^{k+1}\|_{\mathcal{H}}^2.
\]

According to the Theorem 4.1, By the positive definiteness of \( P_m \) and \( G_1 \), it is easy to see that \( \mathcal{N} \) is positive definite. Thus, we have

\[
\sum_{k=0}^{\infty} \|w^k - w^k\|_{\mathcal{N}}^2 \leq \|w^0 - w^*\|_{\mathcal{H}}^2. \tag{55}
\]

Moreover, by the definition of \( \mathcal{H} \) in (12), we have

\[
\|w^k - w^{k+1}\|_{\mathcal{H}}^2 \\
= \|R^k - \bar{R}^k\|_{\mathcal{G}_1}^2 + \|x_m^k - x_m^{k+1}\|_{\mathcal{P}_m}^2 \\
+ \frac{1}{\rho \gamma} \left( \|\rho A_m(x_m^k - x_m^{k+1})\|^2 + \|y^k - y^{k+1}\|^2 + 2(1 - \gamma)\rho (x_m^k - x_m^{k+1})^\top A_m^\top (y^k - y^{k+1}) \right) \\
= \|R^k - \bar{R}^k\|_{\mathcal{G}_1}^2 + \|x_m^k - x_m^{k+1}\|_{\mathcal{P}_m}^2 + \frac{1}{\rho \gamma} \|\rho A_m(x_m^k - x_m^{k+1}) + (y^k - y^{k+1})\|^2 \\
- 2(x_m^k - x_m^{k+1})^\top A_m^\top (y^k - y^{k+1}).
\]

Furthermore, by (6) and (13), we have

\[
y^k - y^{k+1} = -\rho A_m(x_m^k - x_m^{k+1}) + \gamma(y^k - \bar{y}^k). \tag{56}
\]

We thus get

\[
\|w^k - w^{k+1}\|_{\mathcal{H}}^2 = \|R^k - \bar{R}^k\|_{\mathcal{G}_1}^2 + \|x_m^k - x_m^{k+1}\|_{\mathcal{P}_m}^2 \\
+ \frac{\gamma}{\rho} \|y^k - \bar{y}^k\|^2 - 2(x_m^k - x_m^{k+1})^\top A_m^\top (y^k - y^{k+1}). \tag{57}
\]

18
It follows from (45), (50), (55) and (56) that
\[ \sum_{k=1}^{t} \| w^{k} - w^{k+1} \|_{\mathcal{H}}^2 \leq \frac{1}{\sigma \gamma} \sum_{k=1}^{t} \| w^{k} - \bar{w}^{k} \|_{\mathcal{H}}^2 + \sum_{k=1}^{t} \left( \| x^{k-1}_m - x^{k}_m \|_{\mathcal{P}_m}^2 - \| x^{k}_m - x^{k+1}_m \|_{\mathcal{P}_m}^2 \right) \]
\[ \leq \frac{1}{\sigma \gamma} \| w^0 - w^* \|_{\mathcal{H}}^2 + \| x^0_m - x^1_m \|_{\mathcal{P}_m}^2. \] (58)

According to the Theorem 4.1, the sequence \( \{ \| w^{k} - w^{k+1} \|_{\mathcal{H}}^2 \} \) is non-increasing. Thus, we get
\[ t \| w^t - w^{t+1} \|_{\mathcal{H}}^2 \leq \sum_{k=1}^{t} \| w^{k} - w^{k+1} \|_{\mathcal{H}}^2 \]
\[ \leq \frac{1}{\sigma \gamma} \| w^0 - w^* \|_{\mathcal{H}}^2 + \| x^0_m - x^1_m \|_{\mathcal{P}_m}^2. \] (59)

Based on the above inequality, the assertion (53) is proved. \( \square \)

It follows from the conclusion in Theorem 4.5 that we can obtain a worst-case \( \mathcal{O}(1/t) \) convergence rate for Algorithm 1 in a nonergodic sense.

5. Numerical experiments

In this subsection, we report some numerical result for the proposed method by calibrating the correlation matrices. We also compare its numerical performance with the ADMM with Gaussian back substitution (ADMM-G) [20] and an algorithm twisted from generalized ADMM (TADMM) [32].

All experiments are implemented in MATLAB R2010b on a hp-notebook with an Intel Core i5-3340M CPU at 2.70 GHz and 8 GB memory.

