J Optim Theory Appl (2016) 169:1013–1041
DOI 10.1007/s10957-016-0877-2

On the Convergence Properties of a Majorized Alternating Direction Method of Multipliers for Linearly Constrained Convex Optimization Problems with Coupled Objective Functions

Ying Cui[1] · **Xudong Li**[1] · **Defeng Sun**[2] · **Kim-Chuan Toh**[1]

Received: 31 January 2015 / Accepted: 18 January 2016 / Published online: 2 February 2016

Abstract In this paper, we establish the convergence properties for a majorized alternating direction method of multipliers for linearly constrained convex optimization problems, whose objectives contain coupled functions. Our convergence analysis relies on the generalized Mean-Value Theorem, which plays an important role to properly control the cross terms due to the presence of coupled objective functions. Our results, in particular, show that directly applying two-block alternating direction method of multipliers with a large step length of the golden ratio to the linearly constrained convex optimization problem with a quadratically coupled objective function is convergent under mild conditions. We also provide several iteration complexity results for the algorithm.

Keywords Coupled objective function · Convex quadratic programming · Majorization · Iteration complexity · Nonsmooth analysis

Mathematics Subject Classification 90C25 · 68Q25 · 65K05

✉ Defeng Sun
matsundf@nus.edu.sg

Ying Cui
cuiying@nus.edu.sg

Xudong Li
matlixu@nus.edu.sg

Kim-Chuan Toh
mattohkc@nus.edu.sg

[1] Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore, Singapore

[2] Department of Mathematics and Risk Management Institute, National University of Singapore, 10 Lower Kent Ridge Road, Singapore, Singapore

1 Introduction

In this paper, we discuss the convergence of the alternating direction method of multipliers (ADMM) for minimizing linearly constrained convex optimization problems with coupled objective functions. More precisely, the concerned objective function consists of two nonsmooth separable functions and a coupled smooth one. This type of problems arises frequently in many fields of applications, including signal processing, image restoration, machine learning, and etc.

One popular way to solve this class of problems is the well studied classic augmented Lagrangian method (ALM). The ALM minimizes the augmented Lagrangian function with respect to all the decision variables simultaneously, regardless of whether the objective function is coupled or not before updating the Lagrangian multiplier along the (sub-)gradient ascent direction of the dual problem. Numerically, however, to solve the inner subproblems exactly or approximately with a high accuracy in the ALM, which may not be necessary at early stages, can be a time consuming task due to the non-separable structure combined with the two nonsmooth functions in the augmented Lagrangian function.

When the objective function is separable, i.e., the coupled smooth function is not present, one can alleviate the numerical difficulty in the ALM by directly applying the ADMM. The global convergence of the ADMM with separable objective functions has been extensively studied in the literature; see, for examples, [1–5]. For a recent survey, see Eckstein and Yao [6]. Meanwhile, there are very few papers on the ADMM targeting the optimization problems with coupled objective functions except for the work of Hong et al. [7], where the authors studied a majorized multi-block ADMM for linearly constrained optimization problems with non-separable objectives. Hong et al. [7] provided a very general convergence analysis of their majorized ADMM, assuming that the step length is a sufficiently small fixed number or converging to zero, among other conditions. Since a large step length is almost always desired in practice, one needs to develop a new convergence theorem beyond the one in [7].

In this paper, we conduct a thorough convergence analysis on our proposed majorized ADMM (see Sect. 3) when it is applied to the optimization problems with non-separable objective functions. This majorized ADMM reduces to the classic ADMM when the coupled term is absent from the objective. By making use of nonsmooth analysis, especially the generalized Mean-Value Theorem, we are able to establish the global convergence and the iteration complexity for our majorized ADMM with the step length taken as large as 1.618. To the best of our knowledge, this is the first paper providing the convergence properties of the (majorized) ADMM with a large step length for solving linearly constrained convex optimization problems with coupled smooth objective functions.

The remaining parts of our paper are organized as follows. In next section, we state the targeted optimization problems and provide some preliminary results. Section 3 focuses on our framework of a majorized ADMM and two important inequalities for the convergence analysis. In Sect. 4, we prove the global convergence and several iteration complexity results of the proposed algorithm. We conclude our paper in the last section.

2 Preliminaries

Consider the following convex optimization problem:

$$\min_{u,v}[\theta(u,v) := p(u) + q(v) + \phi(u,v) \,|\, \mathcal{A}^*u + \mathcal{B}^*v = c], \tag{1}$$

where $p : \mathcal{U} \to]-\infty, \infty]$, $q : \mathcal{V} \to]-\infty, \infty]$ are two closed and proper convex functions (possibly nonsmooth), $\phi : \mathcal{U} \times \mathcal{V} \to]-\infty, \infty[$ is a smooth and convex function, whose gradient mapping is Lipschitz continuous, $\mathcal{A} : \mathcal{X} \to \mathcal{U}$ and $\mathcal{B} : \mathcal{X} \to \mathcal{V}$ are two given linear operators, $c \in \mathcal{X}$ is a given vector, and $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{X}$ are three real finite dimensional Euclidean spaces, each equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$.

One particular example of problem (1) is the following problem, whose objective is the sum of a quadratic function and a squared distance function to a closed and convex set:

$$\min_{u,v}\left[\frac{1}{2}\left\langle \begin{pmatrix} u \\ v \end{pmatrix}, \widetilde{\mathcal{Q}}\begin{pmatrix} u \\ v \end{pmatrix}\right\rangle + \frac{\rho}{2}\left\| \begin{pmatrix} u \\ v \end{pmatrix} - \Pi_{\mathcal{K}_1}\begin{pmatrix} u \\ v \end{pmatrix}\right\|^2 \,\middle|\, \mathcal{A}^*u + \mathcal{B}^*v = c, \quad u \in \mathcal{K}_2,\ v \in \mathcal{K}_3\right], \tag{2}$$

where $\rho > 0$ is a penalty parameter, $\widetilde{\mathcal{Q}} : \mathcal{U} \times \mathcal{V} \to \mathcal{U} \times \mathcal{V}$ is a self-ajoint positive semidefinite linear operator, $\mathcal{K}_1 \subseteq \mathcal{U} \times \mathcal{V}$, $\mathcal{K}_2 \subseteq \mathcal{U}$ and $\mathcal{K}_3 \subseteq \mathcal{V}$ are closed and convex sets and $\Pi_{\mathcal{K}_1}(\cdot, \cdot)$ denotes the metric projection onto $\mathcal{K}_1$.

The augmented Lagrangian function for problem (1) associated with the parameter $\sigma > 0$ is defined as

$$\begin{aligned}\mathcal{L}_\sigma(u, v; x) := \theta(u, v) + \langle x, \mathcal{A}^*u + \mathcal{B}^*v - c\rangle + \frac{\sigma}{2}\|\mathcal{A}^*u \\ + \mathcal{B}^*v - c\|^2, \quad \forall\, (u, v, x) \in \mathcal{U} \times \mathcal{V} \times \mathcal{X}.\end{aligned} \tag{3}$$

Even if $\phi(\cdot, \cdot)$ is not separable, one can still apply the classic ADMM scheme for solving problem (1) as follows:

$$\begin{aligned} u^{k+1} &= \arg\min_u \mathcal{L}_\sigma(u, v^k; x^k), \\ v^{k+1} &= \arg\min_v \mathcal{L}_\sigma(u^{k+1}, v; x^k), \\ x^{k+1} &= x^k + \tau\sigma(\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c), \end{aligned} \tag{4}$$

where $\tau > 0$ is the step length. However, its convergence analysis is largely non-existent in the literature. One way to deal with the non-separability of $\phi(\cdot, \cdot)$ is to introduce a new variable $w = \begin{pmatrix} u \\ v \end{pmatrix}$. By letting $\widetilde{\mathcal{A}}^* = \begin{pmatrix} \mathcal{A} \\ \mathcal{I}_1 \\ 0 \end{pmatrix}$, $\widetilde{\mathcal{B}}^* = \begin{pmatrix} \mathcal{B} \\ 0 \\ \mathcal{I}_2 \end{pmatrix}$, $\widetilde{\mathcal{C}}^* = \begin{pmatrix} 0 & 0 \\ -\mathcal{I}_1 & 0 \\ 0 & -\mathcal{I}_2 \end{pmatrix}$ and $\tilde{c} = \begin{pmatrix} c \\ 0 \\ 0 \end{pmatrix}$ with identity maps $\mathcal{I}_1 : \mathcal{U} \to \mathcal{U}$ and $\mathcal{I}_2 : \mathcal{V} \to \mathcal{V}$, we can rewrite the optimization problem (2) equivalently as

$$\min_{u,v,w} [\tilde{\theta}(u, v, w) := p(u) + q(v) + \phi(w) \mid \widetilde{\mathcal{A}}^* u + \widetilde{\mathcal{B}}^* v + \widetilde{\mathcal{C}}^* w = \tilde{c}]. \tag{5}$$

For given $\sigma > 0$, the corresponding augmented Lagrangian function for problem (5) is

$$\widetilde{L}_\sigma(u, v, w; x) = \tilde{\theta}(u, v, w) + \langle x, \widetilde{\mathcal{A}}^* u + \widetilde{\mathcal{B}}^* v + \widetilde{\mathcal{C}}^* w - \tilde{c}\rangle + \frac{\sigma}{2}\|\widetilde{\mathcal{A}}^* u + \widetilde{\mathcal{B}}^* v + \widetilde{\mathcal{C}}^* w - \tilde{c}\|^2,$$

where $(u, v, w) \in \mathcal{U} \times \mathcal{V} \times (\mathcal{U} \times \mathcal{V})$ and $x \in \mathcal{X}$. Directly applying the three-block ADMM yields the following framework:

$$\begin{aligned}
u^{k+1} &= \arg\min_u \widetilde{\mathcal{L}}_\sigma(u, v^k, w^k; x^k),\\
v^{k+1} &= \arg\min_v \widetilde{\mathcal{L}}_\sigma(u^{k+1}, v, w^k; x^k),\\
w^{k+1} &= \arg\min_w \widetilde{\mathcal{L}}_\sigma(u^{k+1}, v^{k+1}, w; x^k),\\
x^{k+1} &= x^k + \tau\sigma(\widetilde{\mathcal{A}}^* u^{k+1} + \widetilde{\mathcal{B}}^* v^{k+1} + \widetilde{\mathcal{C}}^* w^{k+1} - \tilde{c}),
\end{aligned}$$

where $\tau > 0$ is the step length. Even though numerically the three-block ADMM works well for many applications, generally it is not a convergent algorithm even if τ is as small as 10^{-8}, as shown in the counterexamples given by Chen et al. [8].

As is mentioned in the Introduction section, Hong et al. [7] proposed to combine the majorization technique within the ADMM framework. When specialized to the two-block case for problem (1), their algorithm works as follows:

$$\begin{aligned}
u^{k+1} &= \arg\min_u [p(u) + \langle x^k, \mathcal{A}^* u\rangle + \hat{h}_1(u; u^k, v^k)],\\
v^{k+1} &= \arg\min_v [q(v) + \langle x^k, \mathcal{B}^* v\rangle + \hat{h}_2(v; u^{k+1}, v^k)],\\
x^{k+1} &= x^k + \alpha_k\sigma(\mathcal{A}^* u^{k+1} + \mathcal{B}^* v^{k+1} - c),
\end{aligned} \tag{6}$$

where $\hat{h}_1(u; u^k, v^k)$ and $\hat{h}_2(v; u^{k+1}, v^k)$ are majorization functions of $\phi(u, v) + \frac{\sigma}{2}\|\mathcal{A}^* u + \mathcal{B}^* v - c\|^2$ at (u^k, v^k) and (u^{k+1}, v^k), respectively, and $\alpha_k > 0$ is the step length.

Similar to Hong et al.'s work [7], our approach also relies on the majorization technique applied to the smooth coupled function $\phi(\cdot, \cdot)$. One difference is that we majorize $\phi(u, v)$ at (u^k, v^k) before the $(k + 1)$th iteration instead of changing the majorization function based on (u^{k+1}, v^k) when updating v^{k+1} as in (6). Interestingly, if $\phi(\cdot, \cdot)$ merely consists of quadratically coupled functions and separable smooth functions, then our majorized ADMM is exactly the same as the one proposed by Hong et al. under a proper choice of the majorization functions. Moreover, for applications such as (2), a potential advantage of our method is that we only need to compute the projection $\Pi_{\mathcal{K}_1}(\cdot, \cdot)$ once in order to compute $\nabla\phi(\cdot, \cdot)$ as a part of the majorization function within one iteration, while the procedure (6) needs to compute $\Pi_{\mathcal{K}_1}(\cdot, \cdot)$ at two different points (u^k, v^k) and (u^{k+1}, v^k).

Next, we summarize several necessary results to be used in our subsequent analysis.

