A MAJORIZED ADMM WITH INDEFINITE PROXIMAL TERMS FOR LINEARLY CONSTRAINED CONVEX COMPOSITE OPTIMIZATION[*]

MIN LI[†], DEFENG SUN[‡], AND KIM-CHUAN TOH[§]

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 nonsmooth 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 the nonergodic sense 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.

Key words. alternating direction method of multipliers, convex composite optimization, indefinite proximal terms, majorization, iteration-complexity

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

DOI. 10.1137/140999025

1. Introduction. We consider the following 2-block convex composite optimization problem:

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

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 $\| \cdot \|$; $p : \mathcal{X} \to (-\infty, +\infty]$ and $q : \mathcal{Y} \to (-\infty, +\infty]$ are two closed proper convex (not necessarily smooth) functions; $f : \mathcal{X} \to (-\infty, +\infty)$ and $g : \mathcal{Y} \to (-\infty, +\infty)$ are two convex functions with Lipschitz continuous gradients on $\mathcal{X}$ and $\mathcal{Y}$, respectively; $A^* : \mathcal{X} \to \mathcal{Z}$ and $B^* : \mathcal{Y} \to \mathcal{Z}$ are the adjoints of the linear operators $A : \mathcal{Z} \to \mathcal{X}$ and $B : \mathcal{Z} \to \mathcal{Y}$, 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:

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

where $\tau \in (0, 2)$ guarantees the convergence. Due to the nonseparability 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 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):

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

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, 14, 17, 12, 13, 9, 8]. For a recent survey, see [10].

By noting the fact that the subproblems in (3) may be difficult to solve and that in many applications f or g is a convex quadratic function, Fazel et al. [11] advocated the use of the following semiproximal ADMM scheme:

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

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. When S and T are two self-adjoint and positive definite linear operators, the above semiproximal ADMM with $\tau = 1$ reduces to the proximal ADMM essentially proposed by Eckstein [8]. He et al. [20] extended the work of Eckstein [8] to monotone variational inequalities to allow σ, S, and T to vary at different iterations. We refer the reader to [11] as well as [7] for a brief history on the development of the semiproximal 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 τ is chosen to be as small as 10^{-8}. This seems to suggest that one has to give up the direct extension of the m-block ($m \geq 3$) ADMM unless one is willing to take a sufficiently small step-length τ as was shown by Hong and Luo in [22] or to take a small penalty parameter σ if at least $m - 2$ blocks in the objective are strongly convex [19, 2, 28, 27, 23]. On the other hand, despite the potential divergence, the directly extended m-block ADMM with $\tau \geq 1$ and an appropriate choice of σ often works very well in practice.

Recently, there has been exciting progress in designing convergent and efficient ADMM-type methods for solving multiblock linear and convex quadratic semidefinite programming problems [34, 25]. The convergence proof of the methods presented in [34] and [25] is via establishing their equivalence to particular cases of the general 2-block semiproximal ADMM considered in [11]. It is this important fact that inspires us to extend the 2-block semiproximal ADMM in [11] to a majorized ADMM with indefinite proximal terms (which we call majorized iPADMM) in this paper. Our new algorithm has two important aspects. First, 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 [36] under the more general setting of Bregman distance functions. The drawback of the Bregman distance function–based ADMM discussed in [36] is that the parameter τ should be small for the global convergence. For example, if we choose the Euclidean distance as the Bregman divergence, then the corresponding parameter τ should be smaller than 1. By focusing on the Euclidean divergence instead of the more general Bregman divergence, we allow τ to stay in the larger interval $(0, (1 + \sqrt{5})/2)$. Related works on using the majorization techniques can also be found in [15, 26]. Second, 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 evidence 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 [34].

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. To the best of our knowledge, 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 τ or the penalty parameter σ.

Besides proving the global convergence of our proposed majorized indefinite proximal ADMM, we also establish its nonergodic iteration-complexity, but not the ergodic iteration-complexity for the sake of saving space. The study of the ergodic iteration-complexity of the classical ADMM is inspired by Nemirovski [31], who proposed a prox-method with $O(1/k)$ iteration-complexity for variational inequalities. Monteiro and Svaiter [29] 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 [30]. He and Yuan [21] provided a simple $O(1/k)$ ergodic iteration-complexity result defined on compact sets for a special semiproximal ADMM scheme (where the x-part uses a semiproximal term, while the y-part does not). Tao and Yuan [35] proved the $O(1/k)$ ergodic iteration-complexity (in terms of [21]) of the ADMM with a logarithmic-quadratic proximal regularization even for $\tau \in (0, (1 + \sqrt{5})/2)$. Ouyang et al. [32] provided an ergodic iteration-complexity for an accelerated linearized ADMM. Wang and Banerjee [36] generalized the ADMM to the 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 of the generated sequence. In section 4, we present the global convergence and the choices of proximal terms. The analysis of the nonergodic iteration-complexity 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 $h\colon \mathcal{X} \to (-\infty, +\infty]$ is defined as $\mathrm{dom}(h) := \{x \in \mathcal{X} \mid h(x) < +\infty\}$.
- The set of all relative interior points of a convex set C is denoted by $\mathrm{ri}(C)$.
- 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 $M : \mathcal{X} \to \mathcal{X}$ 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 1. *Both $f(\cdot)$ and $g(\cdot)$ are smooth convex functions with Lipschitz continuous gradients.*

Under Assumption 1, we know that there exist two self-adjoint and positive semidefinite linear operators Σ_f and Σ_g such that for any $x, x' \in \mathcal{X}$ and any $y, y' \in \mathcal{Y}$,

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

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

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

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

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

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.,

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

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

$$(9) \qquad \widehat{\Sigma}_g \succeq W \succeq \Sigma_g \succeq 0.$$

For further discussions, we need the following constraint qualification.

ASSUMPTION 2. *There exists* $(x_0, y_0) \in \mathrm{ri}(\mathrm{dom}(p) \times \mathrm{dom}(q)) \bigcap P$, *where* $P :=$ $\{(x, y) \in \mathcal{X} \times \mathcal{Y} \mid A^* x + B^* y = c\}$.

Under Assumption 2, it follows from [33, Corollaries 28.2.2 and 28.3.1] that $(\bar{x}, \bar{y}) \in \mathrm{dom}(p) \times \mathrm{dom}(q)$ 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:

$$(10) \quad 0 \in \partial p(\bar{x}) + \nabla f(\bar{x}) + A\bar{z}, \quad 0 \in \partial q(\bar{y}) + \nabla g(\bar{y}) + B\bar{z}, \quad c - A^*\bar{x} - B^*\bar{y} = 0,$$

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 (10) is an optimal solution to the dual of problem (1). Therefore, we call $(\widehat{x}, \widehat{y}, \widehat{z}) \in \mathrm{dom}(p) \times \mathrm{dom}(q) \times \mathcal{Z}$ an ε-approximate KKT point of (1) if it satisfies

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

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, (10) is equivalent to finding a vector $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W} := \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ such that for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have

$$(11) \quad \begin{cases} p(x) - p(\bar{x}) + \langle x - \bar{x}, \nabla f(\bar{x}) + A\bar{z}\rangle \geq 0, \\ q(y) - q(\bar{y}) + \langle y - \bar{y}, \nabla g(\bar{y}) + B\bar{z}\rangle \geq 0, \\ c - A^*\bar{x} - B^*\bar{y} = 0 \end{cases}$$

or, equivalently,

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

which is obtained by using the assumption that f and g are smooth convex functions. We denote by $\mathcal{W}^*$ the solution set of (11), which is nonempty under Assumption 2 and the fact that the solution set of problem (1) is assumed to be nonempty.

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

LEMMA 3. *If a sequence* $\{a_i\} \subseteq \Re$ *obeys* (1) $a_i \geq 0$ *and* (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

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

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

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

$$(13) \qquad \mathcal{L}(x, y; z) := p(x) + f(x) + q(y) + g(y) + \langle z, A^* x + B^* y - c\rangle$$

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:

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

where the two majorized convex functions $\hat{f}$ and $\hat{g}$ 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 $(x^0, y^0, z^0) \in \mathrm{dom}(p) \times \mathrm{dom}(q) \times \mathcal{Z}$. Set $k = 0$, and denote $\hat{r}^0 := A^*x^0 + B^*y^0 - c + \sigma^{-1}z^0$.