We first consider to solve the following matrix optimization problem
\[ \min \left\{ \frac{1}{2} \| X - C \|_{F}^2 | X \in S_{+}^n \cap \mathcal{S}_B \right\}, \] (60)
where
\[ S_{+}^n = \{ H \in \mathcal{R}^{n \times n} | H^\top = H, H \succeq 0 \}, \]
We set 

\[ A_1 = \begin{pmatrix} I \\ I \\ 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} -I \\ 0 \\ I \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 \\ -I \\ -I \end{pmatrix} \]

Then, the above problem (60) could be converted to the following equivalent form

\[
\begin{align*}
\min & \frac{1}{2} \| X_1 - C \|^2 + \frac{1}{2} \| X_2 - C \|^2 + \frac{1}{2} \| X_3 - C \|^2 \\
\text{s.t.} \quad & A_1 X_1 + A_2 X_2 + A_3 X_3 = 0, \\
& X_1, \ X_2 \in S^n_a, \ X_3 \in S_B.
\end{align*}
\]

(61)

We then derive the subproblems when Algorithm 1 is applied to (61):

\[
\left\{ \begin{array}{l}
X_1^{k+1} = \arg\min \left\{ \frac{1}{2} \| X_1 - C \|^2 - X_1^T A_1^T y^k + \frac{\rho}{2} \| A_1 X_1 + A_2 X_2^k + A_3 X_3^k - b \|^2 \\
\quad + \frac{1}{2} \| X_1 - X_1^k \|^2 \| \tilde{p}_1 \| X_1 \in X_1 \}, \\
X_2^{k+1} = \arg\min \left\{ \frac{1}{2} \| X_2 - C \|^2 - X_2^T A_2^T y^k + \frac{\rho}{2} \| A_1 X_1 + A_2 X_2^k + A_3 X_3^k - b \|^2 \\
\quad + \frac{1}{2} \| X_2 - X_2^k \|^2 \| \tilde{p}_2 \| X_2 \in X_2 \}, \\
X_3^{k+1} = \arg\min \left\{ \frac{1}{2} \| X_3 - C \|^2 - X_3^T A_3^T y^k + \frac{\rho}{2} \| A_1 X_1^{k+1} + A_2 X_2^{k+1} \| + (1 - \gamma) (b - A_3 X_3^k) \\
\quad + A_3 X_3 - b \|^2 + \frac{1}{2} \| X_3 - X_3^k \|^2 \| \tilde{p}_3 \| X_3 \in X_3 \}, \\
y^{k+1} = y^k - \rho \left( \gamma (A_1 X_1^{k+1} + A_2 X_2^{k+1}) + (1 - \gamma) (b - A_3 X_3^{k+1}) + A_3 X_3^{k+1} - b \right).
\end{array} \right.
\]

(62)

For the \(X_1\)-subproblem in (62), by first-order optimal condition, we have

\[
X_1^{k+1} = P_{S_{a^n}} \left\{ (P_1 + (1 + 2\beta) \cdot I)^{-1} \left( -\beta A_1^T A_2 X_2^k - \beta A_1^T A_3 X_3^k - A_1^T y^k + C + P_1 X_1^k \right) \right\}.
\]

(63)

where \( P_{S_{a^n}}(A) = \mathcal{U} \Lambda^+ \mathcal{U}^T \) and \( \Lambda^+ = \text{max}(\Lambda, 0), \quad [\mathcal{U}, \Lambda] = \text{eig}(A) \). For the \(X_2\)-subproblem in (62), by simple calculation, we get its solution is

\[
X_2^{k+1} = P_{S_{a^n}} \left\{ (P_2 + (1 + 2\beta) \cdot I)^{-1} \left( -\beta A_2^T A_1 X_1^{k+1} - \beta A_2^T A_3 X_3^k + A_2^T y^k + C + P_2 X_2^k \right) \right\}.
\]

(64)
The $X_3$-subproblem can be solved explicitly via

$$X_3^{k+1} = P_{SB} \left\{ (P_3 + (1 + 2\beta) \cdot I)^{-1} \left( -\beta \gamma (A_3^T A_1 X_1^{k+1} + A_3^T A_2 X_2^{k+1}) 
\quad + \beta (1 - \gamma) A_3^T A_3 X_3^k + A_3^T y^k + C + P_3 X_3^k) \right) \right\} \quad (65)$$

where $S_B = \{ H \in R^{n \times n} | H_L \preceq H \preceq H_U \}$ and $P_{SB}(A) = \min(\max(H_L, A), H_U)$.

