

# A Majorized ADMM with Indefinite Proximal Terms for Linearly Constrained Convex Composite Optimization

Min Li\*, Defeng Sun<sup>†</sup> and Kim-Chuan Toh<sup>‡</sup>

December 5, 2014; Revised: June 22, 2015

## Abstract

This paper presents a majorized alternating direction method of multipliers (ADMM) with indefinite proximal terms for solving linearly constrained 2-block convex composite optimization problems with each block in the objective being the sum of a non-smooth convex function ( $p(x)$  or  $q(y)$ ) and a smooth convex function ( $f(x)$  or  $g(y)$ ), i.e.,  $\min_{x \in \mathcal{X}, y \in \mathcal{Y}} \{p(x) + f(x) + q(y) + g(y) \mid A^*x + B^*y = c\}$ . By choosing the indefinite proximal terms properly, we establish the global convergence, and the iteration-complexity in both the non-ergodic and ergodic senses of the proposed method for the step-length  $\tau \in (0, (1 + \sqrt{5})/2)$ . The computational benefit of using indefinite proximal terms within the ADMM framework instead of the current requirement of positive semidefinite ones is also demonstrated numerically. This opens up a new way to improve the practical performance of the ADMM and related methods.

**Keywords.** Alternating direction method of multipliers, Convex composite optimization, Indefinite proximal terms, Majorization, Iteration-complexity

**AMS subject classifications.** 90C25, 90C33, 65K05

## 1 Introduction

We consider the following 2-block convex composite optimization problem

$$\min_{x \in \mathcal{X}, y \in \mathcal{Y}} \left\{ p(x) + f(x) + q(y) + g(y) \mid A^*x + B^*y = c \right\}, \quad (1)$$

where  $\mathcal{X}$ ,  $\mathcal{Y}$  and  $\mathcal{Z}$  are three real finite dimensional Euclidean spaces each equipped with an inner product  $\langle \cdot, \cdot \rangle$  and its induced norm  $\| \cdot \|$ ;  $p : \mathcal{X} \rightarrow (-\infty, +\infty]$  and  $q : \mathcal{Y} \rightarrow (-\infty, +\infty]$  are two closed

---

\*School of Economics and Management, Southeast University, Nanjing, 210096, China (lmin@seu.edu.cn). The research of this author was supported in part by the National Natural Science Foundation of China (Grant No. 11001053, 71390333), Program for New Century Excellent Talents in University (Grant No. NCET-12-0111) and Qing Lan Project.

<sup>†</sup>Department of Mathematics and Risk Management Institute, National University of Singapore, 10 Lower Kent Ridge Road, Singapore (matsundf@nus.edu.sg). The research of this author was supported in part by the Academic Research Fund (Grant No. R-146-000-207-112).

<sup>‡</sup>Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore (mattohkc@nus.edu.sg). The research of this author was supported in part by the Ministry of Education, Singapore, Academic Research Fund (Grant No. R-146-000-194-112).

proper convex (not necessarily smooth) functions;  $f : \mathcal{X} \rightarrow (-\infty, +\infty)$  and  $g : \mathcal{Y} \rightarrow (-\infty, +\infty)$  are two convex functions with Lipschitz continuous gradients on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively;  $A^* : \mathcal{X} \rightarrow \mathcal{Z}$  and  $B^* : \mathcal{Y} \rightarrow \mathcal{Z}$  are the adjoints of the linear operators  $A : \mathcal{Z} \rightarrow \mathcal{X}$  and  $B : \mathcal{Z} \rightarrow \mathcal{Y}$ , respectively; and  $c \in \mathcal{Z}$ . The solution set of (1) is assumed to be nonempty throughout this paper.

Let  $\sigma \in (0, +\infty)$  be a given parameter. Define the augmented Lagrangian function as

$$\mathcal{L}_\sigma(x, y; z) := p(x) + f(x) + q(y) + g(y) + \langle z, A^*x + B^*y - c \rangle + \frac{\sigma}{2} \|A^*x + B^*y - c\|^2$$

for any  $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ . One may attempt to solve (1) by using the classical augmented Lagrangian method (ALM), which consists of the following iterations:

$$\begin{cases} (x^{k+1}, y^{k+1}) := \operatorname{argmin}_{x \in \mathcal{X}, y \in \mathcal{Y}} \mathcal{L}_\sigma(x, y; z^k), \\ z^{k+1} := z^k + \tau\sigma(A^*x^{k+1} + B^*y^{k+1} - c), \end{cases} \quad (2)$$

where  $\tau \in (0, 2)$  guarantees the convergence. Due to the non-separability of the quadratic penalty term in  $\mathcal{L}_\sigma$ , it is generally a challenging task to solve the joint minimization problem (2) exactly or approximately with a high accuracy (which may not be necessary at the early stage of the ALM). To overcome this difficulty, one may consider the following popular 2-block alternating direction method of multipliers (ADMM) to solve (1):

$$\begin{cases} x^{k+1} := \operatorname{argmin}_{x \in \mathcal{X}} \mathcal{L}_\sigma(x, y^k; z^k), \\ y^{k+1} := \operatorname{argmin}_{y \in \mathcal{Y}} \mathcal{L}_\sigma(x^{k+1}, y; z^k), \\ z^{k+1} := z^k + \tau\sigma(A^*x^{k+1} + B^*y^{k+1} - c), \end{cases} \quad (3)$$

where  $\tau \in (0, (1 + \sqrt{5})/2)$ . The convergence of the 2-block ADMM has long been established under various conditions and the classical literature includes [18, 15, 17, 13, 14, 9, 8]. For a recent survey, see [10].

By noting the facts that the subproblems in (3) may still be difficult to solve and that in many applications  $f$  or  $g$  is a convex quadratic function, Fazel et al. [12] advocated the use of the following semi-proximal ADMM scheme

$$\begin{cases} x^{k+1} := \operatorname{argmin}_{x \in \mathcal{X}} \mathcal{L}_\sigma(x, y^k; z^k) + \frac{1}{2} \|x - x^k\|_S^2, \\ y^{k+1} := \operatorname{argmin}_{y \in \mathcal{Y}} \mathcal{L}_\sigma(x^{k+1}, y; z^k) + \frac{1}{2} \|y - y^k\|_T^2, \\ z^{k+1} := z^k + \tau\sigma(A^*x^{k+1} + B^*y^{k+1} - c), \end{cases} \quad (4)$$

where  $\tau \in (0, (1 + \sqrt{5})/2)$ , and  $S \succeq 0$  and  $T \succeq 0$  are two self-adjoint and positive semidefinite (not necessarily positive definite) linear operators. We refer the readers to [12] as well as [7] for a brief history on the development of the semi-proximal ADMM scheme (4).

The successful applications of the 2-block ADMM in solving various problems to acceptable levels of moderate accuracy have inevitably inspired many researchers' interest in extending the scheme to the general  $m$ -block ( $m \geq 3$ ) case. However, it has been shown very recently by Chen et al. [1] via simple counterexamples that the direct extension of the ADMM to the simplest 3-block case can be divergent even if the step-length  $\tau$  is chosen to be as small as  $10^{-8}$ . This seems to

suggest that one has to give up the direct extension of  $m$ -block ( $m \geq 3$ ) ADMM unless if one is willing to take a sufficiently small step-length  $\tau$  as was shown by Hong and Luo in [21] or to take a small penalty parameter  $\sigma$  if at least  $m - 2$  blocks in the objective are strongly convex [19, 2, 25, 26, 22]. On the other hand, despite the potential divergence, the directly extended  $m$ -block ADMM with  $\tau \geq 1$  and an appropriate choice of  $\sigma$  often works very well in practice.

Recently, there is exciting progress in designing convergent and efficient ADMM type methods for solving multi-block linear and convex quadratic semidefinite programming problems [33, 23]. The convergence proof of the methods presented in [33] and [23] is via establishing their equivalence to particular cases of the general 2-block semi-proximal ADMM considered in [12]. It is this important fact that inspires us to extend the 2-block semi-proximal ADMM in [12] to a majorized ADMM with indefinite proximal terms (which we name it as Majorized iPADMM) in this paper. Our new algorithm has two important aspects. Firstly, we introduce a majorization technique to deal with the case where  $f$  and  $g$  in (1) may not be quadratic or linear functions. The purpose of the majorization is to make the corresponding subproblems in (4) more amenable to efficient computations. We note that a similar majorization technique has also been used by Wang and Banerjee [35] under the more general setting of Bregman distance functions. The drawback of the Bregman distance function based ADMM discussed in [35] is that the parameter  $\tau$  should be small for the global convergence. For example, if we choose the Euclidean distance as the Bregman divergence, then the corresponding parameter  $\tau$  should be smaller than 1. By focusing on the Euclidean divergence instead of the more general Bregman divergence, we allow  $\tau$  to stay in the larger interval  $(0, (1 + \sqrt{5})/2)$ . Secondly and more importantly, we allow the added proximal terms to be indefinite for better practical performance. The introduction of the indefinite proximal terms instead of the commonly used positive semidefinite or positive definite terms is motivated by numerical evidences showing that the former can outperform the latter in the majorized penalty approach for solving rank constrained matrix optimization problems in [16] and in solving linear semidefinite programming problems with a large number of inequality constraints in [33].

Here, we conduct a rigorous study of the conditions under which indefinite proximal terms are allowed within the 2-block ADMM while also establishing the convergence of the algorithm. We have thus provided the necessary theoretical support for the numerical observation just mentioned in establishing the convergence of the indefinite-proximal 2-block ADMM. Interestingly, Deng and Yin [7] mentioned that the matrix  $T$  in the ADMM scheme (4) may be indefinite if  $\tau \in (0, 1)$  though no further developments are given. As far as we are aware of, this is the first paper proving that indefinite proximal terms can be employed within the ADMM framework with convergence guarantee while not making restrictive assumptions on the step-length parameter  $\tau$  or the penalty parameter  $\sigma$ .

Besides establishing the convergence and a simple non-ergodic iteration-complexity of our proposed majorized indefinite-proximal ADMM, we also establish its worst-case  $O(1/k)$  ergodic iteration-complexity. The study of the ergodic iteration-complexity of the classical ADMM is inspired by Nemirovski [29], who proposed a prox-method with  $O(1/k)$  iteration-complexity for variational inequalities. Monteiro and Svaiter [28] analyzed the iteration-complexity of a hybrid proximal extragradient (HPE) method. They also considered the ergodic iteration-complexity of block-decomposition algorithms and the ADMM in [27]. He and Yuan [20] provided a simple and different proof for the  $O(1/k)$  ergodic iteration-complexity for a special semi-proximal ADMM scheme (where the  $x$ -part uses a semi-proximal term while the  $y$ -part does not). Tao and Yuan [34] proved the  $O(1/k)$  ergodic iteration-complexity of the ADMM with a logarithmic-quadratic proximal regu-

larization even for  $\tau \in (0, (1 + \sqrt{5})/2)$ . Ouyang et al. [31] provided an ergodic iteration-complexity for an accelerated linearized ADMM. Wang and Banerjee [35] generalized the ADMM to Bregman function based ADMM, which allows the choice of different Bregman divergences and still has the  $O(1/k)$  iteration-complexity.

The remaining parts of this paper are organized as follows. In Section 2, we summarize some useful results for further analysis. Then, we present our majorized indefinite-proximal ADMM in Section 3, followed by some basic properties on the generated sequence. In Section 4, we present the global convergence and the choices of proximal terms. The analysis of the non-ergodic iteration-complexity as well as the ergodic one is provided in Section 5. In Section 6, we provide some illustrative examples to show the potential numerical efficiency that one can gain from the new scheme when using an indefinite proximal term versus the standard choice of a positive semidefinite proximal term.

### Notation.

- The effective domain of a function  $h: \mathcal{X} \rightarrow (-\infty, +\infty]$  is defined as  $\text{dom}(h) := \{x \in \mathcal{X} \mid h(x) < +\infty\}$ .
- The set of all relative interior points of a convex set  $C$  is denoted by  $\text{ri}(C)$ .
- For convenience, we use  $\|x\|_S^2$  to denote  $\langle x, Sx \rangle$  even if  $S$  is only a self-adjoint linear operator which may be indefinite. If  $M: \mathcal{X} \rightarrow \mathcal{X}$  is a self-adjoint and positive semidefinite linear operator, we use  $M^{\frac{1}{2}}$  to denote the unique self-adjoint and positive semidefinite square root of  $M$ .

## 2 Preliminaries

In this section, we first introduce some notation to be used in our analysis and then summarize some useful preliminaries known in the literature.

Throughout this paper, we assume that the following assumption holds.

**Assumption 2.1.** Both  $f(\cdot)$  and  $g(\cdot)$  are smooth convex functions with Lipschitz continuous gradients.

Under Assumption 2.1, we know that there exist two self-adjoint and positive semidefinite linear operators  $\Sigma_f$  and  $\Sigma_g$  such that for any  $x, x' \in \mathcal{X}$  and any  $y, y' \in \mathcal{Y}$ ,

$$f(x) \geq f(x') + \langle x - x', \nabla f(x') \rangle + \frac{1}{2} \|x - x'\|_{\Sigma_f}^2, \quad (5)$$

$$g(y) \geq g(y') + \langle y - y', \nabla g(y') \rangle + \frac{1}{2} \|y - y'\|_{\Sigma_g}^2; \quad (6)$$

moreover, there exist self-adjoint and positive semidefinite linear operators  $\hat{\Sigma}_f \succeq \Sigma_f$  and  $\hat{\Sigma}_g \succeq \Sigma_g$  such that for any  $x, x' \in \mathcal{X}$  and any  $y, y' \in \mathcal{Y}$ ,

$$f(x) \leq \hat{f}(x; x') := f(x') + \langle x - x', \nabla f(x') \rangle + \frac{1}{2} \|x - x'\|_{\hat{\Sigma}_f}^2, \quad (7)$$

$$g(y) \leq \hat{g}(y; y') := g(y') + \langle y - y', \nabla g(y') \rangle + \frac{1}{2} \|y - y'\|_{\hat{\Sigma}_g}^2. \quad (8)$$

The two functions  $\hat{f}$  and  $\hat{g}$  are called the majorized convex functions of  $f$  and  $g$ , respectively. For any given  $y \in \mathcal{Y}$ , let  $\partial^2 g(y)$  be Clarke's generalized Jacobian of  $\nabla g(\cdot)$  at  $y$ , i.e.,

$$\partial^2 g(y) = \text{conv}\left\{ \lim_{y^k \rightarrow y} \nabla^2 g(y^k) : \nabla^2 g(y^k) \text{ exists} \right\}, \quad (9)$$

where “conv” denotes the convex hull. Then for any given  $y \in \mathcal{Y}$ ,  $W \in \partial^2 g(y)$  is a self-adjoint and positive semidefinite linear operator satisfying

$$\widehat{\Sigma}_g \succeq W \succeq \Sigma_g \succeq 0. \quad (10)$$

For further discussions, we need the following constraint qualification.

**Assumption 2.2.** There exists  $(x_0, y_0) \in \text{ri}(\text{dom}(p) \times \text{dom}(q)) \cap P$ , where  $P := \{(x, y) \in \mathcal{X} \times \mathcal{Y} \mid A^*x + B^*y = c\}$ .

Under Assumption 2.2, it follows from [32, Corollary 28.2.2] and [32, Corollary 28.3.1] that  $(\bar{x}, \bar{y}) \in \text{dom}(p) \times \text{dom}(q)$  is an optimal solution to problem (1) if and only if there exists a Lagrange multiplier  $\bar{z} \in \mathcal{Z}$  such that  $(\bar{x}, \bar{y}, \bar{z})$  satisfies the following Karush-Kuhn-Tucker (KKT) system

$$0 \in \partial p(\bar{x}) + \nabla f(\bar{x}) + A\bar{z}, \quad 0 \in \partial q(\bar{y}) + \nabla g(\bar{y}) + B\bar{z}, \quad c - A^*\bar{x} - B^*\bar{y} = 0, \quad (11)$$

where  $\partial p(\cdot)$  and  $\partial q(\cdot)$  are the subdifferential mappings of  $p$  and  $q$ , respectively. Moreover, any  $\bar{z} \in \mathcal{Z}$  satisfying (11) is an optimal solution to the dual of problem (1). Therefore, we call  $(\hat{x}, \hat{y}, \hat{z}) \in \text{dom}(p) \times \text{dom}(q) \times \mathcal{Z}$  an  $\varepsilon$ -approximate KKT point of (1) if it satisfies

$$d^2(0, \partial p(\hat{x}) + \nabla f(\hat{x}) + A\hat{z}) + d^2(0, \partial q(\hat{y}) + \nabla g(\hat{y}) + B\hat{z}) + \|A^*\hat{x} + B^*\hat{y} - c\|^2 \leq \varepsilon,$$

where  $d(w, S)$  denotes the Euclidean distance of a given point  $w$  to a set  $S$ .

By the assumption that  $p$  and  $q$  are convex functions, (11) is equivalent to finding a vector  $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$  such that for any  $(x, y) \in \mathcal{X} \times \mathcal{Y}$ , we have

$$\begin{cases} p(x) - p(\bar{x}) + \langle x - \bar{x}, \nabla f(\bar{x}) + A\bar{z} \rangle \geq 0, \\ q(y) - q(\bar{y}) + \langle y - \bar{y}, \nabla g(\bar{y}) + B\bar{z} \rangle \geq 0, \\ c - A^*\bar{x} - B^*\bar{y} = 0 \end{cases} \quad (12)$$

or equivalently

$$\begin{cases} (p(x) + f(x)) - (p(\bar{x}) + f(\bar{x})) + \langle x - \bar{x}, A\bar{z} \rangle \geq 0, \\ (q(y) + g(y)) - (q(\bar{y}) + g(\bar{y})) + \langle y - \bar{y}, B\bar{z} \rangle \geq 0, \\ c - A^*\bar{x} - B^*\bar{y} = 0, \end{cases} \quad (13)$$

which are obtained by using the assumption that  $f$  and  $g$  are smooth convex functions.

