A fundamental proof of convergence of alternating direction method of multipliers for weakly convex optimization

Tao Zhang[1] and Zhengwei Shen[1*]

*Correspondence:
ijwmip@ustb.edu.cn
[1] School of Mathematics and Physics, University of Science and Technology Beijing, China, Beijing, China

Abstract

The convergence of the alternating direction method of multipliers (ADMMs) algorithm to convex/nonconvex combinational optimization has been well established in the literature. Due to the extensive applications of a weakly convex function in signal processing and machine learning, in this paper, we investigate the convergence of an ADMM algorithm to the strongly and weakly convex combinational optimization (SWCCO) problem. Specifically, we firstly show the convergence of the iterative sequences of the SWCCO-ADMM under a mild regularity condition; then we establish the $o(1/k)$ sublinear convergence rate of the SWCCO-ADMM algorithm using the same conditions and the linear convergence rate by imposing the gradient Lipschitz continuity condition on the objective function. The techniques used for the convergence analysis in this paper are fundamental, and we accomplish the global convergence without using the Kurdyka–Łojasiewicz (KL) inequality, which is common but complex in the proof of nonconvex ADMM.

Keywords: Convex optimization; Weakly convex function; Ublinear and linear convergence; ADMM

1 Introduction

1.1 Motivation and problem

The alternating direction method of multiplier (ADMM) algorithm, as one of the splitting/decoupling techniques, has been successfully exploited in a wide range of structured sparsity regularization optimization problems in machine learning and inverse problems, such as signal restoration, matrix/tensor competition (or factorization), phase retrieval, compressive sensing, regression analysis or statistical inference. In these applications, one of the most important issues is how to induce a reasonable sparse solution and how to eliminate the disturbance (e.g. noise) while preserving the important features of the data. Nonconvex penalty or regularization imposed on a certain prior being incorporated into the optimization model can efficiently improve these problems. However, in practice, it is hard to obtain a closed-form solution for a generic nonconvex penalty optimization model except the weakly convex (a.k.a. semiconvex) penalty of which including a large class of functions, such as difference-of-convex (DC) functions. In addition, the convergence analysis of ADMM based on the weakly convex also has not been addressed in the literature.

Motivated by these observations, in this paper, we consider the following widely used strongly and weakly convex combinational optimization (SWCCO) problem:

$$\min_x F(x) = f(x) + g(Mx), \tag{1}$$

where $f(\cdot)$ is proper closed strongly convex and $g(\cdot)$ is proper closed weakly convex; M is a linear transformation, such as a wavelet/framelet. This problem is equivalent to the following constrained optimization problem:

$$\min_{x,y} f(x) + g(y) \quad s.t. \quad Mx = y. \tag{2}$$

The corresponding SWCCO-ADMM algorithm, firstly proposed in [1, 2], for (2) is

$$\begin{cases} x^{k+1} = \arg\min_x L_\rho(x, y^k, p^k), \\ y^{k+1} = \arg\min_y L_\rho(x^{k+1}, y, p^k), \\ p^{k+1} = p^k + \rho(Mx^{k+1} - y^{k+1}), \end{cases} \tag{3}$$

where $L_\rho(x, y, p) = f(x) + g(y) + \langle p, Mx - y \rangle + \frac{\rho}{2}\|Mx - y\|^2$ is the augmented Lagrangian function (ALF) with ρ being the penalty parameter and $p \in R^n$ being the Lagrangian multiplier. The first-order optimality conditions for (3) are

$$\begin{cases} -\rho M^T(y^{k+1} - y^k) \in \partial f(x^{k+1}) + M^T p^{k+1}, \\ 0 \in \partial g(y^{k+1}) - p^{k+1}, \\ \rho(Mx^{k+1} - y^{k+1}) = p^{k+1} - p^k, \end{cases} \tag{4}$$

where $M^T(y^{k+1} - y^k)$ and $Mx^{k+1} - y^{k+1}$ (or $p^{k+1} - p^k$) are dual and primal residual, respectively.

The main tasks of this manuscript are to study the convergence and sublinear (or linear) convergence rate for the SWCCO-ADMM algorithm (3) under quite general assumptions.

1.2 Related work

In order to avoid solving a nonconvex optimization problem directly, intuitively, we can separate a part from the strongly convex term of the weakly convex term to make the optimization problem (1) a convex–convex combination as follows:

$$\min_x \underbrace{\left(f(x) - \frac{\rho_2}{2}\|Mx\|^2\right)}_{f_1(x)} + \underbrace{\left(g(Mx) + \frac{\rho_2}{2}\|Mx\|^2\right)}_{g_1(Mx)};$$

$$\min_x \underbrace{\left(f(x) - \frac{\rho_1}{2}\|x\|^2\right)}_{f_2(x)} + \underbrace{\left(g(Mx) + \frac{\rho_1}{2}\|x\|^2\right)}_{g_2(\hat{M}x)};$$

$$\min_x \underbrace{\left(f(x) - \frac{\rho_2}{2}\|x\|^2\right)}_{f_3(x)} + \underbrace{\left(g(Mx) + \frac{\rho_2}{2}\|x\|^2\right)}_{g_3(\hat{M}x)};$$

$$\min_x \underbrace{\left(f(x) - \frac{\rho_1}{2}\|Mx\|^2\right)}_{f_4(x)} + \underbrace{\left(g(Mx) + \frac{\rho_1}{2}\|Mx\|^2\right)}_{g_4(Mx)}.$$

Here ρ_1 and ρ_2 are the strongly convex and weakly convex modulus, respectively; see details in Definition 2.1. However, such a partition and combination is infeasible in most cases, since these two terms in (1) usually play different roles in signal processing, thus such a simple separation will completely change the sense of the original model; moreover, the linear operator involved in these terms will make such a separation a hard situation; and even if it can be separated, the recombined model will become harder to solve as shown in [3, 4]. Also, we cannot ensure the new proximal operators to be easily computed.

The convergence of the forward–backward splitting (FBS) and the Douglas–Rachford splitting (DRS) method corresponding to a SWCCO problem with M being an identity map has been established in [3, 5, 6] very recently. Even though using DRS for the dual problem is equivalent to using ADMM for the primal problem as shown in [7, 8] in the context of convex cases, this is not the case for a nonconvex optimization problem, since the conjugate function of a nonconvex function has not been well defined yet in the literature. Therefore, it is necessary to establish the convergence of the SWCCO-ADMM algorithm.

In the convex case, convergence for ADMM has been well studied in [9–12]. In the nonconvex case, however, it has been merely investigated limited to certain specific assumptions. This is because the iterative sequences (x^k, y^k, p^k) generated by ADMM for a nonconvex optimization problem does not satisfy Féjer monotonicity; moreover, the sufficient decrease condition [13] used to measure the quality of descent methods to nonconvex objective function $F(x)$ is also no longer valid. In order to prove convergence to a nonconvex ADMM algorithm, some unverifiable or unreasonable assumptions have to be imposed for the objective functions or for iterative sequences (x^k, y^k, p^k). In addition, the global convergence and convergence rates of a nonconvex optimization problem usually require $F(x)$ to satisfy the Kurdyka–Łojasiewicz (KL) property at each point, where the KL exponent (i.e. the geometrical properties of the objective function around its stationary point) is not easily to determine [13].

In fact, the key steps to establish convergence of the nonconvex ADMM algorithm is to prove the dual residual $M^T(y^{k+1} - y^k)$ and the primal residual $(Mx^{k+1} - y^{k+1})$ (or $(p^{k+1} - p^k)$) going to zero as $k \to \infty$. In the end, one common method developed in very recently papers [13–17] is to exploit the monotonically nonincreasing of certain type of Lyapunov function (e.g. the ALF [14–16] or its variant [17]) to measure the iterative error. Furthermore, the global convergence of a nonconvex optimization problem can be obtained by using the KL inequality. However, we have to emphasize that using the ALF or its variant as the Lyapunov function leaves trouble handling the nonsmooth objective functions, since the Lyapunov function may no longer be monotonically decreasing in this case. Thus, in [14–17], at least one or more of the objective functions are required to be gradient Lipschitz continuous to guarantee the algorithm convergence, which will greatly limit the applications of the model (1).

On the other hand, the surjective (i.e. full row rank) assumption for the linear mapping M, which is a necessary condition to prove the monotonically nonincreasing property of ALF in [14–17], however, will exclude many interesting applications. In fact, many well-known developed sparse representation operators M do not satisfy the full row rank condition, for example, the framelet or the learned sparse representation redundant dictionary from the given data. Moreover, if we assume $f(x)$ to be gradient Lipschitz continuous and M to be full row rank, then problem (2) actually can be solved by using block co-

Table 1 Regularity conditions to convergence analysis of ADMM for nonconvex optimization problem

	M	f, g	KL property	Lyapunov function
Hong [14]	$M = I$	f and g gradient Lipschitz continuous	*no*	ALF
Li [15]	full row rank	bounded Hessian of f	*yes*	ALF
Wang [16]	full row rank	f gradient Lipschitz continuous g strongly convex	*yes*	variants of ALF
Wang [17]	weak full column rank condition	f gradient Lipschitz continuous g has special structure	*yes*	ALF
Ours	full column rank	f strongly convex g weakly convex	*no*	H^k (shown in Sect. 3)

ordinate descent methods (BCDs) [18] and does not need an ADMM algorithm. Specifically, suppose M to be of full row rank, then there exists a nonsingular matrix M_1 that can split the constraint $Mx = y$ into the following formulation: $M_1 x_1 + M_2 x_2 = y$. That is, $x_1 = M_1^{-1}(y - M_2 x_2)$. In this case, the constrained optimization problem (2) can be reformatted as the following unconstrained optimization problem:

$$\min_{x_1, x_2, y} f(x_1, x_2) + g(y) = \underbrace{f\left(M_1^{-1}(y - M_2 x_2), x_2\right)}_{h(x_2, y)} + g(y). \tag{5}$$

Then we have

$$\min_{x_2, y} h(x_2, y) + \langle 0, x_2 \rangle + g(y), \tag{6}$$

which can be solved by BCD. The regularity conditions used in a convergence analysis for nonconvex ADMM are summarized in Table 1.

