

ON THE LINEAR CONVERGENCE RATE OF GENERALIZED ADMM FOR CONVEX COMPOSITE PROGRAMMING

HAN WANG¹, PEILI LI², YUNHAI XIAO^{2,*}

¹*School of Mathematics and Statistics, Henan University, Kaifeng 475000, China*

²*Center for Applied Mathematics of Henan Province, Henan University, Kaifeng 475000, China*

Abstract. Over the fast few years, the numerical success of the generalized alternating direction method of multipliers (GADMM) proposed by Eckstein and Bertsekas [Math. Program. 1992] has inspired intensive attention in analyzing its theoretical convergence properties. This paper is devoted to the linear convergence rate of the semi-proximal GADMM (sPGADMM) for solving linearly constrained convex composite optimization problems. The semi-proximal terms contained in each subproblem possess the abilities of handling with multi-block problems efficiently. We initially present some important inequalities for the sequence generated by the sPGADMM, and then establish the local linear convergence rate under the assumption of calmness. As a by-product, the global convergence property is also discussed.

Keywords. Alternating direction method of multipliers; Calmness; Error bound condition; Linear convergence rate; Semi-proximal terms.

2020 Mathematics Subject Classification. 90C25.

1. INTRODUCTION

Let \mathcal{X} , \mathcal{Y} , and \mathcal{Z} be finite-dimensional real Euclidean spaces each equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|$, in which, $\mathcal{Y} := \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$ and $\mathcal{Z} := \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_n$ are the Cartesian product of some finite-dimensional real Euclidean spaces. We consider the following linearly constrained convex composite optimization problem

$$\begin{aligned} \min \quad & f_1(y_1) + f_2(y_1, y_2, \dots, y_m) + g_1(z_1) + g_2(z_1, z_2, \dots, z_n) \\ \text{s.t.} \quad & \mathcal{A}_1^* y_1 + \mathcal{A}_2^* y_2 + \dots + \mathcal{A}_m^* y_m + \mathcal{B}_1^* z_1 + \mathcal{B}_2^* z_2 + \dots + \mathcal{B}_n^* z_n = c, \end{aligned} \tag{1.1}$$

where $f_1 : \mathcal{Y}_1 \rightarrow (-\infty, +\infty]$ and $g_1 : \mathcal{Z}_1 \rightarrow (-\infty, +\infty]$ are simple closed proper convex (not necessarily smooth) functions; $f_2 : \mathcal{Y} \rightarrow (-\infty, +\infty)$ and $g_2 : \mathcal{Z} \rightarrow (-\infty, +\infty)$ are continuously differentiable convex quadratic functions; $\mathcal{A}_i^* : \mathcal{Y}_i \rightarrow \mathcal{X}$ and $\mathcal{B}_i^* : \mathcal{Z}_i \rightarrow \mathcal{X}$ are the adjoints of the linear operators $\mathcal{A}_i : \mathcal{X} \rightarrow \mathcal{Y}_i$ and $\mathcal{B}_i : \mathcal{X} \rightarrow \mathcal{Z}_i$, respectively; $c \in \mathcal{X}$ is a given data. For convenience, we denote the linear maps $\mathcal{A} : \mathcal{X} \rightarrow \mathcal{Y}$ and $\mathcal{B} : \mathcal{X} \rightarrow \mathcal{Z}$ such that their adjoint

*Corresponding author.

E-mail address: whmath@henu.edu.cn (H. Wang), lipeli@henu.edu.cn (P. Li), yhxiao@henu.edu.cn (Y. Xiao).
 Received 17 September 2023; Accepted 22 November 2023; Published online 5 March 2024.

maps are given by

$$\mathcal{A}^*y = \sum_{i=1}^m \mathcal{A}_i^*y_i, \quad \forall y \in \mathcal{Y}, \quad \mathcal{B}^*z = \sum_{j=1}^n \mathcal{B}_j^*z_j, \quad \forall z \in \mathcal{Z}.$$

Accordingly, model (1.1) could be transformed into the following form

$$\begin{aligned} \min \quad & f(y) + g(z) \\ \text{s.t.} \quad & \mathcal{A}^*y + \mathcal{B}^*z = c, \end{aligned} \quad (1.2)$$

where $f(y) := f_1(y_1) + f_2(y_1, y_2, \dots, y_m)$ and $g(z) := g_1(z_1) + g_2(z_1, z_2, \dots, z_n)$ are closed proper convex functions in the form of ‘nonsmooth+quadratic’.

Given its structure, problem (1.2) covers a wide range of apparently related formulations in different scientific fields, including deep learning, compressive sensing, sparse coefficient estimation, etc. For solving (1.2), a frequently used benchmark is the alternating direction method of multipliers (ADMM) which was originally proposed by Glowinski & Marroco [19] and Gabay & Mercier [17]. Starting from $(x^0, y^0, z^0) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, the iterative scheme of ADMM for solving (1.2) takes the following form

$$\begin{cases} y^{k+1} = \arg \min_{y \in \mathcal{Y}} f(y) - \langle \mathcal{A}x^k, y \rangle + \frac{\sigma}{2} \|\mathcal{A}^*y + \mathcal{B}^*z^k - c\|^2, \\ z^{k+1} = \arg \min_{z \in \mathcal{Z}} g(z) - \langle \mathcal{B}x^k, z \rangle + \frac{\sigma}{2} \|\mathcal{A}^*y^{k+1} + \mathcal{B}^*z - c\|^2, \\ x^{k+1} = x^k - \tau \sigma (\mathcal{A}^*y^{k+1} + \mathcal{B}^*z^{k+1} - c), \end{cases} \quad (1.3)$$

where $x \in \mathcal{X}$ is a multiplier, $\sigma > 0$ is a penalty parameter, and τ is a steplength within the interval $(0, (1 + \sqrt{5})/2)$. Incidentally, iteration scheme (1.3) is not always well-defined. One can consult a counter-example in [6] for more details.

Interestingly, Gabay [16] showed that the type of ADMM (1.3) with a unit steplength is equivalent to the Douglas-Rachford splitting (DRs) method [9, 26] for the sum of two maximal monotone operators. And then, Eckstein & Bertsekas [12] showed that the DRs is actually an application of the proximal point algorithm (PPA) [32, 33]. For the historical developments and applications of ADMM, one may refer to [4, 13, 18, 39]. In light of this, Eckstein & Bertsekas [12] considered a variant of PPA and then applied it on the minimization problem (1.2) which leads to the following generalized version of ADMM (GADMM):

$$\begin{cases} y^{k+1} = \arg \min_{y \in \mathcal{Y}} f(y) - \langle \mathcal{A}x^k, y \rangle + \frac{\sigma}{2} \|\mathcal{A}^*y + \mathcal{B}^*z^k - c\|^2, \\ z^{k+1} = \arg \min_{z \in \mathcal{Z}} g(z) - \langle \mathcal{B}x^k, z \rangle + \frac{\sigma}{2} \|\rho(\mathcal{A}^*y^{k+1} + \mathcal{B}^*z^k - c) + \mathcal{B}^*(z - z^k)\|^2, \\ x^{k+1} = x^k - \sigma (\rho(\mathcal{A}^*y^{k+1} + \mathcal{B}^*z^k - c) + \mathcal{B}^*(z^{k+1} - z^k)), \end{cases} \quad (1.4)$$

where $\rho \in (0, 2)$ is a relaxation factor. It is quite clear that the GADMM (1.4) with a unit relaxation factor is just the case of ADMM (1.3) with $\tau = 1$. As pointed out by Adona et al. [1] that, the GADMM is actually an instance of a hybrid proximal extra-gradient framework of Monteiro & Svaiter [27].

Comparing with (1.3) and (1.4), we see that GADMM still retains the benefits of ADMM in treating f and g separately, but the relaxation factor may make this method more flexible and efficient. Some earlier numerical studies of GADMM in this shell can be found in [2, 10].

Eckstien [11] suggested adding proximal terms to the subproblems contained in (1.3) to make them easier to solve. Then, Fazel et al. [15] proved that these proximal terms can be relaxed to positive and semi-definite operators (named semi-proximal terms). It should be emphasized that the semi-proximal terms allow the subproblems in (1.3) to be decomposed into some smaller ones and then solved individually. For example, note $f(y) = f_1(y_1) + f_2(y_1, y_2, \dots, y_m)$, then the y -subproblem can be attained via the following process

$$y_m \rightarrow y_{m-1} \rightarrow \dots \rightarrow y_2 \rightarrow y_1 \rightarrow y_2 \dots \rightarrow y_{m-1} \rightarrow y_m \quad (1.5)$$

by fixing other blocks with their latest values, which means that (1.3) has the advantage to resolve the potentially nonsolvability issue of the subproblems. This updating order is called symmetric Gauss-Seidel (sGS) iteration which was firstly appeared in [36] and later was analyzed by Li et al. [23, 25]. For more details on the developments and applications of sGS, one may refer to the Ph.D. thesis [22] and the papers of [21, 5, 24, 40]. From [25, Theorem 1], we know that this updating order (1.5) is actually equivalent to adding a special semi-proximal term, and hence, the convergence can be easily obtained from Fazel et al. [15, Theorem B.1].

Inspired by the success of sGS technique, it is natural to add a pair of semi-proximal terms to the subproblems in (1.4) so that GADMM possesses the abilities of handling multi-block problems in the form of (1.1). Actually, this idea was firstly attempted by Fang et al. [14] in the sense of adding a proximal term to the z -subproblem in (1.4), but the y -subproblem was ignored. This defect was remedied immediately by Adona et al. [1]. Let $\mathcal{S} : \mathcal{Y} \rightarrow \mathcal{Y}$ and $\mathcal{T} : \mathcal{Z} \rightarrow \mathcal{Z}$ be a pair of self-adjoint positive semidefinite (not necessarily positive definite) linear operators, the generalized ADMM with semi-proximal terms (abbr. sPGADMM) for solving the convex composite programming (1.2) can be described as follows:

Algorithm sPGADMM:

Step 0: Let $\sigma \in (0, +\infty)$ and $\rho \in (0, 2)$ be given parameters. Choose $(y^0, z^0, x^0) \in \text{dom}(f) \times \text{dom}(g) \times \mathcal{X}$. For $k = 0, 1, 2, \dots$, do the following steps iteratively:

Step 1: Compute

$$\left\{ \begin{array}{l} y^{k+1} = \arg \min_{y \in \mathcal{Y}} f(y) - \langle \mathcal{A}x^k, y \rangle + \frac{\sigma}{2} \|\mathcal{A}^*y + \mathcal{B}^*z^k - c\|^2 \\ \quad + \frac{1}{2} \|y - y^k\|_{\mathcal{S}}^2, \end{array} \right. \quad (1.6a)$$

$$\left\{ \begin{array}{l} z^{k+1} = \arg \min_{z \in \mathcal{Z}} g(z) - \langle \mathcal{B}x^k, z \rangle + \frac{1}{2} \|z - z^k\|_{\mathcal{T}}^2 \\ \quad + \frac{\sigma}{2} \|\rho(\mathcal{A}^*y^{k+1} + \mathcal{B}^*z^k - c) + \mathcal{B}^*(z - z^k)\|^2, \end{array} \right. \quad (1.6b)$$

$$\left\{ \begin{array}{l} x^{k+1} := x^k - \sigma \left(\rho(\mathcal{A}^*y^{k+1} + \mathcal{B}^*z^k - c) + \mathcal{B}^*(z^{k+1} - z^k) \right). \end{array} \right. \quad (1.6c)$$

Step 2: If a termination criterion is not met, set $k := k + 1$ and go to **Step 1**.

