Iteration complexity analysis of a partial LQP-based alternating direction method of multipliers *

Jianchao Bai† Yuxue Ma‡ Hao Sun § Miao Zhang ¶

Abstract

In this paper, we consider a prototypical convex optimization problem with multi-block variables and separable structures. By adding the Logarithmic Quadratic Proximal (LQP) regularizer with suitable proximal parameter to each of the first grouped subproblems, we develop a partial LQP-based Alternating Direction Method of Multipliers (ADMM-LQP). The dual variable is updated twice with relatively larger stepsizes than the classical region $(0, \frac{1+\sqrt{5}}{2})$. Using a prediction-correction approach to analyze properties of the iterates generated by ADMM-LQP, we establish its global convergence and sublinear convergence rate of $O(1/T)$ in the new ergodic and nonergodic senses, where T denotes the iteration index. We also extend the algorithm to a nonsmooth composite convex optimization and establish similar convergence results as our ADMM-LQP.

Keywords Convex optimization; Alternating direction method of multipliers; Proximal term; Larger stepsize; Convergence complexity
Mathematics Subject Classification(2010) 65K10; 65Y20; 90C25

1 Introduction

Consider the following multi-block separable convex optimization problem

$$\begin{array}{ll} \min & \sum_{i=1}^{p} f_i(x_i) + g(y) \\ \text{s.t.} & \sum_{i=1}^{p} A_i x_i + By = b, \\ & x_i \in \mathbb{R}_+^{m_i},\ y \in \mathcal{Y}\ , i = 1, \cdots, p, \end{array} \tag{1}$$

where $f_i(x_i) : \mathbb{R}_+^{m_i} \to \mathbb{R},\ g(y) : \mathbb{R}^d \to \mathbb{R}$ are closed proper convex functions (but not necessarily smooth/ strongly convex); $A_i \in \mathbb{R}^{n\times m_i}, B \in \mathbb{R}^{n\times d}$ and $b \in \mathbb{R}^n$ are given data; $\mathcal{Y} \subset \mathbb{R}^d$ is a closed convex set and $p \geq 1$ is an integer. Throughout our discussions, the solution set of the problem (1) is assumed to be nonempty and all of matrices $A_i (i = 1, \ldots, p)$ and B have full column

*Authors are ranked by family name with equal contributions. Part of this research was supported by National Natural Science Foundation of China (12001430, 72071158), Fundamental Research Funds for the Central Universities (G2020KY05203), and China Postdoctoral Science Foundation (2020M683545).

†jianchaobai@nwpu.edu.cn, https://teacher.nwpu.edu.cn/jcbai, School of Mathematics and Statistics and the MIIT Key Laboratory of Dynamics and Control of Complex Systems, Northwestern Polytechnical University, Xi'an 710129, China.

‡✉ Corresponding author. Mayuxue708991@163.com, School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, China.

§✉ Corresponding author. hsun@nwpu.edu.cn, School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, China.

¶mzhan33@lsu.edu, Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803-4918, USA

ranks. For convenience of analysis, we denote $A = (A_1, \ldots, A_p)$, $\mathbf{x} = (x_1^\mathsf{T}, \ldots, x_p^\mathsf{T})^\mathsf{T}$, $\mathcal{X} = \mathbb{R}_+^{m_1} \times \cdots \times \mathbb{R}_+^{m_p}$ and $\mathcal{W} = \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^n$.

Problems in the form of (1) cover a large number of practical applications, for example, traffic assignment, network economics, game theoretic and cancer diagnostics problems, see [10, 15, 22, 24, 30]. Here we take three concrete motivating examples:

- **Traffic network equilibrium problem**. Consider a directed traffic network including 25 nodes, 37 links, and 6 origin/destination (O/D) pairs. In order to avoid traffic congestion, it is common to require the link flow to be bounded from above, that is, the path flow x should satisfy $x \in S := \{x \in \mathbb{R}^{55} | A^\mathsf{T} x \leq b, x \geq \mathbf{0}\}$, where $A \in \mathbb{R}^{55\times 37}$ is the path-link incidence matrix and b is the given link capacity vector. Then, the traffic network equilibrium problem can be described to seek a path flow pattern x^* such that

$$x^* \geq \mathbf{0}, \quad \langle x - x^*, f(x^*)\rangle \geq 0, \quad \forall x \in S, \tag{2}$$

  where $f(x^*) = At(A^\mathsf{T} x) - E\eta(E^\mathsf{T} x)$, $E \in \mathbb{R}^{55\times 6}$ is the path-O/D pair incidence matrix, t represents a given link travel cost vector and η denotes the travel disutility determined by O/D pair. As discussed in [15], the traffic equilibrium problem (2) can be mathematically characterized as the form (10), i.e., an equivalent form of Problem (1) with $\theta_2(y) \equiv 0, B = \mathbf{I}, p = 1$ and $\mathcal{Y} = \mathbb{R}_+^{55}$. Performance of the two-block case of our proposed algorithm for solving this problem had been verified in the experiments [30, Section 5].

- **Sparse signal processing problem**. As stated in [32], many signal processing problem in medical imaging, computed tomography and so forth can be mathematically modeled as a large-scale linear equations $\mathcal{A}\mathbf{x} = b$. In order to find a sparse nonnegative solution to this equation with e.g. $\mathcal{A} \in \mathbb{R}^{10000\times 5000}$ and $b \in \mathbb{R}^{10000}$, an effective way is to split $\mathcal{A} = [A_1, A_2, \cdots, A_{10}]$ and $\mathbf{x} = (x_1, x_2, \cdots, x_{10})^\mathsf{T}$ where $x_i \in \mathbb{R}_+^{500}$. Hence, the problem can be converted to the following small-scale minimization problem:

$$\begin{array}{ll} \min & \|x_1\|_1 + \|x_2\|_1 + \cdots + \|x_{10}\|_1 \\ \text{s.t.} & A_1 x_1 + A_2 x_2 + \cdots + A_{10} x_{10} = b,\ x_i \geq \mathbf{0}, i = 1, \cdots, 10. \end{array}$$

  Clearly, this reformulated problem is a multi-block case of Problem (1) with $p = 9, f_i = \|x_i\|_1, g = \|x_{10}\|_1, B = A_{10}$ and $\mathcal{Y} = \mathbb{R}_+^{500}$.

- **Linear programming with box constraints**. Linear programming can date back to the time of Fourier who published a method to solve problems with a system of linear inequalities, see e.g. [3, page 257]. It is a widely studied field in optimization and covers many practical problems, e.g. network flow problems and the model of trolley network [28]. Linear programming problems with box-type constraints can be expressed as

$$\begin{array}{ll} \min & c^\mathsf{T} z \\ \text{s.t.} & Bz = b, l \leq z \leq u, \end{array}$$

  where $c, l, u \in \mathbb{R}^n$, $B \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ are given. Its dual problem reads

$$\begin{array}{ll} \min & u^\mathsf{T} x_1 - l^\mathsf{T} x_2 + b^\mathsf{T} y \\ \text{s.t.} & x_1 - x_2 + B^\mathsf{T} y = -c, \\ & x_1 \geq \mathbf{0}, x_2 \geq \mathbf{0}, y \in \mathbb{R}^m, \end{array}$$

  which is clearly the form of Problem (1) with three blocks.

A prototype method for solving the equality constrained problem in the form of (1) is the Augmented Lagrangian Method (ALM, [14]). For any $\beta > 0$, by constructing the augmented Lagrange function associated with (1):

$$\mathcal{L}_\beta(\mathbf{x}, y, \lambda) = \mathcal{L}(\mathbf{x}, y, \lambda) + \frac{\beta}{2} \|A\mathbf{x} + By - b\|^2, \tag{3}$$

where

$$\mathcal{L}(\mathbf{x}, y, \lambda) = \sum_{i=1}^{p} f_i(x_i) + g(y) - \langle \lambda, A\mathbf{x} + By - b \rangle \,, \tag{4}$$

ALM firstly updates the primal variables by solving the jointed subproblem

$$\min_{(\mathbf{x},y)\in\mathcal{X}\times\mathcal{Y}} \mathcal{L}_\beta(\mathbf{x}, y, \lambda^k)$$

and then updates the Lagrange multipliers by $\lambda^{k+1} = \lambda^k - \beta(A\mathbf{x}^{k+1} + By^{k+1} - b)$. However, ALM does not make full use of the separable structure of the objective functions and hence, could not take advantage of some special properties of each component function; furthermore, in many real applications involving big-data, solving this jointed subproblem would be very expensive or even difficult at each iteration.

An effective approach to overcome the above disadvantages is the Alternating Direction Method of Multipliers (ADMM) which could be regarded as a splitting version of ALM:

$$\begin{cases} x_1^{k+1} & \leftarrow & \arg\min_{x_1\in\mathbb{R}_+^{m_1}} \mathcal{L}_\beta(x_1, x_2^k, \cdots, x_p^k, y^k, \lambda^k), \\ \vdots & & \\ x_p^{k+1} & \leftarrow & \arg\min_{x_p\in\mathbb{R}_+^{m_p}} \mathcal{L}_\beta(x_1^{k+1}, \cdots, x_{p-1}^{k+1}, x_p, y^k, \lambda^k), \\ y^{k+1} & \leftarrow & \arg\min_{y\in\mathcal{Y}} \mathcal{L}_\beta(x_1^{k+1}, \cdots, x_p^{k+1}, y, \lambda^k), \\ \lambda^{k+1} & \leftarrow & \lambda^k - \beta(A\mathbf{x}^{k+1} + By^{k+1} - b). \end{cases} \tag{5}$$

Compared ADMM (5) to ALM, each subproblem of ADMM annihilates the coupled term $\beta\langle A\mathbf{x}, By\rangle$ and makes its iteration independent on other variables, so ADMM can be quite effective for solving large-scale problems with preferred structures. As commented by Boyd et al. [2], "ADMM is at least comparable to very specialized algorithms (even in the serial setting), and in most cases, the simple ADMM algorithm will be efficient enough to be useful." Although the classical ADMM [12, 13] was demonstrated convergent for the two-block separable convex optimization, its direct extension in the Gauss-Seidel scheme (5) is not necessarily convergent [8] without special assumptions on the coefficient matrix in the equality constraints. Besides, it was pointed out by He et al. [17] that the following Jacobi-type extension of ADMM:

$$\begin{cases} x_1^{k+1} & \leftarrow & \arg\min_{x_1\in\mathbb{R}_+^{m_1}} \mathcal{L}_\beta(x_1, x_2^k, \cdots, x_p^k, y^k, \lambda^k), \\ \vdots & & \\ x_p^{k+1} & \leftarrow & \arg\min_{x_p\in\mathbb{R}_+^{m_p}} \mathcal{L}_\beta(x_1^{k}, \cdots, x_{p-1}^{k}, x_p, y^k, \lambda^k), \\ y^{k+1} & \leftarrow & \arg\min_{y\in\mathcal{Y}} \mathcal{L}_\beta(x_1^{k+1}, \cdots, x_p^{k+1}, y, \lambda^k), \\ \lambda^{k+1} & \leftarrow & \lambda^k - \beta(A\mathbf{x}^{k+1} + By^{k+1} - b), \end{cases} \tag{6}$$

is not necessarily convergent either.

Since the development of ADMM in 1970s, a great majority of ADMM-type methods were proposed for solving the multi-block separable convex optimization problems. Particularly, many researchers focused on modifying the above scheme (5) or (6) by adding proper proximal terms into the subproblems of ADMM, which aims to promote the algorithmic convergence and also to overcome the obstacle that each involved subproblem is not easily solvable. For instance, a proximal ADMM using the Gauss-Seidel approach (5) was developed in [18] to solve the 3-block separable convex problem, in which proximal terms $\frac{\tau\beta}{2}\|A_i(\mathbf{x}_i - \mathbf{x}_i^k)\|^2$ were added for the second and third subproblems, and global convergence was guaranteed when the proximal parameter $\tau \in [1, +\infty)$. To deal with the convex problem (1) with N-block variables

($N \geq 3$), a proximal ADMM using the Jacobian approach (6) was proposed in [19], where the terms in the form of $\frac{s\beta}{2}\|A_i(\mathbf{x}_i - \mathbf{x}_i^k)\|^2$ were added to each generated subproblem and its global convergence was established when the proximal parameter $s \in [N-1, +\infty)$. In the recent work [20, 21], the proximal point techniques were further studied and global convergence of the algorithms therein was established with the aid of a prediction-correction approach for the iterative sequence. More recently, by designing a positive proximal term for each x_i-subproblem, Bai et al. [5, the scheme (74)] developed a generalized symmetric ADMM whose dual variable was updated twice with larger stepsizes than the region $(0, \frac{1+\sqrt{5}}{2})$. Under the assumptions that all subdifferential of each component objective function is piecewise linear multi-functions, Bai et al. [6] even established the linear convergence rate of this generalized symmetric ADMM. We refer interested readers to [9, 11, 23, 25, 31, 30] for other proximal ADMMs by exploiting general proximal terms to solve the two block and multiple block cases.