Denote $w := \begin{pmatrix} u \\ v \end{pmatrix}$. Since $\phi(\cdot)$ is assumed to be a convex function with a Lipschitz continuous gradient, $\nabla^2\phi(\cdot)$ exists almost everywhere. Thus, the following Clarke's generalized Hessian at given $w \in \mathcal{U} \times \mathcal{V}$ is well defined [9]:

$$\partial^2\phi(w) := \mathrm{conv}\left\{ \lim_{w^k \to w} \nabla^2\phi(w^k),\ \nabla^2\phi(w^k) \text{ exists} \right\}, \tag{7}$$

where "conv{S}" denotes the convex hull of a given set S. Note that for any $\mathcal{W} \in \partial^2\phi(w)$ with $w \in \mathcal{U} \times \mathcal{V}$, $\mathcal{W}$ is self-adjoint and positive semidefinite. In [10], Hiriart-Urruty et al. provide a second order Mean-Value Theorem for ϕ, which states that for any w' and w in $\mathcal{U} \times \mathcal{V}$, there exists $z \in [w', w]$, which is the line segment connecting w' and w, and $\mathcal{W} \in \partial^2\phi(z)$ such that

$$\phi(w) := \phi(w') + \langle \nabla\phi(w'), w - w' \rangle + \frac{1}{2}\langle w - w', \mathcal{W}(w - w') \rangle .$$

Since $\nabla\phi$ is assumed to be globally Lipschitz continuous, there exist two self-adjoint positive semidefinite linear operators $\mathcal{Q}$ and $\mathcal{H} : \mathcal{U} \times \mathcal{V} \to \mathcal{U} \times \mathcal{V}$ such that for any $w \in \mathcal{U} \times \mathcal{V}$,

$$\mathcal{Q} \preceq \mathcal{W} \preceq \mathcal{Q} + \mathcal{H}, \quad \forall\, \mathcal{W} \in \partial^2\phi(w). \tag{8}$$

Here, $\mathcal{Q}$ and $\mathcal{H}$ are independent of w. Thus, for any $w, w' \in \mathcal{U} \times \mathcal{V}$, we have

$$\phi(w) \geq \phi(w') + \langle \nabla\phi(w'), w - w' \rangle + \frac{1}{2}\|w' - w\|^2_{\mathcal{Q}}, \tag{9}$$

$$\phi(w) \leq \hat{\phi}(w; w') := \phi(w') + \langle \nabla\phi(w'), w - w' \rangle + \frac{1}{2}\|w' - w\|^2_{\mathcal{Q}+\mathcal{H}}. \tag{10}$$

In this paper, we further assume that

$$\mathcal{H} = \mathrm{Diag}\,(\mathcal{D}_1, \mathcal{D}_2), \tag{11}$$

where $\mathcal{D}_1 : \mathcal{U} \to \mathcal{U}$ and $\mathcal{D}_2 : \mathcal{V} \to \mathcal{V}$ are two self-adjoint positive semidefinite linear operators. In fact, this kind of structure naturally appears in applications like (2), where the best possible lower bound of the generalized Hessian is $\widetilde{\mathcal{Q}}$ and the best possible upper bound of the generalized Hessian is $\widetilde{\mathcal{Q}} + \mathcal{I}$, where $\mathcal{I} : \mathcal{U} \times \mathcal{V} \to \mathcal{U} \times \mathcal{V}$ is the identity operator. For this case, the tightest estimation of $\mathcal{H}$ is $\mathcal{I}$, which is block diagonal. Since the coupled function $\phi(u, v)$ consists of two block variables u and v, the operators $\mathcal{Q}$ and $\mathcal{W}$ can be decomposed accordingly as $\mathcal{Q} = \begin{pmatrix} \mathcal{Q}_{11} & \mathcal{Q}_{12} \\ \mathcal{Q}^*_{12} & \mathcal{Q}_{22} \end{pmatrix}$ and $\mathcal{W} = \begin{pmatrix} \mathcal{W}_{11} & \mathcal{W}_{12} \\ \mathcal{W}^*_{12} & \mathcal{W}_{22} \end{pmatrix}$, where $\mathcal{W}_{11},\ \mathcal{Q}_{11} : \mathcal{U} \to \mathcal{U}$ and $\mathcal{W}_{22},\ \mathcal{Q}_{22} : \mathcal{V} \to \mathcal{V}$ are self-adjoint positive semidefinite linear operators, and $\mathcal{W}_{12},\ \mathcal{Q}_{12} : \mathcal{V} \to \mathcal{U}$ are two

linear mappings, whose adjoints are denoted by $\mathcal{W}_{12}^*$ and $\mathcal{Q}_{12}^*$, respectively. Denote $\eta \in [0, 1]$ as a constant that satisfies

$$|\langle u, (\mathcal{W}_{12} - \mathcal{Q}_{12})v\rangle| \le \frac{\eta}{2}(\|u\|_{\mathcal{D}_1}^2 + \|v\|_{\mathcal{D}_2}^2), \quad \forall \mathcal{W} \in \partial^2\phi(u, v),\ u \in \mathcal{U},\ v \in \mathcal{V}. \tag{12}$$

Note that (12) always holds true for $\eta = 1$, according to the Cauchy-Schwarz inequality.

In order to prove the convergence of the proposed majorized ADMM, the following constraint qualification is needed:

Assumption 2.1 There exists $(\hat{u}, \hat{v}) \in \mathrm{ri}\ (\mathrm{dom}(p) \times \mathrm{dom}(q))$ such that $\mathcal{A}^*\hat{u} + \mathcal{B}^*\hat{v} = c$.

Let ∂p and ∂q be the subdifferential mappings of p and q, respectively. Define the set-valued mapping $\mathcal{F}$ by

$$F(u, v, x) := \nabla\phi(w) + \begin{pmatrix} \partial p(u) + \mathcal{A}x \\ \partial q(v) + \mathcal{B}x \end{pmatrix}, \quad \forall (u, v, x) \in \mathcal{U} \times \mathcal{V} \times \mathcal{X}.$$

Under Assumption 2.1,[1] $(\bar{u}, \bar{v})$ is optimal to (1) if and only if there exists $\bar{x} \in \mathcal{X}$ such that the following Karush–Kuhn–Tucker (KKT) condition holds:

$$0 \in F(\bar{u}, \bar{v}, \bar{x}), \quad \mathcal{A}^*\bar{u} + \mathcal{B}^*\bar{v} = c, \tag{13}$$

which is equivalent to the following variational inequality:

$$\begin{aligned}&(p(u) + q(v)) - (p(\bar{u}) + q(\bar{v})) + \langle w - \bar{w}, \nabla\phi(\bar{w})\rangle + \langle u - \bar{u}, \mathcal{A}\bar{x}\rangle + \langle v - \bar{v}, \mathcal{B}\bar{x}\rangle \\ &\quad -\langle x - \bar{x}, \mathcal{A}^*\bar{u} + \mathcal{B}^*\bar{v} - c\rangle \ge 0, \quad \forall (u, v, x) \in \mathcal{U} \times \mathcal{V} \times \mathcal{X}.\end{aligned} \tag{14}$$

Also recall that for any ξ, ζ in the same space and a self-adjoint positive semidefinite operator $\mathcal{G}$, it holds that

$$\langle \xi, \mathcal{G}\zeta\rangle = \frac{1}{2}\left(\|\xi\|_{\mathcal{G}}^2 + \|\zeta\|_{\mathcal{G}}^2 - \|\xi - \zeta\|_{\mathcal{G}}^2\right) = \frac{1}{2}\left(\|\xi + \zeta\|_{\mathcal{G}}^2 - \|\xi\|_{\mathcal{G}}^2 - \|\zeta\|_{\mathcal{G}}^2\right). \tag{15}$$

3 A Majorized Alternating Direction Method of Multipliers with Coupled Objective Functions

In this section, we will first present the framework of our majorized ADMM, and then prove two important inequalities that play an essential role for our convergence analysis.

[1] We may instead directly assume the existence of a KKT point without imposing Assumption 2.1 in our convergence analysis.

Let $\sigma > 0$. For given $w' = (u', v') \in \mathcal{U} \times \mathcal{V}$, define the following majorized augmented Lagrangian function associated with problem (1):

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

where $(w, x) = (u, v, x) \in \mathcal{U} \times \mathcal{V} \times \mathcal{X}$, and the majorized function $\hat{\phi}$ is given by (10). Then, our proposed algorithm works as follows:

Majorized ADMM: A majorized ADMM with coupled objective functions

Choose an initial point $(u^0, v^0, x^0) \in \mathrm{dom}(p) \times \mathrm{dom}(q) \times \mathcal{X}$ and parameters $\tau > 0$. Let $\mathcal{S}$ and $\mathcal{T}$ be given self-adjoint positive semidefinite linear operators. For $k = 0, 1, 2, \cdots$, perform the kth iteration as follows:

Step 1. Compute $u^{k+1} = \arg\min_{u\in\mathcal{U}} [\widehat{\mathcal{L}}_\sigma(u, v^k; (x^k, w^k)) + \frac{1}{2}\|u - u^k\|^2_{\mathcal{S}}]$.

Step 2. Compute $v^{k+1} = \arg\min_{v\in\mathcal{V}} [\widehat{\mathcal{L}}_\sigma(u^{k+1}, v; (x^k, w^k)) + \frac{1}{2}\|v - v^k\|^2_{\mathcal{T}}]$.

Step 3. Compute $x^{k+1} = x^k + \tau\sigma(\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c)$.

In order to simplify subsequent discussions, for $k = 0, 1, 2, \ldots$, define

$$\begin{aligned}
\tilde{x}^{k+1} &:= x^k + \sigma\left(\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right),\\
\Xi_{k+1} &:= \left\|v^{k+1} - v^k\right\|^2_{\mathcal{D}_2+\mathcal{T}} + \eta\left\|u^{k+1} - u^k\right\|^2_{\mathcal{D}_1},\\
\Theta_{k+1} &:= \left\|u^{k+1} - u^k\right\|^2_{\mathcal{S}} + \left\|v^{k+1} - v^k\right\|^2_{\mathcal{T}} + \frac{1}{4}\left\|w^{k+1} - w^k\right\|^2_{\mathcal{Q}},\\
\Gamma_{k+1} &:= \Theta_{k+1} + \min\left(\tau, 1 + \tau - \tau^2\right)\left\|v^{k+1} - v^k\right\|^2_{\sigma\mathcal{B}\mathcal{B}^*}\\
&\quad - \left\|u^{k+1} - u^k\right\|^2_{\eta\mathcal{D}_1} - \left\|v^{k+1} - v^k\right\|^2_{\eta\mathcal{D}_2}
\end{aligned} \tag{16}$$

and denote, for $(u, v, x) \in \mathcal{U} \times \mathcal{V} \times \mathcal{X}$,

$$\begin{aligned}
\Phi_k(u, v, x) &:= \frac{1}{\tau\sigma}\left\|x^k - x\right\|^2 + \left\|u^k - u\right\|^2_{\mathcal{D}_1+\mathcal{S}} + \left\|v^k - v\right\|^2_{\mathcal{Q}_{22}+\mathcal{D}_2+\mathcal{T}}\\
&\quad + \frac{1}{2}\left\|w^k - w\right\|^2_{\mathcal{Q}} + \sigma\left\|\mathcal{A}^*u + \mathcal{B}^*v^k - c\right\|^2,\\
\Psi_k(u, v, x) &:= \Phi_k(u, v, x) + \left\|w^k - w\right\|^2_{\mathcal{Q}}\\
&\quad + \max(1 - \tau, 1 - \tau^{-1})\sigma\left\|\mathcal{A}^*u^k + \mathcal{B}^*v^k - c\right\|^2.
\end{aligned} \tag{17}$$

Proposition 3.1 *Suppose that the solution set of problem (1) is nonempty and Assumption 2.1 holds. Assume that $\mathcal{S}$ and $\mathcal{T}$ are chosen such that the sequence $\{(u^k, v^k, x^k)\}$ is well defined. Then, the following conclusions hold:*

(i) *For $\tau \in]0, 1]$, we have that for any $k \geq 0$ and $(u, v, x) \in \mathcal{U} \times \mathcal{V} \times \mathcal{X}$,*

$$\begin{aligned}
&\left(p\left(u^{k+1}\right) + q\left(v^{k+1}\right)\right) - (p(u) + q(v)) + \left\langle w^{k+1} - w, \nabla\phi(w)\right\rangle \\
&+\left\langle u^{k+1} - u, \mathcal{A}x\right\rangle + \left\langle v^{k+1} - v, \mathcal{B}x\right\rangle \\
&-\left\langle \tilde{x}^{k+1} - x, \mathcal{A}^*u + \mathcal{B}^*v - c\right\rangle + \frac{1}{2}\left(\Phi_{k+1}(u, v, x) - \Phi_k(u, v, x)\right) \\
&\leq -\frac{1}{2}\left(\Theta_{k+1} + \sigma\left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^k - c\right\|^2\right. \\
&\quad \left.+(1-\tau)\sigma\left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\|^2\right).
\end{aligned} \tag{18}$$

(ii) *For $\tau \geq 0$, we have that for any $k \geq 1$ and $(u, v, x) \in \mathcal{U} \times \mathcal{V} \times \mathcal{X}$,*

$$\begin{aligned}
&\left(p\left(u^{k+1}\right) + q\left(v^{k+1}\right)\right) - (p(u) + q(v)) + \left\langle w^{k+1} - w, \nabla\phi(w)\right\rangle \\
&+\left\langle u^{k+1} - u, \mathcal{A}x\right\rangle + \left\langle v^{k+1} - v, \mathcal{B}x\right\rangle \\
&-\left\langle \tilde{x}^{k+1} - x, \mathcal{A}^*u + \mathcal{B}^*v - c\right\rangle + \frac{1}{2}\left(\Psi_{k+1}(u, v, x)\right. \\
&\left.+\Xi_{k+1} - (\Psi_k(u, v, x) + \Xi_k)\right) \\
&\leq -\frac{1}{2}\left(\Gamma_{k+1} + \min(1, 1 + \tau^{-1} - \tau)\sigma\left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\|^2\right).
\end{aligned} \tag{19}$$

Proof In the majorized ADMM iteration scheme, the optimality condition for (u^{k+1}, v^{k+1}) is

$$\begin{aligned}
0 \in\ &\partial p\left(u^{k+1}\right) + \nabla_u\phi\left(w^k\right) + \mathcal{A}\tilde{x}^{k+1} \\
&+ (\mathcal{Q}_{11} + \mathcal{D}_1 + \mathcal{S})\left(u^{k+1} - u^k\right) + \sigma\mathcal{A}\mathcal{B}^*\left(v^k - v^{k+1}\right), \\
0 \in\ &\partial q\left(v^{k+1}\right) + \nabla_v\phi(w^k) + \mathcal{B}\tilde{x}^{k+1} \\
&+ (\mathcal{Q}_{22} + \mathcal{D}_2 + \mathcal{T})\left(v^{k+1} - v^k\right) + \mathcal{Q}_{12}^*\left(u^{k+1} - u^k\right),
\end{aligned} \tag{20}$$

which can be reformulated as

$$\begin{aligned}
&-\mathcal{A}\tilde{x}^{k+1} - \nabla_u\phi\left(w^k\right) - (\mathcal{Q}_{11} + \mathcal{D}_1 + \mathcal{S})\left(u^{k+1} - u^k\right) - \sigma\mathcal{A}\mathcal{B}^*\left(v^k - v^{k+1}\right) \in \partial p(u^{k+1}), \\
&-\mathcal{B}\tilde{x}^{k+1} - \nabla_v\phi\left(w^k\right) - (\mathcal{Q}_{22} + \mathcal{D}_2 + \mathcal{T})\left(v^{k+1} - v^k\right) - \mathcal{Q}_{12}^*\left(u^{k+1} - u^k\right) \in \partial q\left(v^{k+1}\right).
\end{aligned} \tag{21}$$