In this paper, we only focus on the linear convergence rate of the sPGADMM because it plays a pivotal role to measure an algorithm's behavior. We note that there is a smaller number of papers being devoted to the linear convergence rate of the GADMM (1.4) and its related variants.

For example, Fang et al. [14] proved the iteration complexity in both the ergodic and a nonergodic senses for a special case of sPGADMM, and then, the ergodic iteration complexity result was further improved by Adona et al. [1]. Corman & Yuan [7] and Tao & Yuan [38] analyzed the linear convergence rate of GADMM (1.4) from the aspect of generalized PPA, which is only a particular instance of sPGADMM. Peng & Zhang [29] established the linear convergence rate of GADMM (1.4) for the sequence $\{(z^k, x^k)\}$ instead of the whole sequence $\{(y^k, z^k, x^k)\}$ itself. In this shell, Peng et al. [28] proved the linear convergence rate of sPGADMM under certain settings, however, both \mathcal{S} and \mathcal{T} must be positive definite operators. Our focus of this paper is on the sPGADMM scheme (1.6a)-(1.6c) with \mathcal{S} and \mathcal{T} being semi-positive definite, which covers almost all the works aforementioned. We attempt to establish the linear rate of sPGADMM under a calmness condition, which holds automatically for convex composite piecewise linear-quadratic programming. Here, it is worth emphasizing that our convergence analysis is inspired by Han et al. [20] who provided a general linear rate convergence analysis for the semi-proximal ADMM of Fazel et al. [15]. Nevertheless, we show that the proof process is not trivial, because the relaxation factor brings many technical difficulties.

The remaining parts of this paper are organized as follows. Section 2 is divided into two parts: Subsection 2.1 presents some preliminary results on the optimality conditions for problem (1.2), and Subsection 2.2 briefly reviews the concepts of locally upper Lipschitz continuity and the calmness for multi-valued mappings. In Section 3, we provide a particularly useful inequality for the iteration sequence generated by sPGADMM, which is the key to the global convergence of sPGADMM. Section 4 is devoted to building up a general local linear convergence rate under an error bound condition, and then obtaining a global linear convergence rate. Finally, we conclude our paper with some remarks in Section 5.

2. PRELIMINARIES

In this section, we introduce some notations in the context and summarize some useful preliminaries for later analysis.

2.1. Optimality conditions. For any two vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, we use (x, y) to denote their adjunction, i.e., $(x, y) = (x^\top, y^\top)^\top$. For $p \geq 1$, we use $\|x\|_p$ to denote an ℓ_p -norm of a vector x , and $\|x\|$ is short for ℓ_2 -norm. Let \mathcal{G} be a positive definite matrix, the \mathcal{G} -norm of x is denoted by $\|x\|_{\mathcal{G}} := \sqrt{x^\top \mathcal{G} x}$. Given a closed convex set C , the distance of x to C regarding \mathcal{G} -norm is denoted as $\text{dist}_{\mathcal{G}}(x, C) := \inf_{y \in C} \|x - y\|_{\mathcal{G}}$.

Recall that f and g in (1.2) are closed proper convex functions, we denote their subdifferential mappings by ∂f and ∂g , respectively. Moreover, the subdifferential mappings of the closed proper convex functions are maximal monotone [35, Theorem 12.17], i.e., there exist two self-adjoint and positive semidefinite operators Σ_f and Σ_g such that for all $y, y' \in \text{dom}(f)$, $\xi \in \partial f(y)$ and $\xi' \in \partial f(y')$,

$$\langle \xi' - \xi, y' - y \rangle \geq \|y' - y\|_{\Sigma_f}^2, \quad (2.1)$$

and for all $z, z' \in \text{dom}(g)$, $\zeta \in \partial g(z)$ and $\zeta' \in \partial g(z')$,

$$\langle \zeta' - \zeta, z' - z \rangle \geq \|z' - z\|_{\Sigma_g}^2. \quad (2.2)$$

It follows from [34, Corollaries 28.2.2 and 28.3.1] that, $(\bar{y}, \bar{z}) \in \text{ri}(\text{dom}(f) \times \text{dom}(g))$ is an optimal solution to problem (1.2) if and only if there exists a Lagrange multiplier $\bar{x} \in \mathcal{X}$ such

that $(\bar{y}, \bar{z}, \bar{x})$ satisfies the following Karush-Kuhn-Tucker (KKT) system:

$$\mathcal{A}\bar{x} \in \partial f(\bar{y}), \quad \mathcal{B}\bar{z} \in \partial g(\bar{z}) \quad \text{and} \quad \mathcal{A}^*\bar{y} + \mathcal{B}^*\bar{z} - c = 0. \quad (2.3)$$

The solution set to the KKT system (2.3) is denoted by $\bar{\Omega}$. The nonempty of $\bar{\Omega}$ can be guaranteed if a certain constraint qualification such as the Slater condition holds, that is, there exists $(y^0, z^0) \in \text{ri}(\text{dom}(f) \times \text{dom}(g))$ such that $\mathcal{A}^*y^0 + \mathcal{B}^*z^0 = c$, where $\text{ri}(\cdot)$ denotes the relative interior of a given convex set. In this paper, instead of using an explicit constraint qualification, we only make the following assumption on the existence of a KKT point.

Assumption 2.1. The KKT system (2.3) has a nonempty solution set, i.e., $\bar{\Omega} \neq \emptyset$.

For notational convenience, we let $u := (y, z, x) \in \mathcal{U}$ with $\mathcal{U} := \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$ such that $y \in \mathcal{Y}$, $z \in \mathcal{Z}$, and $x \in \mathcal{X}$. Define the KKT mapping $R: \mathcal{U} \rightarrow \mathcal{U}$ as

$$R(u) := \begin{pmatrix} y - \text{Prox}_f(y + \mathcal{A}x) \\ z - \text{Prox}_g(z + \mathcal{B}x) \\ \mathcal{A}^*y + \mathcal{B}^*z - c \end{pmatrix}, \quad \forall u \in \mathcal{U}, \quad (2.4)$$

where $\text{Prox}_f(\cdot)$ represents the proximal mapping of a closed proper convex function f . From optimization theory, it was known that the proximal mappings $\text{Prox}_f(\cdot)$ and $\text{Prox}_g(\cdot)$ are globally Lipschitz continuous, which means that the mapping $R(\cdot)$ is continuous on \mathcal{U} and $R(u) = 0$ if and only if $u \in \bar{\Omega}$. The KKT mapping of (2.4) originated from Han et al. [20], which plays a vital role in our subsequent analysis.

2.2. Calmness. Let $F: \mathcal{X} \rightrightarrows \mathcal{Y}$ be a set-valued mapping with its graph denoted by $\text{gph}F$, and let $\mathbf{B}_{\mathcal{Y}}$ be a unit ball in \mathcal{Y} . In the first place, we give the definition of locally upper Lipschitz continuity.

Definition 2.1 ([30]). The multi-valued mapping $F: \mathcal{X} \rightrightarrows \mathcal{Y}$ is said to be Locally upper Lipschitz continuity at $x^0 \in \mathcal{X}$ if there exist a constant $k_0 > 0$ along with a neighborhood V of x_0 such that

$$F(x) \subseteq F(x^0) + k_0 \|x - x^0\| \mathbf{B}_{\mathcal{Y}}, \quad \forall x \in V.$$

It was known from Robinson [31] that, if a multivalued mapping $F: \mathcal{X} \rightrightarrows \mathcal{Y}$ is piecewise polyhedral, then F is locally upper Lipschitz continuous at any $x^0 \in \mathcal{X}$ with modulus k_0 independent of the choice of x^0 . A closed proper convex function $f: \mathcal{X} \rightarrow (-\infty, +\infty]$ is said to be piecewise linear-quadratic if $\text{dom}(f)$ is an union of finitely many polyhedral sets and on each of these polyhedral sets, f is either an affine or a quadratic function. Sun [37] proved that, f is piecewise linear-quadratic if and only if the graph of its subdifferential mapping ∂f is piecewise polyhedral. For more details, one may refer to Rockafellar & Wets [35, propositions 12.30 and 11.14]

In the second place, we give the definition of calmness for $F: \mathcal{X} \rightrightarrows \mathcal{Y}$ at x^0 for y^0 with $(x^0, y^0) \in \text{gph}F$. For more details on calmness, one can see Dontchev & Rockafellar [8, Section 3.8 (3H)] and Rockafellar & Wets [35].

Definition 2.2 ([8, 35]). Let $(x^0, y^0) \in \text{gph}F$. The multi-valued mapping $F: \mathcal{X} \rightrightarrows \mathcal{Y}$ is said to be calm at x^0 for y^0 with modulus $k_0 \geq 0$ if there exist a neighborhood V of x^0 and a neighborhood W of y^0 such that

$$F(x) \cap W \subseteq F(x^0) + k_0 \|x - x^0\| \mathbf{B}_{\mathcal{Y}}, \quad \forall x \in V.$$

Therefore, we get that the subdifferential mapping $\partial f(= F)$ of a convex piecewise linear-quadratic function f is calm at x^0 for y^0 meeting $(x^0, y^0) \in \text{gph}F$ with modulus $k_0 \geq 0$ independent of the selection of (x^0, y^0) . Besides, it is also known from Dontchev & Rockafellar [8, Theorem 3H.3] that, for any $(x^0, y^0) \in \text{gph}F$, the mapping F is calm at x^0 for y^0 if and only if F^{-1} (the inverse mapping of F) is metrically subregular at y^0 for x^0 , i.e., there exist a constant $k'_0 \geq 0$, a neighborhood W of y^0 , and a neighborhood V of x^0 such that

$$\text{dist}(y, F(x^0)) \leq k'_0 \text{dist}(x^0, F^{-1}(y) \cap V), \quad \forall y \in W, \quad (2.5)$$

where $\text{dist}(\cdot, \cdot)$ represents a distance regarding an appropriate norm.