Step 1. Compute

$$(14) \quad \begin{cases} x^{k+1} := \underset{x \in \mathcal{X}}{\mathrm{argmin}} \ \widehat{\mathcal{L}}_\sigma(x, y^k; (z^k, x^k, y^k)) + \frac{1}{2}\|x - x^k\|_S^2 \\ \qquad = \underset{x \in \mathcal{X}}{\mathrm{argmin}} \ p(x) + \frac{1}{2}\langle x, \mathcal{P}x \rangle + \langle \nabla f(x^k) + \sigma A\hat{r}^k - \mathcal{P}x^k, \ x \rangle, \\ y^{k+1} := \underset{y \in \mathcal{Y}}{\mathrm{argmin}} \ \widehat{\mathcal{L}}_\sigma(x^{k+1}, y; (z^k, x^k, y^k)) + \frac{1}{2}\|y - y^k\|_T^2 \\ \qquad = \underset{y \in \mathcal{Y}}{\mathrm{argmin}} \ q(y) + \frac{1}{2}\langle y, \mathcal{Q}y \rangle + \langle \nabla g(y^k) + \sigma B \\ \qquad\qquad\qquad \times (\hat{r}^k + A^*(x^{k+1} - x^k)) - \mathcal{Q}y^k, \ y \rangle, \\ z^{k+1} := z^k + \tau\sigma(A^*x^{k+1} + B^*y^{k+1} - c). \end{cases}$$

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

Remark 4. 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 (14). Second, they play an important role in ensuring the boundedness of the two generated sequences $\{x^{k+1}\}$ and $\{y^{k+1}\}$. Third, as demonstrated in [25], 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 but 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

$$(15) \quad H_f := \frac{1}{2}\Sigma_f + S + \frac{1}{2}(1-\alpha)\sigma AA^*, \quad M_g := \frac{1}{2}\Sigma_g + T + \min(\tau,\ 1+\tau-\tau^2)\alpha\sigma BB^*,$$

and for $(x, y, z) \in \mathcal{W}$, $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W}^*$, $k = 0, 1, \ldots$, define

$$(16) \quad \begin{cases} \phi_k(x, y, z) := (\tau\sigma)^{-1}\|z^k - z\|^2 + \|x^k - x\|_{\widehat{\Sigma}_f + S}^2 + \|y^k - y\|_{\widehat{\Sigma}_g + T + \sigma BB^*}^2, \\ \bar{\phi}_k := \phi_k(\bar{x}, \bar{y}, \bar{z}) = (\tau\sigma)^{-1}\|z^k - \bar{z}\|^2 + \|x^k - \bar{x}\|_{\widehat{\Sigma}_f + S}^2 \\ \qquad\qquad\qquad\qquad + \|y^k - \bar{y}\|_{\widehat{\Sigma}_g + T + \sigma BB^*}^2, \end{cases}$$

$$(17) \quad \begin{cases} \xi_{k+1} := \|y^{k+1} - y^k\|_{\widehat{\Sigma}_g + T}^2, \\ s_{k+1} := \|x^{k+1} - x^k\|_{\frac{1}{2}\Sigma_f + S}^2 + \|y^{k+1} - y^k\|_{\frac{1}{2}\Sigma_g + T}^2, \\ t_{k+1} := \|x^{k+1} - x^k\|_{H_f}^2 + \|y^{k+1} - y^k\|_{M_g}^2, \end{cases}$$

and

$$(18) \qquad\qquad r^k := A^*x^k + B^*y^k - c, \qquad \tilde{z}^{k+1} := z^k + \sigma r^{k+1}.$$

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

LEMMA 5. (a) *For any vectors* u_1, u_2, v_1, v_2 *in the same Euclidean vector space* $\mathcal{X}$, *we have the identity*

$$(19) \quad \langle u_1 - u_2, v_1 - v_2 \rangle = \frac{1}{2}(\|v_2 - u_1\|^2 - \|v_1 - u_1\|^2) + \frac{1}{2}(\|v_1 - u_2\|^2 - \|v_2 - u_2\|^2).$$

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

$$(20) \qquad \langle u, Gv \rangle = \frac{1}{2}\big(\|u\|_G^2 + \|v\|_G^2 - \|u - v\|_G^2\big) = \frac{1}{2}\big(\|u + v\|_G^2 - \|u\|_G^2 - \|v\|_G^2\big).$$

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

LEMMA 6. *Suppose that Assumption 1 holds. Let* $\{z^{k+1}\}$ *be generated by* (14), *and let* $\{\tilde{z}^{k+1}\}$ *be defined by* (18). *Then for any* $z \in \mathcal{Z}$ *and* $k \geq 0$ *we have*

$$\frac{1}{\sigma}\langle z - \tilde{z}^{k+1}, \tilde{z}^{k+1} - z^k \rangle + \frac{1}{2\sigma}\|z^k - \tilde{z}^{k+1}\|^2$$

$$(21) \qquad = \frac{1}{2\tau\sigma}(\|z^k - z\|^2 - \|z^{k+1} - z\|^2) + \frac{\tau - 1}{2\sigma}\|z^k - \tilde{z}^{k+1}\|^2.$$

Proof. From (14) and (18), we get

$$(22) \qquad\qquad z^{k+1} - z^k = \tau\sigma r^{k+1} \quad \text{and} \quad \tilde{z}^{k+1} - z^k = \sigma r^{k+1}.$$

It follows from (22) that

$$(23) \qquad \tilde{z}^{k+1} - z^k = \frac{1}{\tau}(z^{k+1} - z^k) \quad \text{and} \quad z^{k+1} - \tilde{z}^{k+1} = -(\tau - 1)(z^k - \tilde{z}^{k+1}).$$

By using the first equation in (23), we obtain

$$
\frac{1}{\sigma}\langle z - \tilde{z}^{k+1}, \tilde{z}^{k+1} - z^k\rangle + \frac{1}{2\sigma}\|z^k - \tilde{z}^{k+1}\|^2
$$

(24)
$$
= \frac{1}{\tau\sigma}\langle z - \tilde{z}^{k+1}, z^{k+1} - z^k\rangle + \frac{1}{2\sigma}\|z^k - \tilde{z}^{k+1}\|^2.
$$

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 (19) to the first term of the right-hand side of (24), we obtain

$$
\frac{1}{\sigma}\langle z - \tilde{z}^{k+1}, \tilde{z}^{k+1} - z^k\rangle + \frac{1}{2\sigma}\|z^k - \tilde{z}^{k+1}\|^2
$$
$$
= \frac{1}{2\tau\sigma}\left(\|z^k - z\|^2 - \|z^{k+1} - z\|^2\right) + \frac{1}{2\tau\sigma}\left(\|z^{k+1} - \tilde{z}^{k+1}\|^2 - \|z^k - \tilde{z}^{k+1}\|^2\right)
$$
(25)
$$
+ \tau\|z^k - \tilde{z}^{k+1}\|^2\Big).
$$

By using the second equation in (23), we have

$$
\|z^{k+1} - \tilde{z}^{k+1}\|^2 - \|z^k - \tilde{z}^{k+1}\|^2 + \tau\|z^k - \tilde{z}^{k+1}\|^2
$$
$$
= (\tau - 1)^2\|z^k - \tilde{z}^{k+1}\|^2 - \|z^k - \tilde{z}^{k+1}\|^2 + \tau\|z^k - \tilde{z}^{k+1}\|^2 \;=\; \tau(\tau - 1)\|z^k - \tilde{z}^{k+1}\|^2,
$$

which, together with (25), proves the assertion (21). $\qquad\square$

LEMMA 7. *Suppose that Assumption 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 ξ_k and r^k be defined as in* (17) *and* (18), *respectively. Then for any $k \geq 1$, we have*

$$
(1 - \tau)\sigma\|r^{k+1}\|^2 + \sigma\|A^*x^{k+1} + B^*y^k - c\|^2
$$
$$
\geq \max\left(1 - \tau, 1 - \tau^{-1}\right)\sigma\left(\|r^{k+1}\|^2 - \|r^k\|^2\right) + \left(\xi_{k+1} - \xi_k\right)
$$
(26)
$$
+ \min\left(\tau, 1 + \tau - \tau^2\right)\sigma\left(\tau^{-1}\|r^{k+1}\|^2 + \|B^*(y^{k+1} - y^k)\|^2\right).
$$

Proof. Note that

$$
(1 - \tau)\sigma\|r^{k+1}\|^2 + \sigma\|A^*x^{k+1} + B^*y^k - c\|^2
$$
$$
= (1 - \tau)\sigma\|r^{k+1}\|^2 + \sigma\|(A^*x^{k+1} + B^*y^{k+1} - c) + B^*(y^k - y^{k+1})\|^2
$$
(27)
$$
= (2 - \tau)\sigma\|r^{k+1}\|^2 + \sigma\|B^*(y^{k+1} - y^k)\|^2 + 2\sigma\langle B^*(y^k - y^{k+1}), r^{k+1}\rangle.
$$

First, we shall estimate the term $2\sigma\langle B^*(y^k - y^{k+1}), r^{k+1}\rangle$ in (27). From the first-order optimality condition of (14) and the notation of $\tilde{z}^{k+1}$ defined in (18), we have

(28)
$$
-\nabla g(y^k) - B\tilde{z}^{k+1} - (\widehat{\Sigma}_g + T)(y^{k+1} - y^k) \in \partial q(y^{k+1}),
$$
$$
-\nabla g(y^{k-1}) - B\tilde{z}^k - (\widehat{\Sigma}_g + T)(y^k - y^{k-1}) \in \partial q(y^k).
$$

From Clarke's mean value theorem [3, Proposition 2.6.5], we know that

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

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

(29)
$$
\nabla g(y^k) - \nabla g(y^{k-1}) = W^k(y^k - y^{k-1}).
$$

From (28) 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 (29), gives rise to

$$2\langle y^k - y^{k+1}, B\tilde{z}^{k+1} - B\tilde{z}^k \rangle$$
$$\geq 2\langle \nabla g(y^k) - \nabla g(y^{k-1}), y^{k+1} - y^k \rangle - 2\langle (\widehat{\Sigma}_g + T)(y^k - y^{k-1}), y^{k+1} - y^k \rangle$$
$$+ 2\|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g + T}$$
$$= 2\langle W^k(y^k - y^{k-1}), y^{k+1} - y^k \rangle - 2\langle (\widehat{\Sigma}_g + T)(y^k - y^{k-1}), y^{k+1} - y^k \rangle$$
$$+ 2\|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g + T}$$
$$(30) \qquad = 2\langle (\widehat{\Sigma}_g - W^k + T)(y^{k-1} - y^k), y^{k+1} - y^k \rangle + 2\|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g + T}.$$

