RESEARCH

Open Access

CrossMark

Convergence analysis on a modified generalized alternating direction method of multipliers

Sha Lu[1,2*] and Zengxin Wei[3]

*Correspondence:
lusha_nn@126.com
[1] School of Science, East China
University of Science and
Technology, Shanghai, China
[2] School of Mathematics and
Statistics, Guangxi Teachers
Education University, Nanning,
China
Full list of author information is
available at the end of the article

Abstract

The alternating direction method of multipliers (ADMM) is one of the most powerful and successful methods for solving convex composite minimization problem. The generalized ADMM relaxes both the variables and the multipliers with a common relaxation factor in $(0, 2)$, which has the potential of enhancing the performance of the classic ADMM. Very recently, two different variants of semi-proximal generalized ADMM have been proposed. They allow the weighting matrix in the proximal terms to be positive semidefinite, which makes the subproblems relatively easy to evaluate. One of the variants of semi-proximal generalized ADMMs has been analyzed theoretically, but the convergence result of the other is not known so far. This paper aims to remedy this deficiency and establish its convergence result under some mild conditions in the sense that the relaxation factor is also restricted into $(0, 2)$.

Keywords: Convex optimization; Augmented Lagrangian function; Alternating direction method of multipliers; Semi-proximal terms

1 Introduction

Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be real finite dimensional Euclidean spaces with the inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. In this paper, we consider the following convex composite problem with coupled linear equality constraint:

$$\min_{x \in \mathcal{X}, y \in \mathcal{Y}} \quad f(x) + g(y)$$
$$\text{s.t.} \quad Ax + By = c, \tag{1}$$

where $f : \mathcal{X} \to (-\infty, +\infty]$ and $g : \mathcal{Y} \to (-\infty, +\infty]$ are closed proper convex functions, $A : \mathcal{X} \to \mathcal{Z}$ and $B : \mathcal{Y} \to \mathcal{Z}$ are linear operators, and $c \in \mathcal{Z}$ is given. Many applications arising in various areas may have mathematical models with the form of (1), such as image processing, compressed sensing, and statistical learning. Denote A^* and B^* as the adjoint of A and B, respectively. Then the dual of problem (1) takes the form

$$\max_{\lambda \in \mathcal{Z}} \left\{ -f^*(A^*\lambda) - g^*(B^*\lambda) + \langle \lambda, c \rangle \right\}, \tag{2}$$

where $f^*(\cdot)$ (resp. $g^*(\cdot)$) is a Fenchel conjugate function of f (resp. g). Under Slater's constraint qualification, it is known that $(\bar{x}, \bar{y})$ is a solution to problem (1) if and only if there exists a Lagrangian multiplier $\bar{\lambda}$ such that the triple $(\bar{x}, \bar{y}; \bar{\lambda})$ is a solution to the following Karush–Kuhn–Tucker (KKT) conditions system:

$$
\begin{cases}
0 \in \partial f(x) - A^*\lambda, \\
0 \in \partial g(y) - B^*\lambda, \\
Ax + By - c = 0.
\end{cases}
\tag{3}
$$

The augmented Lagrangian function associated with (1) is defined as

$$
\mathcal{L}_\sigma(x, y; \lambda) := f(x) + g(y) - \langle \lambda, Ax + By - c \rangle + \frac{\sigma}{2}\|Ax + By - c\|^2,
$$

where $\lambda \in \mathcal{Z}$ is a multiplier and $\sigma > 0$ is a penalty parameter. Given (x_k, y_k), the classic augmented Lagrangian algorithm takes the form to derive the next pair (x_{k+1}, y_{k+1}):

$$
\begin{cases}
(x_{k+1}, y_{k+1}) := \arg\min_{x \in \mathcal{X}, y \in \mathcal{Y}} \mathcal{L}_\sigma(x, y; \lambda_k), & \text{(4a)} \\
\lambda_{k+1} := \lambda_k - \sigma(Ax_{k+1} + By_{k+1} - c). & \text{(4b)}
\end{cases}
$$

For solving subproblem (4a), we must minimize the function with strongly coupled quadratic term, which makes it hard to solve especially in large-scale problems. By noticing the individual structure of f and g in problem (1), one effective approach is the alternating direction method of multipliers (abbreviated as ADMM) that for $k = 0, 1, \ldots$,

$$
\begin{cases}
x_{k+1} := \arg\min_{x \in \mathcal{X}} \mathcal{L}_\sigma(x, y_k; \lambda_k), & \text{(5a)} \\
y_{k+1} := \arg\min_{y \in \mathcal{Y}} \mathcal{L}_\sigma(x_{k+1}, y; \lambda_k), & \text{(5b)} \\
\lambda_{k+1} := \lambda_k - \tau\sigma(Ax_{k+1} + By_{k+1} - c), & \text{(5c)}
\end{cases}
$$

where τ is a step-length which can be chosen in $(0, (1 + \sqrt{5})/2)$. The advantage of alternating technique lies in decomposing a large problem into several smaller pieces via its favorable structure, and then solving them accordingly.

The classic ADMM algorithm was originated by Glowinski and Marroco [1], Gabay and Mercier [2] in the mid-1970s. Gabay [3] showed that the classic ADMM with the $\tau = 1$ is a special case of the Douglas–Rachford splitting method for monotone operators in the early 1980s. Later, in [4], Eckstein and Bertsekas showed that the Douglas–Rachford splitting method is actually a special case of the proximal point algorithm. The variant of proximal ADMM was proposed by Eckstein [5], which ensures that each subproblem enjoys a unique solution by introducing an additional proximal term. This technique improves the behavior of the objective functions in the iteration subproblems and thus ameliorates the convergent property of the whole algorithm. He et al. [6] in turn showed that the proximal term can be chosen differently pre-iteration. Furthermore, Fazel et al. [7] gave a deep investigation and proved that the proximal term can be chosen to be positive semidefinite, which allows more flexible applications. One may refer to [8] for a note on the historical development of the ADMM, and some further research on ADMM can be seen in [9, 10], etc.

Another contribution of Eckstein and Bertsekas [4] is the designing of a generalized ADMM based on a generalized proximal point algorithm. Very recently, combining the idea of semi-proximal terms, Xiao et al. [11] proposed a semi-proximal generalized ADMM for convex composite conic programming, and numerically illustrated that their proposed method is very promising for solving doubly nonnegative semi-positive definite programming. The method of Xiao et al. [11] relaxed all the variables with a factor of $(0, 2)$, which has the potential of enhancing the performance of the classic ADMM. Additionally, in [11], Xiao et al. also developed another variant of semi-proximal generalized ADMM with different semi-proximal terms, but its convergence property has not been investigated so far. This paper targets to prove the global convergence of this semi-proximal generalized ADMM under some mild conditions, which may bring some theoretical foundations in some potential practical applications.