It is easy to see that (12) can be rewritten as the following variational inequality problem: find a vector  $\bar{w} := (\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W} := \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$  such that

$$\theta(u) - \theta(\bar{u}) + \langle w - \bar{w}, F(\bar{w}) \rangle \geq 0 \quad \forall w \in \mathcal{W} \quad (14)$$

with

$$u := \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) := p(x) + q(y), \quad w := \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{and} \quad F(w) := \begin{pmatrix} \nabla f(x) + Az \\ \nabla g(y) + Bz \\ c - A^*x - B^*y \end{pmatrix}. \quad (15)$$

We denote by  $\text{VI}(\mathcal{W}, F, \theta)$  the variational inequality problem (14)-(15); and by  $\mathcal{W}^*$  the solution set of  $\text{VI}(\mathcal{W}, F, \theta)$ , which is nonempty under Assumption 2.2 and the fact that the solution set of problem (1) is assumed to be nonempty. Note that the mapping  $F(\cdot)$  in (15) is monotone with respect to  $\mathcal{W}$ . Thus by [11, Theorem 2.3.5], the solution set  $\mathcal{W}^*$  of  $\text{VI}(\mathcal{W}, F, \theta)$  is closed and convex and it can be characterized as follows:

$$\mathcal{W}^* := \bigcap_{w \in \mathcal{W}} \{\tilde{w} \in \mathcal{W} \mid \theta(u) - \theta(\tilde{u}) + \langle w - \tilde{w}, F(w) \rangle \geq 0\}.$$

Similarly as [30, Definition 1], we give the following definition for an  $\varepsilon$ -approximation solution of the variational inequality problem.

**Definition 2.1.**  $\tilde{w} \in \mathcal{W}$  is an  $\varepsilon$ -approximation solution of  $\text{VI}(\mathcal{W}, F, \theta)$  if it satisfies

$$\sup_{w \in \mathcal{B}(\tilde{w})} \{\theta(\tilde{u}) - \theta(u) + \langle \tilde{w} - w, F(w) \rangle\} \leq \varepsilon, \quad \text{where } \mathcal{B}(\tilde{w}) := \{w \in \mathcal{W} \mid \|w - \tilde{w}\| \leq 1\}. \quad (16)$$

Based on this definition, the worst-case  $O(1/k)$  ergodic iteration-complexity of our proposed algorithm will be established in the sense that we can find a  $\tilde{w} \in \mathcal{W}$  such that

$$\theta(\tilde{u}) - \theta(u) + \langle \tilde{w} - w, F(w) \rangle \leq \varepsilon \quad \forall w \in \mathcal{B}(\tilde{w})$$

with  $\varepsilon = O(1/k)$ , after  $k$  iterations.

The following lemma, motivated by [6, Lemma 1.2], is convenient for discussing the non-ergodic iteration-complexity.

**Lemma 2.1.** *If a sequence  $\{a_i\} \subseteq \mathbb{R}$  obeys: (1)  $a_i \geq 0$ ; (2)  $\sum_{i=1}^{\infty} a_i < +\infty$ , then we have  $\min_{1 \leq i \leq k} \{a_i\} = o(1/k)$ .*

*Proof.* Since  $\min_{1 \leq i \leq 2k} \{a_i\} \leq a_j$  for any  $k+1 \leq j \leq 2k$ , we get

$$0 \leq k \cdot \min_{1 \leq i \leq 2k} \{a_i\} \leq \sum_{i=k+1}^{2k} a_i \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

Therefore, we get  $\min_{1 \leq i \leq k} \{a_i\} = o(1/k)$ . The proof is completed.  $\square$

### 3 A majorized ADMM with indefinite proximal terms

Let  $z \in \mathcal{Z}$  be the Lagrange multiplier associated with the linear equality constraint in (1) and let the Lagrangian function of (1) be

$$\mathcal{L}(x, y; z) := p(x) + f(x) + q(y) + g(y) + \langle z, A^*x + B^*y - c \rangle \quad (17)$$

defined on  $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ . Similarly, for given  $(x', y') \in \mathcal{X} \times \mathcal{Y}$ ,  $\sigma \in (0, +\infty)$  and any  $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ , define the majorized augmented Lagrangian function as follows:

$$\widehat{\mathcal{L}}_\sigma(x, y; (z, x', y')) := p(x) + \widehat{f}(x; x') + q(y) + \widehat{g}(y; y') + \langle z, A^*x + B^*y - c \rangle + \frac{\sigma}{2} \|A^*x + B^*y - c\|^2, \quad (18)$$

where the two majorized convex functions  $\widehat{f}$  and  $\widehat{g}$  are defined by (7) and (8), respectively. Our promised majorized ADMM with indefinite proximal terms for solving problem (1) can then be described as in the following.

**Majorized iPADMM: A majorized ADMM with indefinite proximal terms for solving problem (1).**

Let  $\sigma \in (0, +\infty)$  and  $\tau \in (0, +\infty)$  be given parameters. Let  $S$  and  $T$  be given self-adjoint, possibly indefinite, linear operators defined on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively such that

$$\mathcal{P} := \widehat{\Sigma}_f + S + \sigma AA^* \succeq 0 \quad \text{and} \quad \mathcal{Q} := \widehat{\Sigma}_g + T + \sigma BB^* \succeq 0.$$

Choose  $(x^0, y^0, z^0) \in \text{dom}(p) \times \text{dom}(q) \times \mathcal{Z}$ . Set  $k = 0$  and denote  $\widehat{r}^0 := A^*x^0 + B^*y^0 - c + \sigma^{-1}z^0$ .

**Step 1.** Compute

$$\left\{ \begin{array}{ll} x^{k+1} &:= \underset{x \in \mathcal{X}}{\text{argmin}} \widehat{\mathcal{L}}_\sigma(x, y^k; (z^k, x^k, y^k)) + \frac{1}{2} \|x - x^k\|_S^2 \\ &= \underset{x \in \mathcal{X}}{\text{argmin}} p(x) + \frac{1}{2} \langle x, \mathcal{P}x \rangle + \langle \nabla f(x^k) + \sigma A \widehat{r}^k - \mathcal{P}x^k, x \rangle, \\ y^{k+1} &:= \underset{y \in \mathcal{Y}}{\text{argmin}} \widehat{\mathcal{L}}_\sigma(x^{k+1}, y; (z^k, x^k, y^k)) + \frac{1}{2} \|y - y^k\|_T^2 \\ &= \underset{y \in \mathcal{Y}}{\text{argmin}} q(y) + \frac{1}{2} \langle y, \mathcal{Q}y \rangle + \langle \nabla g(y^k) + \sigma B(\widehat{r}^k + A^*(x^{k+1} - x^k)) - \mathcal{Q}y^k, y \rangle, \\ z^{k+1} &:= z^k + \tau \sigma (A^*x^{k+1} + B^*y^{k+1} - c). \end{array} \right. \quad (19)$$

**Step 2.** If a termination criterion is not met, denote  $\widehat{r}^{k+1} := A^*x^{k+1} + B^*y^{k+1} - c + \sigma^{-1}z^{k+1}$ . Set  $k := k + 1$  and go to Step 1.

*Remark 3.1.* In the above Majorized iPADMM for solving problem (1), the presence of the two self-adjoint operators  $S$  and  $T$  first helps to guarantee the existence of solutions for the subproblems in (19). Secondly, they play an important role in ensuring the boundedness of the two generated sequences  $\{x^{k+1}\}$  and  $\{y^{k+1}\}$ . Thirdly, as demonstrated in [23], the introduction of  $S$  and  $T$  is the key for dealing with additionally an arbitrary number of convex quadratic and linear functions. Hence, these two proximal terms are preferred although the choices of  $S$  and  $T$  are very much problem dependent. The general principle is that both  $S$  and  $T$  should be chosen such that  $x^{k+1}$  and  $y^{k+1}$  take larger step-lengths while they are still relatively easy to compute. From a numerical point of view, it is therefore advantageous to pick an indefinite  $S$  or  $T$  whenever possible. The issue on how to choose  $S$  and  $T$  will be discussed in the later sections.

For notational convenience, for given  $\alpha \in (0, 1]$  and  $\tau \in (0, +\infty)$ , denote

$$H_f := \frac{1}{2} \Sigma_f + S + \frac{1}{2} (1 - \alpha) \sigma AA^* \quad \text{and} \quad M_g := \frac{1}{2} \Sigma_g + T + \min(\tau, 1 + \tau - \tau^2) \alpha \sigma BB^*, \quad (20)$$

for  $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ ,  $k = 0, 1, \dots$ , define

$$\phi_k(x, y, z) := (\tau\sigma)^{-1} \|z^k - z\|^2 + \|x^k - x\|_{\Sigma_f + S}^2 + \|y^k - y\|_{\Sigma_g + T}^2 + \sigma \|A^*x + B^*y^k - c\|^2, \quad (21)$$

$$\begin{cases} \xi_{k+1} := \|y^{k+1} - y^k\|_{\Sigma_g + T}^2, \\ s_{k+1} := \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S}^2 + \|y^{k+1} - y^k\|_{\frac{1}{2}\Sigma_g + T}^2, \\ t_{k+1} := \|x^{k+1} - x^k\|_{H_f}^2 + \|y^{k+1} - y^k\|_{M_g}^2 \end{cases} \quad (22)$$

and

$$r^k := A^*x^k + B^*y^k - c, \quad \tilde{z}^{k+1} := z^k + \sigma r^{k+1}. \quad (23)$$

We also recall the following elementary identities which will be used later.

**Lemma 3.1.** (a) For any vectors  $u_1, u_2, v_1, v_2$  in the same Euclidean vector space  $\mathcal{X}$ , we have the identity:

$$\langle u_1 - u_2, v_1 - v_2 \rangle = \frac{1}{2}(\|v_2 - u_1\|^2 - \|v_1 - u_1\|^2) + \frac{1}{2}(\|v_1 - u_2\|^2 - \|v_2 - u_2\|^2). \quad (24)$$

(b) For any vectors  $u, v$  in the same Euclidean vector space  $\mathcal{X}$  and any self-adjoint linear operator  $G : \mathcal{X} \rightarrow \mathcal{X}$ , we have the identity:

$$\langle u, Gv \rangle = \frac{1}{2}(\|u\|_G^2 + \|v\|_G^2 - \|u - v\|_G^2) = \frac{1}{2}(\|u + v\|_G^2 - \|u\|_G^2 - \|v\|_G^2). \quad (25)$$

To prove the global convergence for the Majorized iPADMM, we first present some useful lemmas.

**Lemma 3.2.** Suppose that Assumption 2.1 holds. Let  $\{z^{k+1}\}$  be generated by (19) and let  $\{\tilde{z}^{k+1}\}$  be defined by (23). Then for any  $z \in \mathcal{Z}$  we have for  $k \geq 0$  that

$$\frac{1}{\sigma} \langle z - \tilde{z}^{k+1}, \tilde{z}^{k+1} - z^k \rangle + \frac{1}{2\sigma} \|z^k - \tilde{z}^{k+1}\|^2 = \frac{1}{2\tau\sigma} (\|z^k - z\|^2 - \|z^{k+1} - z\|^2) + \frac{\tau - 1}{2\sigma} \|z^k - \tilde{z}^{k+1}\|^2. \quad (26)$$

*Proof.* From (19) and (23), we get

$$z^{k+1} - z^k = \tau\sigma r^{k+1} \quad \text{and} \quad \tilde{z}^{k+1} - z^k = \sigma r^{k+1}. \quad (27)$$

It follows from (27) that

$$\tilde{z}^{k+1} - z^k = \frac{1}{\tau}(z^{k+1} - z^k) \quad \text{and} \quad z^{k+1} - \tilde{z}^{k+1} = -(\tau - 1)(z^k - \tilde{z}^{k+1}). \quad (28)$$

By using the first equation in (28), we obtain

$$\frac{1}{\sigma} \langle z - \tilde{z}^{k+1}, \tilde{z}^{k+1} - z^k \rangle + \frac{1}{2\sigma} \|z^k - \tilde{z}^{k+1}\|^2 = \frac{1}{\tau\sigma} \langle z - \tilde{z}^{k+1}, z^{k+1} - z^k \rangle + \frac{1}{2\sigma} \|z^k - \tilde{z}^{k+1}\|^2. \quad (29)$$



Now, by taking  $u_1 = z$ ,  $u_2 = \tilde{z}^{k+1}$ ,  $v_1 = z^{k+1}$  and  $v_2 = z^k$  and applying the identity (24) to the first term of the right-hand side of (29), we obtain

$$\begin{aligned} & \frac{1}{\sigma} \langle z - \tilde{z}^{k+1}, \tilde{z}^{k+1} - z^k \rangle + \frac{1}{2\sigma} \|z^k - \tilde{z}^{k+1}\|^2 \\ &= \frac{1}{2\tau\sigma} \left( \|z^k - z\|^2 - \|z^{k+1} - z\|^2 \right) + \frac{1}{2\tau\sigma} \left( \|z^{k+1} - \tilde{z}^{k+1}\|^2 - \|z^k - \tilde{z}^{k+1}\|^2 + \tau \|z^k - \tilde{z}^{k+1}\|^2 \right). \end{aligned} \quad (30)$$

By using the second equation in (28), we have

$$\begin{aligned} & \|z^{k+1} - \tilde{z}^{k+1}\|^2 - \|z^k - \tilde{z}^{k+1}\|^2 + \tau \|z^k - \tilde{z}^{k+1}\|^2 \\ &= (\tau - 1)^2 \|z^k - \tilde{z}^{k+1}\|^2 - \|z^k - \tilde{z}^{k+1}\|^2 + \tau \|z^k - \tilde{z}^{k+1}\|^2 = \tau(\tau - 1) \|z^k - \tilde{z}^{k+1}\|^2, \end{aligned}$$

which, together with (30), proves the assertion (26).  $\square$

**Lemma 3.3.** *Suppose that Assumption 2.1 holds. Assume that  $\frac{1}{2}\widehat{\Sigma}_g + T \succeq 0$ . Let  $\{(x^k, y^k, z^k)\}$  be generated by the Majorized iPADMM and for each  $k$ , let  $\xi_k$  and  $r^k$  be defined as in (22) and (23), respectively. Then for any  $k \geq 1$ , we have*

$$\begin{aligned} & (1 - \tau)\sigma \|r^{k+1}\|^2 + \sigma \|A^* x^{k+1} + B^* y^k - c\|^2 \\ & \geq \max(1 - \tau, 1 - \tau^{-1})\sigma (\|r^{k+1}\|^2 - \|r^k\|^2) + (\xi_{k+1} - \xi_k) \\ & \quad + \min(\tau, 1 + \tau - \tau^2)\sigma (\tau^{-1}\|r^{k+1}\|^2 + \|B^*(y^{k+1} - y^k)\|^2). \end{aligned} \quad (31)$$

*Proof.* Note that

$$\begin{aligned} & (1 - \tau)\sigma \|r^{k+1}\|^2 + \sigma \|A^* x^{k+1} + B^* y^k - c\|^2 \\ &= (1 - \tau)\sigma \|r^{k+1}\|^2 + \sigma \|(A^* x^{k+1} + B^* y^{k+1} - c) + B^*(y^k - y^{k+1})\|^2 \\ &= (2 - \tau)\sigma \|r^{k+1}\|^2 + \sigma \|B^*(y^{k+1} - y^k)\|^2 + 2\sigma \langle B^*(y^k - y^{k+1}), r^{k+1} \rangle. \end{aligned} \quad (32)$$

First, we shall estimate the term  $2\sigma \langle B^*(y^k - y^{k+1}), r^{k+1} \rangle$  in (32). From the first-order optimality condition of (19) and the notation of  $\tilde{z}^{k+1}$  defined in (23), we have

$$\begin{aligned} & -\nabla g(y^k) - B\tilde{z}^{k+1} - (\widehat{\Sigma}_g + T)(y^{k+1} - y^k) \in \partial q(y^{k+1}), \\ & -\nabla g(y^{k-1}) - B\tilde{z}^k - (\widehat{\Sigma}_g + T)(y^k - y^{k-1}) \in \partial q(y^k). \end{aligned} \quad (33)$$

From Clarke's Mean Value Theorem [3, Proposition 2.6.5], we know that

$$\nabla g(y^k) - \nabla g(y^{k-1}) \in \text{conv}\{\partial^2 g([y^{k-1}, y^k])(y^k - y^{k-1})\}.$$

Thus, there exists a self-adjoint and positive semidefinite linear operator  $W^k \in \text{conv}\{\partial^2 g([y^{k-1}, y^k])\}$  such that

$$\nabla g(y^k) - \nabla g(y^{k-1}) = W^k(y^k - y^{k-1}). \quad (34)$$

From (33) and the maximal monotonicity of  $\partial q(\cdot)$ , it follows that

$$\langle y^k - y^{k+1}, [-\nabla g(y^{k-1}) - B\tilde{z}^k - (\widehat{\Sigma}_g + T)(y^k - y^{k-1})] - [-\nabla g(y^k) - B\tilde{z}^{k+1} - (\widehat{\Sigma}_g + T)(y^{k+1} - y^k)] \rangle \geq 0,$$

which, together with (34), gives rise to

$$\begin{aligned}
& 2\langle y^k - y^{k+1}, B\tilde{z}^{k+1} - B\tilde{z}^k \rangle \\
& \geq 2\langle \nabla g(y^k) - \nabla g(y^{k+1}), y^{k+1} - y^k \rangle - 2\langle (\hat{\Sigma}_g + T)(y^k - y^{k-1}), y^{k+1} - y^k \rangle + 2\|y^{k+1} - y^k\|_{\hat{\Sigma}_g + T}^2 \\
& = 2\langle W^k(y^k - y^{k-1}), y^{k+1} - y^k \rangle - 2\langle (\hat{\Sigma}_g + T)(y^k - y^{k-1}), y^{k+1} - y^k \rangle + 2\|y^{k+1} - y^k\|_{\hat{\Sigma}_g + T}^2 \\
& = 2\langle (\hat{\Sigma}_g - W^k + T)(y^{k-1} - y^k), y^{k+1} - y^k \rangle + 2\|y^{k+1} - y^k\|_{\hat{\Sigma}_g + T}^2.
\end{aligned} \tag{35}$$

By using the first elementary identity in (25) and  $W^k \succeq 0$ , we have

$$\begin{aligned}
& 2\langle (\hat{\Sigma}_g - W^k + T)(y^{k-1} - y^k), y^{k+1} - y^k \rangle \\
& = \|y^{k+1} - y^k\|_{\hat{\Sigma}_g - W^k + T}^2 + \|y^k - y^{k-1}\|_{\hat{\Sigma}_g - W^k + T}^2 - \|y^{k+1} - y^{k-1}\|_{\hat{\Sigma}_g - W^k + T}^2 \\
& \geq \|y^{k+1} - y^k\|_{\hat{\Sigma}_g - W^k + T}^2 + \|y^k - y^{k-1}\|_{\hat{\Sigma}_g - W^k + T}^2 - \|y^{k+1} - y^{k-1}\|_{\hat{\Sigma}_g - \frac{1}{2}W^k + T}^2.
\end{aligned}$$