3. GLOBAL CONVERGENCE OF SPGADMM

In order to prove the linear convergence rate, we must establish the global convergence of sPGADMM because it is essential to our subsequent analysis. Certainly, the convergence can be followed directly from the framework of PPA, but there is no proof for optimization problems especially including a pair of semi-definite operators \mathcal{S} and \mathcal{T} in the current literature. In our analysis, the following elementary equality is used times and times again

$$2\langle v_1, \mathcal{G}v_2 \rangle = \|v_1\|_{\mathcal{G}}^2 + \|v_2\|_{\mathcal{G}}^2 - \|v_1 - v_2\|_{\mathcal{G}}^2 = \|v_1 + v_2\|_{\mathcal{G}}^2 - \|v_1\|_{\mathcal{G}}^2 - \|v_2\|_{\mathcal{G}}^2, \quad (3.1)$$

where v_1 and v_2 are vectors in the same finite dimensional real Euclidean space endowed with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$, and \mathcal{G} is an arbitrary self-adjoint positive semi-definite linear operator.

For notational convenience, we denote $y_e = y - \bar{y}$, $z_e = z - \bar{z}$, and $x_e = x - \bar{x}$. Now, we are ready to give some inequalities used in the later analysis.

Lemma 3.1. *Let $\{u^k\}$ be the sequence generated by the sPGADMM scheme (1.6a)-(1.6c). For any $k \geq 0$, the following results hold:*

$$\langle \mathcal{B}^*(z^{k+1} - z^k), x^{k+1} - x^k \rangle \geq \frac{1}{2} \left(\|z^{k+1} - z^k\|_{\mathcal{T}}^2 - \|z^k - z^{k-1}\|_{\mathcal{T}}^2 \right), \quad (3.2)$$

$$\begin{aligned} & \left\langle x_e^{k+1} + \sigma(1 - \rho)\mathcal{B}^*z^{k+1}, \mathcal{A}^*y_e^{k+1} + \mathcal{B}^*z_e^{k+1} \right\rangle + \frac{1}{2}\sigma\rho\|\mathcal{A}^*y_e^{k+1} + \mathcal{B}^*z_e^{k+1}\|^2 \\ & = (2\sigma\rho)^{-1} \left[\|x_e^k + \sigma(1 - \rho)\mathcal{B}^*z_e^k\|^2 - \|x_e^{k+1} + \sigma(1 - \rho)\mathcal{B}^*z_e^{k+1}\|^2 \right], \end{aligned} \quad (3.3)$$

and

$$\begin{aligned} & \langle \mathcal{B}^*(z_e^{k+1} - z_e^k), \mathcal{A}^*y_e^{k+1} \rangle + (2\rho)^{-1}(2 - \rho)\|\mathcal{B}^*(z_e^{k+1} - z_e^k)\|^2 \\ & \leq \frac{1}{2} \left(\|\mathcal{B}^*z_e^k\|^2 - \|\mathcal{B}^*z_e^{k+1}\|^2 \right) + \frac{1}{2}(\sigma\rho)^{-1} \left(\|z^k - z^{k-1}\|_{\mathcal{T}}^2 - \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \right). \end{aligned} \quad (3.4)$$

Proof. Firstly, invoking the first-order optimality condition for (1.6b), we obtain

$$\mathcal{B} \left[x^k - \sigma\rho \left(\mathcal{A}^*y^{k+1} + \mathcal{B}^*z^{k+1} - c \right) - \sigma(1 - \rho)\mathcal{B}^*(z^{k+1} - z^k) \right] - \mathcal{T} \left(z^{k+1} - z^k \right) \in \partial g(z^{k+1}),$$

which, together with (1.6c), implies

$$\mathcal{B}x^{k+1} - \mathcal{T}(z^{k+1} - z^k) \in \partial g(z^{k+1}), \quad (3.5)$$

and $\mathcal{B}x^k - \mathcal{T}(z^k - z^{k-1}) \in \partial g(z^k)$. From the maximal monotonicity of $\partial g(\cdot)$, we have

$$\left\langle \mathcal{B}^* \left(z^{k+1} - z^k \right), x^{k+1} - x^k \right\rangle - \left\langle \mathcal{T} \left(z^{k+1} - z^k \right) - \mathcal{T} \left(z^k - z^{k-1} \right), z^{k+1} - z^k \right\rangle \geq 0.$$

The above inequality yields

$$\begin{aligned}
 \left\langle \mathcal{B}^* \left(z^{k+1} - z^k \right), x^{k+1} - x^k \right\rangle &\geq \left\langle \mathcal{F} \left(z^{k+1} - z^k \right) - \mathcal{F} \left(z^k - z^{k-1} \right), z^{k+1} - z^k \right\rangle \\
 &\geq \left\| z^{k+1} - z^k \right\|_{\mathcal{F}}^2 - \frac{1}{2} \left(\left\| z^{k+1} - z^k \right\|_{\mathcal{F}}^2 + \left\| z^k - z^{k-1} \right\|_{\mathcal{F}}^2 \right) \\
 &= \frac{1}{2} \left(\left\| z^{k+1} - z^k \right\|_{\mathcal{F}}^2 - \left\| z^k - z^{k-1} \right\|_{\mathcal{F}}^2 \right),
 \end{aligned}$$

which indicates that (3.2) is true.

Secondly, it follows from (1.6c) that

$$x^k = x^{k+1} + \sigma \rho \left(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^{k+1} - c \right) + \sigma (1 - \rho) \mathcal{B}^* (z^{k+1} - z^k).$$

It is easy to see that

$$\left[x_e^k + \sigma (1 - \rho) \mathcal{B}^* z_e^k \right] - \left[x_e^{k+1} + \sigma (1 - \rho) \mathcal{B}^* z_e^{k+1} \right] = \sigma \rho \left(\mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1} \right). \quad (3.6)$$

Subsequently, using the relation (3.6) and the elementary equality (3.1), we know that the equality (3.3) holds.

Finally, according to (3.6), we can deduce that

$$\begin{aligned}
 \mathcal{A}^* y_e^{k+1} &= \left(\mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1} \right) - \mathcal{B}^* z_e^{k+1} \\
 &= (\sigma \rho)^{-1} \left[x_e^k - x_e^{k+1} + \sigma (1 - \rho) \left(\mathcal{B}^* z_e^k - \mathcal{B}^* z_e^{k+1} \right) \right] - \mathcal{B}^* z_e^{k+1}.
 \end{aligned}$$

Combining this equality with (3.1), we obtain that

$$\begin{aligned}
 &\left\langle \mathcal{B}^* \left(z_e^{k+1} - z_e^k \right), \mathcal{A}^* y_e^{k+1} \right\rangle \\
 &= (\sigma \rho)^{-1} \left\langle \mathcal{B}^* \left(z_e^{k+1} - z_e^k \right), x_e^k - x_e^{k+1} \right\rangle - \left\langle \mathcal{B}^* \left(z_e^{k+1} - z_e^k \right), \mathcal{B}^* z_e^{k+1} \right\rangle \\
 &\quad + \rho^{-1} (1 - \rho) \left\langle \mathcal{B}^* \left(z_e^{k+1} - z_e^k \right), \mathcal{B}^* \left(z_e^k - z_e^{k+1} \right) \right\rangle \\
 &= (\sigma \rho)^{-1} \left\langle \mathcal{B}^* \left(z_e^{k+1} - z_e^k \right), x_e^k - x_e^{k+1} \right\rangle \\
 &\quad + \frac{1}{2} \left(\left\| \mathcal{B}^* z_e^k \right\|^2 - \left\| \mathcal{B}^* z_e^{k+1} \right\|^2 \right) - (2\rho)^{-1} (2 - \rho) \left\| \mathcal{B}^* \left(z_e^{k+1} - z_e^k \right) \right\|^2,
 \end{aligned}$$

which, together with (3.2), implies the inequality (3.4). \square

Utilizing Lemma 3.1, we can establish the difference of the distance to $\bar{\Omega}$ for two consecutive points of the sequence $\{u^k\}$.

Lemma 3.2. *Let the sequence $\{u^k\}$ be generated by sPGADMM (1.6a)-(1.6c). For any $k \geq 0$, define*

$$\begin{cases} \phi_{k+1} := (\sigma \rho)^{-1} \left\| x_e^{k+1} + \sigma (1 - \rho) \mathcal{B}^* z_e^{k+1} \right\|^2 + \left\| y_e^{k+1} \right\|_{\mathcal{F}}^2 + \left\| z_e^{k+1} \right\|_{\mathcal{F}}^2 \\ \quad + \sigma (2 - \rho) \left\| \mathcal{B}^* z_e^{k+1} \right\|^2 + (2 - \rho) \rho^{-1} \left\| z_e^{k+1} - z_e^k \right\|_{\mathcal{F}}^2, \\ t_{k+1} := 2 \left\| y_e^{k+1} \right\|_{\Sigma_f}^2 + 2 \left\| z_e^{k+1} \right\|_{\Sigma_g}^2 + \left\| y^{k+1} - y^k \right\|_{\mathcal{F}}^2 + \left\| z^{k+1} - z^k \right\|_{\mathcal{F}}^2 \\ \quad + \sigma (2 - \rho)^2 \rho^{-1} \left\| \mathcal{B}^* \left(z^{k+1} - z^k \right) \right\|^2. \end{cases} \quad (3.7)$$

Then, it holds that

$$\phi_k - \phi_{k+1} \geq t_{k+1} + (2 - \rho)\sigma \left\| \mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1} \right\|^2. \quad (3.8)$$

Proof. Notice that (1.6a) can be rewritten as

$$\mathcal{A} \left[x^k - \sigma \left(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^{k+1} - c \right) + \sigma \mathcal{B}^* \left(z^{k+1} - z^k \right) \right] - \mathcal{J} \left(y^{k+1} - y^k \right) \in \partial f(y^{k+1}).$$

It follows from (1.6c) that

$$\begin{aligned} & \mathcal{A} \left[x^{k+1} - \sigma(1 - \rho) \left(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^{k+1} - c \right) + \sigma(2 - \rho) \mathcal{B}^* \left(z^{k+1} - z^k \right) \right] \\ & - \mathcal{J} \left(y^{k+1} - y^k \right) \in \partial f(y^{k+1}). \end{aligned} \quad (3.9)$$

Combining (2.1) and (2.2) with (3.5) and (3.9), we have

$$\begin{aligned} & \left\langle x_e^{k+1}, \mathcal{A}^* y_e^{k+1} \right\rangle - \sigma(1 - \rho) \left\langle \mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1}, \mathcal{A}^* y_e^{k+1} \right\rangle - \left\langle \mathcal{J} \left(y_e^{k+1} - y_e^k \right), y_e^{k+1} \right\rangle \\ & + \sigma(2 - \rho) \left\langle \mathcal{B}^* \left(z_e^{k+1} - z_e^k \right), \mathcal{A}^* y_e^{k+1} \right\rangle \geq \left\| y_e^{k+1} \right\|_{\Sigma_f}^2, \end{aligned}$$

and

$$\left\langle x_e^{k+1}, \mathcal{B}^* z_e^{k+1} \right\rangle - \left\langle \mathcal{J} \left(z_e^{k+1} - z_e^k \right), z_e^{k+1} \right\rangle \geq \left\| z_e^{k+1} \right\|_{\Sigma_g}^2.$$

