

Linear Rate Convergence of the Alternating Direction Method of Multipliers for Convex Composite Programming

Deren Han,^a Defeng Sun,^b Liwei Zhang^c

^a School of Mathematical Sciences, Key Laboratory for NSLSCS of Jiangsu Province, Nanjing Normal University, Nanjing 210023, China;

^b Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong; ^c School of Mathematical Sciences, Dalian University of Technology, Dalian 116023, China

Contact: handeren@njnu.edu.cn (DH); defeng.sun@polyu.edu.hk,  <http://orcid.org/0000-0003-0481-272X> (DS); lwzhang@dlut.edu.cn (LZ)

Received: August 10, 2015

Revised: November 17, 2016

Accepted: May 3, 2017

Published Online in Articles in Advance:
December 15, 2017

MSC2010 Subject Classification: Primary:
90C25, 90C20, 90C22; secondary: 65K05

OR/MS Subject Classification: Primary:
mathematics, convexity; secondary:
programming, nonlinear

<https://doi.org/10.1287/moor.2017.0875>

Copyright: © 2017 INFORMS

Abstract. In this paper, we aim to prove the linear rate convergence of the alternating direction method of multipliers (ADMM) for solving linearly constrained convex composite optimization problems. Under a mild calmness condition, which holds automatically for convex composite piecewise linear-quadratic programming, we establish the global Q-linear rate of convergence for a general semi-proximal ADMM with the dual step-length being taken in $(0, (1 + 5^{1/2})/2)$. This semi-proximal ADMM, which covers the classic one, has the advantage to resolve the potentially nonsolvability issue of the subproblems in the classic ADMM and possesses the abilities of handling the multi-block cases efficiently. We demonstrate the usefulness of the obtained results when applied to two- and multi-block convex quadratic (semidefinite) programming.

Funding: The research of the first author was supported by the National Natural Science Foundation of China [Projects 11625105 and 11431002]; the research of the second author was supported in part by the Academic Research Fund [Grant R-146-000-207-112]; and the research of the third author was supported by the National Natural Science Foundation of China [Projects 11571059 and 91330206].

Keywords: ADMM • calmness • Q-linear convergence • multiblock • composite conic programming

1. Introduction

In this paper, we shall study the Q-linear rate convergence of the alternating direction method of multipliers (ADMM) for solving the following convex composite optimization problem:

$$\min\{\vartheta(y) + g(y) + \varphi(z) + h(z): \mathcal{A}^*y + \mathcal{B}^*z = c, y \in \mathcal{Y}, z \in \mathcal{Z}\}, \quad (1)$$

where \mathcal{Y} and \mathcal{Z} are two finite-dimensional real Euclidean spaces each equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|$, $\vartheta: \mathcal{Y} \rightarrow (-\infty, +\infty]$ and $\varphi: \mathcal{Z} \rightarrow (-\infty, +\infty]$ are two proper closed convex functions, $g: \mathcal{Y} \rightarrow (-\infty, +\infty)$ and $h: \mathcal{Z} \rightarrow (-\infty, +\infty)$ are two continuously differentiable convex functions (e.g., convex quadratic functions), $\mathcal{A}: \mathcal{Y} \rightarrow \mathcal{X}$ and $\mathcal{B}: \mathcal{Z} \rightarrow \mathcal{X}$ are the adjoints of the two linear operators $\mathcal{A}: \mathcal{X} \rightarrow \mathcal{Y}$ and $\mathcal{B}: \mathcal{X} \rightarrow \mathcal{Z}$, respectively, with \mathcal{X} being another real finite-dimensional Euclidean space equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|$ and $c \in \mathcal{X}$ is a given point. To avoid triviality, neither \mathcal{A} nor \mathcal{B} is assumed to be vacuous. For any convex function $\theta: \mathcal{X} \rightarrow (-\infty, +\infty]$, we use $\text{dom } \theta$ to define its effective domain, i.e., $\text{dom } \theta := \{x \in \mathcal{X}: \theta(x) < \infty\}$, $\text{epi } \theta$ to denote its epigraph, i.e., $\text{epi } \theta := \{(x, t) \in \mathcal{X} \times \mathbb{R}: \theta(x) \leq t\}$ and $\theta^*: \mathcal{X} \rightarrow (-\infty, +\infty]$ to represent its Fenchel conjugate, respectively.

The classic ADMM was designed by Glowinski and Marroco [23] and Gabay and Mercier [20] and its construction was much influenced by Rockafellar's works on proximal point algorithms (PPAs) for solving the more general maximal monotone inclusion problems (Rockafellar [37, 38]). The readers may refer to Glowinski [22] for a note on the historical development of the classic ADMM. The convergence analysis for the classic ADMM under certain settings was first conducted by Gabay and Mercier [20], Glowinski [21], and Fortin and Glowinski [17]. For a recent survey on this, see Eckstein and Yao [15].

Our focus of this paper is on the linear rate convergence analysis of the ADMM. This shall be conducted under a more convenient semi-proximal ADMM (in short, sPADMM) setting proposed by Fazel et al. [16] by allowing the dual step-length to be at least as large as the golden ratio of 1.618. This sPADMM, which covers the classic ADMM, has the advantage to resolve the potentially nonsolvability issue of the subproblems in the classic ADMM. But, perhaps more importantly, it possesses the abilities of handling multiblock convex optimization problems. For example, it has been shown most recently that the sPADMM plays a pivotal role in

solving multiblock convex composite semidefinite programming problems (Sun et al. [41], Li et al. [30], Chen et al. [5]) of a low to medium accuracy. We shall come back to this in Section 4.

For any self-adjoint positive semidefinite linear operator $\mathcal{M}: \mathcal{X} \rightarrow \mathcal{X}$, denote $\|x\|_{\mathcal{M}} := \sqrt{\langle x, \mathcal{M}x \rangle}$ and $\text{dist}_{\mathcal{M}}(x, D) = \inf_{x' \in D} \|x' - x\|_{\mathcal{M}}$ for any $x \in \mathcal{X}$ and any set $D \subseteq \mathcal{X}$. We use \mathcal{I} to denote the identity mapping from \mathcal{X} to itself. Let $\sigma > 0$ be a given parameter. Write $\mathfrak{g}_g(\cdot) \equiv \mathfrak{g}(\cdot) + g(\cdot)$ and $\varphi_h(\cdot) \equiv \varphi(\cdot) + h(\cdot)$. The augmented Lagrangian function of problem (1) is defined by

$$\mathcal{L}_{\sigma}(y, z; x) := \mathfrak{g}_g(y) + \varphi_h(z) + \langle x, \mathcal{A}^*y + \mathcal{B}^*z - c \rangle + \frac{\sigma}{2} \|\mathcal{A}^*y + \mathcal{B}^*z - c\|^2, \quad \forall (y, z, x) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}. \quad (2)$$

Then the sPADMM may be described as follows.

sPADMM: A semi-proximal alternating direction method of multipliers for solving the convex optimization problem (1).

Step 0. Input $(y^0, z^0, x^0) \in \text{dom } \mathfrak{g} \times \text{dom } \varphi \times \mathcal{X}$. Let $\tau \in (0, +\infty)$ be a positive parameter (e.g., $\tau \in (0, (1 + \sqrt{5})/2)$), and $\mathcal{S}: \mathcal{Y} \rightarrow \mathcal{Y}$ and $\mathcal{T}: \mathcal{Z} \rightarrow \mathcal{Z}$ be two self-adjoint positive semidefinite, not necessarily positive definite, linear operators. Set $k := 0$.

Step 1. Set

$$\begin{cases} y^{k+1} \in \arg \min \mathcal{L}_{\sigma}(y, z^k; x^k) + \frac{1}{2} \|y - y^k\|_{\mathcal{S}}^2, & (3a) \\ z^{k+1} \in \arg \min \mathcal{L}_{\sigma}(y^{k+1}, z; x^k) + \frac{1}{2} \|z - z^k\|_{\mathcal{T}}^2, & (3b) \end{cases}$$

$$x^{k+1} = x^k + \tau \sigma (\mathcal{A}^*y^{k+1} + \mathcal{B}^*z^{k+1} - c). \quad (3c)$$

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

The sPADMM scheme (3a)–(3c) with $\mathcal{S} = 0$ and $\mathcal{T} = 0$ is nothing but the classic ADMM of Glowinski and Marroco [23] and Gabay and Mercier [20]. When $\mathcal{B} = \mathcal{I}$ and \mathcal{A} is surjective, the global convergence of the classic ADMM with any $\tau \in (0, (1 + \sqrt{5})/2)$ has been established by Glowinski [21] and Fortin and Glowinski [17]. Interestingly, in Gabay [19], Gabay has further shown that the classic ADMM with $\tau = 1$, under the existence condition of a solution to the Karush-Kuhn-Tucker (KKT) system of problem (1), is actually equivalent to the Douglas-Rachford (DR) splitting method applied to a stationary system to the dual of problem (1). Moreover, Eckstein and Berstekas [14] have proven that the DR splitting method can be equivalently represented as a special case of PPA. This is achieved by using a splitting operator constructed by Eckstein in his PhD thesis Eckstein [12], which we will call the Eckstein splitting operator for the convenience of reference. Thus, one may always use known results on the DR splitting method and the PPA to study the properties of the classic ADMM with $\tau = 1$ (this does not apply to the case that $\tau \neq 1$ of course) though the properties on the corresponding Eckstein splitting operator can be much involved.

The above sPADMM scheme (3a)–(3c) with $\mathcal{S} > 0$ and $\mathcal{T} > 0$ was initiated by Eckstein [13] to make the subproblems in (3a) and (3b) easier to solve. In the same paper, Eckstein [13] showed how the sPADMM with $\mathcal{S} > 0$ and $\mathcal{T} > 0$ can be transformed into the framework of PPAs. In He et al. [26], He et al. further studied an inexact version of Eckstein's work in the context of monotone variational inequalities. Using essentially the same variational techniques developed by Glowinski [21] and Fortin and Glowinski [17], Fazel et al. developed an extremely easy-to-use convergence theorem, which covers earlier nice results of Xu and Wu [43] with $\mathcal{S} > 0$ and/or $\mathcal{T} > 0$, for the sPADMM (Fazel et al. [16, Appendix B]) when the dual step-length τ is chosen in $(0, (1 + \sqrt{5})/2)$. In Shefi and Teboulle [40], Shefi and Teboulle conducted a comprehensive study on the iteration complexities, in particular in the ergodic sense, for the sPADMM with $\tau = 1$ and $\mathcal{B} \equiv \mathcal{I}$. Related results for the more general cases can be found, e.g., in Li et al. [28] for the case that the linear operators \mathcal{S} and \mathcal{T} are allowed to be indefinite, and in Cui et al. [8] for the case that the objective function is allowed to have a coupled smooth term. For details on choosing \mathcal{S} and \mathcal{T} , one may refer to the recent PhD thesis of Li [29].

Compared with the large amount of literature¹ mainly being devoted to the applications of the ADMM, there is a much smaller number of papers targeting the linear rate, in particular the Q-linear rate, convergence analysis, though there do exist a number of classic results and several new advancements on the latter. By using the aforementioned connections among the DR splitting method, PPAs, and the classic ADMM with $\tau = 1$, we can derive the corresponding R-linear rate convergence of the ADMM from the works of Lions and Mercier [31] on the DR splitting method with a globally Lipschitz continuous and strongly monotone operator, and Rockafellar [37, 38] and Luque [32] on the convergence rates of the PPAs under various error bound conditions imposed on the Eckstein splitting operator. For example, within this spirit, Eckstein [12] proved

the global R-linear convergence rate of the ADMM with $\tau = 1$ when it is applied to linear programming. In the same vein, one can easily obtain the similar global R-linear convergence rate of the ADMM with $\tau = 1$ for convex piecewise linear-quadratic programming by combining the classic result of Robinson on piecewise polyhedral multivalued mappings Robinson [35] and Sun's subdifferential characterization of convex piecewise linear-quadratic functions (Sun [42]).

