Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 7876949, 12 pages
https://doi.org/10.1155/2020/7876949

Research Article

A Proximal Alternating Direction Method of Multipliers with a Substitution Procedure

Miantao Chao[1], **Yongxin Zhao**[1] **and Dongying Liang**[2]

[1]*Department of Mathematics, Guangxi University, Nanning 530004, China*
[2]*Guangxi Vocational and Technical College of Communications, Nanning 530023, China*

Correspondence should be addressed to Dongying Liang; liangdy_go@126.com

Received 27 September 2019; Accepted 27 March 2020; Published 27 April 2020

Academic Editor: Jerzy Baranowski

Copyright © 2020 Miantao Chao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we considers the separable convex programming problem with linear constraints. Its objective function is the sum of m individual blocks with nonoverlapping variables and each block consists of two functions: one is smooth convex and the other one is convex. For the general case $m \geq 3$, we present a gradient-based alternating direction method of multipliers with a substitution. For the proposed algorithm, we prove its convergence via the analytic framework of contractive-type methods and derive a worst-case $O(1/t)$ convergence rate in nonergodic sense. Finally, some preliminary numerical results are reported to support the efficiency of the proposed algorithm.

1. Introduction

In this paper, we consider the following convex minimization model with linear constraints and separable objective function:

$$\min \left\{ \sum_{i=1}^{m} [f_i(x_i) + g_i(x_i)] \,\middle|\, \sum_{i=1}^{m} A_i x_i = b,\ x_i \in \mathcal{X}_i,\ i = 1, \ldots, m \right\}, \tag{1}$$

where $f_i: \mathcal{R}^{n_i} \longrightarrow \mathcal{R} \cup \{+\infty\}$ $(i = 1, \ldots, m)$ are closed proper convex functions and $g_i: \mathcal{R}^{n_i} \longrightarrow \mathcal{R}$ $(i = 1, \ldots, m)$ are smooth convex functions, $\mathcal{X}_i \subseteq \mathcal{R}^{n_i}$ $(i = 1, \ldots, m)$ are closed convex sets, $A_i \in \mathcal{R}^{l \times n_i}$ $(i = 1, \ldots, m)$ are given matrices, and $b \in \mathcal{R}^l$ is a given vector. Furthermore, we assume that each g_i has Lipschitz-continuous gradient, i.e., there exists $L_i > 0$ such that

$$\left\| \nabla g_i(x) - \nabla g_i(y) \right\| \leq L_i \|x - y\|, \text{ for all } x, y \in \mathcal{X}_i. \tag{2}$$

Throughout the paper, the solution set of (1) is assumed to be nonempty.

A fundamental method for solving (1) in the case of $m = 2$ is the alternating direction method of multipliers (ADMM), which was presented originally in [1, 2]. We refer to [3, 4] for some review papers on ADMM. There are many problems of form (1) with $m \geq 3$ in contemporary applications, such as the robust principal component analysis model [5], the total variation-based image restoration problem [6], the super-resolution image reconstruction problem [7, 8], the multi-stage stochastic programming problem [9], the deblurring Poissonian images problem [10], the latent variable Gaussian graphical model selection [11], the quadratic discriminant analysis model [12], and the electrical engineering [13, 14]. Then, our discussion focuses on (1) in the case of $m \geq 3$.

A natural idea for solving (1) is to extend the ADMM from the special case $m = 2$ to the general case $m \geq 3$. This straightforward extension can be written as follows:

$$\begin{cases} x_i^{k+1} \in \arg\min_{x_i \in \mathcal{X}_i} \left\{ f_i(x_i) + g_i(x_i) + \frac{1}{2} \left\| \sum_{j=1}^{i-1} A_j x_j^{k+1} + A_i x_i \right. \right. \\ \left. \left. + \sum_{j=i+1}^{m} A_j x_j^k - b - \mu \lambda^k \right\|_H^2 \right\}, \quad i = 1, 2, \ldots, m, \\ \lambda^{k+1} = \lambda^k - H\left(\sum_{j=1}^{m} A_j x_j^{k+1} - b \right). \end{cases} \tag{3}$$

The convergence of (3) is proved in some special cases (see [15–17]). Unfortunately, without further conditions, the direct extension of ADMM (3) for the general case $m \geq 3$ may fail to converge (see [18]). In [19, 20], the authors present two convergent semiproximal ADMM for two types of 3-block problems. Recently, He et al. [21] showed that if a new iterate is generated by correcting the output of (3) with a substitution procedure, then the sequence of iterates converges to a solution of (1). Since then, several variants of the ADMM were proposed for solving (1) (see [21–26]).

In (3), all the x_i-related subproblems are in the form of

$$\min \left\{ f_i(x_i) + g_i(x_i) + \frac{1}{2} \left\| A_i x_i - a_i \right\|_H^2 \mid x_i \in \mathcal{X}_i \right\}, \tag{4}$$

with a certain known $a_i \in \mathcal{R}^l$ and a symmetric positive definite matrix H. When A_i is not the identity matrix, problem (4) becomes complicated. A popular technique is to linearize the quadratic term of (4) (see [27, 28]), that is, one can solve the following problem instead of (4):

$$\min \left\{ f_i(x_i) + g_i(x_i) + \frac{1}{2} \left\| \tau_i x_i - c_i \right\|^2 \mid x_i \in \mathcal{X}_i \right\} \tag{5}$$

with a certain known $c_i \in \mathcal{R}^l$. In general, one can solve the following problem instead of (4):

$$\min \left\{ f_i(x_i) + g_i(x_i) + \frac{1}{2} \left\| A_i x_i - a_i \right\|_H^2 + \frac{1}{2} \left\| x_i - x_i^k \right\|_{G_i}^2 \mid x_i \in \mathcal{X}_i \right\}, \tag{6}$$

where x_i^k is the current iteration. If $G_i = \tau_i I_i - A_i^T H A_i \succ 0$, then (6) becomes the form of (5).

Since g_i is smooth, the following problem is easier than (6):

$$\min \left\{ f_i(x_i) + \langle \nabla g_i(x_i^k), x_i - x_i^k \rangle + \frac{1}{2} \left\| A_i x_i - a_i \right\|_H^2 + \frac{1}{2} \cdot \left\| x_i - x_i^k \right\|_{G_i}^2 \mid x_i \in \mathcal{X}_i \right\}. \tag{7}$$

Now, we can give the gradient-based ADMM (G-ADMM) iterative scheme as follows:

$$\begin{cases} \bar{x}_i^k = \arg\min_{x_i \in \mathcal{X}_i} \left\{ f_i(x_i) + \langle \nabla g_i(x_i^k), x_i - x_i^k \rangle + \frac{1}{2} \left\| \sum_{j=1}^{i-1} A_j \bar{x}_j^k + A_i x_i \right. \right. \\ \left. \left. + \sum_{j=i+1}^{m} A_j x_j^k - b - H^{-1} \lambda^k \right\|_H^2 + \frac{1}{2} \left\| x_i - x_i^k \right\|_{G_i}^2 \right\}, \quad i = 1, \ldots, m, \\ \bar{\lambda}^k = \lambda^k - H\left(\sum_{j=1}^{m} A_j \bar{x}_j^k - b \right). \end{cases} \tag{8}$$

In this paper, imal ADMM with a substitution based on (8). In Section 2, we provide some preliminaries for further analysis. Then, we present the gradient-based alternating direction method of multipliers with a substitution (G-ADMM-S) for solving (1) and its convergence is shown in Section 3. In Section 4, we estimate the worst-case iteration complexity for the proposed algorithm in nonergodic sense. In Section 5, some preliminary numerical results are reported to support the efficiency of the proposed algorithm. Finally, some conclusions are given in Section 6.