Therefore, by the convexity of p and q, we have that for any $u \in \mathcal{U}$ and $v \in \mathcal{V}$,

$$\begin{aligned} p(u) \geq\ & p\left(u^{k+1}\right) - \Big\langle u - u^{k+1}, \mathcal{A}\tilde{x}^{k+1} + \nabla_u \phi\left(w^k\right) + (\mathcal{Q}_{11} + \mathcal{D}_1 + \mathcal{S})\left(u^{k+1} - u^k\right) \\ & + \sigma \mathcal{A}\mathcal{B}^*\left(v^k - v^{k+1}\right)\Big\rangle, \\ q(v) \geq\ & q\left(v^{k+1}\right) - \Big\langle v - v^{k+1}, \mathcal{B}\tilde{x}^{k+1} + \nabla_v \phi\left(w^k\right) + (\mathcal{Q}_{22} + \mathcal{D}_2 + \mathcal{T})\left(v^{k+1} - v^k\right) \\ & + \mathcal{Q}_{12}^*\left(u^{k+1} - u^k\right)\Big\rangle. \end{aligned} \tag{22}$$

By taking $(w, w') = (w, w^k)$ and (w^{k+1}, w) in (9), and $(w, w') = (w^{k+1}, w^k)$ in (10), we know that

$$\begin{aligned} \phi(w) &\geq \phi\left(w^k\right) + \left\langle \nabla\phi\left(w^k\right), w - w^k\right\rangle + \tfrac{1}{2}\left\|w - w^k\right\|_{\mathcal{Q}}^2, \\ \phi\left(w^{k+1}\right) &\geq \phi(w) + \left\langle \nabla\phi(w), w^{k+1} - w\right\rangle + \tfrac{1}{2}\left\|w^{k+1} - w\right\|_{\mathcal{Q}}^2, \\ \phi\left(w^{k+1}\right) &\leq \phi\left(w^k\right) + \left\langle \nabla\phi\left(w^k\right), w^{k+1} - w^k\right\rangle + \tfrac{1}{2}\left\|w^{k+1} - w^k\right\|_{\mathcal{Q}+\mathcal{H}}^2. \end{aligned} \tag{23}$$

Putting the above three inequalities together, we get

$$\begin{aligned} \left\langle \nabla\phi\left(w^k\right) - \nabla\phi(w), w^{k+1} - w\right\rangle \geq\ & \frac{1}{2}\left(\left\|w^k - w\right\|_{\mathcal{Q}}^2 + \left\|w^{k+1} - w\right\|_{\mathcal{Q}}^2\right) \\ & - \frac{1}{2}\left\|w^{k+1} - w^k\right\|_{\mathcal{Q}+\mathcal{H}}^2. \end{aligned} \tag{24}$$

Substituting (24) into the sum of the two inequalities in (22) and by the assumption (11) and the identity (15), we can further obtain that

$$\begin{aligned} & \left(p\left(u^{k+1}\right) + q\left(v^{k+1}\right)\right) - (p(u) + q(v)) + \left\langle w^{k+1} - w, \nabla\phi(w)\right\rangle + \left\langle u^{k+1} - u, \mathcal{A}x\right\rangle \\ & + \left\langle v^{k+1} - v, \mathcal{B}x\right\rangle - \left\langle \tilde{x}^{k+1} - x, \mathcal{A}^*u + \mathcal{B}^*v - c\right\rangle \\ & \leq \sigma\left\langle \mathcal{B}^*(v^{k+1} - v^k), \mathcal{A}^*(u^{k+1} - u)\right\rangle + \left\langle v^{k+1} - v^k, \mathcal{Q}_{12}^*(u^{k+1} - u)\right\rangle \\ & \quad - \frac{1}{2}\left(\left\|u^{k+1} - u^k\right\|_{\mathcal{S}}^2 + \left\|u^{k+1} - u\right\|_{\mathcal{D}_1+\mathcal{S}}^2 - \left\|u^k - u\right\|_{\mathcal{D}_1+\mathcal{S}}^2\right) \\ & \quad - \frac{1}{2}\left(\left\|v^{k+1} - v^k\right\|_{\mathcal{T}}^2 + \left\|v^{k+1} - v\right\|_{\mathcal{D}_2+\mathcal{T}}^2 - \left\|v^k - v\right\|_{\mathcal{D}_2+\mathcal{T}}^2\right) \\ & \quad - \frac{1}{2\tau\sigma}\left(\left\|x^{k+1} - x^k\right\|^2 + \left\|x^{k+1} - x\right\|^2 - \left\|x^k - x\right\|^2\right) \\ & \quad - (1-\tau)\sigma\left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\|^2 - \left\|w^{k+1} - w\right\|_{\mathcal{Q}}^2. \end{aligned} \tag{25}$$

(i) Assume that $\tau \in]0, 1]$. Then, we get that

$$\begin{aligned}
&\left\langle v^{k+1} - v^k \mathcal{Q}_{12}^* \left(u^{k+1} - u\right)\right\rangle \\
&= \left\langle \begin{pmatrix} 0 \\ v^{k+1} - v^k \end{pmatrix}, \mathcal{Q}(w^{k+1} - w)\right\rangle \\
&\quad - \left\langle \mathcal{Q}_{22}\left(v^{k+1} - v^k\right), v^{k+1} - v\right\rangle \\
&\leq \frac{1}{2}\left(\left\|v^{k+1} - v^k\right\|_{\mathcal{Q}_{22}}^2 + \left\|w^{k+1} - w\right\|_{\mathcal{Q}}^2\right) \\
&\quad - \frac{1}{2}\left(\left\|v^{k+1} - v^k\right\|_{\mathcal{Q}_{22}}^2 + \left\|v^{k+1} - v\right\|_{\mathcal{Q}_{22}}^2 - \left\|v^k - v\right\|_{\mathcal{Q}_{22}}^2\right) \\
&= \frac{1}{2}\left\|w^{k+1} - w\right\|_{\mathcal{Q}}^2 + \frac{1}{2}\left(\left\|v^k - v\right\|_{\mathcal{Q}_{22}}^2 - \left\|v^{k+1} - v\right\|_{\mathcal{Q}_{22}}^2\right), \qquad (26)
\end{aligned}$$

where the inequality is obtained by the Cauchy-Schwarz inequality. By some simple manipulations we can also see that

$$\begin{aligned}
&\sigma\left\langle \mathcal{B}^*\left(v^{k+1} - v^k\right), \mathcal{A}^*\left(u^{k+1} - u\right)\right\rangle \\
&= \frac{\sigma}{2}\left(\left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\|^2 - \left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^k - c\right\|^2\right) \\
&\quad + \frac{\sigma}{2}\left(\left\|\mathcal{A}^*u + \mathcal{B}^*v^k - c\right\|^2 - \left\|\mathcal{A}^*u + \mathcal{B}^*v^{k+1} - c\right\|^2\right). \qquad (27)
\end{aligned}$$

Finally, by substituting (26) and (27) into (25) and recalling the definition of $\Phi_{k+1}(\cdot,\cdot,\cdot)$ and Θ_{k+1} in (16) and (17), we have that

$$\begin{aligned}
&\left(p\left(u^{k+1}\right) + q\left(v^{k+1}\right)\right) - (p(u) + q(v)) + \left\langle w^{k+1} - w, \nabla\phi(w)\right\rangle + \left\langle u^{k+1} - u, \mathcal{A}x\right\rangle \\
&+ \left\langle v^{k+1} - v, \mathcal{B}x\right\rangle - \left\langle \tilde{x}^{k+1} - x, \mathcal{A}^*u + \mathcal{B}^*v - c\right\rangle + \frac{1}{2}\left(\Phi_{k+1}(u, v, x) - \Phi_k(u, v, x)\right) \\
&\leq -\frac{1}{2}\left(\left\|u^{k+1} - u^k\right\|_{\mathcal{S}}^2 + \left\|v^{k+1} - v^k\right\|_{\mathcal{T}}^2 + \frac{1}{2}\left\|w^{k+1} - w\right\|_{\mathcal{Q}}^2 + \frac{1}{2}\left\|w^k - w\right\|_{\mathcal{Q}}^2\right. \\
&\quad \left. + \sigma\left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^k - c\right\|^2 + (1 - \tau)\sigma\left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\|^2\right) \\
&\leq -\frac{1}{2}\left(\Theta_{k+1} + \sigma\left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^k - c\right\|^2 + (1 - \tau)\sigma\left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\|^2\right),
\end{aligned}$$

where the last inequality comes from the fact that $\frac{1}{2}\|w^{k+1} - w\|_{\mathcal{Q}}^2 + \frac{1}{2}\|w^k - w\|_{\mathcal{Q}}^2 \geq \frac{1}{4}\|w^{k+1} - w^k\|_{\mathcal{Q}}^2$. This completes the proof of part (i).

(ii) Assume that $\tau \geq 0$. In this part, first we shall estimate the following term

$$\sigma\left\langle \mathcal{B}^*\left(v^{k+1} - v^k\right), \mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\rangle + \left\langle v^{k+1} - v^k, \mathcal{Q}_{12}^*\left(u^{k+1} - u\right) + \mathcal{Q}_{22}\left(v^{k+1} - v\right)\right\rangle.$$

It follows from (21) that

$$\begin{aligned} &-\mathcal{B}\tilde{x}^{k+1} - \nabla_v \phi\left(w^k\right) - (\mathcal{Q}_{22} + \mathcal{D}_2 + \mathcal{T})\left(v^{k+1} - v^k\right) - \mathcal{Q}_{12}^*\left(u^{k+1} - u^k\right) \in \partial q\left(v^{k+1}\right), \\ &-\mathcal{B}\tilde{x}^{k} - \nabla_v \phi\left(w^{k-1}\right) - (\mathcal{Q}_{22} + \mathcal{D}_2 + \mathcal{T})\left(v^{k} - v^{k-1}\right) - \mathcal{Q}_{12}^*\left(u^{k} - u^{k-1}\right) \in \partial q\left(v^{k}\right). \end{aligned} \tag{28}$$

Since $\nabla\phi$ is globally Lipschitz continuous, it is known from Clarke's Mean-Value Theorem [9, Proposition 2.6.5] that there exists a self-adjoint and positive semidefinite operator $\mathcal{W}^k \in \mathrm{conv}\{\partial^2\phi([w^{k-1}, w^k])\}$ such that

$$\nabla\phi\left(w^k\right) - \nabla\phi\left(w^{k-1}\right) = \mathcal{W}^k\left(w^k - w^{k-1}\right),$$

where the set $\mathrm{conv}\{\partial^2\phi[w^{k-1}, w^k]\}$ denotes the convex hull of all points $\mathcal{W} \in \partial^2\phi(z)$ for any $z \in [w^{k-1}, w^k]$. Denote $\mathcal{W}^k := \begin{pmatrix} \mathcal{W}_{11}^k & \mathcal{W}_{12}^k \\ (\mathcal{W}_{12}^k)^* & \mathcal{W}_{22}^k \end{pmatrix}$, where $\mathcal{W}_{11}^k : \mathcal{U} \to \mathcal{U}$, $\mathcal{W}_{22}^k : \mathcal{V} \to \mathcal{V}$ are self-adjoint positive semidefinite operators and $\mathcal{W}_{12}^k : \mathcal{U} \to \mathcal{V}$ is a linear operator. By using (28) and the monotonicity of $\partial q(\cdot)$, we obtain that

$$\begin{aligned} &-\left\langle \mathcal{B}\left(\tilde{x}^{k+1} - \tilde{x}^k\right), v^{k+1} - v^k\right\rangle - \left\langle \mathcal{Q}_{22}\left(v^{k+1} - v^k\right) + \mathcal{Q}_{12}^*\left(u^{k+1} - u^k\right), v^{k+1} - v^k\right\rangle \\ &\geq \left\langle \nabla_v\phi\left(w^k\right) - \nabla_v\phi\left(w^{k-1}\right), v^{k+1} - v^k\right\rangle - \left\langle (\mathcal{Q}_{22} + \mathcal{D}_2 + \mathcal{T})\left(v^k - v^{k-1}\right), v^{k+1} - v^k\right\rangle \\ &\quad + \left\|v^{k+1} - v^k\right\|_{\mathcal{T}+\mathcal{D}_2}^2 - \left\langle u^k - u^{k-1}, \mathcal{Q}_{12}\left(v^{k+1} - v^k\right)\right\rangle \\ &= \left\langle u^k - u^{k-1}, \left(\mathcal{W}_{12}^k - \mathcal{Q}_{12}\right)\left(v^{k+1} - v^k\right)\right\rangle \\ &\quad - \left\langle \left(\mathcal{Q}_{22} + \mathcal{D}_2 + \mathcal{T} - \mathcal{W}_{22}^k\right)\left(v^k - v^{k-1}\right), v^{k+1} - v^k\right\rangle + \left\|v^{k+1} - v^k\right\|_{\mathcal{T}+\mathcal{D}_2}^2 \\ &\geq -\frac{\eta}{2}\left(\left\|u^k - u^{k-1}\right\|_{\mathcal{D}_1}^2 + \left\|v^{k+1} - v^k\right\|_{\mathcal{D}_2}^2\right) \\ &\quad - \frac{1}{2}\left(\left\|v^{k+1} - v^k\right\|_{\mathcal{T}+\mathcal{D}_2}^2 + \left\|v^k - v^{k-1}\right\|_{\mathcal{T}+\mathcal{D}_2}^2\right) + \left\|v^{k+1} - v^k\right\|_{\mathcal{T}+\mathcal{D}_2}^2 \\ &= \frac{1}{2}\left\|v^{k+1} - v^k\right\|_{\mathcal{T}+(1-\eta)\mathcal{D}_2}^2 - \frac{1}{2}\left\|v^k - v^{k-1}\right\|_{\mathcal{T}+\mathcal{D}_2}^2 - \frac{\eta}{2}\left\|u^k - u^{k-1}\right\|_{\mathcal{D}_1}^2, \end{aligned}$$

where the second inequality is obtained from (12) and the fact that $\mathcal{W}_{22}^k \succeq \mathcal{Q}_{22}$. Therefore, in light of $\mu_{k+1} = (1-\tau)\sigma\langle\mathcal{B}^*(v^{k+1} - v^k), \mathcal{A}^*u^k + \mathcal{B}^*v^k - c\rangle$, we can estimate the cross term

$$\begin{aligned} &\sigma\left\langle \mathcal{B}^*\left(v^{k+1} - v^k\right), \mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\rangle \\ &\quad + \left\langle \mathcal{Q}_{12}^*\left(u^{k+1} - u\right) + \mathcal{Q}_{22}\left(v^{k+1} - v\right), v^{k+1} - v^k\right\rangle \end{aligned}$$

$$
\begin{aligned}
&= (1-\tau)\sigma\left\langle \mathcal{B}^*\left(v^{k+1}-v^k\right), \mathcal{A}^*u^k+\mathcal{B}^*v^k-c\right\rangle\\
&\quad+\left\langle \mathcal{B}^*\left(v^{k+1}-v^k\right), \tilde{x}^{k+1}-\tilde{x}^k\right\rangle\\
&\quad+\left\langle \mathcal{Q}_{12}^*\left(u^k-u\right)+\mathcal{Q}_{22}\left(v^k-v\right), v^{k+1}-v^k\right\rangle\\
&\quad+\langle \mathcal{Q}_{12}^*\left(u^{k+1}-u^k\right)+\mathcal{Q}_{22}\left(v^{k+1}-v^k\right), v^{k+1}-v^k\rangle\\
&\le \mu_{k+1}+\frac{1}{2}\left(\left\|w^k-w\right\|_{\mathcal{Q}}^2+\left\|v^{k+1}-v^k\right\|_{\mathcal{Q}_{22}}^2\right)\\
&\quad-\frac{1}{2}\left\|v^{k+1}-v^k\right\|_{\mathcal{T}+(1-\eta)\mathcal{D}_2}^2+\frac{1}{2}\left\|v^k-v^{k-1}\right\|_{\mathcal{T}+\mathcal{D}_2}^2\\
&\quad+\frac{\eta}{2}\left\|u^k-u^{k-1}\right\|_{\mathcal{D}_1}^2. \qquad (29)
\end{aligned}
$$

Finally, by the Cauchy-Schwarz inequality we know that

$$
\mu_{k+1}\le\begin{cases}\frac{1}{2}(1-\tau)\sigma\left(\left\|\mathcal{B}^*\left(v^{k+1}-v^k\right)\right\|^2+\left\|\mathcal{A}^*u^k+\mathcal{B}^*v^k-c\right\|^2\right), & \tau\in]0,1],\\ \frac{1}{2}(\tau-1)\sigma\left(\tau\left\|\mathcal{B}^*\left(v^{k+1}-v^k\right)\right\|^2+\tau^{-1}\left\|\mathcal{A}^*u^k+\mathcal{B}^*v^k-c\right\|^2\right), & \tau>1.\end{cases} \qquad (30)
$$

Substituting (29) and (30) into (25) and by some manipulations, we can obtain (19). This completes the proof of part (ii). □

4 Convergence Analysis

With all the preparations given in the previous sections, we can now discuss the main convergence results of our paper.

