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

Tao Zhang

School of Mathematics and Physics, University of Science and Technology Beijing, China, 100083

Zhengwei Shen[*]

School of Mathematics and Physics, University of Science and Technology Beijing, China, 100083

Abstract

The convergence analysis of the alternating direction method of multipliers (ADMM) methods to convex/nonconvex combinational optimization have been well established in literature. Consider the extensive applications of weakly convex function in signal processing and machine learning(e.g. *Special issue: DC-Theory, Algorithms and Applications, Mathematical Programming, 169(1):1-336,2018*), in this paper, we investigate the convergence analysis of ADMM algorithm to the strongly and weakly convex combinational optimization(SWCCO) problem. Specifically, we firstly build 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 to SWCCO-ADMM algorithm using the same conditions and the linear convergence rate by imposing gradient Lipschitz continuous condition to the objective function. The techniques used for convergence analysis in this paper is 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: nonconvex optimization; weakly convex function; ADMM; sublinear and linear convergence.

1. Introduction

---

[*]Corresponding author:ijwmip@ustb.edu.cn

1.1. Motivation and problem

The ADMM algorithm, as one of splitting/decoupling techniques, has been successfully exploited in a wide range of structured sparsity regularization optimization problems in machine learning and inverse problem. Of the most important issues in these applications focus on: (1) how to induce a reasonable sparse solution; (2) how to eliminate the disturbance(e.g. noise) while preserving the important features of the signal. Nonconvex penalty or regularization imposed on certain *a priori* being incorporated into optimization model can efficiently improve upon these two problems. In practice, however, it is hard to obtain a closed-form solution for generic nonconvex penalty except the weakly convex(a.k.a. semiconvex) penalty that including a large class of functions, e.g. difference-of-convex(DC) functions.

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

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

which 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[15, 16], to (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 is the penalty parameter and $p \in \mathbb{R}^n$ is the Lagrangian multiplier. The first-order optimality conditions to (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 $M^T(y^{k+1} - y^k)$, $Mx^{k+1} - y^{k+1}$ (or $p^{k+1} - p^k$) are dual and primal residual respectively.

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

1.2. Related works

In order to avoid solving a nonconvex optimization problem directly, intuitively, we can separate a part from the strongly convex term to the weakly convex term to make optimization problem (1) to be a convex-convex combination. However, it 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, linear operator involved in these terms will make such a separation very hard; and even if it can be separated, the recombined model will become more hard to be solved as shown in [3, 24].

The convergence analysis of the forward-backward splitting(FBS) and the Douglas-Rachford splitting(DRS) method to SWCCO problem with M is an identity map have been established in [2, 3, 18] very recently. Even though using DRS to the dual problem is equivalent to using ADMM to primal problem as shown in [10, 12] in the context of convex cases, this is not the case for nonconvex optimization problem, since the conjugate function of a nonconvex function has not been well defined yet in literature. Therefore, it is necessary to establish the convergence of SWCCO-ADMM algorithm.

In convex case, convergence analysis to ADMM has been well built in[5, 7, 11, 14]. For nonconvex case, however, it has merely been investigated limited to certain specific assumptions. This is because the iterative sequences (x^k, y^k, p^k) generated by ADMM to nonconvex optimization problem does not satisfy the Féjer monotonicity; moreover, the sufficient decrease condition [1] 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 nonconvex ADMM algorithm, some uncheckable or unreasonable assumptions have to be imposed to objective functions or to iterative sequences (x^k, y^k, p^k). In addition, global convergence or convergence rate of nonconvex optimization problem usually require $F(x)$ to satisfy the Kurdyka-Łojasiewicz (KL) property at each point, of which the KL exponent (i.e. the geometrical properties of the objective function around its stationary point) is not easily to be determined [1].