2. Preliminaries

In this section, we provide some preliminaries. Let $\langle x, y \rangle = x^T y$ and $\|x\| = \sqrt{\langle x, x \rangle}$. $G \succ 0 \, (\succcurlyeq 0)$ denotes that G is a positive definite (semidefinite) matrix. For any positive definite matrix G, we denote $\|\cdot\|_G$ as the G-norm. If G is the product of a positive parameter β and the identity matrix I, i.e., $G = \beta I$, we use a simpler notation: $\|\cdot\|_G = \|\cdot\|_\beta$. Let $f: \mathcal{R}^n \longrightarrow (-\infty, +\infty]$. The domain of f denoted by $\operatorname{dom} f := \{x \in \mathcal{R}^n \mid f(x) < +\infty\}$. We say that f is convex if

$$f(tx+(1-t)y) \le tf(x)+(1-t)f(y), \quad \forall x, y \in \mathscr{R}^n, \forall t \in [0,1]. \tag{9}$$

For convex function f, the subdifferential of f is the set-valued operator defined by

$$\partial f(x) := \left\{\xi \in \mathscr{R}^n \mid f(y) \ge f(x) + \langle y-x, \xi\rangle, \quad \forall y \in \operatorname{dom} f\right\}. \tag{10}$$

2.1. Variational Characterizations of (1). Let $\Theta_i(x_i) = f_i(x_i) + g_i(x_i)$, $i = 1,2,\ldots,m$, and $\mathscr{W} := \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_m \times \mathscr{R}^l$. Since all $\Theta_i(x_i)$ are convex functions, by invoking the first-order necessary and sufficient condition for convex programming, one can easily find out that problem (1) is characterized by the following variational inequality: we obtain $w^* = (x_1^*, x_2^*, \ldots, x_m^*, \lambda^*) \in \mathscr{W}$ and $\xi_i^* \in \partial f_i(x_i^*) (i = 1,2,\ldots,m)$ such that

$$\begin{aligned} \langle x_i - x_i^*, \xi_i^* + \nabla g_i(x_i^*) - A_i^T\lambda^*\rangle \ge 0, \\ \langle \lambda - \lambda^*, \sum_{i=1}^m A_i x_i^* - b\rangle \ge 0, \end{aligned} \tag{11}$$

for all $(x_1, x_2, \ldots, x_m, \lambda) \in \mathscr{W}$.

The Lagrange function of (1) is given by

$$L(x_1, x_2, \ldots, x_m, \lambda) = \sum_{i=1}^m \Theta_i(x_i) - \lambda^T\left(\sum_{i=1}^m A_i x_i - b\right), \quad (x_1, x_2, \ldots, x_m, \lambda) \in \mathscr{W}. \tag{12}$$

Let $(x_1^*, x_2^*, \ldots, x_m^*, \lambda^*)$ be a saddle point of the Lagrange function $L(x_1, x_2, \ldots, x_m, \lambda)$. That is, for any $\lambda \in \mathscr{R}^l$ and $x_i \in \mathscr{X}_i$ $(i = 1,2,\ldots,m)$,

$$L(x_1^*, x_2^*, \ldots, x_m^*, \lambda) \le L(x_1^*, x_2^*, \ldots, x_m^*, \lambda^*) \le L(x_1, x_2, \ldots, x_m, \lambda^*). \tag{13}$$

Finding a saddle point of $L(x_1, x_2, \ldots, x_m, \lambda)$ is equivalent to finding a $w^* = (x_1^*, x_2^*, \ldots, x_m^*, \lambda^*) \in \mathscr{W}$ such that

$$\begin{cases} \Theta_1(x_1) - \Theta_1(x_1^*) + (x_1 - x_1^*)^T(-A_1^T\lambda^*) \ge 0, & \forall x_1 \in \mathscr{X}_1, \\ \Theta_2(x_2) - \Theta_2(x_2^*) + (x_2 - x_2^*)^T(-A_2^T\lambda^*) \ge 0, & \forall x_2 \in \mathscr{X}_2, \\ \cdots \\ \Theta_m(x_m) - \Theta_m(x_m^*) + (x_m - x_m^*)^T(-A_m^T\lambda^*) \ge 0, & \forall x_m \in \mathscr{X}_m, \\ (\lambda - \lambda^*)^T\left(\sum_{i=1}^m A_i x_i^* - b\right) \ge 0, & \forall \lambda \in \mathscr{R}^l. \end{cases} \tag{14}$$

Let $x = (x_1, x_2, \ldots, x_m)^T$, $w = (x_1, x_2, \ldots, x_m, \lambda)^T$, $\Theta(x) = \sum_{i=1}^m \Theta_i(x_i)$, and

$$G(w) = \left(-A_1^T\lambda, -A_2^T\lambda, \ldots, -A_m^T\lambda, \sum_{i=1}^m A_i x_i - b\right)^T. \tag{15}$$

Then, (14) can be rewritten as the following variational inequality (VI): we obtain $w^* = (x_1^*, x_2^*, \ldots, x_m^*, \lambda^*) \in \mathscr{W}$ such that

$$\mathrm{VI}(\mathscr{W}, G, \Theta)\colon \Theta(x) - \Theta(x^*) + (w - w^*)^T G(w^*) \ge 0, \quad \forall w \in \mathscr{W}. \tag{16}$$

Let $\mathscr{W}^*$ be the solution set of $\mathrm{VI}(\mathscr{W}, G, \Theta)$. Since we have assumed that the solution set of (1) is nonempty, $\mathscr{W}^*$ is also nonempty. It follows from the definition of $G(w)$ that

$$(w' - w'')^T G(w') = (w' - w'')^T G(w''), \quad \forall w', w'' \in \mathscr{W}. \tag{17}$$

2.2. Some Notations. Let $x^k = (x_1^k, x_2^k, \ldots, x_m^k)^T$, $\overline{x}^k = (\overline{x}_1^k, \overline{x}_2^k, \ldots, \overline{x}_m^k)^T$, $w^k = (x_1^k, \ldots, x_m^k, \lambda^k)^T$, $\overline{w}^k = (\overline{x}_1^k, \ldots, \overline{x}_m^k, \overline{\lambda}^k)^T$, and $v = (x_2, \ldots, x_m, \lambda)^T$. Let H and G_i $(i = 1,2,\ldots,m)$ be given positive definite matrices, $A = (A_1, A_2, \ldots, A_m)$. $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of one matrix, and $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of one matrix. The following notions will be used in the later analysis:

$$M = \begin{pmatrix} A_2^T H A_2 & 0 & \cdots & 0 \\ A_3^T H A_2 & A_3^T H A_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A_m^T H A_2 & A_m^T H A_3 & \cdots & A_m^T H A_m \end{pmatrix},$$

$$\overline{M} = \begin{pmatrix} A_2^T H A_2 & 0 & \cdots & 0 & 0 \\ A_3^T H A_2 & A_3^T H A_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_m^T H A_2 & A_m^T H A_3 & \cdots & A_m^T H A_m & 0 \\ 0 & 0 & \cdots & 0 & H^{-1} \end{pmatrix},$$

$$Q = \begin{pmatrix} A_2^T H A_2 & A_2^T H A_3 & \cdots & A_2^T H A_m & A_2^T \\ A_3^T H A_2 & A_3^T H A_3 & \cdots & A_3^T H A_m & A_3^T \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_m^T H A_2 & A_m^T H A_3 & \cdots & A_m^T H A_m & A_m^T \\ A_2 & A_3 & \cdots & A_m & H^{-1} \end{pmatrix}, \tag{18}$$

$$P = \operatorname{diag}\left(A_2^T H A_2, A_3^T H A_3, \ldots, A_m^T H A_m, H^{-1}\right),$$

$$D_k = \begin{pmatrix} G_1\left(x_1^k - \overline{x}_1^k\right) + \nabla g_1\left(\overline{x}_1^k\right) - \nabla g_1\left(x_1^k\right) \\ \cdots \\ G_i\left(x_i^k - \overline{x}_i^k\right) + \nabla g_i\left(\overline{x}_i^k\right) - \nabla g_i\left(x_i^k\right) + A_i^T H \sum_{j=2}^{i} A_j\left(x_j^k - \overline{x}_j^k\right) \\ \cdots \\ G_m\left(x_m^k - \overline{x}_m^k\right) + \nabla g_m\left(\overline{x}_m^k\right) - \nabla g_m\left(x_m^k\right) + A_m^T H \sum_{j=2}^{m} A_j\left(x_j^k - \overline{x}_j^k\right) \\ H^{-1}\left(\lambda^k - \overline{\lambda}^k\right) \end{pmatrix},$$

$$b_k = \left(w^k - \overline{w}^k\right)^T D_k + \left(\lambda^k - \overline{\lambda}^k\right)^T \sum_{j=2}^{m} A_j\left(x_j^k - \overline{x}_j^k\right). \tag{19}$$