4.1 The Global Convergence

First we prove that under mild conditions, the iteration sequence $\{(u^k, v^k, x^k)\}$ generated by the majorized ADMM with $\tau\in]0, \frac{1+\sqrt{5}}{2}[$ converges to an optimal solution of problem (1) and its dual.

Let $\bar{w}=(\bar{u},\bar{v})\in\mathcal{U}\times\mathcal{V}$ be an optimal solution of (1) and $\bar{x}\in\mathcal{X}$ be the corresponding optimal multiplier. For $k=0,1,2,\ldots$, define

$$
u_e^k:=u^k-\bar{u},\quad v_e^k:=v^k-\bar{v},\quad w_e^k:=w^k-\bar{w},\quad x_e^k:=x^k-\bar{x}.
$$

Theorem 4.1 *Suppose that the solution set of (1) is nonempty and Assumption 2.1 holds. Assume that $\mathcal{S}$ and $\mathcal{T}$ are chosen such that*

$$
\mathcal{Q}_{11}+\sigma\mathcal{A}\mathcal{A}^*+\mathcal{S}\succ 0\quad\textit{and}\quad\mathcal{Q}_{22}+\sigma\mathcal{B}\mathcal{B}^*+\mathcal{T}\succ 0.
$$

(i) *Assume that* $\tau \in]0, 1]$. *If for any* $w = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathcal{U} \times \mathcal{V}$, *it holds that*

$$\langle w, [\mathcal{Q} + Diag\,(\mathcal{S} + (1-\tau)\sigma\mathcal{A}\mathcal{A}^*, \mathcal{T} + (1-\tau)\sigma\mathcal{B}\mathcal{B}^*)]w\rangle = 0 \Rightarrow \|u\|\|v\| = 0, \tag{31}$$

then the generated sequence $\{(u^k, v^k)\}$ *converges to an optimal solution of (1), and* $\{x^k\}$ *converges to the corresponding optimal multiplier.*

(ii) *Assume that* $\tau \in]0, \frac{1+\sqrt{5}}{2}[$. *Under the conditions that*

$$\begin{aligned} \mathcal{M} &:= \frac{1}{4}\mathcal{Q} + Diag\,(\mathcal{S} - \eta\mathcal{D}_1,\ \mathcal{T} - \eta\mathcal{D}_2) \succeq 0, \\ \frac{1}{4}\mathcal{Q}_{11} + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* - \eta\mathcal{D}_1 &\succ 0, \quad \frac{1}{4}\mathcal{Q}_{22} + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* - \eta\mathcal{D}_2 \succ 0, \end{aligned} \tag{32}$$

and for any $w = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathcal{U} \times \mathcal{V}$, *it holds that*

$$\langle w, [\mathcal{M} + \sigma Diag\,(\mathcal{A}\mathcal{A}^*, \mathcal{B}\mathcal{B}^*)]w\rangle = 0 \Rightarrow \|u\|\|v\| = 0, \tag{33}$$

the generated sequence $\{(u^k, v^k)\}$ *converges to an optimal solution of (1) and* $\{x^k\}$ *converges to the corresponding optimal multiplier.*

Proof (i) Let $\tau \in]0, 1]$. By letting $(u, v, x) = (\bar{u}, \bar{v}, \bar{x})$ in (18) and the optimality condition (14), we can obtain that for any $k \geq 0$,

$$\begin{aligned} &\Phi_{k+1}(\bar{u}, \bar{v}, \bar{x}) - \Phi_k(\bar{u}, \bar{v}, \bar{x}) \leq \\ &\quad -\left(\Theta_{k+1} + \sigma\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^k - c\|^2 + (1-\tau)\sigma\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\|^2\right). \end{aligned} \tag{34}$$

The above inequality shows that $\{\Phi_{k+1}(\bar{u}, \bar{v}, \bar{x})\}$ is bounded, which implies that $\{\|x^{k+1}\|\}$, $\{\|w_e^{k+1}\|_{\mathcal{Q}}\}$, $\{\|u_e^{k+1}\|_S\}$ and $\{\|v_e^{k+1}\|_{\mathcal{Q}_{22}+\sigma\mathcal{B}\mathcal{B}^*+\mathcal{T}}\}$ are all bounded. From the positive definiteness of $\mathcal{Q}_{22} + \sigma\mathcal{B}\mathcal{B}^* + \mathcal{T}$, we can see that $\{\|v_e^{k+1}\|\}$ is bounded. By using the inequalities

$$\begin{aligned} \left\|\mathcal{A}^*u_e^{k+1}\right\| &\leq \left\|\mathcal{A}^*u_e^{k+1} + \mathcal{B}^*v_e^{k+1}\right\| + \left\|\mathcal{B}^*v_e^{k+1}\right\| \\ &\leq \tau\sigma\left(\left\|x_e^{k+1}\right\| + \left\|x_e^k\right\|\right) + \left\|\mathcal{B}^*v_e^{k+1}\right\|, \left\|u_e^{k+1}\right\|_{\mathcal{Q}_{11}} \\ &\leq \left\|w_e^{k+1}\right\|_{\mathcal{Q}} + \left\|v_e^{k+1}\right\|_{\mathcal{Q}_{22}}, \end{aligned}$$

we know that the sequence $\{\|u_e^{k+1}\|_{\sigma\mathcal{A}\mathcal{A}^*+\mathcal{Q}_{11}}\}$ is also bounded. Therefore, $\{\|u_e^{k+1}\|_{\mathcal{Q}_{11}+\sigma\mathcal{A}\mathcal{A}^*+\mathcal{S}}\}$ is bounded. By the positive definiteness of $\mathcal{Q}_{11} + \sigma\mathcal{A}\mathcal{A}^* + \mathcal{S}$, we know that $\{\|u_e^{k+1}\|\}$ is bounded. On the whole, the sequence $\{(u^k, v^k, x^k)\}$ is bounded. Thus, there exists a subsequence $\{(u^{k_i}, v^{k_i}, x^{k_i})\}$ converging to a cluster

point, say $(u^\infty, v^\infty, x^\infty)$. Next, we will prove that (u^∞, v^∞) is optimal to (1) and x^∞ is the corresponding optimal multiplier.

The inequality (34) also implies that

$$\lim_{k\to\infty}\left\|\mathcal{A}^*u^{k+1}+\mathcal{B}^*v^k-c\right\|=0,\quad \lim_{k\to\infty}(1-\tau)\left\|\mathcal{A}^*u^{k+1}+\mathcal{B}^*v^{k+1}-c\right\|=0,$$
$$\lim_{k\to\infty}\left\|w^{k+1}-w^k\right\|_{\mathcal{Q}+\mathrm{Diag}\,(\mathcal{S},\mathcal{T})}=0. \tag{35}$$

For $\tau\in]0,1[$, since $\lim_{k\to\infty}\|\mathcal{A}^*u^{k+1}+\mathcal{B}^*v^{k+1}-c\|=0$, by using (35) we see that

$$\lim_{k\to\infty}\left\|\mathcal{A}^*(u^{k+1}-u^k)\right\|\le\lim_{k\to\infty}\left(\left\|\mathcal{A}^*u^{k+1}+\mathcal{B}^*v^k-c\right\|+\left\|\mathcal{A}^*u^k+\mathcal{B}^*v^k-c\right\|\right)=0,$$
$$\lim_{k\to\infty}\left\|\mathcal{B}^*(v^{k+1}-v^k)\right\|\le\lim_{k\to\infty}\left(\left\|\mathcal{A}^*u^{k+1}+\mathcal{B}^*v^{k+1}-c\right\|+\left\|\mathcal{A}^*u^{k+1}+\mathcal{B}^*v^k-c\right\|\right)=0,$$

which implies $\lim_{k\to\infty}\|w^{k+1}-w^k\|_{\mathcal{Q}+\mathrm{Diag}\,(\mathcal{S}+\sigma\mathcal{A}\mathcal{A}^*,\mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*)}=0$. Therefore, for $\tau\in]0,1]$, we have $\lim_{k\to\infty}\|w^{k+1}-w^k\|_{\mathcal{Q}+\mathrm{Diag}\,(\mathcal{S}+(1-\tau)\sigma\mathcal{A}\mathcal{A}^*,\mathcal{T}+(1-\tau)\sigma\mathcal{B}\mathcal{B}^*)}=0$. By picking a subsequence of $\{(u^{k_i},v^{k_i})\}$ if necessary, we can see that either $\lim_{k_i\to\infty}\|u^{k_i+1}-u^{k_i}\|=0$ or $\lim_{k_i\to\infty}\|v^{k_i+1}-v^{k_i}\|=0$ by the condition (31). Without any loss of generality we assume that $\lim_{k_i\to\infty}\|v^{k_i+1}-v^{k_i}\|=0$. Thus,

$$\begin{aligned}\lim_{k_i\to\infty}\left\|\mathcal{A}^*(u^{k_i+1}-u^{k_i})\right\|&\le\lim_{k\to\infty}\Big(\left\|\mathcal{A}^*u^{k_i+1}+\mathcal{B}^*v^{k_i}-c\right\|+\left\|\mathcal{A}^*u^{k_i}+\mathcal{B}^*v^{k_i-1}-c\right\|\\&\qquad+\left\|\mathcal{B}^*\left(v^{k_i}-v^{k_i-1}\right)\right\|\Big)=0,\\\lim_{k_i\to\infty}\left\|u^{k_i+1}-u^{k_i}\right\|_{\mathcal{Q}_{11}}&\le\lim_{k_i\to\infty}\left(\left\|w^{k_i+1}-w^{k_i}\right\|_{\mathcal{Q}}+\left\|v^{k_i+1}-v^{k_i}\right\|_{\mathcal{Q}_{22}}\right)=0.\end{aligned} \tag{36}$$

Therefore, $\lim_{k_i\to\infty}\|u^{k_i+1}-u^{k_i}\|_{\mathcal{Q}_{11}+\mathcal{S}+\sigma\mathcal{A}\mathcal{A}^*}=0$. This implies $\lim_{k_i\to\infty}\|u^{k_i+1}-u^{k_i}\|=0$ by the positive definiteness of $\mathcal{Q}_{11}+\mathcal{S}+\sigma\mathcal{A}\mathcal{A}^*$.

Now, taking limits on both sides of (20) along the subsequence $\{(u^{k_i},v^{k_i},x^{k_i})\}$, and by using the closedness of the graphs of ∂p, ∂q and the continuity of $\nabla\phi$, we obtain

$$0\in F(u^\infty,v^\infty,x^\infty),\quad \mathcal{A}^*u^\infty+\mathcal{B}^*v^\infty=c.$$

This indicates that (u^∞,v^∞) is an optimal solution to (1) and x^∞ is the corresponding optimal multiplier. Since $(u^\infty,v^\infty,x^\infty)$ satisfies (13), all the above arguments involving $(\bar u,\bar v,\bar x)$ can be replaced by $(u^\infty,v^\infty,x^\infty)$. Thus the subsequence $\{\Phi_{k_i}(u^\infty,v^\infty,x^\infty)\}$ converges to 0 as $k_i\to\infty$. Since $\{\Phi_{k_i}(u^\infty,v^\infty,x^\infty)\}$ is non-increasing, we obtain that

$$\lim_{k\to\infty} \Phi_{k+1}\left(u^\infty, v^\infty, x^\infty\right) = \lim_{k\to\infty} (\tau\sigma)^{-1} \left\|x^{k+1} - x^\infty\right\|^2 + \left\|v^{k+1} - v^\infty\right\|^2_{\sigma\mathcal{B}\mathcal{B}^*} + \mathcal{T} + \mathcal{Q}_{22} + \left\|u^{k+1} - u^\infty\right\|^2_{\mathcal{S}} + \left\|w^{k+1} - w^\infty\right\|^2_{\mathcal{Q}} = 0. \tag{37}$$

From this we can immediately get $\lim_{k\to\infty} x^{k+1} = x^\infty$ and $\lim_{k\to\infty} v^{k+1} = v^\infty$. Similar to (36), we have that $\lim_{k\to\infty} \sigma\|\mathcal{A}^*(u^{k+1} - u^\infty)\| = 0$ and $\lim_{k\to\infty} \|u^{k+1} - u^\infty\|_{\mathcal{Q}_{11}} = 0$, which, together with, (37) imply that $\lim_{k\to\infty} \|u^{k+1} - u^\infty\| = 0$ by the positive definiteness of $\mathcal{Q}_{11} + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^*$. Therefore, the whole sequence $\{(u^k, v^k, x^k)\}$ converges to $(u^\infty, v^\infty, x^\infty)$, the unique limit of the sequence. This completes the proof for the first case.