From (10), we know that  $\hat{\Sigma}_g - \frac{1}{2}W^k + T = \frac{1}{2}\hat{\Sigma}_g + \frac{1}{2}(\hat{\Sigma}_g - W^k) + T \succeq \frac{1}{2}\hat{\Sigma}_g + T \succeq 0$ . Then, by using the elementary inequality  $\|u + v\|_G^2 \leq 2\|u\|_G^2 + 2\|v\|_G^2$  for any self-adjoint and positive semidefinite linear operator  $G$ , we get

$$\begin{aligned}
-\|y^{k+1} - y^{k-1}\|_{\hat{\Sigma}_g - \frac{1}{2}W^k + T}^2 & = -\|(y^{k+1} - y^k) + (y^k - y^{k-1})\|_{\hat{\Sigma}_g - \frac{1}{2}W^k + T}^2 \\
& \geq -2\left(\|y^{k+1} - y^k\|_{\hat{\Sigma}_g - \frac{1}{2}W^k + T}^2 + \|y^k - y^{k-1}\|_{\hat{\Sigma}_g - \frac{1}{2}W^k + T}^2\right).
\end{aligned}$$

Substituting the above inequalities into (35), we obtain

$$\begin{aligned}
2\langle y^k - y^{k+1}, B\tilde{z}^{k+1} - B\tilde{z}^k \rangle & \geq -\|y^{k+1} - y^k\|_{\hat{\Sigma}_g + T}^2 - \|y^k - y^{k-1}\|_{\hat{\Sigma}_g + T}^2 + 2\|y^{k+1} - y^k\|_{\hat{\Sigma}_g + T}^2 \\
& = \|y^{k+1} - y^k\|_{\hat{\Sigma}_g + T}^2 - \|y^k - y^{k-1}\|_{\hat{\Sigma}_g + T}^2.
\end{aligned} \tag{36}$$

Thus, by letting  $\mu_{k+1} := (1 - \tau)\sigma\langle B^*(y^k - y^{k+1}), r^k \rangle$ , and using  $\sigma r^{k+1} = (1 - \tau)\sigma r^k + \tilde{z}^{k+1} - \tilde{z}^k$  (see (19) and (23)) and (36), we have

$$\begin{aligned}
2\sigma\langle B^*(y^k - y^{k+1}), r^{k+1} \rangle & = 2\mu_{k+1} + 2\langle y^k - y^{k+1}, B\tilde{z}^{k+1} - B\tilde{z}^k \rangle \\
& \geq 2\mu_{k+1} + \|y^{k+1} - y^k\|_{\hat{\Sigma}_g + T}^2 - \|y^k - y^{k-1}\|_{\hat{\Sigma}_g + T}^2.
\end{aligned} \tag{37}$$

Since  $\tau \in (0, +\infty)$ , from the definition of  $\mu_{k+1}$ , we obtain

$$2\mu_{k+1} \geq \begin{cases} -(1 - \tau)\sigma\|B^*(y^{k+1} - y^k)\|^2 - (1 - \tau)\sigma\|r^k\|^2 & \text{if } \tau \in (0, 1], \\ (1 - \tau)\sigma\tau\|B^*(y^{k+1} - y^k)\|^2 + (1 - \tau)\sigma\tau^{-1}\|r^k\|^2 & \text{if } \tau \in (1, +\infty), \end{cases}$$

which, together with (32), (37) and the notation of  $\xi_k$ , shows that (31) holds. This completes the proof.  $\square$

Now we are ready to present an inequality from which an upper bound for  $\theta(\tilde{u}^{k+1}) - \theta(u) + \langle \tilde{w}^{k+1} - w, F(w) \rangle$  (i.e.,  $\langle p(x^{k+1}) + q(y^{k+1}), -(p(x) + q(y)) \rangle + \langle x^{k+1} - x, \nabla f(x) + Az \rangle + \langle y^{k+1} - y, \nabla g(y) + Bz \rangle + \langle \tilde{z}^{k+1} - z, -(A^*x + B^*y - c) \rangle$ ) with  $\tilde{u}^{k+1} = (x^{k+1}, y^{k+1})$  and  $\tilde{w}^{k+1} = (x^{k+1}, y^{k+1}, \tilde{z}^{k+1})$  can be found for all  $w = (x, y, z) \in \mathcal{W}$ . This inequality is also crucial for analyzing the iteration-complexity for the sequence generated by the Majorized iPADMM.

**Proposition 3.1.** *Suppose that Assumption 2.1 holds. Let  $\{(x^k, y^k, z^k)\}$  be generated by the Majorized iPADMM. For each  $k$  and  $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ , let  $\phi_k(x, y, z)$ ,  $\xi_{k+1}$ ,  $s_{k+1}$ ,  $t_{k+1}$ ,  $r^k$  and  $\tilde{z}^{k+1}$  be defined as in (21), (22) and (23). Then the following results hold:*

(a) *For any  $k \geq 0$  and  $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ , we have*

$$\begin{aligned} & (p(x^{k+1}) + q(y^{k+1})) - (p(x) + q(y)) + \langle x^{k+1} - x, \nabla f(x) + Az \rangle + \langle y^{k+1} - y, \nabla g(y) + Bz \rangle \\ & + \langle \tilde{z}^{k+1} - z, -(A^*x + B^*y - c) \rangle + \frac{1}{2}(\phi_{k+1}(x, y, z) - \phi_k(x, y, z)) \\ & \leq -\frac{1}{2}\left(s_{k+1} + (1 - \tau)\sigma\|r^{k+1}\|^2 + \sigma\|A^*x^{k+1} + B^*y^k - c\|^2\right). \end{aligned} \quad (38)$$

(b) *Assume it holds that  $\frac{1}{2}\widehat{\Sigma}_g + T \succeq 0$ . Then for any  $\alpha \in (0, 1]$ ,  $k \geq 1$  and  $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ , we have*

$$\begin{aligned} & (p(x^{k+1}) + q(y^{k+1})) - (p(x) + q(y)) + \langle x^{k+1} - x, \nabla f(x) + Az \rangle + \langle y^{k+1} - y, \nabla g(y) + Bz \rangle \\ & + \langle \tilde{z}^{k+1} - z, -(A^*x + B^*y - c) \rangle + \frac{1}{2}\left\{[\phi_{k+1}(x, y, z) + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^{k+1}\|^2\right. \\ & \left. + \alpha\xi_{k+1}] - [\phi_k(x, y, z) + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^k\|^2 + \alpha\xi_k]\right\} \\ & \leq -\frac{1}{2}\left\{t_{k+1} + [-\tau + \alpha \min(1 + \tau, 1 + \tau^{-1})]\sigma\|r^{k+1}\|^2\right\}. \end{aligned} \quad (39)$$

*Proof.* By setting  $x = x^{k+1}$  and  $x' = x^k$  in (7), we have

$$f(x^{k+1}) \leq f(x^k) + \langle x^{k+1} - x^k, \nabla f(x^k) \rangle + \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2.$$

Setting  $x' = x^k$  in (5), we get

$$f(x) \geq f(x^k) + \langle x - x^k, \nabla f(x^k) \rangle + \frac{1}{2}\|x^k - x\|_{\Sigma_f}^2 \quad \forall x \in \mathcal{X}.$$

Combining the above two inequalities, we obtain

$$f(x) - f(x^{k+1}) - \frac{1}{2}\|x^k - x\|_{\Sigma_f}^2 + \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 \geq \langle x - x^{k+1}, \nabla f(x^k) \rangle \quad \forall x \in \mathcal{X}. \quad (40)$$

From the first-order optimality condition of (19), for any  $x \in \mathcal{X}$ , we have

$$p(x) - p(x^{k+1}) + \langle x - x^{k+1}, \nabla f(x^k) + A[z^k + \sigma(A^*x^{k+1} + B^*y^k - c)] + (\widehat{\Sigma}_f + S)(x^{k+1} - x^k) \rangle \geq 0. \quad (41)$$

Substituting (40) into (41), we get

$$\begin{aligned} & (p(x) + f(x)) - (p(x^{k+1}) + f(x^{k+1})) \\ & + \langle x - x^{k+1}, A[z^k + \sigma(A^*x^{k+1} + B^*y^k - c)] + (\widehat{\Sigma}_f + S)(x^{k+1} - x^k) \rangle \\ & \geq \frac{1}{2}\|x^k - x\|_{\Sigma_f}^2 - \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 \quad \forall x \in \mathcal{X}. \end{aligned} \quad (42)$$

Using the similar derivation as to get (42), we have for any  $y \in \mathcal{Y}$ ,

$$\begin{aligned}
& (q(y) + g(y)) - (q(y^{k+1}) + g(y^{k+1})) \\
& + \langle y - y^{k+1}, B[z^k + \sigma(A^*x^{k+1} + B^*y^{k+1} - c)] + (\widehat{\Sigma}_g + T)(y^{k+1} - y^k) \rangle \\
& \geq \frac{1}{2}\|y^k - y\|_{\Sigma_g}^2 - \frac{1}{2}\|y^{k+1} - y^k\|_{\widehat{\Sigma}_g}^2 \quad \forall y \in \mathcal{Y}.
\end{aligned} \tag{43}$$

Note that  $\tilde{z}^{k+1} = z^k + \sigma r^{k+1}$ , where  $r^{k+1} = A^*x^{k+1} + B^*y^{k+1} - c$ . Then we have

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

Adding up (42) and (43), and using the above equation, we have for any  $(x, y) \in \mathcal{X} \times \mathcal{Y}$ ,

$$\begin{aligned}
& (p(x) + f(x) + q(y) + g(y)) - (p(x^{k+1}) + f(x^{k+1}) + q(y^{k+1}) + g(y^{k+1})) \\
& + \langle x - x^{k+1}, A\tilde{z}^{k+1} + \sigma AB^*(y^k - y^{k+1}) + (\widehat{\Sigma}_f + S)(x^{k+1} - x^k) \rangle \\
& + \langle y - y^{k+1}, B\tilde{z}^{k+1} + (\widehat{\Sigma}_g + T)(y^{k+1} - y^k) \rangle \\
& \geq \frac{1}{2}\|x^k - x\|_{\Sigma_f}^2 - \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 + \frac{1}{2}\|y^k - y\|_{\Sigma_g}^2 - \frac{1}{2}\|y^{k+1} - y^k\|_{\widehat{\Sigma}_g}^2.
\end{aligned} \tag{44}$$

Now setting  $x = x^{k+1}$ ,  $x' = x$  in (5), and  $y = y^{k+1}$ ,  $y' = y$  in (6), we get

$$f(x^{k+1}) - f(x) + \langle x - x^{k+1}, \nabla f(x) \rangle \geq \frac{1}{2}\|x^{k+1} - x\|_{\Sigma_f}^2 \quad \forall x \in \mathcal{X}, \tag{45}$$

$$g(y^{k+1}) - g(y) + \langle y - y^{k+1}, \nabla g(y) \rangle \geq \frac{1}{2}\|y^{k+1} - y\|_{\Sigma_g}^2 \quad \forall y \in \mathcal{Y}. \tag{46}$$

Let

$$\Delta^k(x, y, z) := \langle x - x^{k+1}, A(\tilde{z}^{k+1} - z) + \sigma AB^*(y^k - y^{k+1}) \rangle + \langle y - y^{k+1}, B(\tilde{z}^{k+1} - z) \rangle.$$

Adding up (44), (45) and (46), and using the elementary inequality  $\frac{1}{2}\|u\|_G^2 + \frac{1}{2}\|v\|_G^2 \geq \frac{1}{4}\|u - v\|_G^2$  for any self-adjoint and positive semidefinite linear operator  $G$ , we have

$$\begin{aligned}
& (p(x) + q(y)) - (p(x^{k+1}) + q(y^{k+1})) + \langle x - x^{k+1}, \nabla f(x) + Az \rangle + \langle y - y^{k+1}, \nabla g(y) + Bz \rangle \\
& + \Delta^k(x, y, z) + \langle x - x^{k+1}, (\widehat{\Sigma}_f + S)(x^{k+1} - x^k) \rangle + \langle y - y^{k+1}, (\widehat{\Sigma}_g + T)(y^{k+1} - y^k) \rangle \\
& \geq \frac{1}{2}\|x^k - x\|_{\Sigma_f}^2 + \frac{1}{2}\|x^{k+1} - x\|_{\Sigma_f}^2 + \frac{1}{2}\|y^k - y\|_{\Sigma_g}^2 + \frac{1}{2}\|y^{k+1} - y\|_{\Sigma_g}^2 \\
& - \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 - \frac{1}{2}\|y^{k+1} - y^k\|_{\widehat{\Sigma}_g}^2 \\
& \geq \frac{1}{4}\|x^{k+1} - x^k\|_{\Sigma_f}^2 + \frac{1}{4}\|y^{k+1} - y^k\|_{\Sigma_g}^2 - \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 - \frac{1}{2}\|y^{k+1} - y^k\|_{\widehat{\Sigma}_g}^2.
\end{aligned} \tag{47}$$

By simple manipulations, we have

$$\begin{aligned}
\Delta^k(x, y, z) & = \langle A^*(x^{k+1} - x), z - \tilde{z}^{k+1} + \sigma B^*(y^{k+1} - y^k) \rangle + \langle B^*(y^{k+1} - y), z - \tilde{z}^{k+1} \rangle \\
& = -\langle A^*x + B^*y - c, z - \tilde{z}^{k+1} \rangle + \langle r^{k+1}, z - \tilde{z}^{k+1} \rangle + \sigma \langle A^*(x^{k+1} - x), B^*(y^{k+1} - y^k) \rangle
\end{aligned} \tag{48}$$

and

$$\sigma \langle A^*(x^{k+1} - x), B^*(y^{k+1} - y^k) \rangle = \sigma \langle (A^*x - c) - (A^*x^{k+1} - c), (-B^*y^{k+1}) - (-B^*y^k) \rangle. \quad (49)$$

Now, by taking  $u_1 = A^*x - c$ ,  $u_2 = A^*x^{k+1} - c$ ,  $v_1 = -B^*y^{k+1}$  and  $v_2 = -B^*y^k$  and applying the identity (24) to the right-hand side of (49), we obtain

$$\begin{aligned} \sigma \langle A^*(x^{k+1} - x), B^*(y^{k+1} - y^k) \rangle &= \frac{\sigma}{2} (\|A^*x + B^*y^k - c\|^2 - \|A^*x + B^*y^{k+1} - c\|^2) \\ &\quad + \frac{\sigma}{2} (\|A^*x^{k+1} + B^*y^{k+1} - c\|^2 - \|A^*x^{k+1} + B^*y^k - c\|^2). \end{aligned}$$

Substituting this into (48) and using the definition of  $\tilde{z}^{k+1}$ , we have

$$\begin{aligned} \Delta^k(x, y, z) &= \langle z - \tilde{z}^{k+1}, -(A^*x + B^*y - c) \rangle + \frac{1}{\sigma} \langle z - \tilde{z}^{k+1}, \tilde{z}^{k+1} - z^k \rangle + \frac{1}{2\sigma} \|z^k - \tilde{z}^{k+1}\|^2 \\ &\quad - \frac{\sigma}{2} \|A^*x^{k+1} + B^*y^k - c\|^2 + \frac{\sigma}{2} (\|A^*x + B^*y^k - c\|^2 - \|A^*x + B^*y^{k+1} - c\|^2). \end{aligned} \quad (50)$$

Applying (26) in Lemma 3.2 to (50), we get

$$\begin{aligned} \Delta^k(x, y, z) &= \langle z - \tilde{z}^{k+1}, -(A^*x + B^*y - c) \rangle + \frac{1}{2\tau\sigma} (\|z^k - z\|^2 - \|z^{k+1} - z\|^2) + \frac{\tau - 1}{2\sigma} \|z^k - \tilde{z}^{k+1}\|^2 \\ &\quad - \frac{\sigma}{2} \|A^*x^{k+1} + B^*y^k - c\|^2 + \frac{\sigma}{2} (\|A^*x + B^*y^k - c\|^2 - \|A^*x + B^*y^{k+1} - c\|^2). \end{aligned} \quad (51)$$

Using the second elementary identity in (25), we obtain that

$$\begin{aligned} &\langle x - x^{k+1}, (\widehat{\Sigma}_f + S)(x^{k+1} - x^k) \rangle + \langle y - y^{k+1}, (\widehat{\Sigma}_g + T)(y^{k+1} - y^k) \rangle \\ &= \frac{1}{2} (\|x^k - x\|_{\widehat{\Sigma}_f + S}^2 - \|x^{k+1} - x\|_{\widehat{\Sigma}_f + S}^2) - \frac{1}{2} \|x^{k+1} - x^k\|_{\widehat{\Sigma}_f + S}^2 \\ &\quad + \frac{1}{2} (\|y^k - y\|_{\widehat{\Sigma}_g + T}^2 - \|y^{k+1} - y\|_{\widehat{\Sigma}_g + T}^2) - \frac{1}{2} \|y^{k+1} - y^k\|_{\widehat{\Sigma}_g + T}^2. \end{aligned}$$

Substituting this and (51) into (47), and using the definitions of  $\tilde{z}^{k+1}$  and  $r^{k+1}$ , we have

$$\begin{aligned} &(p(x) + q(y)) - (p(x^{k+1}) + q(y^{k+1})) + \langle x - x^{k+1}, \nabla f(x) + Az \rangle + \langle y - y^{k+1}, \nabla g(y) + Bz \rangle \\ &\quad + \langle z - \tilde{z}^{k+1}, -(A^*x + B^*y - c) \rangle + \frac{1}{2\tau\sigma} (\|z^k - z\|^2 - \|z^{k+1} - z\|^2) \\ &\quad + \frac{1}{2} (\|x^k - x\|_{\widehat{\Sigma}_f + S}^2 - \|x^{k+1} - x\|_{\widehat{\Sigma}_f + S}^2) + \frac{\sigma}{2} (\|A^*x + B^*y^k - c\|^2 - \|A^*x + B^*y^{k+1} - c\|^2) \\ &\quad + \frac{1}{2} (\|y^k - y\|_{\widehat{\Sigma}_g + T}^2 - \|y^{k+1} - y\|_{\widehat{\Sigma}_g + T}^2) \\ &\geq \frac{1}{2} \left( (1 - \tau)\sigma \|r^{k+1}\|^2 + \sigma \|A^*x^{k+1} + B^*y^k - c\|^2 + \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S}^2 + \|y^{k+1} - y^k\|_{\frac{1}{2}\Sigma_g + T}^2 \right). \end{aligned} \quad (52)$$

Now we can get (38) from (52) immediately by using the notation in (21) and (22). So Part (a) is proved.