Adding both sides of the inequalities, we have

$$\begin{aligned} & \left\langle x_e^{k+1}, \mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1} \right\rangle - \sigma(1 - \rho) \left\langle \mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1}, \mathcal{A}^* y_e^{k+1} \right\rangle \\ & + \sigma(2 - \rho) \left\langle \mathcal{B}^* \left(z_e^{k+1} - z_e^k \right), \mathcal{A}^* y_e^{k+1} \right\rangle - \left\langle \mathcal{J} \left(y_e^{k+1} - y_e^k \right), y_e^{k+1} \right\rangle \\ & - \left\langle \mathcal{J} \left(z_e^{k+1} - z_e^k \right), z_e^{k+1} \right\rangle \\ & \geq \left\| y_e^{k+1} \right\|_{\Sigma_f}^2 + \left\| z_e^{k+1} \right\|_{\Sigma_g}^2, \end{aligned} \quad (3.10)$$

which can be reformulated as

$$\begin{aligned} & \left\langle x_e^{k+1} + \sigma(1 - \rho) \mathcal{B}^* z_e^{k+1}, \mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1} \right\rangle - \sigma(1 - \rho) \left\| \mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1} \right\|^2 \\ & + \sigma(2 - \rho) \left\langle \mathcal{B}^* \left(z_e^{k+1} - z_e^k \right), \mathcal{A}^* y_e^{k+1} \right\rangle - \left\langle \mathcal{J} \left(y_e^{k+1} - y_e^k \right), y_e^{k+1} \right\rangle \\ & - \left\langle \mathcal{J} \left(z_e^{k+1} - z_e^k \right), z_e^{k+1} \right\rangle \\ & \geq \left\| y_e^{k+1} \right\|_{\Sigma_f}^2 + \left\| z_e^{k+1} \right\|_{\Sigma_g}^2. \end{aligned} \quad (3.11)$$

Next, we analyze the left-hand-side of (3.11). Using the elementary equality (3.1), it is easy to see that

$$\left\langle \mathcal{J} \left(y_e^{k+1} - y_e^k \right), y_e^{k+1} \right\rangle = \frac{1}{2} \left(\left\| y_e^{k+1} - y_e^k \right\|_{\mathcal{J}}^2 + \left\| y_e^{k+1} \right\|_{\mathcal{J}}^2 - \left\| y_e^k \right\|_{\mathcal{J}}^2 \right), \quad (3.12)$$

and

$$\left\langle \mathcal{T} \left(z_e^{k+1} - z_e^k \right), z_e^{k+1} \right\rangle = \frac{1}{2} \left(\left\| z_e^{k+1} - z_e^k \right\|_{\mathcal{T}}^2 + \left\| z_e^{k+1} \right\|_{\mathcal{T}}^2 - \left\| z_e^k \right\|_{\mathcal{T}}^2 \right). \quad (3.13)$$

Substituting (3.3), (3.4), (3.12), and (3.13) into (3.11), we have

$$\begin{aligned} & (\sigma\rho)^{-1} \left\| x_e^k + \sigma(1-\rho) \mathcal{B}^* z_e^k \right\|^2 + \left\| y_e^k \right\|_{\mathcal{S}}^2 + \left\| z_e^k \right\|_{\mathcal{T}}^2 + \sigma(2-\rho) \left\| \mathcal{B}^* z_e^k \right\|^2 \\ & + \rho^{-1}(2-\rho) \left\| z_e^k - z_e^{k-1} \right\|_{\mathcal{T}}^2 \\ & \geq (\sigma\rho)^{-1} \left\| x_e^{k+1} + \sigma(1-\rho) \mathcal{B}^* z_e^{k+1} \right\|^2 + \left\| y_e^{k+1} \right\|_{\mathcal{S}}^2 + \left\| z_e^{k+1} \right\|_{\mathcal{T}}^2 + \sigma(2-\rho) \left\| \mathcal{B}^* z_e^{k+1} \right\|^2 \\ & + \rho^{-1}(2-\rho) \left\| z_e^{k+1} - z_e^k \right\|_{\mathcal{T}}^2 + 2 \left\| y_e^{k+1} \right\|_{\Sigma_f}^2 + 2 \left\| z_e^{k+1} \right\|_{\Sigma_g}^2 \\ & + (2-\rho)\sigma \left\| \mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1} \right\|^2 + \left\| y_e^{k+1} - y_e^k \right\|_{\mathcal{S}}^2 + \left\| z_e^{k+1} - z_e^k \right\|_{\mathcal{T}}^2 \\ & + \sigma(2-\rho)^2 \rho^{-1} \left\| \mathcal{B}^* \left(z_e^{k+1} - z_e^k \right) \right\|^2. \end{aligned}$$

Therefore, (3.8) holds by recalling the definitions in (3.7). \square

Inequality (3.8) is essential to establish the global convergence of sPGADMM. Although the global convergence of sPGADMM has been well studied in the literature especially from the respect of generalized PPA framework, we present a detailed proof for the completeness of global convergence.

Theorem 3.1. *Suppose that Assumption 2.1 holds. Assume that $\Sigma_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^*$ and $\Sigma_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^*$ are positive definite. Let the sequence $\{u^k\}$ be generated by sPGADMM scheme (1.6a)-(1.6c). Then $\{y^k, z^k\}$ converges to an optimal solution of (1.2) and $\{x^k\}$ converges to an optimal solution of the corresponding dual problem.*

Proof. In view of $\rho \in (0, 2)$, we obtain from (3.8) that $\{\phi_k\}$ is a nonnegative and monotonically nonincreasing sequence. Consequently, ϕ_k is bounded, so do the following sequences:

$$\left\{ \left\| x_e^k + \sigma(1-\rho) \mathcal{B}^* z_e^k \right\| \right\}, \left\{ \left\| z_e^k \right\|_{\sigma\mathcal{B}\mathcal{B}^*} \right\}, \left\{ \left\| z_e^{k+1} - z_e^k \right\|_{\mathcal{T}} \right\}, \left\{ \left\| z_e^k \right\|_{\mathcal{T}} \right\} \text{ and } \left\{ \left\| y_e^k \right\|_{\mathcal{S}} \right\}. \quad (3.14)$$

Besides, from inequality (3.8), it is clear that

$$\begin{aligned} \lim_{k \rightarrow \infty} \left\| \mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1} \right\| &= 0, \quad \lim_{k \rightarrow \infty} \left\| z_e^{k+1} - z_e^k \right\|_{\sigma\mathcal{B}\mathcal{B}^*} = 0, \quad \lim_{k \rightarrow \infty} \left\| z_e^{k+1} - z_e^k \right\|_{\mathcal{T}} = 0, \\ \lim_{k \rightarrow 0} \left\| z_e^{k+1} \right\|_{\Sigma_g} &= 0, \quad \lim_{k \rightarrow 0} \left\| y_e^{k+1} - y_e^k \right\|_{\mathcal{S}} = 0, \quad \text{and} \quad \lim_{k \rightarrow 0} \left\| y_e^{k+1} \right\|_{\Sigma_f} = 0. \end{aligned} \quad (3.15)$$

Furthermore, by the use of

$$\left\| \mathcal{A}^* y_e^{k+1} \right\| \leq \left\| \mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1} \right\| + \left\| \mathcal{B}^* z_e^{k+1} \right\|, \quad (3.16)$$

we know that $\left\{ \left\| y_e^{k+1} \right\|_{\sigma\mathcal{A}\mathcal{A}^*} \right\}$ is bounded, and so do the sequence $\left\{ \left\| y_e^{k+1} \right\|_{\Sigma_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^*} \right\}$. In light of the positive definiteness of $\Sigma_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^*$, sequence $\left\{ \left\| y_e^k \right\| \right\}$ is also bounded. Similarly, sequences $\left\{ \left\| z_e^k \right\|_{\sigma\mathcal{B}\mathcal{B}^*} \right\}$, $\left\{ \left\| z_e^k \right\|_{\mathcal{T}} \right\}$, and $\left\{ \left\| z_e^k \right\|_{\Sigma_g} \right\}$ are all bounded. Then, $\left\{ \left\| z_e^k \right\| \right\}$ is also bounded because $\Sigma_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^*$ is assumed to be positive definite. The boundedness of

$\{\|x_e^k + \sigma(1-\rho)\mathcal{B}^*z_e^k\|\}$ and $\{\|z_e^k\|\}$ further indicates that $\{\|x_e^k\|\}$ is bounded. According to the above arguments, it yields that $\{(y^k, z^k, x^k)\}$ is bounded.

From the boundedness of $\{(y^k, z^k, x^k)\}$, we know that there exists at least one subsequence which converges to a cluster point. Suppose that $\{(y^{k_i}, z^{k_i}, x^{k_i})\}$ is a subsequence of (y^k, z^k, x^k) converging to $(y^\infty, z^\infty, x^\infty) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$. Taking limits on both sides of (3.5) and (3.9) along the subsequence $\{(y^{k_i}, z^{k_i}, x^{k_i})\}$, using (3.14) and (3.15), and noticing the closedness of the graphs of ∂f and ∂g [3, p.80], we can obtain that

$$\mathcal{A}x^\infty \in \partial f(y^\infty), \quad \mathcal{B}x^\infty \in \partial g(z^\infty), \quad \mathcal{A}^*y^\infty + \mathcal{B}^*z^\infty - c = 0,$$

which implies that $(y^\infty, z^\infty, x^\infty)$ is a solution to the KKT system (2.3). Hence, (y^∞, z^∞) is an optimal solution to problem (1.2) and that x^∞ is an optimal solution to the corresponding dual problem.