The rest of this paper is organized as follows. In Sect. 2, we present some preliminary results and review some variants of ADMMs. In Sect. 3, we establish the global convergence of the generalized semi-proximal ADMM with semi-proximal terms. In Sect. 4, we conclude this paper with some remarks.

2 Preliminaries

In this section, we provide some basic concepts and give a quick review of some variants of generalized ADMMs which will be used in the subsequent developments.

2.1 Basic concepts

Let $\mathcal{E}$ be a finite dimensional real Euclidean space with the inner product and the associated norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. Let $f : \mathcal{E} \to (-\infty, +\infty]$ be a closed proper convex function. The effective domain of f is defined as $\mathrm{dom}\, f = \{x \in \mathcal{E} | f(x) < +\infty\}$. The subdifferential of f is the operator defined as $\partial f(x) = \{x^* | f(z) \geq f(x) + \langle x^*, z - x \rangle, \forall z \in \mathcal{E}\}$, and it is simply denoted by $\partial f(x)$. Obviously, $\partial f(x)$ is a closed convex set while it is not empty. The point-to-set operator $\partial f : x \to \partial f(x)$ is trivially monotone, i.e., for any $x, y \in \mathcal{E}$ such that $\partial f(x)$ and $\partial f(y)$ are not empty, it holds that $\langle x - y, u - v \rangle \geq \|x - y\|_\Sigma^2$ for all $u \in \partial f(x)$ and $v \in \partial f(y)$, where $\Sigma : \mathcal{E} \to \mathcal{E}$ is a self-adjoint positive semidefinite linear operator. The Fenchel conjugate of a function f at $y \in \mathcal{E}$ is defined as

$$f^*(y) := \sup_x \{\langle x, y \rangle - f(x)\} = -\inf_x \{f(x) - \langle x, y \rangle\}.$$

It is well known in [12] that the conjugate function $f^*(y)$ is always convex and closed, proper if and only if f is proper. Furthermore, $(\mathrm{cl} f)^* = f^*$ and $f^{**} = \mathrm{cl} f$, where $\mathrm{cl} f$ denotes the closed function of f, i.e., the epigraph of $\mathrm{cl} f$ is a closure of the epigraph of the convex function f.

Assuming that the KKT system (3) is not empty, then the dual problem (2) can be solved by using the splitting method to solve the following inclusion problem:

$$0 \in (T_1 + T_2)(\lambda) \tag{6}$$

with

$$T_1(\lambda) = c - A \partial f^* (-A^* \lambda) \quad \text{and} \quad T_2(\lambda) = -B \partial g^* (-B^* \lambda). \tag{7}$$

It is easy to see that both T_1 and T_2 are maximal monotone operators. To solve (6), an equivalent form of the generalized proximal point algorithm of Eckstein and Bertsekas [4] with any initial point v_0 is that

$$
\begin{cases}
v_{k+1} = (1 - \rho)v_k + \rho[J_{\sigma T_1}(2J_{\sigma T_2} - I)v_k + (I - J_{\sigma T_2})v_k], \\
\lambda_{k+1} = J_{\sigma T_2}(v_{k+1}),
\end{cases}
\tag{8}
$$

where $\rho \in (0, 2)$ and $J_{\sigma T} = (I + \sigma T)^{-1}$ is the so-called resolvent operator.

2.2 Generalized ADMM

Eckstein and Bertsekas [4] further showed that the iterative framework (8) is equivalent to the following iterative scheme while it is used to solve the minimization problem (1):

$$
\begin{cases}
x_{k+1} := \arg\min_{x \in \mathcal{X}}\{f(x) - \langle \lambda_k, Ax \rangle + \frac{\sigma}{2}\|Ax + By_k - c\|^2\}, & \text{(9a)} \\
y_{k+1} := \arg\min_{y \in \mathcal{Y}}\{g(y) - \langle \lambda_k, By \rangle + \frac{\sigma}{2}\|\rho Ax_{k+1} - (1 - \rho)By_k + By - \rho c\|^2\}, & \text{(9b)} \\
\lambda_{k+1} := \lambda_k - \sigma[\rho Ax_{k+1} - (1 - \rho)By_k + By_{k+1} - \rho c]. & \text{(9c)}
\end{cases}
$$

Obviously, the classic ADMM (5a)–(5c) with $\tau = 1$ is exactly the generalized ADMM with $\rho = 1$.

Furthermore, Chen [13] showed that the generalized ADMM (9a)–(9c) is equivalent to the following ADMM scheme with initial point $\tilde{w}_0 = (\tilde{x}_0, \tilde{y}_0; \tilde{\lambda}_0)$, and the parameter $\rho \in (0, 2)$ is transformed into a relaxation factor:

$$
\begin{cases}
y_k := \arg\min_{y \in \mathcal{Y}} \mathcal{L}_\sigma(\tilde{x}_k, y; \tilde{\lambda}_k), & \text{(10a)} \\
\lambda_k := \tilde{\lambda}_k - \sigma(A\tilde{x}_k + By_k - c), & \text{(10b)} \\
x_k := \arg\min_{x \in \mathcal{X}} \mathcal{L}_\sigma(x, y_k; \lambda_k), & \text{(10c)} \\
\tilde{\omega}_{k+1} := \tilde{\omega}_k + \rho(\omega_k - \tilde{\omega}_k), & \text{(10d)}
\end{cases}
$$

where $\omega_k = (x_k, y_k; \lambda_k)$ and $\tilde{\omega}_k = (\tilde{x}_k, \tilde{y}_k; \tilde{\lambda}_k)$. More details on the equivalence of these two methods are also given in the Ph.D. thesis [13].

2.3 Proximal ADMM

In order to broaden the capability of the classic ADMM, Eckstein [5] added a proximal term to each subproblem, which reduced to

$$
\begin{cases}
x_{k+1} = \arg\min_{x \in \mathcal{X}}\{\mathcal{L}_\sigma(x, y_k; \lambda_k) + \frac{\sigma}{2}\|x - x_k\|_S^2\}, & \text{(11a)} \\
y_{k+1} = \arg\min_{y \in \mathcal{Y}}\{\mathcal{L}_\sigma(x_{k+1}, y; \lambda_k) + \frac{\sigma}{2}\|y - y_k\|_T^2\}, & \text{(11b)} \\
\lambda_{k+1} = \lambda_k + \tau\sigma(Ax_{k+1} + By_{k+1} - c), & \text{(11c)}
\end{cases}
$$

where S and T are positive definite matrices. Moreover, Fazel et al. [7] further illustrated that both weighting matrices can be chosen as positive semidefinite so that it can be applied in more practical situations. For more details on its convergence results, one can refer to [7] and the references therein.