(ii) From the inequality (19) and the optimality condition (14) we know that for any $k \geq 1$,

$$(\Psi_{k+1}(\bar{u}, \bar{v}, \bar{x}) + \Xi_{k+1}) - (\Psi_k(\bar{u}, \bar{v}, \bar{x}) + \Xi_k) \leq -\left(\Gamma_{k+1} + \min\left(1, 1 + \tau^{-1} - \tau\right)\sigma \left\|\mathcal{A}^* u^{k+1} + \mathcal{B}^* v^{k+1} - c\right\|^2\right). \tag{38}$$

By the assumptions $\tau \in]0, \frac{1+\sqrt{5}}{2}[$ and $\mathcal{M} \succeq 0$, we can obtain that $\Gamma_{k+1} \geq 0$ and $\min(1, 1 + \tau^{-1} - \tau) \geq 0$. Then, both $\{\Psi_{k+1}(\bar{u}, \bar{v}, \bar{x})\}$ and $\{\Xi_{k+1}\}$ are bounded. Thus, by a similar approach to case (i), we see that the sequence $\{(u^k, v^k, x^k)\}$ is bounded. Therefore, there exists a subsequence $\{(u^{k_i}, v^{k_i}, x^{k_i})\}$ that converges to a cluster point, say $(u^\infty, v^\infty, x^\infty)$. Next, we will prove that (u^∞, v^∞) is optimal to problem (1) and x^∞ is the corresponding optimal multiplier. The inequality (38) also implies that

$$\lim_{k\to\infty} \left\|x^{k+1} - x^k\right\| = \lim_{k\to\infty} (\tau\sigma)^{-1} \left\|\mathcal{A}^* u^{k+1} + \mathcal{B}^* v^{k+1} - c\right\| = 0,$$
$$\lim_{k\to\infty} \left\|w^{k+1} - w^k\right\|_{\mathcal{M}} = 0, \quad \lim_{k\to\infty} \left\|\mathcal{B}^*\left(v^{k+1} - v^k\right)\right\| = 0.$$

By the relationship

$$\lim_{k\to\infty} \left\|\mathcal{A}^*(u^{k+1} - u^k)\right\| \leq \lim_{k\to\infty} \left(\left\|\mathcal{A}^* u^{k+1} + \mathcal{B}^* v^{k+1} - c\right\| + \left\|\mathcal{A}^* u^k + \mathcal{B}^* v^k - c\right\| + \left\|\mathcal{B}^*\left(v^{k+1} - v^k\right)\right\|\right) = 0,$$

we can further get $\lim_{k\to\infty} \|w^{k+1} - w^k\|_{\mathcal{M}+\mathrm{Diag}\,(\sigma\mathcal{A}\mathcal{A}^*, \sigma\mathcal{B}\mathcal{B}^*)} = 0$. Thus, by the condition (33) and taking a subsequence of $\{(u^{k_i}, v^{k_i})\}$ if necessary, we can get either

$\lim_{k\to\infty} \|u^{k_i+1} - u^{k_i}\| = 0$ or $\lim_{k\to\infty} \|v^{k_i+1} - v^{k_i}\| = 0$. Again, similar to case (i), we can see that, in fact, both of them would hold by the positive definiteness of $\frac{1}{4}\mathcal{Q}_{11} + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* - \eta\mathcal{D}_1$ and $\frac{1}{4}\mathcal{Q}_{22} + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* - \eta\mathcal{D}_{22}$. The remaining proof about the convergence of the whole sequence $\{(u^k, v^k, x^k)\}$ follows exactly the same as in case (i). This completes the proof for the second case. □

Remark 4.1 In Theorem 4.1, for $\tau \in]0, 1]$, a sufficient condition for the convergence is

$$\mathcal{Q} + \text{Diag}\left(\mathcal{S} + (1-\tau)\sigma\mathcal{A}\mathcal{A}^*, \mathcal{T} + (1-\tau)\sigma\mathcal{B}\mathcal{B}^*\right) \succ 0,$$

and for $\tau \in [1, \frac{1+\sqrt{5}}{2}[$, a sufficient condition for the convergence is

$$\frac{1}{4}\mathcal{Q} + \text{Diag}\,(\mathcal{S} - \eta\mathcal{D}_1,\ \mathcal{T} - \eta\mathcal{D}_2) \succeq 0 \quad \text{and}$$
$$\frac{1}{4}\mathcal{Q} + \text{Diag}\left(\mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* - \eta\mathcal{D}_1, \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* - \eta\mathcal{D}_2\right) \succ 0.$$

Remark 4.2 An interesting application of Theorem 4.1 is for the linearly constrained convex optimization problem with a quadratically coupled objective function of the form

$$\phi(w) = \frac{1}{2}\langle w, \widetilde{\mathcal{Q}}w\rangle + f(u) + g(v),$$

where $\widetilde{\mathcal{Q}} : \mathcal{U}\times\mathcal{V} \to \mathcal{U}\times\mathcal{V}$ is a self-adjoint positive semidefinite linear operator, $f : \mathcal{U} \to]-\infty, \infty[$ and $g : \mathcal{V} \to]-\infty, \infty[$ are two convex and smooth functions with Lipschitz continuous gradients. In this case, there exist four self-adjoint positive semidefinite operators $\Sigma_f, \widehat{\Sigma}_f : \mathcal{U} \to \mathcal{U}$ and $\Sigma_g, \widehat{\Sigma}_g : \mathcal{V} \to \mathcal{V}$ such that

$$\Sigma_f \preceq \xi \preceq \widehat{\Sigma}_f, \quad \forall \xi \in \partial^2 f(u),\ u \in \mathcal{U} \quad \text{and} \quad \Sigma_g \preceq \zeta \preceq \widehat{\Sigma}_g, \quad \forall \zeta \in \partial^2 g(v),\ v \in \mathcal{V},$$

where $\partial^2 f$ and $\partial^2 g$ are defined in (7). Then, by letting $\mathcal{Q} = \widetilde{\mathcal{Q}} + \text{Diag}\,(\Sigma_f, \Sigma_g)$ in (9) and $\mathcal{Q} + \mathcal{H} = \widetilde{\mathcal{Q}} + \text{Diag}\,(\widehat{\Sigma}_f, \widehat{\Sigma}_g)$ in (10), we have $\eta = 0$ in (12). This implies that $\mathcal{M} \succeq 0$ always holds in (32). Therefore, for $\tau \in]0, \frac{1+\sqrt{5}}{2}[$, the conditions for the convergence can be equivalently written as

$$\widetilde{\mathcal{Q}}_{11} + \Sigma_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \succ 0,\ \widetilde{\mathcal{Q}}_{22} + \Sigma_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* \succ 0 \tag{39}$$

and

$$\left\langle w, \left[\widetilde{\mathcal{Q}} + \text{Diag}\,\left(\Sigma_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^*, \Sigma_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^*\right)\right] w\right\rangle = 0 \Rightarrow \|u\|\|v\| = 0. \tag{40}$$

A sufficient condition for ensuring (39) and (40) to hold is

$$\widetilde{\mathcal{Q}} + \mathrm{Diag}\left(\Sigma_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^*,\, \Sigma_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^*\right) \succ 0.$$

If $\widetilde{\mathcal{Q}} = 0$, i.e., if the objective function of the original problem (1) is separable, then we will recover the convergence conditions given in [11] for a majorized ADMM with semi-proximal terms.

4.2 The Non-ergodic Iteration Complexity for General Coupled Objective Functions

In this subsection, we will present the non-ergodic iteration complexity for the majorized ADMM in terms of the KKT optimality condition.

Before showing the main result, we first present the following Lemma, which shows the nonincreasing property of the difference between two consecutive iteration points when the step length $\tau = 1$. This property has also been discussed by He and Yuan [12] for the classic ADMM with $\tau = 1$. For $k = 0, 1, 2, \ldots$, denote the following notation:

$$\Delta x^{k+1} := x^{k+1} - x^k, \quad \Delta u^{k+1} := u^{k+1} - u^k, \quad \Delta v^{k+1} := v^{k+1} - v^k,$$
$$\Delta w^{k+1} := w^{k+1} - w^k.$$

Lemma 4.1 *Assume that $\tau = 1$. Then, for any $k \geq 1$, it holds that*

$$\begin{aligned}&\left\|\Delta x^{k+1}\right\|^2_{\sigma^{-1}\mathcal{I}} + \left\|\Delta u^{k+1}\right\|^2_{\mathcal{S}} + \left\|\Delta v^{k+1}\right\|^2_{\mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*+\mathcal{Q}_{22}} + \left\|\Delta w^{k+1}\right\|^2_{\mathcal{Q}+\mathcal{H}}\\ &\leq \left\|\Delta x^{k}\right\|^2_{\sigma^{-1}\mathcal{I}} + \left\|\Delta u^{k}\right\|^2_{\mathcal{S}} + \left\|\Delta v^{k}\right\|^2_{\mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*+\mathcal{Q}_{22}} + \left\|\Delta w^{k}\right\|^2_{\mathcal{Q}+\mathcal{H}}. \end{aligned} \tag{41}$$

Proof For $\tau = 1$ and $k \geq 1$, the optimality conditions at the $(k+1)$th and kth iterations can be written as

$$\begin{aligned}-\mathcal{A}x^{k+1} - \nabla_u\phi\left(w^k\right) - \left(\mathcal{Q}_{11} + \mathcal{D}_1 + \mathcal{S}\right)\Delta u^{k+1} + \sigma\mathcal{A}\mathcal{B}^*\Delta v^{k+1} &\in \partial p\left(u^{k+1}\right),\\ -\mathcal{B}x^{k+1} - \nabla_v\phi\left(w^k\right) - \left(\mathcal{Q}_{22} + \mathcal{D}_2 + \mathcal{T}\right)\Delta v^{k+1} - \mathcal{Q}_{12}^*\Delta u^{k+1} &\in \partial q\left(v^{k+1}\right)\end{aligned}$$

and

$$\begin{aligned}-\mathcal{A}x^{k} - \nabla_u\phi\left(w^{k-1}\right) - \left(\mathcal{Q}_{11} + \mathcal{D}_1 + \mathcal{S}\right)\Delta u^{k} + \sigma\mathcal{A}\mathcal{B}^*\Delta v^{k} &\in \partial p\left(u^{k}\right),\\ -\mathcal{B}x^{k} - \nabla_v\phi\left(w^{k-1}\right) - \left(\mathcal{Q}_{22} + \mathcal{D}_2 + \mathcal{T}\right)\Delta v^{k} - \mathcal{Q}_{12}^*\Delta u^{k} &\in \partial q\left(v^{k}\right).\end{aligned}$$

By the monotonicity of the subdifferential of the convex functions p and q, we have the following inequalities:

$$\begin{aligned}
&\left\langle \Delta u^{k+1}, \mathcal{A}\Delta x^{k+1} + \left(\nabla_u \phi\left(w^k\right) - \nabla_u \phi\left(w^{k-1}\right)\right) + (\mathcal{Q}_{11} + \mathcal{D}_1 + \mathcal{S}) \right.\\
&\quad \left. \times \left(\Delta u^{k+1} - \Delta u^k\right) - \sigma \mathcal{A}\mathcal{B}^* \left(\Delta v^{k+1} - \Delta v^k\right)\right\rangle \le 0,\\
&\langle \Delta v^{k+1}, \mathcal{B}\Delta x^{k+1} + \left(\nabla_v \phi\left(w^k\right) - \nabla_v \phi\left(w^{k-1}\right)\right) + (\mathcal{Q}_{22} + \mathcal{D}_2 + \mathcal{T}) \\
&\quad \times \left(\Delta v^{k+1} - \Delta v^k\right) + \mathcal{Q}_{12}^* \left(\Delta u^{k+1} - \Delta u^k\right)\rangle \le 0.
\end{aligned}$$

Adding the above two inequalities together, and using the fact that

$$\Delta x^{k+1} - \Delta x^k = \sigma(\mathcal{A}^* \Delta u^{k+1} + \mathcal{B}^* \Delta v^{k+1}),$$

we get

$$\begin{aligned}
&\sigma^{-1}\left\langle \Delta x^{k+1}, \Delta x^{k+1} - \Delta x^k\right\rangle + \left\langle \Delta u^{k+1}, \mathcal{S}\left(\Delta u^{k+1} - \Delta u^k\right)\right\rangle \\
&+ \left\langle \Delta v^{k+1}, \mathcal{T}\left(\Delta v^{k+1} - \Delta v^k\right)\right\rangle \\
&+ \left\langle \Delta w^{k+1}, \nabla\phi\left(w^k\right) - \nabla\phi\left(w^{k-1}\right)\right. \\
&\left. + (\mathcal{Q} + \mathcal{H})\left(\Delta w^{k+1} - \Delta w^k\right)\right\rangle \\
&- \left\langle \Delta u^{k+1}, \left(\sigma\mathcal{A}\mathcal{B}^* + \mathcal{Q}_{12}\right)\left(\Delta v^{k+1} - \Delta v^k\right)\right\rangle \le 0. \qquad (42)
\end{aligned}$$

Since $\nabla\phi$ is globally Lipschitz continuous, there exists a self-adjoint positive semi-definite linear operator $\mathcal{W}^k \in \mathrm{conv}\{\partial^2\phi([w^{k-1}, w^k])\}$ such that

$$\nabla\phi\left(w^k\right) - \nabla\phi\left(w^{k-1}\right) = \mathcal{W}^k\left(w^k - w^{k-1}\right). \qquad (43)$$

By using the Cauchy-Schwarz inequality and (15), we see that

$$\begin{aligned}
&\left\langle \Delta u^{k+1}, \sigma\mathcal{A}\mathcal{B}^*\left(\Delta v^{k+1} - \Delta v^k\right)\right\rangle \\
&= \sigma\left\langle \mathcal{A}^*\Delta u^{k+1}, \mathcal{B}^*\Delta v^{k+1}\right\rangle - \sigma\left\langle \mathcal{A}^*\Delta u^{k+1}, \mathcal{B}^*\Delta v^k\right\rangle \\
&\le \frac{1}{2}\left(\sigma\left\|\mathcal{A}^*\Delta u^{k+1} + \mathcal{B}^*\Delta v^{k+1}\right\|^2 - \left\|\Delta u^{k+1}\right\|^2_{\sigma\mathcal{A}\mathcal{A}^*} - \left\|\Delta v^{k+1}\right\|^2_{\sigma\mathcal{B}\mathcal{B}^*}\right) \\
&\quad + \frac{1}{2}\left(\left\|\Delta u^{k+1}\right\|^2_{\sigma\mathcal{A}\mathcal{A}^*} + \left\|\Delta v^k\right\|^2_{\sigma\mathcal{B}\mathcal{B}^*}\right) \\
&= \frac{1}{2}\left\|\Delta x^{k+1} - \Delta x^k\right\|^2_{\sigma^{-1}\mathcal{I}} - \frac{1}{2}\left(\left\|\Delta v^{k+1}\right\|^2_{\sigma\mathcal{B}\mathcal{B}^*} - \left\|\Delta v^k\right\|^2_{\sigma\mathcal{B}\mathcal{B}^*}\right). \qquad (44)
\end{aligned}$$