There are some interesting new developments on the R-linear and/or Q-Linear convergence rate of the ADMM. Apparently, unaware of the above mentioned connections, for convex quadratic programming, Boley [1] provided a local R-linear convergence result for the ADMM with $\tau = 1$ under the conditions of the uniqueness of the optimal solutions to both the primal and dual problems and the strict complementarity; in Han and Yuan [24], Han and Yuan removed the restrictive conditions imposed by Boley and established the local Q-linear rate convergence of the generalized ADMM in the sense of Eckstein and Bertsekas [14] for the sequence $\{(z^k, x^k)\}$. For the more general convex piecewise linear-quadratic programming problems, Yang and Han [44] established the global Q-linear convergence rate for the sequence $\{(z^k, x^k)\}$ and $\{(y^k, z^k, x^k)\}$ of the ADMM and the linearized ADMM (a special case of sPADMM with $\mathcal{P} > 0$ and $\mathcal{T} > 0$) with $\tau = 1$, respectively. We remark that when either $\mathcal{P} > 0$ or $\mathcal{T} > 0$ fails to hold, the convergence analysis in Yang and Han [44] for the linearized ADMM is no longer valid. In Deng and Yin [9], Deng and Yin provided a number of scenarios on both the R-linear and Q-linear rate convergence for the ADMM and sPADMM with $\tau = 1$ under the assumption that either $\vartheta_g(\cdot)$ or $\varphi_h(\cdot)$ is strongly convex with a Lipschitz continuous gradient in addition to the boundedness condition on the generated iteration sequence and others. Deng and Yin also provided a detailed comparison between their most notable R-linear rate convergence result and that of Lions and Mercier [31] on the DR splitting method when applied to a stationary system to the dual of problem (1). Assuming an error bound condition and some others, Hong and Luo [27] provided a global R-linear rate convergence of the multiblock ADMM and its variants with a sufficiently small step-length τ . Theoretically, this may constitute important progress on understanding the convergence and the linear rate of convergence of the ADMM. Computationally, however, this is far from being satisfactory, as in practical implementations one always prefers a larger step-length for achieving numerical efficiency.

In this paper, we aim to resolve the Q-linear rate convergence issue for the sPADMM scheme (3a)–(3c) with $\tau \in (0, (1 + \sqrt{5})/2)$ under mild conditions. Special attention shall be paid to convex composite piecewise linear-quadratic programming and quadratic semidefinite programming. Under a calmness condition only, we provide a global Q-linear rate convergence analysis for the sPADMM with $\tau \in (0, (1 + \sqrt{5})/2)$. This is made possible by constructing an elegant inequality on the iteration sequence via reorganizing the relevant results developed in Fazel et al. [16, Appendix B]. For convex composite piecewise linear-quadratic programming, the global Q-linear convergence rate is obtained with no additional conditions as the calmness assumption holds automatically. By choosing the positive semidefinite linear operators \mathcal{P} and \mathcal{T} properly, in particular $\mathcal{T} = 0$, we demonstrate how the established global Q-linear rate convergence of the sPADMM can be applied to multiblock convex composite quadratic conic programming.

The remaining parts of this paper are organized as follows. In Section 2, we conduct brief discussions on the optimality conditions for problem (1) and on both the locally upper Lipschitz continuity and the calmness for multivalued mappings. Section 3 is divided into two parts with the first part focusing on the derivation of a particularly useful inequality for the iteration sequence generated from the sPADMM. This inequality, which grows out of the results in Fazel et al. [16, Appendix B], is then employed to build up a general Q-linear rate convergence theorem under a calmness condition. Section 4 is about the applications of the Q-linear convergence theorem of the sPADMM to important convex composite quadratic conic programming problems. We make our final conclusions in Section 5.

2. Preliminaries

In this section, we summarize some useful preliminaries for our subsequent analysis.

2.1. Optimality Conditions

For a multifunction $F: \mathcal{Y} \rightrightarrows \mathcal{Y}$, we say that F is monotone if

$$\langle y' - y, \xi' - \xi \rangle \geq 0, \quad \forall \xi' \in F(y'), \quad \forall \xi \in F(y). \quad (4)$$

It is well known that for any proper closed convex function $\theta: \mathcal{X} \rightarrow (-\infty, +\infty]$, $\partial\theta(\cdot)$ is a monotone multivalued function (see Rockafellar [36]), that is, for any $w_1 \in \text{dom } \theta$ and any $w_2 \in \text{dom } \theta$,

$$\langle \xi - \zeta, w_1 - w_2 \rangle \geq 0, \quad \forall \xi \in \partial\theta(w_1), \quad \forall \zeta \in \partial\theta(w_2). \quad (5)$$

In our analysis, we shall often use the optimality conditions for problem (1). Let $(\bar{y}, \bar{z}) \in \text{dom } \vartheta \times \text{dom } \varphi$ be an optimal solution to problem (1). If there exists $\bar{x} \in \mathcal{X}$ such that $(\bar{y}, \bar{z}, \bar{x})$ satisfies the following KKT system:

$$\begin{cases} 0 \in \partial \vartheta(y) + \nabla g(y) + \mathcal{A}x, \\ 0 \in \partial \varphi(z) + \nabla h(z) + \mathcal{B}x, \\ c - \mathcal{A}^*y - \mathcal{B}^*z = 0, \end{cases} \quad (6)$$

then $(\bar{y}, \bar{z}, \bar{x})$ is called a KKT point for problem (1). Denote the solution set to the KKT system (6) by $\bar{\Omega}$. The existence of such KKT points can be guaranteed if a certain constraint qualification such as the Slater condition holds:

$$\exists (y', z') \in \text{ri}(\text{dom } \vartheta \times \text{dom } \varphi) \cap \{(y, z) \in \mathcal{Y} \times \mathcal{Z} : \mathcal{A}^*y + \mathcal{B}^*z = c\},$$

where $\text{ri}(S)$ denotes the relative interior of a given convex set S . In this paper, instead of using an explicit constraint qualification, we make the following blanket assumption on the existence of a KKT point.

Assumption 1. The KKT system (6) has a nonempty solution set.

Denote $u := (y, z, x)$ for $y \in \mathcal{Y}$, $z \in \mathcal{Z}$ and $x \in \mathcal{X}$. Let $\mathcal{U} := \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$. Define the KKT mapping $R: \mathcal{U} \rightarrow \mathcal{U}$ as

$$R(u) := \begin{pmatrix} y - \text{Pr}_{\mathcal{S}}[y - (\nabla g(y) + \mathcal{A}x)] \\ z - \text{Pr}_{\varphi}[z - (\nabla h(z) + \mathcal{B}x)] \\ c - \mathcal{A}^*y - \mathcal{B}^*z \end{pmatrix}, \quad \forall u \in \mathcal{U}, \quad (7)$$

where for any convex function $\theta: \mathcal{X} \rightarrow (-\infty, +\infty]$, $\text{Pr}_{\theta}(\cdot)$ denotes its associated Moreau-Yosida proximal mapping Rockafellar and Wets [39]. If $\theta(\cdot) = \delta_{\mathcal{K}}(\cdot)$, the indicator function over the closed convex set $\mathcal{K} \subseteq \mathcal{X}$, then $\text{Pr}_{\theta}(\cdot) = \Pi_{\mathcal{K}}(\cdot)$, the metric projection operator over \mathcal{K} . Then, since the Moreau-Yosida proximal mappings $\text{Pr}_{\mathcal{S}}(\cdot)$ and $\text{Pr}_{\varphi}(\cdot)$ are both globally Lipschitz continuous with modulus one, the mapping $R(\cdot)$ is at least continuous on \mathcal{U} and

$$\forall u \in \mathcal{U}, \quad R(u) = 0 \Leftrightarrow u \in \bar{\Omega}.$$

2.2. Locally Upper Lipschitz Continuity and Calmness

Let \mathcal{X} and \mathcal{Y} be two finite-dimensional real Euclidean spaces and $F: \mathcal{X} \rightrightarrows \mathcal{Y}$ be a set-valued mapping. Denote the graph of F by $\text{gph } F$. Let $\mathbf{B}_{\mathcal{Y}}$ denote the unit ball in \mathcal{Y} .

Definition 1. The multivalued mapping $F: \mathcal{X} \rightrightarrows \mathcal{Y}$ is said to be *locally upper Lipschitz continuous* at $x^0 \in \mathcal{X}$ with modulus $\kappa_0 > 0$ if there exist a neighborhood V of x^0 such that

$$F(x) \subseteq F(x^0) + \kappa_0 \|x - x^0\| \mathbf{B}_{\mathcal{Y}}, \quad \forall x \in V.$$

The above definition of locally upper Lipschitz continuity was first coined by Robinson [34] for the purpose of developing an implicit function for generalized variational inequalities. In the same paper, he also studied several important properties of multivalued mappings. Recall that the multivalued mapping F is called piecewise polyhedral if $\text{gph } F$ is the union of finitely many polyhedral sets. In one of his seminal papers, Robinson [35] established the following fundamental property on the locally upper Lipschitz continuity of a piecewise polyhedral multivalued mapping.

Proposition 1. If the 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 κ_0 independent of the choice of x^0 .

One important class of piecewise polyhedral multivalued mappings is the subdifferential of convex piecewise linear-quadratic functions. Note that a closed proper convex function $\theta: \mathcal{X} \rightarrow (-\infty, +\infty]$ is said to be piecewise linear-quadratic if $\text{dom } \theta$ is the union of finitely many polyhedral sets and on each of these polyhedral sets, θ is either an affine or a quadratic function. In his PhD thesis, J. Sun (Sun [42]) established the following key characterization on convex piecewise linear-quadratic functions. For a complete proof about this proposition and its extensions, see the monograph written by Rockafellar and Wets [39, Propositions 12.30 and 11.14].

Proposition 2. Let $\theta: \mathcal{X} \rightarrow (-\infty, +\infty]$ be a closed proper convex function. Then θ is piecewise linear quadratic if and only if the graph of $\partial \theta$ is piecewise polyhedral. Moreover, θ is piecewise linear quadratic if and only if θ^* is piecewise linear quadratic.

Next, we give the definition of calmness for $F: \mathcal{X} \rightrightarrows \mathcal{Y}$ at x^0 for y^0 with $(x^0, y^0) \in \text{gph} F$.

Definition 2. Let $(x^0, y^0) \in \text{gph} F$. The multivalued mapping $F: \mathcal{X} \rightrightarrows \mathcal{Y}$ is said to be *calm at x^0 for y^0 with modulus $\kappa_0 \geq 0$* if there exist a neighborhood V of x^0 and a neighborhood W of y^0 such that

$$F(x) \cap W \subseteq F(x^0) + \kappa_0 \|x - x^0\| \mathbf{B}_{\mathcal{Y}}, \quad \forall x \in V.$$

The above definition of calmness is taken from Dontchev and Rockafellar [11, Section 3.8(3H)]. It follows from Proposition 1 that if $F: \mathcal{X} \rightrightarrows \mathcal{Y}$ is piecewise polyhedral, and in particular from Proposition 2 that F is the subdifferential mapping of a convex piecewise linear-quadratic function, then F is calm at x^0 for y^0 satisfying $(x^0, y^0) \in \text{gph} F$ with modulus $\kappa_0 > 0$ independent of the choice of (x^0, y^0) . Furthermore, it is well known, e.g., Dontchev and Rockafellar [11, 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 $\kappa'_0 \geq 0$, a neighborhood W of y^0 , and a neighborhood V of x^0 such that

$$\text{dist}(y, F(x^0)) \leq \kappa'_0 \text{dist}(x^0, F^{-1}(y) \cap V), \quad \forall y \in W. \quad (8)$$

3. A General Theorem on the Q-Linear Rate Convergence

In this section, we shall establish a general theorem on the Q-linear convergence rate of the sPADMM scheme (3a)–(3c).