It is a fact acknowledged that the performance of ADMM-type methods depends significantly on the difficulty of solving each subproblem. For the problem (1), the involved subproblems usually do not have closed-form solutions and need to be solved approximately by inexact methods or some inner iterative algorithms. Fortunately, for the nonnegative linearly constrained problem the Logarithmic Quadratic Proximal (LQP, [1]) regularizer can convert the splitting subproblem to a system of nonlinear equations, which is comparatively easier than the original one. Another merit is that LQP regularization can ensure that the solution of involved subproblem stays strictly within the interior of positive orthant. Mainly motivated by the advantages of LQP regularization and previous researches [5, 27, 29, 30], in this paper we propose a partial LQP-based ADMM (**ADMM-LQP**), which is described formally in the following table, with larger stepsizes of dual variables to solve the general problem (1). In our proposed algorithm ADMM-LQP, $r_i > 0$ plays the role of proximal parameter, $d(\cdot, \cdot)$ is a specified LQP regularizer given by Definition 1 and the dual stepsizes

$$(\alpha, \tau) \in \mathcal{K} := \left\{(\alpha, \tau) \mid 1 > \alpha > -1,\ \alpha + \tau > 0,\ 1 + \alpha + \tau - \alpha\tau - \alpha^2 - \tau^2 > 0\right\}. \tag{7}$$

Initialize $(x_1^0, \cdots, x_p^0, y^0, \lambda^0)$ and choose $(\alpha, \tau) \in \mathcal{K}, \beta > 0$;
While stopping criteria is not satisfied **do**
  For $i = 1, 2, \cdots, p,$ **do**
    $x_i^{k+1} \leftarrow \arg\min_{x_i \in \mathbb{R}_+^{m_i}} \mathcal{L}_\beta(x_1^k, \cdots, x_{i-1}^k, x_i, x_{i+1}^k \cdots, x_p^k, y^k, \lambda^k) + r_i d(x_i, x_i^k)$;
  End for
  $\lambda^{k+\frac{1}{2}} \leftarrow \lambda^k - \alpha\beta(A\mathbf{x}^{k+1} + By^k - b)$;
  $y^{k+1} \leftarrow \arg\min_{y \in \mathcal{Y}} \mathcal{L}_\beta(x_1^{k+1}, \cdots, x_p^{k+1}, y, \lambda^{k+\frac{1}{2}})$;
  $\lambda^{k+1} \leftarrow \lambda^{k+\frac{1}{2}} - \tau\beta(A\mathbf{x}^{k+1} + By^{k+1} - b)$;
  $k \leftarrow k + 1$;
End while

The rest of this paper is organized as follows. In Section 2, we summarize some fundamental preliminaries including the LQP regularization term and its related results, the first-order optimality condition of the problem whose objective function is a smooth function plus nonsmooth function, as well as the variational characterization for the saddle point of the problem and the iterates of the proposed algorithm. In Section 3, we analyze the global convergence of ADMM-LQP with the worst-case $O(1/T)$ convergence rate for a new average iterate $\mathbf{w}_T := \frac{1}{1+T}\sum_{k=\kappa}^{\kappa+T} \tilde{\mathbf{w}}^k$, where $\kappa \geq 0, T > 0$ are integers. Section 4 discusses and analyzes an extended version of ADMM-LQP. Finally, we conclude the paper in Section 5.

Notations Let $\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{m\times n}$ be the sets of real numbers, n dimensional real column vectors and $m \times n$ dimensional real matrices, respectively. In particular, $\mathbb{R}_{++}^n(\mathbb{R}_+^n)$ denotes the set of

n dimensional real positive (nonnegative) vectors. The bold $\mathbf{I}$ denotes the identity matrix and $\mathbf{0}$ denotes zero matrix/vector with proper dimensions. For any symmetric matrix G, we define $\|x\|_G^2 := x^\top Gx$ where the superscript $^\top$ denotes the transpose. Note that G could be indefinite with $x^\top Gx < 0$ for some x. If G is positive definite, then we call $\|x\|_G$ G-weighted norm. The notations $\|\cdot\|, \langle\cdot,\cdot\rangle$ and $\nabla_x f(x,y)$ denote the standard Euclidean norm, inner product, and the partial derivative of the differentiable function $f(x,y)$ at x, respectively.

2 Preliminaries

The forthcoming Definition 1 and Lemma 1, which are very useful for analyzing convergence properties of ADMM-LQP in the sequel sections, can be respectively found in the earlier work [1] and [26]. It is clear from Definition 1 that $d(\cdot,z)$ is a closed proper convex function and $d(\cdot,\cdot)$ is nonnegative. $d(v,z)=0$ if and only if $v=z$. Moreover,

$$\nabla_v d(v,z) = (v-z) + \mu(z - Z^2 v^{-1}),$$

where $Z = \mathrm{diag}(z_1, z_2, \cdots, z_n) \in \mathbb{R}^{n\times n}$, $v^{-1} \in \mathbb{R}^n$ denotes a vector whose jth element is $1/v_j$.

Definition 1 *Let $0<\mu<1$ be a given constant. For $z \in \mathbb{R}^n_{++}$, define LQP regularizer as*

$$d(v,z) := \begin{cases} \sum\limits_{j=1}^{n} \left[\frac{1}{2}(v_j - z_j)^2 + \mu\left(z_j^2 \log\frac{z_j}{v_j} + v_j z_j - z_j^2\right)\right], & \text{if } v \in \mathbb{R}^n_{++}, \\ +\infty, & \text{otherwise}. \end{cases}$$

Lemma 1 *Let $P := \mathrm{diag}(p_1, p_2, \cdots, p_n) \in \mathbb{R}^{n\times n}$ be a positive definite diagonal matrix, $q : \mathbb{R}^n_+ \to \mathbb{R}^n$ be a monotone mapping and $\theta : \mathbb{R}^n \to \mathbb{R}$. Let μ be a given positive constant. For given $\bar{z}, z \in \mathbb{R}^n_{++}$, we define $\overline{Z} := \mathrm{diag}(\bar{z}_1, \bar{z}_2, \cdots, \bar{z}_n)$, $z^{-1} := (1/z_1, 1/z_2, \cdots, 1/z_n)^\top$ and $\phi'(\bar{z},z) := (z-\bar{z}) + \mu(\bar{z} - \bar{Z}^2 z^{-1})$. Then, the following variational inequality*

$$\theta(v) - \theta(z) + \left\langle v - z, q(z) + P\phi'(\bar{z},z)\right\rangle \ge 0, \quad \forall v \in \mathbb{R}^n_+,$$

has a unique positive solution z. Moreover, for a positive solution $z \in \mathbb{R}^n_{++}$, we have

$$\theta(v) - \theta(z) + \langle v - z, q(z)\rangle \ge (1+\mu)\langle \bar{z} - z, P(v-z)\rangle - \mu\|\bar{z} - z\|_P^2, \quad \forall v \in \mathbb{R}^n_+.$$

The following lemma can be found in [21] that is widely used to characterize the first-order optimality conditions of the subproblems in ADMM-LQP.

Lemma 2 *Let $f : \mathbb{R}^m \longrightarrow \mathbb{R}$ and $h : \mathbb{R}^m \longrightarrow \mathbb{R}$ be two convex functions defined on a closed convex set $\Omega \subset \mathbb{R}^m$ and h is differentiable. Suppose that the solution set $\Omega^* = \arg\min\limits_{x\in\Omega}\{f(x) + h(x)\}$ is nonempty. Then, we have*

$$x^* \in \Omega^* \;\; \textit{if and only if} \;\; x^* \in \Omega,\; f(x) - f(x^*) + \langle x - x^*, \nabla h(x^*)\rangle \ge 0,\; \forall x \in \Omega.$$

Now, a point $\mathbf{w}^* = (\mathbf{x}^*, y^*, \lambda^*) \in \mathcal{W}$ is called the saddle-point of (1) if

$$\mathcal{L}(\mathbf{x}^*, y^*, \lambda) \le \mathcal{L}(\mathbf{x}^*, y^*, \lambda^*) \le \mathcal{L}(\mathbf{x}, y, \lambda^*), \quad \forall (\mathbf{x}, y, \lambda) \in \mathcal{W},$$

which is equivalent to

$$\begin{cases} \textbf{For } i = 1, 2, \cdots, p, \\ x_i^* \in \mathbb{R}^{m_i}_+, \;\; f_i(x_i) - f_i(x_i^*) + \left\langle x_i - x_i^*, -A_i^\top \lambda^*\right\rangle \ge 0, \;\; \forall x_i \in \mathbb{R}^{m_i}_+, \\ y^* \in \mathcal{Y}^*, \quad g(y) - g(y^*) + \left\langle y - y^*, -B^\top \lambda^*\right\rangle \ge 0, \qquad \forall y \in \mathcal{Y}, \\ \lambda^* \in \mathbb{R}^n, \quad \left\langle \lambda - \lambda^*, A_1 x_1^* + A_2 x_2^* + \cdots + A_p x_p^* + By^* - b\right\rangle \ge 0, \;\; \forall \lambda \in \mathbb{R}^n. \end{cases}$$

Rewriting these inequalities in a compact Variational Inequality (VI) form, we have

$$\Theta(\mathbf{w}) - \Theta(\mathbf{w}^*) + \langle \mathbf{w} - \mathbf{w}^*, \mathcal{J}(\mathbf{w}^*) \rangle \geq 0, \ \forall \mathbf{w} \in \mathcal{W}, \tag{8}$$

where

$$\Theta(\mathbf{w}) = \sum_{i=1}^{p} f_i(x_i) + g(y), \ \mathbf{w} = \begin{pmatrix} \mathbf{x} \\ y \\ \lambda \end{pmatrix}, \ \mathcal{J}(\mathbf{w}) = \begin{pmatrix} -A^{\mathsf{T}}\lambda \\ -B^{\mathsf{T}}\lambda \\ A\mathbf{x} + By - b \end{pmatrix}. \tag{9}$$

Remark 1 *Let $\mathcal{W}^*$ be the solution set of (8). Then, ADMM-LQP can be also used to solve the following structured variational inequality problem:*

$$\textit{find } \mathbf{w}^* \in \mathcal{W}^* \textit{ such that } \left\langle \mathbf{w} - \mathbf{w}^*, \hat{\mathcal{J}}(\mathbf{w}^*) \right\rangle \geq 0, \ \forall \mathbf{w} \in \mathcal{W}, \tag{10}$$

from which

$$\hat{\mathcal{J}}(\mathbf{w}) = \begin{pmatrix} \theta_1(\mathbf{x}) - A^{\mathsf{T}}\lambda \\ \theta_2(y) - B^{\mathsf{T}}\lambda \\ A\mathbf{x} + By - b \end{pmatrix}, \ \textit{and } \theta_1(\mathbf{x}) \in \partial \sum_{i=1}^{p} f_i(x_i), \ \theta_2(y) \in \partial g(y). \tag{11}$$