Recently, a similar work [4, 19] to this manuscript has investigated the use of the primal–dual hybrid gradient (PDHG) method to solve (1), and also has established convergence. Although the PDHG (a.k.a, the relaxed alternating minimization algorithm (AMA) [20]) is apparently similar to the ADMM, they are quite different. Actually, the PDHG has a deep relationship with the inexact Uzawa method [19, 20].

1.3 Contributions

In this paper, without requiring gradient Lipschitz continuity of any of the objective functions, we establish the convergence of the SWCCO-ADMM algorithm (3); we merely require M to be full column rank. We also establish the sublinear convergence rate with the same assumptions as in convergence analysis, and we establish the linear convergence rate of the SWCCO-ADMM algorithm, which needs some additional regularity assumptions. In addition, we use the nonnegative Lyapunov function defined in [9] to measure the iterative error instead of using ALF, and the global convergence result can be obtained without using the KL inequality. Thus, our proof is more fundamental than the work in [14–17].

We summarize the contributions of this manuscript as follows.

- We prove that the iterative sequences $\{(x^k, y^k, p^k)\}$ of SWCCO-ADMM (3) globally converge to the critical point of (2) under the conditions that the strongly convex terms dominate the weakly convex terms and the penalty parameter ρ is at least larger than at least twice the weakly convex modulus. We also give an example that the

SWCCO-ADMM (3) will diverge if the latter condition is not satisfied. Meanwhile, we prove that the iterative sequence $\{x^k\}$ generated by ADMM (3) converges to an optimal solution of (1).

- We show a sublinear convergence rate $o(1/k)$ of the SWCCO-ADMM algorithm using the same regularity conditions as the convergence analysis. To the best of our knowledge, this is the first sublinear convergence rate result for a nonconvex ADMM algorithm.
- Furthermore, we establish the linear convergence rate for the SWCCO-ADMM algorithm by imposing the gradient Lipschitz continuity on one of the objective functions.

The rest of this paper is organized as follows. In Sect. 2, we list some fundamental definitions which are useful for the following analysis. Then the convergence of ADMM for the case SWCCO is established in Sect. 3. Next, we address the sublinear and linear convergence rate in Sect. 4. We also give a numerical example in Sect. 5. At last, some conclusions are given in Sect. 6.

2 Preliminaries

To start, we list some fundamental definitions and notations for the following analysis.

We denote by $\langle \cdot, \cdot \rangle$ and $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$ the inner product and a norm on finite-dimensional real vectors spaces X and Y, respectively. For a given function $f : R^n \to R \cup \{\infty\}$, we denote $\text{dom} f = \{x : f(x) < +\infty\}$ it being nonempty. The largest and smallest eigenvalue of the linear operator M are defined by $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$, respectively.

Definition 2.1 The function $f : R^n \to R \cup \{\infty\}$ is said to be strongly convex with modulus $\rho_1 > 0$ if $f(x) - \frac{\rho_1}{2}\|x\|^2$ is convex; the function $g : R^n \to R \cup \{\infty\}$ is said to be weakly convex with modulus $\rho_2 > 0$, if $g(y) + \frac{\rho_2}{2}\|y\|^2$ is convex.

The strongly and weakly convex functions have the following properties [21]. Let $f : R^n \to R \cup \{\infty\}$ be a strongly convex function with modulus ρ_1. Then, for $u_1 \in \partial f(x_1)$, $u_2 \in \partial f(x_2)$, we have

$$\langle u_1 - u_2, x_1 - x_2 \rangle \geq \rho_1 \|x_1 - x_2\|^2. \tag{7}$$

For a weakly convex function $g : R^n \to R \cup \{\infty\}$ with modulus ρ_2 and $v_1 \in \partial g(y_1)$, $v_2 \in \partial g(y_2)$, we have

$$\langle v_1 - v_2, y_1 - y_2 \rangle \geq -\rho_2 \|y_1 - y_2\|^2. \tag{8}$$

Definition 2.2 ([22]) Let $f : R^n \to R \cup \{\infty\}$ be a proper lower semi-continuous function being finite at $\bar{x} \in R^n$.

 (i) The Fréchet subdifferential of f at $\bar{x}$, written $\hat{\partial} f(\bar{x})$, is the set

$$\left\{ u \in R^n : \lim_{x \neq \bar{x}} \inf_{x \to \bar{x}} \frac{f(x) - f(\bar{x}) - \langle u, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0 \right\}.$$

 (ii) The limiting subdifferential of f at $\bar{x}$, written $\partial f(\bar{x})$, is the set

$$\left\{ u \in R^n : \exists x^k \to \bar{x}, f(x^k) \to f(\bar{x}), u^k \in \hat{\partial} f(x^k), u^k \to u \right\}.$$

From this definition, let $f : R^n \to R \cup \{\infty\}$ be a proper lower semi-continuous function; we can get the following assertions.

(i) The subdifferential of f at $\bar{x}$ is the set

$$\{u \in R^n : \forall x \in R^n, \langle x - \bar{x}, u \rangle + f(\bar{x}) \leq f(x)\}.$$

(ii) The limiting subdifferential of f is closed, i.e. if $y^k \to \bar{y}$, $v^k \to \bar{v}$, $g(y^k) \to g(\bar{y})$ and $v^k \in \partial g(y^k)$, then $\bar{v} \in \partial g(\bar{y})$.

(iii) Suppose $\text{dom}(f) \cap \text{dom}(g \circ M)$ is not empty, then

$$\partial \big(f(x) + g(Mx)\big) = \partial f(x) + M^T(\partial g)(Mx).$$

Definition 2.3 Let a nonempty set S be the set of critical point of augmented Lagrangian function L_ρ. $(x^*, y^*, z^*) \in S$ is a critical point if

$$-M^T p^* \in \partial f(x^*), \quad p^* \in \partial g(y^*) \quad and \quad M^T x^* - y^* = 0.$$

In order to find linear convergence, we need the following gradient Lipschitz continuous definition.

Definition 2.4 Let f be a differentiable function, then the gradient ∇f is called Lipschitz continuous with modulus $L > 0$ if

$$\big\| \nabla f(x_1) - \nabla f(x_2) \big\| \leq L \| x_1 - x_2 \|, \quad \forall x_1, x_2 \in \text{dom} f.$$

We will often use the following relations, for all vectors $u, v, w \in R^n$:

$$2\langle u - v, w - u \rangle = \| v - w \|^2 - \| u - v \|^2 - \| u - w \|^2; \tag{9}$$

$$\| u + v \|^2 \leq \left(1 + \frac{1}{\gamma} \right) \| u \|^2 + (1 + \gamma) \| v \|^2, \quad \forall \gamma > 0. \tag{10}$$

3 Convergence analysis

In this section we will study the convergence of the SWCCO-ADMM algorithm (3) under the following mild assumptions.

Assumption 3.1

(i) Let $f(x)$ be strongly convex with modulus $\rho_1 > 0$ and $g(y)$ be weakly convex with modulus $\rho_2 > 0$, respectively; and the set of augmented Lagrangian function critical point S is nonempty. For $\forall \mu > 0$, we need

$$\rho_1 - \rho_2 \| M \|^2 - \mu \geq 0. \tag{11}$$

(ii) We also suppose that the penalty parameter ρ in the augmented Lagrangian function satisfies

$$\rho > 2\rho_2 + \frac{8\rho_2^2 \| MM^T \|}{\mu}. \tag{12}$$

Assumption 3.2 $\mathrm{dom}(f) \cap \mathrm{dom}(g \circ M)$ is not empty.

Remark 1　From (ii) of Assumption 3.1, we will see that ρ is larger than $2\rho_2$. This assumption not only ensures the unique solution of the second subproblem in ADMM algorithm (3) as shown later, but also is used to ensure that the Lyapunou function H^k defined in Lemma 3.1 has a sufficient descent property. We also give an example that the iterative sequences of SWCCO-ADMM algorithm (3) will diverge if this assumption is not satisfied in the end of this section.

For the sake of convenience, let (x^*, y^*, p^*) be one of the critical points of the augmented function; $z^{k+1} = -M^T p^k - \rho M^T (M x^{k+1} - y^k)$; $x_e^{k+1} = x^{k+1} - x^*$; $y_e^{k+1} = y^{k+1} - y^*$; $p_e^{k+1} = p^{k+1} - p^*$ and $z_e^{k+1} = z^{k+1} - (-M^T p^*)$ in the following proof. Then by the strong convexity property (7) of $f(x)$ and the weak convexity property (8) of $g(y)$ we have

$$\langle z_e^{k+1}, x_e^{k+1} \rangle \geq \rho_1 \| x_e^{k+1} \|^2 \tag{13}$$

and

$$\langle p_e^{k+1}, y_e^{k+1} \rangle \geq -\rho_2 \| y_e^{k+1} \|^2. \tag{14}$$

By Assumption 3.1, we will obtain the following monotonically nonincreasing property of the nonnegative Lyapunou function $H^k = \frac{\rho}{2} \| y_e^k \|^2 + \frac{1}{2\rho} \| p_e^k \|^2$, which will play an important role in our convergence analysis. Note that x_e^k is not considered in H^k, since the primal variable x can be regarded as an intermediate variable in the iterations of SWCCO-ADMM, while y and p are the essential variables [9].