First we recall the global convergence of the sPADMM from Fazel et al. [16, Appendix B]. Since both $\partial \vartheta$ and $\partial \varphi$ are maximally monotone and g and h are two continuously differentiable convex functions, there exist two self-adjoint and positive semidefinite linear operators Σ_g and Σ_h such that for all $y', y \in \text{dom } \vartheta_g$, $\xi \in \partial \vartheta_g(y)$ and $\xi' \in \partial \vartheta_g(y')$, and for all $z', z \in \text{dom } \varphi_h$, $\zeta \in \partial \varphi_h(z)$ and $\zeta' \in \partial \varphi_h(z')$,

$$\langle \xi' - \xi, y' - y \rangle \geq \|y' - y\|_{\Sigma_g}^2, \quad \langle \zeta' - \zeta, z' - z \rangle \geq \|z' - z\|_{\Sigma_h}^2. \quad (9)$$

For notational convenience, let $\mathcal{E}: \mathcal{X} \rightarrow \mathcal{U} := \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$ be a linear operator such that its adjoint \mathcal{E}^* satisfies $\mathcal{E}^*(y, z, x) = \mathcal{A}^*y + \mathcal{B}^*z$ for any $(y, z, x) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$ and for $u := (y, z, x) \in \mathcal{U}$ and $u' := (y', z', x') \in \mathcal{U}$, define the following function to measure the weighted distance of two points:

$$\theta(u, u') := (\tau\sigma)^{-1} \|x - x'\|^2 + \|y - y'\|_{\mathcal{Y}}^2 + \|z - z'\|_{\mathcal{Z}}^2 + \sigma \|\mathcal{B}^*(z - z')\|^2.$$

Theorem 1, which will be used in the following, is adapted from Appendix B of Fazel et al. [16].

Theorem 1. Let Assumption 1 be satisfied. Suppose that the sPADMM generates a well-defined infinite sequence $\{u^k\}$. Let $\bar{u} = (\bar{y}, \bar{z}, \bar{x}) \in \bar{\Omega}$. For $k \geq 1$, denote

$$\begin{cases} \delta_k := \tau(1 - \tau + \min\{\tau, \tau^{-1}\})\sigma \|\mathcal{B}^*(z^k - z^{k-1})\|^2 + \|z^k - z^{k-1}\|_{\mathcal{Z}}^2, \\ \nu_k := \delta_k + \|y^k - y^{k-1}\|_{\mathcal{Y}}^2 + 2\|y^k - \bar{y}\|_{\Sigma_g}^2 + 2\|z^k - \bar{z}\|_{\Sigma_h}^2. \end{cases} \quad (10)$$

Then, the following results hold:

(i) For any $k \geq 1$,

$$\begin{aligned} & \left[\theta(u^{k+1}, \bar{u}) + \|z^{k+1} - \bar{z}\|_{\mathcal{Z}}^2 + (1 - \min\{\tau, \tau^{-1}\})\sigma \|\mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c\|^2 \right] \\ & \quad - \left[\theta(u^k, \bar{u}) + \|z^k - \bar{z}\|_{\mathcal{Z}}^2 + (1 - \min\{\tau, \tau^{-1}\})\sigma \|\mathcal{E}^*(y^k, z^k, 0) - c\|^2 \right] \\ & \leq -\left[\nu_{k+1} + (1 - \tau + \min\{\tau, \tau^{-1}\})\sigma \|\mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c\|^2 \right]. \end{aligned} \quad (11)$$

(ii) Assume that both $\Sigma_g + \mathcal{S} + \sigma \mathcal{A} \mathcal{A}^*$ and $\Sigma_h + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^*$ are positive definite so that the sequence $\{u^k\}$ is automatically well defined. If $\tau \in (0, (1 + \sqrt{5})/2)$, then the whole sequence $\{(y^k, z^k, x^k)\}$ converges to a KKT point in $\bar{\Omega}$.

Theorem 1 provides global convergence results for the sPADMM under fairly general and mild conditions. For the purpose of obtaining inequality (11) in Theorem 1, one needs to assume that the subproblems in the sPADMM must admit optimal solutions, which can be guaranteed if both $\Sigma_g + \mathcal{S} + \sigma \mathcal{A} \mathcal{A}^*$ and $\Sigma_h + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^*$ are positive definite. Obviously one can always choose positive semidefinite linear operators \mathcal{S} and \mathcal{T} to ensure $\Sigma_g + \mathcal{S} + \sigma \mathcal{A} \mathcal{A}^* > 0$ and $\Sigma_h + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^* > 0$ as $\Sigma_g + \sigma \mathcal{A} \mathcal{A}^* \geq 0$ and $\Sigma_h + \sigma \mathcal{B} \mathcal{B}^* \geq 0$. In the classic ADMM, since both $\mathcal{S} = 0$ and $\mathcal{T} = 0$, one needs to assume that $\Sigma_g + \sigma \mathcal{A} \mathcal{A}^* > 0$ and $\Sigma_h + \sigma \mathcal{B} \mathcal{B}^* > 0$, which hold automatically if the surjectivity of both \mathcal{A} and \mathcal{B} is assumed as in the original ADMM papers Glowinski [21] and Fortin and

Glowinski [17]. An example was constructed by Chen et al. [6] to show that Assumption 1 itself is not enough for ensuring the existence of solutions to all the subproblems in the sPADMM. This example also shows that the statement made in Boyd et al. [3] on the convergence of the classic ADMM without the surjectivity of \mathcal{A} and \mathcal{B} is incorrect. Interestingly, for the case that $\mathcal{S} = 0$ and $\mathcal{T} = 0$, the global convergence results in Theorem 1 may still be valid even if the surjectivity of \mathcal{A} or \mathcal{B} fails to hold if Σ_f and Σ_g are incorporated in the analysis. To illustrate this, let us consider the following convex composite least squares optimization problem:

$$\begin{aligned} \min \quad & \{\tfrac{1}{2}\|Lx - d\|^2 + \phi(x)\} \\ \text{s.t.} \quad & Ax = b, \end{aligned} \quad (12)$$

where $d \in \mathbb{R}^l$, $b \in \mathbb{R}^m$, $L \in \mathbb{R}^{l \times n}$ and $A \in \mathbb{R}^{m \times n}$ are two matrices and $\phi: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ is a proper closed convex function. Without loss of generality, assume that A is of full row rank. The Lagrange dual of (12) can be written as

$$\begin{aligned} \max \quad & \{-\tfrac{1}{2}\|w - d\|^2 + b^T y_E - \phi^*(z)\} \\ \text{s.t.} \quad & L^T w + A^T y_E - z = 0, \end{aligned}$$

which can be equivalently reformulated as

$$\begin{aligned} \min \quad & \{\tfrac{1}{2}\|v\|^2 - b^T y_E + \phi^*(z)\} \\ \text{s.t.} \quad & L^T v + A^T y_E - z = -L^T d. \end{aligned} \quad (13)$$

By treating (v, y_E) as one-variable block and z the other, we can write problem (13) in the form of (1) with

$$g(v, y_E) := \tfrac{1}{2}\|v\|^2 - b^T y_E, \quad \forall (v, y_E) \in \mathbb{R}^l \times \mathbb{R}^m \quad \text{and} \quad h(z) := 0, \quad \forall z \in \mathbb{R}^n. \quad (14)$$

Consequently, one immediately obtains

$$\Sigma_g \equiv \begin{pmatrix} I_l & 0 \\ 0 & 0 \end{pmatrix}, \quad \Sigma_h \equiv 0, \quad (15)$$

where for any positive integer j , I_j denotes the j by j identity matrix. Note that in this case $\mathcal{A} := \begin{pmatrix} L \\ A \end{pmatrix}$ is not necessarily surjective though $\mathcal{B} = -I_n$ is. So the known global convergence analysis for the classic ADMM without using Σ_g may be invalid. However, since both $\Sigma_g + \sigma\mathcal{A}\mathcal{A}^*$ and $\Sigma_h + \sigma\mathcal{B}\mathcal{B}^*$ are positive definite, the global convergence of the classic ADMM (both $\mathcal{S} = 0$ and $\mathcal{T} = 0$) for solving problem (13) follows readily from Theorem 1. Thus, one can see the benefit of exploiting the availability of Σ_g or Σ_h .

For any self-adjoint linear operator $\mathcal{M}: \mathcal{X} \rightarrow \mathcal{X}$, we use $\lambda_{\max}(\mathcal{M})$ to denote its largest eigenvalue. Define

$$\kappa_1 := 3\|\mathcal{S}\|, \quad \kappa_2 := \max\{3\sigma\lambda_{\max}(\mathcal{A}\mathcal{A}^*), 2\|\mathcal{T}\|\}, \quad \kappa_3 := \sigma^{-1} + (1 - \tau)^2\sigma(3\lambda_{\max}(\mathcal{A}\mathcal{A}^*) + 2\lambda_{\max}(\mathcal{B}\mathcal{B}^*)).$$

Let $\kappa_4 := \max\{\kappa_1, \kappa_2, \kappa_3\}$. Let \mathcal{H}_0 be the block-diagonal linear operator defined by

$$\mathcal{H}_0 := \kappa_4 \text{Diag}(\mathcal{S}, \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^*, (\tau^2\sigma)^{-1}\mathcal{I}). \quad (16)$$

The usefulness of the block-diagonal linear operator \mathcal{H}_0 can be found in the following lemma on deriving an upper bound for $\|R(\cdot)\|$ computed at the sequence generated by the sPADMM.

Lemma 1. Let $\{u^k := (y^k, z^k, x^k)\}$ be the infinite sequence generated by the sPADMM scheme (3a)–(3c). Then for any $k \geq 0$,

$$\|u^{k+1} - u^k\|_{\mathcal{H}_0}^2 \geq \|R(u^{k+1})\|^2. \quad (17)$$

Proof. The optimality condition for (3a) is

$$0 \in \partial\vartheta(y^{k+1}) + \nabla g(y^{k+1}) + \mathcal{A}[x^k + \sigma(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^k - c)] + \mathcal{S}(y^{k+1} - y^k). \quad (18)$$

From the definition of x^{k+1} , we have

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

It then follows from (18) that

$$0 \in \partial\vartheta(y^{k+1}) + \nabla g(y^{k+1}) + \mathcal{A}[\sigma\mathcal{B}^*(z^k - z^{k+1}) + x^k + \tau^{-1}(x^{k+1} - x^k)] + \mathcal{S}(y^{k+1} - y^k),$$

which implies

$$y^{k+1} = \text{Pr}_{\mathcal{Y}}(y^{k+1} - (\nabla g(y^{k+1}) + \mathcal{A}[\sigma \mathcal{B}^*(z^k - z^{k+1}) + x^k + \tau^{-1}(x^{k+1} - x^k)] + \mathcal{P}(y^{k+1} - y^k))). \quad (19)$$

Noting that since z^{k+1} is a solution to the subproblem (3b), we have that

$$0 \in \partial \varphi(z^{k+1}) + \nabla h(z^{k+1}) + \mathcal{B}x^k + \sigma \mathcal{B}(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^{k+1} - c) + \mathcal{F}(z^{k+1} - z^k),$$

which is equivalent to

$$0 \in \partial \varphi(z^{k+1}) + \nabla h(z^{k+1}) + \mathcal{B}[x^k + \tau^{-1}(x^{k+1} - x^k)] + \mathcal{F}(z^{k+1} - z^k).$$

Thus, we have

$$z^{k+1} = \text{Pr}_{\mathcal{Z}}(z^{k+1} - (\nabla h(z^{k+1}) + \mathcal{B}[x^k + \tau^{-1}(x^{k+1} - x^k)] + \mathcal{F}(z^{k+1} - z^k))). \quad (20)$$

Note that from (3c),

$$x^{k+1} = x^k + \tau \sigma (\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^{k+1} - c). \quad (21)$$