It is easy to see that

$$Q = \begin{pmatrix} A_2^T H^{1/2} \\ A_3^T H^{1/2} \\ \vdots \\ A_m^T H^{1/2} \\ H^{-1/2} \end{pmatrix} \left(A_2^T H^{\frac{1}{2}}, A_3^T H^{\frac{1}{2}}, \ldots, A_m^T H^{\frac{1}{2}}, H^{-\frac{1}{2}}\right) \succcurlyeq 0. \tag{20}$$

3. Algorithm and Convergence Analysis

In this section, we first describe G-ADMM-S and then prove its convergence via the analytic framework of the contractive-type method [29]. Throughout this section, we assume that $\lambda_{min}(G_i) > L_i$ $(i = 1, 2, \ldots, m)$. We propose the iterative scheme of G-ADMM-S for solving (1) in Algorithm G-ADMM-S:

Let $\gamma \in (0, 2) D_k$ and b_k be defined in (18) and (19), respectively. Start with w^0. With the given iterate w^k, the new iterate w^{k+1} is given as follows:

Step 1 (G-ADMM procedure). Execute scheme (8) to generate $\overline{w}^k$.

Step 2 (substitution procedure). Generate the new iterate w^{k+1} via

$$w^{k+1} = w^k - \alpha^k D_k, \tag{21}$$

where

$$\alpha^k = \gamma \alpha_k \quad \text{with} \quad \alpha_k = \frac{b_k}{\left\|D_k\right\|^2}. \tag{22}$$

Next, we establish the global convergence of Algorithm G-ADMM-S following the analytic framework of contractive-type methods. We outline the proof sketch as follows:

(1) Prove that $-D_k$ is a descent direction of the function $(1/2)\|w - w^*\|^2$ at the point $w = w^k$ whenever

$w^k \neq \overline{w}^k$, where $\overline{w}^k$ is generated by G-ADMM scheme (8) and $w^* \in \mathscr{W}^*$

(2) Prove that the sequence generated by Algorithm G-ADMM-S is contractive with respect to $\mathscr{W}^*$

(3) Establish the convergence

Accordingly, we divide the convergence analysis into three sections to address the claims listed above.

3.1. Verification of the Descent Direction. In this section, we show that $-D_k$ is a descent direction of the function $(1/2)\|w - w^*\|^2$ at the point $w = w^k$ whenever $w^k \neq \overline{w}^k$ and $w^* \in \mathscr{W}^*$. For this purpose, we first prove an important inequality for the output of G-ADMM procedure (8), which will be used often in our further discussion.

Theorem 1. *$\overline{w}^k \in \mathscr{W}$ and*

$$\Theta(x) - \Theta(\overline{x}^k) + (w - \overline{w}^k)^T G(\overline{w}^k) + P_k \geq (w - \overline{w}^k)^T D_k, \quad \forall w \in \mathscr{W}, \tag{23}$$

where $P_k = (w - \overline{w}^k)^T \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \\ 0 \end{pmatrix} H \sum_{j=2}^m A_j (x_j^k - \overline{x}_j^k)$.

Proof. By the optimality condition of the x_i-related subproblem in (8), for $i = 1, 2, \ldots, m$, we have $\overline{x}_i^k \in \mathscr{X}_i$ and

$$0 \in \partial f_i(\overline{x}_i^k) + \nabla g_i(x_i^k) + A_i^T H\left(\sum_{j=1}^{i} A_j \overline{x}_j^k + \sum_{j=i+1}^{m} A_j x_j^k - b - H^{-1}\lambda^k\right) + G_i(\overline{x}_i^k - x_i^k) + \partial\delta(\overline{x}_i^k \mid \mathscr{X}_i), \tag{24}$$

where $\delta(\mathscr{X}_i)$ is the indicator function of the set $\mathscr{X}_i$. Thus, $\overline{w}^k \in \mathscr{W}$ and there exists $\eta \in \partial\delta(\mathscr{X}_i)(\overline{x}_i^k)$ such that

$$-\eta \in \partial f_i(\overline{x}_i^k) + \nabla g_i(\overline{x}_i^k) + \rho = \partial\Theta_i(\overline{x}_i^k) + \rho, \tag{25}$$

where $\rho := \nabla g_i(x_i^k) - \nabla g_i(\overline{x}_i^k) + A_i^T H(\sum_{j=1}^i A_j \overline{x}_j^k + \sum_{j=i+1}^m A_j x_j^k - b - H^{-1}\lambda^k) + G_i(\overline{x}_i^k - x_i^k)$. From the subgradient inequality, one has

$$\begin{aligned} \Theta_i(x_i) - \Theta_i(\overline{x}_i^k) &\geq (x_i - \overline{x}_i^k)^T (-\eta - \rho), \quad \forall x_i \in \mathscr{X}_i, \\ \Theta_i(x_i) - \Theta_i(\overline{x}_i^k) + (x_i - \overline{x}_i^k)^T \rho &\geq -(x_i - \overline{x}_i^k)^T \eta, \quad \forall x_i \in \mathscr{X}_i. \end{aligned} \tag{26}$$

From the definition of $\partial\delta(\mathscr{X}_i)(\overline{x}_i^k)$, one has

$$\Theta_i(x_i) - \Theta_i(\overline{x}_i^k) + (x_i - \overline{x}_i^k)^T \rho \geq 0, \quad \forall x_i \in \mathscr{X}_i. \tag{27}$$

That is,

$$\begin{aligned} 0 \leq \Theta_i(x_i) - \Theta_i(\overline{x}_i^k) + (x_i - \overline{x}_i^k)^T \Bigg\{ &\nabla g_i(x_i^k) - \nabla g_i(\overline{x}_i^k) + A_i^T H\left(\sum_{j=1}^{i} A_j \overline{x}_j^k \right. \\ &\left. + \sum_{j=i+1}^{m} A_j x_j^k - b - H^{-1}\lambda^k\right) + G_i(\overline{x}_i^k - x_i^k) \Bigg\}, \end{aligned} \tag{28}$$

for all $x_i \in \mathscr{X}_i$. Substituting $\overline{\lambda}^k = \lambda^k - H(\sum_{j=1}^m A_j \overline{x}_j^k - b)$ (see (8)) in the above inequality, we obtain

$$\begin{aligned} 0 \leq \Theta_i(x_i) - \Theta_i(\overline{x}_i^k) + (x_i - \overline{x}_i^k)^T \Bigg\{ &\nabla g_i(x_i^k) - \nabla g_i(\overline{x}_i^k) \\ &- A_i^T \overline{\lambda}^k + A_i^T H \sum_{j=i+1}^{m} A_j(x_j^k - \overline{x}_j^k) + G_i(\overline{x}_i^k - x_i^k) \Bigg\}, \end{aligned} \tag{29}$$

for all $x_i \in \mathscr{X}_i$. Summing the above inequality over $i = 1, 2, \ldots, m$, we obtain

$$\Theta(x) - \Theta(\overline{x}^k) + (x - \overline{x}^k)^T (-A^T \overline{\lambda}^k) C_k \geq 0, \tag{30}$$

where

$$C_k = \begin{pmatrix} G_1(\overline{x}_1^k - x_1^k) + \nabla g_1(x_1^k) - \nabla g_1(\overline{x}_1^k) + A_1^T H \sum_{j=2}^{m} A_j(x_j^k - \overline{x}_j^k) \\ \cdots \\ G_i(\overline{x}_i^k - x_i^k) + \nabla g_i(x_i^k) - \nabla g_i(\overline{x}_i^k) + A_i^T H \sum_{j=i+1}^{m} A_j(x_j^k - \overline{x}_j^k) \\ \cdots \\ G_m(\overline{x}_m^k - x_m^k) + \nabla g_m(x_m^k) - \nabla g_m(\overline{x}_m^k) \end{pmatrix}. \tag{31}$$