Lemma 3.1 *Let $H^k = \frac{\rho}{2} \| y_e^k \|^2 + \frac{1}{2\rho} \| p_e^k \|^2$. Then the iterative sequences $\{(x^k, y^k, p^k)\}$ generated by SWCCO-ADMM algorithm* (3) *satisfy*

$$H^k - H^{k+1} \geq \sigma_1 \| x_e^{k+1} \|^2 + \sigma_2 \| y^{k+1} - y^k \|^2 + \sigma_3 \| p^{k+1} - p^k \|^2, \tag{15}$$

where $\sigma_1 = \rho_1 - \rho_2 \| M \|^2 - \mu \geq 0$, $\sigma_2 = \frac{\rho}{2} - \rho_2 > 0$ *and* $\sigma_3 = \frac{1}{2\rho} - \frac{\rho_2}{\rho^2} - \frac{4\rho_2^2 \| M M^T \|}{\rho^2 \mu} > 0$.

Proof　By the optimality condition of (4), we have

$$\begin{cases} -\rho M^T (y_e^{k+1} - y_e^k) - M^T p_e^{k+1} \in \partial f(x^{k+1}) - \partial f(x^*); \\ p_e^{k+1} \in \partial g(y^{k+1}) - \partial g(x^*). \end{cases} \tag{16}$$

Then, by (13) and (14), it follows that

$$\langle -M^T p_e^{k+1} - \rho M^T (y_e^{k+1} - y_e^k), x_e^{k+1} \rangle \geq \rho_1 \| x_e^{k+1} \|^2;$$
$$\langle p_e^{k+1}, y_e^{k+1} \rangle \geq -\rho_2 \| y_e^{k+1} \|^2.$$

Adding these two inequalities, then

$$\langle -M^T p_e^{k+1} - \rho M^T (y_e^{k+1} - y_e^k), x_e^{k+1} \rangle + \langle p_e^{k+1}, y_e^{k+1} \rangle \geq \rho_1 \| x_e^{k+1} \|^2 - \rho_2 \| y_e^{k+1} \|^2. \tag{17}$$

By rearranging the left side of inequality (17), we have

$$
\begin{aligned}
&\langle -M^T p_e^{k+1} - \rho M^T(y_e^{k+1} - y_e^k), x_e^{k+1}\rangle + \langle p_e^{k+1}, y_e^{k+1}\rangle \\
&\quad = -\langle p_e^{k+1}, Mx_e^{k+1}\rangle - \rho\langle y_e^{k+1} - y_e^k, Mx_e^{k+1}\rangle + \langle p_e^{k+1}, y_e^{k+1}\rangle \\
&\quad = -\frac{1}{\rho}\langle p_e^{k+1}, p_e^{k+1} - p_e^k\rangle - \rho\langle y_e^{k+1}, Mx_e^{k+1}\rangle + \rho\langle y_e^k, Mx_e^{k+1}\rangle \\
&\quad = -\frac{1}{2\rho}\left(\|p_e^{k+1}\|^2 + \|\bar{p}_e^{k+1} - p_e^k\|^2 - \|p_e^k\|^2\right) \\
&\qquad - \frac{\rho}{2}\left(\|y_e^{k+1}\|^2 + \|Mx_e^{k+1}\|^2 - \|Mx_e^{k+1} - y_e^{k+1}\|^2\right) \\
&\qquad + \frac{\rho}{2}\left(\|y_e^k\|^2 + \|Mx_e^{k+1}\|^2 - \|Mx_e^{k+1} - y_e^k\|^2\right) \\
&\quad = \frac{1}{2\rho}\left(\|p_e^k\|^2 - \|p_e^{k+1}\|^2\right) + \frac{\rho}{2}\left(\|y_e^k\|^2 - \|y_e^{k+1}\|^2\right) - \frac{\rho}{2}\|Mx_e^{k+1} - y_e^k\|^2,
\end{aligned}
\tag{18}
$$

where the third equation follows from the cosine rule (9). Then, combining (17) and (18), we have

$$
H^k - H^{k+1} \geq \rho_1\|x_e^{k+1}\|^2 - \rho_2\|y_e^{k+1}\|^2 + \frac{\rho}{2}\|Mx_e^{k+1} - y_e^k\|^2.
\tag{19}
$$

Notice that (19) does not imply the nonincreasing property of H^k since we have the second negative term $-\rho_2\|y_e^{k+1}\|^2$ on the right side. Next, we will deal with this term using strong and weak convexity of the functions $f(x)$ and $g(y)$, respectively.

Firstly, the third term $\frac{\rho}{2}\|Mx_e^{k+1} - y_e^k\|^2$ on the right side of (19) can be rewritten as follows:

$$
\begin{aligned}
\frac{\rho}{2}\|Mx_e^{k+1} - y_e^k\|^2 &= \frac{\rho}{2}\|Mx_e^{k+1} - y_e^{k+1} + y_e^{k+1} - y_e^k\|^2 \\
&= \frac{\rho}{2}\|Mx_e^{k+1} - y_e^{k+1}\|^2 + \frac{\rho}{2}\|y_e^{k+1} - y_e^k\|^2 + \rho\langle Mx_e^{k+1} - y_e^{k+1}, y_e^{k+1} - y_e^k\rangle \\
&= \frac{1}{2\rho}\|p^{k+1} - p^k\|^2 + \frac{\rho}{2}\|y^{k+1} - y^k\|^2 + \langle p^{k+1} - p^k, y^{k+1} - p^k\rangle \\
&\geq \frac{1}{2\rho}\|p^{k+1} - p^k\|^2 + \left(\frac{\rho}{2} - \rho_2\right)\|y^{k+1} - y^k\|^2,
\end{aligned}
\tag{20}
$$

where the inequality follows from the weakly convex property (8). Secondly, from the last equation of (3), it follows that

$$
\begin{aligned}
\rho_2\|y_e^{k+1}\|^2 &= \frac{\rho_2}{\rho^2}\|p^{k+1} - p^k - \rho Mx_e^{k+1}\|^2 \\
&= \frac{\rho_2}{\rho^2}\|p^{k+1} - p^k\|^2 + \rho_2\|Mx_e^{k+1}\|^2 - \frac{2\rho_2}{\rho}\langle p^{k+1} - p^k, Mx_e^{k+1}\rangle.
\end{aligned}
\tag{21}
$$

Substituting (20), (21) into (19) and rearranging the terms, we have

$$
\begin{aligned}
H^k - H^{k+1} &\geq \rho_1\|x_e^{k+1}\|^2 - \rho_2\|Mx_e^{k+1}\|^2 \\
&\quad + \left(\frac{1}{2\rho} - \frac{\rho_2}{\rho^2}\right)\|p^{k+1} - p^k\|^2 + \left(\frac{\rho}{2} - \rho_2\right)\|y^{k+1} - y^k\|^2
\end{aligned}
$$

$$+ \frac{2\rho_2}{\rho} \langle p^{k+1} - p^k, Mx_e^{k+1} \rangle$$

$$\geq \rho_1 \|x_e^{k+1}\|^2 - \rho_2 \|Mx_e^{k+1}\|^2 - 2\rho_2\zeta \|x_e^{k+1}\|^2$$

$$+ \left(\frac{1}{2\rho} - \frac{\rho_2}{\rho^2} - \frac{2\rho_2\|MM^T\|}{\rho^2\zeta} \right) \|p^{k+1} - p^k\|^2$$

$$+ \left(\frac{\rho}{2} - \rho_2 \right) \|y^{k+1} - y^k\|^2$$

$$\geq \left(\rho_1 - \rho_2\|M\|^2 - 2\rho_2\zeta \right) \|x_e^{k+1}\|^2 + \left(\frac{1}{2\rho} - \frac{\rho_2}{\rho^2} - \frac{2\rho_2\|MM^T\|}{\rho^2\zeta} \right) \|p^{k+1} - p^k\|^2$$

$$+ \left(\frac{\rho}{2} - \rho_2 \right) \|y^{k+1} - y^k\|^2, \tag{22}$$

where the second inequality follows from the Cauchy–Schwarz inequality, i.e. $\frac{2\rho_2}{\rho} \langle p^{k+1} - p^k, Mx_e^{k+1} \rangle \geq -2\rho_2\zeta \|x_e^{k+1}\|^2 - \frac{2\rho_2\|MM^T\|}{\rho^2\zeta} \|p^{k+1} - p^k\|^2$, $\forall \zeta > 0$. By the arbitrariness of ζ, let $\mu = 2\rho_2\zeta$. It follows from (22) that

$$H^k - H^{k+1} \geq \left(\rho_1 - \rho_2\|M\|^2 - \mu \right) \|x_e^{k+1}\|^2 + \left(\frac{1}{2\rho} - \frac{\rho_2}{\rho^2} - \frac{4\rho_2^2\|MM^T\|}{\rho^2\mu} \right) \|p^{k+1} - p^k\|^2$$

$$+ \left(\frac{\rho}{2} - \rho_2 \right) \|y^{k+1} - y^k\|^2.$$

By Assumption 3.1, we have $\sigma_1 = \rho_1 - \rho_2\|M\|^2 - \mu \geq 0$, $\sigma_2 = \frac{\rho}{2} - \rho_2 > 0$ and $\sigma_3 = \frac{1}{2\rho} - \frac{\rho_2}{\rho^2} - \frac{4\rho_2^2\|MM^T\|}{\rho^2\mu} > 0$. Therefore we complete the proof. $\qquad \square$

Now, we are ready to prove the main convergence result of the SWCCO-ADMM algorithm (3) under Assumption 3.1.