In fact, the key steps to establish convergence analysis to 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 $(p^{k+1} - p^k)$) goes to zero as $k \to \infty$. To this end, one common method developed in very recently papers[1, 22, 23, 28, 29] is to exploit the monotonically nonincreasing of certain type of Lyapunov function(e.g. the ALF[22, 23, 28] or its variant[29]) to measure the iterative error. Furthermore, the global convergence to the stationary point to nonconvex optimization problem can be obtained by using KL inequality. However, we have to emphasize that using the ALF or its variant to be the Lyapunov function has trouble to handle the nonsmooth objective functions, since Lyapunov function may no longer be monotone

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

Condition / Work	M	f, g	KL property	Lyapunov function
Hong [22]	$M = I$	f and g gradient Lipschitz continuous	*no*	ALF
Li [23]	full row rank	bounded Hessian of f	*yes*	ALF
Wang [28]	full row rank	f gradient Lipschitz continuous g strongly convex	*yes*	variants of ALF
Wang [29]	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 section 3)

decreasing in this case. Thus, in[22, 23, 28, 29], at least one or part of objective functions are required to be gradient Lipschitz continuous to guarantee algorithm convergence, which however will great limit the applications of the model (1).

On the other hand, the surjective(i.e. full row rank) assumption to linear mapping M, which is a necessary condition to prove monotonically nonincreasing of ALF in [22, 23, 28, 29], 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 data. Moreover, if we assume $f(x)$ to be gradient Lipschitz continuous and M is full row rank, then problem (2) actually can be solved using block coordinate descent methods(BCD) [4] and do not need ADMM algorithm. Specifically, suppose M to be full row rank, then there exists nonsingular matrix M_1 can split the constraint $Mx = y$ into the following formulation $M_1x_1 + M_2x_2 = y$. That is $x_1 = M_1^{-1}(y - M_2x_2)$. In this case, the constrained optimization problem (2) can be reformatted as the following unconstrained optimization problem

$$\min_{x_1,x_2,y} f(x_1, x_2) + g(y) = \underbrace{f(M_1^{-1}(y - M_2x_2), x_2)}_{h(x_2,y)} + g(y) \tag{5}$$

is equivalent to

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

which can be solve by BCD. The regularity conditions used in convergence analysis to nonconvex ADMM summarized in Table 1.

Very recently, a similar works [24, 27] compared with this manuscript, has been investigated to use the primal-dual hybrid gradient(PDHG) method to solve (1), and the convergence analysis is also well established in this literature. Although,

4

the PDHG in [24], which is a relaxed alternating minimization algorithm(AMA)[13], is apparently similar to the ADMM, they are in fact quite different. Actually, it is has a deep relationship with the inexact Uzawa method[13, 27].

1.3. Contributions

In this paper, without requiring gradient Lipschitz continuous of any(or part) of objective functions, we will establish the convergence analysis of SWCCO-ADMM algorithm (3) merely need M to be full column rank. We also establish the sublinear convergence rate with the similar assumptions of that used to convergence analysis, and establish the linear convergence rate of SWCCO-ADMM algorithm need some additional regularity assumptions. In addition, we use the nonnegative Lyapunov function defined in [5] to measure the iterative error instead of using ALF, and global convergence result can be obtained without using KL inequality. Thus, our proof is relatively more fundamental than the works in[22, 23, 28, 29].

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 dominate the weakly convex terms and the penalty parameter ρ is at least larger than two times of weakly convex modulus. We also give an example that the SWCCO-ADMM (3) will diverge if the later condition do not be satisfied. Meanwhile, we prove the iterative sequence $\{x^k\}$ generated by ADMM (3) converges to an optimal solution of (1).

- We build sublinear convergence rate $o(1/k)$ for the SWCCO-ADMM algorithm using the same regularity conditions as the convergence analysis. To the best our knowledge, this is the first sublinear convergence rate result for nonconvex ADMM algorithm.

- Furthermore, we establish the linear convergence rate to the SWCCO-ADMM algorithm by imposing the gradient Lipschitz continuous on one of the objective function.

The rest of this paper is organized in the following. In section 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 section 3. Next, we build the sublinear and linear convergence rate in section 4. Lastly, some conclusions are given in section 5.

2. Preliminaries

In this part, we list some fundamental definitions and notations for the following analysis.