It can be observed that the subproblems in the generalized ADMM schemes (10a)–(10d) may not admit solutions because A or B is not assumed to be row full-rank. One natural way to fix this problem is to add proximal terms to these subproblems. Very recently, Xiao et al. [11] suggested a couple of approaches to achieve this purpose. One of them is to add the semi-proximal terms $\frac{1}{2}\|x - x_{k-1}\|_S^2$ and $\frac{1}{2}\|y - y_{k-1}\|_T^2$ to the subproblems for computing x_k and y_k, i.e.,

$$
\begin{cases}
x_k := \arg\min_{x \in \mathcal{X}} \mathcal{L}_\sigma(x, \tilde{y}_k; \tilde{\lambda}_k) + \frac{1}{2}\|x - x_{k-1}\|_S^2, & \text{(12a)} \\[2mm]
\lambda_k := \tilde{\lambda}_k - \sigma(Ax_k + B\tilde{y}_k - c), & \text{(12b)} \\[2mm]
y_k := \arg\min_{y \in \mathcal{Y}} \mathcal{L}_\sigma(x_k, y; \lambda_k) + \frac{1}{2}\|y - y_{k-1}\|_T^2, & \text{(12c)} \\[2mm]
\tilde{\omega}_{k+1} = \tilde{\omega}_k + \rho(\omega_k - \tilde{\omega}_k). & \text{(12d)}
\end{cases}
$$

Another one is to add the proximal terms $\frac{1}{2}\|x - \tilde{x}_k\|_S$ and $\frac{1}{2}\|y - \tilde{y}_k\|_T^2$, i.e.,

$$
\begin{cases}
x_k := \arg\min_{x \in \mathcal{X}} \mathcal{L}_\sigma(x, \tilde{y}_k; \tilde{\lambda}_k) + \frac{1}{2}\|x - \tilde{x}_k\|_S^2, & \text{(13a)} \\[2mm]
\lambda_k := \tilde{\lambda}_k - \sigma(Ax_k + B\tilde{y}_k - c), & \text{(13b)} \\[2mm]
y_k := \arg\min_{y \in \mathcal{Y}} \mathcal{L}_\sigma(x_k, y; \lambda_k) + \frac{1}{2}\|y - \tilde{y}_k\|_T^2, & \text{(13c)} \\[2mm]
\tilde{\omega}_{k+1} = \tilde{\omega}_k + \rho(\omega_k - \tilde{\omega}_k). & \text{(13d)}
\end{cases}
$$

Compared with the traditional proximal approach (12a)–(12d), the semi-proximal terms in (13a)–(13d) are more natural in the sense that the most recently updated values of variables are involved. Actually, the global convergence of the iterative framework (13a)–(13d) has been analyzed in [11], and the corresponding numerical results illustrated that the proposed method can solve these problems not only effectively but also efficiently. In this paper, we particularly concentrate on the convergence analysis of the corresponding algorithm based on the former iterative framework to solve the separable convex minimization problem (1).

3 Global convergence

This section is devoted to analyzing the global convergence of the generalized ADMM based on the iterative framework (12a)–(12d) Since f and g are both closed proper convex functions, it is known that ∂f and ∂g are maximal monotone mappings [14], and then there exist a couple of self-adjoint positive semidefinite linear operators $\Sigma_f : \mathcal{X} \to \mathcal{X}$ and $\Sigma_g : \mathcal{Y} \to \mathcal{Y}$ such that, for any $x, x' \in \mathcal{X}$ and $y, y' \in \mathcal{Y}$ with $u \in \partial f(x)$, $u' \in \partial f(x')$, $v \in \partial g(y)$, and $v' \in \partial g(y')$,

$$
\langle u - u', x - x' \rangle \geq \|x - x'\|_{\Sigma_f}^2 \quad \text{and} \quad \langle v - v', y - y' \rangle \geq \|y - y'\|_{\Sigma_g}^2. \tag{14}
$$

First, we state the detailed steps of the generalized ADMM with semi-proximal terms (abbreviate it as sPGADM) as follows.

Algorithm sPGADM Set $\rho \in (0, 2)$, $\sigma > 0$. Choose $S : \mathcal{X} \to \mathcal{X}$, $T : \mathcal{Y} \to \mathcal{Y}$ such that $\sum_f + S + A^*A \succ 0$ and $\sum_g + T + B^*B \succ 0$. Input an initial point $\tilde{\omega}_0 = (\tilde{x}_0, \tilde{y}_0; \tilde{\lambda}_0) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. For $k = 1, 2, \ldots,$

Step 1 (main step). Compute

$$
\begin{cases}
x_k := \arg\min_{x \in \mathcal{X}} \{ f(x) - \langle \tilde{\lambda}_k, Ax \rangle + \frac{\sigma}{2} \| Ax + B\tilde{y}_k - c \|^2 + \frac{1}{2} \| x - x_{k-1} \|_S^2 \}, & \text{(15a)} \\
\lambda_k := \tilde{\lambda}_k - \sigma(Ax_k + B\tilde{y}_k - c), & \text{(15b)} \\
y_k := \arg\min_{y \in \mathcal{Y}} \{ g(y) - \langle \lambda_k, By \rangle + \frac{\sigma}{2} \| Ax_k + By - c \|^2 + \frac{1}{2} \| y - y_{k-1} \|_T^2 r \}. & \text{(15c)}
\end{cases}
$$

Step 2 (relaxation step). Compute

$$
\tilde{\omega}_{k+1} = \tilde{\omega}_k + \rho(\omega_k - \tilde{\omega}_k). \tag{16}
$$

Before deducing the convergence property of sPGADM, we do some preparations to facilitate the later analysis. Firstly, we make the following assumption.

Assumption A There exists at least one vector $(\bar{x}, \bar{y}; \bar{\lambda}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ such that the KKT system (3) is satisfied.

We now let $\{(x_k, y_k; \lambda_k)\}$ be the sequence generated by sPGADM and $(\bar{x}, \bar{y}; \bar{\lambda})$ be a solution of the KKT system (3). For a more convenient discussion, we denote $x_k^e = x_k - \bar{x}$, $y_k^e = y_k - \bar{y}$, and $\lambda_k^e = \lambda_k - \bar{\lambda}$.