Theorem 3.2 *Under Assumption 3.1 suppose the iterative sequence $\{x^k\}$ of SWCCO-ADMM (3) is bounded. Let $\{(x^*, y^*, p^*)\}$ be a critical point in S. Then:*

  (i) *The iterative sequences $\{(y^k, p^k)\}$ generated by SWCCO-ADMM (3) converge to $\{(y^*, p^*)\}$ and $Mx^k \to Mx^*$, as $k \to \infty$.*

  (ii) *Under Assumption 3.2 if one of the following conditions holds, i.e. M is full column rank or $\sigma_1 = \rho_1 - \rho_2\|M\|^2 - \mu > 0$, then $\{x^k\}$ converges to a optimal solution of SWCCO problem (1).*

Proof (i) Adding both sides of (15) in Lemma (3.1), from $k = 0$ to $k = \infty$, it follows that

$$\sigma_2 \sum_{k=0}^{\infty} \|y^{k+1} - y^k\|^2 + \sigma_3 \sum_{k=0}^{\infty} \|p^{k+1} - p^k\|^2 \leq H^0 - H^\infty < \infty.$$

Thus,

$$\|p^{k+1} - p^k\| \to 0 \quad and \quad \|y^{k+1} - y^k\| \to 0 \quad as\ k \to \infty. \tag{23}$$

Since $\{(x^*, y^*, p^*)\}$ is a critical point and $H^k \leq H^0$ follows from Lemma (3.1), it follows that y^k and p^k are bounded. Next, from the boundedness of $\{(x^k, y^k, p^k)\}$ and supposing that

$\{(\bar{x}, \bar{y}, \bar{p})\}$ is an accumulation point of $\{(x^k, y^k, p^k)\}$; there exists a subsequence $\{(x^{k_j}, y^{k_j}, p^{k_j})\}$ that converges to the $\{(\bar{x}, \bar{y}, \bar{p})\}$, i.e.,

$$\lim_{j \to \infty} (x^{k_j}, y^{k_j}, p^{k_j}) = (\bar{x}, \bar{y}, \bar{p}).$$

Also, from (23), we have $p^{k_j+1} - p^{k_j} = Mx^{k_j+1} - y^{k_j+1} \to 0$; after passing to limits, we obtain

$$M\bar{x} - \bar{y} = 0. \tag{24}$$

Now, we will show that $\{(\bar{x}, \bar{y}, \bar{p})\}$ is a critical point in the following. By taking the limit on the first equation of the optimality conditions (4) along with $\{(x^{k_j}, y^{k_j}, p^{k_j})\}$, and from the closeness of ∂f, we obtain

$$-M^T \bar{p} \in \partial f(\bar{x}). \tag{25}$$

Since y^{k_j+1} is a minimizer of $L_\rho(x^{k_j}, y, p^{k_j})$, we have

$$L_\rho\left(x^{k_j}, y^{k_j+1}, p^{k_j}\right) \le L_\rho\left(x^{k_j}, \bar{y}, p^{k_j}\right).$$

Taking the limit of the above inequality, we get

$$\limsup_{j \to \infty} L_\rho\left(x^{k_j}, y^{k_j+1}, p^{k_j}\right) \le L_\rho(\bar{x}, \bar{y}, \bar{p}).$$

Next, from the lower semicontinuity of L_ρ, we also have

$$\liminf_{j \to \infty} L_\rho\left(x^{k_j}, y^{k_j+1}, p^{k_j}\right) \ge L_\rho(\bar{x}, \bar{y}, \bar{p}).$$

Combining the above two inequalities, we conclude $g(y^{k_j+1}) = g(\bar{y})$, and from assertion (ii) of definition (2.2)

$$\bar{p} \in \partial g(\bar{x}). \tag{26}$$

This together with (24) and (25), shows $\{(\bar{x}, \bar{y}, \bar{p})\}$ is a critical point follows from the optimality conditions (4).

Without loss of generality, let $\{(x^*, y^*, p^*)\} = \{(\bar{x}, \bar{y}, \bar{p})\}$. Again from the nonincreasing monotonicity and boundedness of H^k, we know H^k is convergent. Since $\lim_{j \to \infty}(x^{k_j}, y^{k_j}, p^{k_j}) = (x^*, y^*, p^*)$, we say $H^k \to 0$, i.e., $y^k \to y^*$ and $p^k \to p^*$. Again due to $p^{k+1} = p^k + \rho(Mx^{k+1} - y^{k+1})$, we get $Mx^k \to Mx^*$. Finally, we complete the proof that $\{(y^k, p^k)\}$ converges to $\{(y^*, p^*)\}$ and $Mx^k \to Mx^*$.

(ii) When $\sigma_1 = \rho_1 - \rho_2\|M\|^2 - \mu > 0$, adding both sides of (15) from $k = 0$ to $k = \infty$, we get

$$\sigma_1 \sum_{k=0}^{\infty} \left\|x_e^{k+1}\right\|^2 \le H^0 - H^\infty < \infty,$$

hence $x^k \to x^*$.

When M has full column rank, i.e. $M^T M \succeq \gamma I$ for some $\gamma > 0$, then

$$\gamma \left\| x^{k+1} - x^* \right\| \leq \left\| M x_e^{k+1} \right\| = \left\| \frac{1}{\rho} (p^{k+1} - p^k) + y_e^{k+1} \right\| \leq \frac{1}{\rho} \left\| p^{k+1} - p^k \right\| + \left\| y_e^{k+1} \right\|.$$

Hence, we can also get $x^k \to x^*$ as follows from $p^{k+1} - p^k \to 0$ and $y^{k+1} - y^* \to 0$.

Last, due to Assumption 3.2 and the assertion (iii) of Definition 2.2, we have

$$\begin{aligned}
0 &= -M^T p^* + M^T p^* \\
&\in \partial f(x^*) + M^T \partial g(y^*) \\
&= \partial f(x^*) + M^T \partial(g)(M x^*) \\
&= \partial (f(x^*) + g(M x^*)),
\end{aligned}$$

i.e., x^* is a critical point of $f(x) + g(Mx)$, then x^* is the solution of (1) because of the convexity of $f(x) + g(Mx)$ and x^k converges to an optimal solution of $f(x) + g(Mx)$. □

Remark 2 The boundedness assumption of the iterative sequence x^k can be verified if one of the following three conditions is satisfied: (a) $\sigma_1 = \rho_1 - \rho_2 \|M\|^2 - \mu > 0$; (b) M has full column rank; (c) f is coercive.

How to tune the penalty parameter ρ in ADMM is a big issue and we expect the domain range of parameter ρ to be tuned to be as wide as possible; however, in most cases (even for convex ADMM algorithm), we have to accept the fact that only a relatively smaller range of ρ can be available. In general, we have the largest range (i.e. $(0, +\infty)$) of ρ to two-blocks convex optimization problems. For multi-block convex optimization problems, ρ has to be chosen to be larger than zero but less than a given constant, even under the assumption that one function is strongly convex as shown in [23]. In the context of nonconvexity, to ensure the sufficient descent of the selected Lyapunov function, ρ has to be larger than a constant that depends on the given regularity conditions [14–17]. Turning to the SWCCO problem (2) addressed in this paper, we will set ρ to be larger than at least twice the weakly convex modulus, i.e. $\rho > 2\rho_2$ in order to guarantee the nonincreasing of the Lyapunov function. We say this condition is necessary and the following example shows that the iterative sequence will diverge if $\rho \leq 2\rho_2$.

Example 3.3 Consider the optimization problem

$$\min_x \frac{a}{2} x^2 - \frac{b}{2} x^2,$$

where $\rho_1 = a > b = \rho_2 > 0$. We use ADMM (3) to solve this question, which is equivalent to

$$\begin{cases} \min_{x,y} \frac{a}{2} x^2 + (-\frac{b}{2}) y^2, \\ s.t. \quad x = y. \end{cases}$$

Note that $\frac{a}{2} x^2$ is strongly convex with modulus a and $(-\frac{b}{2}) y^2$ is weakly convex with modulus 1, respectively. Therefore, from ADMM (3), we can get

$$0 = a x^{k+1} + p^k + \rho (x^{k+1} - y^k);$$

$$0 = -by^{k+1} - p^{k+1};$$

$$p^{k+1} = p^k + \rho\left(x^{k+1} - y^{k+1}\right).$$

Rearranging the above equations, we get

$$y^{k+1} = \frac{1}{\rho - b}\left(\rho\frac{b+\rho}{a+\rho} - b\right)y^k.$$

It is obvious that y^k will diverge if $|\frac{1}{\rho-b}(\rho\frac{b+\rho}{a+\rho} - b)| > 1$. Next, we let $a \to \infty$ and $\rho \in (b, 2b)$, we can get $|\frac{1}{\rho-b}(\rho\frac{b+\rho}{a+\rho} - b)| > 1$, so y^k diverges.