5

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

Definition 2.1. *Function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is said to be strongly convex with modulus $\rho_1 > 0$ if $f(x) - \frac{\rho_1}{2}\|x\|^2$ is convex; function $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ 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 functions have the following properties [25]. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be 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 weakly convex function $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ 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. *[26] Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be proper lower semi-continuous function and finite at $\bar{x} \in \mathbb{R}^n$.*
(i) The Fréchet subdifferential of f at $\bar{x}$, written by $\hat{\partial} f(\bar{x})$, is the set

$$\{u \in \mathbb{R}^n : \liminf_{\substack{x \neq \bar{x} \\ x \to \bar{x}}} \frac{f(x) - f(\bar{x}) - \langle u, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0\}.$$

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

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

From this definition, let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a proper lower semi-continuous function we can get the following assertions.
(i) The subdifferential of f at $\bar{x}$ is the set

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

(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 $dom(f) \cap dom(g \circ M) \neq \varnothing$, then

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

6

Definition 2.3. *Let $S \neq \varnothing$ be the set of critical point of augmented Lagrangian function L_ρ. A triple $(x^*, y^*, z^*) \in S$ 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 build the 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*

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

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

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

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

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 augmented Lagrangian function satisfy,

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

Assumption 3.2. *$dom(f) \cap dom(g \circ M) \neq \varnothing$.*

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

7

For the sake of convenience, let a triple (x^*, y^*, p^*) be one of the critical point of 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 processing. Then by strongly convex property (7) of $f(x)$ and weakly convex 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 primal variable x can be regarded as an intermediate variable in the iterations of SWCCO-ADMM while y and p are the essential variables[5].

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) *satisfies*

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

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

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

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

Adding these two inequalities, then

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

8

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

$$\langle -M^T p_e^{k+1} - \rho M^T(y_e^{k+1} - y_e^k), x_e^{k+1} \rangle + \langle p_e^{k+1}, y_e^{k+1} \rangle$$
$$= -\langle p_e^{k+1}, M x_e^{k+1} \rangle - \rho \langle y_e^{k+1} - y_e^k, M x_e^{k+1} \rangle + \langle p_e^{k+1}, y_e^{k+1} \rangle$$
$$= -\frac{1}{\rho} \langle p_e^{k+1}, p_e^{k+1} - p_e^k \rangle - \rho \langle y_e^{k+1}, M x_e^{k+1} \rangle + \rho \langle y_e^k, M x_e^{k+1} \rangle$$
$$= -\frac{1}{2\rho}(\|p_e^{k+1}\|^2 + \|\bar{p}_e^{k+1} - p_e^k\|^2 - \|p_e^k\|^2) - \frac{\rho}{2}(\|y_e^{k+1}\|^2 + \|M x_e^{k+1}\|^2 - \|M x_e^{k+1} - y_e^{k+1}\|^2)$$
$$\quad + \frac{\rho}{2}(\|y_e^k\|^2 + \|M x_e^{k+1}\|^2 - \|M x_e^{k+1} - y_e^k\|^2)$$
$$= \frac{1}{2\rho}(\|p_e^k\|^2 - \|p_e^{k+1}\|^2) + \frac{\rho}{2}(\|y_e^k\|^2 - \|y_e^{k+1}\|^2) - \frac{\rho}{2}\|M x_e^{k+1} - y_e^k\|^2, \qquad (18)$$

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

$$H^k - H^{k+1} \geq \rho_1 \|x_e^{k+1}\|^2 - \rho_2 \|y_e^{k+1}\|^2 + \frac{\rho}{2}\|M x_e^{k+1} - y_e^k\|^2. \qquad (19)$$

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

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

$$\frac{\rho}{2}\|M x_e^{k+1} - y_e^k\|^2 = \frac{\rho}{2}\||M x_e^{k+1} - y_e^{k+1} + y_e^{k+1} - y_e^k\|^2$$
$$= \frac{\rho}{2}\||M x_e^{k+1} - y_e^{k+1}\|^2 + \frac{\rho}{2}\|y_e^{k+1} - y_e^k\|^2 + \rho \langle M x_e^{k+1} - y_e^{k+1}, y_e^{k+1} - y_e^k \rangle$$
$$= \frac{1}{2\rho}\|p^{k+1} - p^k\|^2 + \frac{\rho}{2}\|y^{k+1} - y^k\|^2 + \langle p^{k+1} - p^k, y^{k+1} - p^k \rangle$$
$$\geq \frac{1}{2\rho}\|p^{k+1} - p^k\|^2 + (\frac{\rho}{2} - \rho_2)\|y^{k+1} - y^k\|^2, \qquad (20)$$