Then, by adding the following term

$$\left(x-\overline{x}^k\right)^T\begin{pmatrix} G_1\left(x_1^k-\overline{x}_1^k\right)+\nabla g_1\left(\overline{x}_1^k\right)-\nabla g_1\left(x_1^k\right) \\ \cdots \\ G_i\left(x_i^k-\overline{x}_i^k\right)+\nabla g_i\left(\overline{x}_i^k\right)-\nabla g_i\left(x_i^k\right)+A_i^T H\sum_{j=2}^{i} A_j\left(x_j^k-\overline{x}_j^k\right) \\ \cdots \\ G_m\left(x_m^k-\overline{x}_m^k\right)+\nabla g_m\left(\overline{x}_m^k\right)-\nabla g_m\left(x_m^k\right)+A_m^T H\sum_{j=2}^{m} A_j\left(x_j^k-\overline{x}_j^k\right) \end{pmatrix} \tag{32}$$

to both sides of (30), we get

$$\begin{aligned} &\Theta(x)-\Theta\left(\overline{x}^k\right)+\left(x-\overline{x}^k\right)^T\left\{\left(-A^T\overline{\lambda}^k\right)+\begin{pmatrix} A_1^T H\sum_{j=2}^{m} A_j\left(x_j^k-\overline{x}_j^k\right) \\ \cdots \\ A_i^T H\sum_{j=2}^{m} A_j\left(x_j^k-\overline{x}_j^k\right) \\ \cdots \\ A_m^T H\sum_{j=2}^{m} A_j\left(x_j^k-\overline{x}_j^k\right) \end{pmatrix}\right\}\geq \\ &\left(x-\overline{x}^k\right)^T\begin{pmatrix} G_1\left(x_1^k-\overline{x}_1^k\right)+\nabla g_1\left(\overline{x}_1^k\right)-\nabla g_1\left(x_1^k\right) \\ \cdots \\ G_i\left(x_i^k-\overline{x}_i^k\right)+\nabla g_i\left(\overline{x}_i^k\right)-\nabla g_i\left(x_i^k\right)+A_i^T H\sum_{j=2}^{i} A_j\left(x_j^k-\overline{x}_j^k\right) \\ \cdots \\ G_m\left(x_m^k-\overline{x}_m^k\right)+\nabla g_m\left(\overline{x}_m^k\right)-\nabla g_m\left(x_m^k\right)+A_m^T H\sum_{j=2}^{m} A_j\left(x_j^k-\overline{x}_j^k\right) \end{pmatrix}. \end{aligned} \tag{33}$$

Since $\overline{\lambda}^k=\lambda^k-H(\sum_{j=1}^m A_j\overline{x}_j^k-b)$, we have

$$\left(\sum_{j=1}^{m} A_j\overline{x}_j^k-b\right)+H^{-1}\left(\overline{\lambda}^k-\lambda^k\right)=0. \tag{34}$$

Combining the above two formulas, we have

$$\Theta(x)-\Theta\left(\overline{x}^k\right)+\left(w-\overline{w}^k\right)^T\left(G\left(\overline{w}^k\right)+H_k\right)\geq\left(w-\overline{w}^k\right)^T Q_k,\quad \forall w\in\mathcal{W}, \tag{35}$$

where

$$\begin{aligned} H_k&=\left(A_1^T H\sum_{j=2}^{m} A_j\left(x_j^k-\overline{x}_j^k\right),A_2^T H\sum_{j=2}^{m} A_j\left(x_j^k-\overline{x}_j^k\right),\ldots,A_m^T H\sum_{j=2}^{m} A_j\left(x_j^k-\overline{x}_j^k\right),0\right)^T, \\ Q_k&=\begin{pmatrix} G_1\left(x_1^k-\overline{x}_1^k\right)+\nabla g_1\left(\overline{x}_1^k\right)-\nabla g_1\left(x_1^k\right) \\ \cdots \\ G_i\left(x_i^k-\overline{x}_i^k\right)+\nabla g_i\left(\overline{x}_i^k\right)-\nabla g_i\left(x_i^k\right)+A_i^T H\sum_{j=2}^{i} A_j\left(x_j^k-\overline{x}_j^k\right) \\ \cdots \\ G_m\left(x_m^k-\overline{x}_m^k\right)+\nabla g_m\left(\overline{x}_m^k\right)-\nabla g_m\left(x_m^k\right)+A_m^T H\sum_{j=2}^{m} A_j\left(x_j^k-\overline{x}_j^k\right) \\ H^{-1}\left(\lambda^k-\overline{\lambda}^k\right) \end{pmatrix}. \end{aligned} \tag{36}$$

Using the notations of $G(\overline{w}^k)$ (see (15)) and D_k (see (18)), assertion (23) is proved. □

Based on assertion (23), we can get the following result.

Corollary 1.

$$\left(\overline{w}^k - w^*\right)^T D_k \geq \left(\lambda^k - \overline{\lambda}^k\right)^T \sum_{j=2}^{m} A_j\left(x_j^k - \overline{x}_j^k\right), \quad \forall w^* \in \mathcal{W}^*. \tag{37}$$

Proof. It follows from (23) that

$$\begin{aligned}\left(\overline{w}^k - w^*\right)^T D_k \geq{}& \Theta\left(\overline{x}^k\right) - \Theta\left(x^*\right) + \left(\overline{w}^k - w^*\right)^T G\left(\overline{w}^k\right) \\ &+ \left(\overline{w}^k - w^*\right)^T \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \end{pmatrix} H \sum_{j=2}^{m} A_j\left(x_j^k - \overline{x}_j^k\right).\end{aligned} \tag{38}$$

Using (17) and the optimality of w^*, we have

$$\begin{aligned}\Theta\left(\overline{x}^k\right) - \Theta\left(x^*\right) + \left(\overline{w}^k - w^*\right)^T G\left(\overline{w}^k\right) ={}& \Theta\left(\overline{x}^k\right) - \Theta\left(x^*\right) \\ &+ \left(\overline{w}^k - w^*\right)^T G\left(w^*\right) \geq 0.\end{aligned} \tag{39}$$

Thus,

$$\left(\overline{w}^k - w^*\right)^T D_k \geq \left[H\left(\sum_{j=1}^{m} A_j\overline{x}_j^k - \sum_{j=1}^{m} A_j x_j^*\right)\right]^T \sum_{j=2}^{m} A_j\left(x_j^k - \overline{x}_j^k\right). \tag{40}$$

Since $\sum_{j=1}^{m} A_j x_j^* = b$ and $H(\sum_{j=1}^{m}(A_j\overline{x}_j^k - b)) = \lambda^k - \overline{\lambda}^k$,

$$\left(\overline{w}^k - w^*\right)^T D_k \geq \left(\lambda^k - \overline{\lambda}^k\right)^T \sum_{j=2}^{m} A_j\left(x_j^k - \overline{x}_j^k\right). \tag{41}$$

The next theorem implies that $-D_k$ is a descent direction of the function $(1/2)\|w - w^*\|^2$ at the point $w = w^k$ whenever $w^k \neq \overline{w}^k$. □

Theorem 2. *For all* $w^* \in \mathcal{W}^*$,

$$\begin{aligned}\left(w^k - w^*\right)^T D_k &\geq b_k \\ &\geq \sum_{j=1}^{m}\left(\left\|x_j^k - \overline{x}_j^k\right\|_{G_j}^2 - L_i\left\|x_j^k - \overline{x}_j^k\right\|\right) + \frac{1}{2}\left\|v^k - \overline{v}^k\right\|_{(P+Q)}^2 \\ &\geq \sum_{j=1}^{m}\left(\lambda_{\min}\left(G_j\right) - L_i\right)\left\|x_j^k - \overline{x}_j^k\right\|^2 + \frac{1}{2}\left\|\lambda^k - \overline{\lambda}^k\right\|_{H^{-1}}^2.\end{aligned} \tag{42}$$