To implement Algorithm 1, we use the stopping criterion

$$\max \left\{ \frac{\|X_1^{k+1} - X_1^k\|}{\|X_1^k - X_0\|}, \frac{\|X_2^{k+1} - X_2^k\|}{\|X_2^k - X_0\|}, \frac{\|X_3^{k+1} - X_3^k\|}{\|X_3^k - X_0\|}, \frac{\|y^{k+1} - y^k\|}{\|y^k - y^0\|} \right\} < \text{tol},$$

where $\text{tol} = 10^{-6}$. We set $P_1 = P_2 = P_3 = \frac{1}{2} I$, $C = \text{rand}(n, n)$, $C = (C' + C) - \text{ones}(n, n) + \text{eye}(n)$, $H_U = \text{ones}(n, n) \cdot 0.1$ and $H_L = -H_U$.

We first test the sensitivity of $\gamma$ for the L-GADMM (6) and take the matrix optimization model (60) with $n = 50$. And we fix $\beta = 1$ and choose different values of $\gamma$ in the interval $(0, 2)$. We repeat each scheme ten times and report some numerical results for the above experiment, and we plot them in Fig. 5.1.

![Fig.1 Sensitivity test on the relaxation factor $\gamma$](image)

From Fig.1, the bigger $\gamma$ often lead to better numerical results since we can see that both the number of iterations and computing time in seconds have decreasing tendency as the relaxation factor $\gamma$ increases. Then, we compare Algorithm 1 with ADMM-G [20], TADMM [32] for their averaged performances.
To further observe the convergence of the three tested algorithms, in Fig. 5.2, we plot the evolutions of the objective function value when the three methods are applied to solve (61).

From Fig. 2, it is easy to see that the proposed algorithm provides faster convergence than that provided by ADMM-G and TADMM; and the figure also shows that the Algorithm 1 has better objective function than that provided by ADMM-G.
Furthermore, we give some numerical results under different dimensions (i.e. \(n=100, 300, 500\)) in Table 1. The data in this table shows that achieving the same level of objective function values, Algorithm 1 with \(\gamma = 1.9\) is faster than Algorithm 1 with \(\gamma = 1\); and both the Algorithm 1 with \(\gamma = 1.9\) and Algorithm 1 with \(\gamma = 1\) are much faster than ADMM-G. Clearly, our proposed method with \(\gamma = 1.9\) are also much faster than TADMM. Moreover, we also give the iterations, computing time in seconds and the epsilon when the stopping criterion is achieved.

According to the above numerical experiment, therefore, the Algorithm 1 is an efficient method for the matrix optimization problem (61).

|          | Alg 1 (\(\gamma = 1\)) | Alg 1 (\(\gamma = 1.9\)) | TADMM  | ADMM-G  |
|----------|-------------------------|--------------------------|--------|---------|
| n=100    |                         |                          |        |         |
| Iterations | 234                    | 110                      | 217    | 905     |
| CPU times(s) | 8.302651               | 4.198581                 | 7.945454 | 33.137704 |
| Objective value | 71.43                 | 71.43                    | 71.43  | 71.69   |
| Epsilon    | 9.988e-007             | 7.122e-007               | 9.915e-007 | 9.989e-007 |
| n=300    |                         |                          |        |         |
| Iterations | 369                    | 196                      | 345    | 448     |
| CPU times(s) | 179.435412             | 100.498216               | 158.495029 | 213.224576 |
| Objective value | 256.90                | 256.90                   | 256.90 | 257.17  |
| Epsilon    | 8.148e-007             | 9.681e-007               | 9.992e-007 | 9.940e-007 |
| n=500    |                         |                          |        |         |
| Iterations | 428                    | 222                      | 395    | 505     |
| CPU times(s) | 856.182117             | 435.490322               | 732.217027 | 941.7561791 |
| Objective value | 430.70                | 430.70                   | 430.70 | 430.91  |
| Epsilon    | 9.958e-007             | 9.821e-007               | 9.957e-007 | 9.998e-007 |

Table 1: Numerical comparison between the averaged performance of Algorithm 1 and TADMM, ADMM-G
6. Conclusions

In this paper, we propose a linearized GADMM to solve multi-block separable convex optimization problem. Under some mild conditions, we give the proof of the global convergence of Algorithm 1 and establish the worst-case $O(1/t)$ convergence rate in both the ergodic and nonergodic senses. Our numerical experiment shows that the proposed method is efficient for solving the matrix optimization problem.

Appendices: manuscript data

Appendix A: Comparison of three algorithms in different dimensions

1. $n=100$

   (1)PGADMM10 scheme $n=100$
   PGADMM10 scheme $k=234$ $eps=9.988e-007$
   Elapsed time is 8.302651 seconds.

   (2)PGADMM11 scheme $n=100$
   PGADMM11 scheme $k=110$ $eps=7.122e-007$
   Elapsed time is 4.198581 seconds.

   (3)TADMM $n=100$
   TADMM $k=217$ $eps=9.915e-007$
   Elapsed time is 7.945454 seconds.