Then, by combining (19)–(21) and noticing of the Lipschitz continuity of the Moreau-Yosida proximal mappings, we obtain from the definition of $R(\cdot)$ in (7) that

$$\begin{aligned} \|R(u^{k+1})\|^2 &\leq \| -\mathcal{P}(y^{k+1} - y^k) + \sigma \mathcal{A} \mathcal{B}^*(z^{k+1} - z^k) + (1 - \tau^{-1}) \mathcal{A}(x^{k+1} - x^k) \|^2 \\ &\quad + \| -\mathcal{F}(z^{k+1} - z^k) + (1 - \tau^{-1}) \mathcal{B}(x^{k+1} - x^k) \|^2 + (\tau \sigma)^{-2} \|x^{k+1} - x^k\|^2 \\ &\leq [3\|\mathcal{P}\| \|y^{k+1} - y^k\|_{\mathcal{Y}}^2 + 3\sigma^2 \lambda_{\max}(\mathcal{A} \mathcal{A}^*) \|\mathcal{B}^*(z^{k+1} - z^k)\|^2 + 3(1 - \tau^{-1})^2 \|\mathcal{A}(x^{k+1} - x^k)\|^2] \\ &\quad + [2\|\mathcal{F}\| \|z^{k+1} - z^k\|_{\mathcal{Z}}^2 + 2(1 - \tau^{-1})^2 \|\mathcal{B}(x^{k+1} - x^k)\|^2 + (\tau \sigma)^{-2} \|x^{k+1} - x^k\|^2] \\ &\leq \kappa_1 \|y^{k+1} - y^k\|_{\mathcal{Y}}^2 + \kappa_2 \|z^{k+1} - z^k\|_{\mathcal{Z} + \sigma \mathcal{B} \mathcal{B}^*}^2 + \kappa_3 (\tau^2 \sigma)^{-1} \|x^{k+1} - x^k\|^2, \end{aligned}$$

which immediately implies (17). \square

For any $\tau \in (0, +\infty)$, define

$$s_{\tau} := \frac{5 - \tau - 3 \min\{\tau, \tau^{-1}\}}{4} \quad \text{and} \quad t_{\tau} := \frac{1 - \tau + \min\{\tau, \tau^{-1}\}}{2}.$$

Note that one can easily compute the following

$$1/4 \leq s_{\tau} \leq 5/4 \quad \text{and} \quad 0 < t_{\tau} \leq 1/2, \quad \forall \tau \in (0, (1 + \sqrt{5})/2). \quad (22)$$

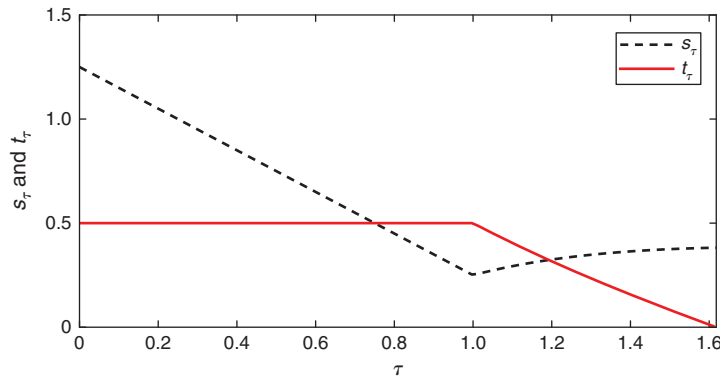
See Figure 1 for the values of s_{τ} and t_{τ} for $\tau \in [0, (1 + \sqrt{5})/2]$.

Denote the following two self-adjoint linear operators for our subsequent developments:

$$\mathcal{M} := \text{Diag}(\mathcal{P} + \Sigma_g, \mathcal{F} + \Sigma_h + \sigma \mathcal{B} \mathcal{B}^*, (\tau \sigma)^{-1} \mathcal{F}) + s_{\tau} \sigma \mathcal{E} \mathcal{E}^*, \quad (23)$$

$$\mathcal{H} := \text{Diag}(\mathcal{P} + \frac{1}{2} \Sigma_g, \mathcal{F} + \frac{1}{2} \Sigma_h + 2t_{\tau} \tau \sigma \mathcal{B} \mathcal{B}^*, t_{\tau} (\tau^2 \sigma)^{-1} \mathcal{F}) + \frac{1}{4} t_{\tau} \sigma \mathcal{E} \mathcal{E}^*. \quad (24)$$

Figure 1. (Color online) The values of s_{τ} and t_{τ} for $\tau \in [0, (1 + \sqrt{5})/2]$.



Then we immediately get the following relation

$$\kappa_4 \mathcal{H} \geq \min\{2\tau, 1\} t_\tau \mathcal{H}_0 + \frac{1}{4} \kappa_4 t_\tau \sigma \mathcal{E} \mathcal{E}^*, \quad \forall \tau \in (0, +\infty). \quad (25)$$

The operator \mathcal{M} will be used to define the weighted distance from an iterate to the KKT points while \mathcal{H} will be employed to measure the weighted distance between two consecutive iterates. The next proposition will answer the needed positive definiteness of these two linear operators, which is made possible due to the introduction of the last term in (23) and (24), respectively.

Proposition 3. Let $\tau \in (0, (1 + \sqrt{5})/2)$. Then

$$(\Sigma_g + \mathcal{P} + \sigma \mathcal{A} \mathcal{A}^* > 0 \text{ and } \Sigma_h + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^* > 0) \Leftrightarrow \mathcal{M} > 0 \Leftrightarrow \mathcal{H} > 0.$$

Proof. Since, in view of (22), it is obvious that $\mathcal{M} > 0 \Leftrightarrow \mathcal{H} > 0$, we only need to show that

$$(\Sigma_g + \mathcal{P} + \sigma \mathcal{A} \mathcal{A}^* > 0 \text{ and } \Sigma_h + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^* > 0) \Leftrightarrow \mathcal{M} > 0.$$

First, we show that $\Sigma_g + \mathcal{P} + \sigma \mathcal{A} \mathcal{A}^* > 0$ and $\Sigma_h + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^* > 0 \Leftrightarrow \mathcal{M} > 0$. Suppose that $\Sigma_g + \mathcal{P} + \sigma \mathcal{A} \mathcal{A}^* > 0$ and $\Sigma_h + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^* > 0$, but there exists a vector $0 \neq d := (d_y, d_z, d_x) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$ such that $\langle d, \mathcal{M}d \rangle = 0$. By using the definition of \mathcal{M} and (22), we have

$$d_x = 0, \quad (\Sigma_h + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^*)d_z = 0, \quad (\Sigma_g + \mathcal{P})d_y = 0 \quad \text{and} \quad \mathcal{E}^*(d_y, d_z, 0) = 0,$$

which, together with the assumption that $\Sigma_g + \mathcal{P} + \sigma \mathcal{A} \mathcal{A}^* > 0$ and $\Sigma_h + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^* > 0$, imply $d = 0$. This contradiction shows that $\mathcal{M} > 0$.

Next, suppose that $\mathcal{M} > 0$. Since $s_\tau > 0$ and for any $d = (0, d_z, 0) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$, $\langle d, \mathcal{M}d \rangle = \langle d_z, (\Sigma_h + \mathcal{T} + (1 + s_\tau) \cdot \sigma \mathcal{B} \mathcal{B}^*)d_z \rangle$, we know that $\Sigma_h + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^* > 0$. Similarly, since for any $d = (d_y, 0, 0) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$, $\langle d, \mathcal{M}d \rangle = \langle d_y, (\Sigma_g + \mathcal{P} + s_\tau \sigma \mathcal{A} \mathcal{A}^*)d_y \rangle$, we know that $\Sigma_g + \mathcal{P} + \sigma \mathcal{A} \mathcal{A}^* > 0$. So the proof is completed. \square

Based on Proposition 3, we are ready to develop the promised key inequality needed for proving the Q-linear rate of convergence for the sPADMM.

Proposition 4. Let $\tau \in (0, (1 + \sqrt{5})/2]$ and $\{(y^k, z^k, x^k)\}$ be an infinite sequence generated by the sPADMM. Then for any $\bar{u} = (\bar{y}, \bar{z}, \bar{x}) \in \bar{\Omega}$ and any $k \geq 1$,

$$\|u^{k+1} - \bar{u}\|_{\mathcal{M}}^2 + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \leq (\|u^k - \bar{u}\|_{\mathcal{M}}^2 + \|z^k - z^{k-1}\|_{\mathcal{T}}^2) - \|u^{k+1} - u^k\|_{\mathcal{H}}^2. \quad (26)$$

Consequently, we have for all $k \geq 1$,

$$\text{dist}_{\mathcal{M}}^2(u^{k+1}, \bar{\Omega}) + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \leq (\text{dist}_{\mathcal{M}}^2(u^k, \bar{\Omega}) + \|z^k - z^{k-1}\|_{\mathcal{T}}^2) - \|u^{k+1} - u^k\|_{\mathcal{H}}^2. \quad (27)$$

Proof. Let $\bar{u} = (\bar{y}, \bar{z}, \bar{x}) \in \bar{\Omega}$ be fixed but arbitrarily chosen. From part (i) of Theorem 1, we have for $k \geq 1$ that

$$\begin{aligned} & (\tau\sigma)^{-1} \|x^{k+1} - \bar{x}\|^2 + \|y^{k+1} - \bar{y}\|_{\mathcal{T}}^2 + \|z^{k+1} - \bar{z}\|_{\mathcal{T}}^2 + \sigma \|\mathcal{B}^*(z^{k+1} - \bar{z})\|^2 + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \\ & + (1 - \min\{\tau, \tau^{-1}\}) \sigma \|\mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c\|^2 \\ & \leq (\tau\sigma)^{-1} \|x^k - \bar{x}\|^2 + \|y^k - \bar{y}\|_{\mathcal{T}}^2 + \|z^k - \bar{z}\|_{\mathcal{T}}^2 + \sigma \|\mathcal{B}^*(z^k - \bar{z})\|^2 + \|z^k - z^{k-1}\|_{\mathcal{T}}^2 \\ & + (1 - \min\{\tau, \tau^{-1}\}) \sigma \|\mathcal{E}^*(y^k, z^k, 0) - c\|^2 \\ & - \{ \sigma[\tau - \tau^2 + \tau \min\{\tau, \tau^{-1}\}] \|\mathcal{B}^*(z^{k+1} - z^k)\|^2 + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 + \|y^{k+1} - y^k\|_{\mathcal{T}}^2 \\ & + 2\|y^{k+1} - \bar{y}\|_{\Sigma_g}^2 + 2\|z^{k+1} - \bar{z}\|_{\Sigma_h}^2 + (1 - \tau + \min\{\tau, \tau^{-1}\}) \sigma \|\mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c\|^2 \}. \end{aligned} \quad (28)$$

By reorganizing the terms in (28), we obtain

$$\begin{aligned} & (\tau\sigma)^{-1} \|x^{k+1} - \bar{x}\|^2 + \|y^{k+1} - \bar{y}\|_{\mathcal{T}}^2 + \|z^{k+1} - \bar{z}\|_{\mathcal{T}}^2 + \sigma \|\mathcal{B}^*(z^{k+1} - \bar{z})\|^2 + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \\ & + \frac{1}{4}(5 - \tau - 3 \min\{\tau, \tau^{-1}\}) \sigma \|\mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c\|^2 + \|y^{k+1} - \bar{y}\|_{\Sigma_g}^2 + \|z^{k+1} - \bar{z}\|_{\Sigma_h}^2 \\ & \leq (\tau\sigma)^{-1} \|x^k - \bar{x}\|^2 + \|y^k - \bar{y}\|_{\mathcal{T}}^2 + \|z^k - \bar{z}\|_{\mathcal{T}}^2 + \sigma \|\mathcal{B}^*(z^k - \bar{z})\|^2 + \|z^k - z^{k-1}\|_{\mathcal{T}}^2 \\ & + \frac{1}{4}(5 - \tau - 3 \min\{\tau, \tau^{-1}\}) \sigma \|\mathcal{E}^*(y^k, z^k, 0) - c\|^2 + \|y^k - \bar{y}\|_{\Sigma_g}^2 + \|z^k - \bar{z}\|_{\Sigma_h}^2 \end{aligned}$$

$$- \{2t_\tau \sigma \tau \|\mathcal{B}^*(z^{k+1} - z^k)\|^2 + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 + \|y^{k+1} - y^k\|_{\mathcal{S}}^2 + \|y^{k+1} - \bar{y}\|_{\Sigma_g}^2 + \|y^k - \bar{y}\|_{\Sigma_g}^2 + \|z^{k+1} - \bar{z}\|_{\Sigma_h}^2 + \|z^k - \bar{z}\|_{\Sigma_h}^2 + \frac{1}{2}(1 - \tau + \min\{\tau, \tau^{-1}\})\sigma \|\mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c\|^2 + \frac{1}{4}(1 - \tau + \min\{\tau, \tau^{-1}\})\sigma [\|\mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c\|^2 + \|\mathcal{E}^*(y^k, z^k, 0) - c\|^2]\};$$

or equivalently

$$\begin{aligned} & (\tau\sigma)^{-1}\|x^{k+1} - \bar{x}\|^2 + \|y^{k+1} - \bar{y}\|_{\mathcal{S}}^2 + \|z^{k+1} - \bar{z}\|_{\mathcal{T}}^2 + \sigma \|\mathcal{B}^*(z^{k+1} - \bar{z})\|^2 + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \\ & + s_\tau \sigma \|\mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c\|^2 + \|y^{k+1} - \bar{y}\|_{\Sigma_g}^2 + \|z^{k+1} - \bar{z}\|_{\Sigma_h}^2 \\ & \leq (\tau\sigma)^{-1}\|x^k - \bar{x}\|^2 + \|y^k - \bar{y}\|_{\mathcal{S}}^2 + \|z^k - \bar{z}\|_{\mathcal{T}}^2 + \sigma \|\mathcal{B}^*(z^k - \bar{z})\|^2 + \|z^k - z^{k-1}\|_{\mathcal{T}}^2 \\ & + s_\tau \sigma \|\mathcal{E}^*(y^k, z^k, 0) - c\|^2 + \|y^k - \bar{y}\|_{\Sigma_g}^2 + \|z^k - \bar{z}\|_{\Sigma_h}^2 \\ & - \{2t_\tau \sigma \tau \|\mathcal{B}^*(z^{k+1} - z^k)\|^2 + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 + \|y^{k+1} - y^k\|_{\mathcal{S}}^2 + \|y^{k+1} - \bar{y}\|_{\Sigma_g}^2 + \|y^k - \bar{y}\|_{\Sigma_g}^2 \\ & + \|z^{k+1} - \bar{z}\|_{\Sigma_h}^2 + \|z^k - \bar{z}\|_{\Sigma_h}^2 + t_\tau \sigma \|\mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c\|^2 \\ & + \frac{1}{2}t_\tau \sigma [\|\mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c\|^2 + \|\mathcal{E}^*(y^k, z^k, 0) - c\|^2]\}. \end{aligned} \quad (29)$$