Why should the penalty parameter ρ be larger than a given parameter in the nonconvex case in order to ensure convergence of the ADMM algorithm? This is an interesting and open problem for further research. In comparison with the case that both of the two objective functions are convex, the penalty parameter of which is merely greater than zero, the SWCCO optimization problem essentially is a convex question, but the setting of the penalty parameter ρ also follows the rules of the nonconvex case, i.e. it is greater than a given positive constant. The main reason for this setting is that the SWCCO problem involves a nonconvex term.

4 Sublinear and linear convergence rate analysis

Compared with the large amount of convergence analysis results for convex/nonconvex ADMM algorithms, there are merely a limited number of references in the literature [24–30] that have investigated the convergence rate as regards the convex optimization problems, not to speak of the nonconvex optimization problems. Specifically, the worst-case $O(1/k)$ convergence rate is established in [25] for the classic ADMM. The authors in [24] have investigated the dual objective function of the classic ADMM admitting $O(1/k)$ and $O(1/k^2)$ convergence rate for the accelerated version. Very recently, the authors in [26] have established the $o(1/k)$ convergence rate to multi-blocks ADMM and the linear convergence rate to multi-block ADMM is established in [28, 29] but requiring a strongly convex, gradient Lipschitz continuous function or some additional assumptions for the objective function. Without smoothness assumptions for the objective function, the authors in [27] have established Q-linear convergence for the more general convex piecewise linear-quadratic programming problems solved by the ADMM and the linearized ADMM, respectively. Subsequently, this global Q-linear convergence rate has been extended to a general semi-proximal ADMM in [30] for solving convex composite piecewise linear-quadratic programming and quadratic semidefinite programming. In this section, for the SWCCO-ADMM algorithm (3) applied to problem (1), we will show the sublinear convergence rate $o(1/k)$ only under Assumption 3.1 and linear convergence rate result under further assumptions.

4.1 Sublinear convergence rate analysis

In this section, we extend the sublinear convergence rate results of the multi-block convex ADMM in [26] to the SWCCO-ADMM (3) motivated by the preceding observation that the *primal residual* $Mx^{k+1} - y^{k+1}$ (*or* $p^{k+1} - p^k$) and *dual residual* $y^{k+1} - y^k$ can be used to measure the optimality of the iterations of the SWCCO-ADMM. For simplicity, we

denote $\bar{x}_e^{k+1} = x^{k+1} - x^k$, $\bar{y}_e^{k+1} = y^{k+1} - y^k$ and $\bar{p}_e^{k+1} = p^{k+1} - p^k$. Let us start the proof with a basic lemma.

Lemma 4.1 *If a nonnegative sequence $\{a^k\}_{k=0}^{k=\infty} \subset R$ is monotonically nonincreasing and obeys $\sum_{k=0}^{\infty} a^k < \infty$, then $\{a^k\}_{k=0}^{k=\infty}$ enjoys $o(1/k)$ sublinear convergence rate.*

Proof Adding a^k from $k = t$ to $k = 2t$,

$$ta^{2t} \le \sum_{k=t}^{2t} a^k = \sum_{k=0}^{2t} a^k - \sum_{k=0}^{t} a^k.$$

Then taking $t \to \infty$, we have $\sum_{k=0}^{2t} a^k - \sum_{k=0}^{t} a^k \to 0$; therefore, $a^k = o(1/k)$. □

Thus, the key step to proving the sublinear convergence of SWCCO-ADMM is to verify that $h^k = \frac{\rho}{2}\|y^{k+1} - y^k\|^2 + \frac{1}{2\rho}\|p^{k+1} - p^k\|^2$ is monotonically nonincreasing and $\sum_{k=0}^{\infty} h^k$ is bounded.

Theorem 4.2 *Suppose Assumption 3.1 holds, then the iterative sequences $\{(x^k, y^k, p^k)\}$ generated by SWCCO-ADMM (3) admits*

$$\frac{\rho}{2}\|y^{k+1} - y^k\|^2 + \frac{1}{2\rho}\|p^{k+1} - p^k\|^2 = o(1/k). \tag{27}$$

Proof Firstly, we will prove $\frac{\rho}{2}\|y^{k+1} - y^k\|^2 + \frac{1}{2\rho}\|p^{k+1} - p^k\|^2$ is monotonically nonincreasing. By the optimality condition (4), we get $-M^T p^{k+1} - \rho M^T(y^{k+1} - y^k) \in \partial f(x^{k+1})$ and $p^{k+1} \in \partial g(y^{k+1})$. From the strong convexity property (7) of f and the weakly convex property (8) of g, we obtain

$$\langle -M^T \bar{p}_e^{k+1} - \rho M^T(\bar{y}_e^{k+1} - \bar{y}_e^k), \bar{x}_e^{k+1} \rangle = \langle -\bar{p}_e^{k+1} - \rho(\bar{y}_e^{k+1} - \bar{y}_e^k), M\bar{x}_e^{k+1} \rangle \ge \rho_1 \|\bar{x}_e^{k+1}\|^2;$$

$$\langle \bar{p}_e^{k+1}, \bar{y}_e^{k+1} \rangle \ge -\rho_2 \|\bar{y}_e^{k+1}\|^2.$$

Adding the above two relations and rearranging them, we get

$$
\begin{aligned}
&-\langle \bar{p}_e^{k+1}, M\bar{x}_e^{k+1} - \bar{y}_e^{k+1} \rangle - \rho\langle \bar{y}_e^{k+1} - \bar{y}_e^k, M\bar{x}_e^{k+1} \rangle \\
&= -\frac{1}{\rho}\langle \bar{p}_e^{k+1}, \bar{p}_e^{k+1} - \bar{p}_e^k \rangle - \rho\langle \bar{y}_e^{k+1}, M\bar{x}_e^{k+1} \rangle + \rho\langle \bar{y}_e^k, M\bar{x}_e^{k+1} \rangle \\
&= -\frac{1}{2\rho}\left(\|\bar{p}_e^{k+1}\|^2 + \|\bar{p}_e^{k+1} - \bar{p}_e^k\|^2 - \|\bar{p}_e^k\|^2\right) \\
&\quad -\frac{\rho}{2}\left(\|\bar{y}_e^{k+1}\|^2 + \|M\bar{x}_e^{k+1}\|^2 - \|M\bar{x}_e^{k+1} - \bar{y}_e^{k+1}\|^2\right) \\
&\quad +\frac{\rho}{2}\left(\|\bar{y}_e^k\|^2 + \|M\bar{x}_e^{k+1}\|^2 - \|M\bar{x}_e^{k+1} - \bar{y}_e^k\|^2\right) \\
&= \frac{1}{2\rho}\left(\|\bar{p}_e^k\|^2 - \|\bar{p}_e^{k+1}\|^2\right) + \frac{\rho}{2}\left(\|\bar{y}_e^k\|^2 - \|\bar{y}_e^{k+1}\|^2\right) - \frac{\rho}{2}\|M\bar{x}_e^{k+1} - \bar{y}_e^k\|^2 \\
&\ge \rho_1 \|\bar{x}_e^{k+1}\|^2 - \rho_2 \|\bar{y}_e^{k+1}\|^2,
\end{aligned}
$$

where the first and third equations follow from the relation $p^{k+1} = p^k + \rho(Mx^{k+1} - y^{k+1})$; the second equation follows from the relation $2\langle a, b \rangle = a^2 + b^2 - (a - b)^2$.

Let $h^k = \frac{1}{2\rho}\|\bar{p}_e^k\|^2 + \frac{\rho}{2}\|\bar{y}_e^k\|^2$. From the above inequality, we get

$$h^k - h^{k+1} \geq \rho_1 \|\bar{x}_e^{k+1}\|^2 - \rho_2 \|\bar{y}_e^{k+1}\|^2 + \frac{\rho}{2}\|M\bar{x}_e^{k+1} - \bar{y}_e^k\|^2.$$

Using similar proof methods as shown in Theorem 3.2, if follows that

$$h^k - h^{k+1} \geq \sigma_1 \|\bar{x}_e^{k+1}\|^2 + \sigma_2 \|\bar{y}_e^{k+1} - \bar{y}_e^k\|^2 + \sigma_3 \|\bar{p}_e^{k+1} - \bar{p}_e^k\|^2.$$

Since the right side of the above inequality is nonnegative, h^k is monotonically nonincreasing.

Next, we verify the boundedness of $\sum_{k=0}^{\infty} h^k$. From Lemma 3.1, we know

$$\sigma_2 \sum_{k=0}^{\infty} \|y^{k+1} - y^k\|^2 + \sigma_3 \sum_{k=0}^{\infty} \|p^{k+1} - p^k\|^2 < \infty.$$

Therefore,

$$\sum_{k=0}^{\infty} \frac{\rho}{2} \|y^{k+1} - y^k\|^2 < \infty \quad and \quad \sum_{k=0}^{\infty} \frac{1}{2\rho} \|p^{k+1} - p^k\|^2 < \infty.$$

Adding the above two relations, we have

$$\sum_{k=0}^{\infty} \frac{\rho}{2} \|y^{k+1} - y^k\|^2 + \frac{1}{2\rho} \|p^{k+1} - p^k\|^2 < \infty.$$

Hence we get the boundedness of $\sum_{k=0}^{\infty} h^k$ and the $o(1/k)$ sublinear convergence rate following from Lemma 4.1.   □

4.2 Linear convergence rate

For the convex generalized ADMM algorithm, the linear convergence rate has been established in [27, 29, 30] if appropriate regularity conditions are satisfied. In this section, we will investigate that the SWCCO-ADMM algorithm also admits a linear convergence rate based on some mild regularity conditions. The main idea to prove the linear convergence rate of an SWCCO-ADMM algorithm is to set up the relation

$$H^k \geq (1 + \tau)H^{k+1}, \tag{28}$$

where the parameter $\tau > 0$ and H^k is the Lyapunov function defined in Lemma 3.1. We first list the assumptions used to establish the linear convergence rate as follows.