By substituting (43) and (44) into (42) and some simple calculations, we have that

$$
\begin{aligned}
&\left(\left\|\Delta x^{k+1}\right\|^2_{\sigma^{-1}\mathcal{I}} - \left\|\Delta x^{k}\right\|^2_{\sigma^{-1}\mathcal{I}} + \left\|\Delta x^{k+1}-\Delta x^{k}\right\|^2_{\sigma^{-1}\mathcal{I}}\right)\\
&+\left(\left\|\Delta u^{k+1}\right\|^2_{\mathcal{S}} - \left\|\Delta u^{k}\right\|^2_{\mathcal{S}} + \left\|\Delta u^{k+1}-\Delta u^{k}\right\|^2_{\mathcal{S}}\right)\\
&+\left(\left\|\Delta v^{k+1}\right\|^2_{\mathcal{T}} - \left\|\Delta v^{k}\right\|^2_{\mathcal{T}} + \left\|\Delta v^{k+1}-\Delta v^{k}\right\|^2_{\mathcal{T}}\right)\\
&+2\left\|\Delta w^{k+1}\right\|^2_{\mathcal{Q}+\mathcal{H}}\\
&-\left(\left\|\Delta w^{k+1}\right\|^2_{\mathcal{Q}+\mathcal{H}-\mathcal{W}^k} - \left\|\Delta w^{k+1}-\Delta w^{k}\right\|^2_{\mathcal{Q}+\mathcal{H}-\mathcal{W}^k} + \left\|\Delta w^{k}\right\|^2_{\mathcal{Q}+\mathcal{H}-\mathcal{W}^k}\right)\\
&-\left(\left\|\Delta w^{k+1}\right\|^2_{\mathcal{Q}} + \left\|\Delta v^{k}\right\|^2_{\mathcal{Q}_{22}} - \left\|\Delta v^{k+1}\right\|^2_{\mathcal{Q}_{22}}\right)\\
&-\left(\left\|\Delta x^{k+1}-\Delta x^{k}\right\|^2_{\sigma^{-1}\mathcal{I}} - \left\|\Delta v^{k+1}\right\|^2_{\sigma\mathcal{B}\mathcal{B}^*} + \left\|\Delta v^{k}\right\|^2_{\sigma\mathcal{B}\mathcal{B}^*}\right) \le 0,
\end{aligned}
$$

which can be recast as

$$
\begin{aligned}
&\left\|\Delta x^{k+1}\right\|^2_{\sigma^{-1}\mathcal{I}} + \left\|\Delta u^{k+1}\right\|^2_{\mathcal{S}} + \left\|\Delta v^{k+1}\right\|^2_{\mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*+\mathcal{Q}_{22}} + \left\|\Delta w^{k+1}\right\|^2_{\mathcal{Q}+\mathcal{H}}\\
&\le \left\|\Delta x^{k}\right\|^2_{\sigma^{-1}\mathcal{I}} + \left\|\Delta u^{k}\right\|^2_{\mathcal{S}} + \left\|\Delta v^{k}\right\|^2_{\mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*+\mathcal{Q}_{22}} + \left\|\Delta w^{k}\right\|^2_{\mathcal{Q}+\mathcal{H}}\\
&\quad-\left(\left\|\Delta u^{k+1}-\Delta u^{k}\right\|^2_{\mathcal{S}} + \left\|\Delta v^{k+1}-\Delta v^{k}\right\|^2_{\mathcal{T}} + \left\|\Delta w^{k+1}-\Delta w^{k}\right\|^2_{\mathcal{Q}+\mathcal{H}-\mathcal{W}^k}\right.\\
&\quad\left.+\left\|\Delta w^{k+1}\right\|^2_{\mathcal{W}^k-\mathcal{Q}} + \left\|\Delta w^{k}\right\|^2_{\mathcal{W}_k}\right)\\
&\le \left\|\Delta x^{k}\right\|^2_{\sigma^{-1}\mathcal{I}} + \left\|\Delta u^{k}\right\|^2_{\mathcal{S}} + \left\|\Delta v^{k}\right\|^2_{\mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*+\mathcal{Q}_{22}} + \left\|\Delta w^{k}\right\|^2_{\mathcal{Q}+\mathcal{H}},
\end{aligned}
$$

where the last inequality is obtained by the relationship (8).

Theorem 4.2 *Suppose that the solution set of (1) is nonempty and Assumption 2.1 holds. Assume that one of the following conditions holds:*

(i) $\tau \in]0,1]$, $\mathcal{O}_1 := \frac{1}{4}\mathcal{Q} + Diag\left(\mathcal{S}+(1-\tau)\sigma\mathcal{A}\mathcal{A}^*, \mathcal{T}+(1-\tau)\sigma\mathcal{B}\mathcal{B}^*\right) \succ 0$;

(ii) $\tau \in]0, \frac{1+\sqrt{5}}{2}[$, $\frac{1}{4}\mathcal{Q} + Diag\ (\mathcal{S}-\eta\mathcal{D}_1, \mathcal{T}-\eta\mathcal{D}_2) \succeq 0$, $\mathcal{O}_2 := \frac{1}{4}\mathcal{Q} + Diag\,(\mathcal{S}+\sigma\mathcal{A}\mathcal{A}^*-\eta\mathcal{D}_1, \mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*-\eta\mathcal{D}_2) \succ 0$.

Then, there exists a constant C only depending on the initial point and the optimal solution set, such that the sequence $\{(u^k, v^k, x^k)\}$ generated by the majorized ADMM satisfies that for $k \ge 1$,

$$
\min_{1\le i\le k}\left[dist^2\left(0, F\left(u^{i+1}, v^{i+1}, x^{i+1}\right)\right) + \left\|\mathcal{A}^*u^{i+1}+\mathcal{B}^*v^{i+1}-c\right\|^2\right] \le C/k. \quad (45)
$$

and for the limiting case we have that

$$\lim_{k\to\infty} k\left(\min_{1\le i\le k}\left[dist^2\left(0, F\left(u^{i+1}, v^{i+1}, x^{i+1}\right)\right) + \left\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\right\|^2\right]\right) = 0. \tag{46}$$

Furthermore, when the step length $\tau = 1$, it holds that

$$\lim_{k\to\infty} k\left(dist^2\left(0, F\left(u^{k+1}, v^{k+1}, x^{k+1}\right)\right) + \left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\|^2\right) = 0. \tag{47}$$

i.e., the " $\min_{1\le i\le k}$ *" can be removed from* (46).

Proof From the optimality condition for (u^{k+1}, v^{k+1}), we know that

$$\begin{pmatrix} -(1-\tau)\sigma\mathcal{A}\left(\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right) - \sigma\mathcal{A}\mathcal{B}^*\left(v^k - v^{k+1}\right) - \mathcal{S}\left(u^{k+1} - u^k\right) \\ +\mathcal{Q}_{12}\left(v^{k+1} - v^k\right) - (1-\tau)\sigma\mathcal{B}\left(\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right) - \mathcal{T}\left(v^{k+1} - v^k\right) \end{pmatrix}$$
$$-(\mathcal{Q}+\mathcal{H})\left(w^{k+1} - w^k\right) + \nabla\phi\left(w^{k+1}\right) - \nabla\phi\left(w^k\right) \in F\left(u^{k+1}, v^{k+1}, x^{k+1}\right).$$

Therefore, we can obtain that

$$\begin{aligned}
&\mathrm{dist}^2\left(0, F\left(u^{k+1}, v^{k+1}, x^{k+1}\right)\right) + \left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\|^2 \\
&\le 5\left\|\sigma\mathcal{A}\mathcal{B}^*\left(v^{k+1} - v^k\right)\right\|^2 \\
&\quad +5(1-\tau)^2\sigma^2\left(\|\mathcal{A}\|^2 + \|\mathcal{B}\|^2\right)\left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\|^2 \\
&\quad +5\left\|(\mathcal{Q}+\mathcal{H})\left(w^{k+1} - w^k\right) - \nabla\phi\left(w^{k+1}\right) + \nabla\phi\left(w^k\right)\right\|^2 \\
&\quad +5\left\|\mathcal{Q}_{12}\left(v^{k+1} - v^k\right)\right\|^2 + 5\left\|\mathcal{T}\left(v^{k+1} - v^k\right)\right\|^2 \\
&\quad +5\left\|\mathcal{S}\left(u^{k+1} - u^k\right)\right\|^2 + \left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\|^2 \\
&\le 5\sigma\|\mathcal{A}\|^2\left\|v^{k+1} - v^k\right\|^2_{\sigma\mathcal{B}\mathcal{B}^*} \\
&\quad +\left(5(1-\tau)^2\sigma^2\left(\left\|\mathcal{A}\right\|^2 + \left\|\mathcal{B}\right\|^2\right) + 1\right)\left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\|^2 \\
&\quad +5\left\|\sqrt{\mathcal{Q}_{12}^*\mathcal{Q}_{12}}\right\|\left\|v^{k+1} - v^k\right\|^2_{\sqrt{\mathcal{Q}_{12}^*\mathcal{Q}_{12}}} \\
&\quad +5\|\mathcal{H}\|\left\|w^{k+1} - w^k\right\|^2_{\mathcal{H}} + 5\|\mathcal{S}\|\left\|u^{k+1} - u^k\right\|^2_{\mathcal{S}} \\
&\quad +5\|\mathcal{T}\|\left\|v^{k+1} - v^k\right\|^2_{\mathcal{T}} \le C_1\left\|w^{k+1} - w^k\right\|^2_{\widehat{\mathcal{O}}} \\
&\quad +C_2\left\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\right\|^2,
\end{aligned} \tag{48}$$

where

$$C_1 = 5\max(\sigma\|\mathcal{A}\|^2, \|\sqrt{\mathcal{Q}_{12}^*\mathcal{Q}_{12}}\|, \|\mathcal{H}\|, \|\mathcal{S}\|, \|\mathcal{T}\|),$$
$$C_2 = 5(1-\tau)^2\sigma^2(\|\mathcal{A}\|^2 + \|\mathcal{B}\|^2) + 1,$$
$$\widehat{\mathcal{O}} = \mathcal{H} + \mathrm{Diag}\,(\mathcal{S}, \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* + \sqrt{\mathcal{Q}_{12}^*\mathcal{Q}_{12}}),$$

and the second inequality comes from the fact that there exists some $\mathcal{W}^k \in \mathrm{conv}\{\partial^2\phi([w^{k-1}, w^k])\}$ such that

$$\begin{aligned}&\|(\mathcal{Q}+\mathcal{H})(w^{k+1}-w^k) - \nabla\phi(w^{k+1}) + \nabla\phi(w^k)\|^2\\ &\quad = \|(\mathcal{Q}+\mathcal{H}-\mathcal{W}^k)(w^{k+1}-w^k)\|^2 \le \|\mathcal{H}\|\|w^{k+1}-w^k\|_{\mathcal{H}}^2.\end{aligned}$$

Next, we will estimate the upper bounds for $\|w^{k+1}-w^k\|_{\widehat{\mathcal{O}}}^2$ and $\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\|^2$ by only involving the initial point and the optimal solution set under the two different conditions.

First, assume condition (i) holds. For $\tau \in]0, 1]$, by using (34) in the proof of Theorem 4.1, we have that,

$$\begin{aligned}&\|w^{i+1}-w^i\|_{\frac{1}{4}\mathcal{Q}+\mathrm{Diag}\,(\mathcal{S},\mathcal{T})}^2 + \sigma\|\mathcal{A}^*u^{i+1}\\ &\quad +\mathcal{B}^*v^i - c\|^2 + (1-\tau)\sigma\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2\\ &\le \Phi_i(\bar{u}, \bar{v}, \bar{x}) - \Phi_{i+1}(\bar{u}, \bar{v}, \bar{x}), \quad i \ge 1,\end{aligned}$$

which implies

$$\begin{aligned}&\sum_{i=1}^{k}(\|w^{i+1}-w^i\|_{\frac{1}{4}\mathcal{Q}+\mathrm{Diag}\,(\mathcal{S},\mathcal{T})}^2 + \sigma\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^i - c\|^2 + (1-\tau)\sigma\|\mathcal{A}^*u^{i+1}\\ &\quad +\mathcal{B}^*v^{i+1} - c\|^2 \le \Phi_1(\bar{u}, \bar{v}, \bar{x}) - \Phi_{k+1}(\bar{u}, \bar{v}, \bar{x}) \le \Phi_1(\bar{u}, \bar{v}, \bar{x}).\end{aligned}$$

This shows that

$$\begin{aligned}&\sum_{i=1}^{k}\|w^{i+1}-w^i\|_{\frac{1}{4}\mathcal{Q}+\mathrm{Diag}(\mathcal{S},\mathcal{T})}^2 \le \Phi_1(\bar{u}, \bar{v}, \bar{x}),\\ &\sum_{i=1}^{k}\sigma\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^i - c\|^2 \le \Phi_1(\bar{u}, \bar{v}, \bar{x}),\\ &\sum_{i=1}^{k}(1-\tau)\sigma\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2 \le \Phi_1(\bar{u}, \bar{v}, \bar{x}).\end{aligned} \tag{49}$$

From the above three inequalities we can also get that

$$(1-\tau)\sum_{i=1}^{k}\|u^{i+1}-u^{i}\|^2_{\sigma\mathcal{A}\mathcal{A}^*}\le(1-\tau)\sum_{i=1}^{k}(2\sigma\|\mathcal{A}^*u^{i+1}+\mathcal{B}^*v^{i}-c\|^2$$
$$+2\sigma\|\mathcal{A}^*u^{i}+\mathcal{B}^*v^{i}-c\|^2)\le 2(2-\tau)\Phi_1(\bar{u},\bar{v},\bar{x}),$$
$$(1-\tau)\sum_{i=1}^{k}\|v^{i+1}-v^{i}\|^2_{\sigma\mathcal{B}\mathcal{B}^*}\le(1-\tau)\sum_{i=1}^{k}(2\sigma\|\mathcal{A}^*u^{i+1}+\mathcal{B}^*v^{i}$$
$$-c\|^2+2\sigma\|\mathcal{A}^*u^{i+1}+\mathcal{B}^*v^{i+1}-c\|^2)\le 2(2-\tau)\Phi_1(\bar{u},\bar{v},\bar{x}).$$