We show that $(y^\infty, z^\infty, x^\infty)$ is actually the unique limit of $\{(y^k, z^k, x^k)\}$. One can replace $(\bar{y}, \bar{z}, \bar{x})$ with $(y^\infty, z^\infty, x^\infty)$ in (2.3) because $(y^\infty, z^\infty, x^\infty)$ is also satisfied from the above arguments. It is obvious that for $\rho \in (0, 2)$ the subsequence $\{\phi_{k_i}\}$ converges to 0 as $k_i \rightarrow \infty$. Because the subsequence comes from nonincreasing sequences, we find $\lim_{k \rightarrow \infty} \phi_k = 0$. Consequently, it holds that

$$\lim_{k \rightarrow \infty} \left(\|y_e^k\|_{\mathcal{S}} + \|y_e^{k+1}\|_{\Sigma_f} \right) + \left(\|z_e^k\|_{\sigma\mathcal{B}\mathcal{B}^*} + \|z_e^k\|_{\mathcal{T}} + \|z_e^k\|_{\Sigma_g} \right) = 0.$$

Thus, one can get $\lim_{k \rightarrow \infty} z^k = z^\infty$ as the operator $\Sigma_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^*$ is positive definite. This, combines with (3.16) implies that

$$\lim_{k \rightarrow \infty} \left(\|y_e^k\|_{\sigma\mathcal{A}\mathcal{A}^*} + \|y_e^k\|_{\mathcal{S}} + \|y_e^{k+1}\|_{\Sigma_f} \right) = 0,$$

which, from the fact that $\Sigma_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^*$ being positive definite, shows that $\lim_{k \rightarrow \infty} y^k = y^\infty$. Then, employing (3.14), we acquire that $\lim_{k \rightarrow \infty} x^k = x^\infty$. Therefore, it gets that the whole sequence $\{(y^k, z^k, x^k)\}$ converges to $(y^\infty, z^\infty, x^\infty)$ in the case of $\rho \in (0, 2)$. \square

4. LINEAR CONVERGENCE OF SPGADMM

This section is devoted to proving the local linear convergence rate of sPGADMM for convex composite optimization problem (1.2). For this purpose, we list some notations which will be used in the following analysis. First, for any $\rho \in (0, 2)$, we define

$$l_\rho := \begin{cases} \frac{1}{\rho}, & \rho \in \left(0, \frac{1+\sqrt{5}}{2}\right), \\ \frac{2-\rho}{\rho-1}, & \rho \in \left(\frac{1+\sqrt{5}}{2}, 2\right), \end{cases} \quad h_\rho := \begin{cases} 1 - \min\{\rho, \rho^{-1}\}, & \rho \in \left(0, \frac{1+\sqrt{5}}{2}\right), \\ 2-\rho, & \rho \in \left(\frac{1+\sqrt{5}}{2}, 2\right), \end{cases}$$

which implies that

$$l_\rho > 0, \quad 0 < \rho l_\rho < 2, \quad \text{and} \quad 0 < h_\rho < 2(2-\rho) \quad \forall \rho \in (0, 2).$$

Secondly, we define

$$m_\rho := \rho^{-1} \left(2 \min\{\rho, \rho^{-1}\} - \min\{1, \rho^2\} \right) l_\rho,$$

$$n_\rho := \rho^{-1}(2-\rho) \left[2(2-\rho) - h_\rho \right],$$

and

$$o_\rho := (2-\rho)(2-\rho l_\rho).$$

Thirdly, we define

$$\mathcal{M}_\rho := \begin{pmatrix} \mathcal{T} + \Sigma_g + \rho^{-1} \sigma \mathcal{B} \mathcal{B}^* & \rho^{-1}(1-\rho) \mathcal{B} \\ \rho^{-1}(1-\rho) \mathcal{B}^* & (\sigma \rho)^{-1} \mathcal{I} \end{pmatrix},$$

and

$$\mathcal{M} := \begin{pmatrix} \mathcal{S} + \Sigma_f & 0 \\ 0 & \mathcal{M}_\rho \end{pmatrix} + \frac{1}{4} o_\rho \sigma \varepsilon \varepsilon^*, \quad (4.1)$$

and

$$\mathcal{H} := \begin{pmatrix} \mathcal{S} + \Sigma_f & 0 & 0 \\ 0 & \mathcal{T} + \Sigma_g + \frac{1}{2} n_\rho \sigma \mathcal{B} \mathcal{B}^* & 0 \\ 0 & 0 & (2\sigma)^{-1} m_\rho \mathcal{I} \end{pmatrix} + \frac{1}{8} o_\rho \sigma \varepsilon \varepsilon^*, \quad (4.2)$$

where \mathcal{I} is an identity operator within an appropriate space, $\varepsilon : \mathcal{X} \rightarrow \mathcal{U} := \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$ is a linear operator such that its adjoint ε^* satisfies

$$\varepsilon^*(y, z, x) = \mathcal{A}^* y + \mathcal{B}^* z, \quad \forall (y, z, x) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}. \quad (4.3)$$

Finally, we define

$$k_1 := 3 \|\mathcal{S}\|, \quad k_2 := \max \left\{ \|\mathcal{T}\|, \rho^{-2} \left(3 \lambda_{\max}(\mathcal{A}^* \mathcal{A}) \sigma + 2(1-\rho)^2 \sigma^{-1} \right) \right\},$$

and

$$k_3 := 3(1-\rho)^2 \rho^{-2} \sigma \lambda_{\max}(\mathcal{A}^* \mathcal{A}) + 2\sigma^{-2} \rho^{-2}.$$

Then, we let $k_4 := \max \{k_1, k_2, k_3\}$. Let \mathcal{H}_0 be a block-diagonal linear operator defined by

$$\mathcal{H}_0 := k_4 \begin{pmatrix} \mathcal{S} & 0 & 0 \\ 0 & \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^* & 0 \\ 0 & 0 & (2\sigma)^{-1} \mathcal{I} \end{pmatrix}.$$

Because $\rho \in (0, 2)$, and $\Sigma_f + \mathcal{S} + \sigma \mathcal{A} \mathcal{A}^*$ and $\Sigma_g + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^*$ are assumed to be positive definite, it is easy to see that \mathcal{M}_ρ is positive definite, \mathcal{M} , \mathcal{H}_0 and \mathcal{H} are positive semidefinite.

We now give a lemma to estimate the upper bound of the sequence $\{R(u^k)\}$.

Lemma 4.1. *Let $\{u^k\}$ be the sequence generated by sPGADMM scheme ((1.6a)-(1.6c)). Then, for any $k \geq 0$, we have*

$$\|u^{k+1} - u^k\|_{\mathcal{H}_0}^2 \geq \|R(u^{k+1})\|^2. \quad (4.4)$$

Proof. It follows from (3.5) and (3.9) that

$$\begin{cases} y^{k+1} = \text{Prox}_f \left(y^{k+1} + \mathcal{A} [x^{k+1} - (1-\rho) \sigma (\mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1}) \right. \\ \quad \left. + \sigma (2-\rho) \mathcal{B}^* (z^{k+1} - z^k)] - \mathcal{S} (y^{k+1} - y^k) \right), \\ z^{k+1} = \text{Prox}_g \left(z^{k+1} + \mathcal{B} x^{k+1} - \mathcal{T} (z^{k+1} - z^k) \right). \end{cases} \quad (4.5)$$

Then, combining (4.5) and (1.6c), and noticing the Lipschitz continuity property of Moreau-Yosida proximal mapping, we can get from the definition of $R(\cdot)$ (2.4) that

$$\begin{aligned}
\|R(u^{k+1})\|^2 &\leq \left\| \mathcal{S}(y^{k+1} - y^k) + (1 - \rho)\rho^{-1}\mathcal{A}(x^k - x^{k+1}) - \rho^{-1}\sigma\mathcal{A}\mathcal{B}^*(z^k - z^{k+1}) \right\|^2 \\
&\quad + \left\| \mathcal{T}(z^{k+1} - z^k) \right\|^2 + \left\| (\sigma\rho)^{-1}(x^k - x^{k+1}) + \rho^{-1}(1 - \rho)\mathcal{B}^*(z^k - z^{k+1}) \right\|^2 \\
&\leq 3\|\mathcal{S}\| \left\| y^{k+1} - y^k \right\|_{\mathcal{S}}^2 + 3(1 - \rho)^2\rho^{-2}\lambda_{\max}(\mathcal{A}^*\mathcal{A}) \left\| x^k - x^{k+1} \right\|^2 \\
&\quad + 3\rho^{-2}\sigma^2\lambda_{\max}(\mathcal{A}^*\mathcal{A}) \left\| \mathcal{B}^*(z^{k+1} - z^k) \right\|^2 + \|\mathcal{T}\| \left\| z^{k+1} - z^k \right\|_{\mathcal{T}}^2 \\
&\quad + 2(\sigma\rho)^{-2} \left\| x^k - x^{k+1} \right\|^2 + 2\rho^{-2}(1 - \rho)^2 \left\| \mathcal{B}^*(z^k - z^{k+1}) \right\|^2 \\
&\leq k_1 \left\| y^{k+1} - y^k \right\|_{\mathcal{S}}^2 + k_2 \left\| z^{k+1} - z^k \right\|_{\sigma\mathcal{B}\mathcal{B}^* + \mathcal{T}}^2 + k_3 \left\| x^k - x^{k+1} \right\|^2.
\end{aligned}$$

Recalling the definition of \mathcal{H}_0 , one can readily see that inequality (4.4) holds. This completes the proof. \square

In the following analysis, the \mathcal{M} defined in (4.1) is used to measure the distance from an iterate to the KKT point, while the \mathcal{H} given in (4.2) is used to serve the distance between two consecutive iterates. The positive definiteness of the linear operators \mathcal{M} and \mathcal{H} are proved at the following lemma.

Lemma 4.2. *Let $\rho \in (0, 2)$. Then, it holds that*

$$\left(\Sigma_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \succ 0 \text{ and } \Sigma_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* \succ 0 \right) \iff \mathcal{M} \succ 0 \iff \mathcal{H} \succ 0,$$

where the symbol ' $\succ 0$ ' means that the associated operator is positive definite.

Proof. We only prove the first equivalence because the second one can be obtained in a similar way. Suppose on the contrary that there exists a non-zero vector $d := (d_y, d_z, d_x) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$ such that $\langle d, \mathcal{M}d \rangle = 0$ in the cases of $\Sigma_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \succ 0$ and $\Sigma_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* \succ 0$. Using the definition of \mathcal{M} in (4.1), it means that

$$\langle d_y, (S + \Sigma_f)d_y \rangle = 0, \quad \langle (d_z, d_x), \mathcal{M}_\rho(d_z, d_x) \rangle = 0, \quad \text{and} \quad \mathcal{A}^*d_y + \mathcal{B}^*d_z = 0,$$

which, combines with the fact that $\mathcal{M}_\rho \succ 0$ and that $\Sigma_f + S + \sigma\mathcal{A}\mathcal{A}^* \succ 0$, it yields that $d = 0$. This result contradicts to the assumption $d \neq 0$, and hence, $\mathcal{M} \succ 0$. Conversely, suppose that $\mathcal{M} \succ 0$. Because $o_\rho > 0$ is a positive scalar, for any $d := (d_y, 0, 0) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$, it holds that $\langle d_y, (\mathcal{S} + \Sigma_f + \frac{1}{4}o_\rho\sigma\mathcal{A}\mathcal{A}^*)d_y \rangle > 0$, which means $\Sigma_f + S + \sigma\mathcal{A}\mathcal{A}^* \succ 0$. In a similar way, for any $d := (0, d_z, 0) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$, we get $\langle d_z, [\mathcal{T} + \Sigma_g + (\rho^{-1} + \frac{1}{4}o_\rho)\sigma\mathcal{B}\mathcal{B}^*]d_z \rangle > 0$, which means that $\Sigma_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* \succ 0$. In a summary, the first equivalence is proved. \square

Based on Lemmas 4.1 and 4.2, we are ready to present a key inequality required for the coming linear convergence rate analysis.