By using the first elementary identity in (20) and $W^k \succeq 0$, we have

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

From (9), we know that $\widehat{\Sigma}_g - \frac{1}{2}W^k + T = \frac{1}{2}\widehat{\Sigma}_g + \frac{1}{2}(\widehat{\Sigma}_g - W^k) + T \succeq \frac{1}{2}\widehat{\Sigma}_g + T \succeq 0$. Then, by using the elementary inequality $\|u + v\|^2_G \leq 2\|u\|^2_G + 2\|v\|^2_G$ for any self-adjoint and positive semidefinite linear operator G, we get

$$-\|y^{k+1} - y^{k-1}\|^2_{\widehat{\Sigma}_g - \frac{1}{2}W^k + T} = -\|(y^{k+1} - y^k) + (y^k - y^{k-1})\|^2_{\widehat{\Sigma}_g - \frac{1}{2}W^k + T}$$
$$\geq -2\Big(\|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g - \frac{1}{2}W^k + T} + \|y^k - y^{k-1}\|^2_{\widehat{\Sigma}_g - \frac{1}{2}W^k + T} \Big).$$

Substituting the above inequalities into (30), we obtain

$$2\langle y^k - y^{k+1}, B\tilde{z}^{k+1} - B\tilde{z}^k \rangle$$
$$\geq -\|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g + T} - \|y^k - y^{k-1}\|^2_{\widehat{\Sigma}_g + T} + 2\|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g + T}$$
$$(31) \qquad = \|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g + T} - \|y^k - y^{k-1}\|^2_{\widehat{\Sigma}_g + T}.$$

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 (14) and (18)) and (31), we have

$$2\sigma\langle B^*(y^k - y^{k+1}), r^{k+1} \rangle = 2\mu_{k+1} + 2\langle y^k - y^{k+1}, B\tilde{z}^{k+1} - B\tilde{z}^k \rangle$$
$$(32) \qquad \geq 2\mu_{k+1} + \|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g + T} - \|y^k - y^{k-1}\|^2_{\widehat{\Sigma}_g + T}.$$

Since $\tau \in (0, +\infty)$, from the definition of μ_{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 (27), (32), and the notation of ξ_k, shows that (26) holds. This completes the proof. □

PROPOSITION 8. *Suppose that Assumption* 1 *holds. Let* $\{(x^k, y^k, z^k)\}$ *be generated by the majorized iPADMM. For each* k *and* $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W}^*$, *let* $\bar{\phi}_k$, ξ_{k+1}, s_{k+1}, t_{k+1}, r^k, *and* $\tilde{z}^{k+1}$ *be defined as in* (16), (17), *and* (18). *Then, for any* $\tau \in (0, +\infty)$ *and* $k \geq 0$, *we have*

$$(33) \qquad \bar{\phi}_k - \bar{\phi}_{k+1} \geq s_{k+1} + (1-\tau)\sigma\|r^{k+1}\|^2 + \sigma\|A^*x^{k+1} + B^*y^k - c\|^2.$$

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

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

Combining the above two inequalities, we obtain

$$(34) \quad f(x) - f(x^{k+1}) - \frac{1}{2}\|x^k - x\|_{\Sigma_f}^2 + \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 \geq \langle x - x^{k+1}, \nabla f(x^k)\rangle \quad \forall x \in \mathcal{X}.$$

From the first-order optimality condition of (14), for any $x \in \mathcal{X}$ we have

$$p(x) - p(x^{k+1}) + \langle x - x^{k+1}, \nabla f(x^k) + A[z^k + \sigma(A^*x^{k+1} + B^*y^k - c)]$$
$$(35) \qquad\qquad + (\widehat{\Sigma}_f + S)(x^{k+1} - x^k)\rangle \geq 0.$$

Adding (34) and (35), for any $x \in \mathcal{X}$ we get

$$\big(p(x) + f(x)\big) - \big(p(x^{k+1}) + f(x^{k+1})\big)$$
$$+ \big\langle x - x^{k+1}, A[z^k + \sigma(A^*x^{k+1} + B^*y^k - c)] + (\widehat{\Sigma}_f + S)(x^{k+1} - x^k)\big\rangle$$
$$(36) \qquad \geq \frac{1}{2}\|x^k - x\|_{\Sigma_f}^2 - \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2.$$

Using a derivation similar to that used to obtain (36), for any $y \in \mathcal{Y}$ we have

$$\big(q(y) + g(y)\big) - \big(q(y^{k+1}) + g(y^{k+1})\big)$$
$$+ \big\langle y - y^{k+1}, B[z^k + \sigma(A^*x^{k+1} + B^*y^{k+1} - c)] + (\widehat{\Sigma}_g + T)(y^{k+1} - y^k)\big\rangle$$
$$(37) \qquad \geq \frac{1}{2}\|y^k - y\|_{\Sigma_g}^2 - \frac{1}{2}\|y^{k+1} - y^k\|_{\widehat{\Sigma}_g}^2.$$

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 (36) and (37), setting $x = \bar{x}$ and $y = \bar{y}$, and using the above equation, we have

$$\big(p(\bar{x}) + f(\bar{x}) + q(\bar{y}) + g(\bar{y})\big) - \big(p(x^{k+1}) + f(x^{k+1}) + q(y^{k+1}) + g(y^{k+1})\big)$$
$$+ \big\langle \bar{x} - x^{k+1}, A\tilde{z}^{k+1} + \sigma AB^*(y^k - y^{k+1}) + (\widehat{\Sigma}_f + S)(x^{k+1} - x^k)\big\rangle$$
$$+ \big\langle \bar{y} - y^{k+1}, B\tilde{z}^{k+1} + (\widehat{\Sigma}_g + T)(y^{k+1} - y^k)\big\rangle$$
$$(38) \qquad \geq \frac{1}{2}\|x^k - \bar{x}\|_{\Sigma_f}^2 - \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 + \frac{1}{2}\|y^k - \bar{y}\|_{\Sigma_g}^2 - \frac{1}{2}\|y^{k+1} - y^k\|_{\widehat{\Sigma}_g}^2.$$

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

$$(39) \qquad f(x^{k+1}) - f(\bar{x}) + \langle \bar{x} - x^{k+1}, \nabla f(\bar{x}) \rangle \geq \frac{1}{2}\|x^{k+1} - \bar{x}\|_{\Sigma_f}^2,$$

$$(40) \qquad g(y^{k+1}) - g(\bar{y}) + \langle \bar{y} - y^{k+1}, \nabla g(\bar{y}) \rangle \geq \frac{1}{2}\|y^{k+1} - \bar{y}\|_{\Sigma_g}^2.$$

Let

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

Adding (38), (39), and (40), 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

$$\big(p(\bar{x}) + q(\bar{y})\big) - \big(p(x^{k+1}) + q(y^{k+1})\big) + \langle \bar{x} - x^{k+1}, \nabla f(\bar{x}) + A\bar{z} \rangle$$
$$+ \langle \bar{y} - y^{k+1}, \nabla g(\bar{y}) + B\bar{z} \rangle + \bar{\Delta}^k + \langle \bar{x} - x^{k+1}, (\widehat{\Sigma}_f + S)(x^{k+1} - x^k) \rangle$$
$$+ \langle \bar{y} - y^{k+1}, (\widehat{\Sigma}_g + T)(y^{k+1} - y^k) \rangle$$
$$\geq \frac{1}{2}\|x^k - \bar{x}\|_{\Sigma_f}^2 + \frac{1}{2}\|x^{k+1} - \bar{x}\|_{\Sigma_f}^2 + \frac{1}{2}\|y^k - \bar{y}\|_{\Sigma_g}^2 + \frac{1}{2}\|y^{k+1} - \bar{y}\|_{\Sigma_g}^2$$
$$- \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 - \frac{1}{2}\|y^{k+1} - y^k\|_{\widehat{\Sigma}_g}^2$$
$$(41) \quad \geq \frac{1}{4}\|x^{k+1} - x^k\|_{\Sigma_f}^2 + \frac{1}{4}\|y^{k+1} - y^k\|_{\Sigma_g}^2 - \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 - \frac{1}{2}\|y^{k+1} - y^k\|_{\widehat{\Sigma}_g}^2.$$

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

$$\big(p(x^{k+1}) + q(y^{k+1})\big) - \big(p(\bar{x}) + q(\bar{y})\big)$$
$$(42) \qquad + \langle x^{k+1} - \bar{x}, \nabla f(\bar{x}) + A\bar{z} \rangle + \langle y^{k+1} - \bar{y}, \nabla g(\bar{y}) + B\bar{z} \rangle \geq 0.$$

Adding (41) and (42), we get

$$\bar{\Delta}^k + \langle \bar{x} - x^{k+1}, (\widehat{\Sigma}_f + S)(x^{k+1} - x^k) \rangle + \langle \bar{y} - y^{k+1}, (\widehat{\Sigma}_g + T)(y^{k+1} - y^k) \rangle$$
$$(43) \quad \geq \frac{1}{4}\|x^{k+1} - x^k\|_{\Sigma_f}^2 + \frac{1}{4}\|y^{k+1} - y^k\|_{\Sigma_g}^2 - \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 - \frac{1}{2}\|y^{k+1} - y^k\|_{\widehat{\Sigma}_g}^2.$$