With the notation of operator $\mathcal{O}_1$ we have that

$$\begin{aligned}\sum_{i=1}^{k}\|w^{i+1}-w^{i}\|^2_{\mathcal{O}_1}&=\sum_{i=1}^{k}\|w^{i+1}-w^{i}\|^2_{\frac{1}{4}\mathcal{Q}+\mathrm{Diag}(\mathcal{S},\mathcal{T})}\\&\quad+\sum_{i=1}^{k}\|w^{i+1}-w^{i}\|^2_{(1-\tau)\mathrm{Diag}(\sigma\mathcal{A}\mathcal{A}^*,\mathcal{B}\mathcal{B}^*)}\\&\le(9-4\tau)\Phi_1(\bar{u},\bar{v},\bar{x}).\end{aligned}\tag{50}$$

If $\tau\in]0,1[$, then we further have that

$$\sum_{i=1}^{k}\|\mathcal{A}^*u^{i+1}+\mathcal{B}^*v^{i+1}-c\|^2\le(1-\tau)^{-1}\sigma^{-1}\Phi_1(\bar{u},\bar{v},\bar{x}).\tag{51}$$

If $\tau=1$, then by the condition that $\mathcal{O}_1=\dfrac{1}{4}\mathcal{Q}+\mathrm{Diag}\,(\mathcal{S},\mathcal{T})\succ 0$, we have that

$$\begin{aligned}&\sum_{i=1}^{k}\|\mathcal{A}^*u^{i+1}+\mathcal{B}^*v^{i+1}-c\|^2\\&\le\sum_{i=1}^{k}(2\|\mathcal{A}^*u^{i+1}+\mathcal{B}^*v^{i}-c\|^2+2\|v^{i+1}-v^{i}\|^2_{\mathcal{B}\mathcal{B}^*})\\&\le\sum_{i=1}^{k}(2\|\mathcal{A}^*u^{i+1}+\mathcal{B}^*v^{i}-c\|^2+2\|\mathcal{O}_1^{-\frac{1}{2}}\mathrm{Diag}\,(0,\mathcal{B}\mathcal{B}^*)\mathcal{O}_1^{-\frac{1}{2}}\|\|w^{i+1}-w^{i}\|^2_{\mathcal{O}_1})\\&\le(2\sigma^{-1}+(18-8\tau)\|\mathcal{O}_1^{-\frac{1}{2}}\mathrm{Diag}\,(0,\mathcal{B}\mathcal{B}^*)\mathcal{O}_1^{-\frac{1}{2}}\|)\Phi_1(\bar{u},\bar{v},\bar{x}),\end{aligned}\tag{52}$$

where the second inequality is obtained by the fact that for any ξ, a self-adjoint positive definite operator $\mathcal{G}$ with square root $\mathcal{G}^{\frac{1}{2}}$ and a self-adjoint positive semidefinite operator $\widehat{G}$ defined in the same Euclidean space, it always holds that $\|\xi\|^2_{\widehat{\mathcal{G}}}=\langle\xi,\widehat{\mathcal{G}}\xi\rangle=\langle\xi,(\mathcal{G}^{\frac{1}{2}}\mathcal{G}^{-\frac{1}{2}})\widehat{\mathcal{G}}(\mathcal{G}^{-\frac{1}{2}}\mathcal{G}^{\frac{1}{2}})\xi\rangle=\langle\mathcal{G}^{\frac{1}{2}}\xi,(\mathcal{G}^{-\frac{1}{2}}\widehat{\mathcal{G}}\mathcal{G}^{-\frac{1}{2}})\mathcal{G}^{\frac{1}{2}}\xi\rangle\le\|\mathcal{G}^{-\frac{1}{2}}\widehat{\mathcal{G}}\mathcal{G}^{-\frac{1}{2}}\|\|\xi\|^2_{\mathcal{G}}$.

Therefore, by using (48), (50) and the positive definiteness of operator $\mathcal{O}_1$, we know that

$$\begin{aligned}&\min_{1\le i\le k}[\mathrm{dist}^2(0, F(u^{i+1}, v^{i+1}, x^{i+1})) + \|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2]\\ &\le (\sum_{i=1}^{k}(\mathrm{dist}^2(0, F(u^{i+1}, v^{i+1}, x^{i+1})) + \|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2))/k\\ &\le C\Phi_1(\bar{u}, \bar{v}, \bar{x})/k,\end{aligned}$$

where

$$C = \begin{cases} C_1(9-4\tau)\|\mathcal{O}_1^{-\frac{1}{2}}\widehat{\mathcal{O}}\mathcal{O}_1^{-\frac{1}{2}}\| + C_2(1-\tau)^{-1}\sigma^{-1}, & \tau \in]0, 1[,\\ C_1(9-4\tau)\|\mathcal{O}_1^{-\frac{1}{2}}\widehat{\mathcal{O}}\mathcal{O}_1^{-\frac{1}{2}}\| + C_2(2\sigma^{-1} + (18-8\tau)\|\mathcal{O}_1^{-\frac{1}{2}}\mathrm{Diag}\,(0, \mathcal{B}\mathcal{B}^*)\mathcal{O}_1^{-\frac{1}{2}}\|), & \tau = 1.\end{cases}$$

To prove the limiting case (46), by using inequalities (50), (51), (52) and [11, Lemma2.1], we have that

$$\min_{1\le i\le k}\|w^{i+1} - w^i\|^2_{\mathcal{O}_1} = o(1/k), \quad \min_{1\le i\le k}\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2 = o(1/k),$$

which, together with (48), imply that

$$\begin{aligned}&\lim_{k\to\infty} k(\min_{1\le i\le k}[\mathrm{dist}^2(0, F(u^{i+1}, v^{i+1}, x^{i+1})) + \|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2])\\ &\le \lim_{k\to\infty} k(\min_{1\le i\le k}[C_1\|\mathcal{O}_1^{-\frac{1}{2}}\widehat{\mathcal{O}}\mathcal{O}_1^{-\frac{1}{2}}\|\|w^{i+1} - w^i\|^2_{\mathcal{O}_1} + C_2\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2]) = 0.\end{aligned}$$

Next, we shall prove (47) under the condition (i) and $\tau = 1$. By (49), (52) and the positive definiteness of the self-adjoint linear operator $\frac{1}{4}\mathcal{Q} + \mathrm{Diag}\,(\mathcal{S}, \mathcal{T})$, we know that

$$\sum_{i=1}^{\infty}\|w^{i+1} - w^i\|^2 < \infty, \quad \sum_{i=1}^{\infty}\|x^{i+1} - x^i\|^2 < \infty. \tag{53}$$

Then, by using Lemma 4.1, (53) and [13, Lemma1.2], we know that

$$\lim_{k\to\infty} k(\|w^{k+1} - w^k\|^2_{\mathcal{Q}+\mathcal{H}+\mathrm{Diag}\,(\mathcal{S}, \mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*+\mathcal{Q}_{22})} + \|x^{k+1} - x^k\|^2_{\sigma^{-1}\mathcal{T}}) = 0.$$

Since $\mathcal{Q} + \mathcal{H} + \mathrm{Diag}\,(\mathcal{S}, \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* + \mathcal{Q}_{22}) \succeq \mathcal{Q} + \mathrm{Diag}\,(\mathcal{S}, \mathcal{T}) \succ 0$, we obtain that

$$\lim_{k\to\infty} k\|w^{k+1} - w^k\|^2 = 0, \quad \lim_{k\to\infty} k\|x^{k+1} - x^k\|^2_{\sigma^{-1}\mathcal{T}} = 0,$$

which, together with (48), imply

$$\lim_{k\to\infty} k(\mathrm{dist}^2(0, F(u^{k+1}, v^{k+1}, x^{k+1})) + \|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\|^2) = 0.$$

It completes the proof of the conclusions under condition (i).

For $\tau \in]0, \frac{1+\sqrt{5}}{2}[$, let $\rho(\tau) = \min(\tau, 1+\tau-\tau^2)$. We know from (32) that for $\tau \in]0, \frac{1+\sqrt{5}}{2}[$ and any $k \geq 1$,

$$\begin{aligned}&\sum_{i=1}^{k} \|w^{i+1} - w^i\|^2_{\frac{1}{4}\mathcal{Q}+\mathrm{diag}\,(\mathcal{S}-\eta\mathcal{D}_1, \mathcal{T}-\eta\mathcal{D}_2)} + \rho(\tau)\|v^{i+1} - v^i\|^2_{\sigma\mathcal{B}\mathcal{B}^*}\\ &\quad + \frac{\rho(\tau)}{\tau}\sigma\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2 \leq\\ &\quad (\Psi_1(\bar{u}, \bar{v}, \bar{x}) + \Xi_1) - (\Psi_{k+1}(\bar{u}, \bar{v}, \bar{x}) + \Xi_{k+1}) \leq \Psi_1(\bar{u}, \bar{v}, \bar{x}) + \Xi_1.\end{aligned}$$

Thus, by the positive semidefiniteness of $\frac{1}{4}\mathcal{Q} + \mathrm{Diag}\,(\mathcal{S}-\eta\mathcal{D}_1, \mathcal{T}-\eta\mathcal{D}_2)$, we can get that

$$\begin{aligned}&\sum_{i=1}^{k} \|w^{i+1} - w^i\|^2_{\frac{1}{4}\mathcal{Q}+\mathrm{diag}\,(\mathcal{S}-\eta\mathcal{D}_1, \mathcal{T}-\eta\mathcal{D}_2)} \leq \Psi_1(\bar{u}, \bar{v}, \bar{x}) + \Xi_1,\\ &\sum_{i=1}^{k} \|v^{i+1} - v^i\|^2_{\sigma\mathcal{B}\mathcal{B}^*} \leq (\Psi_1(\bar{u}, \bar{v}, \bar{x}) + \Xi_1)/\rho(\tau),\\ &\sum_{i=1}^{k} \sigma\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2 \leq \tau(\Psi_1(\bar{u}, \bar{v}, \bar{x}) + \Xi_1)/\rho(\tau),\end{aligned} \tag{54}$$

which implies

$$\begin{aligned}&\sum_{i=1}^{k} \|u^{i+1} - u^i\|^2_{\sigma\mathcal{A}\mathcal{A}^*} \leq \sum_{i=1}^{k}(3\sigma\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2 + 3\sigma\|\mathcal{A}^*u^i\\ &\quad + \mathcal{B}^*v^i - c\|^2 + 3\|v^{i+1} - v^i\|^2_{\sigma\mathcal{B}\mathcal{B}^*})\\ &\leq (6\tau+2)(\Psi_1(\bar{u}, \bar{v}, \bar{x}) + \Xi_1)/\rho(\tau).\end{aligned} \tag{55}$$

Combining (54) and (55) one can find that

$$\begin{aligned}\sum_{i=1}^{k} \|w^{i+1} - w^i\|^2_{\mathcal{O}_2} &= \sum_{i=1}^{k} \|w^{i+1} - w^i\|^2_{\frac{1}{4}\mathcal{Q}+\mathrm{diag}\,(\mathcal{S}-\eta\mathcal{D}_1, \mathcal{T}-\eta\mathcal{D}_2)}\\ &\quad + \sum_{i=1}^{k} \|w^{i+1} - w^i\|^2_{\mathrm{Diag}\,(\sigma\mathcal{A}\mathcal{A}^*, \sigma\mathcal{B}\mathcal{B}^*)}\\ &\leq (1 + (6\tau+3)/\rho(\tau))(\Psi_1(\bar{u}, \bar{v}, \bar{x}) + \Xi_1).\end{aligned} \tag{56}$$

Therefore, by using (48), (54), (56), and recalling the positive definiteness of operator $\mathcal{O}_2$, we finally have that

$$\begin{aligned}
&\min_{1\le i\le k}[\mathrm{dist}^2(0, F(u^{i+1}, v^{i+1}, x^{i+1})) + \|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2] \\
&\le \Big(\sum_{i=1}^{k}(\mathrm{dist}^2(0, F(u^{i+1}, v^{i+1}, x^{i+1})) + \|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2\Big)/k \\
&\le C'(\Psi_1(\bar{u}, \bar{v}, \bar{x}) + \Xi_1)/k,
\end{aligned}$$

where $C' = C_1\|\mathcal{O}_2^{-\frac{1}{2}}\widehat{\mathcal{O}}\mathcal{O}_2^{-\frac{1}{2}}\|(1 + (6\tau + 3)/\rho(\tau)) + C_2\sigma^{-1}\tau/\rho(\tau)$. The limiting property (46) and (47) can be derived in the same way as for the case under condition (i).

This completes the proof of Theorem 4.2. □

Remark 4.3 Theorem 4.2 gives the non-ergodic complexity of the KKT optimality condition, which does not seem to be known even for the classic ADMM with separable objective functions. For the latter, related results about the non-ergodic iteration complexity for the primal feasibility and the objective functions of the special classic ADMM with $\tau = 1$ can be found in Davis and Yin [14]. When $\tau \neq 1$, instead of showing the behavior of the current kth iteration point, we provide a non-ergodic complexity property on the "best point among the first k iterations", indicating that the iteration sequence may satisfy the $O(1/k)$ tolerance of the KKT system before the kth step. Thus, it may be of some interest to see whether the slightly better result with the "$\min_{1\le i\le k}$" being removed from (46) holds for $\tau \neq 1$.