Assumption 4.1 M is full column rank and $g(\cdot)$ is gradient Lipschitz continuous with modulus L_g; Assumption 3.1 holds, but with $\sigma_1 = \rho_1 - \rho_2\|M\|^2 - \mu > 0$ (not ≥ 0).

Assumption 4.2 M is invertible, $f(x)$ is gradient Lipschitz continuous with modulus L_f; Assumption 3.1 holds, but with $\sigma_1 = \rho_1 - \rho_2\|M\|^2 - \mu > 0$ (not ≥ 0).

Remark 3 When M is a wavelet transform and $g(x) = \log_2(1 + x^2)$, Assumption 4.1 will easily be satisfied. When M is the identity matrix and $f(x) = \frac{1}{2}\|v - Hx\|^2$, Assumption 4.2 will easily be satisfied. In fact, we can obtain a linear convergence rate if one of Assumption 4.1 and Assumption 4.2 is satisfied in Theorem 4.5. Assumption 3.1 ensures the convergence of the SWCCO-ADMM algorithm. Because $\|x_e^k\|^2$ is an important measure of $\|y_e^k\|^2$ and $\|p^k - p^{k-1}\|^2$, we assume the condition $\sigma_1 > 0$ to ensure $\sigma_1\|x_e^k\|^2 > 0$. From Theorem 4.5, it can be seen that if $\sigma_1 = 0$, then $H^k \geq (1 + \tau)H^{k+1}$ will become $H^k \geq H^{k+1}$, in this case, we will not obtain a linear convergence rate. Meanwhile, as shown in [29], the full rank of M and the gradient Lipschitz property of f and g are used to show a linear convergence rate for a convex generalized ADMM algorithm.

From Lemma 3.1, we first note that the monotonically nonincreasing inequality (15) as regards H^k holds under Assumptions 4.1 or 4.2. In order to establish Eq. (28), we just prove $H^k - H^{k+1} \geq \tau H^{k+1}$. By using inequality (10) with $\gamma = 1$, we first have the following relation:

$$
\begin{aligned}
\left\|y_e^{k+1}\right\|^2 &= \left\|Mx_e^{k+1} + \frac{1}{\rho}\left(p^k - p^{k+1}\right)\right\|^2 \\
&\leq 2\left\|Mx_e^{k+1}\right\|^2 + \frac{2}{\rho^2}\left\|p^{k+1} - p^k\right\|^2 \\
&\leq 2\|M\|^2\left\|x_e^{k+1}\right\|^2 + \frac{2}{\rho^2}\left\|p^{k+1} - p^k\right\|^2.
\end{aligned}
\tag{29}
$$

To show the linear convergence rate, we need the following lemma.

Lemma 4.3 *Suppose $g(y)$ is gradient Lipschitz continuous with modulus L_g, then the iterative sequences generated by SWCCO-ADMM* (3) *satisfy*

$$
\left\|p^{k+1} - p^*\right\|^2 \leq L_g^2\left\|y^{k+1} - y^*\right\|^2 \quad \left(i.e., \left\|p_e^{k+1}\right\|^2 \leq L_g^2\left\|y_e^{k+1}\right\|^2\right).
$$

Proof According to the optimality condition (4), we have $p^{k+1} = \nabla g(y^{k+1})$. Hence, noticing that $p^* = \nabla g(y^*)$, we obtain

$$
\left\|p^{k+1} - p^*\right\|^2 = \left\|\nabla g\left(y^{k+1}\right) - \nabla g\left(y^*\right)\right\|^2 \leq L_g^2\left\|y^{k+1} - y^*\right\|^2. \qquad \square
$$

Lemma 4.4 *Suppose M is full row rank and $f(x)$ is gradient Lipschitz continuous with modulus L_f, then the iterative sequences generated by SWCCO-ADMM* (3) *satisfy*

$$
\left\|p^{k+1} - p^*\right\|^2 \leq \frac{2L_f^2}{\lambda_{\min}(MM^T)}\left\|x^{k+1} - x^*\right\|^2 + \frac{2\rho^2\lambda_{\max}(MM^T)}{\lambda_{\min}(MM^T)}\left\|y^{k+1} - y^k\right\|^2.
$$

Proof According to the optimality condition (4), we have $0 = \nabla f(x^{k+1}) + M^T p^{k+1} + \rho M^T(y^{k+1} - y^k)$ and $0 = \nabla f(x^*) + M^T p^*$. Hence, we have

$$
\begin{aligned}
\lambda_{\min}\left(MM^T\right)\left\|p^{k+1} - p^*\right\|^2 &\leq \left\|M^T p^{k+1} - M^T p^*\right\|^2 \\
&= \left\|\nabla f\left(x^{k+1}\right) - \nabla f\left(x^*\right) + \rho M^T\left(y^{k+1} - y^k\right)\right\|^2
\end{aligned}
$$

$$\leq 2\left\|\nabla f\left(x^{k+1}\right) - \nabla f\left(x^*\right)\right\|^2 + 2\rho^2\left\|M^T\left(y^{k+1} - y^k\right)\right\|^2$$

$$\leq 2L_f^2\left\|x^{k+1} - x^*\right\|^2 + 2\rho^2\lambda_{\max}\left(MM^T\right)\left\|y^{k+1} - y^k\right\|^2,$$

where the second inequality follows from Eq. (10) with $\tau = 1$ and the last inequality follows from the fact that $\nabla f(x)$ is Lipschitz continuous with constant L_f. Since M is full row rank, so $\lambda_{\max}(MM^T) \geq \lambda_{\min}(MM^T) > 0$. Then, dividing both sides of the above inequality by $\lambda_{\min}(MM^T)$, we complete the proof. $\square$

Next, we state the linear convergence rate result of the SWCCO-ADMM algorithm in the following theorem.

Theorem 4.5 *Suppose Assumption* 4.1 *or Assumption* 4.2 *holds, the iterative sequences generated by SWCCO-ADMM* (3) *converge linearly to a critical point.*

Proof On the one hand, if Assumption 4.1 is set up, we have $\|M\|^2 > 0$ due to M being full column rank. Then

$$\frac{1}{2}\left\{\sigma_1\left\|x_e^{k+1}\right\|^2 + \sigma_2\left\|y^{k+1} - y^k\right\|^2 + \sigma_3\left\|p^{k+1} - p^k\right\|^2\right\}$$

$$\geq \frac{1}{2}\left\{\sigma_1\left\|x_e^{k+1}\right\|^2 + \sigma_3\left\|p^{k+1} - p^k\right\|^2\right\}$$

$$= \frac{1}{2}\left\{\left(\frac{\sigma_1}{2\|M\|^2}\frac{2}{\rho}\right)\left(2\|M\|^2\frac{\rho}{2}\right)\left\|x_e^{k+1}\right\|^2 + \left(\sigma_3\frac{\rho^2}{2}\frac{2}{\rho}\right)\left(\frac{2}{\rho^2}\frac{\rho}{2}\right)\left\|p^{k+1} - p^k\right\|^2\right\}$$

$$\geq \frac{\min\left\{\frac{\sigma_1}{2\|M\|^2}\frac{2}{\rho}, \sigma_3\frac{\rho^2}{2}\frac{2}{\rho}\right\}}{2}\left\{\left(2\|M\|^2\frac{\rho}{2}\right)\left\|x_e^{k+1}\right\|^2 + \left(\frac{2}{\rho^2}\frac{\rho}{2}\right)\left\|p^{k+1} - p^k\right\|^2\right\}$$

$$\geq \tau_1\frac{\rho}{2}\left\|y_e^{k+1}\right\|^2, \tag{30}$$

where $\tau_1 = \dfrac{\min\left\{\frac{\sigma_1}{2\|M\|^2}\frac{2}{\rho}, \sigma_3\frac{\rho^2}{2}\frac{2}{\rho}\right\}}{2}$ and the last inequality follows from Eq. (29). Since $\nabla g(y)$ is Lipschitz continuous with constant L_g, and by Lemma 4.3, we have

$$\frac{1}{2}\left\{\sigma_1\left\|x_e^{k+1}\right\|^2 + \sigma_2\left\|y^{k+1} - y^k\right\|^2 + \sigma_3\left\|p^{k+1} - p^k\right\|^2\right\} \geq \tau_1\frac{\rho}{2}\left\|y_e^{k+1}\right\|^2$$

$$\geq \tau_1\frac{\rho}{2}\frac{1}{L_g^2}2\rho\frac{1}{2\rho}\left\|p_e^{k+1}\right\|^2$$

$$= \tau_2\frac{1}{2\rho}\left\|p_e^{k+1}\right\|^2, \tag{31}$$

where $\tau_2 = \tau_1\frac{\rho}{2}\frac{1}{L_g^2}2\rho$. Adding (30) and (31), we have

$$\sigma_1\left\|x_e^{k+1}\right\|^2 + \sigma_2\left\|y^{k+1} - y^k\right\|^2 + \sigma_3\left\|p^{k+1} - p^k\right\|^2$$

$$= 2\frac{1}{2}\left\{\sigma_1\left\|x_e^{k+1}\right\|^2 + \sigma_2\left\|y^{k+1} - y^k\right\|^2 + \sigma_3\left\|p^{k+1} - p^k\right\|^2\right\}$$

$$\geq \tau_1\frac{\rho}{2}\left\|y_e^{k+1}\right\|^2 + \tau_2\frac{1}{2\rho}\left\|p_e^{k+1}\right\|^2$$

$$\geq \tau'\left(\frac{\rho}{2}\left\|y_e^k\right\|^2 + \frac{1}{2\rho}\left\|p_e^k\right\|^2\right) = \tau'H^{k+1},$$

where $\tau' = \min\{\tau_1, \tau_2\}$. So according to Lemma 3.1, we have

$$H^k - H^{k+1} \geq \sigma_1 \|x_e^{k+1}\|^2 + \sigma_2 \|y^{k+1} - y^k\|^2 + \sigma_3 \|p^{k+1} - p^k\|^2 \geq \tau' H^{k+1}.$$