Lemma 4.3. *Let $\rho \in (0, 2)$ and $\{u^k\}$ be a sequence generated by sPGADMM scheme ((1.6a)-(1.6c)). Then, for any $\bar{u} = (\bar{y}, \bar{z}, \bar{x}) \in \bar{\Omega}$ and any $k \geq 0$, we have*

$$\begin{aligned} & \left\| u^k - \bar{u} \right\|_{\mathcal{M}}^2 + (2 - \rho) \rho^{-1} \left\| z^k - z^{k-1} \right\|_{\mathcal{F}}^2 \\ & \geq \left(\left\| u^{k+1} - \bar{u} \right\|_{\mathcal{M}}^2 + (2 - \rho) \rho^{-1} \left\| z^{k+1} - z^k \right\|_{\mathcal{F}}^2 \right) + \left\| u^{k+1} - u^k \right\|_{\mathcal{H}}^2. \end{aligned} \quad (4.6)$$

Consequently, it also holds for $k \geq 0$ that

$$\begin{aligned} & \text{dist}_{\mathcal{M}}^2 \left(u^k, \bar{\Omega} \right) + (2 - \rho) \rho^{-1} \left\| z^k - z^{k-1} \right\|_{\mathcal{F}}^2 \\ & \geq \left(\text{dist}_{\mathcal{M}}^2 \left(u^{k+1}, \bar{\Omega} \right) + (2 - \rho) \rho^{-1} \left\| z^{k+1} - z^k \right\|_{\mathcal{F}}^2 \right) + \left\| u^{k+1} - u^k \right\|_{\mathcal{H}}^2. \end{aligned} \quad (4.7)$$

Proof. From (1.6c), it holds that

$$\begin{aligned} & \left\| x^{k+1} - x^k \right\|^2 = (\sigma \rho)^2 \left\| \mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1} \right\|^2 - \sigma^2 (1 - \rho)^2 \left\| \mathcal{B}^* (z^{k+1} - z^k) \right\|^2 \\ & + 2(1 - \rho) \left\langle x^k - x^{k+1}, \sigma \mathcal{B}^* (z^{k+1} - z^k) \right\rangle \\ & \leq (\sigma \rho)^2 \left\| \mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1} \right\|^2 - \sigma^2 (1 - \rho)^2 \left\| \mathcal{B}^* (z^{k+1} - z^k) \right\|^2 \\ & + (1 - \min \{ \rho, \rho^{-1} \}) \left\| x^{k+1} - x^k \right\|^2 + \rho (\max \{ \rho, \rho^{-1} \} - 1) \sigma^2 \left\| \mathcal{B}^* (z^{k+1} - z^k) \right\|^2. \end{aligned}$$

Let $\bar{u} = (\bar{y}, \bar{z}, \bar{x}) \in \bar{\Omega}$. Using Lemma 3.2 and above inequality, it yields that

$$\begin{aligned} & (\sigma \rho)^{-1} \left\| x_e^k + \sigma (1 - \rho) \mathcal{B}^* z_e^k \right\|^2 + \left\| y_e^k \right\|_{\mathcal{S}}^2 + \left\| z_e^k \right\|_{\mathcal{F}}^2 + \sigma (2 - \rho) \left\| \mathcal{B}^* z_e^k \right\|^2 \\ & + \rho^{-1} (2 - \rho) \left\| z^k - z^{k-1} \right\|_{\mathcal{F}}^2 \\ & \geq (\sigma \rho)^{-1} \left\| x_e^{k+1} + \sigma (1 - \rho) \mathcal{B}^* z_e^{k+1} \right\|^2 + \left\| y_e^{k+1} \right\|_{\mathcal{S}}^2 + \left\| z_e^{k+1} \right\|_{\mathcal{F}}^2 \\ & + \sigma (2 - \rho) \left\| \mathcal{B}^* z_e^{k+1} \right\|^2 + \rho^{-1} (2 - \rho) \left\| z^{k+1} - z^k \right\|_{\mathcal{F}}^2 \\ & + (2\sigma \rho)^{-1} (2 \min \{ \rho, \rho^{-1} \} - \min \{ 1, \rho^2 \}) l_\rho \left\| x^{k+1} - x^k \right\|^2 \\ & + \left\| y^{k+1} - y^k \right\|_{\mathcal{S}}^2 + \left\| z^{k+1} - z^k \right\|_{\mathcal{F}}^2 + (2 - \rho)^2 \rho^{-1} \sigma \left\| \mathcal{B}^* (z^{k+1} - z^k) \right\|^2 \end{aligned}$$

$$\begin{aligned}
& + 2 \left\| y_e^{k+1} \right\|_{\Sigma_f}^2 + 2 \left\| z_e^{k+1} \right\|_{\Sigma_g}^2 + (2 - \rho) \sigma \left\| \varepsilon^* \left(y^{k+1}, z^{k+1}, 0 \right) - c \right\|^2 \\
& - \frac{1}{2} (2 - \rho) \rho \sigma l_\rho \left\| \varepsilon^* \left(y^{k+1}, z^{k+1}, x^{k+1} \right) - c \right\|^2 \\
& - \frac{1}{2} (2 - \rho) (1 - \min \{ \rho, \rho^{-1} \}) \sigma l_\rho \left\| \mathcal{B}^* \left(z^{k+1} - z^k \right) \right\|^2 \\
& = (\sigma \rho)^{-1} \left\| x_e^{k+1} + \sigma (1 - \rho) \mathcal{B}^* z_e^{k+1} \right\|^2 + \left\| y_e^{k+1} \right\|_{\mathcal{J}}^2 + \left\| z_e^{k+1} \right\|_{\mathcal{J}}^2 + \sigma (2 - \rho) \left\| \mathcal{B}^* z_e^{k+1} \right\|^2 \\
& + \rho^{-1} (2 - \rho) \left\| z^{k+1} - z^k \right\|_{\mathcal{J}}^2 + (2\sigma)^{-1} m_\rho \left\| x^{k+1} - x^k \right\|^2 \\
& + \left\| y^{k+1} - y^k \right\|_{\mathcal{J}}^2 + \left\| z^{k+1} - z^k \right\|_{\mathcal{J}}^2 + \frac{1}{2} n_\rho \sigma \left\| \mathcal{B}^* \left(z^{k+1} - z^k \right) \right\|^2 \\
& + \frac{1}{2} o_\rho \sigma \left\| \varepsilon^* \left(y^{k+1}, z^{k+1}, x^{k+1} \right) - c \right\|^2 + 2 \left\| y_e^{k+1} \right\|_{\Sigma_f}^2 + 2 \left\| z_e^{k+1} \right\|_{\Sigma_g}^2,
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
& (\sigma \rho)^{-1} \left\| x_e^k + \sigma (1 - \rho) \mathcal{B}^* z_e^k \right\|^2 + \left\| y_e^k \right\|_{\mathcal{J}}^2 + \left\| z_e^k \right\|_{\mathcal{J}}^2 + \sigma (2 - \rho) \left\| \mathcal{B}^* z_e^k \right\|^2 \\
& + \rho^{-1} (2 - \rho) \left\| z^k - z^{k-1} \right\|_{\mathcal{J}}^2 + \left\| y_e^k \right\|_{\Sigma_f}^2 + \left\| z_e^k \right\|_{\Sigma_g}^2 + \frac{1}{4} o_\rho \sigma \left\| \varepsilon^* \left(y^k, z^k, x^k \right) - c \right\|^2 \\
& \geq \left[(\sigma \rho)^{-1} \left\| x_e^{k+1} + \sigma (1 - \rho) \mathcal{B}^* z_e^{k+1} \right\|^2 + \left\| y_e^{k+1} \right\|_{\mathcal{J}}^2 + \left\| z_e^{k+1} \right\|_{\mathcal{J}}^2 + \sigma (2 - \rho) \left\| \mathcal{B}^* z_e^{k+1} \right\|^2 \right. \\
& + \rho^{-1} (2 - \rho) \left\| z^{k+1} - z^k \right\|_{\mathcal{J}}^2 + \left\| y_e^{k+1} \right\|_{\Sigma_f}^2 + \left\| z_e^{k+1} \right\|_{\Sigma_g}^2 \\
& \left. + \frac{1}{4} o_\rho \sigma \left\| \varepsilon^* \left(y^{k+1}, z^{k+1}, x^{k+1} \right) - c \right\|^2 \right] + (2\sigma)^{-1} m_\rho \left\| x^{k+1} - x^k \right\|^2 + \left\| y^{k+1} - y^k \right\|_{\mathcal{J}}^2 \\
& + \left\| z^{k+1} - z^k \right\|_{\mathcal{J}}^2 + \frac{1}{2} n_\rho \sigma \left\| \mathcal{B}^* \left(z^{k+1} - z^k \right) \right\|^2 + \left\| y_e^k \right\|_{\Sigma_f}^2 + \left\| y_e^{k+1} \right\|_{\Sigma_f}^2 + \left\| z_e^k \right\|_{\Sigma_g}^2 \\
& + \left\| z_e^{k+1} \right\|_{\Sigma_g}^2 + \frac{1}{4} o_\rho \sigma \left\| \varepsilon^* \left(y^{k+1}, z^{k+1}, x^{k+1} \right) - c \right\|^2 + \frac{1}{4} o_\rho \sigma \left\| \varepsilon^* \left(y^k, z^k, x^k \right) - c \right\|^2.
\end{aligned} \tag{4.8}$$

Moreover, from (4.3) we have

$$\begin{aligned}
\varepsilon^* \left(y^{k+1}, z^{k+1}, x^{k+1} \right) - c & = \mathcal{A}^* \left(y^{k+1} - \bar{y} \right) + \mathcal{B}^* \left(z^{k+1} - \bar{z} \right), \\
\varepsilon^* \left(y^k, z^k, x^k \right) - c & = \mathcal{A}^* \left(y^k - \bar{y} \right) + \mathcal{B}^* \left(z^k - \bar{z} \right),
\end{aligned}$$

and we also have

$$\begin{aligned}
 & \left\| y^{k+1} - \bar{y} \right\|_{\Sigma_f}^2 + \left\| y^k - \bar{y} \right\|_{\Sigma_f}^2 \geq \frac{1}{2} \left\| y^{k+1} - y^k \right\|_{\Sigma_f}^2, \\
 & \left\| z^{k+1} - \bar{z} \right\|_{\Sigma_g}^2 + \left\| z^k - \bar{z} \right\|_{\Sigma_g}^2 \geq \frac{1}{2} \left\| z^{k+1} - z^k \right\|_{\Sigma_g}^2, \\
 & \left\| \mathcal{E}^* \left(y^{k+1}, z^{k+1}, x^{k+1} \right) - c \right\|^2 + \left\| \mathcal{E}^* \left(y^k, z^k, x^k \right) - c \right\|^2 \\
 & \geq \frac{1}{2} \left\| \mathcal{E}^* \left(y^{k+1} - y^k, z^{k+1} - z^k, x^{k+1} - x^k \right) \right\|^2.
 \end{aligned}$$