Here, ∂g denotes its limiting-subdifferential. If f_i and g are differentiable, then solving Problem (1) is equivalent to solving (10) with $\theta_1(\mathbf{x}) = \sum_{i=1}^{p} \nabla f_i(x_i), \theta_2(y) = \nabla g(y)$. Besides, the inequality (8) can be also rewritten as

$$\mathit{VI}(\Theta, \mathcal{J}, \mathcal{W}): \ \Theta(\mathbf{w}) - \Theta(\mathbf{w}^*) + \langle \mathbf{w} - \mathbf{w}^*, \mathcal{J}(\mathbf{w}) \rangle \geq 0, \ \forall \mathbf{w} \in \mathcal{W} \tag{12}$$

because the affine mapping $\mathcal{J}$ is skew-symmetric and satisfies

$$\langle \mathbf{w} - \bar{\mathbf{w}}, \mathcal{J}(\mathbf{w}) - \mathcal{J}(\bar{\mathbf{w}}) \rangle \equiv 0 \ \ \forall \mathbf{w}, \bar{\mathbf{w}} \in \mathcal{W}. \tag{13}$$

Due to the assumption that the solution set of (1) is nonempty, the solution set $\mathcal{W}^*$ is also nonempty and convex. Moreover, it can be expressed as (see [16]):

$$\mathcal{W}^* = \bigcap_{\mathbf{w} \in \mathcal{W}} \{ \widehat{\mathbf{w}} \in \mathcal{W} \mid \Theta(\mathbf{w}) - \Theta(\widehat{\mathbf{w}}) + \langle \mathbf{w} - \widehat{\mathbf{w}}, \mathcal{J}(\mathbf{w}) \rangle \geq 0 \}.$$

Note that if the iterates generated by ADMM-LQP satisfy VI$(\Theta, \mathcal{J}, \mathcal{W})$ with an extra term converging to a fixed value as k goes to infinity, then global convergence of our ADMM-LQP could be demonstrated theoretically. To verify this conjecture, we introduce the following auxiliary notations to simplify analysis:

$$\tilde{\mathbf{x}}^k = \begin{pmatrix} \tilde{x}_1^k \\ \tilde{x}_2^k \\ \vdots \\ \tilde{x}_p^k \end{pmatrix} = \begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ \vdots \\ x_p^{k+1} \end{pmatrix}, \ \tilde{y}^k = y^{k+1}, \ \tilde{\mathbf{w}}^k = \begin{pmatrix} \tilde{\mathbf{x}}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} \mathbf{x}^{k+1} \\ y^{k+1} \\ \tilde{\lambda}^k \end{pmatrix}, \tag{14}$$

where

$$\tilde{\lambda}^k = \lambda^k - \beta(A\mathbf{x}^{k+1} + By^k - b). \tag{15}$$

Lemma 3 *(Prediction step) The iterates generated by ADMM-LQP satisfy*

$$\Theta(\mathbf{w}) - \Theta(\tilde{\mathbf{w}}^k) + \left\langle \mathbf{w} - \tilde{\mathbf{w}}^k, \mathcal{J}(\tilde{\mathbf{w}}^k) + Q(\tilde{\mathbf{w}}^k - \mathbf{w}^k) \right\rangle + \left\| \mathbf{w}^k - \tilde{\mathbf{w}}^k \right\|_N^2 \geq 0 \tag{16}$$

for any $\mathbf{w} \in \mathcal{W}$ *where*

$$N = \begin{bmatrix} N_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

with $N_1 = \mu \operatorname{diag}(r_1\mathbf{I}, r_2\mathbf{I}, \cdots, r_p\mathbf{I})$*, and*

$$Q = \begin{bmatrix} Q_1 & \mathbf{0} \\ \mathbf{0} & Q_2 \end{bmatrix} \tag{17}$$

with

$$Q_1 = \begin{bmatrix} (1+\mu)r_1\mathbf{I} & -\beta A_1^\mathsf{T} A_2 & \cdots & -\beta A_1^\mathsf{T} A_p \\ -\beta A_2^\mathsf{T} A_1 & (1+\mu)r_2\mathbf{I} & \cdots & -\beta A_2^\mathsf{T} A_p \\ \vdots & \vdots & \ddots & \vdots \\ -\beta A_p^\mathsf{T} A_1 & -\beta A_p^\mathsf{T} A_2 & \cdots & (1+\mu)r_p\mathbf{I} \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \beta B^\mathsf{T} B & -\alpha B^\mathsf{T} \\ -B & \frac{1}{\beta}\mathbf{I} \end{bmatrix}. \tag{18}$$

Proof. Using Lemma 2 and previous notations, the first-order optimality condition of x_i-subproblem is

$$\begin{aligned} &\tilde{x}_i^k \in \mathbb{R}_+^{m_i}, \quad f_i(x_i) - f_i(\tilde{x}_i^k) + \langle x_i - \tilde{x}_i^k, -A_i^\mathsf{T}\lambda^k + \\ &\beta A_i^\mathsf{T}\left(A_i\tilde{x}_i^k + \sum_{r=1, r\neq i}^{p} A_r x_r^k + By^k - b\right) + r_i \nabla_{x_i^k} d(x_i^k, \tilde{x}_i^k)\Big\rangle \geq 0 \end{aligned} \tag{19}$$

for any $x_i \in \mathbb{R}_+^{m_i}$. Applying Lemma 1 with $\theta(v) = f_i(x_i), v = x_i, z = \tilde{x}_i^k, \bar{z} = x_i^k, P = r_i\mathbf{I}, q(z) = -A_i^\mathsf{T}\lambda^k + \beta A_i^\mathsf{T}\left(A_i\tilde{x}_i^k + \sum_{r=1,r\neq i}^{p} A_r x_r^k + By^k - b\right)$ to (19), we obtain

$$\begin{aligned} &f_i(x_i) - f_i(\tilde{x}_i^k) + \left\langle x_i - \tilde{x}_i^k, -A_i^\mathsf{T}\tilde{\lambda}^k + \beta A_i^\mathsf{T} \sum_{r=1, r\neq i}^{p} A_r(x_r^k - \tilde{x}_r^k) \right\rangle \\ &\geq (1+\mu)r_i \langle x_i - \tilde{x}_i^k, x_i^k - \tilde{x}_i^k\rangle - \mu r_i \left\|x_i^k - \tilde{x}_i^k\right\|^2. \end{aligned} \tag{20}$$

Meanwhile, the first-order optimality condition of y-subproblems is

$$\tilde{y}^k \in \mathcal{Y}, \quad g(y) - g(\tilde{y}^k) + \left\langle y - \tilde{y}^k, -B^\mathsf{T}\lambda^{k+\frac{1}{2}} + \beta B^\mathsf{T}\left(A\tilde{\mathbf{x}}^k + B\tilde{y}^k - b\right)\right\rangle \geq 0 \tag{21}$$

for any $y \in \mathcal{Y}$. By the way of generating $\lambda^{k+\frac{1}{2}}$ and (15), it holds

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) = \tilde{\lambda}^k + (1-\alpha)(\lambda^k - \tilde{\lambda}^k). \tag{22}$$

Plugging it into (21) and using an equivalent reformulation of (15), i.e.,

$$\beta(A\tilde{\mathbf{x}}^k + B\tilde{y}^k - b) = \lambda^k - \tilde{\lambda}^k + \beta B(\tilde{y}^k - y^k), \tag{23}$$

are to achieve

$$g(y) - g(\tilde{y}^k) + \left\langle y - \tilde{y}^k, -B^\mathsf{T}\tilde{\lambda}^k + \alpha B^\mathsf{T}(\lambda^k - \tilde{\lambda}^k) + \beta B^\mathsf{T} B(\tilde{y}^k - y^k)\right\rangle \geq 0. \tag{24}$$

Notice that, the equation (23) is equivalent to

$$\left\langle \lambda - \tilde{\lambda}^k, A\tilde{\mathbf{x}}^k + B\tilde{y}^k - b + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) - B(\tilde{y}^k - y^k)\right\rangle \geq 0 \tag{25}$$

for any $\lambda \in \mathbb{R}^n$. Combining these inequalities (20), (24)-(25) and the definitions of matrices N and Q, the proof is completed. ■

Lemma 3 shows that the iterates generated by ADMM-LQP can be characterized as a variational inequality with the aid of an immediate variable $\tilde{\mathbf{w}}$. We call $\tilde{\mathbf{w}}$ the predicting variable and $\mathbf{w}^{k+1}$ the correcting variable. Moreover, they satisfy the following relationship:

Lemma 4 *(Correction step) The iterates* $\tilde{\mathbf{w}}$ *defined by (14) and* $\mathbf{w}^{k+1}$ *generated by ADMM-LQP satisfy*

$$\mathbf{w}^{k+1} = \mathbf{w}^k - M(\mathbf{w}^k - \tilde{\mathbf{w}}^k), \tag{26}$$

where

$$M = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\tau\beta B & (\alpha+\tau)\mathbf{I} \end{bmatrix}.$$

Proof. By the updates of λ^{k+1} and $\lambda^{k+\frac{1}{2}}$, it can be deduced that

$$\begin{aligned}
\lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \tau\beta(A\mathbf{x}^{k+1} + By^{k+1} - b) \\
&= \lambda^{k+\frac{1}{2}} - \tau\beta(A\mathbf{x}^{k+1} + By^{k} - b) + \tau\beta B(y^k - y^{k+1}) \\
&= \lambda^{k+\frac{1}{2}} - \tau(\lambda^k - \tilde{\lambda}^k) + \tau\beta B(y^k - y^{k+1}) \\
&= \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) - \tau(\lambda^k - \tilde{\lambda}^k) + \tau\beta B(y^k - y^{k+1}) \\
&= \lambda^k - [(\alpha+\tau)(\lambda^k - \tilde{\lambda}^k) - \tau\beta B(y^k - \tilde{y}^k)].
\end{aligned}$$

The above equality together with (14) immediately implies (26). ■

3 Convergence analysis of ADMM-LQP

3.1 Basic properties of $\{\mathbf{w}^k - \mathbf{w}^*\}$

In the following, we begin to analyze some properties of the sequences $\{\mathbf{w}^k - \mathbf{w}^*\}$ under a special H-weighted norm. Followed by (13), (16) can be rewritten as

$$\Theta(\mathbf{w}) - \Theta(\tilde{\mathbf{w}}^k) + \langle \mathbf{w} - \tilde{\mathbf{w}}^k, \mathcal{J}(\mathbf{w}) \rangle \geq \left\langle \mathbf{w} - \tilde{\mathbf{w}}^k, Q(\mathbf{w}^k - \tilde{\mathbf{w}}^k) \right\rangle - \left\| \mathbf{w}^k - \tilde{\mathbf{w}}^k \right\|_N^2. \tag{27}$$

For any $\alpha + \tau > 0$, the matrix M involved in Lemma 4 is nonsingular. So, by (26) and a direct calculation, the first term on the right-hand side of (27) becomes

$$\left\langle \mathbf{w} - \tilde{\mathbf{w}}^k, Q(\mathbf{w}^k - \tilde{\mathbf{w}}^k) \right\rangle = \left\langle \mathbf{w} - \tilde{\mathbf{w}}^k, H(\mathbf{w}^k - \mathbf{w}^{k+1}) \right\rangle \tag{28}$$

with

$$H = QM^{-1} = \begin{bmatrix} Q_1 & \mathbf{0} \\ \mathbf{0} & H_2 \end{bmatrix}, \tag{29}$$

where Q_1 is given by (18) and

$$H_2 = \begin{bmatrix} (1 - \frac{\alpha\tau}{\alpha+\tau})\beta B^\mathsf{T} B & -\frac{\alpha}{\alpha+\tau} B^\mathsf{T} \\ -\frac{\alpha}{\alpha+\tau} B & \frac{1}{(\alpha+\tau)\beta}\mathbf{I} \end{bmatrix}.$$

Next, we provide a sufficient condition to ensure the positive definiteness of H.