To prove Part (b), assume that  $\frac{1}{2}\widehat{\Sigma}_g + T \succeq 0$ . Using the definition of  $r^k$  and the Cauchy-Schwarz inequality, we get

$$\begin{aligned}
\sigma \|A^* x^{k+1} + B^* y^k - c\|^2 &= \sigma \|r^k + A^*(x^{k+1} - x^k)\|^2 \\
&= \sigma \|r^k\|^2 + \sigma \|A^*(x^{k+1} - x^k)\|^2 + 2\sigma \langle A^*(x^{k+1} - x^k), r^k \rangle \\
&\geq (1-2)\sigma \|r^k\|^2 + (1-\frac{1}{2})\sigma \|A^*(x^{k+1} - x^k)\|^2 \\
&= -\sigma \|r^k\|^2 + \frac{1}{2}\sigma \|A^*(x^{k+1} - x^k)\|^2.
\end{aligned}$$

By using the definition of  $s_{k+1}$ ,  $r^{k+1} = (\tau\sigma)^{-1}(z^{k+1} - z^k)$  and the above formula, for any  $\alpha \in (0, 1]$ , we get

$$\begin{aligned}
&(1-\alpha)[s_{k+1} + (1-\tau)\sigma \|r^{k+1}\|^2 + \sigma \|A^* x^{k+1} + B^* y^k - c\|^2] \\
&\geq -(1-\alpha)\tau\sigma \|r^{k+1}\|^2 + (1-\alpha)\|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S + \frac{1}{2}\sigma AA^*}^2 + (1-\alpha)\|y^{k+1} - y^k\|_{\frac{1}{2}\Sigma_g + T}^2 \\
&\quad + (1-\alpha)\sigma(\|r^{k+1}\|^2 - \|r^k\|^2).
\end{aligned} \tag{53}$$

By using the definition of  $s_{k+1}$  and (31) in Lemma 3.3, for any  $\alpha \in (0, 1]$ , we have

$$\begin{aligned}
&\alpha[s_{k+1} + (1-\tau)\sigma \|r^{k+1}\|^2 + \sigma \|A^* x^{k+1} + B^* y^k - c\|^2] \\
&\geq \alpha\|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S}^2 + \alpha\|y^{k+1} - y^k\|_{\frac{1}{2}\Sigma_g + T}^2 + \max(1-\tau, 1-\tau^{-1})\sigma\alpha(\|r^{k+1}\|^2 - \|r^k\|^2) \\
&\quad + \alpha(\xi_{k+1} - \xi_k) + \min(\tau, 1+\tau-\tau^2)\sigma\alpha(\tau^{-1}\|r^{k+1}\|^2 + \|B^*(y^{k+1} - y^k)\|^2).
\end{aligned} \tag{54}$$

Adding up (53) and (54), we obtain for any  $\alpha \in (0, 1]$  that

$$\begin{aligned}
&s_{k+1} + (1-\tau)\sigma \|r^{k+1}\|^2 + \sigma \|A^* x^{k+1} + B^* y^k - c\|^2 \\
&\geq \|x^{k+1} - x^k\|_{H_f}^2 + \|y^{k+1} - y^k\|_{M_g}^2 + (1-\alpha\min(\tau, \tau^{-1}))\sigma(\|r^{k+1}\|^2 - \|r^k\|^2) \\
&\quad + \alpha(\xi_{k+1} - \xi_k) + (-\tau + \alpha\min(1+\tau, 1+\tau^{-1}))\sigma\|r^{k+1}\|^2.
\end{aligned} \tag{55}$$

Using the notation in (21) and (22), we know from (52) and (55) that (39) holds. The proof is completed.  $\square$

*Remark 3.2.* Suppose that  $B$  is vacuous,  $q \equiv 0$  and  $g \equiv 0$ . Then for any  $\tau \in (0, +\infty)$  and  $k \geq 0$ , we have  $y^{k+1} = y^0 = \bar{y}$ . Since  $B$  is vacuous, by using the definition of  $r^{k+1}$ , we have

$$(1-\tau)\sigma \|r^{k+1}\|^2 + \sigma \|A^* x^{k+1} + B^* y^k - c\|^2 = (2-\tau)\sigma \|r^{k+1}\|^2.$$

By observing that the terms concerning  $y$  in (52) cancel out, we can easily from (52) and the above equation to get

$$\begin{aligned}
&p(x^{k+1}) - p(x) + \langle x^{k+1} - x, \nabla f(x) + Az \rangle + \langle \tilde{z}^{k+1} - z, -(A^* x - c) \rangle \\
&\quad + \frac{1}{2\tau\sigma}(\|z^{k+1} - z\|^2 - \|z^k - z\|^2) + \frac{1}{2}(\|x^{k+1} - x\|_{\widehat{\Sigma}_f + S}^2 - \|x^k - x\|_{\widehat{\Sigma}_f + S}^2) \\
&\leq -\frac{1}{2}\left((2-\tau)\sigma \|r^{k+1}\|^2 + \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S}^2\right).
\end{aligned} \tag{56}$$

Similarly as for (53), for any  $\alpha \in (0, 1]$ , we have

$$\begin{aligned} & (1 - \alpha) \left[ (2 - \tau) \sigma \|r^{k+1}\|^2 + \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S}^2 \right] \\ & \geq - (1 - \alpha) \tau \sigma \|r^{k+1}\|^2 + (1 - \alpha) \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S + \frac{1}{2}\sigma AA^*}^2 + (1 - \alpha) \sigma (\|r^{k+1}\|^2 - \|r^k\|^2). \end{aligned}$$

Adding  $\alpha \left[ (2 - \tau) \sigma \|r^{k+1}\|^2 + \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S}^2 \right]$  to both sides of the above inequality, we get

$$\begin{aligned} & (2 - \tau) \sigma \|r^{k+1}\|^2 + \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S}^2 \\ & \geq \|x^{k+1} - x^k\|_{H_f}^2 + (1 - \alpha) \sigma (\|r^{k+1}\|^2 - \|r^k\|^2) + (2\alpha - \tau) \sigma \|r^{k+1}\|^2. \end{aligned}$$

Substituting this into (56), we obtain

$$\begin{aligned} & p(x^{k+1}) - p(x) + \langle x^{k+1} - x, \nabla f(x) + Az \rangle + \langle z^{k+1} - z, -(A^*x - c) \rangle + \frac{1}{2} \left\{ [(\tau\sigma)^{-1} \|z^{k+1} - z\|^2 \right. \\ & \quad \left. + \|x^{k+1} - x\|_{\frac{1}{2}\Sigma_f + S}^2 + (1 - \alpha) \sigma \|r^{k+1}\|^2 \right] - [(\tau\sigma)^{-1} \|z^k - z\|^2 + \|x^k - x\|_{\frac{1}{2}\Sigma_f + S}^2 + (1 - \alpha) \sigma \|r^k\|^2] \Big\} \\ & \leq -\frac{1}{2} \left[ \|x^{k+1} - x^k\|_{H_f}^2 + (2\alpha - \tau) \sigma \|r^{k+1}\|^2 \right]. \end{aligned} \quad (57)$$

## 4 Convergence analysis

In this section, we analyze the convergence for the Majorized iPADMM for solving problem (1). We first prove its global convergence and then give some choices of proximal terms.

### 4.1 The global convergence

Now we are ready to establish the convergence results for the Majorized iPADMM for solving (1).

**Theorem 4.1.** *Suppose that Assumptions 2.1 and 2.2 hold. Let  $H_f$  and  $M_g$  be defined by (20). Let  $\{(x^k, y^k, z^k)\}$  be generated by the Majorized iPADMM. For each  $k$  and  $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ , let  $\phi_k(x, y, z)$ ,  $\xi_{k+1}$ ,  $s_{k+1}$ ,  $t_{k+1}$ ,  $r^k$  and  $\tilde{z}^{k+1}$  be defined as in (21), (22) and (23). For  $k = 0, 1, \dots$ , denote*

$$\bar{\phi}_k := \phi_k(\bar{x}, \bar{y}, \bar{z}) = (\tau\sigma)^{-1} \|z^k - \bar{z}\|^2 + \|x^k - \bar{x}\|_{\frac{1}{2}\Sigma_f + S}^2 + \|y^k - \bar{y}\|_{\frac{1}{2}\Sigma_g + T + \sigma BB^*}^2, \quad (58)$$

where  $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W}^*$ . Then the following results hold:

(a) For any  $\eta \in (0, 1/2)$  and  $k \geq 0$ , we have

$$\begin{aligned} & \left( \bar{\phi}_k + \beta \sigma \|r^k\|^2 \right) - \left( \bar{\phi}_{k+1} + \beta \sigma \|r^{k+1}\|^2 \right) \\ & \geq \left( \frac{1}{\tau^2 \sigma} \|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S + \eta \sigma AA^*}^2 + \|y^{k+1} - y^k\|_{\frac{1}{2}\Sigma_g + T + \eta \sigma BB^*}^2 \right) \\ & \quad - \left( \frac{\tau + \beta}{\tau^2 \sigma} \|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f - \frac{1}{2}\Sigma_f}^2 + \|y^{k+1} - y^k\|_{\frac{1}{2}\Sigma_g - \frac{1}{2}\Sigma_g + \eta \sigma BB^*}^2 \right), \end{aligned} \quad (59)$$

where

$$\beta := \frac{\eta(1 - \eta)}{1 - 2\eta}. \quad (60)$$

In addition, assume that for some  $\eta \in (0, 1/2)$ ,

$$\widehat{\Sigma}_f + S + \eta\sigma AA^* \succ 0 \quad \text{and} \quad \widehat{\Sigma}_g + T + \eta\sigma BB^* \succ 0 \quad (61)$$

and the following condition holds:

$$\sum_{k=0}^{\infty} (\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 + \|y^{k+1} - y^k\|_{\widehat{\Sigma}_g + \sigma BB^*}^2 + \|r^{k+1}\|^2) < +\infty. \quad (62)$$

Then the sequence  $\{(x^k, y^k)\}$  converges to an optimal solution of problem (1) and  $\{z^k\}$  converges to an optimal solution of the dual of problem (1).

(b) Assume it holds that

$$\frac{1}{2}\widehat{\Sigma}_g + T \succeq 0. \quad (63)$$

Then, for any  $\alpha \in (0, 1]$  and  $k \geq 1$ , we have

$$\begin{aligned} & \left[ \bar{\phi}_k + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^k\|^2 + \alpha\xi_k \right] - \left[ \bar{\phi}_{k+1} + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^{k+1}\|^2 + \alpha\xi_{k+1} \right] \\ & \geq t_{k+1} + (-\tau + \alpha \min(1 + \tau, 1 + \tau^{-1}))\sigma\|r^{k+1}\|^2. \end{aligned} \quad (64)$$

In addition, assume that  $\tau \in (0, (1 + \sqrt{5})/2)$  and for some  $\alpha \in (\tau/\min(1 + \tau, 1 + \tau^{-1}), 1]$ ,

$$\widehat{\Sigma}_f + S \succeq 0, \quad H_f \succeq 0, \quad \frac{1}{2}\Sigma_f + S + \sigma AA^* \succ 0 \quad \text{and} \quad M_g \succ 0. \quad (65)$$

Then, the sequence  $\{(x^k, y^k)\}$  converges to an optimal solution of problem (1) and  $\{z^k\}$  converges to an optimal solution of the dual of problem (1).

*Proof.* Note that  $A^*\bar{x} + B^*\bar{y} - c = 0$ . Then we have

$$\begin{aligned} \bar{\phi}_k := \phi_k(\bar{x}, \bar{y}, \bar{z}) &= (\tau\sigma)^{-1}\|z^k - \bar{z}\|^2 + \|x^k - \bar{x}\|_{\widehat{\Sigma}_f + S}^2 + \|y^k - \bar{y}\|_{\widehat{\Sigma}_g + T}^2 + \sigma\|A^*\bar{x} + B^*y^k - c\|^2 \\ &= (\tau\sigma)^{-1}\|z^k - \bar{z}\|^2 + \|x^k - \bar{x}\|_{\widehat{\Sigma}_f + S}^2 + \|y^k - \bar{y}\|_{\widehat{\Sigma}_g + T + \sigma BB^*}^2. \end{aligned}$$

Recall that for any  $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W}^*$ , we have

$$\begin{aligned} & (p(x^{k+1}) + q(y^{k+1})) - (p(\bar{x}) + q(\bar{y})) + \langle x^{k+1} - \bar{x}, \nabla f(\bar{x}) + A\bar{z} \rangle + \langle y^{k+1} - \bar{y}, \nabla g(\bar{y}) + B\bar{z} \rangle \\ & + \langle \bar{z}^{k+1} - \bar{z}, -(A^*\bar{x} + B^*\bar{y} - c) \rangle \geq 0. \end{aligned} \quad (66)$$

In the following, we will consider Part (a) and Part (b) separately.

**Proof of Part (a).** Setting  $(x, y, z) = (\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W}^*$  in (38) and using the relation (66), we get

$$\bar{\phi}_k - \bar{\phi}_{k+1} \geq s_{k+1} + (1 - \tau)\sigma\|r^{k+1}\|^2 + \sigma\|A^*x^{k+1} + B^*y^k - c\|^2. \quad (67)$$



Using the definition of  $r^k$  and the Cauchy-Schwarz inequality, for a given  $\beta > 0$  defined by (60), we get

$$\begin{aligned}
\sigma \|A^* x^{k+1} + B^* y^k - c\|^2 &= \sigma \|r^k + A^*(x^{k+1} - x^k)\|^2 \\
&= \sigma \|r^k\|^2 + \sigma \|A^*(x^{k+1} - x^k)\|^2 + 2\sigma \langle A^*(x^{k+1} - x^k), r^k \rangle \\
&\geq [1 - (1 + \beta)] \sigma \|r^k\|^2 + \left(1 - \frac{1}{1 + \beta}\right) \sigma \|A^*(x^{k+1} - x^k)\|^2 \\
&= -\beta \sigma \|r^k\|^2 + \frac{\beta}{1 + \beta} \sigma \|A^*(x^{k+1} - x^k)\|^2.
\end{aligned}$$

By using the definition of  $s_{k+1}$ ,  $r^{k+1} = (\tau\sigma)^{-1}(z^{k+1} - z^k)$  and the above formula, we obtain

$$\begin{aligned}
&s_{k+1} + (1 - \tau)\sigma \|r^{k+1}\|^2 + \sigma \|A^* x^{k+1} + B^* y^k - c\|^2 \\
&\geq \frac{1 - \tau - \beta}{\tau^2 \sigma} \|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S + \frac{\beta}{1+\beta}\sigma AA^*}^2 + \|y^{k+1} - y^k\|_{\frac{1}{2}\Sigma_g + T}^2 \\
&\quad + \beta \sigma (\|r^{k+1}\|^2 - \|r^k\|^2).
\end{aligned} \tag{68}$$

Recall that

$$\frac{\beta}{1 + \beta} = \frac{\frac{\eta(1-\eta)}{1-2\eta}}{1 + \frac{\eta(1-\eta)}{1-2\eta}} = \eta \frac{1 - \eta}{1 - \eta - \eta^2} > \eta.$$

By simple manipulations, we get

$$\begin{aligned}
&\frac{1 - \tau - \beta}{\tau^2 \sigma} \|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S + \frac{\beta}{1+\beta}\sigma AA^*}^2 + \|y^{k+1} - y^k\|_{\frac{1}{2}\Sigma_g + T}^2 \\
&\geq \left( \frac{1}{\tau^2 \sigma} \|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S + \eta\sigma AA^*}^2 + \|y^{k+1} - y^k\|_{\frac{1}{2}\Sigma_g + T + \eta\sigma BB^*}^2 \right) \\
&\quad - \left( \frac{\tau + \beta}{\tau^2 \sigma} \|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f - \frac{1}{2}\Sigma_f}^2 + \|y^{k+1} - y^k\|_{\frac{1}{2}\Sigma_g - \frac{1}{2}\Sigma_g + \eta\sigma BB^*}^2 \right).
\end{aligned}$$

Substituting this and (68) into (67), we get (59).