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

$$\rho_2\|y_e^{k+1}\|^2 = \frac{\rho_2}{\rho^2}\|p^{k+1} - p^k - \rho M x_e^{k+1}\|^2$$
$$= \frac{\rho_2}{\rho^2}\|p^{k+1} - p^k\|^2 + \rho_2\|M x_e^{k+1}\|^2 - \frac{2\rho_2}{\rho}\langle p^{k+1} - p^k, M x_e^{k+1} \rangle. \qquad (21)$$

9

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

$$H^k - H^{k+1} \geq \rho_1 \|x_e^{k+1}\|^2 - \rho_2 \|Mx_e^{k+1}\|^2 + (\frac{1}{2\rho} - \frac{\rho_2}{\rho^2})\|p^{k+1} - p^k\|^2 + (\frac{\rho}{2} - \rho_2)\|y^{k+1} - y^k\|^2$$

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

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

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

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

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

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

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

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. $\square$

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

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

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

(ii) Under Assumption (3.2) and if one of the following conditions hold, 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 side of (15) in Lemma (3.1), from $k = 0$ to $k = \infty$, it follows that

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

10

Thus,
$$\|p^{k+1} - p^k\| \to 0 \quad and \quad \|y^{k+1} - y^k\| \to 0 \quad as \quad k \to \infty. \qquad (23)$$

Since $\{(x^*, y^*, p^*)\}$ is a critical point and $H^k \le H^0$ 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 suppose that $\{(\bar{x}, \bar{y}, \bar{p})\}$ be a accumulation point of $\{(x^k, y^k, p^k)\}$; hence, there exists a subsequence $\{(x^{k_j}, y^{k_j}, p^{k_j})\}$ 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
$$M\bar{x} - \bar{y} = 0. \qquad (24)$$

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

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

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

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

Taking limit of the above inequality, we get

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

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

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

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

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

This together with (24) and (25), we know $\{(\bar{x}, \bar{y}, \bar{p})\}$ 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 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

11

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 of $\{(y^k, p^k)\}$ converges to $\{(y^*, p^*)\}$ and $Mx^k \to Mx^*$.

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

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

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

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

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

Hence, we can also get $x^k \to x^*$ 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

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

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

Remark 3.2. *The boundedness assumption of the iterative sequence x^k can be verified if one of the following three conditions is satisfied, i.e. (a) $\sigma_1 = \rho_1 - \rho_2 \|M\|^2 - \mu > 0$; (b) M has full column rank; (c) f is coercive.*

How to tune penalty parameter ρ in ADMM is a big issue and we expect the domain range of parameter ρ to be tuned is as wide as possible; however, in most cases(even to 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-blocks 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 being strongly convex as shown in [6]. In the context of nonconvex, 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[22, 23, 28, 29]. Turn to the SWCCO problem (2) processing in this paper, we will set ρ at least larger than two times of weakly convex modulus, i.e. $\rho > 2\rho_2$ in order to guarantee the nonincreasing of Lyapunov function. We say this condition is necessary and the following example show that the iterative sequence will diverge if $\rho \leq 2\rho_2$.

Example 1. *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

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

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

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

Rearranging the above equations, we get

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

It's 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 the penalty parameter ρ should be larger than a given parameter in the nonconvex case in order to ensure ADMM algorithm convergence? This is an interesting and open problem for further research. In comparison to the case that both of the two objective functions are convex, the penalty parameter of which merely greater than zero, the SWCCO optimization problem essentially is a convex question, but the setting of penalty parameter ρ also follows the rules of the nonconvex case, i.e. greater than a given positive constant. The mainly 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 to convex/nonconvex ADMM algorithms, there are merely a limited number of literature[17, 20, 8, 30, 21, 9, 19] have investigated the convergence rate about the convex optimization problems, not to speak of the nonconvex optimization problems. Specifically, the worst-case $O(1/k)$ convergence rate is established in [20] to classic ADMM. The authors in [17] 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 [8] have established the $o(1/k)$ convergence rate to multi-blocks ADMM and the linear convergence rate to multi-block ADMM is established in[21, 9] but requiring strongly convex, gradient Lipschitz continuous or some additional assumptions to objective function. Without smooth assumptions to objective function, authors in [30] 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 [19] for solving convex composite piecewise linear-quadratic programming and quadratic semidefinite programming. In this section, for the SWCCO-ADMM algorithm (3) to problem(1), we will build sublinear convergence rate $o(1/k)$ only under the Assumption (3.1) and linear convergence rate result under the furthermore assumptions.