Lemma 5 *For any* $0 < \mu < 1$ *and* $\gamma > \frac{p-1}{1+\mu}$*, the matrix* H *defined in (29) is symmetric positive definite if*

$$r_i \geq \gamma\beta \left\| A_i^\mathsf{T} A_i \right\|, i = 1, 2, \cdots, p, \quad \alpha < 1 \text{ and } \alpha + \tau > 0 \tag{30}$$

Proof. It follows from the known conditions that

$$H \succeq \begin{bmatrix} \tilde{Q}_1 & \mathbf{0} \\ \mathbf{0} & H_2 \end{bmatrix} := \tilde{H},$$

where

$$\tilde{Q}_1 = \beta \begin{bmatrix} (1+\mu)\gamma A_1^\mathsf{T} A_1 & -A_1^\mathsf{T} A_2 & \cdots & -A_1^\mathsf{T} A_p \\ -A_2^\mathsf{T} A_1 & (1+\mu)\gamma A_2^\mathsf{T} A_2 & \cdots & -A_2^\mathsf{T} A_p \\ \vdots & \vdots & \ddots & \vdots \\ -A_p^\mathsf{T} A_1 & -A_p^\mathsf{T} A_2 & \cdots & (1+\mu)\gamma A_p^\mathsf{T} A_p \end{bmatrix}.$$

So, H is positive definite if $\tilde{H}$ is positive definite. By the block structure of $\tilde{H}$, we only need to show that both $\tilde{Q}_1$ and H_2 are positive definite. Note that $\tilde{Q}_1$ can be decomposed as

$$\tilde{Q}_1 = \beta \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_p \end{bmatrix}^\mathsf{T} \tilde{Q}_{1,0} \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_p \end{bmatrix} \tag{31}$$

where

$$\tilde{Q}_{1,0} = \begin{bmatrix} \gamma(1+\mu)\mathbf{I} & -\mathbf{I} & \cdots & -\mathbf{I} \\ -\mathbf{I} & \gamma(1+\mu)\mathbf{I} & \cdots & -\mathbf{I} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{I} & -\mathbf{I} & \cdots & \gamma(1+\mu)\mathbf{I} \end{bmatrix}.$$

For any $\gamma > \frac{p-1}{1+\mu}$, then $\tilde{Q}_{1,0}$ is strictly diagonally dominant and thus positive definite. So, by (31) $\tilde{Q}_1$ is positive definite under the conditions that $\gamma > \frac{p-1}{1+\mu}$ and all $A_i (i = 1, \ldots, p)$ have full column ranks. In addition, we have

$$H_2 = \begin{bmatrix} \beta^{\frac{1}{2}} B & \\ & \beta^{-\frac{1}{2}} \mathbf{I} \end{bmatrix}^\mathsf{T} H_{2,0} \begin{bmatrix} \beta^{\frac{1}{2}} B & \\ & \beta^{-\frac{1}{2}} \mathbf{I} \end{bmatrix}$$

where

$$H_{2,0} = \frac{1}{\alpha+\tau} \begin{bmatrix} (\alpha+\tau-\alpha\tau)\mathbf{I} & -\alpha\mathbf{I} \\ -\alpha\mathbf{I} & \mathbf{I} \end{bmatrix}.$$

$H_{2,0}$ is positive definite if (α, τ) satisfy (30). So, H_2 is positive definite if the last two conditions of (30) hold and the matrix B has full column rank.

Summarizing the above discussions, if the conditions in (30) hold, then the matrix H is positive definite. This completes the whole proof. ■

Theorem 1 *The iterates generated by ADMM-LQP satisfy*

$$\begin{aligned} &\Theta(\mathbf{w}) - \Theta(\tilde{\mathbf{w}}^k) + \langle \mathbf{w} - \tilde{\mathbf{w}}^k, \mathcal{J}(\mathbf{w}) \rangle \\ &\geq \frac{1}{2}\left\{ \left\|\mathbf{w} - \mathbf{w}^{k+1}\right\|_H^2 - \left\|\mathbf{w} - \mathbf{w}^k\right\|_H^2 \right\} + \frac{1}{2}\left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_G^2, \quad \forall \mathbf{w} \in \mathcal{W}, \end{aligned} \tag{32}$$

where

$$G = Q^\mathsf{T} + Q - M^\mathsf{T} H M - 2N. \tag{33}$$

Proof. Taking $(a, b, c, d) = (\mathbf{w}, \tilde{\mathbf{w}}^k, \mathbf{w}^k, \mathbf{w}^{k+1})$ in the following identity

$$(a-b)^\mathsf{T} H(c-d) = \frac{1}{2}\left\{\|a-d\|_H^2 - \|a-c\|_H^2\right\} + \frac{1}{2}\left\{\|c-b\|_H^2 - \|d-b\|_H^2\right\}$$

gives

$$\begin{aligned} (\mathbf{w} - \tilde{\mathbf{w}}^k)^\mathsf{T} H(\mathbf{w}^k - \mathbf{w}^{k+1}) = &\frac{1}{2}\left\{\left\|\mathbf{w} - \mathbf{w}^{k+1}\right\|_H^2 - \left\|\mathbf{w} - \mathbf{w}^k\right\|_H^2\right\} \\ &+ \frac{1}{2}\left\{\left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_H^2 - \left\|\mathbf{w}^{k+1} - \tilde{\mathbf{w}}^k\right\|_H^2\right\}. \end{aligned} \tag{34}$$

By a simple deduction, we have

$$
\begin{aligned}
&\frac{1}{2}\left\{\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_H^2-\left\|\mathbf{w}^{k+1}-\tilde{\mathbf{w}}^k\right\|_H^2\right\}-\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_N^2\\
&=\frac{1}{2}\left\{\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_H^2-\left\|\mathbf{w}^{k+1}-\mathbf{w}^k+\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_H^2\right\}-\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_N^2\\
&=\frac{1}{2}\left\{\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_H^2-\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k-M(\mathbf{w}^k-\tilde{\mathbf{w}}^k)\right\|_H^2\right\}-\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_N^2 \qquad (35)\\
&=\frac{1}{2}\left\langle\mathbf{w}^k-\tilde{\mathbf{w}}^k,\left(HM+(HM)^\mathsf{T}-M^\mathsf{T}HM-2N\right)(\mathbf{w}^k-\tilde{\mathbf{w}}^k)\right\rangle\\
&=\frac{1}{2}\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_{Q^\mathsf{T}+Q-M^\mathsf{T}HM-2N}^2,
\end{aligned}
$$

where the second and the last equalities follow from (26) and (29), respectively. Finally, (32) follows from (27)-(28), (34)-(35) and the definition of G in (33). ■

In fact, setting $\mathbf{w}=\mathbf{w}^*\in\mathcal{W}^*$ in (32) we could have

$$
\begin{aligned}
&\frac{1}{2}\left\{\left\|\mathbf{w}^*-\mathbf{w}^k\right\|_H^2-\left\|\mathbf{w}^*-\mathbf{w}^{k+1}\right\|_H^2-\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_G^2\right\}\\
&\geq\Theta(\tilde{\mathbf{w}}^k)-\Theta(\mathbf{w}^*)+\left\langle\tilde{\mathbf{w}}^k-\mathbf{w}^*,\mathcal{J}(\mathbf{w}^*)\right\rangle\geq 0,
\end{aligned}
$$

from which the following theorem holds immediately.

Theorem 2 *The iterates generated by ADMM-LQP satisfy*

$$\left\|\mathbf{w}^{k+1}-\mathbf{w}^*\right\|_H^2\leq\left\|\mathbf{w}^k-\mathbf{w}^*\right\|_H^2-\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_G^2,\quad\forall\mathbf{w}^*\in\mathcal{W}^*. \qquad (36)$$

Observing that if the matrix G in Theorem 2 is positive definite, then similar convergence results to [4, 15, 16, 19] hold for our ADMM-LQP. However, the matrix G is not necessarily positive definite for any (α,τ) satisfying (7). Therefore, it is full of necessity to estimate the lower bound of $\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_G^2$ for further analysis.

3.2 Lower bound of $\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_G^2$

In this subsection, we focus on estimating the lower bound of $\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_G^2$ by combining the structure of G as well as the first-order optimality condition of y-subproblem. By a direct calculation we have

$$G=\begin{bmatrix}G_1 & \mathbf{0}\\ \mathbf{0} & G_2\end{bmatrix}, \qquad (37)$$

where

$$G_1=\begin{bmatrix}(1-\mu)r_1\mathbf{I} & -\beta A_1^\mathsf{T}A_2 & \cdots & -\beta A_1^\mathsf{T}A_p\\ -\beta A_2^\mathsf{T}A_1 & (1-\mu)r_2\mathbf{I} & \cdots & -\beta A_2^\mathsf{T}A_p\\ \vdots & \vdots & \ddots & \vdots\\ -\beta A_p^\mathsf{T}A_1 & -\beta A_p^\mathsf{T}A_2 & \cdots & (1-\mu)r_p\mathbf{I}\end{bmatrix}, G_2=\begin{bmatrix}(1-\tau)\beta B^\mathsf{T}B & (\tau-1)B^\mathsf{T}\\ (\tau-1)B & \frac{2-\alpha-\tau}{\beta}\mathbf{I}\end{bmatrix}.$$

In the following, we define a set of auxiliary variables

$$\mathbf{E}^{k+1}=A\mathbf{x}^{k+1}+By^{k+1}-b,\quad\mathbf{E}_{\mathbf{x}}^{k+1}=\mathbf{x}^{k+1}-\mathbf{x}^k,\quad\text{and}\quad\mathbf{E}_y^{k+1}=y^{k+1}-y^k. \qquad (38)$$

Lemma 6 *For any $\gamma>\frac{p-1}{1-\mu}$, if $r_i\geq\gamma\beta\left\|A_i^\mathsf{T}A_i\right\|$ for $i=1,2,\cdots,p$, then there exists a constant $\xi_1>0$ such that*

$$
\begin{aligned}
\left\|\mathbf{w}^k-\tilde{\mathbf{w}}^k\right\|_G^2\geq&\xi_1\sum_{i=1}^p\left\|A_i\mathbf{E}_{x_i}^{k+1}\right\|^2+\beta(1-\alpha)\left\|B\mathbf{E}_y^{k+1}\right\|^2\\
&+\beta(2-\alpha-\tau)\left\|\mathbf{E}^{k+1}\right\|^2-2\beta(1-\alpha)\left\langle\mathbf{E}^{k+1},B\mathbf{E}_y^{k+1}\right\rangle.
\end{aligned} \qquad (39)
$$

Proof. Since $r_i \geq \gamma\beta \left\|A_i^\mathsf{T} A_i\right\|$, we deduce

$$G_1 = \begin{bmatrix} (1-\mu)r_1\mathbf{I} & -\beta A_1^\mathsf{T} A_2 & \cdots & -\beta A_1^\mathsf{T} A_p \\ -\beta A_2^\mathsf{T} A_1 & (1-\mu)r_2\mathbf{I} & \cdots & -\beta A_2^\mathsf{T} A_p \\ \vdots & \vdots & \ddots & \vdots \\ -\beta A_p^\mathsf{T} A_1 & -\beta A_p^\mathsf{T} A_2 & \cdots & (1-\mu)r_p\mathbf{I} \end{bmatrix}$$

$$\succeq \beta \begin{bmatrix} (1-\mu)\gamma A_1^\mathsf{T} A_1 & -A_1^\mathsf{T} A_2 & \cdots & -A_1^\mathsf{T} A_p \\ -A_2^\mathsf{T} A_1 & (1-\mu)\gamma A_2^\mathsf{T} A_2 & \cdots & -A_2^\mathsf{T} A_p \\ \vdots & \vdots & \ddots & \vdots \\ -A_p^\mathsf{T} A_1 & -A_p^\mathsf{T} A_2 & \cdots & (1-\mu)\gamma A_p^\mathsf{T} A_p \end{bmatrix} = \tilde{G}_1.$$

So, it is clear that

$$G \succeq \begin{bmatrix} \tilde{G}_1 & \mathbf{0} \\ \mathbf{0} & G_2 \end{bmatrix} := \tilde{G} \quad \Longrightarrow \quad \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_G^2 \geq \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_{\tilde{G}}^2.$$

Then, by the structure of $\tilde{G}$ the following equality holds:

$$\begin{aligned} \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_{\tilde{G}}^2 =& \beta \left\| \begin{pmatrix} A_1(x_1^k - x_1^{k+1}) \\ A_2(x_2^k - x_2^{k+1}) \\ \vdots \\ A_p(x_p^k - x_p^{k+1}) \end{pmatrix} \right\|_{\bar{G}_{1,0}}^2 + \beta(1-\tau)\left\|B\mathbf{E}_y^{k+1}\right\|^2 \\ &+ 2(1-\tau)\left\langle \lambda^k - \tilde{\lambda}^k, B\mathbf{E}_y^{k+1} \right\rangle + \frac{2-\alpha-\tau}{\beta}\left\|\lambda^k - \tilde{\lambda}^k\right\|^2, \end{aligned}$$

where

$$\bar{G}_{1,0} = \begin{bmatrix} \gamma(1-\mu)\mathbf{I} & -\mathbf{I} & \cdots & -\mathbf{I} \\ -\mathbf{I} & \gamma(1-\mu)\mathbf{I} & \cdots & -\mathbf{I} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{I} & -\mathbf{I} & \cdots & \gamma(1-\mu)\mathbf{I} \end{bmatrix}.$$