Proof. It follows from (37) that

$$\left(w^k - w^*\right)^T D_k \geq \left(w^k - \overline{w}^k\right)^T D_k + \left(\lambda^k - \overline{\lambda}^k\right)^T \sum_{j=2}^{m} A_j\left(x_j^k - \overline{x}_j^k\right). \tag{43}$$

That is, $(w^k - w^*)^T D_k \geq b_k$. Now, we treat the first term of the right-hand side of (43):

$$\begin{aligned}\left(w^k - \overline{w}^k\right)^T D_k ={}& \sum_{i=1}^{m}\left[\left\|x_i^k - \overline{x}_i^k\right\|_{G_i}^2 + \left(x_i^k - \overline{x}_i^k\right)^T\left(\nabla g_i\left(\overline{x}_i^k\right) - \nabla g_i\left(x_i^k\right)\right)\right] \\ &+ \left(v^k - \overline{v}^k\right)^T \overline{M}\left(v^k - \overline{v}^k\right) \\ \geq{}& \sum_{i=1}^{m}\left(\left\|x_i^k - \overline{x}_i^k\right\|_{G_i}^2 - L_i\left\|x_j^k - \overline{x}_j^k\right\|\right) + \left(v^k - \overline{v}^k\right)^T \overline{M}\left(v^k - \overline{v}^k\right) \\ \geq{}& \sum_{i=1}^{m}\left(\lambda_{\min}\left(G_i\right) - L_i\right)\left\|x_i^k - \overline{x}_i^k\right\|^2 + \left(v^k - \overline{v}^k\right)^T \overline{M}\left(v^k - \overline{v}^k\right),\end{aligned} \tag{44}$$

where the first inequality follows from the Lipschitz continuous of ∇g_i. Then, let us deal with the second term of the right-hand side of (43):

$$\left(\lambda^k - \overline{\lambda}^k\right)^T \sum_{j=2}^{m} A_j\left(x_j^k - \overline{x}_j^k\right) = \left(v^k - \overline{v}^k\right)^T \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ A_2 & A_3 & \cdots & A_m & 0 \end{pmatrix} \left(v^k - \overline{v}^k\right). \tag{45}$$

Thus,

$$\left(v^k-\tilde{v}^k\right)^T\overline{M}\left(v^k-\tilde{v}^k\right)+\left(\lambda^k-\tilde{\lambda}^k\right)^T\sum_{j=2}^{m}A_j\left(x_j^k-\tilde{x}_j^k\right)$$
$$=\left(v^k-\tilde{v}^k\right)^T M_1\left(v^k-\tilde{v}^k\right)$$
$$=\frac{1}{2}\left(v^k-\tilde{v}^k\right)^T M_2\left(v^k-\tilde{v}^k\right) \tag{46}$$
$$=\frac{1}{2}\left\|v^k-\tilde{v}^k\right\|_P^2+\frac{1}{2}\left\|v^k-\tilde{v}^k\right\|_Q^2$$
$$\geq\frac{1}{2}\left\|\lambda^k-\tilde{\lambda}^k\right\|_{H^{-1}}^2,$$

where

$$M_1=\begin{pmatrix} A_2^THA_2 & 0 & \cdots & 0 & 0\\ A_3^THA_2 & A_3^THA_3 & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ A_m^THA_2 & A_m^THA_3 & \cdots & A_m^THA_m & 0\\ A_2 & A_3 & \cdots & A_m & H^{-1}\end{pmatrix},$$
$$M_2=\begin{pmatrix} 2A_2^THA_2 & A_2^THA_3 & \cdots & A_2^THA_m & A_2^k\\ A_3^THA_2 & 2A_3^THA_3 & \cdots & A_3^THA_m & A_3^k\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ A_m^THA_2 & A_m^THA_3 & \cdots & 2A_m^THA_m & A_m^k\\ A_2 & A_3 & \cdots & A_m & 2H^{-1}\end{pmatrix}. \tag{47}$$

□

The assertion follows from the above three formulas.

Since $H\succ 0$ and $\lambda_{\min}(G_i)>L_i$, whenever $w^k-\overline{w}^k\neq 0$, assertion (42) shows the positivity of the term $(w^k-w^*)^TD_k$, and thus, the direction $-D_k$ is a descent direction of the function $(1/2)\|w-w^*\|^2$ at the point $w=w^k$.

3.2. Contractive Property. In this section, we show that the sequence $\{w^k\}$ generated by Algorithm G-ADMM-S is contractive with respect to the set $\mathscr{W}^*$.

Since the direction $-D_k$ is a descent direction of the function $(1/2)\|w-w^*\|^2$ at the point $w=w^k$, the new iterate w^{k+1} can be generated by

$$w^{k+1}=w^k-\alpha D_k. \tag{48}$$

Thus,

$$\begin{aligned}\left\|w^k-w^*\right\|^2-\left\|w^{k+1}-w^*\right\|^2&=\left\|w^k-w^*\right\|^2-\left\|\left(w^k-w^*\right)-\alpha D_k\right\|^2\\&=2\alpha\left(w^k-w^*\right)^TD_k-\alpha^2\left\|D_k\right\|^2\\&\geq 2\alpha b_k-\alpha^2\left\|D_k\right\|^2,\quad\forall w^*\in\mathscr{W}^*,\end{aligned} \tag{49}$$

where the inequality follows from the first inequality of (42).

Let $q(\alpha)=2\alpha b_k-\alpha^2\|D_k\|^2$. Note that right-hand side of (49), i.e., $q(\alpha)$, is a quadratic function of α, and

$$q(\alpha)>0\Longleftrightarrow 0<\alpha<\frac{2b_k}{\left\|D_k\right\|^2}. \tag{50}$$

In order to obtain the closest proximity to $\mathscr{W}^*$, we are in the desire to maximize this quadratic function and this promotes us to take the optimal value of α as

$$\alpha=\alpha_k:=\frac{b_k}{\left\|D_k\right\|^2}. \tag{51}$$

With this choice of step size, it follows from (49) that

$$\left\|w^k-w^*\right\|^2\leq\left\|w^k-w^*\right\|^2-\frac{b_k^2}{\left\|D_k\right\|^2},\quad\forall w^*\in\mathscr{W}^*. \tag{52}$$

Let

$$\begin{aligned}T_1&=\max\left\{\lambda_{\max}\left(H^{-1}\right),\lambda_{\max}(G_i)+L_i,1\leq i\leq m\right\}\\&\quad+\sqrt{\lambda_{\max}(M^TM)}>0,\\T_2&=\min\left\{\frac{\lambda_{\min}(H^{-1})}{2},\lambda_{\min}(G_i)-L_i,1\leq i\leq m\right\}>0.\end{aligned} \tag{53}$$

From the definition of D_k (see, (18)), one has

$$\begin{aligned}\left\|D_k\right\|&\leq\left\|\overline{G}_k\right\|+\left\|M\left(v^k-\tilde{v}^k\right)\right\|\\&\leq\Big[\max\left\{\lambda_{\max}\left(H^{-1}\right),\lambda_{\max}(G_i)+L_i,1\leq i\leq m\right\}\\&\quad+\sqrt{\lambda_{\max}(M^TM)}\Big]\left\|w^k-\overline{w}^k\right\|\\&\leq T_1\left\|w^k-\overline{w}^k\right\|,\end{aligned} \tag{54}$$

where

$$\overline{G}_k=\begin{pmatrix} G_1\left(x_1^k-\overline{x}_1^k\right)+\nabla g_1\left(\overline{x}_1^k\right)-\nabla g_1\left(x_1^k\right)\\ G_2\left(x_2^k-\overline{x}_2^k\right)+\nabla g_2\left(\overline{x}_2^k\right)-\nabla g_2\left(x_2^k\right)\\ \cdots\\ G_m\left(x_m^k-\overline{x}_m^k\right)+\nabla g_m\left(\overline{x}_m^k\right)-\nabla g_m\left(x_m^k\right)\\ H^{-1}\left(\lambda^k-\overline{\lambda}^k\right)\end{pmatrix}. \tag{55}$$

It follows from (19) and (42) that

$$\begin{aligned}b_k&\geq\min\left\{\frac{\lambda_{\min}(H^{-1})}{2},\lambda_{\min}(G_i)-L_i,1\leq i\leq m\right\}\left\|w^k-\overline{w}^k\right\|^2\\&\geq T_2\left\|w^k-\overline{w}^k\right\|^2.\end{aligned} \tag{56}$$

It is easy to see that $\alpha_k\geq(T_2/T_1^2)$.