The first-order optimality condition of (15a) can be expressed as

$$
A^* \tilde{\lambda}_k - \sigma A^* (Ax_k + B\tilde{y}_k - c) - S(x_k - x_{k-1}) \in \partial f(x_k),
$$

which combined with (15b) yields

$$
A^* \lambda_k - S(x_k - x_{k-1}) \in \partial f(x_k) \tag{17}
$$

and

$$
A^* \lambda_{k+1} - S(x_{k+1} - x_k) \in \partial f(x_{k+1}). \tag{18}
$$

Since $\bar{x}$ and $\bar{\lambda}$ satisfy the KKT system (3), then we obtain from (14) that

$$
\langle A^* \lambda_{k+1} - S(x_{k+1} - x_k) - A^* \bar{\lambda}, x_{k+1} - \bar{x} \rangle \geq \| x_{k+1} - \bar{x} \|_{\Sigma_f}^2,
$$

or, equivalently, by noting the definition of x_k^e,

$$
\langle \lambda_{k+1}^e, Ax_{k+1}^e \rangle - \langle S(x_{k+1} - x_k), x_{k+1}^e \rangle \geq \| x_{k+1}^e \|_{\Sigma_f}^2. \tag{19}
$$

Similarly, the first-order optimality condition of (15c) can also be described as

$$
B^* \lambda_k - \sigma B^* (Ax_k + By_k - c) - T(y_k - y_{k-1}) \in \partial g(y_k). \tag{20}
$$

Thus, from the monotone property (14), we get

$$
\langle B^* \lambda_k - \sigma B^* (Ax_k + By_k - c) - T(y_k - y_{k-1}) - B^* \bar{\lambda}, y_k - \bar{y} \rangle \geq \| y_k - \bar{y} \|_{\Sigma_g}^2,
$$

or, equivalently,

$$\langle \lambda_k^e, By_k^e \rangle - \langle \sigma(Ax_k + By_k - c), By_k^e \rangle - \langle T(y_k - y_{k-1}), y_k^e \rangle \geq \|y_k^e\|_{\Sigma_g}^2. \tag{21}$$

Adding two sides of (19) and (21) implies

$$\langle \lambda_{k+1}^e, Ax_{k+1}^e \rangle + \langle \lambda_k^e, By_k^e \rangle - \langle \sigma(Ax_k + By_k - c), By_k^e \rangle$$
$$- \langle S(x_{k+1} - x_k), x_{k+1}^e \rangle - \langle T(y_k - y_{k-1}), y_k^e \rangle$$
$$\geq \|x_{k+1}^e\|_{\Sigma_f}^2 + \|y_k^e\|_{\Sigma_g}^2,$$

which can be equivalently rewritten as

$$\langle \lambda_{k+1}^e, Ax_{k+1}^e + By_k^e \rangle + \langle \lambda_k^e - \lambda_{k+1}^e, By_k^e \rangle - \langle \sigma(Ax_k + By_k - c), By_k^e \rangle$$
$$- \langle S(x_{k+1} - x_k), x_{k+1}^e \rangle - \langle T(y_k - y_{k-1}), y_k^e \rangle$$
$$\geq \|x_{k+1}^e\|_{\Sigma_f}^2 + \|y_k^e\|_{\Sigma_g}^2. \tag{22}$$

Note that the first term on the left-hand side of (22) can be reorganized as

$$\langle \lambda_{k+1}^e, Ax_{k+1}^e + By_k^e \rangle = \langle \lambda_{k+1}^e + \sigma(1 - \rho)Ax_{k+1}^e, Ax_{k+1}^e + By_k^e \rangle$$
$$- \sigma(1 - \rho)\langle Ax_{k+1}^e + By_k^e, Ax_{k+1}^e + By_k^e \rangle$$
$$+ \sigma(1 - \rho)\langle By_k^e, Ax_{k+1}^e + By_k^e \rangle. \tag{23}$$

Then, together with $\lambda_k^e - \lambda_{k+1}^e = \lambda_k - \lambda_{k+1}$, (22) is transformed equally into

$$\langle \lambda_{k+1}^e + \sigma(1 - \rho)Ax_{k+1}^e, Ax_{k+1}^e + By_k^e \rangle + \big[\sigma(1 - \rho)\langle By_k^e, Ax_{k+1}^e + By_k^e \rangle$$
$$+ \langle \lambda_k^e - \lambda_{k+1}^e, By_k^e \rangle - \langle \sigma(Ax_k + By_k - c), By_k^e \rangle \big]$$
$$- \langle S(x_{k+1} - x_k), x_{k+1}^e \rangle - \langle T(y_k - y_{k-1}), y_k^e \rangle$$
$$\geq \|x_{k+1}^e\|_{\Sigma_f}^2 + \|y_k^e\|_{\Sigma_g}^2 + \sigma(1 - \rho)\|Ax_{k+1}^e + By_k^e\|^2. \tag{24}$$

The following two lemmas play a fundamental role in our convergence analysis.

Lemma 3.1 *Let $\{(x_k, y_k; \lambda_k)\}$ be the sequence generated from algorithm sPGADM, and $\{\bar{x}, \bar{y}; \bar{\lambda}\}$ is a solution of the KKT system. For $k = 1, 2, \ldots,$ it holds that*

$$\langle \lambda_{k+1}^e + \sigma(1 - \rho)Ax_{k+1}^e, Ax_{k+1}^e + By_k^e \rangle$$
$$= -\frac{\sigma\rho}{2}\|Ax_{k+1}^e + By_k^e\|^2 - \frac{1}{2\sigma\rho}\big[\|\lambda_{k+1}^e + \sigma(1 - \rho)Ax_{k+1}^e\|^2$$
$$- \|\lambda_k^e - \sigma(1 - \rho)Ax_k^e\|^2 \big]. \tag{25}$$

Proof From (15b) and (16), we have

$$\tilde{\lambda}_{k+1} = \lambda_k + (\rho - 1)(\lambda_k - \tilde{\lambda}_k) = \lambda_k - \sigma(\rho - 1)(Ax_k + B\tilde{y}_k - c)$$

and

$$
\begin{aligned}
\lambda_{k+1} &= \tilde{\lambda}_{k+1} - \sigma(Ax_{k+1} + B\tilde{y}_{k+1} - c) \\
&= \lambda_k - \sigma(\rho - 1)(Ax_k + B\tilde{y}_k - c) - \sigma\left[Ax_{k+1} + B(\tilde{y}_k + \rho(y_k - \tilde{y}_k)) - c\right] \\
&= \lambda_k - \sigma\rho(Ax_{k+1} + By_k - c) - \sigma(\rho - 1)(Ax_k - Ax_{k+1}).
\end{aligned}
\tag{26}
$$