Now assume that (61) and (62) hold. For any given  $\eta \in (0, 1/2)$ , using the definitions of  $\bar{\phi}_{k+1}$ ,  $r^{k+1}$  and  $\beta$ ,  $A^* \bar{x} + B^* \bar{y} = c$  and the Cauchy-Schwarz inequality, we get

$$\begin{aligned}
\bar{\phi}_{k+1} + \beta \sigma \|r^{k+1}\|^2 &= (\tau\sigma)^{-1} \|z^{k+1} - \bar{z}\|^2 + \|x^{k+1} - \bar{x}\|_{\frac{1}{2}\Sigma_f + S}^2 + \|y^{k+1} - \bar{y}\|_{\frac{1}{2}\Sigma_g + T + \sigma BB^*}^2 \\
&\quad + \beta \sigma \|A^*(x^{k+1} - \bar{x}) + B^*(y^{k+1} - \bar{y})\|^2 \\
&\geq (\tau\sigma)^{-1} \|z^{k+1} - \bar{z}\|^2 + \|x^{k+1} - \bar{x}\|_{\frac{1}{2}\Sigma_f + S}^2 + \|y^{k+1} - \bar{y}\|_{\frac{1}{2}\Sigma_g + T + \sigma BB^*}^2 \\
&\quad + \beta \left(1 - \frac{\eta}{1 - \eta}\right) \|x^{k+1} - \bar{x}\|_{\sigma AA^*}^2 + \beta \left(1 - \frac{1 - \eta}{\eta}\right) \|y^{k+1} - \bar{y}\|_{\sigma BB^*}^2 \\
&= (\tau\sigma)^{-1} \|z^{k+1} - \bar{z}\|^2 + \|x^{k+1} - \bar{x}\|_{\frac{1}{2}\Sigma_f + S + \eta\sigma AA^*}^2 + \|y^{k+1} - \bar{y}\|_{\frac{1}{2}\Sigma_g + T + \eta\sigma BB^*}^2.
\end{aligned} \tag{69}$$

Set

$$\zeta_k := \frac{\tau + \beta}{\tau^2 \sigma} \|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f - \frac{1}{2}\Sigma_f}^2 + \|y^{k+1} - y^k\|_{\frac{1}{2}\Sigma_g - \frac{1}{2}\Sigma_g + \eta\sigma BB^*}^2$$

and

$$v_k := \frac{1}{\tau^2 \sigma} \|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|_{\widehat{\Sigma}_f + S + \eta \sigma AA^*}^2 + \|y^{k+1} - y^k\|_{\widehat{\Sigma}_g + T + \eta \sigma BB^*}^2.$$

From (62) and  $z^{k+1} = z^k + \tau \sigma r^{k+1}$ , we get  $\sum_{k=0}^{\infty} \zeta_k < +\infty$ . Since for some  $\eta \in (0, 1/2)$

$$\widehat{\Sigma}_f + S + \eta \sigma AA^* \succ 0 \quad \text{and} \quad \widehat{\Sigma}_g + T + \eta \sigma BB^* \succ 0,$$

it follows from (69) and (59) that

$$0 \leq \bar{\phi}_{k+1} + \beta \sigma \|r^{k+1}\|^2 \leq \bar{\phi}_k + \beta \sigma \|r^k\|^2 + \zeta_k \leq \bar{\phi}_0 + \beta \sigma \|r^0\|^2 + \sum_{j=0}^k \zeta_j. \quad (70)$$

Thus the sequence  $\{\bar{\phi}_{k+1} + \beta \sigma \|r^{k+1}\|^2\}$  is bounded. From (69), we see that the three sequences  $\{\|z^{k+1} - \bar{z}\|\}$ ,  $\{\|x^{k+1} - \bar{x}\|_{\widehat{\Sigma}_f + S + \eta \sigma AA^*}\}$  and  $\{\|y^{k+1} - \bar{y}\|_{\widehat{\Sigma}_g + T + \eta \sigma BB^*}\}$  are all bounded. Since  $\widehat{\Sigma}_f + S + \eta \sigma AA^* \succ 0$  and  $\widehat{\Sigma}_g + T + \eta \sigma BB^* \succ 0$ , the sequence  $\{(x^k, y^k, z^k)\}$  is also bounded. Using (59), we get for any  $k \geq 0$ ,

$$v_k \leq (\bar{\phi}_k + \beta \sigma \|r^k\|^2) - (\bar{\phi}_{k+1} + \beta \sigma \|r^{k+1}\|^2) + \zeta_k,$$

and hence

$$\sum_{j=0}^k v_j \leq (\bar{\phi}_0 + \beta \sigma \|r^0\|^2) - (\bar{\phi}_{k+1} + \beta \sigma \|r^{k+1}\|^2) + \sum_{j=0}^k \zeta_j < +\infty.$$

Again, since  $\widehat{\Sigma}_f + S + \eta \sigma AA^* \succ 0$  and  $\widehat{\Sigma}_g + T + \eta \sigma BB^* \succ 0$ , from the definition of  $v_k$ , we get

$$\lim_{k \rightarrow \infty} \|r^{k+1}\| = \lim_{k \rightarrow \infty} (\tau \sigma)^{-1} \|z^{k+1} - z^k\| = 0, \quad (71)$$

$$\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} \|y^{k+1} - y^k\| = 0. \quad (72)$$

Recall that the sequence  $\{(x^k, y^k, z^k)\}$  is bounded. There is a subsequence  $\{(x^{k_i}, y^{k_i}, z^{k_i})\}$  which converges to a cluster point, say  $(x^\infty, y^\infty, z^\infty)$ . We next show that  $(x^\infty, y^\infty)$  is an optimal solution to problem (1) and  $z^\infty$  is a corresponding Lagrange multiplier.

Taking limits on both sides of (42) and (43) along the subsequence  $\{(x^{k_i}, y^{k_i}, z^{k_i})\}$ , using (71) and (72), we obtain that

$$\begin{cases} (p(x) + f(x)) - (p(x^\infty) + f(x^\infty)) + \langle x - x^\infty, Az^\infty \rangle \geq 0, \\ (q(y) + g(y)) - (q(y^\infty) + g(y^\infty)) + \langle y - y^\infty, Bz^\infty \rangle \geq 0, \\ c - A^*x^\infty - B^*y^\infty = 0, \end{cases}$$

i.e.,  $(x^\infty, y^\infty, z^\infty)$  satisfies (13). Hence  $(x^\infty, y^\infty)$  is an optimal solution to problem (1) and  $z^\infty$  is a corresponding Lagrange multiplier.

To complete the proof of Part (a), we show that  $(x^\infty, y^\infty, z^\infty)$  is actually the unique limit of  $\{(x^k, y^k, z^k)\}$ . Replacing  $(\bar{x}, \bar{y}, \bar{z})$  by  $(x^\infty, y^\infty, z^\infty)$  in (70), for any  $k \geq k_i$ , we have

$$\phi_{k+1}(x^\infty, y^\infty, z^\infty) + \beta \sigma \|r^{k+1}\|^2 \leq \phi_{k_i}(x^\infty, y^\infty, z^\infty) + \beta \sigma \|r^{k_i}\|^2 + \sum_{j=k_i}^k \zeta_j. \quad (73)$$

Note that

$$\lim_{i \rightarrow \infty} \phi_{k_i}(x^\infty, y^\infty, z^\infty) + \beta \sigma \|r^{k_i}\|^2 = 0 \quad \text{and} \quad \sum_{j=0}^{\infty} \zeta_j < +\infty.$$

Therefore, we get

$$\lim_{k \rightarrow \infty} \phi_{k+1}(x^\infty, y^\infty, z^\infty) + \beta \sigma \|r^{k+1}\|^2 = 0.$$

Then from (69), we obtain

$$\lim_{k \rightarrow \infty} ((\tau\sigma)^{-1} \|z^{k+1} - z^\infty\|^2 + \|x^{k+1} - x^\infty\|_{\widehat{\Sigma}_f + S + \eta\sigma AA^*}^2 + \|y^{k+1} - y^\infty\|_{\widehat{\Sigma}_g + T + \eta\sigma BB^*}^2) = 0.$$

Since  $\widehat{\Sigma}_f + S + \eta\sigma AA^* \succ 0$  and  $\widehat{\Sigma}_g + T + \eta\sigma BB^* \succ 0$ , we also have that  $\lim_{k \rightarrow \infty} x^k = x^\infty$  and  $\lim_{k \rightarrow \infty} y^k = y^\infty$ . Therefore, we have shown that the whole sequence  $\{(x^k, y^k, z^k)\}$  converges to  $(x^\infty, y^\infty, z^\infty)$  for any  $\tau \in (0, +\infty)$ .

**Proof of Part (b).** By using (66), we can get (64) from (39) immediately. Assume that  $\tau \in (0, (1 + \sqrt{5})/2)$  and  $\alpha \in (\tau/\min(1 + \tau, 1 + \tau^{-1}), 1]$ . Then, we have  $1 - \alpha \min(\tau, \tau^{-1}) > 0$  and  $-\tau + \alpha \min(1 + \tau, 1 + \tau^{-1}) > 0$ . Since

$$\widehat{\Sigma}_g \succeq \Sigma_g, \quad M_g = \frac{1}{2}\Sigma_g + T + \min(\tau, 1 + \tau - \tau^2)\sigma\alpha BB^* \succ 0 \quad \text{and} \quad \min(\tau, 1 + \tau - \tau^2) \leq 1,$$

we have

$$\widehat{\Sigma}_g + T \succeq \frac{1}{2}\widehat{\Sigma}_g + T \succeq 0 \quad \text{and} \quad \widehat{\Sigma}_g + T + \sigma BB^* \succeq \frac{1}{2}\Sigma_g + T + \min(\tau, 1 + \tau - \tau^2)\sigma\alpha BB^* \succ 0.$$

Note that  $H_f \succeq 0$  and  $M_g \succ 0$ . Then we obtain that  $\bar{\phi}_{k+1} \geq 0, t_{k+1} \geq 0, \xi_{k+1} \geq 0$ . From (27) and (64), we see immediately that the sequence  $\{\bar{\phi}_{k+1} + \xi_{k+1}\}$  is bounded,

$$\lim_{k \rightarrow \infty} t_{k+1} = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} \|z^{k+1} - z^k\| = \lim_{k \rightarrow \infty} \tau\sigma \|r^{k+1}\| = 0, \quad (74)$$

which, together with (22) and (65), imply that

$$\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\|_{H_f} = 0, \quad \lim_{k \rightarrow \infty} \|y^{k+1} - y^k\| = 0 \quad (75)$$

and

$$\|A^*(x^{k+1} - x^k)\| \leq \|r^{k+1}\| + \|r^k\| + \|B^*(y^{k+1} - y^k)\| \rightarrow 0, \quad k \rightarrow \infty. \quad (76)$$

Thus, from (75) and (76) we obtain that

$$\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S + \sigma AA^*}^2 = \lim_{k \rightarrow \infty} (\|x^{k+1} - x^k\|_{H_f}^2 + \frac{1}{2}(1 + \alpha)\sigma \|A^*(x^{k+1} - x^k)\|^2) = 0.$$

Since  $\frac{1}{2}\Sigma_f + S + \sigma AA^* \succ 0$ , we also get

$$\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0. \quad (77)$$

By the definition of  $\bar{\phi}_{k+1}$ , we see that the three sequences  $\{\|z^{k+1} - \bar{z}\|\}$ ,  $\{\|x^{k+1} - \bar{x}\|_{\widehat{\Sigma}_f + S}\}$  and  $\{\|y^{k+1} - \bar{y}\|_{\widehat{\Sigma}_g + T + \sigma BB^*}\}$  are all bounded. Since  $\widehat{\Sigma}_g + T + \sigma BB^* \succ 0$ , the sequence  $\{\|y^{k+1}\|\}$  is bounded. Note that  $A^*\bar{x} + B^*\bar{y} = c$ . Furthermore, by using

$$\begin{aligned} \|A^*(x^{k+1} - \bar{x})\| &\leq \|A^*x^{k+1} + B^*y^{k+1} - (A^*\bar{x} + B^*\bar{y})\| + \|B^*(y^{k+1} - \bar{y})\| \\ &= \|r^{k+1}\| + \|B^*(y^{k+1} - \bar{y})\|, \end{aligned}$$

we also know that the sequence  $\{\|A^*(x^{k+1} - \bar{x})\|\}$  is bounded, and so is the sequence  $\{\|x^{k+1} - \bar{x}\|_{\widehat{\Sigma}_f + S + \sigma AA^*}\}$ . This shows that the sequence  $\{\|x^{k+1}\|\}$  is also bounded since  $\widehat{\Sigma}_f + S + \sigma AA^* \succeq \frac{1}{2}\widehat{\Sigma}_f + S + \sigma AA^* \succ 0$ . Thus, the sequence  $\{(x^k, y^k, z^k)\}$  is bounded.

Since the sequence  $\{(x^k, y^k, z^k)\}$  is bounded, there is a subsequence  $\{(x^{k_i}, y^{k_i}, z^{k_i})\}$  which converges to a cluster point, say  $(x^\infty, y^\infty, z^\infty)$ . We next show that  $(x^\infty, y^\infty)$  is an optimal solution to problem (1) and  $z^\infty$  is a corresponding Lagrange multiplier.

Taking limits on both sides of (42) and (43) along the subsequence  $\{(x^{k_i}, y^{k_i}, z^{k_i})\}$ , using (74), (75) and (77), we obtain that

$$\begin{cases} (p(x) + f(x)) - (p(x^\infty) + f(x^\infty)) + \langle x - x^\infty, Az^\infty \rangle \geq 0, \\ (q(y) + g(y)) - (q(y^\infty) + g(y^\infty)) + \langle y - y^\infty, Bz^\infty \rangle \geq 0, \\ c - A^*x^\infty - B^*y^\infty = 0, \end{cases}$$

i.e.,  $(x^\infty, y^\infty, z^\infty)$  satisfies (13). Thus  $(x^\infty, y^\infty)$  is an optimal solution to problem (1) and  $z^\infty$  is a corresponding Lagrange multiplier.

To complete the proof of Part (b), we show that  $(x^\infty, y^\infty, z^\infty)$  is actually the unique limit of  $\{(x^k, y^k, z^k)\}$ . As in the proof of (73) in Part (a), we can apply the inequality (64) with  $(\bar{x}, \bar{y}, \bar{z}) = (x^\infty, y^\infty, z^\infty)$  to show that

$$\lim_{k \rightarrow \infty} \phi_{k+1}(x^\infty, y^\infty, z^\infty) = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} \|\xi_{k+1}\| = 0.$$

Hence

$$\lim_{k \rightarrow \infty} ((\tau\sigma)^{-1}\|z^{k+1} - z^\infty\|^2 + \|x^{k+1} - x^\infty\|_{\widehat{\Sigma}_f + S}^2 + \|y^{k+1} - y^\infty\|_{\widehat{\Sigma}_g + T + \sigma BB^*}^2) = 0,$$

$$\begin{aligned} \|A^*(x^{k+1} - x^\infty)\| &\leq \|A^*x^{k+1} + B^*y^{k+1} - (A^*x^\infty + B^*y^\infty)\| + \|B^*(y^{k+1} - y^\infty)\| \\ &= \|r^{k+1}\| + \|B^*(y^{k+1} - y^\infty)\| \rightarrow 0, \quad k \rightarrow \infty \end{aligned}$$

and

$$\lim_{k \rightarrow \infty} \|x^{k+1} - x^\infty\|_{\widehat{\Sigma}_f + S + \sigma AA^*}^2 = 0.$$

Using that fact that  $\widehat{\Sigma}_f + S + \sigma AA^*$  and  $\widehat{\Sigma}_g + T + \sigma BB^*$  are both positive definite, we have  $\lim_{k \rightarrow \infty} x^k = x^\infty$  and  $\lim_{k \rightarrow \infty} y^k = y^\infty$ . Therefore, we have shown that the whole sequence  $\{(x^k, y^k, z^k)\}$  converges to  $(x^\infty, y^\infty, z^\infty)$  if  $\tau \in (0, (1 + \sqrt{5})/2)$ . The proof is completed.  $\square$

*Remark 4.3.* In practice, Part (a) of Theorem 4.1 can be applied in a more heuristic way by using any sufficient condition to guarantee (62) holds. If this sufficient condition does not hold, then one

can just use the conditions in Part (b). The conditions on  $S$  and  $T$  in Part (b) for the case that  $\tau = 1$  can be written as for some  $\alpha \in (1/2, 1]$ ,

$$\widehat{\Sigma}_f + S \succeq 0, \quad \frac{1}{2}\Sigma_f + S + \frac{1}{2}(1 - \alpha)\sigma AA^* \succeq 0, \quad \frac{1}{2}\Sigma_f + S + \sigma AA^* \succ 0$$

and

$$\frac{1}{2}\widehat{\Sigma}_g + T \succeq 0, \quad \frac{1}{2}\Sigma_g + T + \alpha\sigma BB^* \succ 0;$$

and these conditions for the case that  $\tau = 1.618$  can be replaced by, for some  $\alpha \in [0.99998, 1]$ ,

$$\widehat{\Sigma}_f + S \succeq 0, \quad \frac{1}{2}\Sigma_f + S + \frac{1}{2}(1 - \alpha)\sigma AA^* \succeq 0, \quad \frac{1}{2}\Sigma_f + S + \sigma AA^* \succ 0$$

and

$$\frac{1}{2}\widehat{\Sigma}_g + T \succeq 0, \quad \frac{1}{2}\Sigma_g + T + 0.000075\alpha\sigma BB^* \succ 0.$$

*Remark 4.4.* Suppose that  $B$  is vacuous,  $q \equiv 0$  and  $g \equiv 0$ . Then for any  $\tau \in (0, +\infty)$  and  $k \geq 0$ , we have  $y^{k+1} = y^0 = \bar{y}$ . Similarly as in Part (b) of Theorem 4.1, using (57) we have for any  $\alpha \in (0, 1]$  and  $k \geq 0$  that

$$\begin{aligned} & \left\{ (\tau\sigma)^{-1} \|z^k - \bar{z}\|^2 + \|x^k - \bar{x}\|_{\widehat{\Sigma}_f + S}^2 + (1 - \alpha)\sigma \|r^k\|^2 \right\} \\ & - \left\{ (\tau\sigma)^{-1} \|z^{k+1} - \bar{z}\|^2 + \|x^{k+1} - \bar{x}\|_{\widehat{\Sigma}_f + S}^2 + (1 - \alpha)\sigma \|r^{k+1}\|^2 \right\} \\ & \geq \|x^{k+1} - x^k\|_{H_f}^2 + (2\alpha - \tau)\sigma \|r^{k+1}\|^2. \end{aligned}$$

In addition, assume that  $\tau \in (0, 2)$  and for some  $\alpha \in (\tau/2, 1]$ ,

$$\widehat{\Sigma}_f + S \succeq 0, \quad H_f = \frac{1}{2}\Sigma_f + S + \frac{1}{2}(1 - \alpha)\sigma AA^* \succeq 0, \quad \frac{1}{2}\Sigma_f + S + \sigma AA^* \succ 0.$$

Then, the sequence  $\{x^k\}$  converges to an optimal solution of problem (1) and  $\{z^k\}$  converges to an optimal solution of the dual of problem (1).

## 4.2 Choices of proximal terms

Let  $\mathcal{G} : \mathcal{X} \rightarrow \mathcal{X}$  be any given self-adjoint linear operator with  $\dim(\mathcal{X}) = n$ , the dimension of  $\mathcal{X}$ . We shall first introduce a majorization technique to find a self-adjoint positive definite linear operator  $\mathcal{M}$  such that  $\mathcal{M} \succeq \mathcal{G}$  and  $\mathcal{M}^{-1}$  is easy to calculate. Suppose that  $\mathcal{G}$  has the following spectral decomposition

$$\mathcal{G} = \sum_{i=1}^n \lambda_i u_i u_i^*,$$

where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ , with  $\lambda_l > 0$  for some  $1 \leq l \leq n$ , are the eigenvalues of  $\mathcal{G}$  and  $u_i$ ,  $i = 1, \dots, n$  are the corresponding mutually orthogonal unit eigenvectors. Then, for a small  $l$ , we can design a practically useful majorization for  $\mathcal{G}$  as follows:

$$\mathcal{G} \preceq \mathcal{M} := \sum_{i=1}^l \lambda_i u_i u_i^* + \lambda_l \sum_{i=l+1}^n u_i u_i^* = \lambda_l I + \sum_{i=1}^l (\lambda_i - \lambda_l) u_i u_i^*. \quad (78)$$

Note that  $\mathcal{M}^{-1}$  can be easily obtained as follows:

$$\mathcal{M}^{-1} = \sum_{i=1}^l \lambda_i^{-1} u_i u_i^* + \lambda_l^{-1} \sum_{i=l+1}^n u_i u_i^* = \lambda_l^{-1} I + \sum_{i=1}^l (\lambda_i^{-1} - \lambda_l^{-1}) u_i u_i^*. \quad (79)$$

Thus, we only need to compute the first  $l$  eigen-pairs  $(\lambda_i, u_i)$ ,  $i = 1, \dots, l$  of  $\mathcal{G}$  for computing  $\mathcal{M}$  and  $\mathcal{M}^{-1}$ .