Therefore,

$$H^k \geq (1 + \tau') H^{k+1},$$

which implies that the iterative sequences of SWCCO-ADMM (3) converge linearly to a critical point under Assumption 4.1.

One the other hand, if Assumption 4.2 is set up, then

$$
\begin{aligned}
&\frac{1}{2} \left\{ \sigma_1 \|x_e^{k+1}\|^2 + \sigma_2 \|y^{k+1} - y^k\|^2 + \sigma_3 \|p^{k+1} - p^k\|^2 \right\} \\
&\quad \geq \frac{1}{2} \left\{ \sigma_1 \|x_e^{k+1}\|^2 + \sigma_2 \|y^{k+1} - y^k\|^2 \right\} \\
&\quad = \frac{1}{2} \left\{ \left(\sigma_1 \frac{\lambda_{\min}(MM^T)}{2L_2^2} 2\rho \right) \left(\frac{2L_2^2}{\lambda_{\min}(MM^T)} \frac{1}{2\rho} \right) \|x^{k+1} - x^*\|^2 \right. \\
&\qquad \left. + \left(\sigma_2 \frac{\lambda_{\min}(MM^T)}{2\rho^2 \lambda_{\max}(MM^T)} 2\rho \right) \left(\frac{2\rho^2 \lambda_{\max}(MM^T)}{\lambda_{\min}(MM^T)} \frac{1}{2\rho} \right) \|y^{k+1} - y^k\|^2 \right\} \\
&\quad \geq \frac{\min\{\sigma_1 \frac{\lambda_{\min}(MM^T)}{2L_2^2} 2\rho, \sigma_2 \frac{\lambda_{\min}(MM^T)}{2\rho^2 \lambda_{\max}(MM^T)} 2\rho\}}{2} \left\{ \frac{2L_2^2}{\lambda_{\min}(MM^T)} \frac{1}{2\rho} \|x^{k+1} - x^*\|^2 \right. \\
&\qquad \left. + \frac{2\rho^2 \lambda_{\max}(MM^T)}{\lambda_{\min}(MM^T)} \frac{1}{2\rho} \|y^{k+1} - y^k\|^2 \right\} \\
&\quad \geq \tau_3 \frac{1}{2\rho} \|p^{k+1} - p^*\|^2 = \tau_3 \frac{1}{2\rho} \|p_e^{k+1}\|^2,
\end{aligned}
\tag{32}
$$

where $\tau_3 = \dfrac{\min\{\sigma_1 \frac{\lambda_{\min}(MM^T)}{2L_2^2} 2\rho, \sigma_2 \frac{\lambda_{\min}(MM^T)}{2\rho^2 \lambda_{\max}(MM^T)} 2\rho\}}{2}$ and the last inequality follows from Lemma 4.4. Then, since M is full column rank, (30) holds. Next, adding (30) and (32), we have

$$
\begin{aligned}
&\sigma_1 \|x_e^{k+1}\|^2 + \sigma_2 \|y^{k+1} - y^k\|^2 + \sigma_3 \|p^{k+1} - p^k\|^2 \\
&\quad = 2 \frac{1}{2} \left\{ \sigma_1 \|x_e^{k+1}\|^2 + \sigma_2 \|y^{k+1} - y^k\|^2 + \sigma_3 \|p^{k+1} - p^k\|^2 \right\} \\
&\quad \geq \tau_1 \frac{\rho}{2} \|y_e^{k+1}\|^2 + \tau_3 \frac{1}{2\rho} \|p_e^{k+1}\|^2 \\
&\quad \geq \tau'' \left(\frac{\rho}{2} \|y_e^{k+1}\|^2 + \frac{1}{2\rho} \|p_e^{k+1}\|^2 \right) = \tau'' H^{k+1},
\end{aligned}
$$

where $\tau'' = \min\{\tau_1, \tau_3\}$. Using Lemma 3.1 again, we have

$$H^k \geq (1 + \tau'') H^{k+1}.$$

Hence, the iterative sequences of SWCCO-ADMM (3) converge linearly to a critical point. $\square$

Remark 4 If $\sigma_1 = 0$, $\tau' = 0$ and $\tau'' = 0$.

5 Numerical example

In this section, in order to confirm the convergence and the efficiency of the suggested SWCCO-ADMM algorithm we carry out a numerical experiment on a realistic optimization problem of image deblurring, by comparing with the classical ADMM with l_1 penalty function. Mathematically, the problem of image deblurring can be formulated as

$$f = Hu + \eta, \tag{33}$$

where H is certain type of blur kernel, such as Gaussian or motion blur; η is a certain kind of additional disturbance, such as Gaussian noise with zero-mean and variance σ^2. The aim of image deblurring is to recover the latent high-quality image u from the degraded (i.e. noisy or blurred) image f.

The common sparse regularized based model to (33) is

$$\min_u \frac{1}{2}\|Hu - f\|_2^2 + \lambda \Phi(Wu), \quad \lambda > 0, \tag{34}$$

where W is wavelet frame transform satisfying $W^T W = I$; $\Phi(\cdot)$ is the regularizer aiming to induce sparse solutions, such as the classical convex regularizer and ℓ_1-norm $\|\cdot\|_1$. In this paper, we exploit the following nonconvex sparsity-inducing regularizer, i.e. the following a-weakly convex function $g(x)$ (a.k.a. the generalized minimax-concave penalty function):

$$g(x) = \begin{cases} \tau|x| - \frac{x^2}{2a} & if\ |x| < \frac{\tau}{a}, \\ \frac{\tau^2}{2a} & if\ |x| \geq \frac{\tau}{a}, \end{cases} \tag{35}$$

considered in [5] to induce a more unbiased sparse solution. The corresponding proximal mapping (a.k.a. firm shrinkage) to this weakly convex function can be written as
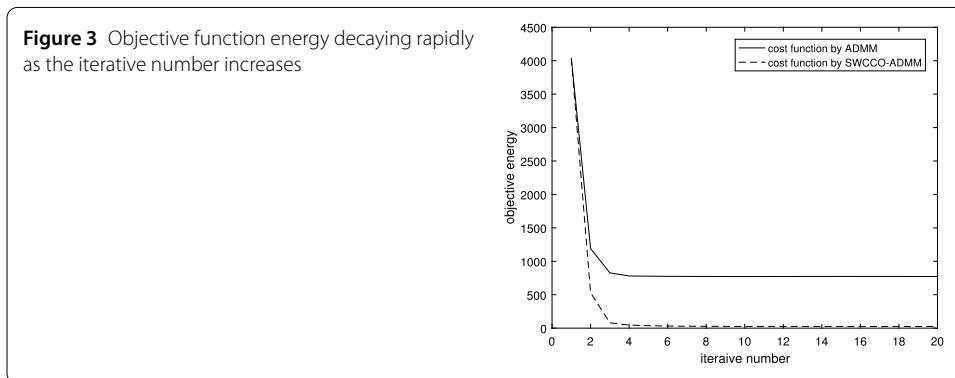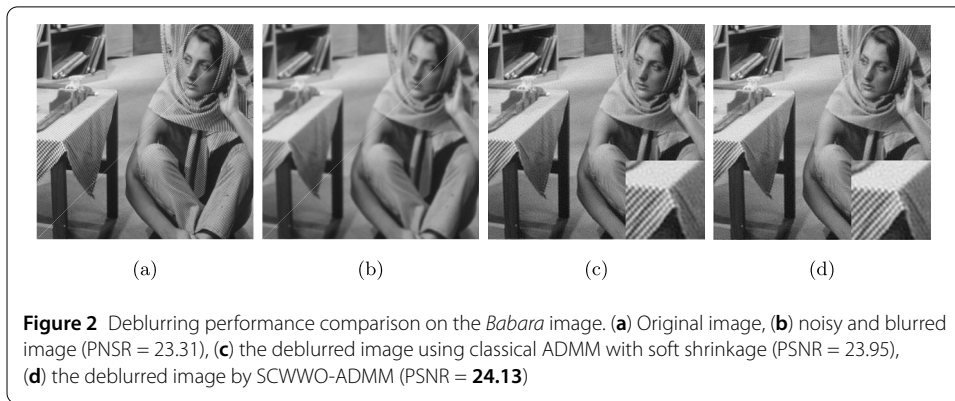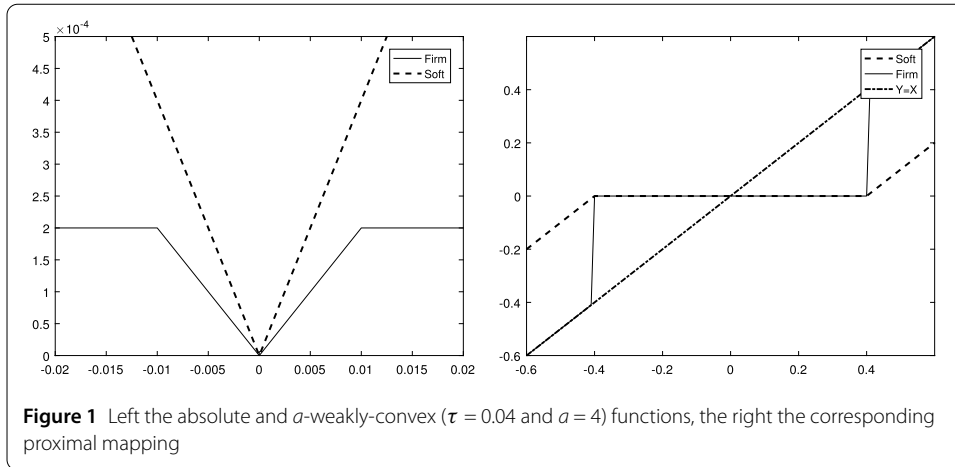
$$\mathrm{prox}_g(x) = \begin{cases} 0, & |x| < \lambda\tau, \\ (1 - \lambda a)^{-1}(x - \lambda\tau), & \lambda\tau \leq |x| < \frac{\tau}{a}, \\ x, & \frac{\tau}{a} \leq |x|, \end{cases} \tag{36}$$