   (4)ADMM with Gaussian back substitution $n=100$
   ADMM with Gaussian back substitution $k=905$ $eps=9.989e-007$
   Elapsed time is 33.137704 seconds.

2. $n=300$
(1) PGADMM10 scheme \( n = 300 \)
PGADMM10 scheme \( k = 369 \) \( \text{eps} = 8.148 \times 10^{-7} \)
Elapsed time is 179.435412 seconds.

(2) PGADMM11 scheme \( n = 300 \)
PGADMM11 scheme \( k = 196 \) \( \text{eps} = 9.681 \times 10^{-7} \)
Elapsed time is 100.498216 seconds.

(3) TADMM \( n = 300 \)
TADMM \( k = 345 \) \( \text{eps} = 9.992 \times 10^{-7} \)
Elapsed time is 158.495029 seconds.

(4) ADMM with Gaussian back substitution \( n = 300 \)
ADMM with Gaussian back substitution \( k = 448 \) \( \text{eps} = 9.940 \times 10^{-7} \)
Elapsed time is 213.224576 seconds.

3. \( n = 500 \)

(1) PGADMM10 scheme \( n = 500 \)
PGADMM10 scheme \( k = 428 \) \( \text{eps} = 9.958 \times 10^{-7} \)
Elapsed time is 856.182117 seconds.

(2) PGADMM11 scheme \( n = 500 \)
PGADMM11 scheme \( k = 222 \) \( \text{eps} = 9.821 \times 10^{-7} \)
Elapsed time is 435.490322 seconds.

(3) TADMM \( n = 500 \)
TADMM \( k = 395 \) \( \text{eps} = 9.957 \times 10^{-7} \)
Elapsed time is 732.217027 seconds.

(4) ADMM with Gaussian back substitution \( n = 500 \)
ADMM with Gaussian back substitution \( k = 505 \) \( \text{eps}=9.998e-007 \)
Elapsed time is 941.756179 seconds.

Appendix B: Relationship between CPU time, number of iterations, and parameter \( \Gamma \)

1. PGADMM1 scheme \( n = 50 \)
PGADMM2 scheme \( k = 755 \) \( \text{eps}=9.999e-007 \)
Elapsed time is 8.285959 seconds.

2. PGADMM2 scheme \( n = 50 \)
PGADMM2 scheme \( k = 393 \) \( \text{eps}=9.999e-007 \)
Elapsed time is 4.141057 seconds.

3. PGADMM3 scheme \( n = 50 \)
PGADMM3 scheme \( k = 265 \) \( \text{eps}=9.958e-007 \)
Elapsed time is 2.756718 seconds.

4. PGADMM scheme \( n = 50 \)
PGADMM4 scheme \( k = 198 \) \( \text{eps}=9.878e-007 \)
Elapsed time is 2.011761 seconds.

5. PGADMM5 scheme \( n = 50 \)
PGADMM5 scheme \( k = 157 \) \( \text{eps}=9.794e-007 \)
Elapsed time is 1.481666 seconds.

6. PGADMM6 scheme \( n = 50 \)
PGADMM6 scheme \( k = 124 \) \( \text{eps}=9.666e-007 \)
Elapsed time is 1.152508 seconds.

7. PGADMM7 scheme \( n = 50 \)
PGADMM7 scheme \( k = 102 \) \( \text{eps}=9.688e-007 \)
Elapsed time is 0.950794 seconds.

8. PGADMM8 scheme n= 50
PGADMM8 scheme k= 87 eps=8.249e-007
Elapsed time is 0.822915 seconds.

9. PGADMM9 scheme n= 50
PGADMM9 scheme k= 74 eps=8.863e-007
Elapsed time is 0.658812 seconds.

References

[1] Tao, M., Yuan, X.M.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM J. Optim. 21(1), 57-81 (2011)

[2] Bardsley, J., Knepper, S., Nagy, J.: Structured linear algebra problems in adaptive optics imaging. Adv. Comput. Math. 35, 103-117 (2011)

[3] Bose, N., Boo, K.: High-resolution image reconstruction with multisensors. Int. J. Imaging Syst. Technol. 9, 294-304 (1998)

[4] Setzer, S., Steidl, G., Tebuber, T.: Deblurring Poissonian images by split Bregman techniques. J. Vis. Commun. Image Represent. 21, 193-199 (2010)

[5] Peaceman, D.H., Rachford, H.H.: The numerical solution of parabolic elliptic differential equations. SIAM J. Appl. Math. 3, 28-41 (1955)

[6] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3, 1-122 (2010)