4.1. Sublinear convergence rate analysis

In this section, we extend the sublinear convergence rate results of multi-block convex ADMM in [8] to the SWCCO-ADMM (3) motivated by the preceding observation that *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's us star the proof with a basic lemma.

Lemma 4.1. *If a nonnegative sequence $\{a^k\}_{k=0}^{k=\infty} \subset \mathbb{R}$ 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 $k = 2t$,

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

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

Thus, the key step to prove 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.

Theroem 4.1. *Suppose Assumption 3.1 holds, then the iterative sequences $\{(x^k, y^k, p^k)\}$ generated by SWCCO-ADMM (3) admits*

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

14

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 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 strongly convex property (7) of f and the weakly convex property (8) of g, we obtain

$$\langle -M^T \bar{p}_e^{k+1} - \rho M^T(\bar{y}_e^{k+1} - \bar{y}_e^k), \bar{x}_e^{k+1} \rangle = \langle -\bar{p}_e^{k+1} - \rho(\bar{y}_e^{k+1} - \bar{y}_e^k), M\bar{x}_e^{k+1} \rangle \geq \rho_1 \|\bar{x}_e^{k+1}\|^2;$$
$$\langle \bar{p}_e^{k+1}, \bar{y}_e^{k+1} \rangle \geq -\rho_2 \|\bar{y}_e^{k+1}\|^2.$$

Adding the above two relations and rearranging them, we get

$$- \langle \bar{p}_e^{k+1}, M\bar{x}_e^{k+1} - \bar{y}_e^{k+1} \rangle - \rho\langle \bar{y}_e^{k+1} - \bar{y}_e^k, M\bar{x}_e^{k+1} \rangle$$
$$= -\frac{1}{\rho}\langle \bar{p}_e^{k+1}, \bar{p}_e^{k+1} - \bar{p}_e^k \rangle - \rho\langle \bar{y}_e^{k+1}, M\bar{x}_e^{k+1} \rangle + \rho\langle \bar{y}_e^k, M\bar{x}_e^{k+1} \rangle$$
$$= -\frac{1}{2\rho}(\|\bar{p}_e^{k+1}\|^2 + \|\bar{p}_e^{k+1} - \bar{p}_e^k\|^2 - \|\bar{p}_e^k\|^2) - \frac{\rho}{2}(\|\bar{y}_e^{k+1}\|^2 + \|M\bar{x}_e^{k+1}\|^2 - \|M\bar{x}_e^{k+1} - \bar{y}_e^{k+1}\|^2)$$
$$+ \frac{\rho}{2}(\|\bar{y}_e^k\|^2 + \|M\bar{x}_e^{k+1}\|^2 - \|M\bar{x}_e^{k+1} - \bar{y}_e^k\|^2)$$
$$= \frac{1}{2\rho}(\|\bar{p}_e^k\|^2 - \|\bar{p}_e^{k+1}\|^2) + \frac{\rho}{2}(\|\bar{y}_e^k\|^2 - \|\bar{y}_e^{k+1}\|^2) - \frac{\rho}{2}\|M\bar{x}_e^{k+1} - \bar{y}_e^k\|^2$$
$$\geq \rho_1 \|\bar{x}_e^{k+1}\|^2 - \rho_2 \|\bar{y}_e^{k+1}\|^2,$$

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 the similar proof ways as shown in Theorem 3.1, 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, hence, 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,

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

15

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 follows from Lemma 4.1. □

4.2. Linear convergence rate

To convex generalized ADMM algorithm, linear convergence rate has been established in [9, 19, 30] if appropriate regularity conditions are satisfied. In this section, we will investigate that the SWCCO-ADMM algorithm also admits linear convergence rate based on some mild regularity conditions. The main idea to prove linear convergence rate of SWCCO-ADMM algorithm is to set up the relation

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

where 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) .*

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

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

To build the linear convergence rate, we need the following Lemma.