By simple manipulations, and using the definition of $\bar{\Delta}^k$ and $A^*\bar{x} + B^*\bar{y} - c = 0$, we have

$$\bar{\Delta}^k = \langle A^*(x^{k+1} - \bar{x}), \bar{z} - \tilde{z}^{k+1} + \sigma B^*(y^{k+1} - y^k) \rangle + \langle B^*(y^{k+1} - \bar{y}), \bar{z} - \tilde{z}^{k+1} \rangle$$
$$= -\langle A^*\bar{x} + B^*\bar{y} - c, \bar{z} - \tilde{z}^{k+1} \rangle + \langle r^{k+1}, \bar{z} - \tilde{z}^{k+1} \rangle$$
$$+ \sigma \langle A^*(x^{k+1} - \bar{x}), B^*(y^{k+1} - y^k) \rangle$$
$$(44) \quad = \langle r^{k+1}, \bar{z} - \tilde{z}^{k+1} \rangle + \sigma \langle A^*(x^{k+1} - \bar{x}), B^*(y^{k+1} - y^k) \rangle$$

and

$$\sigma \langle A^*(x^{k+1} - \bar{x}), B^*(y^{k+1} - y^k) \rangle$$
$$(45) \qquad = \sigma \langle (-B^*\bar{y}) - (A^*x^{k+1} - c), (-B^*y^{k+1}) - (-B^*y^k) \rangle.$$

Now, by taking $u_1 = -B^*\bar{y}$, $u_2 = A^*x^{k+1} - c$, $v_1 = -B^*y^{k+1}$, and $v_2 = -B^*y^k$ and applying the identity (19) to the right-hand side of (45), we obtain

$$\sigma\langle A^*(x^{k+1} - \bar{x}),\ B^*(y^{k+1} - y^k)\rangle = \frac{\sigma}{2}\big(\|B^*y^k - B^*\bar{y}\|^2 - \|B^*y^{k+1} - B^*\bar{y}\|^2\big)$$
$$+ \frac{\sigma}{2}\big(\|A^*x^{k+1} + B^*y^{k+1} - c\|^2 - \|A^*x^{k+1} + B^*y^k - c\|^2\big).$$

Substituting this into (44) and using the definition of $\tilde{z}^{k+1}$, we have

$$\bar{\Delta}^k = \frac{1}{\sigma}\langle \bar{z} - \tilde{z}^{k+1}, \tilde{z}^{k+1} - z^k\rangle + \frac{1}{2\sigma}\|z^k - \tilde{z}^{k+1}\|^2 - \frac{\sigma}{2}\|A^*x^{k+1} + B^*y^k - c\|^2$$
$$(46) \qquad + \frac{\sigma}{2}\big(\|B^*y^k - B^*\bar{y}\|^2 - \|B^*y^{k+1} - B^*\bar{y}\|^2\big).$$

Setting $z = \bar{z}$ in (21), and applying it to (46), we get

$$\bar{\Delta}^k = \frac{1}{2\tau\sigma}\big(\|z^k - \bar{z}\|^2 - \|z^{k+1} - \bar{z}\|^2\big) + \frac{\tau - 1}{2\sigma}\|z^k - \tilde{z}^{k+1}\|^2$$
$$(47) \qquad - \frac{\sigma}{2}\|A^*x^{k+1} + B^*y^k - c\|^2 + \frac{\sigma}{2}\big(\|B^*y^k - B^*\bar{y}\|^2 - \|B^*y^{k+1} - B^*\bar{y}\|^2\big).$$

Using the second elementary identity in (20), we obtain that

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

Substituting this and (47) into (43), and using the definitions of $\tilde{z}^{k+1}$ and r^{k+1}, we have

$$\frac{1}{2\tau\sigma}\big(\|z^k - \bar{z}\|^2 - \|z^{k+1} - \bar{z}\|^2\big) + \frac{1}{2}\big(\|x^k - \bar{x}\|^2_{\widehat{\Sigma}_f + S} - \|x^{k+1} - \bar{x}\|^2_{\widehat{\Sigma}_f + S}\big)$$
$$+ \frac{1}{2}\big(\|y^k - \bar{y}\|^2_{\widehat{\Sigma}_g + T + \sigma BB^*} - \|y^{k+1} - \bar{y}\|^2_{\widehat{\Sigma}_g + T + \sigma BB^*}\big)$$
$$\geq \frac{1}{2}\Big((1 - \tau)\sigma\|r^{k+1}\|^2 + \sigma\|A^*x^{k+1} + B^*y^k - c\|^2 + \|x^{k+1} - x^k\|^2_{\frac{1}{2}\Sigma_f + S}$$
$$(48) \qquad + \|y^{k+1} - y^k\|^2_{\frac{1}{2}\Sigma_g + T}\Big).$$

Using the notation in (16) and (17), we get (33) from (48) immediately. The proof is complete. □

Remark 9. 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 (48) cancel out, from (48) and the above equation we can easily get

$$\frac{1}{\tau\sigma}(\|z^k - \bar{z}\|^2 - \|z^{k+1} - \bar{z}\|^2) + \big(\|x^k - \bar{x}\|^2_{\widehat{\Sigma}_f + S} - \|x^{k+1} - \bar{x}\|^2_{\widehat{\Sigma}_f + S}\big)$$
$$(49) \qquad \geq (2 - \tau)\sigma\|r^{k+1}\|^2 + \|x^{k+1} - x^k\|^2_{\frac{1}{2}\Sigma_f + S}.$$

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 10. *Suppose that Assumptions* 1 *and* 2 *hold. Let* H_f *and* M_g *be defined by* (15). *Let* $\{(x^k, y^k, z^k)\}$ *be generated by the majorized iPADMM. For each* k *and* $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W}^*$, *let* $\bar{\phi}_k$, ξ_{k+1}, s_{k+1}, t_{k+1}, r^k, *and* $\tilde{z}^{k+1}$ *be defined as in* (16), (17), *and* (18). *Then the following results hold:*

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

$$\left(\overline{\phi}_k + \beta\sigma\|r^k\|^2\right) - \left(\overline{\phi}_{k+1} + \beta\sigma\|r^{k+1}\|^2\right)$$

$$\geq \left(\frac{1}{\tau^2\sigma}\|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|^2_{\widehat{\Sigma}_f + S + \eta\sigma AA^*} + \|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g + T + \eta\sigma BB^*}\right)$$

$$(50) \quad -\left(\frac{\tau + \beta}{\tau^2\sigma}\|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|^2_{\widehat{\Sigma}_f - \frac{1}{2}\Sigma_f} + \|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g - \frac{1}{2}\Sigma_g + \eta\sigma BB^*}\right),$$

where

$$(51) \qquad\qquad \beta := \frac{\eta(1 - \eta)}{1 - 2\eta}.$$

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

$$(52) \qquad \widehat{\Sigma}_f + S + \eta\sigma AA^* \succ 0 \qquad and \qquad \widehat{\Sigma}_g + T + \eta\sigma BB^* \succ 0,$$

and the following condition holds:

$$(53) \qquad \sum_{k=0}^{\infty} \left(\|x^{k+1} - x^k\|^2_{\widehat{\Sigma}_f} + \|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g + \sigma BB^*} + \|r^{k+1}\|^2\right) < +\infty.$$

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*

$$(54) \qquad\qquad \frac{1}{2}\widehat{\Sigma}_g + T \succeq 0.$$

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

$$\left[\overline{\phi}_k + \left(1 - \alpha\min(\tau, \tau^{-1})\right)\sigma\|r^k\|^2 + \alpha\xi_k\right]$$

$$- \left[\overline{\phi}_{k+1} + \left(1 - \alpha\min(\tau, \tau^{-1})\right)\sigma\|r^{k+1}\|^2 + \alpha\xi_{k+1}\right]$$

$$(55) \qquad \geq t_{k+1} + \left(-\tau + \alpha\min(1 + \tau, 1 + \tau^{-1})\right)\sigma\|r^{k+1}\|^2.$$

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

$$(56) \quad \widehat{\Sigma}_f + S \succeq 0, \qquad H_f \succeq 0, \qquad \frac{1}{2}\Sigma_f + S + \sigma AA^* \succ 0, \qquad and \qquad M_g \succ 0.$$

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

Proof of part (a). Using the definition of r^k and the Cauchy–Schwarz inequality, for a given $\beta > 0$ defined by (51), we get

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

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

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

$$
\geq \frac{1-\tau-\beta}{\tau^2\sigma}\|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|^2_{\frac{1}{2}\Sigma_f + S + \frac{\beta}{1+\beta}\sigma AA^*}
$$

(57)
$$
+ \|y^{k+1} - y^k\|^2_{\frac{1}{2}\Sigma_g + T} + \beta\sigma\left(\|r^{k+1}\|^2 - \|r^k\|^2\right).
$$

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

$$
\frac{1-\tau-\beta}{\tau^2\sigma}\|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|^2_{\frac{1}{2}\Sigma_f + S + \frac{\beta}{1+\beta}\sigma AA^*} + \|y^{k+1} - y^k\|^2_{\frac{1}{2}\Sigma_g + T}
$$

$$
\geq \left(\frac{1}{\tau^2\sigma}\|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|^2_{\widehat{\Sigma}_f + S + \eta\sigma AA^*} + \|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g + T + \eta\sigma BB^*}\right)
$$

$$
- \left(\frac{\tau+\beta}{\tau^2\sigma}\|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|^2_{\widehat{\Sigma}_f - \frac{1}{2}\Sigma_f} + \|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_g - \frac{1}{2}\Sigma_g + \eta\sigma BB^*}\right).
$$

Substituting this and (57) into (33), we get (50).