Next, we show that the sequence $\{w^k\}$ generated by Algorithm G-ADMM-S is contractive with respect to the set $\mathscr{W}^*$.

Theorem 3. *Let the sequence $\{w^k\}$ be generated by the proposed Algorithm G-ADMM-S. Then,*

$$\left\|w^{k+1} - w^*\right\|^2 \le \left\|w^k - w^*\right\|^2 - \frac{\gamma(2-\gamma)T_2^2}{T_1^2}\left\|\overline{w}^k - w^k\right\|^2, \quad \forall w^* \in \mathcal{W}^*. \tag{57}$$

Proof. Using (21) and (22), we obtain

$$\begin{aligned}\left\|w^{k+1} - w^*\right\|^2 &= \left\|w^k - w^*\right\|^2 - 2\gamma\alpha_k\left(w^k - w^*\right)^T D_k \\ &\quad + (\gamma\alpha_k)^2\left\|D_k\right\|^2 \\ &\le \left\|w^k - w^*\right\|^2 - \left[2\gamma\alpha_k b_k - (\gamma\alpha_k)^2\left\|D_k\right\|^2\right] \\ &= \left\|w^k - w^*\right\|^2 - \gamma(2-\gamma)\alpha_k b_k \\ &= \left\|w^k - w^*\right\|^2 - \frac{\gamma(2-\gamma)b_k^2}{\left\|D_k\right\|^2} \\ &\le \left\|w^k - w^*\right\|^2 - \frac{\gamma(2-\gamma)T_2^2}{T_1^2}\left\|\overline{w}^k - w^k\right\|^2,\end{aligned} \tag{58}$$

where the first inequality follows from the first inequality of (42) and the second inequality follows from (54)–(56). □

3.3. Convergence Result. In this section, we establish the global convergence for Algorithm G-ADMM-S based on the analytic framework of contraction methods in [29].

Theorem 4(global convergence). *Let the sequence $\{w^k\}$ be generated by Algorithm G-ADMM-S. Then, there exists $w^\infty \in \mathcal{W}^*$ such that*

$$\lim_{k\longrightarrow\infty} w^k = w^\infty. \tag{59}$$

Proof. It follows from (57) that the sequence $\{w^k\}$ is bounded and

$$\gamma(2-\gamma)\sum_{k=0}^{\infty}\frac{T_2^2}{T_1^2}\left\|\overline{w}^k - w^k\right\|^2 \le \left\|w^0 - w^*\right\|^2, \tag{60}$$

which implies that $\lim_{k\longrightarrow\infty}\|\overline{w}^k - w^k\| = 0$.

Since $\{w^k\}$ is bounded, the sequence $\{w^k\}$ has at least one cluster point and we denote it by $w^\infty = (x_1^\infty, x_2^\infty, \ldots, x_m^\infty, \lambda^\infty)$. In addition, let $\{w^{k_j}\}$ be the subsequence converging to w^∞. Since $\lim_{k\longrightarrow\infty}\|w^k - \overline{w}^k\| = 0$, $\{\overline{w}^{k_j}\}$ converges to w^∞.

By taking the limit over k_j in (15), we have that

$$\Theta(x) - \Theta(x^\infty) + (w - w^\infty)G(w) \ge 0, \quad \forall w \in \mathcal{W}. \tag{61}$$

Therefore, w^∞ is a solution point of $VI(\mathcal{W}, G, \Theta)$. By using (57), we have

$$\left\|w^{k+1} - w^\infty\right\|^2 \le \left\|w^k - w^\infty\right\|^2 - \frac{\gamma(2-\gamma)T_2^2}{T_1^2}\left\|\overline{w}^k - w^k\right\|^2, \tag{62}$$

and thus, $\lim_{k\longrightarrow\infty} w^k = w^\infty$. □

4. Iteration Complexity

In this section, we will show that after t iterations of Algorithm G-ADMM-S, we can ensure that

$$\min_{0\le k\le t}\left\{\left\|w^k - w^k\right\|^2\right\} \le \varepsilon, \tag{63}$$

where $\varepsilon = O(1/t)$. Thus, a worst-case $O(1/\varepsilon)$ iteration complexity is established in nonergodic sense for Algorithm G-ADMM-S.

Theorem 5. *Let the sequence $\{w^k\}$ be generated by Algorithm G-ADMM-S. Then,*

$$\min_{0\le k\le t}\left\{\left\|\overline{w}^k - w^k\right\|^2\right\} \le \frac{1}{c(t+1)}\left\|w^0 - w^*\right\|^2, \quad \forall w^* \in \mathcal{W}^*, \tag{64}$$

where $c(\gamma(2-\gamma)T_2^2/T_1^2)$.

Proof. It follows from (57) that

$$c\sum_{k=0}^{\infty}\left\|\overline{w}^k - w^k\right\|^2 \le \left\|w^0 - w^*\right\|^2, \quad \forall w^* \in \mathcal{W}^*. \tag{65}$$

Thus, for any integer $t > 0$, we obtain

$$c\sum_{k=0}^{t}\left\|\overline{w}^k - w^k\right\|^2 \le \left\|w^0 - w^*\right\|^2, \quad \forall w^* \in \mathcal{W}^*, \tag{66}$$

and consequently, we obtain assertion (64). □

Recall that $\mathcal{W}^*$ is convex and closed under our assumptions (see, Theorem 2.3.5 in [30]). Let

$$d := \inf\left\{\left\|w^0 - w^*\right\|^2 \,\middle|\, w^* \in \mathcal{W}^*\right\}. \tag{67}$$

For any given $\varepsilon > 0$, inequality (64) indicates that Algorithm G-ADMM-S requires at most $\lfloor d/(c\varepsilon)\rfloor$ iterations to fulfill the requirement $\|\overline{w}^k - w^k\|^2 \le \varepsilon$.

5. Numerical Results

To investigate the numerical performance of the proposed algorithm, we apply it to solve a convex quadratic programming and a nonlinear convex programming with separable structure and report some preliminary numerical results. All codes were written by Matlab 2016a, and all the numerical experiments were conducted on a Dell desktop computer with Intel Pentium Intel (R) Core processor 3.30 GHz and 4 GB memory.