Lemma 4.2. *Suppose $g(y)$ is gradient Lipschitz continuous with modulus L_g, then the iterative sequences generated by SWCCO-ADMM (3) satisfy*

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

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

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

$\square$

Lemma 4.3. *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*

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

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

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

where the second inequality follows from relation (10) with $\tau = 1$ and the last inequality follows from 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 SWCCO-ADMM algorithm in the following theorem.

Theroem 4.2. *Suppose Assumption 4.1 or Assumption 4.2 holds, then 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 is full column rank. Then,

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

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

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

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

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

where $\tau_1 = \dfrac{min\{\frac{\sigma_1}{2\|M\|^2}\frac{2}{\rho}, \sigma_3\frac{\rho^2}{2}\frac{2}{\rho}\}}{2}$ and the last inequality follows from relation (29). Since $\nabla g(y)$ is Lipschitz continuous with constant L_g , and by Lemma 4.2, we have

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

$$\geq \tau_1\frac{\rho}{2}\frac{1}{L_g^2}2\rho\frac{1}{2\rho}\|p_e^{k+1}\|^2 = \tau_2\frac{1}{2\rho}\|p_e^{k+1}\|^2 \tag{31}$$

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

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

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

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

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

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 the Assumption 4.1.

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

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

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

$$= \frac{1}{2}\left\{(\sigma_1\frac{\lambda_{min}(MM^T)}{2L_2^2}2\rho)(\frac{2L_2^2}{\lambda_{min}(MM^T)}\frac{1}{2\rho})\|x^{k+1} - x^*\|^2\right.$$

$$\left. + (\sigma_2\frac{\lambda_{min}(MM^T)}{2\rho^2\lambda_{max}(MM^T)}2\rho)(\frac{2\rho^2\lambda_{max}(MM^T)}{\lambda_{min}(MM^T)}\frac{1}{2\rho})\|y^{k+1} - y^k\|^2\right\}$$

$$\geq \frac{min\{\sigma_1\frac{\lambda_{min}(MM^T)}{2L_2^2}2\rho, \sigma_2\frac{\lambda_{min}(MM^T)}{2\rho^2\lambda_{max}(MM^T)}2\rho\}}{2}\left\{\frac{2L_2^2}{\lambda_{min}(MM^T)}\frac{1}{2\rho}\|x^{k+1} - x^*\|^2\right.$$

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

$$\geq \tau_3\frac{1}{2\rho}\|p^{k+1} - p^*\|^2 = \tau_3\frac{1}{2\rho}\|p_e^{k+1}\|^2. \tag{32}$$

where $\tau_3 = \frac{min\{\sigma_1\frac{\lambda_{min}(MM^T)}{2L_2^2}2\rho, \sigma_2\frac{\lambda_{min}(MM^T)}{2\rho^2\lambda_{max}(MM^T)}2\rho\}}{2}$ and the last inequality follows from Lemma 4.3. Then, since M is full column rank, so (30) holds. Next, adding (30) and (32), we have

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

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

$$\geq \tau_1\frac{\rho}{2}\|y_e^{k+1}\|^2 + \tau_3\frac{1}{2\rho}\|p_e^{k+1}\|^2$$

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

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. $\square$

5. Conclusions

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

function. Meanwhile, we also established 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 process showed that the penalty parameter ρ has to be chosen at least larger than two times of weakly convex modulus. We state that this is because the existence of weakly convex term in the objective function. To extend the SWCCO-ADMM to the multi-block composition of strongly convex and 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 strongly convexity dominates the weakly convexity in the objective function, i.e. the objective function is strongly convex. And, the convergence analysis of SWCCO-ADMM algorithm with a weak assumption is our future works.

References

[1] H. Attouch, J. Bolte, and F.B. Svaiter. 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.

[2] I. Bayram. On the convergence of the iterative shrinkage/thresholding algorithm with a weakly convex penalty. *IEEE Trans. Signal Proces.*, 64(6):1597–1608, 2016.

[3] I. Bayram and I.W. Selesnick. The douglas-rachford algorithm for weakly convex penalties. *Oper. Res. Lett.*, 44(3):379–382, 2015.