Using equalities

$$\begin{aligned} \mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c &= \mathcal{A}^*(y^{k+1} - \bar{y}) + \mathcal{B}^*(z^{k+1} - \bar{z}), \\ \mathcal{E}^*(y^k, z^k, 0) - c &= \mathcal{A}^*(y^k - \bar{y}) + \mathcal{B}^*(z^k - \bar{z}), \\ \mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c &= (\tau\sigma)^{-1}(x^{k+1} - x^k) \end{aligned}$$

and inequalities

$$\begin{aligned} \|y^{k+1} - \bar{y}\|_{\Sigma_g}^2 + \|y^k - \bar{y}\|_{\Sigma_g}^2 &\geq \frac{1}{2}\|y^{k+1} - y^k\|_{\Sigma_g}^2, \quad \|z^{k+1} - \bar{z}\|_{\Sigma_h}^2 + \|z^k - \bar{z}\|_{\Sigma_h}^2 \geq \frac{1}{2}\|z^{k+1} - z^k\|_{\Sigma_h}^2, \\ \|\mathcal{E}^*(y^{k+1}, z^{k+1}, 0) - c\|^2 + \|\mathcal{E}^*(y^k, z^k, 0) - c\|^2 &\geq \frac{1}{2}\|\mathcal{A}^*(y^{k+1} - y^k) + \mathcal{B}^*(z^{k+1} - z^k)\|^2, \end{aligned}$$

we obtain from (29) and the definitions of s_τ and t_τ that for any $\tau \in (0, (1 + \sqrt{5})/2]$,

$$\begin{aligned} & (\tau\sigma)^{-1}\|x^{k+1} - \bar{x}\|^2 + \|y^{k+1} - \bar{y}\|_{\mathcal{S}}^2 + \|z^{k+1} - \bar{z}\|_{\mathcal{T}}^2 + \sigma \|\mathcal{B}^*(z^{k+1} - \bar{z})\|^2 \\ & + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 + s_\tau \sigma \|\mathcal{A}^*(y^{k+1} - \bar{y}) + \mathcal{B}^*(z^{k+1} - \bar{z})\|^2 + \|y^{k+1} - \bar{y}\|_{\Sigma_g}^2 + \|z^{k+1} - \bar{z}\|_{\Sigma_h}^2 \\ & \leq (\tau\sigma)^{-1}\|x^k - \bar{x}\|^2 + \|y^k - \bar{y}\|_{\mathcal{S}}^2 + \|z^k - \bar{z}\|_{\mathcal{T}}^2 + \sigma \|\mathcal{B}^*(z^k - \bar{z})\|^2 \\ & + \|z^k - z^{k-1}\|_{\mathcal{T}}^2 + s_\tau \sigma \|\mathcal{A}^*(y^k - \bar{y}) + \mathcal{B}^*(z^k - \bar{z})\|^2 + \|y^k - \bar{y}\|_{\Sigma_g}^2 + \|z^k - \bar{z}\|_{\Sigma_h}^2 \\ & - \{2t_\tau \sigma \tau \|\mathcal{B}^*(z^{k+1} - z^k)\|^2 + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 + \|y^{k+1} - y^k\|_{\mathcal{S}}^2 + \frac{1}{2}\|y^{k+1} - y^k\|_{\Sigma_g}^2 \\ & + \frac{1}{2}\|z^{k+1} - z^k\|_{\Sigma_h}^2 + t_\tau (\tau^2 \sigma)^{-1}\|x^{k+1} - x^k\|^2 + \frac{1}{4}t_\tau \sigma \|\mathcal{A}^*(y^{k+1} - y^k) + \mathcal{B}^*(z^{k+1} - z^k)\|^2\}, \end{aligned}$$

which shows that (26) holds. By noting that $\bar{\Omega}$ is a nonempty closed convex set and (26) holds for any $\bar{u} \in \bar{\Omega}$, we immediately get (27). \square

Now, we can establish the Q-linear rate of convergence of the sPADMM under a calmness condition on R^{-1} at the origin for some KKT point.

Theorem 2. Let $\tau \in (0, (1 + \sqrt{5})/2)$. Let \mathcal{S} and \mathcal{T} be chosen such that $\Sigma_g + \mathcal{S} + \sigma \mathcal{A} \mathcal{A}^* > 0$ and $\Sigma_h + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^* > 0$. Then there exists a KKT point $\bar{u} := (\bar{y}, \bar{z}, \bar{x}) \in \bar{\Omega}$ such that the whole sequence $\{(y^k, z^k, x^k)\}$ generated by the sPADMM converges to \bar{u} . Assume that R^{-1} is calm at the origin for \bar{u} with modulus $\eta > 0$, i.e., there exists $r > 0$ such that

$$\text{dist}(u, \bar{\Omega}) \leq \eta \|R(u)\|, \quad \forall u \in \{u \in \mathcal{U} : \|u - \bar{u}\| \leq r\}. \quad (30)$$

Then there exists an integer $\bar{k} \geq 1$ such that for all $k \geq \bar{k}$,

$$\text{dist}_{\mathcal{M}}^2(u^{k+1}, \bar{\Omega}) + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \leq \mu [\text{dist}_{\mathcal{M}}^2(u^k, \bar{\Omega}) + \|z^k - z^{k-1}\|_{\mathcal{T}}^2], \quad (31)$$

where

$$\mu := (1 + 2\kappa)^{-1}(1 + \kappa) < 1 \quad \text{and} \quad \kappa := \min\{2\tau, 1\}t_\tau(\eta^2\kappa_4\lambda_{\max}(\mathcal{M}))^{-1} > 0.$$

Moreover, there exists a positive number $\varsigma \in [\mu, 1)$ such that for all $k \geq 1$,

$$\text{dist}_{\mathcal{M}}^2(u^{k+1}, \bar{\Omega}) + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \leq \varsigma [\text{dist}_{\mathcal{M}}^2(u^k, \bar{\Omega}) + \|z^k - z^{k-1}\|_{\mathcal{T}}^2]. \quad (32)$$

Proof. From part (ii) of Theorem 1 we already know that the whole sequence $\{(y^k, z^k, x^k)\}$ generated by the sPADMM converges to a KKT point in $\bar{\Omega}$, say $\bar{u} = (\bar{y}, \bar{z}, \bar{x})$. Then there exists $\bar{k} \geq 1$ such that for all $k \geq \bar{k}$,

$$\|u^{k+1} - \bar{u}\| \leq r.$$

Thus, by using Lemma 1 and (30), we know that for all $k \geq \bar{k}$,

$$\text{dist}^2(u^{k+1}, \bar{\Omega}) \leq \eta^2 \|R(u^{k+1})\|^2 \leq \eta^2 \|u^{k+1} - u^k\|_{\mathcal{H}_0}^2. \quad (33)$$

From the definition of \mathcal{H} , we have for all $k \geq 0$,

$$\|z^{k+1} - z^k\|_{\mathcal{J}}^2 \leq \|u^{k+1} - u^k\|_{\mathcal{H}}^2.$$

It follows from (25) and (33) that for all $k \geq \bar{k}$,

$$\|u^{k+1} - u^k\|_{\mathcal{H}}^2 \geq \min\{2\tau, 1\} t_{\tau} \kappa_4^{-1} \|u^{k+1} - u^k\|_{\mathcal{H}_0}^2 \geq \min\{2\tau, 1\} t_{\tau} \kappa_4^{-1} \eta^{-2} \text{dist}^2(u^{k+1}, \bar{\Omega}) \geq \kappa \text{dist}_{\mathcal{M}}^2(u^{k+1}, \bar{\Omega}). \quad (34)$$

Let $\kappa_5 = (1 + \kappa)^{-1}$. From (27) in Proposition 4 and (34), we have for all $k \geq \bar{k}$,

$$\begin{aligned} \text{dist}_{\mathcal{M}}^2(u^{k+1}, \bar{\Omega}) + \|z^{k+1} - z^k\|_{\mathcal{J}}^2 - \{\text{dist}_{\mathcal{M}}^2(u^k, \bar{\Omega}) + \|z^k - z^{k-1}\|_{\mathcal{J}}^2\} &\leq -((1 - \kappa_5) \|u^{k+1} - u^k\|_{\mathcal{H}}^2 + \kappa_5 \|u^{k+1} - u^k\|_{\mathcal{H}}^2) \\ &\leq -((1 - \kappa_5) \|z^{k+1} - z^k\|_{\mathcal{J}}^2 + \kappa_5 \kappa \text{dist}_{\mathcal{M}}^2(u^{k+1}, \bar{\Omega})). \end{aligned} \quad (35)$$

Then we obtain from (35) that for all $k \geq \bar{k}$,

$$(1 + \kappa_5 \kappa) \text{dist}_{\mathcal{M}}^2(u^{k+1}, \bar{\Omega}) + (2 - \kappa_5) \|z^{k+1} - z^k\|_{\mathcal{J}}^2 \leq \text{dist}_{\mathcal{M}}^2(u^k, \bar{\Omega}) + \|z^k - z^{k-1}\|_{\mathcal{J}}^2.$$

By noting that $1 + \kappa_5 \kappa = 2 - \kappa_5 = \mu^{-1}$, we obtain the estimate (31).

By combining (31) with Lemma 1, (27) in Proposition 4 and (25), we can obtain directly that there exists a positive number $\varsigma \in [\mu, 1)$ such that (32) holds for all $k \geq 1$. The proof is completed. \square

Theorem 2 provides a very general result on the Q-linear rate of convergence for the sPADMM. As one can see that the key assumption made in this theorem is the calmness condition (30), which may not hold in general (see the next section for more detailed discussions on this). However, if R^{-1} is piecewise polyhedral, this calmness condition holds automatically. Since R^{-1} is piecewise polyhedral if and only if R itself is piecewise polyhedral, we can obtain the following from Proposition 1 and the proof of Theorem 2.