Thus,

$$
\begin{aligned}
&\left(\lambda_{k+1}^e + \sigma(1-\rho)Ax_{k+1}^e\right) - \left(\lambda_k^e + \sigma(1-\rho)Ax_k^e\right) \\
&\quad = (\lambda_{k+1} - \lambda_k) + \sigma(1-\rho)(Ax_{k+1} - Ax_k) \\
&\quad = -\sigma\rho(Ax_{k+1} + By_k - c) - \sigma(\rho-1)(Ax_k - Ax_{k+1}) + \sigma(1-\rho)(Ax_{k+1} - Ax_k) \\
&\quad = -\sigma\rho(Ax_{k+1} + By_k - c).
\end{aligned}
$$

By using (26), we get

$$
\left(\lambda_{k+1}^e + \sigma(1-\rho)Ax_{k+1}^e\right) + \sigma\rho(Ax_{k+1} + By_k - c) = \lambda_k^e + \sigma(1-\rho)Ax_k^e.
$$

Together with the basic relation

$$
2\langle u, v \rangle = \|u\|^2 + \|v\|^2 - \|u - v\|^2 = \|u + v\|^2 - \|u\|^2 - \|v\|^2,
\tag{27}
$$

it implies that

$$
\begin{aligned}
&2\langle \lambda_{k+1}^e + \sigma(1-\rho)Ax_{k+1}^e, \sigma\rho(Ax_{k+1} + By_k - c)\rangle \\
&\quad = \left\|\lambda_k^e + \sigma(1-\rho)Ax_k^e\right\|^2 - \left\|\lambda_{k+1}^e + \sigma(1-\rho)Ax_{k+1}^e\right\|^2 \\
&\qquad - \sigma^2\rho^2\|Ax_{k+1} + By_k - c\|^2,
\end{aligned}
$$

which is equivalent to

$$
\begin{aligned}
&\langle \lambda_{k+1}^e + \sigma(1-\rho)Ax_{k+1}^e, Ax_{k+1} + By_k - c\rangle \\
&\quad = \frac{1}{2\sigma\rho}\left\|\lambda_k^e + \sigma(1-\rho)Ax_k^e\right\|^2 - \frac{1}{2\sigma\rho}\left\|\lambda_{k+1}^e + \sigma(1-\rho)Ax_{k+1}^e\right\|^2 \\
&\qquad - \frac{\sigma\rho}{2}\|Ax_{k+1} + By_k - c\|^2.
\end{aligned}
\tag{28}
$$

Since $\{\bar{x}, \bar{y}; \bar{\lambda}\}$ satisfies the KKT system (3) that

$$
A\bar{x} + B\bar{y} = c, \qquad Ax_{k+1} + By_k - c = Ax_{k+1}^e + By_k^e,
$$

then from (28) we get (25). □

Lemma 3.2 *Let* $\{(x_k, y_k; \lambda_k)\}$ *be a sequence generated from algorithm sPGADM, and* $\{\bar{x}, \bar{y}; \bar{\lambda}\}$ *be a solution of the KKT system. For* $k = 1, 2, \ldots$, *it holds that*

$$
\sigma(1-\rho)\langle By_k^e, Ax_{k+1}^e + By_k^e \rangle + \langle \lambda_k^e - \lambda_{k+1}^e, By_k^e \rangle - \langle \sigma(Ax_k + By_k - c), By_k^e \rangle
$$

$$
\leq -\frac{2-\rho}{\rho}\|x_{k+1} - x_k\|_{\Sigma_f}^2 - \frac{2-\rho}{2\rho}\|x_{k+1} - x_k\|_S^2 + \frac{2-\rho}{2\rho}\|x_k - x_{k-1}\|_S^2
$$

$$
- \frac{\sigma(2-\rho)^2}{2\rho}\|Ax_{k+1}^e - Ax_k^e\|^2 + \frac{\sigma(2-\rho)}{2}\|Ax_k^e\|^2 - \frac{\sigma(2-\rho)}{2}\|Ax_{k+1}^e\|^2. \tag{29}
$$

Proof By using the elementary relation (27), we have

$$
\langle S(x_{k+1} - x_k) - S(x_k - x_{k-1}), x_{k+1} - x_k \rangle
$$

$$
= \|x_{k+1} - x_k\|_S^2 - \langle S(x_k - x_{k-1}), x_{k+1} - x_k \rangle
$$

$$
\geq \|x_{k+1} - x_k\|_S^2 - \frac{1}{2}\left(\|x_{k+1} - x_k\|_S^2 + \|x_k - x_{k-1}\|_S^2\right)
$$

$$
= \frac{1}{2}\|x_{k+1} - x_k\|_S^2 - \frac{1}{2}\|x_k - x_{k-1}\|_S^2. \tag{30}
$$

Similar to (19), from (14), (17), and (18), we get

$$
\langle A^*(\lambda_{k+1} - \lambda_k) - S(x_{k+1} - x_k) + S(x_k - x_{k-1}), x_{k+1} - x_k \rangle \geq \|x_{k+1} - x_k\|_{\Sigma_f}^2.
$$

Then

$$
\langle \lambda_{k+1} - \lambda_k, A(x_{k+1} - x_k) \rangle - \langle S(x_{k+1} - x_k) - S(x_k - x_{k-1}), x_{k+1} - x_k \rangle \geq \|x_{k+1} - x_k\|_{\Sigma_f}^2,
$$

which together with (30) implies

$$
\langle \lambda_{k+1} - \lambda_k, A(x_{k+1} - x_k) \rangle
$$

$$
\geq \frac{1}{2}\|x_{k+1} - x_k\|_S^2 - \frac{1}{2}\|x_k - x_{k-1}\|_S^2 + \|x_{k+1} - x_k\|_{\Sigma_f}^2. \tag{31}
$$

Notice that from (26)

$$
\sigma(Ax_{k+1}^e + By_k^e) = -\rho^{-1}\left[\lambda_{k+1} - \lambda_k + \sigma(1-\rho)(Ax_{k+1} - Ax_k)\right].
$$