Since $0 < \mu < 1$, $\gamma > \frac{p-1}{1-\mu}$ and $\bar{G}_{1,0}$ is positive definite, there exists a $\xi_1 > 0$ such that

$$\beta \left\| \begin{pmatrix} A_1(x_1^k - x_1^{k+1}) \\ A_2(x_2^k - x_2^{k+1}) \\ \vdots \\ A_p(x_p^k - x_p^{k+1}) \end{pmatrix} \right\|_{\bar{G}_{1,0}}^2 \geq \xi_1 \sum_{i=1}^p \left\|A_i \mathbf{E}_{x_i}^{k+1}\right\|^2.$$

The definition of $\tilde{\lambda}^k$ in (15) shows $\lambda^k - \tilde{\lambda}^k = \beta\left[\mathbf{E}^{k+1} - B\mathbf{E}_y^{k+1}\right]$. So, by (38), we see that

$$2(1-\tau)\left\langle \lambda^k - \tilde{\lambda}^k, B\mathbf{E}_y^{k+1} \right\rangle = 2\beta(1-\tau)\left\langle \mathbf{E}^{k+1}, B\mathbf{E}_y^{k+1} \right\rangle + 2\beta(\tau-1)\left\|B\mathbf{E}_y^{k+1}\right\|^2,$$

and

$$\begin{aligned} &\frac{2-\alpha-\tau}{\beta}\left\|\lambda^k - \tilde{\lambda}^k\right\|^2 = \beta(2-\alpha-\tau)\left\|\mathbf{E}^{k+1}\right\|^2 \\ &+ \beta(2-\alpha-\tau)\left\|B\mathbf{E}_y^{k+1}\right\|^2 - 2\beta(2-\alpha-\tau)\left\langle \mathbf{E}^{k+1}, B\mathbf{E}_y^{k+1} \right\rangle. \end{aligned}$$

The above discussions illustrate that

$$\begin{aligned} \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_{\tilde{G}}^2 \geq& \xi_1 \sum_{i=1}^p \left\|A_i \mathbf{E}_{x_i}^{k+1}\right\|^2 + \beta(1-\alpha)\left\|B\mathbf{E}_y^{k+1}\right\|^2 \\ &+ \beta(2-\alpha-\tau)\|\mathbf{E}^{k+1}\|^2 - 2\beta(1-\alpha)\left\langle \mathbf{E}^{k+1}, B\mathbf{E}_y^{k+1} \right\rangle. \end{aligned}$$

This completes the proof. ■

Lemma 7 *Suppose $\alpha > -1$. Then the iterates generated by ADMM-LQP satisfy*

$$\left\langle \mathbf{E}^{k+1}, -B\mathbf{E}_y^{k+1} \right\rangle \geq \frac{\tau - 1}{1+\alpha} \left\langle \mathbf{E}^k, B\mathbf{E}_y^{k+1} \right\rangle - \frac{\alpha}{1+\alpha} \left\| B\mathbf{E}_y^{k+1} \right\|^2. \tag{40}$$

Proof. The optimality condition of y-subproblem is

$$y^{k+1} \in \mathcal{Y}, \quad g(y) - g(y^{k+1}) + \left\langle y - y^{k+1}, -B^\top \lambda^{k+\frac{1}{2}} + \beta B^\top \mathbf{E}^{k+1} \right\rangle \geq 0$$

for any $y \in \mathcal{Y}$. Setting $y = y^k$ in the above inequality gives

$$g(y^k) - g(y^{k+1}) - \left\langle \mathbf{E}_y^{k+1}, -B^\top \lambda^{k+\frac{1}{2}} + \beta B^\top \mathbf{E}^{k+1} \right\rangle \geq 0. \tag{41}$$

Similarly, the optimality condition of y-subproblem at the previous iteration is

$$g(y) - g(y^k) + \left\langle y - y^k, -B^\top \lambda^{k-\frac{1}{2}} + \beta B^\top \mathbf{E}^k \right\rangle \geq 0.$$

Letting $y = y^{k+1}$ in the above inequality, we have

$$g(y^{k+1}) - g(y^k) + \left\langle \mathbf{E}_y^{k+1}, -B^\top \lambda^{k-\frac{1}{2}} + \beta B^\top \mathbf{E}^k \right\rangle \geq 0. \tag{42}$$

Then, adding (41) and (42) together is to obtain

$$-\left\langle B\mathbf{E}_y^{k+1}, \lambda^{k-\frac{1}{2}} - \lambda^{k+\frac{1}{2}} + \beta(\mathbf{E}^{k+1} - \mathbf{E}^k) \right\rangle \geq 0. \tag{43}$$

Notice by the updates of $\lambda^{k+\frac{1}{2}}$ and λ^k, i.e., $\lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(A\mathbf{x}^{k+1} + By^k - b)$ and $\lambda^k = \lambda^{k-\frac{1}{2}} - \tau\beta\mathbf{E}^k$, that

$$\lambda^{k-\frac{1}{2}} - \lambda^{k+\frac{1}{2}} = \tau\beta\mathbf{E}^k + \alpha\beta\mathbf{E}^{k+1} - \alpha\beta B\mathbf{E}_y^{k+1}. \tag{44}$$

Substituting (44) into the left-hand side of (43), we get

$$\left\langle -B\mathbf{E}_y^{k+1}, (\alpha+1)\mathbf{E}^{k+1} + (\tau-1)\mathbf{E}^k - \alpha B\mathbf{E}_y^{k+1} \right\rangle \geq 0$$

which is equivalent to (40). ■

The following result provides a concrete form of the lower bound of $\left\| \mathbf{w}^k - \tilde{\mathbf{w}}^k \right\|_G^2$, which is the key technique to analyze convergence properties of the proposed algorithm.

Theorem 3 *(Lower bound estimation) Let the sequences $\{\mathbf{w}^k\}$ be generated by ADMM-LQP and $\{\tilde{\mathbf{w}}^k\}$ be defined by (14). For any*

$$\gamma > \frac{p-1}{1-\mu} \quad \textit{and} \quad (\alpha, \tau) \in \mathcal{K},$$

where $0 < \mu < 1$ and $\mathcal{K}$ is defined by (7), there exist $\xi_1, \xi_2 > 0, \xi_3 \geq 0$ such that

$$\left\| \mathbf{w}^k - \tilde{\mathbf{w}}^k \right\|_G^2 \geq \xi_1 \sum_{i=1}^{p} \left\| A_i \mathbf{E}_{x_i}^{k+1} \right\|^2 + \xi_2 \left\| \mathbf{E}^{k+1} \right\|^2 + \xi_3 \left(\left\| \mathbf{E}^{k+1} \right\|^2 - \left\| \mathbf{E}^k \right\|^2 \right).$$

Proof. First of all, it follows from (39) and (40) that there exists a $\xi_1 > 0$ such that

$$\begin{aligned} \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_G^2 \geq \xi_1 \sum_{i=1}^{p} \left\|A_i \mathbf{E}_{x_i}^{k+1}\right\|^2 + \frac{\beta(1-\alpha)^2}{1+\alpha} \left\|B\mathbf{E}_y^{k+1}\right\|^2 \\ + \beta(2-\alpha-\tau)\left\|\mathbf{E}^{k+1}\right\|^2 + \frac{2\beta(1-\alpha)(\tau-1)}{1+\alpha}\left\langle \mathbf{E}^k, B\mathbf{E}_y^{k+1}\right\rangle . \end{aligned} \tag{45}$$

By the Cauchy-Schwartz inequality, the following holds

$$2(1-\alpha)(\tau-1)\left\langle \mathbf{E}^k, B\mathbf{E}_y^{k+1}\right\rangle \geq -(1-\tau)^2\left\|\mathbf{E}^k\right\|^2 - (1-\alpha)^2\left\|B\mathbf{E}_y^{k+1}\right\|^2 .$$

Substituting it into (45) gives

$$\begin{aligned} &\left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_G^2 \geq \xi_1 \sum_{i=1}^{p} \left\|A_i \mathbf{E}_{x_i}^{k+1}\right\|^2 \\ &+ \beta\left(2-\alpha-\tau-\frac{(1-\tau)^2}{1+\alpha}\right)\left\|\mathbf{E}^{k+1}\right\|^2 + \frac{\beta(1-\tau)^2}{1+\alpha}\left(\left\|\mathbf{E}^{k+1}\right\|^2 - \left\|\mathbf{E}^k\right\|^2\right). \end{aligned}$$

For any $(\alpha, \tau) \in \mathcal{K}$ and $\beta > 0$, it holds that

$$\xi_2 := \beta\left(2-\alpha-\tau-\frac{(1-\tau)^2}{1+\alpha}\right) > 0 \quad \text{and} \quad \xi_3 := \frac{\beta(1-\tau)^2}{1+\alpha} \geq 0,$$

which confirms the conclusion. ■

3.3 Global convergence and sublinear convergence rate

In this part, we focus on analyzing the global convergence of the proposed ADMM-LQP and establishing its sublinear convergence rate. The following corollary is obtained directly from the precious Theorems 1, 2 and 3, respectively.

Corollary 3.1 *Suppose that conditions of Theorem 3 hold. Then, there exist $\xi_1, \xi_2 > 0$ and $\xi_3 \geq 0$ such that the iterates generated by ADMM-LQP satisfy*

$$\begin{aligned} &\left\|\mathbf{w}^{k+1} - \mathbf{w}^*\right\|_H^2 + \xi_3\left\|\mathbf{E}^{k+1}\right\|^2 \\ &\leq \left\|\mathbf{w}^k - \mathbf{w}^*\right\|_H^2 + \xi_3\left\|\mathbf{E}^k\right\|^2 - \xi_1\sum_{i=1}^{p}\left\|A_i\mathbf{E}_{x_i}^{k+1}\right\|^2 - \xi_2\left\|\mathbf{E}^{k+1}\right\|^2 \end{aligned} \tag{46}$$

and

$$\begin{aligned} &\Theta(\mathbf{w}) - \Theta(\tilde{\mathbf{w}}^k) + \left\langle \mathbf{w} - \tilde{\mathbf{w}}^k, \mathcal{J}(\mathbf{w})\right\rangle \\ &\geq \frac{1}{2}\left\{\left\|\mathbf{w} - \mathbf{w}^{k+1}\right\|_H^2 + \xi_3\left\|\mathbf{E}^{k+1}\right\|^2\right\} - \frac{1}{2}\left\{\left\|\mathbf{w} - \mathbf{w}^k\right\|_H^2 + \xi_3\left\|\mathbf{E}^k\right\|^2\right\}. \end{aligned} \tag{47}$$

Theorem 4 *(Global convergence) Suppose the conditions of Theorem 3 hold. Then,*
(i) $\lim_{k\to\infty} \sum_{i=1}^{p} \left\|A_i(x_i^k - x_i^{k+1})\right\|^2 = 0$ *and* $\lim_{k\to\infty}\left\|A\mathbf{x}^{k+1} + By^{k+1} - b\right\|^2 = 0$;
(ii) any accumulation point of $\{\mathbf{w}^k\}$ is a solution of VI$(\Theta, \mathcal{J}, \mathcal{W})$;
(iii) there exists a point $\mathbf{w}^\infty \in \mathcal{W}^*$ such that $\lim_{k\to\infty} \mathbf{w}^k = \mathbf{w}^\infty$.

Proof. The statement (i) can be proved by summing (46) over $k = 0, 1, \ldots, \infty$ directly:

$$\sum_{k=0}^{\infty}\left(\xi_1\sum_{i=1}^{p}\left\|A_i\mathbf{E}_{x_i}^{k+1}\right\|^2 + \xi_2\left\|\mathbf{E}^{k+1}\right\|^2\right) \leq \left\|\mathbf{w}^0 - \mathbf{w}^*\right\|_H^2 + \xi_3\left\|\mathbf{E}^0\right\|^2 < \infty, \tag{48}$$

since $\xi_1, \xi_2 > 0$.