Corollary 1. Let $\tau \in (0, (1 + \sqrt{5})/2)$. Suppose that $\bar{\Omega} \neq \emptyset$ and that both $\Sigma_g + \mathcal{S} + \sigma \mathcal{A} \mathcal{A}^*$ and $\Sigma_h + \mathcal{T} + \sigma \mathcal{B} \mathcal{B}^*$ are positive definite. Assume that the mapping $R: \mathcal{U} \rightarrow \mathcal{U}$ is piecewise polyhedral. Then there exist a constant $\hat{\eta} > 0$ such that the infinite sequence $\{(y^k, z^k, x^k)\}$ generated from the sPADMM satisfies for all $k \geq 1$,

$$\text{dist}(u^k, \bar{\Omega}) \leq \hat{\eta} \|R(u^k)\|, \quad (36)$$

$$\text{dist}_{\mathcal{M}}^2(u^{k+1}, \bar{\Omega}) + \|z^{k+1} - z^k\|_{\mathcal{J}}^2 \leq \hat{\mu} [\text{dist}_{\mathcal{M}}^2(u^k, \bar{\Omega}) + \|z^k - z^{k-1}\|_{\mathcal{J}}^2], \quad (37)$$

where

$$\hat{\mu} := (1 + 2\hat{\kappa})^{-1} (1 + \hat{\kappa}) < 1 \quad \text{and} \quad \hat{\kappa} := \min\{2\tau, 1\} t_{\tau} (\hat{\eta}^2 \kappa_4 \lambda_{\max}(\mathcal{M}))^{-1} > 0.$$

Proof. Since $\bar{\Omega} \neq \emptyset$ and $R^{-1}: \mathcal{U} \rightarrow \mathcal{U}$ is piecewise polyhedral, it follows from Proposition 1 that there exist two constants $\eta > 0$ and $\rho > 0$ such that

$$\text{dist}(u, \bar{\Omega}) \leq \eta \|R(u)\|, \quad \forall u \in \{u \in \mathcal{U}: \|R(u)\| \leq \rho\}.$$

Moreover, from the proof of Theorem 2, we know that there exists a constant $r > 0$ such that the sequence $\{(y^k, z^k, x^k)\}$ generated by the sPADMM converges to a KKT point $\bar{u} \in \bar{\Omega}$ with $\|u^k - \bar{u}\| \leq r$ for all $k \geq 0$. Since for those u^k such that $\|R(u^k)\| > \rho$, it holds that

$$\text{dist}(u^k, \bar{\Omega}) \leq \|u^k - \bar{u}\| \leq r < r(\rho^{-1} \|R(u^k)\|), \quad (38)$$

we know that (36) holds with $\hat{\eta} := \max\{\eta, r/\rho\}$. The inequality (37) can then be proved similarly as for (31) in Theorem 2. \square

Before we move to the next section, let us compare the results in the above corollary with those obtained in the most recent paper by Yang and Han [44], where the authors considered the following two cases with $R(\cdot)$ being a piecewise polyhedral mapping:

- (1) The classic ADMM with $\mathcal{S} = 0$, $\mathcal{T} = 0$, and $\tau = 1$. Both \mathcal{A} and \mathcal{B} are assumed to be surjective.
- (2) The linearized ADMM with $\tau = 1$ and two positive definite linear operators: $\mathcal{S} = \gamma_1 \mathcal{I} - \sigma \mathcal{A} \mathcal{A}^*$ and $\mathcal{T} = \gamma_2 \mathcal{I} - \sigma \mathcal{B} \mathcal{B}^*$, where $\gamma_1 > \sigma \lambda_{\max}(\mathcal{A} \mathcal{A}^*)$ and $\gamma_2 > \sigma \lambda_{\max}(\mathcal{B} \mathcal{B}^*)$. Again \mathcal{A} and \mathcal{B} are assumed to be surjective.

For Case (1), Yang and Han [44] proved the global Q-linear rate of convergence of the sequence $\{(z^k, x^k)\}$ while we proved in Corollary 1 the global Q-linear rate of convergence of the sequence $\{(y^k, z^k, x^k)\}$ for any $\tau \in (0, (1 + \sqrt{5})/2)$. Interestingly, the global Q-linear rate of convergence result in Corollary 1 is still valid even if the surjectivity of \mathcal{A} or \mathcal{B} fails to hold as the availability of Σ_g and Σ_h can be exploited (cf. problem (13)).

For Case (2), Yang and Han [44] proved the global Q-linear convergence rate of the whole sequence $\{(y^k, z^k, x^k)\}$. We also proved the same thing but with one major difference: unlike Yang and Han [44] we neither need to assume the surjectivity of \mathcal{A} or \mathcal{B} nor we need to assume \mathcal{S} or \mathcal{T} to be positive definite. In fact, the analysis in Yang and Han [44] breaks down when $\gamma_1 \rightarrow \sigma \lambda_{\max}(\mathcal{A} \mathcal{A}^*)$ or $\gamma_2 \rightarrow \sigma \lambda_{\max}(\mathcal{B} \mathcal{B}^*)$ even if both \mathcal{A} and \mathcal{B} are surjective. On the other hand, it is easy to see that our results in Corollary 1 are still valid with $\mathcal{S} = \sigma \lambda_{\max}(\mathcal{A} \mathcal{A}^*) \mathcal{I} - \sigma \mathcal{A} \mathcal{A}^*$ and $\mathcal{T} = \sigma \lambda_{\max}(\mathcal{B} \mathcal{B}^*) \mathcal{I} - \sigma \mathcal{B} \mathcal{B}^*$. Here, the main reason that we can obtain the Q-linear convergence results as in Corollary 1 is due to the availability of the key inequality (26) proven in Proposition 4 via the construction of the two linear operators \mathcal{M} and \mathcal{H} in (23) and (24), respectively. More importantly, the freedom of choices of the positive semidefinite linear operators \mathcal{S} and \mathcal{T} in our model allows us to efficiently deal with even multi-block convex composite quadratic conic optimization problems as shall be demonstrated in the next section.

4. Applications to Convex Composite Quadratic Conic Programming

In this section we shall demonstrate how the Q-linear rate convergence results proven in the last section can be applied to the following convex composite quadratic conic programming:

$$\begin{aligned} \min \quad & \frac{1}{2} \langle x, \mathcal{Q}x \rangle + \langle c, x \rangle + \phi(x) \\ \text{s.t.} \quad & \mathcal{A}x = b, \quad x \in \mathcal{K}, \end{aligned} \quad (39)$$

where $c \in \mathcal{X}$, $b \in \mathbb{R}^m$, $\mathcal{Q}: \mathcal{X} \rightarrow \mathcal{X}$ is a self-adjoint positive semidefinite linear operator, $\mathcal{A}: \mathcal{X} \rightarrow \mathbb{R}^m$ is a linear operator, \mathcal{K} is a closed convex cone in \mathcal{X} , and $\phi: \mathcal{X} \rightarrow (-\infty, +\infty]$ is a proper closed convex function. Here, we assume that $\phi^*(\cdot)$ can be computed relatively easily. If \mathcal{K} is polyhedral and ϕ is piecewise linear-quadratic, problem (39) is called the convex composite piecewise linear-quadratic programming. Note that for the latter, the first quadratic term in the objective function of problem (39) could be absorbed in the piecewise linear-quadratic function ϕ . However, this should be avoided as it is more efficient to deal with this quadratic term separately.

By introducing an additional variable $d \in \mathcal{X}$, we can rewrite problem (39) equivalently as

$$\begin{aligned} \min \quad & \frac{1}{2} \langle x, \mathcal{Q}x \rangle + \langle c, x \rangle + \delta_{\mathcal{K}}(x) + \phi(d) \\ \text{s.t.} \quad & \mathcal{A}x = b, \quad x - d = 0. \end{aligned} \quad (40)$$

Obviously, problem (40) is in the form of (1). Let the polar of \mathcal{K} be defined by $\mathcal{K}^\circ := \{x' \in \mathcal{X} : \langle x', x \rangle \leq 0, \forall x \in \mathcal{K}\}$. Denote the dual cone of \mathcal{K} by $\mathcal{K}^* := -\mathcal{K}^\circ$. The Lagrange dual of problem (40) takes the form of

$$\begin{aligned} \max \quad & \inf_{x \in \mathcal{X}} \left\{ \frac{1}{2} \langle x, \mathcal{Q}x \rangle + \langle v, x \rangle + \langle b, y \rangle - \phi^*(-z) \right\} \\ \text{s.t.} \quad & s + \mathcal{A}^*y + v + z = c, \quad s \in \mathcal{K}^*, \end{aligned}$$

which is equivalent to

$$\begin{aligned} \min \quad & \delta_{\mathcal{K}^*}(s) - \langle b, y \rangle + \frac{1}{2} \langle w, \mathcal{Q}w \rangle + \phi^*(-z) \\ \text{s.t.} \quad & s + \mathcal{A}^*y - \mathcal{Q}w + z = c, \quad w \in \mathcal{W}, \end{aligned} \quad (41)$$

where \mathcal{W} is any linear subspace in \mathcal{X} containing $\text{Range } \mathcal{Q}$, the range space of \mathcal{Q} , e.g., $\mathcal{W} = \mathcal{X}$ or $\mathcal{W} = \text{Range } \mathcal{Q}$. When $\mathcal{W} = \mathcal{X}$, problem (41) is better known as the Wolfe dual to problem (40) (see Fujiwara et al. [18] for discussions on the Wolfe dual of conventional nonlinear programming and Qi [33] on nonlinear semidefinite programming). So when $\text{Range } \mathcal{Q} \subseteq \mathcal{W} \neq \mathcal{X}$, one may call problem (41) the restricted Wolfe dual to problem (40). One particularly useful case is the restricted Wolfe dual with $\mathcal{W} = \text{Range } \mathcal{Q}$. The dual problem (41) has four

natural variable blocks and can be written in the form of (1) in several different ways. The cases that we are interested in applying the sPADMM to problem (41) are the following:

Case 1. If $\mathcal{H} \neq \mathcal{X}$, then (s, y, w) is treated as one variable block and z the other block.

Case 2. If $\mathcal{H} = \mathcal{X}$, then (w, y) is treated as one variable block and z the other block.

Here we shall only discuss Case (1) as Case (2) can be done similarly in a simpler manner.

4.1. The Primal Case

First, we consider the application of the sPADMM to the primal problem (40). Let $\mathcal{U} := \mathcal{X} \times \mathcal{X} \times \mathbb{R}^m \times \mathcal{X}$. The augmented Lagrangian function \mathcal{L}_σ^P for problem (40) is defined as follows:

$$\begin{aligned} \mathcal{L}_\sigma^P(x, d; y, z) := & \frac{1}{2} \langle x, \mathcal{Q}x \rangle + \langle c, x \rangle + \delta_{\mathcal{H}}(x) + \phi(d) + \langle y, \mathcal{A}x - b \rangle + \langle z, x - d \rangle \\ & + \frac{\sigma}{2} (\|\mathcal{A}x - b\|^2 + \|x - d\|^2), \quad \forall (x, d, y, z) \in \mathcal{U}. \end{aligned}$$

Then the sPADMM for solving problem (40) can be stated in the following way.

sPADMM: A semi-proximal alternating direction method of multipliers for solving the convex optimization problem (40).

Step 0. Input $(x^0, d^0, y^0, z^0) \in \mathcal{H} \times \text{dom } \phi \times \mathbb{R}^m \times \mathcal{X}$. Let $\tau \in (0, +\infty)$ be a positive parameter (e.g., $\tau \in (0, (1 + \sqrt{5})/2)$). Define $\mathcal{S}: \mathcal{X} \rightarrow \mathcal{X}$ to be any self-adjoint positive semidefinite linear operator such that $\mathcal{Q} + \mathcal{S} + \sigma \mathcal{A}^* \mathcal{A} > 0$. Set $k := 0$.

Step 1. Set

$$\begin{cases} x^{k+1} = \arg \min \mathcal{L}_\sigma^P(x, d^k; y^k, z^k) + \frac{1}{2} \|x - x^k\|_{\mathcal{S}}^2, \\ d^{k+1} = \arg \min \mathcal{L}_\sigma^P(x^{k+1}, d; y^k, z^k), \\ y^{k+1} = y^k + \tau \sigma (\mathcal{A}x^{k+1} - b) \text{ and } z^{k+1} = z^k + \tau \sigma (x^{k+1} - d^{k+1}). \end{cases}$$

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

In the above sPADMM for solving the convex optimization problem (40), we need to choose $\mathcal{S} \geq 0$ satisfying $\mathcal{Q} + \mathcal{S} + \sigma \mathcal{A}^* \mathcal{A} > 0$ such that the subproblems on the x -part are relatively easy to solve, e.g., one can take $\mathcal{S} = \lambda_{\max}(\mathcal{Q} + \sigma \mathcal{A}^* \mathcal{A}) \mathcal{I} - (\mathcal{Q} + \sigma \mathcal{A}^* \mathcal{A})$. Note that if one takes $\mathcal{S} = \gamma_1 \mathcal{I} - \sigma \mathcal{A}^* \mathcal{A}$ with $\gamma_1 > \sigma \lambda_{\max}(\mathcal{A}^* \mathcal{A})$, as discussed in Yang and Han [44], the subproblems on the x -part may still be difficult to solve unless \mathcal{Q} is simple, e.g., $\mathcal{Q} = 0$ or \mathcal{I} .