Now assume that (52) and (53) hold. For any given $\eta \in (0, 1/2)$, using the definitions of $\overline{\phi}_{k+1}$, r^{k+1}, and β, $A^*\bar{x} + B^*\bar{y} = c$, and the Cauchy–Schwarz inequality, we get

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

(58)
$$
= (\tau\sigma)^{-1}\|z^{k+1} - \bar{z}\|^2 + \|x^{k+1} - \bar{x}\|^2_{\widehat{\Sigma}_f + S + \eta\sigma AA^*} + \|y^{k+1} - \bar{y}\|^2_{\widehat{\Sigma}_g + T + \eta\sigma BB^*}.
$$

Set

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

and

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

From (53) and $z^{k+1} = z^k + \tau\sigma r^{k+1}$, we get $\sum_{k=0}^{\infty} \zeta_k < +\infty$. Since for some $\eta \in (0, 1/2)$

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

it follows from (58) and (50) that

$$(59) \qquad 0 \le \overline{\phi}_{k+1} + \beta\sigma\|r^{k+1}\|^2 \le \overline{\phi}_k + \beta\sigma\|r^k\|^2 + \zeta_k \ \le \ \overline{\phi}_0 + \beta\sigma\|r^0\|^2 + \sum_{j=0}^{k} \zeta_j.$$

Thus the sequence $\{\overline{\phi}_{k+1} + \beta\sigma\|r^{k+1}\|^2\}$ is bounded. From (58), 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 (50), we get for any $k \ge 0$,

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

and hence

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

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

$$(60) \qquad \lim_{k\to\infty} \|r^{k+1}\| \ = \ \lim_{k\to\infty} (\tau\sigma)^{-1}\|z^{k+1} - z^k\| = 0,$$

$$(61) \qquad \lim_{k\to\infty} \|x^{k+1} - x^k\| = 0 \quad \text{and} \quad \lim_{k\to\infty} \|y^{k+1} - y^k\| = 0.$$

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^∞, y^∞) is an optimal solution to problem (1) and z^∞ is a corresponding Lagrange multiplier.

Taking limits on both sides of (36) and (37) along the subsequence $\{(x^{k_i}, y^{k_i}, z^{k_i})\}$, using (60) and (61), we obtain that

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

i.e., $(x^\infty, y^\infty, z^\infty)$ satisfies (12). Hence (x^∞, y^∞) is an optimal solution to problem (1), and z^∞ 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 (59), for any $k \ge k_i$, we have

$$(62) \qquad \phi_{k+1}(x^\infty, y^\infty, z^\infty) + \beta\sigma\|r^{k+1}\|^2 \ \le \ \phi_{k_i}(x^\infty, y^\infty, z^\infty) + \beta\sigma\|r^{k_i}\|^2 + \sum_{j=k_i}^{k} \zeta_j.$$

Note that

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

Therefore, we get

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

Then from (58), we obtain

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

Since $\widehat{\Sigma}_f + S + \eta\sigma AA^* \succ 0$ and $\widehat{\Sigma}_g + T + \eta\sigma BB^* \succ 0$, we also have that $\lim_{k\to\infty} x^k = x^\infty$ and $\lim_{k\to\infty} y^k = y^\infty$. 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). 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

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

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

$$(1-\alpha)\big[s_{k+1} + (1-\tau)\sigma\|r^{k+1}\|^2 + \sigma\|A^*x^{k+1} + B^*y^k - c\|^2\big]$$

$$\geq -(1-\alpha)\tau\sigma\|r^{k+1}\|^2 + (1-\alpha)\|x^{k+1}-x^k\|^2_{\frac{1}{2}\Sigma_f+S+\frac{1}{2}\sigma AA^*}$$

$$\text{(63)} \qquad + (1-\alpha)\|y^{k+1}-y^k\|^2_{\frac{1}{2}\Sigma_g+T} + (1-\alpha)\sigma\big(\|r^{k+1}\|^2 - \|r^k\|^2\big).$$

By using the definition of s_{k+1} and (26) in Lemma 7, for any $\alpha \in (0,1]$, we have

$$\alpha\big[s_{k+1} + (1-\tau)\sigma\|r^{k+1}\|^2 + \sigma\|A^*x^{k+1} + B^*y^k - c\|^2\big]$$

$$\geq \alpha\|x^{k+1}-x^k\|^2_{\frac{1}{2}\Sigma_f+S} + \alpha\|y^{k+1}-y^k\|^2_{\frac{1}{2}\Sigma_g+T}$$

$$+ \max\big(1-\tau, 1-\tau^{-1}\big)\sigma\alpha\big(\|r^{k+1}\|^2 - \|r^k\|^2\big) + \alpha\big(\xi_{k+1} - \xi_k\big)$$

$$\text{(64)} \qquad + \min\big(\tau, 1+\tau-\tau^2\big)\sigma\alpha\big(\tau^{-1}\|r^{k+1}\|^2 + \|B^*(y^{k+1}-y^k)\|^2\big).$$

Adding (63) and (64), we obtain for any $\alpha \in (0,1]$ that

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

$$\geq \|x^{k+1}-x^k\|^2_{H_f} + \|y^{k+1}-y^k\|^2_{M_g} + \big(1-\alpha\min(\tau, \tau^{-1})\big)\sigma\big(\|r^{k+1}\|^2 - \|r^k\|^2\big)$$

$$\text{(65)} \qquad + \alpha\big(\xi_{k+1} - \xi_k\big) + \big(-\tau + \alpha\min(1+\tau, 1+\tau^{-1})\big)\sigma\|r^{k+1}\|^2.$$

Using the notation in (16) and (17), we know from (33) and (65) that

$$
\left[\overline{\phi}_{k+1} + \left(1 - \alpha\min(\tau, \tau^{-1})\right)\sigma\|r^{k+1}\|^2 + \alpha\xi_{k+1}\right]
$$
$$
- \left[\overline{\phi}_k + \left(1 - \alpha\min(\tau,\ \tau^{-1})\right)\sigma\|r^k\|^2 + \alpha\xi_k\right]
$$
$$
(66) \qquad \leq -\left\{t_{k+1} + \left[-\tau + \alpha\min(1+\tau,\ 1+\tau^{-1})\right]\sigma\|r^{k+1}\|^2\right\}
$$

holds. We can get (55) from (66) 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 $\overline{\phi}_{k+1} \geq 0$, $t_{k+1} \geq 0$, $\xi_{k+1} \geq 0$. From (22) and (55), we see immediately that the sequence $\{\overline{\phi}_{k+1} + \xi_{k+1}\}$ is bounded,

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

which, together with (17) and (56), imply that

$$
(68) \qquad \lim_{k\to\infty}\|x^{k+1} - x^k\|_{H_f} = 0, \qquad \lim_{k\to\infty}\|y^{k+1} - y^k\| = 0,
$$

and

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

Thus, from (68) and (69) we obtain that

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

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

$$
(70) \qquad \lim_{k\to\infty}\|x^{k+1} - x^k\| = 0.
$$

By the definition of $\overline{\phi}_{k+1}$, 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

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

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 $\widehat{\Sigma}_f + S + \sigma AA^* \succeq \frac{1}{2}\Sigma_f + S + \sigma AA^* \succ 0$. 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^∞, y^∞) is an optimal solution to problem (1), and z^∞ is a corresponding Lagrange multiplier.

Taking limits on both sides of (36) and (37) along the subsequence $\{(x^{k_i}, y^{k_i}, z^{k_i})\}$, using (67), (68), and (70), we obtain that

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

i.e., $(x^\infty, y^\infty, z^\infty)$ satisfies (12). Thus (x^∞, y^∞) is an optimal solution to problem (1), and z^∞ 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 (62) in part (a), we can apply the inequality (55) with $(\bar{x}, \bar{y}, \bar{z}) = (x^\infty, y^\infty, z^\infty)$ to show that

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

Hence

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

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

and

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

Using the fact that $\widehat{\Sigma}_f + S + \sigma AA^*$ and $\widehat{\Sigma}_g + T + \sigma BB^*$ are both positive definite, we have $\lim_{k \to \infty} x^k = x^\infty$ and $\lim_{k \to \infty} y^k = y^\infty$. 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 complete. $\qquad \square$

Remark 11. In practice, part (a) of Theorem 10 can be applied in a more heuristic way by using any sufficient condition to guarantee (53) 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 12. 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 (63), for any $\alpha \in (0, 1]$, we have

$$(1 - \alpha)\big[(2 - \tau)\sigma\|r^{k+1}\|^2 + \|x^{k+1} - x^k\|^2_{\frac{1}{2}\Sigma_f + S}\big]$$

$$\geq -(1 - \alpha)\tau\sigma\|r^{k+1}\|^2 + (1 - \alpha)\|x^{k+1} - x^k\|^2_{\frac{1}{2}\Sigma_f + S + \frac{1}{2}\sigma AA^*}$$

$$+ (1 - \alpha)\sigma\big(\|r^{k+1}\|^2 - \|r^k\|^2\big).$$

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

$$(2 - \tau)\sigma\|r^{k+1}\|^2 + \|x^{k+1} - x^k\|^2_{\frac{1}{2}\Sigma_f + S}$$

$$\geq \|x^{k+1} - x^k\|^2_{H_f} + (1 - \alpha)\sigma\big(\|r^{k+1}\|^2 - \|r^k\|^2\big) + (2\alpha - \tau)\sigma\|r^{k+1}\|^2.$$

Substituting this into (49), we obtain

$$\Big\{(\tau\sigma)^{-1}\|z^k - \bar{z}\|^2 + \|x^k - \bar{x}\|^2_{\widehat{\Sigma}_f + S} + (1 - \alpha)\sigma\|r^k\|^2\Big\}$$

$$- \Big\{(\tau\sigma)^{-1}\|z^{k+1} - \bar{z}\|^2 + \|x^{k+1} - \bar{x}\|^2_{\widehat{\Sigma}_f + S} + (1 - \alpha)\sigma\|r^{k+1}\|^2\Big\}$$

$$\geq \|x^{k+1} - x^k\|^2_{H_f} + (2\alpha - \tau)\sigma\|r^{k+1}\|^2.$$

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

$$(71) \qquad \mathcal{G} \preceq \mathcal{M} := \sum_{i=1}^l \lambda_i u_i u_i^* + \lambda_l \sum_{i=l+1}^n u_i u_i^* = \lambda_l I + \sum_{i=1}^l (\lambda_i - \lambda_l) u_i u_i^*.$$

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

$$(72) \qquad \mathcal{M}^{-1} = \sum_{i=1}^{l} \lambda_i^{-1} u_i u_i^* + \lambda_l^{-1} \sum_{i=l+1}^{n} u_i u_i^* = \lambda_l^{-1} I + \sum_{i=1}^{l} (\lambda_i^{-1} - \lambda_l^{-1}) u_i u_i^*.$$

Thus, we need only compute the first l eigenpairs (λ_i, u_i), $i = 1, \ldots, l$, of $\mathcal{G}$ for computing $\mathcal{M}$ and $\mathcal{M}^{-1}$.