Now, we prove (ii). It follows from (i) , (38), and the full column rank assumption on all the matrices A_i that

$$\lim_{k\to\infty} x_i^k - x_i^{k+1} = \mathbf{0} \quad \text{and} \quad \lim_{k\to\infty} A\mathbf{x}^{k+1} + By^{k+1} - b = \mathbf{0} \tag{49}$$

for all $i = 1, \dots, p$. By the definitions of $\tilde{\mathbf{x}}^k$ and $\tilde{\lambda}^k$ in (14) and (15), we have

$$\lambda^k - \tilde{\lambda}^k = \beta A(\tilde{\mathbf{x}}^k - \mathbf{x}^k) + \beta(A\mathbf{x}^k + By^k - b).$$

Combining the above equation with (49), it holds

$$\lim_{k\to\infty} \lambda^k - \tilde{\lambda}^k = \mathbf{0}.$$

The above discussions together with (23) as well as the full column rank assumption on B show $\lim\limits_{k\to\infty} y^k - \tilde{y}^k = \mathbf{0}$. Hence, we have

$$\lim_{k\to\infty} \mathbf{w}^k - \tilde{\mathbf{w}}^k = \mathbf{0}. \tag{50}$$

Suppose $\hat{\mathbf{w}} = (\hat{\mathbf{x}}, \hat{y}, \hat{\lambda})$ is an accumulation point of $\{\mathbf{w}^k\}$, that is, there is a subsequence $\{\mathbf{w}^{k_j}\}$ converging to $\hat{\mathbf{w}}$ as j goes to infinity. Then, we have

$$\lim_{j\to\infty} \mathbf{w}^{k_j} = \hat{\mathbf{w}} = \lim_{k\to\infty} \tilde{\mathbf{w}}^k.$$

So, for any fixed $\mathbf{w} \in \mathcal{W}$, taking $k = k_j$ in (16) and letting $j \to \infty$, we have

$$\Theta(\mathbf{w}) - \Theta(\hat{\mathbf{w}}) + \langle \mathbf{w} - \hat{\mathbf{w}}, \mathcal{J}(\hat{\mathbf{w}})\rangle \geq 0, \ \forall \mathbf{w} \in \mathcal{W}.$$

Therefore, $\hat{\mathbf{w}} \in \mathcal{W}^*$ is a solution of VI$(\Theta, \mathcal{J}, \mathcal{W})$.

Next, we prove (iii). Followed by (46) and the positive definiteness of H, the sequence $\{\mathbf{w}^k\}$ is uniformly bounded. So, there exists a subsequence $\{\mathbf{w}^{k_j}\}$ converging to a point $\mathbf{w}^\infty = (\mathbf{x}^\infty, y^\infty, \lambda^\infty) \in \mathcal{W}$. Recalling the proof of (ii), we know $\lim\limits_{j\to\infty} \mathbf{w}^{k_j} = \mathbf{w}^\infty$. So $\mathbf{w}^\infty \in \mathcal{W}^*$ is a solution point of VI$(\Theta, \mathcal{J}, \mathcal{W})$. Since (46) holds for any $\mathbf{w}^* \in \mathcal{W}^*$, by (46) again and $\mathbf{w}^\infty \in \mathcal{W}^*$, we have for all $l \geq k_j$ that

$$\left\|\mathbf{w}^l - \mathbf{w}^\infty\right\|_H^2 + \xi_3 \left\|\mathbf{E}^l\right\|^2 \leq \left\|\mathbf{w}^{k_j} - \mathbf{w}^\infty\right\|_H^2 + \xi_3 \left\|\mathbf{E}^{k_j}\right\|^2.$$

Combining this inequality with (i), (50) and the positive definiteness of H is to obtain $\lim\limits_{l\to\infty} \mathbf{w}^l = \mathbf{w}^\infty$. Therefore, the whole sequence $\{\mathbf{w}^k\}$ converges to $\mathbf{w}^\infty$. ■

Theorem 4 illustrates that our proposed algorithm ADMM-LQP is globally convergent. In the following, we will show its sublinear convergence rate for the ergodic iterates that seems to appear originally in [7]:

$$\mathbf{w}_T := \frac{1}{1+T} \sum_{k=\kappa}^{\kappa+T} \tilde{\mathbf{w}}^k \quad \text{for any } \kappa \geq 0. \tag{51}$$

Theorem 5 *(Ergodic convergence rate) Suppose that conditions of Theorem 3 hold. Then, for any integers $\kappa \geq 0, T > 0$ and for all $k \in [\kappa, \kappa+T]$, there exists a constant $\xi_3 \geq 0$ such that*

$$\Theta(\mathbf{w}_T) - \Theta(\mathbf{w}) + \langle \mathbf{w}_T - \mathbf{w}, \mathcal{J}(\mathbf{w})\rangle \leq \frac{1}{2(1+T)} \left\{\|\mathbf{w} - \mathbf{w}^\kappa\|_H^2 + \xi_3 \|A\mathbf{x}^\kappa + By^\kappa - b\|^2\right\}. \tag{52}$$

Proof. Summing the inequality (47) over k between κ and $\kappa+T$ gives

$$\begin{aligned}&\sum_{k=\kappa}^{\kappa+T}\Theta(\tilde{\mathbf{w}}^k)-(1+T)\Theta(\mathbf{w})+\left\langle\sum_{k=\kappa}^{\kappa+T}\tilde{\mathbf{w}}^k-(1+T)\mathbf{w},\mathcal{J}(\mathbf{w})\right\rangle\\ \le&\frac{1}{2}\left\{\|\mathbf{w}-\mathbf{w}^\kappa\|_H^2+\xi_3\|A\mathbf{x}^\kappa+By^\kappa-b\|^2\right\},\quad\forall\mathbf{w}\in\mathcal{W}.\end{aligned}\tag{53}$$

By the convexity of Θ and the definition of $\mathbf{w}_T$, we have

$$\Theta(\mathbf{w}_T)\le\frac{1}{1+T}\sum_{k=\kappa}^{\kappa+T}\Theta(\tilde{\mathbf{w}}^k).$$

Dividing (53) by $(1+T)$ and invoking the above inequality, the conclusion (52) is confirmed. ■

Remark 2 *In the following, we further investigate another variant of the sublinear convergence rate in terms of the objective function value error and the feasibility error, which looks a bit more intuitive. For any $\xi>0$, we define $\Gamma_\xi:=\{\lambda|\ \xi\ge\|\lambda\|\}$ and*

$$\delta_\xi:=\inf_{\mathbf{u}^*\in\mathcal{X}^*\times\mathcal{Y}^*}\sup_{\lambda\in\Gamma_\xi}\left\|\begin{pmatrix}\mathbf{u}^*-\mathbf{u}^\kappa\\ \lambda-\lambda^\kappa\end{pmatrix}\right\|_H^2+\xi_3\|A\mathbf{x}^\kappa+By^\kappa-b\|^2,$$

where $\mathbf{u}^*=(\mathbf{x}^*;y^*)$. By setting $\mathbf{w}:=(\mathbf{u}^*,\lambda)$ into (52) and using the definitions of $\mathbf{w}_T$ and $\mathcal{J}(\mathbf{w})$, we get

$$\begin{aligned}&\Theta(\mathbf{w}_T)-\Theta(\mathbf{w})+\langle\mathbf{w}_T-\mathbf{w},\mathcal{J}(\mathbf{w})\rangle\\ =&\Theta(\mathbf{u}_T)-\Theta(\mathbf{u}^*)-\lambda^\top A(\mathbf{x}_T-\mathbf{x}^*)-\lambda^\top B(y_T-y^*)+(\lambda_T-\lambda)^\top(A\mathbf{x}^*+By^*-b)\\ =&\Theta(\mathbf{u}_T)-\Theta(\mathbf{u}^*)-\lambda^\top(A\mathbf{x}_T+By_T-b)\\ \le&\frac{1}{2(1+T)}\left\{\left\|\begin{pmatrix}\mathbf{u}^*-\mathbf{u}^\kappa\\ \lambda-\lambda^\kappa\end{pmatrix}\right\|_H^2+\xi_3\|A\mathbf{x}^\kappa+By^\kappa-b\|^2\right\}\end{aligned}$$

where the second equality uses the fact $A\mathbf{x}^*+By^*=b$. Then, it follows from the above inequality that

$$\begin{aligned}&\Theta(\mathbf{u}_T)-\Theta(\mathbf{u}^*)+\xi\|A\mathbf{x}_T+By_T-b\|\\ =&\sup_{\lambda\in\Gamma_\xi}\left\{\Theta(\mathbf{u}_T)-\Theta(\mathbf{u}^*)-\lambda^\top(A\mathbf{x}_T+By_T-b)\right\}\\ \le&\frac{1}{2(1+T)}\left\{\inf_{\mathbf{u}^*\in\mathcal{X}^*\times\mathcal{Y}^*}\sup_{\lambda\in\Gamma_\xi}\left\|\begin{pmatrix}\mathbf{u}^*-\mathbf{u}^\kappa\\ \lambda-\lambda^\kappa\end{pmatrix}\right\|_H^2+\xi_3\|A\mathbf{x}^\kappa+By^\kappa-b\|^2\right\}\\ =&\frac{\delta_\xi}{2(1+T)}.\end{aligned}\tag{54}$$

Analogous to the analysis of (54) together with (8), we must have

$$\Theta(\mathbf{u}_T)-\Theta(\mathbf{u}^*)-(\lambda^*)^\top(A\mathbf{x}_T+By_T-b)\ge0$$

showing that $\Theta(\mathbf{u}_T)-\Theta(\mathbf{u}^*)\ge-\|\lambda^*\|\|A\mathbf{x}_T+By_T-b\|$. Taking $\xi=2\|\lambda^*\|+1$ in (54) gives

$$\begin{aligned}&(\|\lambda^*\|+1)\|A\mathbf{x}_T+By_T-b\|\\ \le&\Theta(\mathbf{u}_T)-\Theta(\mathbf{u}^*)+(2\|\lambda^*\|+1)\|A\mathbf{x}_T+By_T-b\|\le\frac{\delta_\xi}{2(1+T)},\end{aligned}$$

that is,

$$\|A\mathbf{x}_T + By_T - b\| \le \frac{\delta_\xi}{2(1+T)(\|\lambda^*\|+1)}.$$

So, we will also get $\Theta(\mathbf{u}_T) - \Theta(\mathbf{u}^*) \ge -\|\lambda^*\| \, \|A\mathbf{x}_T + By_T - b\| \ge -\frac{\delta_\xi}{2(1+T)}$ *which shows*

$$\mid \Theta(\mathbf{u}_T) - \Theta(\mathbf{u}^*) \mid \le \frac{\delta_\xi}{2(1+T)}.$$

Now, we would investigate the sublinear convergence rate of our ADMM-LQP for the primal residuals, that is, the nonergodic convergence rate.