In the following, we only discuss how to choose proximal terms for the  $x$ -part. The discussions for the  $y$ -part are similar, and are thus omitted here. It follows from (19) that

$$x^{k+1} = \operatorname{argmin}_{x \in \mathcal{X}} p(x) + \frac{1}{2} \langle x, \mathcal{P}x \rangle + \langle \nabla f(x^k) + \sigma A \hat{r}^k - \mathcal{P}x^k, x \rangle. \quad (80)$$

**Example 4.1:**  $p \not\equiv 0$  and  $A \neq 0$ . Choose  $\alpha \in (\tau / \min(1 + \tau, 1 + \tau^{-1}), 1]$  such that  $2\hat{\Sigma}_f - \Sigma_f \succeq (1 - \alpha)\sigma AA^*$ . Define

$$\rho_0 := \lambda_{\max} \left( \hat{\Sigma}_f - \frac{1}{2} \Sigma_f + \frac{1}{2} (1 + \alpha) \sigma AA^* \right). \quad (81)$$

Note that  $\rho_0 \geq \lambda_{\max} \left( \hat{\Sigma}_f - \frac{1}{2} \Sigma_f \right)$ . Let  $\rho$  be any positive number such that

$$\begin{cases} \rho \geq \rho_0 & \text{if } AA^* \succ 0, \\ \rho \geq \rho_0 \text{ and } \rho > \lambda_{\max} \left( \hat{\Sigma}_f - \frac{1}{2} \Sigma_f \right) & \text{otherwise.} \end{cases} \quad (82)$$

A particular choice which we will consider later in the numerical experiments is:

$$\rho = 1.01 \rho_0. \quad (83)$$

Choose

$$S := -\frac{1}{2} \left[ \Sigma_f + (1 - \alpha) \sigma AA^* \right] + \left[ \rho I - \left( \hat{\Sigma}_f - \frac{1}{2} \Sigma_f + \frac{1}{2} (1 + \alpha) \sigma AA^* \right) \right] = \rho I - \hat{\Sigma}_f - \sigma AA^*, \quad (84)$$

where  $\rho$  is defined in (82). Then,  $S$ , which obviously may be indefinite, satisfies (65), and

$$\mathcal{P} = \hat{\Sigma}_f + S + \sigma AA^* = \rho I \succ 0.$$

One interesting special case is  $\hat{\Sigma}_f = \Sigma_f = Q$  for some self-adjoint linear operator  $Q \succeq 0$ . By taking  $\alpha = 1$  and  $\rho = \frac{1}{2} \lambda_{\max}(Q) + \sigma \lambda_{\max}(AA^*)$ , we have

$$S = \frac{1}{2} \lambda_{\max}(Q) I - Q + \sigma [\lambda_{\max}(AA^*) I - AA^*].$$

**Example 4.2:**  $p \equiv 0$ . Let  $\alpha \in (\tau / \min(1 + \tau, 1 + \tau^{-1}), 1]$  such that  $2\hat{\Sigma}_f - \Sigma_f \succeq (1 - \alpha)\sigma AA^*$ . Choose  $\mathcal{G}$  such that

$$\begin{cases} \mathcal{G} = \hat{\Sigma}_f - \frac{1}{2} \Sigma_f + \frac{1}{2} (1 + \alpha) \sigma AA^* & \text{if } AA^* \succ 0, \\ \mathcal{G} \succeq \hat{\Sigma}_f - \frac{1}{2} \Sigma_f + \frac{1}{2} (1 + \alpha) \sigma AA^* \text{ and } \mathcal{G} \succ \hat{\Sigma}_f - \frac{1}{2} \Sigma_f & \text{otherwise.} \end{cases}$$

Let  $\mathcal{M} \succ 0$  be the majorization of  $\mathcal{G}$  as in (78). Choose

$$S := -\frac{1}{2} \left[ \Sigma_f + (1 - \alpha) \sigma A A^* \right] + \left[ \mathcal{M} - (\widehat{\Sigma}_f - \frac{1}{2} \Sigma_f + \frac{1}{2} (1 + \alpha) \sigma A A^*) \right].$$

Certainly,  $S$ , which may be indefinite, satisfies (65), and

$$\mathcal{P} = S + \widehat{\Sigma}_f + \sigma A A^* = \mathcal{M} \succ 0.$$

By using (78) and (79), one can compute  $\mathcal{P}$  and  $\mathcal{P}^{-1}$  at a low cost if  $l$  is a small integer number, for example  $1 \leq l \leq 6$ . One special case is  $A A^* \succ 0$  and  $\widehat{\Sigma}_f = \Sigma_f = Q$  for some self-adjoint linear operator  $Q \succeq 0$ . By taking  $\alpha = 1$ ,  $\mathcal{G} = \frac{1}{2} Q + \sigma A A^*$ , and  $\mathcal{M}$  to be a majorization of  $\mathcal{G}$  as defined in (78), we have

$$S = \mathcal{M} - (Q + \sigma A A^*).$$

**Example 4.3:**  $p$  can be decomposed into two separate parts  $p(x) := p_1(x_1) + p_2(x_2)$  with  $x := (x_1, x_2)$ . For simplicity, we assume

$$\Sigma_f = Q := \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}^* & Q_{22} \end{pmatrix} \quad \text{and} \quad \widehat{\Sigma}_f = Q + \text{Diag}(D_1, D_2),$$

where  $D_1$  and  $D_2$  are two self-adjoint and positive semidefinite linear operators. Define

$$\mathcal{M} := \text{Diag}(\mathcal{M}_1, \mathcal{M}_2),$$

where

$$\mathcal{M}_1 := D_1 + \frac{1}{2} (Q_{11} + (Q_{12} Q_{12}^*)^{\frac{1}{2}}) + \sigma (A_1 A_1^* + (A_1 A_2^* A_2 A_1^*)^{\frac{1}{2}})$$

and

$$\mathcal{M}_2 := D_2 + \frac{1}{2} (Q_{22} + (Q_{12}^* Q_{12})^{\frac{1}{2}}) + \sigma (A_2 A_2^* + (A_2 A_1^* A_1 A_2^*)^{\frac{1}{2}}).$$

If  $A_1 A_1^* + (A_1 A_2^* A_2 A_1^*)^{\frac{1}{2}} \succ 0$  and  $A_2 A_2^* + (A_2 A_1^* A_1 A_2^*)^{\frac{1}{2}} \succ 0$ , we can choose

$$S := \mathcal{M} - Q - \text{Diag}(D_1, D_2) - \sigma A A^*;$$

otherwise we can add a block diagonal self-adjoint positive definite linear operator to  $S$ . Then one can see that  $S$ , which again may be indefinite, satisfies (65) for  $\alpha = 1$ , by using the fact that for any given linear operator  $X$  from  $\mathcal{X}$  to another finite dimensional real Euclidean space, it holds that

$$\begin{pmatrix} & X \\ X^* & \end{pmatrix} \preceq \begin{pmatrix} (X X^*)^{\frac{1}{2}} & \\ & (X^* X)^{\frac{1}{2}} \end{pmatrix}.$$

Thus, from the block diagonal structure of  $\mathcal{P} = \mathcal{M}$ , we can see that solving the subproblem for  $x$  can be split into solving two separate subproblems for the  $x_1$ -part and  $x_2$ -part, respectively. If the subproblem for either the  $x_1$ - or the  $x_2$ - part is still difficult to solve, one may add a self-adjoint positive semidefinite linear operator to  $\mathcal{M}_1$  or  $\mathcal{M}_2$ , respectively, to make the subproblem easier to solve. We refer the readers to Examples 4.1 and 4.2 for possible choices of such linear operators in different scenarios.

In Examples 4.1-4.3, we list various choices of proximal terms in different situations. Nevertheless, they are far from being exhaustive. For example, if  $p \neq 0$ ,  $x := (x_1, \dots, x_m)$ ,  $p(x) = p_1(x_1)$  and  $f(x)$  is a convex quadratic function, one may construct Schur complement based or more general symmetric Gauss-Seidel based proximal terms to derive convergent ADMMs for solving some interesting multi-block conic optimization problems [23, 24].

## 5 The analysis of iteration-complexity

In this section, we will present the iteration-complexity including non-ergodic and ergodic senses of the Majorized iPADMM.

### 5.1 The non-ergodic iteration-complexity

In this subsection, we will present the non-ergodic iteration-complexity of an  $\varepsilon$ -approximate KKT point for the Majorized iPADMM. For related results, see the work of Davis and Yin [5] on the operator-splitting scheme with separable objective functions and the work of Cui et al. [4] on the majorized ADMM with coupled objective functions.

**Theorem 5.1.** *Assume that Assumptions 2.1 and 2.2 hold. Let  $\{(x^i, y^i, z^i)\}$  be generated by the Majorized iPADMM. Assume that  $\tau \in (0, (1 + \sqrt{5})/2)$  and for some  $\alpha \in (\tau/\min(1 + \tau, 1 + \tau^{-1}), 1]$ ,*

$$\widehat{\Sigma}_f + S \succeq 0, \quad H_f \succ 0, \quad \frac{1}{2}\widehat{\Sigma}_g + T \succeq 0 \quad \text{and} \quad M_g \succ 0,$$

where  $H_f$  and  $M_g$  are defined in (20). Then, we have

$$\min_{1 \leq i \leq k} \left\{ d^2(0, \partial p(x^{i+1}) + \nabla f(x^{i+1}) + Az^{i+1}) + d^2(0, \partial q(y^{i+1}) + \nabla g(y^{i+1}) + Bz^{i+1}) \right. \\ \left. + \|A^*x^{i+1} + B^*y^{i+1} - c\|^2 \right\} = o(1/k) \quad (85)$$

and

$$\min_{1 \leq i \leq k} \left| (p(x^i) + f(x^i) + q(y^i) + g(y^i)) - (p(\bar{x}) + f(\bar{x}) + q(\bar{y}) + g(\bar{y})) \right| = o(1/\sqrt{k}), \quad (86)$$

where  $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$  satisfies (11).

**Proof.** For each  $i$ , let  $\bar{\phi}_i$  be defined by (58) and

$$a_i := \frac{-\tau + \alpha \min(1 + \tau, 1 + \tau^{-1})}{\tau^2 \sigma} \|z^{i+1} - z^i\|^2 + \|x^{i+1} - x^i\|_{H_f}^2 + \|y^{i+1} - y^i\|_{M_g}^2.$$

Since  $\tau \in (0, (1 + \sqrt{5})/2)$ ,  $\alpha \in (\tau/\min(1 + \tau, 1 + \tau^{-1}), 1]$ ,  $\widehat{\Sigma}_f + S \succeq 0$  and  $\widehat{\Sigma}_g + T \succeq \frac{1}{2}\widehat{\Sigma}_g + T \succeq 0$ , we have  $-\tau + \alpha \min(1 + \tau, 1 + \tau^{-1}) > 0$ ,  $a_i \geq 0$  and  $\bar{\phi}_i + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^i\|^2 + \alpha\xi_i \geq 0$ , for any  $i \geq 1$ . It follows from (64) and the definitions of  $t_{i+1}$  and  $r^{i+1}$  that for any  $i \geq 1$  we have

$$a_i = t_{i+1} + (-\tau + \alpha \min(1 + \tau, 1 + \tau^{-1}))\sigma\|r^{i+1}\|^2 \\ \leq \left[ \bar{\phi}_i + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^i\|^2 + \alpha\xi_i \right] - \left[ \bar{\phi}_{i+1} + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^{i+1}\|^2 + \alpha\xi_{i+1} \right].$$

For any  $k \geq 1$ , summing the above inequality over  $i = 1, \dots, k$ , we obtain

$$\sum_{i=1}^k a_i \leq \bar{\phi}_1 + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^1\|^2 + \alpha\xi_1.$$

From the above inequality, we have  $\sum_{i=1}^\infty a_i < +\infty$ . Then, by Lemma 2.1 we get  $\min_{1 \leq i \leq k} \{a_i\} = o(1/k)$ , that is

$$\min_{1 \leq i \leq k} \{ \|z^{i+1} - z^i\|^2 + \|x^{i+1} - x^i\|^2 + \|y^{i+1} - y^i\|^2 \} = o(1/k). \quad (87)$$



It follows from the first-order optimality condition of (19) that

$$\nabla f(x^i) + A[z^i + \sigma(A^*x^{i+1} + B^*y^i - c)] + (\widehat{\Sigma}_f + S)(x^{i+1} - x^i) \in -\partial p(x^{i+1}).$$

And then from the definition of  $z^{i+1}$  in (19), we have

$$\begin{aligned} & \nabla f(x^{i+1}) - \nabla f(x^i) + (1 - \tau^{-1})A(z^{i+1} - z^i) + \sigma AB^*(y^{i+1} - y^i) - (\widehat{\Sigma}_f + S)(x^{i+1} - x^i) \\ & \in \partial p(x^{i+1}) + \nabla f(x^{i+1}) + Az^{i+1}. \end{aligned} \quad (88)$$

Similarly, we get

$$\nabla g(y^{i+1}) - \nabla g(y^i) + (1 - \tau^{-1})B(z^{i+1} - z^i) - (\widehat{\Sigma}_g + T)(y^{i+1} - y^i) \in \partial q(y^{i+1}) + \nabla g(y^{i+1}) + Bz^{i+1}. \quad (89)$$

It follows from (19) that

$$\|A^*x^{i+1} + B^*y^{i+1} - c\|^2 = (\tau\sigma)^{-2}\|z^{i+1} - z^i\|^2. \quad (90)$$

By using the Cauchy-Schwarz inequality, (88), (89) and (90), we have

$$\begin{aligned} & d^2(0, \partial p(x^{i+1}) + \nabla f(x^{i+1}) + Az^{i+1}) + d^2(0, \partial q(y^{i+1}) + \nabla g(y^{i+1}) + Bz^{i+1}) \\ & + \|\nabla f(x^{i+1}) - \nabla f(x^i) + (1 - \tau^{-1})A(z^{i+1} - z^i) + \sigma AB^*(y^{i+1} - y^i) - (\widehat{\Sigma}_f + S)(x^{i+1} - x^i)\|^2 \\ & \leq 4\|\nabla f(x^{i+1}) - \nabla f(x^i)\|^2 + 4(1 - \tau^{-1})^2\|A(z^{i+1} - z^i)\|^2 + 4\sigma^2\|AB^*(y^{i+1} - y^i)\|^2 \\ & + 4\|(\widehat{\Sigma}_f + S)(x^{i+1} - x^i)\|^2 + 3\|\nabla g(y^{i+1}) - \nabla g(y^i)\|^2 + 3(1 - \tau^{-1})^2\|B(z^{i+1} - z^i)\|^2 \\ & + 3\|(\widehat{\Sigma}_g + T)(y^{i+1} - y^i)\|^2 + (\tau\sigma)^{-2}\|z^{i+1} - z^i\|^2. \end{aligned}$$

It follows from the above inequality, Assumption 2.1 and (87) that the assertion (85) is proved.

By using (13), for any  $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ , we obtain

$$(p(x) + f(x) + q(y) + g(y)) - (p(\bar{x}) + f(\bar{x}) + q(\bar{y}) + g(\bar{y})) + \langle \bar{z}, A^*x + B^*y - c \rangle \geq 0.$$

Setting  $x = x^i$  and  $y = y^i$  in the above inequality, we get

$$(p(x^i) + f(x^i) + q(y^i) + g(y^i)) - (p(\bar{x}) + f(\bar{x}) + q(\bar{y}) + g(\bar{y})) \geq -\langle \bar{z}, A^*x^i + B^*y^i - c \rangle. \quad (91)$$

Note that  $p$ ,  $f$ ,  $q$  and  $g$  are convex functions and the sequence  $\{(x^i, y^i, z^i)\}$  generated by the Majorized iPADMM is bounded. For any  $u \in \partial p(x^i)$  and  $v \in \partial q(y^i)$ , using  $A^*\bar{x} + B^*\bar{y} = c$ , we obtain

$$\begin{aligned} & (p(\bar{x}) + f(\bar{x}) + q(\bar{y}) + g(\bar{y})) - (p(x^i) + f(x^i) + q(y^i) + g(y^i)) \\ & \geq \langle u + \nabla f(x^i), \bar{x} - x^i \rangle + \langle v + \nabla g(y^i), \bar{y} - y^i \rangle \\ & = \langle u + \nabla f(x^i) + Az^i, \bar{x} - x^i \rangle + \langle z^i, A^*x^i - A^*\bar{x} \rangle + \langle v + \nabla g(y^i) + Bz^i, \bar{y} - y^i \rangle \\ & \quad + \langle z^i, B^*y^i - B^*\bar{y} \rangle \\ & = \langle u + \nabla f(x^i) + Az^i, \bar{x} - x^i \rangle + \langle v + \nabla g(y^i) + Bz^i, \bar{y} - y^i \rangle + \langle z^i, A^*x^i + B^*y^i - c \rangle, \end{aligned}$$

which, together with (85) and (91), implies (86). The proof is complete.  $\square$

## 5.2 The ergodic iteration-complexity

In this subsection, we shall establish a worst-case ergodic iteration-complexity for the sequence  $\{(x^i, y^i, z^i)\}$  generated by the Majorized iPADMM. Recall that  $\tilde{z}^{i+1}$  is defined by (23). Let

$$\hat{x}^k = \frac{1}{k} \sum_{i=1}^k x^{i+1}, \quad \hat{y}^k = \frac{1}{k} \sum_{i=1}^k y^{i+1} \quad \text{and} \quad \hat{z}^k = \frac{1}{k} \sum_{i=1}^k \tilde{z}^{i+1}.$$

**Lemma 5.1.** *Suppose that Assumption 2.1 holds. Assume that  $\tau \in (0, (1 + \sqrt{5})/2)$  and for some  $\alpha \in (\tau/\min(1 + \tau, 1 + \tau^{-1}), 1]$ ,*

$$\hat{\Sigma}_f + S \succeq 0, \quad H_f \succeq 0, \quad \frac{1}{2}\hat{\Sigma}_g + T \succeq 0 \quad \text{and} \quad M_g \succ 0,$$

where  $H_f$  and  $M_g$  are defined in (20). Then, for any  $k \geq 1$  and  $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ , we have

$$\begin{aligned} & (p(\hat{x}^k) + q(\hat{y}^k)) - (p(x) + q(y)) + \langle \hat{x}^k - x, \nabla f(x) + Az \rangle + \langle \hat{y}^k - y, \nabla g(y) + Bz \rangle \\ & + \langle \hat{z}^k - z, -(A^*x + B^*y - c) \rangle \leq \frac{\phi_1(x, y, z) + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^1\|^2 + \alpha\xi_1}{2k}. \end{aligned} \quad (92)$$