5.1. Quadratic Programming Problem. First, we consider the following quadratic programming problem:

$$\min \sum_{i=1}^{3}\left(\frac{1}{2}x_i^T M_i x_i + q_i^T x_i\right)$$
$$\text{s.t.}\quad A_1x_1 + x_2 + A_3x_3 = b, \tag{68}$$
$$x_1 \in \mathscr{R}^{n_1},\quad B_L \le x_1 \le B_U,\quad x_2 \in \mathscr{R}^{n_2},$$
$$\|x_2\| \le r,\quad x_3 \in \mathscr{R}_+^{n_3}.$$

In the experiments, we set $f_i(x_i) = q_1^T x_i$ and $g_i(x_1) = (1/2)x_i^T M_i x_i$, $(i = 1, 2, 3)$. We set the matrix $M_1 = I_{n_1\times n_1}$ and construct the rest of matrices M_i $(i = 2, 3)$ in a way similar to [25, 31]. That is, $M_i = V_i^T V_i + \tau I_{n_i}$, where V_i are random matrices and

$$\tau = \frac{\lambda_{\max}(V_i^T V_i) - t\cdot\lambda_{\min}(V_i^T V_i)}{t-1}. \tag{69}$$

In our tests, we set $t = 10^3$ and generate the matrices $V_i = \text{rand}((n_i/5), n_i)$ in Matlab function. In the experiments, we set $[B_L, B_U] = [0, 10]$ and the radius $r = 10$. For the linear constraint, the entries of $A_i \in \mathscr{R}^{l\times n_i}$ $(i = 1, 3)$ are uniformly distributed in $(0, 1)$ with the density 0.1 and $l = n_2$. x_i^*'s are given generated by $x_i^* = \text{sprand}(n_i, 1, 0.5)$ $(i = 1, 2, 3)$ in the Matlab function. In order to guarantee the feasibility of the problem, we set $q_i = -M_i x_i^*$, $(i = 1, 2, 3)$ and $b = A_1x_1^* + x_2^* + A_3x_3^*$. Thus, (x_1^*, x_2^*, x_3^*) is an optimal solution of (68). The Algorithm G-ADMM-S is compared with the PPSM-C in [25]. The initial iteration points are the zero vectors $x_i^0 = 0_{n_i\times 1}$ $(i = 1, 2, 3)$ and $\lambda^0 = 0_{l\times 1}$ for all tested algorithms. We set a maximal number of 20000 for iteration of the proposed algorithms with a modified stopping criterion as follows:

$$\max_{i=1,2,3}\left\{\frac{\|x_i^k - \overline{x}_i^k\|}{\|x_i^k\|}, \frac{\|\lambda^k - \overline{\lambda}^k\|}{\|\lambda^k\|}\right\} \le 10^{-2}. \tag{70}$$

Now, we specify the choices of parameters to implement these algorithms. First, we set $H = \beta I$ with $\beta = 0.01$ and the relaxation parameter $\gamma = 1.8$ for all tested algorithms. For "PPSM-C″, we set $r_i = \|M_i\|_F + 0.15\|A_i^T A_i\|$ $(i = 1, 2, 3)$, where $\|\cdot\|_F$ represents the Frobenius norm. For G-ADMM-S, we consider two cases of the matrices G_i $(i = 1, 2, 3)$:

Case 1: $G_i = r_i I_{n_i\times n_i} - \beta A_i^T A_i$ with $r_i = \|M_i\|_F + 0.15\|A_i^T A_i\|_F$

Case 2: $G_i = r_i I_{n_i\times n_i} - \beta A_i^T A_i$ with $r_i = \|M_i\|_F + \beta\|A_i^T A_i\|_F$

In order to investigate the stability and efficiency of our algorithms, we test 16 groups of problems with random data. Some preliminary numerical results are reported in Table 1. Since they are synthetic examples with random date, for each scenario, we test 10 times and report the average performance. Specifically, we report the number of iterations ("Iter.") and the computing time in seconds ("Time") for all the tested methods. The data in Table 1 show that Algorithm G-ADMM-S $(G_i = r_i I_{n_i\times n_i} - \beta A_i^T A_i$ with $r_i = \|M_i\|_F + \beta\|A_i^T\ A_i\|_F)$ is more efficient than the rest of algorithms for the test problems.

Table 1: Numerical results.

(n_1, n_2, n_3)	PPSM-C		G-ADMM-S Case 1		G-ADMM-S Case 2	
	Iter.	Time	Iter.	Time	Iter.	Time
(500, 500, 500)	1253	6.065	1589	6.290	1340	5.302
(500, 600, 500)	1057	5.483	1744	7.628	1523	6.671
(600, 500, 600)	1560	9.296	1825	8.589	1445	6.753
(600, 600, 600)	1580	10.307	1915	10.142	1642	8.625
(600, 700, 600)	1641	11.432	2042	11.817	1782	10.313
(700, 600, 700)	1745	13.129	2223	13.157	1871	11.087
(700, 700, 700)	2036	16.875	2357	16.304	1998	13.820
(700, 800, 700)	2857	25.090	2628	19.335	2308	17.168
(800, 700, 800)	2428	23.296	2634	19.875	2210	16.638
(800, 800, 800)	2809	28.659	2804	23.453	2381	19.620
(800, 900, 800)	4048	43.482	3016	28.678	2592	24.695
(900, 800, 900)	3204	41.398	3178	31.030	2675	26.080
(900, 900, 900)	3881	47.465	3332	36.036	2820	30.255
(900, 1000, 900)	10000	129.147	3790	44.253	3224	37.557
(1000, 900, 1000)	4547	71.910	3771	46.183	3140	38.250
(1000, 1000, 1000)	10000	151.273	4131	53.813	3482	45.660

5.2. Nonlinear Convex Programming Problem. In this section, we consider the following nonlinear convex programming problem:

$$\min\left(\|x_1\|_1 - \frac{1}{2}\ln\left(\|x_1\|^2 + 1\right)\right) + \left(q^T x_2 + \frac{1}{2}x_2^T M x_2\right)$$
$$+\left(\frac{1}{2}\|x_3\|^2 - \sum_{i=1}^{n_3}\cos x_{3,i}\right)$$
$$\text{s.t.}\quad A_1x_1 + A_2x_2 + A_3x_3 = b,$$
$$x_1 \in \mathscr{R}^{n_1},\ 0 \le x_1,\ x_2 \in \mathscr{R}^{n_2},\ x_3 \in \mathscr{R}^{n_3},\ |x_{3,i}| \le \frac{\pi}{2},\ i =, 1, \ldots, n_3, \tag{71}$$

where $\|x\|_1 = \sum_{i=1}^{n}|x_i|$ and M is a positive matrix. In the experiments, we set $f_1(x_1) = \|x_1\|_1, g_1(x_1) = -(1/2)\ln(\|x_1\|^2 + 1)$, $f_2(x_2) = q^T x_2, g_2(x_2) = (1/2)x_2^T\ Mx_2, f_3(x_3) = (1/2)\ \|x_3\|^2$, and $g_3(x_3) = -\sum_{i=1}^{n_3}\cos x_{3,i}$. It is easy to see that, when the PPSM-C in [25] solves (71), there is no explicit solution to the subproblems. In this section, we only use the Algorithm G-ADMM-S to solve (71). In the experiments, the entries of $A_i \in \mathscr{R}^{l\times n_i}$ $(i = 1, 3)$ are uniformly distributed in $(0, 1)$ with the density 0.1 and $l = n_2$. $x_1^* = 0_{n_1\times 1}$, $x_3^* = 0_{n_3\times 1}$, and x_2^* is given generated by $x_2^* = \text{sprand}(n_2, 1, 0.5)$ in the Matlab function. We set the matrix M in a way similar to [25, 31]. In order to guarantee the feasibility of the problem, we set $q = -Mx_2^*$ and $b = A_1x_1^* + A_2x_2^* + A_3x_3^*$. Thus, (x_1^*, x_2^*, x_3^*) is an optimal solution of (71). The initial iteration points are the zero vectors $x_i^0 = 0_{n_i\times 1}$ $(i = 1, 2, 3)$ and $\lambda^0 = 0_{l\times 1}$ for all tested algorithms. We set a maximal number of 20000 for iteration of the proposed algorithms with a modified stopping criterion as follows:

Table 2: Numerical results.