4.3 The Ergodic Iteration Complexity for General Coupled Objective Functions

In this subsection, we will discuss the ergodic iteration complexity of the majorized ADMM for solving problem (1). For $k = 1, 2, \cdots$, denote

$$\hat{x}^k := \frac{1}{k}\sum_{i=1}^{k}\tilde{x}^{i+1}, \quad \hat{u}^k := \frac{1}{k}\sum_{i=1}^{k}u^{i+1}, \quad \hat{v}^k := \frac{1}{k}\sum_{i=1}^{k}v^{i+1}, \quad \hat{w}^k := (\hat{u}^k, \hat{v}^k)$$

and

$$\begin{aligned}
&\Lambda_{k+1} := \|u_e^{k+1}\|^2_{\mathcal{D}_1+\mathcal{S}} + \|v_e^{k+1}\|^2_{\mathcal{D}_2+\mathcal{T}+\mathcal{Q}_{22}+\sigma\mathcal{B}\mathcal{B}^*} + (\tau\sigma)^{-1}\|x^{k+1}\|^2, \\
&\overline{\Lambda}_{k+1} := \Lambda_{k+1} + \Xi_{k+1} + \|w_e^{k+1}\|^2_{\mathcal{Q}} + \max(1-\tau, 1-\tau^{-1})\sigma\|\mathcal{A}^*u^{k+1} + \mathcal{B}^*v^{k+1} - c\|^2.
\end{aligned}$$

Theorem 4.3 *Suppose that $\mathcal{S}$ and $\mathcal{T}$ are chosen such that*

$$\mathcal{Q}_{11} + \sigma\mathcal{A}\mathcal{A}^* + \mathcal{S} \succ 0 \quad \textit{and} \quad \mathcal{Q}_{22} + \sigma\mathcal{B}\mathcal{B}^* + \mathcal{T} \succ 0.$$

Assume that either (a) $\tau \in]0, 1]$ and (31) holds, or (b) $\tau \in]0, \frac{1+\sqrt{5}}{2}[$ and (32) and (33) hold. Then, there exist constants D_1 and D_2 that depend only on the initial point and the optimal solution set such that for $k \geq 1$, the following conclusions hold:

(i)

$$\|\mathcal{A}^*\hat{u}^k + \mathcal{B}^*\hat{v}^k - c\| \leq D_1/k. \tag{57}$$

(ii) *For case (b), if we further assume that $\mathcal{S} - \eta\mathcal{D}_1 \succeq 0$ and $\mathcal{T} - \eta\mathcal{D}_2 \succeq 0$, then*

$$|\theta(\hat{u}^k, \hat{v}^k) - \theta(\bar{u}, \bar{v})| \leq D_2/k. \tag{58}$$

The inequality (58) holds for case (a) without additional assumptions.

Proof (i) Under the conditions for case (a), (34) indicates that $\{\Phi_{k+1}(\bar{u}, \bar{v}, \bar{x})\}$ is a non-increasing sequence, which implies that

$$(\tau\sigma)^{-1}\|x^{k+1} - \bar{x}\|^2 \leq \Phi_{k+1}(\bar{u}, \bar{v}, \bar{x}) \leq \Phi_1(\bar{u}, \bar{v}, \bar{x}).$$

Similarly, under the conditions for case (b), we can get from (38) that

$$(\tau\sigma)^{-1}\|x^{k+1} - \bar{x}\|^2 \leq \Psi_{k+1}(\bar{u}, \bar{v}, \bar{x}) + \Xi_{k+1} \leq \Psi_1(\bar{u}, \bar{v}, \bar{x}) + \Xi_1.$$

Therefore, in terms of the ergodic primal feasibility, we have that

$$\begin{aligned}
&\|\mathcal{A}^*\hat{u}^k + \mathcal{B}^*\hat{v}^k - c\|^2 \\
&= \|\frac{1}{k}\sum_{i=1}^{k}(\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c)\|^2 \\
&= \|(\tau\sigma)^{-1}(x^{k+1} - x^1)\|^2/k^2 \\
&\leq 2\|(\tau\sigma)^{-1}(x^{k+1} - \bar{x})\|^2/k^2 + 2\|(\tau\sigma)^{-1}(x^1 - \bar{x})\|^2/k^2 \leq C_3/k^2,
\end{aligned} \tag{59}$$

where

$$C_3 = \begin{cases} 2(\tau\sigma)^{-1}\Phi_1(\bar{u}, \bar{v}, \bar{x}) + 2\|(\tau\sigma)^{-1}(x^1 - \bar{x})\|^2 & \text{for case (a)}, \\ 2(\tau\sigma)^{-1}(\Psi_1(\bar{u}, \bar{v}, \bar{x}) + \Xi_1) + 2\|(\tau\sigma)^{-1}(x^1 - \bar{x})\|^2 & \text{for case (b)}. \end{cases}$$

Then, by taking the square root on inequality (59), we can obtain (57).

(ii) For the complexity of primal objective functions, first, we know from (13) that

$$\begin{aligned}
p(u) &\geq p(\bar{u}) + \langle -\mathcal{A}\bar{x} - \nabla_u\phi(\bar{w}), u - \bar{u}\rangle, \quad \forall u \in \mathcal{U}, \\
q(v) &\geq q(\bar{v}) + \langle -\mathcal{B}\bar{x} - \nabla_v\phi(\bar{w}), v - \bar{v}\rangle, \quad \forall v \in \mathcal{V}.
\end{aligned}$$

Therefore, summing them up and by noting $\mathcal{A}^*\bar{u} + \mathcal{B}^*\bar{v} = c$ and the convexity of function ϕ, we have that

$$\begin{aligned}
\theta(u, v) - \theta(\bar{u}, \bar{v}) &\geq -\langle \bar{x}, \mathcal{A}^*u + \mathcal{B}^*v - c\rangle + \phi(w) - \phi(\bar{w}) - \langle \nabla\phi(\bar{w}), w - \bar{w}\rangle \\
&\geq -\langle \bar{x}, \mathcal{A}^*u + \mathcal{B}^*v - c\rangle, \quad \forall u \in \mathcal{U}, v \in \mathcal{V}.
\end{aligned}$$

Thus, with $(u, v) = (\hat{u}^k, \hat{v}^k)$, it holds that

$$\begin{aligned}\theta(\hat{u}^k, \hat{v}^k) - \theta(\bar{u}, \bar{v}) &\geq -\langle \bar{x}, \mathcal{A}^*\hat{u}^k + \mathcal{B}^*\hat{v}^k - c\rangle \\ &\geq -\frac{1}{2}\left(\frac{1}{k}\|\bar{x}\|^2 + k\|\mathcal{A}^*\hat{u}^k + \mathcal{B}^*\hat{v}^k - c\|^2\right) \\ &\geq -\frac{1}{2}(\|\bar{x}\|^2 + C_3)/k, \end{aligned} \tag{60}$$

where C_3 is the same constant as in (59).

For the reverse part, by (9) and (10) we can obtain that for any $i \geq 1$,

$$\begin{aligned}\phi(w^{i+1}) &\leq \phi(w^i) + \langle \nabla\phi(w^i), w^{i+1} - w^i\rangle + \frac{1}{2}\|w^{i+1} - w^i\|^2_{\mathcal{Q}+\mathcal{H}}, \\ \phi(\bar{w}) &\geq \phi(w^i) + \langle \nabla\phi(w^i), \bar{w} - w^i\rangle + \frac{1}{2}\|\bar{w} - w^i\|^2_{\mathcal{Q}},\end{aligned}$$

which indicates that

$$\phi(w^{i+1}) - \phi(\bar{w}) \leq \langle \nabla\phi(w^i), w^{i+1} - \bar{w}\rangle + \frac{1}{2}\|w^{i+1} - w^i\|^2_{\mathcal{Q}+\mathcal{H}} - \frac{1}{2}\|w^i - \bar{w}\|^2_{\mathcal{Q}}. \tag{61}$$

Thus, (22) and (61) imply that for $\tau \in]0, 1]$ and any $i \geq 1$,

$$\begin{aligned}&\theta(u^{i+1}, v^{i+1}) - \theta(\bar{u}, \bar{v}) \\ &\leq \frac{1}{2}\|w^{i+1} - w^i\|^2_{\mathcal{Q}+\mathcal{H}} - \frac{1}{2}\|w^i - \bar{w}\|^2_{\mathcal{Q}} + \langle \bar{w} - w^{i+1}, \mathcal{Q}(w^{i+1} - w^i)\rangle \\ &\quad + \langle \bar{u} - u^{i+1}, \mathcal{A}\tilde{x}^{i+1}\rangle + \langle \bar{v} - v^{i+1}, \mathcal{B}\tilde{x}^{i+1}\rangle - \langle v^{i+1} - v^i, \mathcal{Q}^*_{12}(\bar{u} - u^{i+1})\rangle \\ &\quad + \sigma\langle \mathcal{A}^*(u^{i+1} - \bar{u}), \mathcal{B}^*(v^{i+1} - v^i)\rangle + \langle \bar{u} - u^{i+1}, (\mathcal{D}_1 + \mathcal{S})(u^{i+1} - u^i)\rangle \\ &\quad + \langle \bar{v} - v^{i+1}, (\mathcal{D}_2 + \mathcal{T})(v^{i+1} - v^i)\rangle \\ &\leq \frac{1}{2}(\Lambda_i - \Lambda_{i+1}) - \frac{1}{2}(\|u^{i+1} - u^i\|^2_{\mathcal{S}} + \|v^{i+1} - v^i\|^2_{\mathcal{T}} + \sigma\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^i - c\|^2 \\ &\quad + \sigma(1 - \tau)\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2) \leq \frac{1}{2}(\Lambda_i - \Lambda_{i+1}).\end{aligned} \tag{62}$$

Therefore, summing up the above inequalities over $i = 1, \cdots k$ and by using the convexity of function θ we can obtain that

$$\theta(\hat{u}_k, \hat{v}_k) - \theta(\bar{u}, \bar{v}) \leq (\Lambda_1(u, v, x) - \Lambda_{k+1}(u, v, x))/2k \leq \Lambda_1/2k. \tag{63}$$

The inequalities (60) and (63) indicate that (58) holds for case (a).

Next, assume that the conditions for case (b) hold. Similar to (62), we have that

$$\begin{aligned}\theta(u^{i+1}, v^{i+1}) - \theta(\bar{u}, \bar{v}) \\ \leq \frac{1}{2}(\overline{\Lambda}_i - \overline{\Lambda}_{i+1}) - \frac{1}{2}(\|u^{i+1} - u^i\|^2_{\mathcal{S}-\eta\mathcal{D}_1} + \|v^{i+1} - v^i\|^2_{\mathcal{T}+\min(\tau,1+\tau-\tau^2)\sigma\mathcal{B}\mathcal{B}^*-\eta\mathcal{D}_2} \\ + \min(1, 1+\tau^{-1}-\tau)\sigma\|\mathcal{A}^*u^{i+1} + \mathcal{B}^*v^{i+1} - c\|^2).\end{aligned}$$

By the assumptions that $\mathcal{S} - \eta\mathcal{D}_1 \succeq 0$ and $\mathcal{T} - \eta\mathcal{D}_2 \succeq 0$, we can obtain that

$$\theta(\hat{u}^k, \hat{v}^k) - \theta(\bar{u}, \bar{v}) \leq (\overline{\Lambda}_1 - \overline{\Lambda}_{k+1})/2k \leq \overline{\Lambda}_1/2k. \tag{64}$$

Thus, by (63) and (64) we can obtain (58). □

Below we make a remark about the results in Theorem 4.3.

Remark 4.4 The results in Theorem 4.3, which are on the ergodic complexity of the primal feasibility and the objective function, respectively, are extended from the work of Davis and Yin [14] on the classic ADMM with separable objective functions. However, there is no corresponding result available on the dual problem. Therefore, it will be very interesting to see if one can develop a more explicit ergodic complexity result containing all the three parts in the KKT condition.

5 Conclusions

In this paper, we establish the convergence properties for the majorized ADMM with a large step length to solve linearly constrained convex programming, whose objective function includes a coupled smooth function. From Theorem 4.1, one can see the influence of the coupled objective on the convergence condition. For $\tau \in]0, \frac{1+\sqrt{5}}{2}[$, a joint condition like (31) or (33) is needed to analyze the behaviour of the iteration sequence. One can further observe that the parameter η, which controls the off-diagonal term of the generalized Hessian, also affects the choice of proximal operators $\mathcal{S}$ and $\mathcal{T}$. However, as is pointed out in Remark 4.2, when the coupled function is convex quadratic, $\eta = 0$ and the corresponding influence would disappear. Although, in this paper we focus on the two-block case, it is not hard to see that, with the help of the Schur complement technique introduced in [15], one can apply our majorized ADMM to solve large scale convex optimization problems with many smooth blocks.

Acknowledgments The authors would like to thank Caihua Chen at the Nanjing University for discussions on the iteration complexity described in the paper, and Bo Chen at the National University of Singapore for the comments on the global convergence conditions in Theorem 4.1. The research of Defeng Sun was supported in part by the Academic Research Fund (Grant No. R-146-000-207-112). The research of Kim-Chuan Toh was supported in part by the Ministry of Education, Singapore, Academic Research Fund (Grant No. R-146-000-194-112).

References

1. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin M., Glowinski R. (eds.) Augmented Lagrangian Methods: Applications to the Solution of Boundary Problems, pp. 299–331. North–Holland, Amsterdam (1983)
2. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**(1), 17–40 (1976)
3. Glowinski, R.: Lectures on Numerical Methods for Non-linear Variational Problems, Tata Institute of Fundamental Research Lectures on Mathematics and Physics, Notes by Vijayasundaram, G and Adimurthi, M, vol. **65**. Springer, Berlin (1980)
4. Glowinski, R., Marroco, A.: Approximation by finite elements of order one and solution by penalization-duality of a class of nonlinear dirichlet problems. Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique **9**(2), 41–76 (1975)
5. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**(1), 293–318 (1992)
6. Eckstein, J., Yao, W.: Understanding the convergence of the alternating direction method of multipliers: theoretical and computational perspectives. Pac. J. Optim. **11**(4), 619–644 (2014)
7. Hong, M., Chang, T.H., Wang, X., Razaviyayn, M., Ma, S., Luo, Z.Q.: A block successive upper bound minimization method of multipliers for linearly constrained convex optimization. arXiv preprint arXiv:1401.7079 (2014)
8. Chen, C., He, B., Ye, Y., Yuan, X.: The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. Math. Program. **155**(1), 57–79 (2016)
9. Clarke, F.H.: Optimization and Nonsmooth Analysis, vol. 5. SIAM, Philadelphia (1990)
10. Hiriart-Urruty, J.B., Strodiot, J.J., Nguyen, V.H.: Generalized hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data. Appl. Math. Optim. **11**(1), 43–56 (1984)
11. Li, M., Sun, D., Toh, K.C.: A majorized admm with indefinite proximal terms for linearly constrained convex composite optimization. SIAM J. Optim. (to appear)
12. He, B., Yuan, X.: On non-ergodic convergence rate of Douglas–Rachford alternating direction method of multipliers. Numerische Mathematik **130**(3), 567–577 (2012)
13. Deng, W., Lai, M., Peng, Z., Yin, W.: Parallel multi-block admm with o(1/k) convergence. arXiv preprint arXiv:1312.3040 (2013)
14. Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. arXiv preprint arXiv:1406.4834 (2014)
15. Li, X., Sun, D., Toh, K.C.: A schur complement based semi-proximal admm for convex quadratic conic programming and extensions. Math. Program. **155**(1), 333–373 (2016)