*Proof.* By the assumptions that  $\tau \in (0, (1 + \sqrt{5})/2)$ ,  $\alpha \in (\tau/\min(1 + \tau, 1 + \tau^{-1}), 1]$ ,  $\hat{\Sigma}_f + S \succeq 0$ ,  $H_f \succeq 0$ ,  $\frac{1}{2}\hat{\Sigma}_g + T \succeq 0$  and  $M_g \succ 0$ , from (21) and (22), for any  $i \geq 1$ , we get

$$\phi_i(x, y, z) + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^i\|^2 + \alpha\xi_i \geq 0 \quad \text{and} \quad t_{i+1} + (-\tau + \alpha \min(1 + \tau, 1 + \tau^{-1}))\sigma\|r^{i+1}\|^2 \geq 0.$$

Then, it follows from (39) that

$$\begin{aligned} & (p(x^{i+1}) + q(y^{i+1})) - (p(x) + q(y)) + \langle x^{i+1} - x, \nabla f(x) + Az \rangle + \langle y^{i+1} - y, \nabla g(y) + Bz \rangle \\ & + \langle \tilde{z}^{i+1} - z, -(A^*x + B^*y - c) \rangle \\ & \leq \frac{1}{2} \left\{ [\phi_i(x, y, z) + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^i\|^2 + \alpha\xi_i] - [\phi_{i+1}(x, y, z) + (1 - \alpha \min(\tau, \tau^{-1}))\right. \\ & \quad \left. \times \sigma\|r^{i+1}\|^2 + \alpha\xi_{i+1}] \right\}. \end{aligned}$$

Summing the above inequalities over  $i = 1, \dots, k$ , we obtain

$$\begin{aligned} & \sum_{i=1}^k \left( p(x^{i+1}) + q(y^{i+1}) \right) - k(p(x) + q(y)) + \left\langle \sum_{i=1}^k x^{i+1} - kx, \nabla f(x) + Az \right\rangle \\ & + \left\langle \sum_{i=1}^k y^{i+1} - ky, \nabla g(y) + Bz \right\rangle + \left\langle \sum_{i=1}^k \tilde{z}^{i+1} - kz, -(A^*x + B^*y - c) \right\rangle \\ & \leq \frac{1}{2} [\phi_1(x, y, z) + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^1\|^2 + \alpha\xi_1]. \end{aligned}$$

Since  $(\hat{x}^k, \hat{y}^k, \hat{z}^k)$  is a convex combination of  $(x^2, y^2, \tilde{z}^2), \dots, (x^{k+1}, y^{k+1}, \tilde{z}^{k+1})$ , for any  $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ , we obtain that

$$\begin{aligned} & \frac{1}{k} \sum_{i=1}^k \left( p(x^{i+1}) + q(y^{i+1}) \right) - (p(x) + q(y)) + \left\langle \hat{x}^k - x, \nabla f(x) + Az \right\rangle + \left\langle \hat{y}^k - y, \nabla g(y) + Bz \right\rangle \\ & + \left\langle \hat{z}^k - z, -(A^*x + B^*y - c) \right\rangle \leq \frac{1}{2k} [\phi_1(x, y, z) + (1 - \alpha \min(\tau, \tau^{-1}))\sigma\|r^1\|^2 + \alpha\xi_1]. \end{aligned}$$

By using the convexity of  $p(\cdot)$  and  $q(\cdot)$ , we obtain

$$p(\hat{x}^k) + q(\hat{y}^k) \leq \frac{1}{k} \sum_{i=1}^k (p(x^{i+1}) + q(y^{i+1})).$$

The assertion (92) then follows from the above two inequalities immediately.  $\square$

**Theorem 5.2.** *Suppose that Assumptions 2.1 and 2.2 hold. Assume that  $\tau \in (0, (1 + \sqrt{5})/2)$  and for some  $\alpha \in (\tau/\min(1 + \tau, 1 + \tau^{-1}), 1]$ ,*

$$\hat{\Sigma}_f + S \succeq 0, \quad H_f \succeq 0, \quad \frac{1}{2}\Sigma_f + S + \sigma AA^* \succ 0, \quad \frac{1}{2}\hat{\Sigma}_g + T \succeq 0 \quad \text{and} \quad M_g \succ 0,$$

where  $H_f$  and  $M_g$  are defined in (20). Then the Majorized iPADMM has a worst-case  $O(1/k)$  ergodic iteration-complexity.

*Proof.* From (92) in Lemma 5.1 and the definitions of  $\phi_1(x, y, z)$ ,  $r^1$  and  $\xi_1$ , we know

$$\begin{aligned} & (p(\hat{x}^k) + q(\hat{y}^k)) - (p(x) + q(y)) + \langle \hat{x}^k - x, \nabla f(x) + Az \rangle + \langle \hat{y}^k - y, \nabla g(y) + Bz \rangle \\ & \quad + \langle \hat{z}^k - z, -(A^*x + B^*y - c) \rangle \\ & \leq \frac{1}{2k} \left( (\tau\sigma)^{-1} \|z^1 - z\|^2 + \|x^1 - x\|_{\hat{\Sigma}_f + S}^2 + \|y^1 - y\|_{\hat{\Sigma}_g + T}^2 + \sigma \|A^*x + B^*y^1 - c\|^2 \right. \\ & \quad \left. + (1 - \alpha \min(\tau, \tau^{-1})) \tau^{-2} \sigma^{-1} \|z^1 - z^0\|^2 + \alpha \|y^1 - y^0\|_{\hat{\Sigma}_g + T}^2 \right). \end{aligned} \quad (93)$$

Note that  $\hat{\Sigma}_f + S \succeq 0$  and  $\hat{\Sigma}_g + T \succeq \frac{1}{2}\hat{\Sigma}_g + T \succeq 0$ . By using the Cauchy-Schwarz inequality and  $z^{k+1} - z^1 = k\tau\sigma(A^*\hat{x}^k + B^*\hat{y}^k - c)$ , for any  $w \in \mathcal{B}(\hat{w}^k)$  and  $\bar{w} := (\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W}_0^*$ , which is a nonempty compact subset of  $\mathcal{W}^*$ , we have

$$\begin{aligned} & (\tau\sigma)^{-1} \|z^1 - z\|^2 + \|x^1 - x\|_{\hat{\Sigma}_f + S}^2 + \|y^1 - y\|_{\hat{\Sigma}_g + T}^2 + \sigma \|A^*x + B^*y^1 - c\|^2 \\ & = (\tau\sigma)^{-1} \|z^1 - \hat{z}^k + \hat{z}^k - z\|^2 + \|x^1 - \hat{x}^k + \hat{x}^k - x\|_{\hat{\Sigma}_f + S}^2 + \|y^1 - \hat{y}^k + \hat{y}^k - y\|_{\hat{\Sigma}_g + T}^2 \\ & \quad + \sigma \|(A^*\hat{x}^k + B^*\hat{y}^k - c) + (A^*x - A^*\hat{x}^k) + (B^*y^1 - B^*\hat{y}^k)\|^2 \\ & \leq 2(\tau\sigma)^{-1} \|z^1 - \hat{z}^k\|^2 + 2\|x^1 - \hat{x}^k\|_{\hat{\Sigma}_f + S}^2 + 2\|y^1 - \hat{y}^k\|_{\hat{\Sigma}_g + T + \sigma BB^*}^2 + 4\sigma \|A^*\hat{x}^k + B^*\hat{y}^k - c\|^2 \\ & \quad + 2(\tau\sigma)^{-1} \|\hat{z}^k - z\|^2 + 2\|\hat{x}^k - x\|_{\hat{\Sigma}_f + S + 2\sigma AA^*}^2 + 2\|\hat{y}^k - y\|_{\hat{\Sigma}_g + T}^2 \\ & \leq 2[(\tau\sigma)^{-1} \|z^1 - \hat{z}^k\|^2 + \|x^1 - \hat{x}^k\|_{\hat{\Sigma}_f + S}^2 + \|y^1 - \hat{y}^k\|_{\hat{\Sigma}_g + T + \sigma BB^*}^2] + \frac{4}{k^2 \tau^2 \sigma} \|z^{k+1} - z^1\|^2 + 2D_1 \\ & \leq 4[(\tau\sigma)^{-1} \|\bar{z} - \hat{z}^k\|^2 + \|\bar{x} - \hat{x}^k\|_{\hat{\Sigma}_f + S}^2 + \|\bar{y} - \hat{y}^k\|_{\hat{\Sigma}_g + T + \sigma BB^*}^2] \\ & \quad + \frac{8}{k^2 \tau^2 \sigma} \|\bar{z} - z^{k+1}\|^2 + 2D_1 + 4D_2, \end{aligned} \quad (94)$$

where the constants  $D_1$  and  $D_2$  are defined, respectively, by

$$D_1 := \sup_{k \geq 1} \max_{w \in \mathcal{B}(\hat{w}^k)} \left\{ (\tau\sigma)^{-1} \|\hat{z}^k - z\|^2 + \|\hat{x}^k - x\|_{\hat{\Sigma}_f + S + 2\sigma AA^*}^2 + \|\hat{y}^k - y\|_{\hat{\Sigma}_g + T}^2 \right\}$$

and

$$D_2 := \max_{(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W}_0^*} \left\{ (\tau\sigma)^{-1} \|z^1 - \bar{z}\|^2 + \|x^1 - \bar{x}\|_{\hat{\Sigma}_f+S}^2 + \|y^1 - \bar{y}\|_{\hat{\Sigma}_g+T+\sigma BB^*}^2 + \frac{2}{\tau^2\sigma} \|z^1 - \bar{z}\|^2 \right\}.$$

Let

$$D_3 := (1 - \alpha \min(\tau, \tau^{-1})) \tau^{-2} \sigma^{-1} \|z^1 - z^0\|^2 + \alpha \|y^1 - y^0\|_{\hat{\Sigma}_g+T}^2.$$

It then follows from Part (b) of Theorem 4.1 that for any  $i \geq 1$  and any  $\bar{w} := (\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W}^*$ , we have

$$(\tau\sigma)^{-1} \|\bar{z} - z^{i+1}\|^2 + \|\bar{x} - x^{i+1}\|_{\hat{\Sigma}_f+S}^2 + \|\bar{y} - y^{i+1}\|_{\hat{\Sigma}_g+T+\sigma BB^*}^2 \leq D_2 + D_3, \quad (95)$$

which, together with the convexity of the quadratic function  $\|\cdot\|^2$ , implies that for any  $k \geq 1$  and  $\bar{w} := (\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W}^*$ ,

$$(\tau\sigma)^{-1} \left\| \bar{z} - \frac{1}{k} \sum_{i=1}^k z^{i+1} \right\|^2 + \|\bar{x} - \hat{x}^k\|_{\hat{\Sigma}_f+S}^2 + \|\bar{y} - \hat{y}^k\|_{\hat{\Sigma}_g+T+\sigma BB^*}^2 \leq D_2 + D_3. \quad (96)$$

Then, by using the fact that for  $k \geq 2$ ,

$$\hat{z}^k = \frac{1}{k} \sum_{i=1}^k \hat{z}^{i+1} = \tau^{-1} \frac{1}{k} \sum_{i=1}^k z^{i+1} + (1 - \tau^{-1}) \left( \frac{1}{k} z^1 + \frac{k-1}{k} \frac{1}{k-1} \sum_{i=1}^{k-1} z^{i+1} \right),$$

we obtain that for any  $k \geq 2$  and any  $\bar{w} := (\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W}^*$ ,

$$\begin{aligned} & (\tau\sigma)^{-1} \|\bar{z} - \hat{z}^k\|^2 \\ &= (\tau\sigma)^{-1} \left\| \tau^{-1} \left( \frac{1}{k} \sum_{i=1}^k z^{i+1} - \bar{z} \right) + (1 - \tau^{-1}) \left( \frac{1}{k} z^1 + \frac{k-1}{k} \frac{1}{k-1} \sum_{i=1}^{k-1} z^{i+1} - \bar{z} \right) \right\|^2 \\ &\leq 2(\tau\sigma)^{-1} \tau^{-2} \left\| \frac{1}{k} \sum_{i=1}^k z^{i+1} - \bar{z} \right\|^2 + 2(\tau\sigma)^{-1} (1 - \tau^{-1})^2 \left\| \left( \frac{1}{k} z^1 + \frac{k-1}{k} \frac{1}{k-1} \sum_{i=1}^{k-1} z^{i+1} - \bar{z} \right) \right\|^2 \\ &\leq 2[\tau^{-2} + (1 - \tau^{-1})^2] (D_2 + D_3). \end{aligned} \quad (97)$$

It follows from (96) and (97) that

$$\begin{aligned} & 4[(\tau\sigma)^{-1} \|\bar{z} - \hat{z}^k\|^2 + \|\bar{x} - \hat{x}^k\|_{\hat{\Sigma}_f+S}^2 + \|\bar{y} - \hat{y}^k\|_{\hat{\Sigma}_g+T+\sigma BB^*}^2] \\ &\leq 8[\tau^{-2} + (1 - \tau^{-1})^2] (D_2 + D_3) + 4(D_2 + D_3). \end{aligned}$$

By using (95), we have

$$\frac{8}{k^2 \tau^2 \sigma} \|\bar{z} - z^{k+1}\|^2 \leq 8\tau^{-1} (\tau\sigma)^{-1} \|\bar{z} - z^{k+1}\|^2 \leq 8\tau^{-1} (D_2 + D_3).$$

Therefore, from (93), (94) and the above two inequalities, we know that for any  $k \geq 2$ ,

$$\begin{aligned} & (p(\hat{x}^k) + q(\hat{y}^k)) - (p(x) + q(y)) + \langle \hat{x}^k - x, \nabla f(x) + Az \rangle + \langle \hat{y}^k - y, \nabla g(y) + Bz \rangle \\ &+ \langle \hat{z}^k - z, -(A^*x + B^*y - c) \rangle \leq \frac{D}{2k}, \end{aligned}$$

where the constant  $D$  is given by

$$D := 8[\tau^{-2} + (1 - \tau^{-1})^2 + \tau^{-1}](D_2 + D_3) + 2D_1 + 8D_2 + 5D_3.$$

Therefore, for any given  $\varepsilon > 0$ , after at most  $k := \lceil \frac{D}{2\varepsilon} \rceil$  iterations, we have

$$\begin{aligned} & (p(\hat{x}^k) + q(\hat{y}^k)) - (p(x) + q(y)) + \langle \hat{x}^k - x, \nabla f(x) + Az \rangle + \langle \hat{y}^k - y, \nabla g(y) + Bz \rangle \\ & + \langle \hat{z}^k - z, -(A^*x + B^*y - c) \rangle \leq \varepsilon \quad \forall w \in \mathcal{B}(\hat{w}^k), \end{aligned}$$

which means that  $(\hat{x}^k, \hat{y}^k, \hat{z}^k)$  is an approximate solution of  $\text{VI}(\mathcal{W}, F, \theta)$  with an accuracy of  $O(1/k)$ . That is, a worst-case  $O(1/k)$  ergodic iteration-complexity of the Majorized iPADMM is established. The proof is completed.  $\square$

## 6 Numerical experiments

We consider the following problem to illustrate the benefit which can be brought about by using an indefinite proximal term instead of the standard requirement of a positive semidefinite proximal term in applying the semi-proximal ADMM:

$$\min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} \left\{ \frac{1}{2} \langle x, Qx \rangle - \langle b, x \rangle + \frac{\chi}{2} \|\Pi_{\mathbb{R}_+^m}(d - Hx)\|^2 + \varrho \|x\|_1 + \delta_{\mathbb{R}_+^m}(y) \mid Hx + y = c \right\}, \quad (98)$$

where  $\|x\|_1 := \sum_{i=1}^n |x_i|$ ,  $\delta_{\mathbb{R}_+^m}(\cdot)$  is the indicator function of  $\mathbb{R}_+^m$ ,  $Q$  is an  $n \times n$  symmetric and positive semidefinite matrix (may not be positive definite),  $H \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^n$ ,  $c \in \mathbb{R}^m$  and  $\varrho > 0$  are given data. In addition,  $d \leq c$  are given vectors,  $\chi$  is a nonnegative penalty parameter, and  $\Pi_{\mathbb{R}_+^m}(\cdot)$  denotes the projection onto  $\mathbb{R}_+^m$ . Observe that when the parameter  $\chi$  is chosen to be positive, one may view the term  $\frac{\chi}{2} \|\Pi_{\mathbb{R}_+^m}(d - Hx)\|^2$  as the penalty for failing to satisfy the soft constraint  $Hx - d \geq 0$ .

For problem (98), it can be expressed in the form (1) by taking

$$f(x) = \frac{1}{2} \langle x, Qx \rangle - \langle b, x \rangle + \frac{\chi}{2} \|\Pi_{\mathbb{R}_+^m}(d - Hx)\|^2, \quad p(x) = \varrho \|x\|_1, \quad g(y) \equiv 0, \quad q(y) = \delta_{\mathbb{R}_+^m}(y)$$

with  $A^* = H$ ,  $B^* = I$ . The KKT system for problem (98) is given by:

$$Hx + y - c = 0, \quad \nabla f(x) + H^* \xi + v = 0, \quad y \geq 0, \quad \xi \geq 0, \quad y \circ \xi = 0, \quad v \in \partial \varrho \|x\|_1, \quad (99)$$

where “ $\circ$ ” denotes the elementwise product. In our numerical experiments, we apply the Majorized iPADMM to solve the problem (98) by using both the step-length parameters  $\tau = 1.618$  and  $\tau = 1$ . We stop the iPADMM based on the following relative residual on the KKT system:

$$\max \left\{ \frac{\|Hx^{k+1} + y^{k+1} - c\|}{1 + \|c\|}, \frac{\|\nabla f(x^{k+1}) + H^* \xi^{k+1} + v^{k+1}\|}{1 + \|b\|} \right\} \leq 10^{-6}. \quad (100)$$

Note that in the process of computing the iterates  $x^{k+1}$  and  $y^{k+1}$  from the iPADMM, we can generate the corresponding dual variables  $\xi^{k+1}$  and  $v^{k+1}$  which satisfy the complementarity conditions in (99). Thus we need not check the complementarity conditions in (99) since they are satisfied exactly.