Based on the above relations, one can see from (4.8) for any $\rho \in (0, 2)$ that

$$\begin{aligned}
 & (\sigma\rho)^{-1} \left\| x_e^k + \sigma(1-\rho) \mathcal{B}^* z_e^k \right\|^2 + \left\| y_e^k \right\|_{\mathcal{F}}^2 + \left\| z_e^k \right\|_{\mathcal{F}}^2 + \sigma(2-\rho) \left\| \mathcal{B}^* z_e^k \right\|^2 \\
 & + \rho^{-1}(2-\rho) \left\| z^k - z^{k-1} \right\|_{\mathcal{F}}^2 + \left\| y_e^k \right\|_{\Sigma_f}^2 + \left\| z_e^k \right\|_{\Sigma_g}^2 + \frac{1}{4} o_\rho \sigma \left\| \mathcal{E}^* \left(y_e^k, z_e^k, x_e^k \right) \right\|^2 \\
 & \geq \left[(\sigma\rho)^{-1} \left\| x_e^{k+1} + \sigma(1-\rho) \mathcal{B}^* z_e^{k+1} \right\|^2 + \left\| y_e^{k+1} \right\|_{\mathcal{F}}^2 + \left\| z_e^{k+1} \right\|_{\mathcal{F}}^2 + \sigma(2-\rho) \left\| \mathcal{B}^* z_e^{k+1} \right\|^2 \right. \\
 & \left. + \rho^{-1}(2-\rho) \left\| z^{k+1} - z^k \right\|_{\mathcal{F}}^2 + \left\| y_e^{k+1} \right\|_{\Sigma_f}^2 + \left\| z_e^{k+1} \right\|_{\Sigma_g}^2 \right. \\
 & \left. + \frac{1}{4} o_\rho \sigma \left\| \mathcal{E}^* \left(y_e^{k+1}, z_e^{k+1}, x_e^{k+1} \right) \right\|^2 \right] + (2\sigma)^{-1} m_\rho \left\| x^{k+1} - x^k \right\|^2 \\
 & + \left\| y^{k+1} - y^k \right\|_{\mathcal{F}}^2 + \left\| z^{k+1} - z^k \right\|_{\mathcal{F}}^2 + \frac{1}{2} n_\rho \sigma \left\| \mathcal{B}^* \left(z^{k+1} - z^k \right) \right\|^2 \\
 & + \frac{1}{2} \left\| y^{k+1} - y^k \right\|_{\Sigma_f}^2 + \frac{1}{2} \left\| z^{k+1} - z^k \right\|_{\Sigma_g}^2 \\
 & + \frac{1}{8} o_\rho \sigma \left\| \mathcal{E}^* \left(y^{k+1} - y^k, z^{k+1} - z^k, x^{k+1} - x^k \right) \right\|^2,
 \end{aligned} \tag{4.9}$$

which implies inequality (4.6) holds. Because $\bar{\Omega}$ is assumed to be nonempty and (4.6) holds for any $\bar{u} \in \bar{\Omega}$, we obtain the desired result (4.7) immediately. \square

To prove the local linear convergence rate of sPGADMM, we need the following calmness assumption on R^{-1} at the original point for a KKT point.

Assumption 4.1. The inverse mapping of KKT mapping $R(\cdot)$ defined in (2.4) is calm at the origin for \bar{u} with modulus $\kappa > 0$ in the sense that there exists a constant $r > 0$ such that

$$\text{dist}(u, \bar{\Omega}) \leq \kappa \|R(u)\|, \quad \forall u \in \{u \in \mathcal{U} \mid \|u - \bar{u}\| \leq r\}.$$

Now, we are ready to present the local linear convergence rate of sPGADMM.

Theorem 4.1. *Suppose that Assumptions 2.1 and 4.1 hold. Assume that $\Sigma_f + \mathcal{S} + \sigma \mathcal{A} \mathcal{A}^* \succ 0$ and $\Sigma_g + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^* \succ 0$. Then the sequence $\{(y^k, z^k, x^k)\}$ generated by sPGADMM converges to a KKT point $\bar{u} = (\bar{y}, \bar{z}, \bar{x}) \in \bar{\Omega}$, and there exists a $\bar{k} \geq 0$ such that for all $k \geq \bar{k}$ it holds that*

$$\text{dist}_{\bar{\mathcal{M}}}^2(u^{k+1}, \bar{\Omega}) + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \leq \vartheta \left[\text{dist}_{\bar{\mathcal{M}}}^2(u^k, \bar{\Omega}) + \|z^k - z^{k-1}\|_{\mathcal{T}}^2 \right], \quad (4.10)$$

where $\bar{\mathcal{M}} := \rho(2-\rho)^{-1} \mathcal{M}$ and $\vartheta := \left[1 + k_5 \rho (\rho + (2-\rho)k_5)^{-1} \right]^{-1}$ with

$$k_5 := \min \left\{ 1, m_\rho, \frac{1}{2} n_\rho \right\} k_4^{-1} \rho (2-\rho)^{-1} \kappa^{-2} \lambda_{\max}(\bar{\mathcal{M}})^{-1},$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a given matrix. Moreover, there also exists a positive number $\eta \in [\vartheta, 1)$ such that for all $k \geq \bar{k}$,

$$\text{dist}_{\bar{\mathcal{M}}}^2(u^{k+1}, \bar{\Omega}) + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \leq \eta \left[\text{dist}_{\bar{\mathcal{M}}}^2(u^k, \bar{\Omega}) + \|z^k - z^{k-1}\|_{\mathcal{T}}^2 \right]. \quad (4.11)$$

Proof. From Theorem 3.1, we know that the sequence $\{(y^k, z^k, x^k)\}$ generated by sPGADMM converges to a KKT point $\bar{u} = (\bar{y}, \bar{z}, \bar{x}) \in \bar{\Omega}$, which shows that there exist $\bar{k} \geq 0$ and $r > 0$ such that, for any $k \geq \bar{k}$, $\|u^{k+1} - \bar{u}\| \leq r$. Subsequently, from Assumption 4.1 and Lemma 4.1, it obtains for all $k \geq \bar{k}$ that

$$\text{dist}^2(u^{k+1}, \bar{\Omega}) \leq \kappa^2 \|R(u^{k+1})\|^2 \leq \kappa^2 \|u^{k+1} - u^k\|_{\mathcal{H}_0}^2. \quad (4.12)$$

Denoting $\bar{\mathcal{H}} := \rho(2-\rho)^{-1} \mathcal{H}$ and noticing $\bar{\mathcal{M}} := \rho(2-\rho)^{-1} \mathcal{M}$, one sees that (4.7) can be simplified as

$$\text{dist}_{\bar{\mathcal{M}}}^2(u^k, \bar{\Omega}) + \|z^k - z^{k-1}\|_{\mathcal{T}}^2 \geq \left(\text{dist}_{\bar{\mathcal{M}}}^2(u^{k+1}, \bar{\Omega}) + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \right) + \|u^{k+1} - u^k\|_{\bar{\mathcal{H}}}^2. \quad (4.13)$$

Recalling the structure of \mathcal{H} , we know for all $k \geq 0$ that

$$\|u^{k+1} - u^k\|_{\bar{\mathcal{H}}}^2 \geq \rho(2-\rho)^{-1} \|z^{k+1} - z^k\|_{\mathcal{T}}^2. \quad (4.14)$$

Observing the structures of $\bar{\mathcal{H}}$ and \mathcal{H}_0 , it holds that

$$k_4 \bar{\mathcal{H}} \geq \min \left\{ 1, m_\rho, \frac{1}{2} n_\rho \right\} \rho(2-\rho)^{-1} \mathcal{H}_0 + \frac{1}{8} k_4 \rho(2-\rho)^{-1} o_\rho \sigma \varepsilon \varepsilon^*, \quad \forall \rho \in (0, 2). \quad (4.15)$$

Combining (4.12) and (4.15), we obtain for $k \geq \bar{k}$ that

$$\begin{aligned} \|u^{k+1} - u^k\|_{\bar{\mathcal{H}}}^2 &\geq \min \left\{ 1, m_\rho, \frac{1}{2} n_\rho \right\} k_4^{-1} \rho(2-\rho)^{-1} \|u^{k+1} - u^k\|_{\mathcal{H}_0}^2 \\ &\geq \min \left\{ 1, m_\rho, \frac{1}{2} n_\rho \right\} k_4^{-1} \rho(2-\rho)^{-1} \kappa^{-2} \text{dist}^2(u^{k+1}, \bar{\Omega}) \\ &\geq k_5 \text{dist}_{\bar{\mathcal{M}}}^2(u^{k+1}, \bar{\Omega}). \end{aligned} \quad (4.16)$$

Let $k_6 := \rho(\rho + (2 - \rho)k_5)^{-1}$. Accordingly, for all $k \geq \bar{k}$, it follows from (4.16), (4.14) and (4.13) that

$$\begin{aligned} & \text{dist}_{\bar{\mathcal{M}}}^2(u^k, \bar{\Omega}) + \|z^k - z^{k-1}\|_{\mathcal{Z}}^2 \\ & \geq \text{dist}_{\bar{\mathcal{M}}}^2(u^{k+1}, \bar{\Omega}) + \|z^{k+1} - z^k\|_{\mathcal{Z}}^2 + k_6 \|u^{k+1} - u^k\|_{\mathcal{H}}^2 + (1 - k_6) \|u^{k+1} - u^k\|_{\mathcal{H}}^2 \\ & \geq (1 + k_5 k_6) \text{dist}_{\bar{\mathcal{M}}}^2(u^{k+1}, \bar{\Omega}) + \left[1 + (1 - k_6)\rho(2 - \rho)^{-1}\right] \|z^{k+1} - z^k\|_{\mathcal{Z}}^2. \end{aligned}$$

From the fact that $1 + k_5 k_6 = 1 + (1 - k_6)\rho(2 - \rho)^{-1}$, we obtain the assertion (4.10) holds with $\vartheta = (1 + k_5 k_6)^{-1}$. Because $\vartheta < 1$, there must exist a positive number $\eta \in [\vartheta, 1)$ such that (4.11) holds. This completes the proof. \square

From Theorem 4.1, we know that the calmness assumption on R^{-1} is needed to ensure the local linear convergence rate of sPGADMM. From the preliminaries reviewed in Subsection 2.2 we know that, R^{-1} is calm if it is a piecewise polyhedral multivalued mapping, and particularly, it is the subdifferential mapping of a convex piecewise linear-quadratic function. Using the fact that R^{-1} is piecewise polyhedral if and only if R itself is piecewise polyhedral, we conclude that the calmness property holds automatically for convex composite optimization problem (1.2), as well as the following corollary.