In the following, we discuss how to choose proximal terms only for the x-part. The discussions for the y-part are similar and are thus omitted here. It follows from (14) that

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

Example 13. $p \not\equiv 0$ and $A \neq 0$. Choose $\alpha \in (\tau / \min(1 + \tau, 1 + \tau^{-1}), 1]$ such that $2\widehat{\Sigma}_f - \Sigma_f \succeq (1 - \alpha)\sigma AA^*$. Define

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

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

$$(73) \qquad \begin{cases} \rho \geq \rho_0 & \text{if } AA^* \succ 0, \\[2mm] \rho \geq \rho_0 \ \text{and} \ \rho > \lambda_{\max}(\widehat{\Sigma}_f - \frac{1}{2}\Sigma_f) & \text{otherwise.} \end{cases}$$

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

$$(74) \qquad\qquad\qquad\qquad \rho = 1.01\rho_0.$$

Choose

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

$$(75) \qquad = \rho I - \widehat{\Sigma}_f - \sigma AA^*,$$

where ρ is defined in (73). Then, S, which obviously may be indefinite, satisfies (56), and

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

One interesting special case is $\widehat{\Sigma}_f = \Sigma_f = Q$ 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

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

Example 14. $p \equiv 0$. Let $\alpha \in (\tau / \min(1 + \tau, 1 + \tau^{-1}), 1]$ such that $2\widehat{\Sigma}_f - \Sigma_f \succeq (1 - \alpha)\sigma AA^*$. Choose $\mathcal{G}$ such that

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

Let $\mathcal{M} \succ 0$ be the majorization of $\mathcal{G}$ as in (71). Choose

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

Certainly, S, which may be indefinite, satisfies (56), and

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

By using (71) and (72), one can compute $\mathcal{P}$ and $\mathcal{P}^{-1}$ at a low cost if l is a small integer number, for example, $1 \le l \le 6$. One special case is $AA^* \succ 0$ and $\widehat{\Sigma}_f = \Sigma_f = Q$ for some self-adjoint linear operator $Q \succeq 0$. By taking $\alpha = 1$, $\mathcal{G} = \frac{1}{2}Q + \sigma AA^*$, and $\mathcal{M}$ to be a majorization of $\mathcal{G}$ as defined in (71), we have

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

Example 15. 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

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

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

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

where

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

and

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

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

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

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 (56) 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

$$\left(\begin{array}{cc} & X \\ X^* & \end{array}\right) \preceq \left(\begin{array}{cc} (XX^*)^{\frac{1}{2}} & \\ & (X^*X)^{\frac{1}{2}} \end{array}\right).$$

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 the 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 reader to Examples 13 and 14 for possible choices of such linear operators in different scenarios.

In Examples 13–15, we list various choices of proximal terms in different situations, where in order for S to be indefinite, f does not need to be strongly convex, nor does A need to be surjective. Nevertheless, the choices are far from exhaustive. For example, if $p \not\equiv 0$, $x := (x_1, \ldots, x_m)$, $p(x) = p_1(x_1)$, 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 multiblock conic optimization problems [25, 24].

5. The analysis of nonergodic iteration-complexity. In this section, we will present the nonergodic iteration-complexity of an ε-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 16. *Assume that Assumptions* 1 *and* 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* $\alpha \in (\tau/\min(1+\tau, 1+\tau^{-1}),\ 1]$,

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

where H_f *and* M_g *are defined in* (15). *Then, we have*

$$\min_{1 \le i \le k} \Big\{ d^2\big(0,\ \partial p(x^{i+1}) + \nabla f(x^{i+1}) + Az^{i+1}\big) + d^2\big(0,\ \partial q(y^{i+1}) + \nabla g(y^{i+1})$$

$$(76) \qquad\qquad + Bz^{i+1}\big) + \|A^*x^{i+1} + B^*y^{i+1} - c\|^2 \Big\} = o(1/k)$$

and

$$(77) \quad \min_{1 \le i \le k} \Big| \big(p(x^i) + f(x^i) + q(y^i) + g(y^i)\big) - \big(p(\bar{x}) + f(\bar{x}) + q(\bar{y}) + g(\bar{y})\big) \Big| = o(1/\sqrt{k}),$$

where $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{W}^*$.

Proof. For each i, let $\overline{\phi}_i$ be defined by (16), and let

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

Since $\tau \in (0, (1+\sqrt{5})/2)$, $\alpha \in (\tau/\min(1+\tau, 1+\tau^{-1}),\ 1]$, $\widehat{\Sigma}_f + S \succeq 0$, $H_f \succ 0$, $\widehat{\Sigma}_g + T \succeq \frac{1}{2}\widehat{\Sigma}_g + T \succeq 0$, and $M_g \succ 0$, we have $-\tau + \alpha \min(1+\tau,\ 1+\tau^{-1}) > 0$, $a_i \ge 0$, and $\overline{\phi}_i + \big(1 - \alpha \min(\tau,\ \tau^{-1})\big)\sigma\|r^i\|^2 + \alpha\xi_i \ge 0$ for any $i \ge 1$. It follows from (55) and the definitions of t_{i+1} and r^{i+1} that for any $i \ge 1$ we have

$$a_i = t_{i+1} + \big(-\tau + \alpha \min(1+\tau, 1+\tau^{-1})\big)\sigma\|r^{i+1}\|^2$$

$$\le \Big[\overline{\phi}_i + \big(1 - \alpha \min(\tau,\ \tau^{-1})\big)\sigma\|r^i\|^2 + \alpha\xi_i\Big]$$

$$\quad - \Big[\overline{\phi}_{i+1} + \big(1 - \alpha \min(\tau,\ \tau^{-1})\big)\sigma\|r^{i+1}\|^2 + \alpha\xi_{i+1}\Big].$$

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

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

From the above inequality, we have $\sum_{i=1}^\infty a_i < +\infty$. Then, by Lemma 3 and using $H_f \succ 0$ and $M_g \succ 0$, we get $\min_{1 \le i \le k}\{a_i\} = o(1/k)$, that is,

$$(78) \qquad \min_{1 \le i \le k} \big\{ \|z^{i+1} - z^i\|^2 + \|x^{i+1} - x^i\|^2 + \|y^{i+1} - y^i\|^2 \big\} = o(1/k).$$

It follows from the first-order optimality condition of (14) 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 (14), we have

$$
\nabla f(x^{i+1}) - \nabla f(x^i) + (1 - \tau^{-1})A(z^{i+1} - z^i) + \sigma AB^*(y^{i+1} - y^i)
$$

(79) $\qquad - (\widehat{\Sigma}_f + S)(x^{i+1} - x^i) \in \partial p(x^{i+1}) + \nabla f(x^{i+1}) + Az^{i+1}.$

Similarly, we get

$$
\nabla g(y^{i+1}) - \nabla g(y^i) + (1 - \tau^{-1})B(z^{i+1} - z^i) - (\widehat{\Sigma}_g + T)(y^{i+1} - y^i)
$$

(80) $\qquad \in \partial q(y^{i+1}) + \nabla g(y^{i+1}) + Bz^{i+1}.$

It follows from (14) that

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

By using the Cauchy–Schwarz inequality, (79), (80), and (81), we have

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

It follows from the above inequality, Assumption 1, and (78) that the assertion (76) is proved.

By using (12), for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we obtain

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

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

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

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

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

which, together with (76) and (82), implies (77). The proof is complete. $\qquad \square$

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 semiproximal ADMM:

(83)
$$
\min_{x \in \Re^n,\, y \in \Re^m} \left\{ \frac{1}{2}\langle x, Qx \rangle - \langle b, x \rangle + \frac{\chi}{2}\|\Pi_{\Re_+^m}(D(d - Hx))\|^2 + \varrho\|x\|_1 + \delta_{\Re_+^m}(y) \ \Big| \ Hx + y = c \right\},
$$

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

For problem (83), it can be expressed in the form (1) by taking

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

with $A^* = H$, $B^* = I$. The KKT system for problem (83) is given by

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

where "$\circ$" denotes the elementwise product. In our numerical experiments, we apply the majorized iPADMM to solve problem (83) by using both of the step-length parameters $\tau = 1.618$ and $\tau = 1$. We stop the majorized iPADMM based on the following relative residual on the KKT system:

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

Note that in the process of computing the iterates x^{k+1} and y^{k+1} from the majorized iPADMM, we can generate the corresponding dual variables ξ^{k+1} and v^{k+1} which satisfy the complementarity conditions in (84). Thus we need not check the complementarity conditions in (84) since they are satisfied exactly.