Combining it with (31), we obtain

$$
\sigma\langle Ax_k^e - Ax_{k+1}^e, Ax_{k+1}^e + By_k^e \rangle
$$

$$
= -\rho^{-1}\langle Ax_k^e - Ax_{k+1}^e, \lambda_{k+1} - \lambda_k + \sigma(1-\rho)(Ax_{k+1} - Ax_k)\rangle
$$

$$
= \rho^{-1}\langle Ax_{k+1}^e - Ax_k^e, \lambda_{k+1} - \lambda_k \rangle + \rho^{-1}\sigma(1-\rho)\|Ax_{k+1}^e - Ax_k^e\|^2
$$

$$
\geq \frac{1}{\rho}\|x_{k+1} - x_k\|_{\Sigma_f}^2 + \frac{1}{2\rho}\|x_{k+1} - x_k\|_S^2 - \frac{1}{2\rho}\|x_k - x_{k-1}\|_S^2
$$

$$
+ \frac{\sigma(1-\rho)}{\rho}\|Ax_{k+1}^e - Ax_k^e\|^2. \tag{32}
$$

Also from (26)

$$
\begin{aligned}
&\sigma(1-\rho)\big(Ax^e_{k+1} + By^e_k\big) + \big(\lambda^e_k - \lambda^e_{k+1}\big) - \sigma(Ax_k + By_k - c) \\
&= \sigma(1-\rho)\big(Ax^e_{k+1} + By^e_k\big) + \big[\sigma\rho(Ax_{k+1} + By_k - c) + \sigma(\rho-1)(Ax_k - Ax_{k+1})\big] \\
&\quad - \sigma(Ax_k + By_k - c) \\
&= \sigma(\rho-2)\big(Ax^e_k - Ax^e_{k+1}\big).
\end{aligned}
\tag{33}
$$

Then, since $\rho \in (0,2)$, by using (33), (32), and (27) successively, we deduce that

$$
\begin{aligned}
&\sigma(1-\rho)\big\langle Ax^e_{k+1} + By^e_k, By^e_k\big\rangle + \big\langle \lambda^e_k - \lambda^e_{k+1}, By^e_k\big\rangle - \sigma\big\langle Ax_k + By_k - c, By^e_k\big\rangle \\
&= \sigma(\rho-2)\big\langle Ax^e_k - Ax^e_{k+1}, By^e_k\big\rangle \\
&= -\sigma(2-\rho)\big\langle Ax^e_k - Ax^e_{k+1}, Ax^e_{k+1} + By^e_k\big\rangle + \sigma(2-\rho)\big\langle Ax^e_k - Ax^e_{k+1}, Ax^e_{k+1}\big\rangle \\
&\le -(2-\rho)\bigg[\frac{1}{\rho}\|x_{k+1} - x_k\|^2_{\Sigma_f} + \frac{1}{2\rho}\|x_{k+1} - x_k\|^2_S - \frac{1}{2\rho}\|x_k - x_{k-1}\|^2_S \\
&\quad + \frac{\sigma(1-\rho)}{\rho}\big\|Ax^e_{k+1} - Ax^e_k\big\|^2\bigg] \\
&\quad + \frac{\sigma(2-\rho)}{2}\big[\big\|Ax^e_k\big\|^2 - \big\|Ax^e_k - Ax^e_{k+1}\big\|^2 - \big\|Ax^e_{k+1}\big\|^2\big] \\
&= -\frac{2-\rho}{\rho}\|x_{k+1} - x_k\|^2_{\Sigma_f} - \frac{2-\rho}{2\rho}\|x_{k+1} - x_k\|^2_S + \frac{2-\rho}{2\rho}\|x_k - x_{k-1}\|^2_S \\
&\quad - \frac{\sigma(2-\rho)^2}{2\rho}\big\|Ax^e_{k+1} - Ax^e_k\big\|^2 + \frac{\sigma(2-\rho)}{2}\big\|Ax^e_k\big\|^2 - \frac{\sigma(2-\rho)}{2}\big\|Ax^e_{k+1}\big\|^2.
\end{aligned}
\tag{34}
$$

Thus, (29) is true and the proof is completed. □

Now, based on Lemmas 3.1 and 3.2, we define ϕ_k $(k > 0)$ as

$$
\begin{aligned}
\phi_k := {}&\frac{1}{\sigma\rho}\big\|\lambda^e_k + \sigma(1-\rho)Ax^e_k\big\|^2 + \big\|x^e_k\big\|^2_S + \big\|y^e_{k-1}\big\|^2_T \\
&+ \frac{2-\rho}{\rho}\|x_k - x_{k-1}\|^2_S + \sigma(2-\rho)\big\|Ax^e_k\big\|^2.
\end{aligned}
\tag{35}
$$

The following theorem shows that the sequence $\{\phi_k\}_{k>0}$ is monotonically decreasing and the algorithm sPGADM is globally convergent.

Theorem 3.3 *Assume that the solution set of* (1) *is nonempty and that there exists a vector* $(\bar{x}, \bar{y}; \bar{\lambda})$ *satisfying the KKT system* (3). *Let* $\phi_k(k > 0)$ *be defined as* (35), *and let the sequence* $\{(x_k, y_k; \lambda_k)\}$ *be generated from algorithm sPGADM. Then, for* $k = 1, 2, \ldots,$

$$
\begin{aligned}
&\phi_k - \phi_{k+1} \\
&\ge 2\big\|x^e_{k+1}\big\|^2_{\Sigma_f} + 2\big\|y^e_k\big\|^2_{\Sigma_g} + \sigma(2-\rho)\big\|Ax^e_{k+1} + By^e_k\big\|^2 + \|x_{k+1} - x_k\|^2_S \\
&\quad + \|y_k - y_{k-1}\|^2_T + \frac{2(2-\rho)}{\rho}\|x_{k+1} - x_k\|^2_{\Sigma_f} + \frac{\sigma(2-\rho)^2}{\rho}\big\|Ax^e_k - Ax^e_{k+1}\big\|^2.
\end{aligned}
\tag{36}
$$

Furthermore, the sequence $\{(x_k, y_k; \lambda_k)\}$ *converges to a KKT solution* $(\bar{x}, \bar{y}; \bar{\lambda})$ *to problem* (1).