where λ is the shrinkage parameter as set in the formulation of Eq. (34).

From Fig. 1 for the regularizers and the associated proximal mappings, we can see the firm shrinkage does not underestimate large magnitudes, i.e. the features (edges) of the solution u.

In the experiments, we set the standard deviation of Gaussian blurring kernel H to be 2, the variance σ_n^2 of Gaussian white noise η to be 2. In addition, we also compare our numerical results implemented using the proposed SWCCO-ADMM (3) with the classical ℓ_1-norm ADMM algorithms. Both algorithms are implemented on Matlab2014a with the same parameters setting as occur in the ADMM framework; and we set the iteration numbers to be 20. The performance evaluation criteria of the proposed algorithms are quantitatively measured by means of the peak signal-to-noise ratio (PSNR).

From Fig. 2, it can be seen that the proposed SCWWO-ADMM can achieve a better PNSR than the classical ADMM in the same parameters setting.

Figure 1 Left the absolute and *a*-weakly-convex ($\tau = 0.04$ and $a = 4$) functions, the right the corresponding proximal mapping

(a)　　　　　　　　(b)　　　　　　　　(c)　　　　　　　　(d)

Figure 2 Deblurring performance comparison on the *Babara* image. (**a**) Original image, (**b**) noisy and blurred image (PNSR = 23.31), (**c**) the deblurred image using classical ADMM with soft shrinkage (PSNR = 23.95), (**d**) the deblurred image by SCWWO-ADMM (PSNR = **24.13**)

Figure 3 Objective function energy decaying rapidly as the iterative number increases

The dependence of the energy of the objective function on the iterative number for a Barbara image shown in Fig. 3, which indicates that our proposed SWCCO-ADMM algorithm has convergence in practice and has a faster convergence rate than the classical ADMM.

6 Conclusions

In this paper, we have investigated the convergence for the alternating direction method of multipliers algorithm for the minimization of a combinational optimization problem, which consists of a strongly convex function and a weakly convex function. Meanwhile,

we also established the sublinear ($o(1/k)$) and linear convergence rate for this SWCCO-ADMM algorithm for the first time. In order to guarantee the algorithm convergence, the proof shows that the penalty parameter ρ has to be chosen larger than at least twice the weakly convex modulus. We state that this is because the existence of the weakly convex term in the objective function. To extend the SWCCO-ADMM to the multi-block composition of a strongly convex and a weakly convex function will be an interesting topic for the future research. The convergence analysis in this manuscript is based on the fact that the strong convexity dominates the weak convexity in the objective function, i.e. the objective function is strongly convex. A convergence analysis of the SWCCO-ADMM algorithm with a weak assumption is our future work.

Availability of data and materials
Not applicable.

Competing interests
The authors declare that they have no competing interests.

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

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

References
 1. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**(1), 17–40 (1976)
 2. Glowinski, R., Marrocco, A.: Sur l'approximation, par éléments finis d'ordre 1, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. J. Equine Vet. Sci. **2**(R2), 41–76 (1975)
 3. Bayram, I., Selesnick, I.W.: The Douglas–Rachford algorithm for weakly convex penalties. Oper. Res. Lett. **44**(3), 379–382 (2015)
 4. Möllenhoff, T., Strekalovskiy, E., Moeller, M., Cremers, D.: The primal-dual hybrid gradient method for semiconvex splittings. SIAM J. Imaging Sci. **8**(2), 827–857 (2015)
 5. Bayram, I.: On the convergence of the iterative shrinkage/thresholding algorithm with a weakly convex penalty. IEEE Trans. Signal Process. **64**(6), 1597–1608 (2016)
 6. Guo, K., Han, D.R., Yuan, X.M.: Convergence analysis of Douglas–Rachford splitting method for "strongly + weakly" convex programming. SIAM J. Numer. Anal. **55**(4), 1549–1577 (2017)
 7. Eckstein, J.: Splitting methods for monotone operators with applications to parallel optimization. Massachusetts Institute of Technology (1989)
 8. Eckstein, J., Wang, Y.: Understanding the convergence of the alternating direction method of multipliers: theoretical and computational perspectives. Pac. J. Optim. **11**(4), 619–644 (2015)
 9. Boyd, S., Parikh, N., Chu, E., Peleato, B.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)
10. Chen, L., Sun, D.F., Toh, K.C.: A note on the convergence of ADMM for linearly constrained convex optimization problems. Comput. Optim. Appl. **66**(2), 1–17 (2016)
11. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**(1–3), 293–318 (1992)
12. Fazel, M., Pong, T.K., Sun, D.F., Tseng, P.: Hankel matrix rank minimization with applications to system identification and realization. SIAM J. Matrix Anal. Appl. **34**(3), 946–977 (2013)
13. Attouch, H., Bolte, J., Svaiter, F.B.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods. Math. Program. **137**(1–2), 91–129 (2013)
14. Hong, M.Y., Luo, Z.Q., Razaviyayn, M.: Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. In: IEEE Int. Conf. Acoustics Speech. Signal Proc., pp. 3836–3840 (2015)
15. Li, G.Y., Pong, T.K.: Global convergence of splitting methods for nonconvex composite optimization. SIAM J. Optim. **25**(4), 2434–2460 (2015)

16. Wang, F.H.: Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems. Arxiv preprint arXiv:1410.8625 (2014)
17. Wang, Y., Yin, W.T., Zeng, J.S.: Global convergence of ADMM in nonconvex nonsmooth optimization. J. Sci. Comput., 1–35 (2018)
18. Beck, A., Tetruashvili, L.: On the convergence of block coordinate descent type methods. SIAM J. Optim. **23**(4), 2037–2060 (2013)
19. Sun, T., Barrio, R., Cheng, L., Jiang, H.: Precompact convergence of the nonconvex primal-dual hybird gradient algorithm. J. Comput. Appl. Math. **330**, 15–27 (2018)
20. Esser, E., Zhang, X., Chan, T.F.: A gernal framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. **3**(4), 1015–1046 (2010)
21. Rockafellar, R.T.: Convex Analysis, pp. 5–101. Princeton University Press
22. Rockafellar, R.T., Wets, J.B.R.: Variational Analysis. Springer
23. Cai, X.J., Han, D.R., Yuan, X.M.: On the convergence of the direct extension of ADMM for three-block separable convex minimization models with one strongly convex function. Comput. Optim. Appl. **66**(1), 39–73 (2017)
24. Goldstein, T., O'Donoghue, B., Setzer, S.: Fast alternating direction optimization methods. SIAM J. Imaging Sci. **7**(3), 225–231 (2014)
25. He, B.S., Yuan, X.M.: On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method. SIAM J. Numer. Anal. **50**(2), 700–709 (2012)
26. Deng, W., Lai, M.J., Peng, Z.M., Yin, W.T.: Parallel multi-block ADMM with $o(1/k)$ convergence. J. Sci. Comput. **71**(2), 712–736 (2017)
27. Yang, W.H., Han, D.R.: Linear convergence of the alternating direction method of multipliers for a class of convex optimization problems. SIAM J. Numer. Anal. **54**(2), 625–640 (2016)
28. Hong, M.Y., Luo, Z.Q.: On the linear convergence of the alternating direction method of multipliers. Math. Program. **162**(1–2), 1–35 (2012)
29. Deng, W.: On the global and linear convergence of the generalized alternating direction method of multipliers. J. Sci. Comput. **66**(3), 889–916 (2016)
30. Han, D.R., Sun, D.F.: Linear rate convergence of the alternating direction method of multipliers for convex composite problemming. Math. Oper. Res. **43**(2), 622–637 (2018)