In our numerical experiments, for a given pair (n, m), we generate the data for (83) randomly as

```
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 (83) 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 (83).

6.1. The case where the parameter χ in (83) 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 (75) with

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

where ρ_1 is the constant given in (74), i.e.,

$$(87) \qquad \rho_1 := 1.01\lambda_{\max}\left(\widehat{\Sigma}_f - \frac{1}{2}\Sigma_f + \frac{1}{2}(1+\alpha)\sigma AA^*\right).$$

In the notation of (75), we fix $\alpha = 1$ when we choose $\tau = 1.618$. In this case, we get $\widehat{\Sigma}_f + S = \rho_1 I - \sigma AA^* \succeq \widehat{\Sigma}_f - \frac{1}{2}\Sigma_f \succeq 0$. When we choose $\tau = 1$, we need to restrict α to the interval $(0.5, 1]$. In this case, we also pick $\alpha = 1$, and the condition $\widehat{\Sigma}_f + S \succeq \widehat{\Sigma}_f - \frac{1}{2}\Sigma_f - \frac{1-\alpha}{2}\sigma AA^* \succeq 0$ is satisfied.

With the above choices of the proximal terms, we have that

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

Furthermore, the conditions (54) and (56) in Theorem 10 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 10 as follows:

$$(88) \qquad S := \rho_2 I - \widehat{\Sigma}_f - \sigma AA^*, \qquad T := 0,$$

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

$$(89) \qquad \rho_2 := \lambda_{\max}\left(\frac{1}{2}Q + \gamma_2(1-\eta)\sigma AA^*\right).$$

Observe that ρ_2 is a smaller constant than ρ_1 and hence the proximal term S in (88) is more indefinite than the one in (86). In this case, we can easily check that the conditions in (52) for S and T are satisfied. In particular, $\widehat{\Sigma}_f + S + \eta\sigma AA^* \succeq (1-\eta)\sigma\left(\lambda_{\max}(AA^*)I - AA^*\right) + (\gamma_2 - 1)(1-\eta)\sigma\lambda_{\max}(AA^*)I \succ 0$. However, in order to ensure that the iPADMM with the proximal terms chosen in (88) is convergent, we need to monitor the residual

$$(90) \qquad R^{k+1} := \|x^{k+1} - x^k\|_{\widehat{\Sigma}_f}^2 + \|y^{k+1} - y^k\|_{\widehat{\Sigma}_g + \sigma BB^*}^2 + \|r^{k+1}\|^2$$

in condition (53) as follows. At the kth 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 (88), but with the parameter γ_2 in (89) increased by a factor of 1.1. Obviously, the number of such restarts is finite since eventually ρ_2 will be larger than ρ_1 in (87). In our numerical runs, we encountered such a restart only once when testing all the instances.

Table 1 reports the comparison between the iPADMM (whose proximal term S is the indefinite matrix mentioned in (86)) and the semiproximal ADMM in [11] (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 (88)) and the sPADMM.

We can see from the results in both tables that the iPADMM can sometimes bring about a 40–50% reduction in the number of iterations needed to solve problem (83) as compared to the sPADMM. In addition, the iPADMM using the more aggressive indefinite proximal term S in (88) sometimes takes substantially fewer iterations to solve the problems as compared to the more conservative choice in (86).

TABLE 1
Comparison between the number of iterations (time in seconds) taken by iPADMM (for S given by (86)) and sPADMM (for $S = \lambda_{\max}(Q+\sigma AA^*)I - (Q+\sigma AA^*)$) to solve problem (83), with $\chi = 0$, to the required accuracy stated in (85).

Dim. of H ($m \times n$)	sPADMM $\tau = 1.618$	iPADMM $\tau = 1.618$	Ratio (%)	sPADMM $\tau = 1$	iPADMM $\tau = 1$	Ratio (%)
2000×1000	8977 (34.9)	8571 (33.1)	95.5 (94.8)	11077 (43.4)	9382 (36.1)	84.7 (83.1)
2000×2000	1807 (18.9)	1406 (16.1)	77.8 (85.2)	1970 (20.0)	1398 (15.6)	71.0 (77.9)
2000×4000	1224 (43.9)	714 (29.5)	58.3 (67.2)	1254 (43.9)	784 (31.7)	62.5 (72.1)
2000×8000	1100 (113.9)	613 (83.1)	55.7 (73.0)	1103 (115.4)	647 (84.4)	58.7 (73.2)
4000×2000	9863 (181.4)	9222 (168.3)	93.5 (92.8)	11919 (219.2)	10413 (188.4)	87.4 (86.0)
4000×4000	2004 (104.2)	1273 (71.9)	63.5 (69.1)	2140 (110.8)	1490 (80.7)	69.6 (72.9)
4000×8000	1408 (191.5)	829 (133.8)	58.9 (69.9)	1396 (193.1)	865 (138.0)	62.0 (71.5)
4000×16000	1745 (589.9)	915 (397.1)	52.4 (67.3)	1911 (648.6)	982 (407.0)	51.4 (62.8)
8000×4000	10681 (902.0)	10288 (871.6)	96.3 (96.6)	13187 (1114.7)	11458 (969.8)	86.9 (87.0)
8000×8000	2037 (419.2)	1220 (266.7)	59.9 (63.6)	2085 (426.8)	1296 (277.4)	62.2 (65.0)
8000×16000	1594 (798.1)	917 (549.7)	57.5 (68.9)	1696 (842.6)	994 (569.0)	58.6 (67.5)

TABLE 2
Comparison between the number of iterations (time in seconds) taken by iPADMM (for S given by (88)) and sPADMM (for $S = \lambda_{\max}(Q+\sigma AA^*)I - (Q+\sigma AA^*)$) to solve problem (83), with $\chi = 0$, to the required accuracy stated in (85).

Dim. of H ($m \times n$)	sPADMM $\tau = 1.618$	iPADMM $\tau = 1.618$	Ratio (%)	sPADMM $\tau = 1$	iPADMM $\tau = 1$	Ratio (%)
2000×1000	8977 (36.0)	6216 (25.3)	69.2 (70.2)	11077 (41.0)	8094 (30.0)	73.1 (73.2)
2000×2000	1807 (20.1)	1001 (12.6)	55.4 (62.6)	1970 (20.7)	1119 (12.8)	56.8 (61.9)
2000×4000	1224 (44.0)	659 (29.3)	53.8 (66.5)	1254 (43.9)	734 (31.0)	58.5 (70.7)
2000×8000	1100 (115.7)	593 (80.8)	53.9 (69.9)	1103 (114.1)	626 (85.4)	56.8 (74.9)
4000×2000	9863 (181.7)	7150 (131.3)	72.5 (72.3)	11919 (218.0)	8762 (158.3)	73.5 (72.6)
4000×4000	2004 (104.4)	1148 (65.6)	57.3 (62.8)	2140 (110.1)	1280 (71.0)	59.8 (64.5)
4000×8000	1408 (192.2)	805 (134.4)	57.2 (69.9)	1396 (191.3)	846 (136.6)	60.6 (71.4)
4000×16000	1747 (588.0)	900 (392.0)	51.5 (66.7)	1911 (658.0)	930 (403.9)	48.7 (61.4)
8000×4000	10681 (900.8)	8111 (689.5)	75.9 (76.5)	13187 (1111.1)	9587 (821.0)	72.7 (73.9)
8000×8000	2037 (420.2)	1109 (248.8)	54.4 (59.2)	2085 (423.8)	1221 (269.2)	58.6 (63.5)
8000×16000	1594 (802.2)	902 (539.5)	56.6 (67.3)	1696 (845.0)	942 (551.5)	55.5 (65.3)

6.2. The case where the parameter χ in (83) is positive. In the second case, we consider problem (83) where the parameter χ is set to $\chi = 2\varrho$. 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 $\widehat{\Sigma}_f = Q + \chi H^* D^2 H$, $\Sigma_f = Q$, $\widehat{\Sigma}_g = 0 = \Sigma_g$. For problem (83) with $\chi = 2\varrho$, the indefinite proximal term S given in (86) is too conservative due to the fact that $\widehat{\Sigma}_f$ is substantially "larger" than Σ_f. In order to realize the full potential of allowing for an indefinite proximal term, we make use of the condition (53) in part (a) of Theorem 10 by choosing S and T as follows:

$$(91) \qquad S := \rho_3 I - \widehat{\Sigma}_f - \sigma AA^*, \qquad T := 0,$$

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

$$(92) \qquad \rho_3 := \lambda_{\max}\Big(\frac{1}{2}Q + \big((1-\eta)\sigma + \gamma_3\chi\big)AA^*\Big).$$

Table 3

Comparison between the number of iterations (time in seconds) taken by the majorized iPADMM (for S given by (91)) and majorized sPADMM (for $S = \lambda_{\max}(\widehat{\Sigma}_f + \sigma AA^*)I - (\widehat{\Sigma}_f + \sigma AA^*)$) to solve problem (83), with $\chi = 2\varrho$, to the required accuracy stated in (85).

Dim. of H ($m \times n$)	sPADMM $\tau = 1.618$	iPADMM $\tau = 1.618$	Ratio (%)	sPADMM $\tau = 1$	iPADMM $\tau = 1$	Ratio (%)
2000×1000	16030 (91.0)	4417 (30.5)	27.6 (33.5)	16638 (94.7)	4951 (31.3)	29.8 (33.1)
2000×2000	2091 (33.8)	574 (12.5)	27.5 (37.0)	2111 (33.6)	647 (12.9)	30.6 (38.4)
2000×4000	1292 (74.0)	427 (29.8)	33.0 (40.3)	1311 (74.0)	479 (31.8)	36.5 (42.9)
2000×8000	972 (173.3)	414 (86.5)	42.6 (50.0)	988 (173.4)	455 (94.2)	46.1 (54.3)
4000×2000	14499 (421.0)	4717 (142.6)	32.5 (33.9)	14750 (429.3)	5616 (168.5)	38.1 (39.2)
4000×4000	1599 (141.0)	514 (51.8)	32.1 (36.7)	1677 (148.5)	638 (61.0)	38.0 (41.1)
4000×8000	1076 (253.7)	416 (113.2)	38.7 (44.6)	1110 (275.3)	476 (119.1)	42.9 (43.3)
4000×16000	1082 (814.1)	526 (382.5)	48.6 (47.0)	1109 (816.6)	549 (381.0)	49.5 (46.7)
8000×4000	10528 (1336.1)	4193 (530.6)	39.8 (39.7)	11096 (1401.4)	4596 (584.2)	41.4 (41.7)
8000×8000	1260 (475.9)	543 (205.8)	43.1 (43.2)	1318 (483.3)	644 (229.9)	48.9 (47.6)