To apply Theorem 2 and Corollary 1 to prove the Q-linear convergence rate of the sPADMM for solving problem (40), we need to know under what conditions the required calmness assumption for problem (40) holds. Next, we shall discuss this issue in two situations: (1) \mathcal{H} is polyhedral and $\phi(\cdot)$ is piecewise linear quadratic; and (2) \mathcal{H} is the nonpolyhedral cone \mathbb{S}_+^n , which is the cone of all n by n symmetric and positive semidefinite matrices.

The KKT optimality conditions for problem (40) take the form of

$$0 \in \mathcal{Q}x + c + \partial \delta_{\mathcal{H}}(x) + \mathcal{A}^* y + z, \quad 0 \in \partial \phi(d) - z, \quad \mathcal{A}x - b = 0, \quad x - d = 0. \quad (42)$$

Define the KKT mapping $R: \mathcal{U} \rightarrow \mathcal{U}$ by

$$R(x, d, y, z) := \begin{pmatrix} x - \Pi_{\mathcal{H}}[x - (\mathcal{Q}x + c + \mathcal{A}^* y + z)] \\ d - \text{Pr}_{\phi}(d + z) \\ b - \mathcal{A}x \\ d - x \end{pmatrix}, \quad \forall (x, d, y, z) \in \mathcal{U}. \quad (43)$$

Then $(x, d, y, z) \in \mathcal{U}$ satisfies (42) if and only if $R(x, d, y, z) = 0$.

If \mathcal{H} is polyhedral and $\phi(\cdot)$ is piecewise linear quadratic, then things are much easier as in this case Proposition 2 implies that both $\Pi_{\mathcal{H}}(\cdot)$ and $\text{Pr}_{\phi}(\cdot)$ are piecewise polyhedral, and so are R and R^{-1} . Thus, from the discussions in Section 2 we know that in this case, R^{-1} is calm at the origin for any KKT point, if exists, to problem (40) with a modulus independent of the choice of the KKT points. Therefore, for any $\tau \in (0, (1 + \sqrt{5})/2)$, as long as \mathcal{H} is a polyhedral set, $\phi(\cdot)$ is a piecewise linear-quadratic convex function and problem (40) has at least one KKT point, the infinite sequence $\{(x^k, d^k, y^k, z^k)\}$ generated by the sPADMM converges to a KKT point of problem (40) globally at a Q-linear rate.

When \mathcal{K} is nonpolyhedral, things are more complicated regardless of the properties of ϕ . This can be seen from the following convex quadratic semidefinite programming (SDP) example constructed by Bonnans and Shapiro [2, Example 4.54].

Example 1. Consider

$$\begin{aligned} \min \quad & x_1 + x_1^2 + x_2^2 \\ \text{s.t.} \quad & X - \text{Diag}(x) = \varepsilon B, \quad X \in \mathbb{S}_+^2, \end{aligned} \quad (44)$$

where $x = (x_1, x_2) \in \mathbb{R}^2$, $\text{Diag}(x)$ is the 2×2 diagonal matrix whose i th diagonal element is x_i , $i = 1, 2$, B is any 2 by 2 nondiagonal symmetric matrix, and ε is a scalar parameter. When $\varepsilon = 0$, the optimization problem (44) has a unique KKT point $(\bar{X}, \bar{x}, \bar{Y}) \in \mathbb{S}_+^2 \times \mathbb{R}^2 \times (-\mathbb{S}_+^2)$ with $(\bar{X}, \bar{x}) = (0, 0)$ and $\bar{Y} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$.

Bonnans and Shapiro [2, Example 4.54] showed that for any given $\varepsilon \geq 0$, problem (44) has a KKT point $(\bar{X}(\varepsilon), \bar{x}(\varepsilon), \bar{Y}(\varepsilon)) \in \mathbb{S}_+^2 \times \mathbb{R}^2 \times (-\mathbb{S}_+^2)$ with $\bar{x}_2(\varepsilon)$ of the order $\varepsilon^{2/3}$ as $\varepsilon \downarrow 0$. Then, in view of Cui et al. [7, Proposition 2.4], we know that R^{-1} cannot be calm at the origin for $(\bar{X}, \bar{x}, \bar{Y})$ even if the unperturbed problem has a strongly convex objective function with a unique KKT point.

Example 1 shows that for a nonpolyhedral set \mathcal{K} , unlike the polyhedral case, we need additional conditions for guaranteeing the calmness property for problem (40). At the moment, not many results are available when \mathcal{K} is a general nonpolyhedral cone. However, most recently several interesting results on the calmness property have been obtained for the following convex composite quadratic semidefinite programming

$$\begin{aligned} \min \quad & \{\tfrac{1}{2}\langle X, \mathcal{Q}X \rangle + \langle C, X \rangle + \delta_{\mathcal{P}}(X)\} \\ \text{s.t.} \quad & \mathcal{A}X = b, \quad X \in \mathbb{S}_+^n, \end{aligned} \quad (45)$$

where $C \in \mathbb{S}^n$, $b \in \mathbb{R}^m$, $\mathcal{Q}: \mathbb{S}^n \rightarrow \mathbb{S}^n$ is a self-adjoint positive semidefinite linear operator, $\mathcal{A}: \mathbb{S}^n \rightarrow \mathbb{R}^m$ is a linear operator, \mathcal{P} is a simple nonempty convex polyhedral set in \mathbb{S}^n , and $\langle \cdot, \cdot \rangle$ is the usual trace inner product.

Firstly, in Han et al. [25], the authors proved that if problem (45) has a unique KKT point, then the mapping R^{-1} is calm at the origin for this KKT point if and only if the no-gap second order sufficient conditions in terms of Bonnans and Shapiro [2] hold for both the primal and its restricted Wolfe-dual problems. Thus, the reason for the lack of the calmness property of R^{-1} for Example 1 is due to the fact that the no-gap second order sufficient condition for the dual of the unperturbed problem fails to hold. The above characterization has led Ding et al. [10] to study the calmness property at an isolated KKT point for a class of nonconvex conic programming problems with \mathcal{K} being a C^2 -cone reducible set, which is rich enough to include the polyhedral set, the second order cone, the positive semidefinite cone \mathbb{S}_+^n , and their Cartesian products (Bonnans and Shapiro [2]).

Secondly, sufficient conditions for ensuring the metric subregularity of R or equivalently the calmness of R^{-1} have been provided by Cui et al. [7] even if problem (45) may admit multiple KKT points. Here, instead of presenting these sufficient conditions in Cui et al. [7], we shall quote an example used in Cui et al. [7] to illustrate the calmness property of R^{-1} .

Example 2. Consider the following convex quadratic SDP problem:

$$\begin{aligned} \min \quad & \{\tfrac{1}{2}(\langle I_2, X \rangle - 1)^2\} \\ \text{s.t.} \quad & \langle A, X \rangle + x = 1, \quad X \in \mathbb{S}_+^2, x \in \mathbb{R}_+, \end{aligned} \quad (46)$$

whose dual (in its equivalent minimization form) can be written as

$$\begin{aligned} \min \quad & \{-y + \tfrac{1}{2}w^2 + w + \delta_{\mathbb{S}_+^2}(S) + \delta_{\mathbb{R}_+}(s)\} \\ \text{s.t.} \quad & yA + wI_2 + S = 0, \quad s - y = 0, \end{aligned} \quad (47)$$

where $A = \begin{pmatrix} -2 & 1 \\ 1 & 1 \end{pmatrix}$. Problem (47) has a unique optimal solution $(\bar{y}, \bar{w}, \bar{S}, \bar{s}) = (0, 0, 0, 0)$. The set of all optimal solutions to problem (46) is given by

$$\{(X, x) \in \mathbb{S}_+^2 \times \mathbb{R}_+ \mid \langle A, X \rangle + x = 1, \langle I_2, X \rangle = 1\}. \quad (48)$$

One can easily check that for Example 2 the sufficient conditions made in Cui et al. [7] for ensuring the metric subregularity of R hold. Thus, for this example R^{-1} is calm at the origin for any KKT point.

4.2. The Dual Case

In this subsection we turn to the dual problem (41). As mentioned earlier, problem (41) has four natural variable-blocks. Since the directly extended ADMM to the multi-block case may be divergent even the dual step-length τ is taken to be as small as 10^{-8} (Chen et al. [4]), one needs new ideas to deal with problem (41). Here, we will adopt the symmetric Gauss-Seidel (sGS) technique invented by Li et al. [30]. For details on the sGS technique, see Li [29]. Most recent research has shown that it is much more efficient to solve the dual problem (41) rather than its primal counterpart (40) in the context of semidefinite programming and convex quadratic semidefinite programming (Chen et al., Li et al., Sun et al. [41, 30, 29, 5]). At the first glance, this seems to be counterintuitive, as problem (41) looks much more complicated than the primal problem (40). The key point for more efficiently dealing with the dual problem is to intelligently combine the above mentioned sGS technique with the sPADMM.

The augmented Lagrangian function \mathcal{L}_σ^D for problem (41) is defined as follows:

$$\begin{aligned} \mathcal{L}_\sigma^D(s, y, w, z; x) := & \delta_{\mathcal{H}^*}(s) - \langle b, y \rangle + \frac{1}{2} \langle w, \mathcal{Q}w \rangle + \phi^*(-z) + \langle x, s + \mathcal{A}^*y - \mathcal{Q}w + z - c \rangle \\ & + \frac{\sigma}{2} \|s + \mathcal{A}^*y - \mathcal{Q}w + z - c\|^2, \quad \forall (s, y, w, z, x) \in \mathcal{X} \times \mathcal{R}^m \times \mathcal{W} \times \mathcal{Z} \times \mathcal{X}. \end{aligned}$$

Then the sGS technique based sPADMM, in short sGS-sPADMM, considered by Li et al. [30] for solving the multi-block problem (41) can be stated as in the following. At the first glance, the sGS-sPADMM does not seem to fall within the scheme (3a)–(3c). However, it has been shown in Li et al. [30] that it is indeed a special case of (3a)–(3c) through the construction of special semi-proximal terms.

sGS-sPADMM: A symmetric Gauss-Seidel based semi-proximal alternating direction method of multipliers for solving problem (41).

Step 0. Input $(s^0, y^0, w^0, z^0, x^0) \in \mathcal{H}^* \times \mathcal{R}^m \times \mathcal{W} \times (-\text{dom } \phi^*) \times \mathcal{X}$. Let $\tau \in (0, +\infty)$ be a positive parameter (e.g., $\tau \in (0, (1 + \sqrt{5})/2)$). Choose any two self-adjoint positive semidefinite linear operators $\mathcal{S}_1: \mathcal{R}^m \rightarrow \mathcal{R}^m$ and $\mathcal{S}_2: \mathcal{W} \rightarrow \mathcal{W}$ satisfying $\mathcal{S}_1 + \sigma \mathcal{A} \mathcal{A}^* > 0$ and $\mathcal{S}_2 + \mathcal{Q} + \sigma \mathcal{Q}^2 > 0$. Set $k := 0$.