Theorem 6 *(Nonergodic convergence rate) Suppose that conditions of Theorem 3 hold. Then, for any integer* $k > 0$*, there exists an integer* $t \le k$ *such that*

$$\left\|\mathbf{x}^t - \mathbf{x}^{t-1}\right\|^2 \le \frac{\vartheta}{k}, \quad \textit{and} \quad \left\|y^t - y^{t-1}\right\|^2 \le \frac{\vartheta}{k}, \tag{55}$$

where $\vartheta > 0$ *is a constant depending on the problem data and the parameters of ADMM-LQP.*

Proof. Let $k > 0$ be any fixed constant and $t \in [1, k]$ be an integer such that

$$\begin{aligned} &\xi_1 \sum_{i=1}^p \left\|A_i(x_i^{t-1} - x_i^t)\right\|^2 + \xi_2 \left\|\mathbf{E}^t\right\|^2 \\ =& \min\left\{\xi_1 \sum_{i=1}^p \left\|A_i(x_i^{l-1} - x_i^l)\right\|^2 + \xi_2 \left\|\mathbf{E}^l\right\|^2 : l = 1, 2, \cdots, k\right\}. \end{aligned}$$

Then, we deduce by (48) that

$$\xi_1 \sum_{i=1}^p \left\|A_i(x_i^{t-1} - x_i^t)\right\|^2 + \xi_2 \left\|\mathbf{E}^t\right\|^2 \le \frac{\vartheta}{k},$$

and $\vartheta > 0$ is a generic constant only depending on the problem data and the parameters of ADMM-LQP. Hence, the left inequality in (55) is obtained. The right inequality in (55) holds by combining the equality (23) and the result $\|\mathbf{E}^t\|^2 \le \vartheta/k$. This completes the proof. ■

4 Extension of ADMM-LQP

In this part, we consider the following structured convex optimization problem

$$\begin{array}{ll} \min & \sum\limits_{i=1}^p f_i(x_i) + g(y) + h(y) \\ \text{s.t.} & \sum\limits_{i=1}^p A_i x_i + By = b, \\ & x_i \in \mathbb{R}_+^{m_i}, \, y \in \mathbb{R}^d \, , i = 1, \cdots, p, \end{array} \tag{56}$$

where $f_i(i = 1, \cdots, p)$ are the same as before, $h(y)$ is nonsmooth but $g(y)$ is smooth. We assume the gradient of g satisfies the Lipschitz condition, i.e., for any $y, \bar{y} \in \mathbb{R}^d$, there exists a constant $L_g > 0$ such that $\|g(y) - g(\bar{y})\| \le L_g \|y - \bar{y}\|$. This inequality, by a Taylor expansion, implies

$$g(y) \le g(\bar{y}) + \langle \nabla g(\bar{y}), y - \bar{y} \rangle + \frac{L_g}{2} \|y - \bar{y}\|^2. \tag{57}$$

The function $h(y)$ is usually used to promote a data structure different from the structure promoted by $g(y)$ at the solution. Particularly, constraints of the form $y \in \mathcal{Y}$ where $\mathcal{Y}$ is a closed convex set, can be incorporated in the optimization problem by letting $h(y)$ be the indicator function of $\mathcal{Y}$, that is, $h(y) = \infty$ when $y \notin \mathcal{Y}$ and otherwise $h(y) = 0$.

Using the popular linearization technique, the y-subproblem in ADMM-LQP can be updated by the following

$$y^{k+1} \leftarrow \arg\min_{y\in\mathbb{R}^d} \left\{ \begin{array}{l} h(y) + \left\langle y, \nabla g(y^k) - B^\top \lambda^{k+\frac{1}{2}} \right\rangle + \\ \frac{\beta}{2} \left\| A\mathbf{x}^{k+1} + By - b \right\|^2 + \frac{1}{2} \left\| y - y^k \right\|_D^2 \end{array} \right\}, \tag{58}$$

where D is a determined proximal matrix. An obvious advantage of using the proximal term $\frac{1}{2}\left\|y - y^k\right\|_D^2$, where $D = \sigma\mathbf{I} - \beta B^\top B$ and $\sigma \geq \beta\|B^\top B\|$, is that it could transform the y-subproblem into the proximal mapping:

$$\mathbf{prox}_{h,\sigma}(y_c^k) = \arg\min h(y) + \frac{\sigma}{2}\left\|y - y_c^k\right\|^2.$$

Here $y_c^k = y^k - \frac{\nabla g(y^k) - B^\top\lambda^{k+\frac{1}{2}} + \beta B^\top(A\mathbf{x}^{k+1} + By^k - b)}{\sigma}$.

In the following, we briefly analyze the convergence of our extended algorithm named **eADMM-LQP** for solving problem (56). Since the x_i-subproblem updates in the same way as before, we just need to analyze the first-order optimality condition of the subproblem in (58).

By the convexity of $g(y)$ and (57), we have

$$\begin{aligned} g(\tilde{y}^k) &\leq g(y^k) + \left\langle \nabla g(y^k), \tilde{y}^k - y^k \right\rangle + \frac{L_g}{2}\left\|\tilde{y}^k - y^k\right\|^2 \\ &= g(y^k) + \left\langle \nabla g(y^k), y - y^k + \tilde{y}^k - y \right\rangle + \frac{L_g}{2}\left\|\tilde{y}^k - y^k\right\|^2 \\ &\leq g(y) + \left\langle \nabla g(y^k), \tilde{y}^k - y \right\rangle + \frac{L_g}{2}\left\|\tilde{y}^k - y^k\right\|^2. \end{aligned}$$

The first-order optimality condition of y-subproblem is

$$\begin{aligned} &h(y) - h(y^{k+1}) + \Big\langle y - y^{k+1}, \nabla g(y^k) - B^\top\lambda^{k+\frac{1}{2}} \\ &+\beta B^\top(A\mathbf{x}^{k+1} + By^{k+1} - b) + D(y^{k+1} - y^k)\Big\rangle \geq 0, \quad \forall y \in \mathbb{R}^d, \end{aligned}$$

which, by (14)-(15) and (22), can be rewritten as

$$\begin{aligned} &h(y) - h(\tilde{y}^k) + \left\langle y - \tilde{y}^k, \nabla g(y^k) \right\rangle \\ &+ \left\langle y - \tilde{y}^k, -B^\top\tilde{\lambda}^k + \alpha B^\top(\lambda^k - \tilde{\lambda}^k) + (\beta B^\top B + D)(\tilde{y}^k - y^k) \right\rangle \geq 0. \end{aligned}$$

The above inequalities show

$$\begin{aligned} &\left\{h(y) + g(y) - h(\tilde{y}^k) - g(\tilde{y}^k)\right\} + \left\langle y - \tilde{y}^k, -B^\top\tilde{\lambda}^k \right\rangle \\ &\geq \left\langle \tilde{y}^k - y, \alpha B^\top(\lambda^k - \tilde{\lambda}^k) + (\beta B^\top B + D)(\tilde{y}^k - y^k) \right\rangle - \frac{L_g}{2}\left\|\tilde{y}^k - y^k\right\|^2, \end{aligned}$$

which, by combining the previous inequalities (20) and (25), gives

$$\Theta(\mathbf{w}) - \Theta(\tilde{\mathbf{w}}^k) + \left\langle \mathbf{w} - \tilde{\mathbf{w}}^k, \mathcal{J}(\tilde{\mathbf{w}}^k) + Q(\tilde{\mathbf{w}}^k - \mathbf{w}^k) \right\rangle + \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_N^2 \geq 0 \tag{59}$$

for any $\mathbf{w} \in \mathcal{W}$, where the notations $\Theta, \mathbf{w}, \tilde{\mathbf{w}}^k, \mathcal{J}, Q$ are the same as that in Lemma 3, but the right-lower block of Q is

$$Q_2 = \begin{bmatrix} \beta B^\top B + D & -\alpha B^\top \\ -B & \frac{1}{\beta}\mathbf{I} \end{bmatrix}, \tag{60}$$

and

$$N = \begin{bmatrix} N_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{L_g}{2}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Note that if taking $D = \sigma\mathbf{I} - \beta B^\top B$ and $\sigma \geq \beta\|B^\top B\|$, then the Q_2 given by (60) is similar to that in (18). Next, we use this form to simplify convergence analysis of eADMM-LQP.

Theorem 7 *Let $D = \sigma\mathbf{I} - \beta B^\top B$ and $\sigma \geq \beta\|B^\top B\|$. Then the iterates generated by eADMM-LQP satisfy (32) and (36) with Q_2 given by (60).*

Proof. Similar to the proof of Theorem 1, we can show

$$\begin{aligned} \Theta(\mathbf{w}) - \Theta(\tilde{\mathbf{w}}^k) + \left\langle \mathbf{w} - \tilde{\mathbf{w}}^k, \mathcal{J}(\mathbf{w}) \right\rangle \geq \frac{1}{2}\left\{ \left\|\mathbf{w} - \mathbf{w}^{k+1}\right\|_H^2 - \left\|\mathbf{w} - \mathbf{w}^k\right\|_H^2 \right\} \\ + \frac{1}{2}\left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_G^2, \end{aligned} \tag{61}$$

where $H = \mathrm{diag}(Q_1, H_2)$ with

$$H_2 = \begin{bmatrix} (1 - \frac{\alpha\tau}{\alpha+\tau})\beta B^\top B + D & -\frac{\alpha}{\alpha+\tau}B^\top \\ -\frac{\alpha}{\alpha+\tau}B & \frac{1}{(\alpha+\tau)\beta}\mathbf{I} \end{bmatrix},$$

and $G = \mathrm{diag}(G_1, G_2)$ with

$$G_2 = \begin{bmatrix} (1-\tau)\beta B^\top B + D - L_g\mathbf{I} & (\tau - 1)B^\top \\ (\tau-1)B & \frac{2-\alpha-\tau}{\beta}\mathbf{I} \end{bmatrix}.$$

Finally, setting $\mathbf{w} = \mathbf{w}^* \in \mathcal{W}^*$ in (32) ensures the inequality (36). ■

Analogous to the analytical techniques in Section 3.2, we next estimate the lower bound of $\left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_G^2$. In fact, by a similar way to the proof of Lemma 6 together with the previous notations in (38), there exists a constant $\xi_1 > 0$ such that

$$\begin{aligned} \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_G^2 \geq \xi_1 \sum_{i=1}^p \left\|A_i\mathbf{E}_{x_i}^{k+1}\right\|^2 + \beta(1-\alpha)\left\|B\mathbf{E}_y^{k+1}\right\|^2 \\ + \beta(2-\alpha-\tau)\left\|\mathbf{E}^{k+1}\right\|^2 + \left\|\mathbf{E}_y^{k+1}\right\|_{D-L_g\mathbf{I}}^2 + 2\beta(1-\alpha)\left\langle \mathbf{E}^{k+1}, -B\mathbf{E}_y^{k+1}\right\rangle. \end{aligned} \tag{62}$$

To further investigate the lower bound of $\left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_G^2$, we need a lemma as follows.

Lemma 8 *Suppose $\alpha > -1$. Then the iterates generated by eADMM-LQP satisfy*

$$\begin{aligned} \left\langle \mathbf{E}^{k+1}, -B\mathbf{E}_y^{k+1}\right\rangle \geq \frac{\tau-1}{1+\alpha}\left\langle \mathbf{E}^k, B\mathbf{E}_y^{k+1}\right\rangle - \frac{1}{\beta(1+\alpha)}\left\langle \mathbf{E}_y^{k+1}, D\mathbf{E}_y^k\right\rangle \\ + \frac{1}{\beta(1+\alpha)}\left\|\mathbf{E}_y^{k+1}\right\|_D^2 - \frac{\alpha}{1+\alpha}\left\|B\mathbf{E}_y^{k+1}\right\|^2 - \frac{L_g}{2\beta(1+\alpha)}\left(\left\|\mathbf{E}_y^k\right\|^2 + \left\|\mathbf{E}_y^{k+1}\right\|^2\right). \end{aligned} \tag{63}$$

Proof. It follows from the optimality condition of y-subproblem together with the previous notations in (38) that

$$\begin{aligned} & h(y) + g(y) - h(y^{k+1}) - g(y^{k+1}) \\ & + \left\langle y - y^{k+1}, -B^\top\lambda^{k+\frac{1}{2}} + \beta B^\top\mathbf{E}^{k+1} + D\mathbf{E}_y^{k+1} \right\rangle + \frac{L_g}{2}\left\|\mathbf{E}_y^{k+1}\right\|^2 \geq 0. \end{aligned}$$

Letting $y = y^k$ in the above inequality gives

$$\begin{aligned} & h(y^k) + g(y^k) - h(y^{k+1}) - g(y^{k+1}) \\ & - \left\langle \mathbf{E}_y^{k+1}, -B^\top \lambda^{k+\frac{1}{2}} + \beta B^\top \mathbf{E}^{k+1} + D\mathbf{E}_y^{k+1} \right\rangle + \frac{L_g}{2} \left\| \mathbf{E}_y^{k+1} \right\|^2 \geq 0. \end{aligned} \tag{64}$$