8000×16000	983 (1148.3)	443 (486.5)	45.1 (42.4)	991 (1104.4)	499 (496.8)	50.4 (45.0)

In this case, we can easily check that the conditions in (52) 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 (91) is convergent, we need to monitor the residual, defined similarly as in (90), in condition (53) as follows. At the kth 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 (91), but with the parameter γ_3 in (92) increased by a factor of 1.1. As before, the number of such restarts is finite since eventually ρ_3 will be larger than ρ_1 in (87). 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 (91)) 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 problem (83) 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 for more research on augmented Lagrangian function–based methods beyond their traditional domain.

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*, Math. Program., 155 (2016), pp. 57–79, doi:10.1007/s10107-014-0826-5.
[2] C. H. Chen, Y. Shen, and Y. F. You, *On the convergence analysis of the alternating direction method of multipliers with three blocks*, Abstr. Appl. Anal., 2013 (2013), 183961, doi:10.1155/2013/183961.

[3] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, Classics Appl. Math. 5, SIAM, Philadel-
     phia, 1990.
[4] Y. CUI, X. D. LI, D. F. SUN, AND K.-C. TOH, *On the convergence properties of a ma-
     jorized alternating direction method of multipliers for linearly constrained convex op-
     timization problems with coupled objective functions*, J. Optim. Theory Appl., (2016),
     doi:10.1007/s10957-016-0877-2.
[5] D. DAVIS AND W. YIN, *Convergence Rate Analysis of Several Splitting Schemes*, preprint,
     arXiv:1406.4834v3 [math.OC], 2015.
[6] W. DENG, M.-J. LAI, Z. M. PENG, AND W. YIN, *Parallel Multi-Block ADMM with o(1/k)
     Convergence*, preprint, arXiv:1312.3040v2 [math.OC], 2014.
[7] W. DENG AND W. YIN, *On the global and linear convergence of the generalized al-
     ternating direction method of multipliers*, J. Sci. Comput., 66 (2016), pp. 889–916,
     doi:10.1007/s10915-015-0048-x.
[8] J. ECKSTEIN, *Some saddle-function splitting methods for convex programming*, Optim. Methods
     Softw., 4 (1994), pp. 75–83, doi:10.1080/10556789408805578.
[9] J. ECKSTEIN AND D. P. BERTSEKAS, *On the Douglas-Rachford splitting method and the proximal
     point algorithm for maximal monotone operators*, Math. Programming, 55 (1992), pp. 293–
     318, doi:10.1007/BF01581204.
[10] J. ECKSTEIN AND W. YAO, *Understanding the convergence of the alternating direction method
     of multipliers: Theoretical and computational perspectives*, Pac. J. Optim., 11 (2015),
     pp. 619–644.
[11] M. FAZEL, T. K. PONG, D. F. SUN, AND P. TSENG, *Hankel matrix rank minimization with ap-
     plications to system identification and realization*, SIAM J. Matrix Anal. Appl., 34 (2013),
     pp. 946–977, doi:10.1137/110853996.
[12] M. FORTIN AND R. GLOWINSKI, *On decomposition-coordination methods using an augmented
     Lagrangian*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of
     Boundary-Value Problems, M. Fortin and R. Glowinski, eds., North-Holland, Amsterdam,
     1983, pp. 97–146.
[13] D. GABAY, *Applications of the method of multipliers to variational inequalities*, in Augmented
     Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems,
     M. Fortin and R. Glowinski, eds., North-Holland, Amsterdam, 1983, pp. 299–331.
[14] D. GABAY AND B. MERCIER, *A dual algorithm for the solution of nonlinear variational
     problems via finite element approximations*, Comput. Math. Appl., 2 (1976), pp. 17–40,
     doi:10.1016/0898-1221(76)90003-1.
[15] X. GAO, B. JIANG, AND S. ZHANG, *On the information-adaptive variants of the ADMM:
     An iteration complexity perspective*, Optimization Online, 2014, http://www.optimization-
     online.org/DB_HTML/2014/11/4633.html.
[16] Y. GAO AND D. F. SUN, *A Majorized Penalty Approach for Calibrating Rank Constrained Cor-
     relation Matrix Problems*, Technical report, Department of Mathematics, National Univer-
     sity of Singapore, Singapore, 2010.
[17] R. GLOWINSKI, *Lectures on Numerical Methods for Nonlinear Variational Problems. Notes by
     M. G. Vijayasundaram and M. Adimurthi*, Tata Inst. Fund. Res. Lectures on Math. and
     Phys. 65, Tata Institute of Fundamental Research, Bombay, 1980.
[18] R. GLOWINSKI AND A. MARROCO, *Sur l'approximation, par éléments finis d'ordre un, et la
     résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires*,
     Rev. Française Automat. Informat. Recherche Opérationelle Sér. Rouge Anal. Numér., 9
     (1975), pp. 41–76; available online at http://www.numdam.org/item?id=M2AN_1975_9_
     2_41_0.
[19] D. R. HAN AND X. M. YUAN, *A note on the alternating direction method of multipliers*, J.
     Optim. Theory Appl., 155 (2012), pp. 227–238, doi:10.1007/s10957-012-0003-z.
[20] B. S. HE, L.-Z. LIAO, D. R. HAN, AND H. YANG, *A new inexact alternating directions
     method for monotone variational inequalities*, Math. Program., 92 (2002), pp. 103–118,
     doi:10.1007/s101070100280.
[21] B. S. HE AND X. M. YUAN, *On the O(1/n) convergence rate of the Douglas–Rachford
     alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709,
     doi:10.1137/110836936.
[22] M. Y. HONG AND Z.-Q. LUO, *On the Linear Convergence of the Alternating Direction Method
     of Multipliers*, preprint, arXiv:1208.3922v3 [math.OC], 2013.
[23] 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-Pac. J. Oper. Res., 32 (2015),
     1550024, doi:10.1142/S0217595915500244.
[24] X. D. LI, *A Two-Phase Augmented Lagrangian Method for Convex Composite Quadratic Pro-
     gramming*, Ph.D. thesis, Department of Mathematics, National University of Singapore,
     Singapore, 2015.

[25] X. D. LI, D. F. SUN, AND K.-C. TOH, *A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions*, Math. Program., 155 (2016), pp. 333–373, doi:10.1007/s10107-014-0850-5.

[26] T. LIN, S. MA, AND S. ZHANG, *An extragradient-based alternating direction method for convex minimization*, Found. Comput. Math., (2015), doi:10.1007/s10208-015-9282-8.

[27] T. LIN, S. MA, AND S. ZHANG, *On the global linear convergence of the ADMM with multiblock variables*, SIAM J. Optim., 25 (2015), pp. 1478–1497, doi:10.1137/140971178.

[28] T.-Y. LIN, S.-Q. MA, AND S.-Z. ZHANG, *On the sublinear convergence rate of multi-block ADMM*, J. Oper. Res. Soc. China, 3 (2015), pp. 251–274, doi:10.1007/s40305-015-0092-0.

[29] 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 J. Optim., 20 (2010), pp. 2755–2787, doi:10.1137/090753127.

[30] R. D. C. MONTEIRO AND B. F. SVAITER, *Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers*, SIAM J. Optim., 23 (2013), pp. 475–507, doi:10.1137/110849468.

[31] 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 J. Optim., 15 (2004), pp. 229–251, doi:10.1137/S1052623403425629.

[32] Y. OUYANG, Y. CHEN, G. LAN, AND E. PASILIAO, JR., *An accelerated linearized alternating direction method of multipliers*, SIAM J. Imaging Sci., 8 (2015), pp. 644–681, doi:10.1137/14095697X.

[33] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[34] D. F. SUN, K.-C. TOH, AND L. YANG, *A convergent 3-block semiproximal alternating direction method of multipliers for conic programming with 4-type constraints*, SIAM J. Optim., 25 (2015), pp. 882–915, doi:10.1137/140964357.

[35] M. TAO AND X. M. YUAN, *On the $O(1/t)$ convergence rate of alternating direction method with logarithmic-quadratic proximal regularization*, SIAM J. Optim., 22 (2012), pp. 1431–1448, doi:10.1137/110847639.

[36] H. WANG AND A. BANERJEE, *Bregman alternating direction method of multipliers*, in Advances in Neural Information Processing Systems 27, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, eds., Curran Associates, Red Hook, NY, 2014, pp. 2816–2824; available online at http://papers.nips.cc/paper/5612-bregman-alternating-direction-method-of-multipliers.pdf.