Proof From (27), we have

$$\langle S(x_{k+1} - x_k), x_{k+1}^e \rangle = \frac{1}{2}\big[\|x_{k+1} - x_k\|_S^2 + \|x_{k+1} - \bar{x}\|_S^2 - \|x_k - \bar{x}\|_S^2\big], \tag{37}$$

$$\langle T(y_k - y_{k-1}), y_k^e \rangle = \frac{1}{2}\big[\|y_k - y_{k-1}\|_T^2 + \|y_k - \bar{y}\|_T^2 - \|y_{k-1} - \bar{y}\|_T^2\big]. \tag{38}$$

Substituting (25), (29), (37), and (38) into the four parts of the left-hand side of (24) respectively, we get

$$\begin{aligned}
&\Big[-\frac{\sigma\rho}{2}\big\|Ax_{k+1}^e + By_k^e\big\|^2 - \frac{1}{2\sigma\rho}\big(\big\|\lambda_{k+1}^e + \sigma(1-\rho)Ax_{k+1}^e\big\|^2 \\
&\quad - \big\|\lambda_k^e - \sigma(1-\rho)Ax_k^e\big\|^2\big)\Big] - \frac{2-\rho}{\rho}\|x_{k+1} - x_k\|_{\Sigma_f}^2 \\
&\quad - \frac{2-\rho}{2\rho}\big(\|x_{k+1} - x_k\|_S^2 - \|x_k - x_{k-1}\|_S^2\big) - \frac{\sigma(2-\rho)^2}{2\rho}\big\|Ax_{k+1}^e - Ax_k^e\big\|^2 \\
&\quad + \frac{\sigma(2-\rho)}{2}\big(\big\|Ax_k^e\big\|^2 - \big\|Ax_{k+1}^e\big\|^2\big) - \frac{1}{2}\big(\|x_{k+1} - x_k\|_S^2 + \big\|x_{k+1}^e\big\|_S^2 - \big\|x_k^e\big\|_S^2\big) \\
&\quad - \frac{1}{2}\big(\|y_k - y_{k-1}\|_T^2 + \big\|y_k^e\big\|_T^2 - \big\|y_{k-1}^e\big\|_T^2\big) \\
&\geq \langle \lambda_{k+1}^e + \sigma(1-\rho)Ax_{k+1}^e, Ax_{k+1}^e + By_k^e \rangle + \big[\sigma(1-\rho)\langle By_k^e, Ax_{k+1}^e + By_k^e \rangle \\
&\quad + \langle \lambda_k^e - \lambda_{k+1}^e, By_k^e \rangle - \langle \sigma(Ax_k + By_k - c), By_k^e \rangle\big] \\
&\quad - \langle S(x_{k+1} - x_k), x_{k+1}^e \rangle - \langle T(y_k - y_{k-1}), y_k^e \rangle \\
&\geq \big\|x_{k+1}^e\big\|_{\Sigma_f}^2 + \big\|y_k^e\big\|_{\Sigma_g}^2 + \sigma(1-\rho)\big\|Ax_{k+1}^e + By_k^e\big\|^2. \tag{39}
\end{aligned}$$

By the definition of $\phi_k\,(k > 0)$ and some simple deformations, it is easy to see that (39) implies (36).

Because $\rho \in (0, 2)$, then (35) and (36) deduce that the sequence $\{\phi_k\}_{k>0}$ is nonnegative and monotonically nonincreasing. Thus, the boundedness of ϕ_k means that each part on the right hand side of (35), such as $\{\|\lambda_k^e + \sigma(1-\rho)Ax_k^e\|\}$, $\{\|x_k^e\|_S\}$, $\{\|y_{k-1}^e\|_T\}$, $\{\|x_k - x_{k-1}\|_S\}$, and $\{\|Ax_k^e\|\}$, is bounded. Also, from (36), it gives that when $k \to \infty$,

$$\begin{aligned}
&\big\|x_{k+1}^e\big\|_{\Sigma_f} \to 0, \qquad \big\|y_k^e\big\|_{\Sigma_g} \to 0, \qquad \big\|Ax_{k+1}^e + By_k^e\big\| \to 0, \\
&\|x_{k+1} - x_k\|_S \to 0, \qquad \|y_k - y_{k-1}\|_T \to 0, \\
&\|x_{k+1} - x_k\|_{\Sigma_f} \to 0, \qquad \big\|Ax_k^e - Ax_{k+1}^e\big\| \to 0. \tag{40}
\end{aligned}$$

From the boundedness of $\{\|x_k^e\|_S\}$, $\{\|x_k^e\|_{\Sigma_f}\}$, and $\{\|Ax_k^e\|\}$, we know that $\{\|x_k\|_{\Sigma_f+S+A^*A}\}$ is bounded. Furthermore, since $\Sigma_f + S + A^*A \succ 0$ in sPGADM, we get $\{\|x_k\|\}$ is also bounded. Similarly, we can deduce that $\{\|y_k\|_{\Sigma_g+T+B^*B}\}$ is bounded due to the fact that $\{\|y_k^e\|_T\}$, $\{\|y_k^e\|_{\Sigma_g}\}$, $\{\|Ax_k^e\|\}$, and $\{\|Ax_{k+1}^e + By_k^e\|\}$ are all bounded. Thus, from $\Sigma_g + T + B^*B \succ 0$, $\{\|y_k\|\}$ is bounded. Also, the boundedness of $\{\|\lambda_k\|\}$ can come from the boundedness of $\{\|\lambda_k^e + \sigma(1-\rho)Ax_k^e\|\}$ and $\{\|Ax_k^e\|\}$.

The boundedness of the sequence $\{(x_k, y_k; \lambda_k)\}$ implies that there exists at least one convergent subsequence; for simplicity we denote it as

$$\lim_{k_i \to \infty} \{(x_{k_i}, y_{k_i}; \lambda_{k_i})\} = (x_\infty, y_\infty; \lambda_\infty), \quad \{k_i\} \subseteq \{0, 1, \ldots\}.$$

By using (17) and (20), we obtain

$$A^* \lambda_{k_i} - S(x_{k_i} - x_{k_i-1}) \in \partial f(x_{k_i}), \tag{41}$$

$$B^* \lambda_{k_i} - \sigma B^*(Ax_{k_i} + By_{k_i} - c) - T(y_{k_i} - y_{k_i-1}) \in \partial g(y_{k_i}). \tag{42}$$

Because f and g are closed proper convex functions, the nonempty sets ∂f and ∂g are closed. By noticing that $\|Ax_k + By_k - c\|^2 \leq \|Ax_{k+1}^e + By_k^e\|^2 + \|Ax_k^e - Ax_{k+1}^e\|^2$, and as mentioned before,

$$\left\|Ax_{k+1}^e + By_k^e\right\| \to 0, \qquad \left\|Ax_k^e - Ax_{k+1}^e\right\| \to 0, \tag{43}$$

$$\|x_{k+1} - x_k\|_S \to 0, \qquad \|y_k - y_{k-1}\|_T \to 0, \tag{44}$$

we take limits with k_i on both sides of (41) and (42). It implies that $(x_\infty, y_\infty; \lambda_\infty)$ satisfies the KKT condition:

$$A^* \lambda_\infty \in \partial f(x_\infty), \qquad B^* \lambda_\infty \in \partial g(y_\infty), \qquad Ax_\infty + By_\infty - c = 0.$$