[4] A. Beck and L. Tetruashvili. On the convergence of block coordinate descent type methods. *SIAM J. Optim.*, 23(4):2037–2060, 2013.

[5] S. Boyd, N. Parikh, E. Chu, and B. Peleato. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[6] X.J. Cai, D.R. Han, and X.M. Yuan. 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.

[7] L. Chen, D.F. Sun, and K.C. Toh. A note on the convergence of admm for linearly constrained convex optimization problems. *Comput. Optim. Appl.*, 66(2):1–17, 2016.

[8] W. Deng, M.J. Lai, Z.M. Peng, and W.T. Yin. Parallel multi-block admm with o (1 / k) convergence. *J. Sci. Comput.*, 71(2):712–736, 2017.

[9] W. Deng and W.T. Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. *J. Sci. Comput.*, 66(3):889–916, 2016.

[10] J. Eckstein. Splitting methods for monotone operators with applications to parallel optimization. *Massachusetts Institute of Technology*, 1989.

[11] J. Eckstein and D.P. Bertsekas. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.*, 55(1-3):293–318, 1992.

[12] J. Eckstein and Y. Wang. Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. *Pac. J. Optim.*, 11(4):619–644, 2015.

[13] E. Esser, X. Zhang, and T.F. Chan. 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.

[14] M. Fazel, T.K. Pong, D.F. Sun, and P. Tseng. Hankel matrix rank minimization with applications to system identification and realization. *SIAM J. Matrix Anal. Appl.*, 34(3):946–977, 2013.

[15] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.*, 2(1):17–40, 1976.

[16] R. Glowinski and A. Marrocco. 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(R-2):41–76, 1975.

[17] T. Goldstein, B. O'Donoghue, and S. Setzer. Fast alternating direction optimization methods. *SIAM J. Imaging Sci.*, 7(3):225–231, 2014.

[18] K. Guo, D.R. Han, and X.M. Yuan. Convergence analysis of douglas-rachford splitting method for "strongly + weakly" convex programming. *SIAM J. Numerical Anal.*, 55(4):1549–1577, 2017.

[19] D. R. Han, D.F. Sun, and L.W.Zhang. Linear rate convergence of the alternating direction method of multipliers for convex composite problemming. *Math. Oper. Res.*, 43(2):622–637, 2018.

[20] B.S. He and X.M. Yuan. On the $o(1/n)$ convergence rate of the douglas-rachford alternating direction method. *SIAM J. Numerical Anal.*, 50(2):700–709, 2012.

[21] M.Y. Hong and Z.Q. Luo. On the linear convergence of the alternating direction method of multipliers. *Math. Program.*, 162(1-2):1–35, 2012.

[22] M.Y. Hong, Z.Q. Luo, and M. Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. In *IEEE Int. Conf. Acoustics Speech. Signal Proc.*, pages 3836–3840, 2015.

[23] G.Y. Li and T.K. Pong. Global convergence of splitting methods for nonconvex composite optimization. *SIAM J. Optim.*, 25(4):2434–2460, 2015.

[24] T. Möllenhoff, E. Strekalovskiy, M. Moeller, and D. Cremers. The primal-dual hybrid gradient method for semiconvex splittings. *SIAM J. Imaging Sci.*, 8(2):827–857, 2015.

[25] R.T. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[26] R.T. Rockafellar and J.B. Roger Wets. *Variational analysis.* Springer Science & Business Media, 2009.

[27] T.Sun, R.Barrio, L.Cheng, and H.Jiang. Precompact convergence of the nonconvex primal-dual hybird gradient algorithm. *J. Compu. Appl. Math.*, 330:15–27, 2018.

[28] F.H. Wang, Z.B. Xu, and H.K. Xu. Convergence of bregman alternating direction method with multipliers for nonconvex composite problems. *arXiv preprint arXiv:1410.8625*, 2014.

[29] Y. Wang, W.T. Yin, and J.S. Zeng. Global convergence of admm in nonconvex nonsmooth optimization. *J. Sci. Comput.*, pages 1–35, 2018.

[30] W.H. Yang and D.R. Han. Linear convergence of the alternating direction method of multipliers for a class of convex optimization problems. *SIAM J. Numerical Ana.*, 54(2):625–640, 2016.