(n_1, n_2, n_3)	Case 1 ($\mu = \beta$)			Case 2 ($\mu = 0.15$)		
	Iter.	Time	f-error	Iter.	Time	f-error
(600, 600, 600)	3760.0	42.487	0.000211	3535.6	39.895	0.000207
(600, 700, 600)	5036.0	68.516	0.000196	5205.0	71.399	0.000200
(800, 700, 800)	3598.0	59.179	0.000211	3842.8	64.059	0.000211
(800, 800, 800)	5620.0	109.576	0.000202	5114.4	99.910	0.000206
(800, 900, 800)	7206.0	170.436	0.000192	7015.2	158.227	0.000199
(1000, 900, 1000)	5655.0	148.923	0.000203	5351.4	140.452	0.000205
(1000, 1000, 1000)	7577.0	223.832	0.000202	7529.2	229.729	0.000198

$$\max\left\{\left\|x_1^k\right\|, \frac{\left\|x_2^k - \overline{x}_2^k\right\|}{\left\|x_2^k\right\|}, \left\|x_3^k\right\|, \left\|A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b\right\|\right\} \leq 10^{-3}. \tag{72}$$

Now, we specify the choices of parameters to implement these algorithms. We set $H = \beta I$ with $\beta = 0.01$, the relaxation parameter $\gamma = 1.8$, $r_1 = n1 + \beta\|A_1^T A_1\|$, $r_2 = \|M\|_F + \beta\|A_2^T A_2\|$, $r_3 = n3 + \beta\|A_3^T A_3\|$, and $G_i = r_i I_{n_i \times n_i} - \mu A_i^T A_i$ $(i = 1, 2, 3)$. We consider two cases of the parameter μ: Case 1: $\mu = \beta$; Case 2: $\mu = 0.15$.

We test 7 groups of problems with random data. Numerical results are reported in Table 2. For each scenario, we test 5 times and report the average performance. Specifically, we report the number of iterations (“Iter.”), the computing time in seconds (“Time”), and the absolute error of function value (“f-error”). The numerical results show that Algorithm G-ADMM-S is effective.

6. Conclusion

In this paper, for the linearly constrained separable convex programming, whose objective function is the sum of m individual blocks with nonoverlapping variables and each block is convex, we present a gradient-based ADMM with a substitution in the case $m \geq 3$. We have analysed its convergence and iteration complexity. The preliminary numerical results have shown the efficiency of the proposed algorithm.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The first author would like to extend his sincere gratitude to his doctoral supervisor, Pro. Caozong Cheng, for his guidance.

References

[1] D. Gabay and B. Mercier, “A dual algorithm for the solution of nonlinear variational problems via finite element approximation,” *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.

[2] R. Glowinski and A. Marroco, “Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires,” *Revue Française D’automatique, Informatique, Recherche Opérationnelle. Analyse Numérique*, vol. 9, no. R2, pp. 41–76, 1975.

[3] 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*, vol. 3, no. 1, pp. 1–122, 2010.

[4] R. Glowinski, “On alternating direction methods of multipliers: a historical perspective,” *Computational Methods in Applied Sciences*, vol. 34, pp. 59–82, 2014.

[5] M. Tao and X. M. Yuan, “Recovering low-rank and sparse components of matrices from incomplete and noisy observations,” *SIAM Journal on Optimization*, vol. 21, no. 1, pp. 57–81, 2011.

[6] M. K. Ng, P. Weiss, and X. M. Yuan, “Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods,” *SIAM Journal on Scientific Computing*, vol. 32, no. 5, pp. 2710–2736, 2010.

[7] N. K. Bose and K. J. Boo, “High-resolution image reconstruction with multisensors,” *International Journal of Imaging Systems and Technology*, vol. 9, no. 4, pp. 294–304, 1998.

[8] M. Ng, F. Wang, and X. M. Yuan, “Fast minimization methods for solving constrained total-variation superresolution image reconstruction,” *Multidimensional Systems and Signal Processing*, vol. 22, no. 1–3, pp. 259–286, 2011.

[9] C. H. Rosa, *Pathways of Economic Development in an Uncertain Environment: A Finite Scenario Approach to the U.S. Region under Carbon Emission Restrictions, WP-94-41*, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1994.

[10] S. Setzer, G. Steidl, and T. Teuber, “Deblurring Poissonian images by split Bregman techniques,” *Journal of Visual Communication and Image Representation*, vol. 21, no. 3, pp. 193–199, 2010.

[11] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, “Latent variable graphical model selection via convex optimization,” *The Annals of Statistics*, vol. 40, no. 4, pp. 1935–1967, 2012.

[12] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, p. 544, Wiley-Interscience, Hoboken, NJ, USA, 2004.

[13] L. F. Yang, J. Y. Luo, Y. Xu, Z. R. Zhang, and Z. Y. Dong, “A distributed dual consensus ADMM based on partition for DC-DOPF with carbon emission trading,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1858–1872, 2020.

[14] L. F. Yang, J. B. Jian, K. Meng, Y. Xu, and Z. Dong, “A novel projected two-binary-variable formulation for unit

commitment in power systems," *Applied Energy*, vol. 187, pp. 732–745, 2017.

[15] M. T. Chao and C. Z. Cheng, "A note on the convergence of alternating proximal gradient method," *Applied Mathematics and Computation*, vol. 228, pp. 258–263, 2014.

[16] D. R. Han and X. M. Yuan, "A note on the alternating direction method of multipliers," *Journal of Optimization Theory and Applications*, vol. 155, no. 1, pp. 227–238, 2012.

[17] M. Y. Hong and Z. Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *Mathematical Programming*, vol. 162, no. 1-2, pp. 165–199, 2017.

[18] C. H. Chen, B. S. He, Y. Ye, and X. M. Yuan, "The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent," *Mathematical Programming*, vol. 155, no. 1-2, pp. 57–79, 2016.

[19] M. Li, D. F. Sun, and K. C. Toh, "A convergent 3-block semiproximal ADMM for convex minimization problems with one strongly convex block," *Asia-Pacific Journal of Operational Research*, vol. 32, p. 19, 2015.

[20] D. F. Sun, K. C. Toh, and L. Q. Yang, "A convergent 3-block semiproximal alternating direction method of multipliers for conic programming with 4-type constraints," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 882–915, 2015.

[21] B. S. He, M. Tao, and X. M. Yuan, "Alternating direction method with Gaussian back substitution for separable convex programming," *SIAM Journal on Optimization*, vol. 22, no. 2, pp. 313–340, 2012.

[22] M. T. Chao, C. Z. Cheng, and D. Y. Liang, "A proximal block minimization method of multipliers with a substitution procedure," *Optimization Methods and Software*, vol. 30, no. 4, pp. 825–842, 2015.

[23] M. T. Chao, C. Z. Cheng, and H. B. Zhang, "*A linearized alternating direction method of multipliers with substitution procedure*, Asia Pac," *Journal of Operations Research*, vol. 32, p. 19, 2015.

[24] D. R. Han, X. M. Yuan, W. X. Zhang, and X. J. Cai, "An ADM-based splitting method for separable convex programming," *Computational Optimization and Applications*, vol. 54, no. 2, pp. 343–369, 2013.

[25] D. R. Han, H. J. He, and L. L. Xu, "A proximal parallel splitting method for minimizing sum of convex functions with linear constraints," *Journal of Computational and Applied Mathematics*, vol. 256, pp. 36–51, 2014.

[26] B. S. He, M. Tao, and X. M. Yuan, "A splitting method for separable convex programming," *IMA Journal of Numerical Analysis*, vol. 34, pp. 1–33, 2014.

[27] P. L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM Journal on Numerical Analysis*, vol. 16, no. 6, pp. 964–979, 1979.

[28] G. B. Passty, "Ergodic convergence to a zero of the sum of monotone operators in Hilbert space," *Journal of Mathematical Analysis and Applications*, vol. 72, no. 2, pp. 383–390, 1979.

[29] E. Blum and W. Oettli, *Mathematische Optimierung, Econometrics and Operations Research XX*, Springer-Verlag, New York, NY, USA, 1975.

[30] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vol. I, Springer Series in Operations Research, Springer Verlag, New York, NY, USA, 2003.

[31] P. T. Harker and J. S. Pang, "A damped-Newton method for the linear complementarity problem," in *Computational Solution of Nonlinear Systems of Equations*, E. L. Allgower and K. Georg, Eds., pp. 265–284, American Math. Soc., Providence, RI, USA, 1990.