To complete the whole proof, now we will show that $(x_\infty, y_\infty; \lambda_\infty)$ is the unique limit of the sequence $\{(x_k, y_k; \lambda_k)\}$. In fact, since $(x_\infty, y_\infty; \lambda_\infty)$ satisfies the KKT condition, without loss of generality, we can let $(\bar{x}, \bar{y}; \bar{\lambda}) = (x_\infty, y_\infty; \lambda_\infty)$. Thus, from the definition of ϕ_k in (35), there exists a subsequence $\{(x_{k_i}, y_{k_i}; \lambda_{k_i})\}$ such that

$$\lim_{k_i \to \infty} \phi_{k_i} = 0.$$

Together with the nonincreasing and boundedness of $\{\phi_k\}$, we know that $\{\phi_k\}$ converges to zero itself. By (35), it turns out that when $k \to \infty$,

$$\begin{aligned} &\left\|\lambda_k^e + \sigma(1-\rho)Ax_k^e\right\| \to 0, \qquad \left\|x_k^e\right\|_S \to 0, \qquad \left\|y_{k-1}^e\right\|_T \to 0, \\ &\|x_k - x_{k-1}\|_S \to 0, \qquad \left\|Ax_k^e\right\| \to 0. \end{aligned} \tag{45}$$

Thus, $\lim_{k \to \infty} \lambda_k = \bar{\lambda}$ since $0 \leq \|\lambda_k^e\| \leq \|\lambda_k^e + \sigma(1-\rho)Ax_k^e\| + \sigma|1-\rho|\|Ax_k^e\|$. Noticing that $\|x_k^e\|_{\Sigma_f} \to 0$ in (40), we get

$$\lim_{k \to \infty} \left(\left\|x_k^e\right\|_{\Sigma_f} + \left\|x_k^e\right\|_S + \left\|Ax_k^e\right\|\right) = 0.$$

Thus, from $\Sigma_f + S + A^*A \succ 0$ and $x_k^e = x_k - \bar{x}$, it comes true that

$$\lim_{k \to \infty} x_k = \bar{x}.$$

Also, from (45) together with $\|y_k^e\|_{\Sigma_g} \to 0$ and $\|Ax_{k+1}^e + By_k^e\| \to 0$ in (40), we have

$$\lim_{k\to\infty} \left(\left\|y_k^e\right\|_{\Sigma_g} + \left\|y_k^e\right\|_T + \left\|By_k^e\right\| \right) = 0.$$

Then, by $\Sigma_g + T + B^*B \succ 0$ and $y_k^e = y_k - \bar{y}$, it deduces that $\lim_{k\to\infty} y_k = \bar{y}$. The above discussion concludes that the whole sequence $\{(x_k, y_k; \lambda_k)\}$ converges to $(\bar{x}, \bar{y}; \bar{\lambda})$. The proof is completed.　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　□

4　Conclusions

The generalized ADMM, as an important variant of ADMM, is derived from the generalized proximal point algorithm while it is used to solve the sum of maximal monotone operators inclusion problems. Recently, it was shown that the generalized ADMM is also equivalent to the unit step-length ADMM but with additional relaxation steps based on a factor within $(0, 2)$. Combining the idea of semi-proximal terms, Xiao et al. [11] proposed a semi-proximal generalized ADMM and numerically illustrated that their proposed method is very promising for semi-positive definite programming. Additionally, Xiao et al. [11] introduced another variant of semi-proximal generalized ADMM with different semi-proximal terms, but its convergence property has not been investigated so far. This study aimed to remedy this deficiency and established its convergence result under some mild conditions in the sense that the relaxation factor is also restricted into $(0, 2)$. More precisely, if $\rho \in (0, 2)$, theoretical analysis has shown that the proposed algorithm converges globally by assuming that the optimal solutions set is nonempty and the matrices $\Sigma_f + \mathcal{S} + A^*A$ and $\Sigma_g + \mathcal{T} + B^*B$ are both positive definite. The result is quite in accord with the standard semi-proximal ADMM [11]. The paper paid more attention to analyzing the generalized semi-proximal ADMM for solving separable convex minimization. However, it has not been tested with different factor values of ρ for performance comparing. This should be our further task to investigate.

Authors' contributions
All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Author details
[1]School of Science, East China University of Science and Technology, Shanghai, China. [2]School of Mathematics and Statistics, Guangxi Teachers Education University, Nanning, China. [3]School of Mathematics and Information Science, Guangxi University, Nanning, China.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Glowinski, R., Marroco, A.: Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualité, d'une classe de problemes de Dirichlet nonlineares. Rev. Fr. Autom. Inform. Rech. Opér., Anal. Numér. **9**, 41–76 (1975)
2. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**, 17–40 (1976)

3. Gabay, D.: Applications of the method of multipliers to variational inequalities in augmented Lagrangian methods: applications to the numerical solution of boundary-value problems. In: Fortin, M., Glowinski, R. (eds.) Studies in Mathematics and Its Applications, vol. 15, pp. 299–331. Elsevier, Amsterdam (1983)
4. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**, 293–318 (1992)
5. Eckstein, J.: Some saddle-function splitting methods for convex programming. Optim. Methods Softw. **4**, 75–83 (1994)
6. He, B.S., Liao, L.Z., Han, D.R., Yang, H.: A new inexact alternating directions method for monotone variational inequalities. Math. Program. **92**, 103–118 (2002)
7. Fazel, M., Pong, T.K., Sun, D.F., Tseng, P.: Hankel matrix rank minimization with applications in system identification and realization. SIAM J. Matrix Anal. Appl. **34**, 946–977 (2013)
8. Glowinski, R.: On alternating direction methods of multipliers: a historical perspective. In: Fitzgibbon, W., Kuznetsov, Y.A., Neittaanmaki, P., Pironneau, O. (eds.) Modeling, Simulation and Optimization for Science and Technology, pp. 59–82. Springer, Berlin (2014)
9. He, B.S., Yuan, X.M.: On the $o(1/n)$ convergence rate of the alternating direction method. SIAM J. Numer. Anal. **50**, 700–709 (2012)
10. He, B.S., Yuan, X.M.: Block-wise alternating direction method of multipliers for multiple-block convex programming and beyond. SMAI J. Comput. Math. **1**, 145–174 (2015)
11. Xiao, Y., Chen, L., Li, D.: A generalized alternating direction method of multipliers with semi-proximal terms for convex composite conic programming. Math. Program. Comput. (2018, in press). https://doi.org/10.1007/s12532-018-0134-9
12. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
13. Chen, C.H.: Numerical algorithms for a class of matrix norm approximation problems. PhD thesis, Nanjing University, Department of Mathematics, (2012)
14. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, New York (1998)