In our numerical experiments, for a given pair of  $(n, m)$ , we generate the data for (98) randomly as follows:

```

Q1= sprandn(floor(0.1*n),n,0.1); Q = Q1'*Q1;
H = sprandn(m,n,0.2); xx = randn(n,1); c = H*xx + max(randn(m,1),0); b = Q*xx;

```

and we set the parameter  $\varrho = 5\sqrt{n}$ . We take  $d = c - 5e$ , where  $e$  is the vector of all ones. As we can see,  $Q$  is positive semidefinite but not positive definite. Note that for the data generated from the above scheme, the optimal solution  $(x^*, y^*)$  of (98) has the property that both  $x^*$  and  $y^*$  are nonzero vectors but each has a significant portion of zero components. We have tested other schemes to generate the data but the corresponding optimal solution is not interesting enough in that  $y^*$  is usually the zero vector.

In the next two subsections, we consider two separate cases for evaluating the performance of the Majorized iPADMM in solving (98).

### 6.1 The case where the parameter $\chi$ in (98) is zero

In the first case, we set the penalty parameter  $\chi = 0$ . Hence  $f(x)$  is simply a convex quadratic function, and we can omit the word “Majorized” since there is no majorization on  $f$  or  $g$ . For this case, we have that  $\widehat{\Sigma}_f = Q = \Sigma_f$  and  $\widehat{\Sigma}_g = 0 = \Sigma_g$ . We consider the following two choices of the proximal terms for the iPADMM.

(a) The proximal terms in the iPADMM are chosen according to (84) with

$$S := \rho_1 I - \widehat{\Sigma}_f - \sigma AA^*, \quad T := 0, \quad (101)$$

where  $\rho_1$  is the constant given in (83) i.e.,

$$\rho_1 := 1.01\lambda_{\max}\left(\widehat{\Sigma}_f - \frac{1}{2}\Sigma_f + \frac{1}{2}(1 + \alpha)\sigma AA^*\right). \quad (102)$$

In the notation of (84), we fix  $\alpha := \min(1.01\tau / \min(1 + \tau, 1 + \tau^{-1}), 1)$ . With the above choices of the proximal terms, we have that

$$\widehat{\Sigma}_f + S + \sigma AA^* = \rho_1 I, \quad \widehat{\Sigma}_g + T + \sigma BB^* = \sigma I.$$

Furthermore, the conditions (63) and (65) in Theorem 4.1 are satisfied. Therefore the convergence of the iPADMM is ensured even though  $S$  is an indefinite matrix.

(b) The proximal terms in the iPADMM are chosen based on Part (a) of Theorem 4.1 as follows:

$$S := \rho_2 I - \widehat{\Sigma}_f - \sigma AA^*, \quad T := 0, \quad (103)$$

where for a chosen parameter  $\eta \in (0, 1/2)$ , say  $\eta := 0.49$ , and  $\gamma_2 := 1.1$ ,

$$\rho_2 := \lambda_{\max}\left(\frac{1}{2}Q + \gamma_2(1 - \eta)\sigma AA^*\right). \quad (104)$$

Observe that  $\rho_2$  is a smaller constant than  $\rho_1$  and hence the proximal term  $S$  in (103) is more indefinite than the one in (101). In this case, we can easily check that the conditions in (61) for  $S$  and  $T$  are satisfied. In particular,  $\widehat{\Sigma}_f + S + \eta\sigma AA^* \succeq (1 - \eta)\sigma(\lambda_{\max}(AA^*)I - AA^*) + (\gamma_2 - 1)(1 - \eta)\sigma\lambda_{\max}(AA^*)I \succ 0$ . However, in order to ensure that the iPADMM with the proximal terms chosen in (103) is convergent, we need to monitor the residual

$$R^{k+1} := \|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 + \|y^{k+1} - y^k\|_{\widehat{\Sigma}_g + \sigma BB^*}^2 + \|r^{k+1}\|^2 \quad (105)$$

Table 1: Comparison between the number of iterations taken by iPADMM (for  $S$  given by (101)) and sPADMM (for  $S = \lambda_{\max}(Q + \sigma AA^*)I - (Q + \sigma AA^*)$ ) to solve the problem (98), with  $\chi = 0$ , to the required accuracy stated in (100).

dim. of $H$ ( $m \times n$ )	sPADMM $\tau = 1.618$	iPADMM $\tau = 1.618$	ratio (%)	sPADMM $\tau = 1$	iPADMM $\tau = 1$	ratio (%)
$2000 \times 1000$	8975	8571	95.5	11077	9317	84.1
$2000 \times 2000$	1807	1406	77.8	1970	1399	71.0
$2000 \times 4000$	1224	714	58.3	1254	767	61.2
$2000 \times 8000$	1100	613	55.7	1103	642	58.2
$4000 \times 2000$	9863	9222	93.5	11919	10141	85.1
$4000 \times 4000$	2004	1273	63.5	2140	1468	68.6
$4000 \times 8000$	1408	829	58.9	1396	860	61.6
$4000 \times 16000$	1749	913	52.2	1771	938	53.0
$8000 \times 4000$	10681	10288	96.3	13187	11111	84.3
$8000 \times 8000$	2037	1220	59.9	2085	1298	62.3
$8000 \times 16000$	1594	917	57.5	1630	956	58.7

in condition (62) as follows. At the  $k$ th iteration, if  $\sum_{j=1}^{k+1} R^j \geq 50$  and  $R^{k+1} \geq 10/(k+1)^{1.1}$ , restart the iPADMM (using the best iterate generated so far as the initial point) with a new  $S$  as in (103) but the parameter  $\gamma_2$  in (104) is increased by a factor of 1.1. Obviously, the number of such restarts is finite since eventually  $\rho_2$  will be larger than  $\rho_1$  in (102). In our numerical runs, we only encounter such a restart once when testing all the instances.

Table 1 reports the comparison between the iPADMM (whose proximal term  $S$  is the indefinite matrix mentioned in (101)) and the semi-proximal ADMM in [12] (denoted as sPADMM) where the proximal term  $S$  is replaced by the positive semidefinite matrix  $\lambda_{\max}(Q + \sigma AA^*)I - (Q + \sigma AA^*)$ . Table 2 is the same as Table 1 except that the comparison is between the iPADMM (whose indefinite proximal term  $S$  is given by (103)) and the sPADMM.

We can see from the results in both tables that the iPADMM can sometimes bring about 40–50% reduction in the number of iterations needed to solve the problem (98) as compared to the sPADMM. In addition, the iPADMM using the more aggressive indefinite proximal term  $S$  in (103) sometimes takes substantially less iterations to solve the problems as compared to the more conservative choice in (101).

## 6.2 The case where the parameter $\chi$ in (98) is positive

In the second case, we consider problem (98) where the parameter  $\chi$  is set to  $\chi = 2\rho$ . In this case, a majorization on  $f$  is necessary in order for the corresponding subproblem in the iPADMM or sPADMM to be solved efficiently. In this case, we have  $\hat{\Sigma}_f = Q + \chi H^*H$ ,  $\Sigma_f = Q$ ,  $\hat{\Sigma}_g = 0 = \Sigma_g$ . For problem (98) with  $\chi = 2\rho$ , the indefinite proximal term  $S$  given in (101) is too conservative due to the fact that  $\hat{\Sigma}_f$  is substantially “larger” than  $\Sigma_f$ . In order to realize the full potential of allowing for an indefinite proximal term, we make use of the condition (62) in Part (a) of Theorem

Table 2: Comparison between the number of iterations taken by iPADMM (for  $S$  given by (103)) and sPADMM (for  $S = \lambda_{\max}(Q + \sigma AA^*)I - (Q + \sigma AA^*)$ ) to solve the problem (98), with  $\chi = 0$ , to the required accuracy stated in (100).

dim. of $H$ ( $m \times n$ )	sPADMM $\tau = 1.618$	iPADMM $\tau = 1.618$	ratio (%)	sPADMM $\tau = 1$	iPADMM $\tau = 1$	ratio (%)
$2000 \times 1000$	8975	6216	69.3	11077	8094	73.1
$2000 \times 2000$	1807	1001	55.4	1970	1119	56.8
$2000 \times 4000$	1224	659	53.8	1254	734	58.5
$2000 \times 8000$	1100	593	53.9	1103	626	56.8
$4000 \times 2000$	9863	7312	74.1	11919	8762	73.5
$4000 \times 4000$	2004	1148	57.3	2140	1280	59.8
$4000 \times 8000$	1408	805	57.2	1396	846	60.6
$4000 \times 16000$	1749	900	51.5	1771	929	52.5
$8000 \times 4000$	10681	8047	75.3	13187	9587	72.7
$8000 \times 8000$	2037	1109	54.4	2085	1221	58.6
$8000 \times 16000$	1594	902	56.6	1630	940	57.7

4.1 by choosing  $S$  and  $T$  as follows:

$$S := \rho_3 I - \widehat{\Sigma}_f - \sigma AA^*, \quad T := 0, \quad (106)$$

where for a chosen parameter  $\eta \in (0, 1/2)$ , say  $\eta := 0.49$ , and  $\gamma_3 := 0.25$ ,

$$\rho_3 := \lambda_{\max}\left(\frac{1}{2}Q + ((1 - \eta)\sigma + \gamma_3\chi)AA^*\right). \quad (107)$$

In this case, we can easily check that the conditions in (61) for  $S$  and  $T$  are satisfied. In particular,  $\widehat{\Sigma}_f + S + \eta\sigma AA^* \succeq (1 - \eta)\sigma(\lambda_{\max}(AA^*)I - AA^*) + \gamma_3\chi\lambda_{\max}(AA^*)I \succ 0$ . Again, in order to ensure that the Majorized iPADMM with the proximal terms chosen in (106) is convergent, we need to monitor the residual, defined similarly as in (105), in condition (62) as follows. At the  $k$ th iteration, if  $\sum_{j=1}^{k+1} R^j \geq 50$  and  $R^{k+1} \geq 10/(k+1)^{1.1}$ , restart the Majorized iPADMM (using the best iterate generated so far as the initial point) with a new  $S$  as in (106) but the parameter  $\gamma_3$  in (107) is increased by a factor of 1.1. As before, the number of such restarts is finite since eventually  $\rho_3$  will be larger than  $\rho_1$  in (102). In our numerical runs, each of the tested instances encounters such a restart at most once.

Table 3 reports the comparison between the Majorized iPADMM (whose proximal term  $S$  is the indefinite matrix given in (106)) and the Majorized sPADMM where the proximal term  $S$  is replaced by the positive semidefinite matrix  $\lambda_{\max}(\widehat{\Sigma}_f + \sigma AA^*)I - (\widehat{\Sigma}_f + \sigma AA^*)$ . We can see from the results in the table that the iPADMM can achieve the dramatic reduction of about 50–70% in the number of iterations needed to solve the problem (98) as compared to the sPADMM.

The numerical results in this section serve to demonstrate the benefit one can get by using an indefinite proximal term in the Majorized iPADMM. Naturally, this calls on more research on augmented Lagrangian function based methods beyond their traditional domain.



Table 3: Comparison between the number of iterations taken by the Majorized iPADMM (for  $S$  given by (106)) and Majorized sPADMM (for  $S = \lambda_{\max}(\hat{\Sigma}_f + \sigma AA^*)I - (\hat{\Sigma}_f + \sigma AA^*)$ ) to solve the problem (98), with  $\chi = 2\rho$ , to the required accuracy stated in (100).

dim. of $H$ ( $m \times n$ )	sPADMM $\tau = 1.618$	iPADMM $\tau = 1.618$	ratio (%)	sPADMM $\tau = 1$	iPADMM $\tau = 1$	ratio (%)
$2000 \times 1000$	16030	5056	31.5	16638	5936	35.7
$2000 \times 2000$	2091	630	30.1	2111	708	33.5
$2000 \times 4000$	1292	428	33.1	1311	481	36.7
$2000 \times 8000$	972	372	38.3	988	423	42.8
$4000 \times 2000$	14499	5248	36.2	14750	5628	38.2
$4000 \times 4000$	1599	568	35.5	1677	665	39.7
$4000 \times 8000$	1077	430	39.9	1100	506	46.0
$4000 \times 16000$	1082	428	39.6	1109	500	45.1
$8000 \times 4000$	10528	4016	38.1	10782	4515	41.9
$8000 \times 8000$	1260	542	43.0	1318	645	48.9
$8000 \times 16000$	983	439	44.7	991	526	53.1

**Acknowledgements.** The authors would like to thank Ms. Ying Cui at Department of Mathematics, National University of Singapore for her suggestion of using Clarke’s Mean Value Theorem to analyze the behavior of the algorithm described in the paper.

## References

- [1] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent, Mathematical Programming, Series A, DOI 10.1007/s10107-014-0826-5, (2014).
- [2] C. H. Chen, Y. Shen and Y. F. You, On the convergence analysis of the alternating direction method of multipliers with three blocks, Abstract and Applied Analysis, **2013**, Article ID 183961, 7 pages (2013).
- [3] F. H. Clarke, Optimization and Nonsmooth Analysis, 2nd edition, Classics in Applied Mathematics, vol. 5, Society for Industrial and Applied Mathematics, Philadelphia, (1990).
- [4] Y. Cui, X. D. Li, D. F. Sun and K.-C. Toh, On the convergence properties of a majorized ADMM for linearly constrained convex optimization problems with coupled objective functions, arXiv:1502.00098v1, (2015).
- [5] D. Davis and W. Yin, Convergence rate analysis of several splitting schemes, arXiv:1406.4834v3, (2014).
- [6] W. Deng, M.-J. Lai, Z. M. Peng and W. Yin, Parallel multi-block ADMM with  $o(1/k)$  convergence, arXiv:1312.3040v2, (2014).
- [7] W. Deng and W. Yin, On the global and linear convergence of the generalized alternating direction method of multipliers, Rice University CAAM Technical Report TR12-14, (2012).
- [8] J. Eckstein, Some saddle-function splitting methods for convex programming, Optimization Methods and Software, **4**, pp. 75–83 (1994).
- [9] J. Eckstein and D. P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, Mathematical Programming, **55**, pp. 293–318 (1992).
- [10] J. Eckstein and W. Yao, Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives, RUTCOR Research Report, (2014).

- [11] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Series in Operations Research, Springer Verlag, New York, (2003).
- [12] M. Fazel, T. K. Pong, D. F. Sun and P. Tseng, Hankel matrix rank minimization with applications to system identification and realization, *SIAM Journal on Matrix Analysis and Applications*, **34**(3), pp. 946–977 (2013).
- [13] M. Fortin and R. Glowinski, On decomposition-coordination methods using an augmented Lagrangian. In M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Solution of Boundary Problems*. North-Holland, Amsterdam, 1983.
- [14] D. Gabay, Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Solution of Boundary Problems*. North-Holland, Amsterdam, 1983.
- [15] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximations, *Computers & Mathematics with Applications*, **2**, pp. 17–40 (1976).
- [16] Y. Gao and D. F. Sun, A majorized penalty approach for calibrating rank constrained correlation matrix problems, Technical Report, Department of Mathematics, National University of Singapore, 2010.
- [17] R. Glowinski, Lectures on numerical methods for nonlinear variational problems, vol. 65 of *Tata Institute of Fundamental Research Lectures on Mathematics and Physics*, Tata Institute of Fundamental Research, Bombay, 1980. Notes by M. G. Vijayasundaram and M. Adimurthi.
- [18] R. Glowinski and A. Marroco, Sur l’approximation, par elements finis d’ordre un, et la resolution, par penalisation-dualit’e, d’une classe de problemes de Dirichlet non lineares, *Revue Francaise d’Automatique, Informatique et Recherche Op’erationelle*. 9 (R-2), pp. 41–76 (1975).
- [19] D. R. Han and X. M. Yuan, A note on the alternating direction method of multipliers, *Journal of Optimization Theory and Applications*, **155**(1), pp. 227–238 (2012).
- [20] B. S. He and X. M. Yuan, On the  $O(1/n)$  convergence rate of the Douglas-Rachford alternating direction method, *SIAM Journal on Numerical Analysis*, **50**, pp. 700–709 (2012).
- [21] M. Y. Hong and Z. Q. Luo, On the linear convergence of the alternating direction method of multipliers, *arXiv:1208.3922*, (2012).
- [22] M. Li, D. F. Sun and K.-C. Toh, A convergent 3-block semi-proximal ADMM for convex minimization problems with one strongly convex block, *Asia-Pacific Journal of Operational Research*, **32**(3), 1550024 (19 pages) (2015).
- [23] X. D. Li, D. F. Sun and K.-C. Toh, A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions, accepted by *Mathematical Programming, Series A*, DOI: 10.1007/s10107-014-0850-5 (2014).
- [24] X. D. Li, A two-phase augmented Lagrangian method for convex composite quadratic programming, PhD thesis, Department of Mathematics, National University of Singapore, 2015.
- [25] T. Y. Lin, S. Q. Ma and S. Z. Zhang, On the convergence rate of multi-block ADMM, *arXiv:1408.4265v1*, (2014).
- [26] T. Y. Lin, S. Q. Ma and S. Z. Zhang, On the global linear convergence of the ADMM with multi-block variables, *arXiv:1408.4266v1*, (2014).
- [27] R. D. C. Monteiro and B. F. Svaiter, Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers, *SIAM Journal on Optimization*, **23**(1), pp. 475–507 (2013).
- [28] R. D. C. Monteiro and B. F. Svaiter, On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean, *SIAM Journal on Optimization*, **20**(6), pp. 2755–2787 (2010).
- [29] A. Nemirovski, Prox-method with rate of convergence  $O(1/t)$  for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems, *SIAM Journal on Optimization*, **15**(1), pp. 229–251 (2004).
- [30] Y. Nesterov, Gradient methods for minimizing composite objective function, *Mathematical Programming, Ser. B*, **140**(1), pp. 125–161 (2013).
- [31] Y. Ouyang, Y. Chen, G. Lan and E. Pasiliao Jr., An accelerated linearized alternating direction method of multipliers, *SIAM Journal on Imaging Sciences*, **8**(1), pp. 644–681 (2015).
- [32] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, Princeton (1970).

- [33] D. F. Sun, K.-C. Toh and L. Yang, A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type constraints, *SIAM Journal on Optimization*, **25**(2), pp. 882–915 (2015).
- [34] M. Tao and X. M. Yuan, On the  $O(1/t)$  convergence rate of alternating direction method with logarithmic-quadratic proximal regularization, *SIAM Journal on Optimization*, **22**, pp. 1431–1448 (2012).
- [35] H. Wang and A. Banerjee, Bregman alternating direction method of multipliers, arXiv:1306.3203v3, (2014).