[7] Candés, E., Li, X., Ma, Y., Wright, J.: Robust principal component analysis J. ACM 58, 1-37 (2011)
[8] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K.: Sparsity and smoothness via the fused LASSO. J. R. Stat. Soc. B 67, 91-108 (2005)

[9] Dinh, Q., Necoara, I., Savorgnan, C., Diehl, M.: An inexact perturbed path-following method for Lagrangian decomposition in large-scale separable convex optimization. SIAM J. Optim. 23, 95-125 (2013)

[10] Eckstein, J.: Augmented Lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results. Tech. Rep. 32-2012, Rutgers University (2012)

[11] Nedić, A., Ozdaglar, A.: Subgradient methods for saddle point problems. J. Optim. Theory Appl. 142, 205-228 (2009)

[12] Dinh, D., Savorgnan, C., Diehl, M.: Combining Lagrangian decomposition and excessive gap smoothing technique for solving large-scale separable convex optimization problems. Comput. Optim. Appl. 55, 75-111 (2013)

[13] Tsiaflakis, P., Diehl, M., Moonen, M.: Distributed spectrum management algorithms for multiuser dsl networks. IEEE Trans. Signal Process. 56, 4825-4843 (2008)

[14] Glowinski, R., Marrocco, A.: Approximation par éléments finis d’ordre un et résolution par pénalisation-dualité d’une classe de problèmes non linéaires. R.A.I.R.O. R2, 41-76 (1975)

[15] D. Gabay, B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximations, Comput. Math. Appl. 2(1976) 17-40.

[16] Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Num. Analy. 16, 964-979 (1979)

[17] Sun, M., Liu, J.: Generalized Peaceman-Rachford splitting method for separable convex programming with applications to image processing, J. Appl. Math. Comput., 51:605-622 (2016)
[18] Sun, M., Liu, J.: A proximal Peaceman-Rachford splitting method for compressive sensing, J. Appl. Math. Comput., 50:349-363 (2016)

[19] He, B.S., Liu, H., Wang, Z.R., Yuan, X.M.: A strictly contractive Peaceman-Rachford splitting method for convex programming. SIAM J. Optim. 24(3), 1011-1040 (2014)

[20] He, B.S., Tao, M., Yuan, X.M.: Alternating direction method with Gaussian back substitution for separable convex programming. SIAM J. Optim. 22, 313-340 (2012)

[21] Cai, X., Gu, G., He, B., Yuan, X.: A proximal point algorithm revisit on alternating direction method of multipliers. Sci. China Math. 56(10), 2179-2186 (2013)

[22] Eckstein, J.: Parallel alternating direction multiplier decomposition of convex programs. J. Optim. Theory Appli. 80(1), 39-62 (1994)

[23] Eckstein, J., Bertsekas, D.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. 55, 293-318 (1992)

[24] He, B., Liao, L.-Z., Han, D.R., Yang, H.: A new inexact alternating directions method for monotone variational inequalities. Math. Program. 92, 103-118 (2002)

[25] Ng, M.K., Wang, F., Yuan, X.: Inexact alternating direction methods for image recovery. SIAM J. Sci. Comput. 33(4), 1643-1668 (2011)

[26] Wang, X., Yuan, X.: The linearized alternating direction method of multipliers for Dantzig Selector. SIAM J. Sci. Comput. 34, 2782-2811 (2012)

[27] Gabay, D.: Applications of the method of multipliers to variational inequalities, Augmented Lagrange Methods: applications to the solution of boundary-valued problems. Fortin, M. Glowinski, R. eds. North Holland pp. 299-331 (1983)
[28] Martinet, B.: Regularisation, d’inquations variationelles par approximations successives. Rev. Francaise d’Inform. Recherche Oper. 4, 154-159 (1970)

[29] Ethan X, Fang, Bingsheng He, Han Liu, Xiaoming Yuan, Generalized alternating direction method of multipliers: new theoretical insights and applications. Math. Prog. Comp. 7, 149-187 (2015)

[30] Zhang, X.Q., Burger, M., Osher, S.: A unified primal-dual algorithm framework based on Bregman iteration. J. Sci. Comput. 6, 20-46 (2010)

[31] Yang, J., Yuan, X.: Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. Math. Comput. 82, 301-329 (2013)

[32] Jin Jiang Wang, Wen Song, An algorithm twisted from generalized ADMM for multi-block separable convex minimization models. J Comput Appl Math. 309, 342-358 (2017)

[33] Bingsheng He, Xiaoming Yuan, Linearized alternating direction method of multipliers with Gaussian back substitution for separable convex programming. Numerical Algebra, Control and Optimization. 3(2), 247-260 (2013)