Step 1. Set

$$\begin{cases} w^{k+1/2} = \arg \min \mathcal{L}_\sigma^D(s^k, y^k, w, z^k; x^k) + \frac{1}{2} \|w - w^k\|_{\mathcal{S}_2}^2, \\ y^{k+1/2} = \arg \min \mathcal{L}_\sigma^D(s^k, y, w^{k+1/2}, z^k; x^k) + \frac{1}{2} \|y - y^k\|_{\mathcal{S}_1}^2, \\ s^{k+1} = \arg \min \mathcal{L}_\sigma^D(s, y^{k+1/2}, w^{k+1/2}, z^k; x^k), \\ y^{k+1} = \arg \min \mathcal{L}_\sigma^D(s^{k+1}, y, w^{k+1/2}, z^k; x^k) + \frac{1}{2} \|y - y^k\|_{\mathcal{S}_1}^2, \\ w^{k+1} = \arg \min \mathcal{L}_\sigma^D(s^{k+1}, y^{k+1}, w, z^k; x^k) + \frac{1}{2} \|w - w^k\|_{\mathcal{S}_2}^2, \\ z^{k+1} = \arg \min \mathcal{L}_\sigma^D(s^{k+1}, y^{k+1}, w^{k+1}, z; x^k), \\ x^{k+1} = x^k + \tau \sigma (s^{k+1} + \mathcal{A}^* y^{k+1} - \mathcal{Q} w^{k+1} + z^{k+1} - c). \end{cases}$$

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

As mentioned earlier, the global convergence of Algorithm sGS-sPADMM is established by Li et al. [30] through converting it into an equivalent sPADMM scheme (3a)–(3c) for solving a particular problem of the form (1). To illustrate how this is achieved, for simplicity, we assume that $\mathcal{A}: \mathcal{X} \rightarrow \mathcal{R}^m$ is surjective and $\mathcal{W} = \text{Range } \mathcal{Q}$ so that we can take $\mathcal{S}_1 = 0$ and $\mathcal{S}_2 = 0$, i.e., there are no proximal terms in the above Algorithm sGS-sPADMM. Note that the self-dual linear operator \mathcal{Q} is always positive definite from $\text{Range } \mathcal{Q}$ to itself even if it is only positive semidefinite from \mathcal{X} to \mathcal{X} .

Define the self-adjoint positive semidefinite linear operator (to be interpreted as in the matrix format) $\mathcal{S}: \mathcal{X} \times \mathcal{R}^m \times \mathcal{W} \rightarrow \mathcal{X} \times \mathcal{R}^m \times \mathcal{W}$ by

$$\mathcal{S} = \sigma \begin{pmatrix} 0 & 0 & 0 \\ \mathcal{A} & 0 & 0 \\ -\mathcal{Q} & -\mathcal{Q} \mathcal{A}^* & 0 \end{pmatrix} \begin{pmatrix} \mathcal{I} & 0 & 0 \\ 0 & (\mathcal{A} \mathcal{A}^*)^{-1} & 0 \\ 0 & 0 & (\mathcal{Q}^2 + \sigma^{-1} \mathcal{Q})^{-1} \end{pmatrix} \begin{pmatrix} 0 & \mathcal{A}^* & -\mathcal{Q} \\ 0 & 0 & -\mathcal{A} \mathcal{Q} \\ 0 & 0 & 0 \end{pmatrix}. \quad (49)$$

Then, with (s, y, w) being treated as one variable block and z the other block, the above sGS-sPADMM for solving problem (41) reduces to the sPADMM scheme (3a)–(3c), where the proximal terms \mathcal{S} being given by (49) and $\mathcal{T} = 0$ (Li et al. [30]). We remark here that although the linear operator \mathcal{S} looks complicated, one never needs to compute it in the numerical implementations and it is introduced only for connecting Algorithm sGS-sPADMM to the general scheme (3a)–(3c).

One may further note that at each iteration, either the w -part or the y -part needs to be solved twice, which seems to suggest that Algorithm sGS-sPADMM is more costly compared to the nonconvergent directly extended

ADMM. However, the extra cost is minimum as the coefficient matrix for solving either part is identical through all the iterations.

By using the above connection, just as for the primal case, one can use Theorem 2 and Corollary 1 to derive the Q-linear rate convergence of the infinite sequence $\{(s^k, y^k, w^k, z^k, x^k)\}$ generated by Algorithm sGS-sPADMM if Assumption 1 and the required calmness condition hold for problem (41) and $\tau \in (0, (1 + \sqrt{5})/2)$. On the calmness condition, one may conduct similar discussions as in Section 4.1, but start from the dual problem (41). For brevity, we omit the details here. As a final note to this section, we comment that in all the above applications, the linear operator $\mathcal{T} \equiv 0$ while the linear operator \mathcal{S} may take various values, which are often to be positive semidefinite only.

5. Conclusions

In this paper, we have provided a road map for analyzing the Q-linear convergence rate of the sPADMM for solving linearly constrained convex composite optimization problems. One significant feature of our approach is that it only relies on a very mild calmness property. This allows us to obtain a more or less complete picture on the Q-linear rate convergence analysis for solving the convex composite piecewise linear-quadratic programming. More importantly, it also allows us to derive Q-linearly convergent results of the sPADMM for solving convex composite quadratic semidefinite programming. Along this line, perhaps, the most important issue left unanswered is to provide weaker sufficient conditions for ensuring the calmness property for convex composite optimization problems with nonpolyhedral cone constraints. Another important issue is to develop similar results for the inexact version of the sPADMM, which is often more useful in practice. Given the recent progress made on the inexact symmetric Gauss-Seidel based sPADMM in Chen et al. [5], it does not seem to be difficult to extend our analysis to the inexact sPADMM in a parallel way.

Acknowledgments

The authors would like to thank Ying Cui, Xudong Li, Yangjing Zhang, and the two referees for their helpful suggestions on improving the quality of this work.

Endnote

¹For example, according to Google Scholar, the survey paper by Boyd et al. [3] on the applications of the ADMM with $\tau = 1$ has been cited 5,000 times as of May 5, 2017.

References

- [1] Boley D (2013) Local linear convergence of ADMM on quadratic or linear programs. *SIAM J. Optim.* 23(4):2183–2207.
- [2] Bonnans JF, Shapiro A (2000) *Perturbation Analysis of Optimization Problems* (Springer, New York).
- [3] Boyd S, Parikh N, Chu E, Peleato B, Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* 3(1):1–122.
- [4] Chen CH, He BS, Ye YY, Yuan XM (2016) The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Math. Programming* 155(1–2):57–79.
- [5] Chen L, Sun DF, Toh K-C (2017) An efficient inexact symmetric Gauss-Seidel-based majorized ADMM for high-dimensional convex composite conic programming. *Math. Programming* 161(1–2):237–270.
- [6] Chen L, Sun DF, Toh K-C (2017) A note on the convergence of ADMM for linearly constrained convex optimization problems. *Comput. Optim. Appl.* 66:327–343.
- [7] Cui Y, Sun DF, Toh K-C (2016) On the asymptotic superlinear convergence of the augmented Lagrangian method for semidefinite programming with multiple solutions. Preprint arXiv:1610.00875.
- [8] Cui Y, Li XD, Sun DF, Toh K-C (2016) On the convergence properties of a majorized ADMM for linearly constrained convex optimization problems with coupled objective functions. *J. Optim. Theory Appl.* 169(3):1013–1041.
- [9] Deng W, Yin WT (2016) On the global and linear convergence of the generalized alternating direction method of multipliers. *J. Scientific Comput.* 66(3):889–916.
- [10] Ding C, Sun DF, Zhang LW (2017) Characterization of the robust isolated calmness for a class of conic programming problems. *SIAM J. Optim.* 27(1):67–90.
- [11] Dontchev AL, Rockafellar RT (2009) *Implicit Functions and Solution Mappings* (Springer, New York).
- [12] Eckstein J (1989) Splitting Methods for Monotone Operators with Applications to Parallel Optimization. PhD Thesis, MIT, Cambridge, MA.
- [13] Eckstein J (1994) Some saddle-function splitting methods for convex programming. *Optim. Methods and Software* 4(1):75–83.
- [14] Eckstein J, Bertsekas DP (1992) On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming* 55(1–3):293–318.
- [15] Eckstein J, Yao W (2015) Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. *Pacific J. Optim.* 11(4):619–644.
- [16] Fazel M, Pong TK, Sun DF, Tseng P (2013) Hankel matrix rank minimization with applications to system identification and realization. *SIAM J. Matrix Anal. Appl.* 34(3):946–977.

- [17] Fortin M, Glowinski R (1983) On decomposition-coordination methods using an augmented Lagrangian. Fortin M, Glowinski R, eds. *Augmented Lagrangian Methods: Applications to the Solution of Boundary Problems*. Studies in Mathematics And its Applications, Vol. 15 (Elsevier, Amsterdam), 97–146.
- [18] Fujiwara O, Han S-P, Mangasarian OL (1984) Local duality of nonlinear programs. *SIAM J. Control Optim.* 22(1):162–169.
- [19] Gabay D (1983) Applications of the method of multipliers to variational inequalities. Fortin M, Glowinski R, eds. *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*. Studies in Mathematics And its Applications, Vol. 15 (Elsevier, Amsterdam), 299–331.
- [20] Gabay D, Mercier B (1976) A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.* 2:17–40.
- [21] Glowinski R (1980) Lectures on numerical methods for nonlinear variational problems. (Springer, Berlin).
- [22] Glowinski R (2014) On alternating direction methods of multipliers: A historical perspective. Fitzgibbon W, Kuznetsov YA, Neittaanmaki P, Pironneau O, eds. *Modeling, Simulation and Optimization for Science and Technology* (Springer, Netherlands), 59–82.
- [23] Glowinski R, Marroco A (1975) Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *Revue française d'atomique, informatique recherche opérationnelle. Analyse Numérique* 9(R2):41–76.
- [24] Han DR, Yuan XM (2013) Local linear convergence of the alternating direction method of multipliers for quadratic programs. *SIAM J. Numerical Anal.* 51(6):3446–3457.
- [25] Han DR, Sun DF, Zhang LW (2015) On the isolated calmness and strong regularity for convex composite quadratic semidefinite programming, November 2016; revised from the second part of arXiv:1508.02134, August.
- [26] He BS, Liao LZ, Han DR, Yang H (2002) A new inexact alternating directions method for monotone variational inequalities. *Math. Programming* 92(1):103–118.
- [27] Hong M, Luo ZQ (2017) On the linear convergence of alternating direction method of multipliers. *Math. Programming* 162(1–2):165–199.
- [28] Li M, Sun DF, Toh K-C (2016) A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization. *SIAM J. Optim.* 26(2):922–950.
- [29] Li XD (2015) A two-phase augmented Lagrangian method for convex composite quadratic programming. PhD Thesis, Department of Mathematics, National University of Singapore.
- [30] Li XD, Sun DF, Toh K-C (2016) A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions. *Math. Programming* 155(1–2):333–373.
- [31] Lions PL, Mercier B (1979) Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numerical Anal.* 16(6):964–979.
- [32] Luque FJ (1984) Asymptotic convergence analysis of the proximal point algorithm. *SIAM J. Control Optim.* 22(2):277–293.
- [33] Qi HD (2009) Local duality of nonlinear semidefinite programming. *Math. Oper. Res.* 34(1):124–141.
- [34] Robinson SM (1976) An implicit-function theorem for generalized variational inequalities. Technical Summary Report 1672, Mathematics Research Center, University of Wisconsin-Madison; available from National Technical Information Service under Accession ADA031952.
- [35] Robinson SM (1981) Some continuity properties of polyhedral multifunctions. König H, Korte B, Ritter K, eds. *Mathematical Programming at Oberwolfach*. Mathematical Programming Studies, Vol. 14 (Springer, Berlin), 206–214.
- [36] Rockafellar RT (1970) *Convex Analysis* (Princeton University Press, Princeton, NJ).
- [37] Rockafellar RT (1976) Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* 1(2):97–116.
- [38] Rockafellar RT (1976) Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* 14(5):877–898.
- [39] Rockafellar RT, Wets RJ-B (1998) *Variational Analysis* (Springer, Berlin).
- [40] Shefi R, Teboulle M (2014) Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. *SIAM J. Optim.* 24(1):269–297.
- [41] Sun DF, Toh K-C, Yang LQ (2015) A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type constraints. *SIAM J. Optim.* 25(2):882–915.
- [42] Sun J (1986) On monotropic piecewise quadratic programming. PhD Thesis, Department of Mathematics, University of Washington, Seattle.
- [43] Xu MH, Wu T (2011) A class of linearized proximal alternating direction methods. *J. Optim. Theory Appl.* 151(2):321–337.
- [44] Yang WH, Han DR (2016) Linear convergence of alternating direction method of multipliers for a class of convex optimization problems. *SIAM J. Numerical Anal.* 54(2):625–640.