Corollary 4.1. *Let $\rho \in (0, 2)$. Suppose that the set $\bar{\Omega}$ is nonempty and that both $\Sigma_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^*$ and $\Sigma_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^*$ are positive definite. Let $u^k = \{(y^k, z^k, x^k)\}$ be the sequence generated by Algorithm sPGADMM that converges to $\bar{u} \in \bar{\Omega}$. Assume that the mapping $R(\cdot)$ is piecewise polyhedral. Then, there exists a constant $\hat{\kappa} > 0$ such that for all $k \geq 0$,*

$$\text{dist}(u^k, \bar{\Omega}) \leq \hat{\kappa} \|R(u^k)\|, \quad (4.17)$$

and

$$\text{dist}_{\bar{\mathcal{M}}}^2(u^{k+1}, \bar{\Omega}) + \|z^{k+1} - z^k\|_{\mathcal{Z}}^2 \leq \hat{\vartheta} \left[\text{dist}_{\bar{\mathcal{M}}}^2(u^k, \bar{\Omega}) + \|z^k - z^{k-1}\|_{\mathcal{Z}}^2 \right], \quad (4.18)$$

where

$$\hat{\vartheta} = \left[1 + \hat{k}_5 \rho (\rho + (2 - \rho) \hat{k}_5)^{-1}\right]^{-1}, \text{ and } \hat{k}_5 = \min \left\{ 1, m_\rho, \frac{1}{2} n_\rho \right\} k_4^{-1} \rho (2 - \rho)^{-1} \hat{\kappa}^{-2} \lambda_{\max}(\bar{\mathcal{M}})^{-1}.$$

Proof. Because $\bar{\Omega}$ is nonempty and R^{-1} is piecewise polyhedral, we know that there exist constants $\kappa > 0$ and $\zeta > 0$ such that

$$\text{dist}(u, \bar{\Omega}) \leq \kappa \|R(u)\|, \quad \forall u \in \{u \in \mathcal{U} \mid \|R(u)\| \leq \zeta\}.$$

Besides, according to the proof of Theorem 4.1, there exists a constant $r > 0$ such that for all $k \geq 1$, the sequence $\{(y^k, z^k, x^k)\}$ generated by sPGADMM converges to a KKT point $\bar{u} = (\bar{y}, \bar{z}, \bar{x})$ if $\|u^k - \bar{u}\| \leq r$. With these u^k satisfying $\|R(u^k)\| > \zeta$, we have

$$\text{dist}(u^k, \bar{\Omega}) \leq \|u^k - \bar{u}\| \leq r < r(\zeta^{-1} \|R(u^k)\|).$$

Then (4.17) holds immediately with $\hat{\kappa} := \max\{\kappa, r\zeta^{-1}\}$ for all $k \geq 1$. The inequality (4.18) can be proved in a similar way to the one in Theorem 4.1. \square

To end this section, we note that the Corollary 4.1 is the global linear convergence rate of sPGADMM, and for convex composite quadratic programming, it holds with no additional conditions because the error bound assumption holds automatically.

5. CONCLUSIONS

The generalized ADMM for convex composite optimization problems was firstly proposed by Eckstein & Bertsekas [12] in 1992, and has been widely analyzed and implemented over the fast few decades. The convergence of this method can be obtained from the framework of PPA, but the proof from the aspect of optimization is relatively fewer. This paper considered a sPGADMM which covers almost all the existed generalized ADMMs as special cases, but importantly, the subproblems would be solved more efficiently with the help of the semi-proximal term. This paper focused on proving the linear convergence rate of sPGADMM under the condition that the mapping $R^{-1}(\cdot)$ is calmness at the original point for \bar{u} with a suitable modulus, because this condition is very mild and holds automatically for the considered problem (1.2). The required condition is same to the work of Han et al. [20] in establishing the linear convergence rate of the semi-proximal ADMM of Fazel et al. [15]. To some extent, our paper can be considered as an extension of the work of Han et al. [20] to the generalized variant of ADMM. However, we must note that, this extension is not trivial, because the relaxation factor has caused many obstacles for convergence analysis. Certainly, this is the motivation as well as the contribution of this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [Grant No. 11971149, 12271217, and 12301420], and the National Natural Science Foundation of Henan Province [Grant No. 232300421018].

REFERENCES

- [1] V.A. Adona, M. LN Goncalves, J.G. Melo, Iteration-complexity analysis of a generalized alternating direction method of multipliers, *J. Global Optim.* 73 (2019), 331-348.
- [2] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Computer Science and Applied Mathematics, Elsevier, 1982. doi: 10.1016/C2013-0-10366-2.
- [3] J.M. Borwein, A.S. Lewis, *Convex Analysis and Nonlinear Optimization*, Springer, New York, 2006. doi: 10.1007/978-0-387-31256-9.
- [4] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Mach. Learn.* 3 (2011), 1-122.
- [5] L. Chen, D.F. Sun, K.C. Toh, An efficient inexact symmetric Gauss–Seidel based majorized ADMM for high-dimensional convex composite conic programming, *Math. Program.* 161 (2017), 237-270.
- [6] L. Chen, D.F. Sun, K.C. Toh, A note on the convergence of ADMM for linearly constrained convex optimization problems, *Comput. Optim. Appl.* 66 (2017), 327-343.
- [7] E. Corman, X. Yuan, A generalized proximal point algorithm and its convergence rate, *SIAM J. Optim.* 24 (2014), 1614-1638.
- [8] A.L. Dontchev, R.T. Rockafellar, *Implicit Functions and Solution Mappings*, Springer, New York, 2009. doi: 10.1007/978-0-387-87821-8.
- [9] J. Douglas, H.H. Rachford, On the numerical solution of heat conduction problems in two and three space variables, *Trans. Maer. Math. Sco.* 82 (1956), 421-439.
- [10] J. Eckstein, Parallel alternating direction multiplier decomposition of convex programs, *J. Optim. Theory Appl.* 80 (1994), 39-62.

- [11] J. Eckstein, Some Saddle-function splitting methods for convex programming, *Optim. Meth. Softw.* 4 (1994), 75-83.
- [12] J. Eckstein, D.P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Program.* 55 (1992), 293-318.
- [13] J. Eckstein, W. Yao, Augmented Lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results, *Rutcor Research Reports* 32 (3): 44 (2012).
- [14] E.X. Fang, B.S. He, H. Liu, X.M. Yuan, Generalized alternating direction method of multipliers: new theoretical insights and applications, *Math Program. Comput.* 7 (2015), 149-187.
- [15] M. Fazel, T.K. Pong, D.F. Sun, P. Tsend, Hankel matrix rank minimization with applications to system identification and realization, *SIAM J. Matrix Anal. Appl.* 34 (2013), 946-977.
- [16] D. Gabay, Applications of the method of multipliers to variational inequalities, *Stud. Math. Appl.* 15 (1983), 299-331.
- [17] D. Gabay, B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation, *Comput. Math. Appl.* 2 (1976), 17-40.
- [18] R. Glowinski, *On Alternating Direction Methods of Multipliers: A Historical Perspective*, Modeling, Simulation and Optimization for Science and Technology, 2014.
- [19] R. Glowinski, A. Marroco, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires, *ESAIM: Math. Model. Numer. Anal.* 9 (1975), 41-76.
- [20] D.R. Han, D.F. Su, L.W. Zhang, Linear rate convergence of the alternating direction method of multipliers for convex composite programming, *Math. Oper. Res.* 43 (2018), 622-637.
- [21] M. Li, S.F. Sun, K.C. Toh, A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization, *SIAM J. Optim.* 26 (2016), 922-950.
- [22] X.D. Li, A two-phase augmented Lagrangian method for convex composite quadratic programming, PhD thesis, Department of Mathematics, National University of Singapore, 2015.
- [23] X.D. Li, S.F. Sun, K.C. Toh, A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions, *Math. Program.* 155 (2016), 333-373.
- [24] X.D. Li, D.F. Sun, K.C. Toh, QSDPNAL: A two-phase augmented Lagrangian method for convex quadratic semidefinite programming, *Math. Program. Comput.* 10 (2018), 703-743.
- [25] X.D. Li, D.F. Sun, K.C. Toh, A block symmetric Gauss-Seidel decomposition theorem for convex composite quadratic programming and its applications, *Math. Program.* 175 (2019), 395-418.
- [26] P.L. Lions, B. Mercier. Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.* 16 (1979), 964-979.
- [27] R.D.C. Monteiro, B.F. Svaiter, On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean, *SIAM J. Optim.* 20 (2010), 2755-2787.
- [28] J.W. Peng, D.X. Liu, X.Q. Zhang, The linear convergence analysis of the linearized version of the generalized alternating direction method of multipliers for convex optimization problems, arXiv:2202.09610, 2022.
- [29] J.W. Peng, X.Q. Zhang, Linear convergence rate of the generalized alternating direction method of multipliers for a class of convex minimization problems, Corpus ID: 52793813, 2017.
- [30] S.M. Robinson, An implicit-function theorem for generalized variational inequalities, Technical report, Mathematics Research Center, University of Wisconsin-Madison, 1976.
- [31] S.M. Robinson, Some continuity properties of polyhedral multifunctions, *Mathematical Programming at Oberwolfach*, pp. 206-214, 1981.
- [32] R.R. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Math. Oper. Res.* 1 (1976), 97-116.
- [33] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.* 14 (1976), 877-898.
- [34] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.
- [35] R.T. Rockafellar, R.J-B. Wets, *Variational Analysis*, Springer, 1998.
- [36] D.F. Sun, K.C. Toh, L.Q. Yang, A convergent 3-block semiproximal alternating direction method of multipliers for conic programming with 4-type constraints, *SIAM J. Optim.* 25 (2015), 882-915.

- [37] J. Su, On monotropic piecewise quadratic programming (Network, algorithm, convex programming, decomposition method), University of Washington, 1986.
- [38] M. Tao, X. Yuan, On the optimal linear convergence rate of a generalized proximal point algorithm, *J. Sci. Computing*, 74 (2018), 826-850.
- [39] C. Wu, Q. Wang, Y. Xiao, On the equivalence of peaceman-rachford splitting method and some typical alternating direction method of multipliers, *J. Nonlinear Convex Anal.* 19 (2018), 1933–1944.
- [40] Y. Xiao, L. Chen, D. Li, A generalized alternating direction method of multipliers with semi-proximal terms for convex composite conic programming, *Math. Program. Comput.* 10 (2018), 533–555.