Similarly, the optimality condition of y-subproblem at the k-th iteration is

$$\begin{aligned} & h(y) + g(y) - h(y^k) - g(y^k) \\ & + \left\langle y - y^k, -B^\top \lambda^{k-\frac{1}{2}} + \beta B^\top \mathbf{E}^k + D\mathbf{E}_y^k \right\rangle + \frac{L_g}{2} \left\| \mathbf{E}_y^k \right\|^2 \geq 0, \end{aligned}$$

which, by setting $y = y^{k+1}$, shows

$$\begin{aligned} & h(y^{k+1}) + g(y^{k+1}) - h(y^k) - g(y^k) \\ & + \left\langle \mathbf{E}_y^{k+1}, -B^\top \lambda^{k-\frac{1}{2}} + \beta B^\top \mathbf{E}^k + D\mathbf{E}_y^k \right\rangle + \frac{L_g}{2} \left\| \mathbf{E}_y^k \right\|^2 \geq 0. \end{aligned} \tag{65}$$

Combining (64) and (65), we have

$$\begin{aligned} & \left\langle \mathbf{E}_y^{k+1}, B^\top (\lambda^{k+\frac{1}{2}} - \lambda^{k-\frac{1}{2}}) + \beta B^\top (\mathbf{E}^k - \mathbf{E}^{k+1}) + D(\mathbf{E}_y^k - \mathbf{E}_y^{k+1}) \right\rangle \\ & + \frac{L_g}{2} \left(\left\| \mathbf{E}_y^k \right\|^2 + \left\| \mathbf{E}_y^{k+1} \right\|^2 \right) \geq 0. \end{aligned} \tag{66}$$

Then, substituting (44) into the left-hand side of (66), we deduce

$$\begin{aligned} & \left\langle -B\mathbf{E}_y^{k+1}, (\tau - 1)\beta \mathbf{E}^k + (1+\alpha)\beta \mathbf{E}^{k+1} \right\rangle + \alpha\beta \left\| B\mathbf{E}_y^{k+1} \right\|^2 - \left\| \mathbf{E}_y^{k+1} \right\|_D^2 \\ & + \left\langle \mathbf{E}_y^{k+1}, D\mathbf{E}_y^k \right\rangle + \frac{L_g}{2} \left(\left\| \mathbf{E}_y^k \right\|^2 + \left\| \mathbf{E}_y^{k+1} \right\|^2 \right) \geq 0, \end{aligned}$$

which is equivalent to (63). ■

Theorem 8 *(Lower bound estimation) Let $D = \sigma \mathbf{I} - \beta B^\top B$ with $\sigma \geq \beta \|B^\top B\| + \frac{3-\alpha}{1+\alpha} L_g$, and the parameters (γ, μ) satisfy the conditions in Theorem 3. Then, for any $(\alpha, \tau) \in \mathcal{K}$, there exist $\xi_1, \xi_2 > 0$ and $\xi_3 \geq 0$ such that*

$$\left\| \mathbf{w}^k - \tilde{\mathbf{w}}^k \right\|_G^2 \geq \xi_1 \sum_{i=1}^p \left\| A_i \mathbf{E}_{x_i}^{k+1} \right\|^2 + \xi_2 \left\| \mathbf{E}^{k+1} \right\|^2 + \xi_3 \left(\left\| \mathbf{E}^{k+1} \right\|^2 - \left\| \mathbf{E}^k \right\|^2 \right).$$

Proof. Plugging (63) into (62) and using the Cauchy-Schwartz inequality, we have

$$\begin{aligned} & \left\| \mathbf{w}^k - \tilde{\mathbf{w}}^k \right\|_G^2 \\ \geq & \xi_1 \sum_{i=1}^p \left\| A_i \mathbf{E}_{x_i}^{k+1} \right\|^2 + \beta(1-\alpha) \left\| B\mathbf{E}_y^{k+1} \right\|^2 + \beta(2-\alpha-\tau) \left\| \mathbf{E}^{k+1} \right\|^2 + \left\| \mathbf{E}_y^{k+1} \right\|_{D - L_g \mathbf{I}}^2 \\ & - \frac{\beta}{1+\alpha} \left[(1-\tau)^2 \left\| \mathbf{E}^k \right\|^2 + (1-\alpha)^2 \left\| B\mathbf{E}_y^{k+1} \right\|^2 \right] - \frac{1-\alpha}{1+\alpha} \left[\left\| \mathbf{E}_y^k \right\|_D^2 + \left\| \mathbf{E}_y^{k+1} \right\|_D^2 \right] \\ & + \frac{2(1-\alpha)}{1+\alpha} \left\| \mathbf{E}_y^{k+1} \right\|_D^2 - \frac{2\beta\alpha(1-\alpha)}{1+\alpha} \left\| B\mathbf{E}_y^{k+1} \right\|^2 - \frac{(1-\alpha)L_g}{1+\alpha} \left[\left\| \mathbf{E}_y^{k+1} \right\|^2 + \left\| \mathbf{E}_y^k \right\|^2 \right] \\ = & \xi_1 \sum_{i=1}^p \left\| A_i \mathbf{E}_{x_i}^{k+1} \right\|^2 + \beta \left(2 - \alpha - \tau - \frac{(1-\tau)^2}{1+\alpha} \right) \left\| \mathbf{E}^{k+1} \right\|^2 \\ & + \frac{\beta(1-\tau)^2}{1+\alpha} \left[\left\| \mathbf{E}^{k+1} \right\|^2 - \left\| \mathbf{E}^k \right\|^2 \right] + R_{xy}, \end{aligned}$$

where

$$
\begin{aligned}
R_{xy} &= \frac{1-\alpha}{1+\alpha}\left[\left\|\mathbf{E}_y^{k+1}\right\|_D^2 - \left\|\mathbf{E}_y^k\right\|_D^2\right] + \left\|\mathbf{E}_y^{k+1}\right\|_{D-L_g\mathbf{I}}^2 - \frac{(1-\alpha)L_g}{1+\alpha}\left[\left\|\mathbf{E}_y^{k+1}\right\|^2 + \left\|\mathbf{E}_y^k\right\|^2\right] \\
&= \frac{2}{1+\alpha}\left\|\mathbf{E}_y^{k+1}\right\|_{D-L_g\mathbf{I}}^2 - \frac{1-\alpha}{1+\alpha}\left\|\mathbf{E}_y^k\right\|_{D+L_g\mathbf{I}}^2.
\end{aligned}
$$

Since $D = \sigma\mathbf{I} - \beta B^\top B$, then for any $\sigma \geq \beta\|B^\top B\| + \frac{3-\alpha}{1+\alpha}L_g$, we have $D \succeq \frac{3-\alpha}{1+\alpha}L_g\mathbf{I}$ and thus $R_{xy} \geq 0$. So, it holds

$$
\left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_G^2 \geq \xi_1\sum_{i=1}^p\left\|A_i\mathbf{E}_{x_i}^{k+1}\right\|^2 + \xi_2\left\|\mathbf{E}^{k+1}\right\|^2 + \xi_3\left[\left\|\mathbf{E}^{k+1}\right\|^2 - \left\|\mathbf{E}^k\right\|^2\right],
$$

where $\xi_i (i = 1, 2, 3)$ are the same as that in Theorem 3. ■

Finally, based on the above discussions and the similar analysis to Section 3.3, the algorithm eADMM-LQP converges globally with a sublinear convergence rate in the ergodic and nonergodic sense.

5 Concluding remarks

In this article, we have proposed a partial LQP-based ADMM with larger stepsizes of dual variables as in [5] to solve a family of separable convex minimization problems. The so-called LQP regularization term is added to the first grouped subproblems with proper proximal parameters but it is not added to the final subproblem. With the aid of prediction-correction technique, we establish the global convergence of the proposed algorithm and its sublinear convergence rate in terms of the objective residual and the primal residual. The proposed algorithm is also extended to solve a nonsmoooth composite convex optimization, where the second subproblem is updated by using the linearization technique and proximal technique. We analyze the global sublinear convergence properties of the extended method in a brief way.

Notice that our result about the ergodic convergence rate is established for a new average iterate $\mathbf{w}_T := \frac{1}{1+T}\sum_{k=\kappa}^{\kappa+T}\tilde{\mathbf{w}}^k$ for any integer $\kappa \geq 0$, which is more general than the traditional iterate $\mathbf{w}_T := \frac{1}{1+T}\sum_{k=0}^{T}\tilde{\mathbf{w}}^k$ in the literature of ADMM. This makes it possible to use some accelerated techniques in practical programming, see the experiments in [7] for more details. Another observation is that the constrained set $\mathcal{Y}$ is any closed convex set, so our problem and algorithm are more general than two existing researches in [27, 30] and thus have a large number of potential applications. How to extend our algorithm to the nonseparable nonconvex optimization problem and how to solve the subproblem inexactly are very interesting topics. Finally, whether the proposed algorithm could use indefinite proximal term needs our further investigations in the future work.

References

[1] A. Auslender, M. Teboulle, S. Ben-Tiba. A logarithmic-quadratic proximal method for variational inequalities, Comput. Math. Appl. 12 (1999) 31-40.

[2] S. Boyd, N. Parikh, E. Chu, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (2010) 1-122.

[3] S. Boyd, L. Vandenberghe. Convex Optimization, Cambridge University Press, Cambridge, 2004.

[4] J. Bai, H. Zhang, J. Li. A parameterized proximal point algorithm for separable convex optimization, Optim. Lett. 12 (2018) 1589-1608.

[5] J. Bai, J. Li, F. Xu, H. Zhang. Generalized symmetric ADMM for separable convex optimization, Comput. Optim. Appl. 70 (2018) 129-170.

[6] J. Bai, X. Chang, J. Li, F. Xu. Convergence revisit on generalized symmetric ADMM, Optimization, 70 (2021) 149-168.

[7] J. Bai, W. Hager, H. Zhang. An inexact accelerated stochastic ADMM for separable convex optimization, Optimization Online, (2020) pp. 1-32. http://www.optimization-online.org/DB_HTML/2020/08/7990.html.

[8] C. Chen, B. He, Y. Ye, X. Yuan. The direct extension of ADMM for multi-block minimization problems is not necessarily convergent, Math. Program. 155 (2016) 57-79.

[9] J. Chen, Y. Wang, H. He, Y. Lv. Convergence analysis of positive-indefinite proximal ADMM with a Glowinski's relaxation factor, Numer. Algor. 83 (2020) 1415-1440.

[10] F. Facchinei, J. Pang. Finite-Dimensional Variational Inequalities and Complementarity Problems, Vol. I, Springer-Verlag, New York, 2003.

[11] M. Fazel, T. Pong, D. Sun, P. Tseng. Hankel matrix rank minimization with applications to system identification and realization, SIAM J. Matrix Anal. Appl. 34 (2013) 946-977.

[12] D. Gabay, B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. 2 (1976) 17-40.

[13] R. Glowinski, A. Marroco. Approximation parélḿents finis d'rdre un et résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires, Rev. Fr. Autom. Inform. Rech. Opér. Anal. Numér. 2 (1975) 41-76.

[14] M. Hestenes. Multiplier and gradient methods, J. Optim. Theory Appl. 4 (1969) 303-320.

[15] B. He, Y. Xu, X. Yuan. A logarithmic-quadratic proximal prediction-correction method for structured monotone variational inequalities, Comput. Optim. Appl. 35 (2006) 19-46.

[16] B. He, X. Yuan. On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method, SIAM J. Numer. Anal. 50 (2012) 700-709.

[17] B. He, L. Hou, X. Yuan. On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming, SIAM J. Optim. 25 (2015) 2274-2312.

[18] B. He. Modified alternating directions method of multipliers for convex optimization with three separable functions (in Chinese), Oper. Res. Trans. 19 (2015) 57-70.

[19] B. He, H. Xu, X. Yuan. On the proximal Jacobian decomposition of ALM for multiple-block separable convex minimization problems and its relationship to ADMM, J. Sci. Comput. 66 (2016) 1204-1217.

[20] B. He, X. Yuan. Block-wise alternating direction method of multipliers for multiple-block convex programming and beyond, SMAI J. Comput. Math. 1 (2015) 145-174.

[21] B. He, F. Ma, X. Yuan. Convergence study on the symmetric version of admm with larger step sizes, SIAM J. Imaging Sci. 9 (2016) 1467-1501.

[22] V. Hryhorenko, D. Klyushin, S. Lyashko. Multiblock ADMM in Machine Learning, 2019 IEEE Int. Conf. Adv. Trends Inf. Theory, ATIT 2019-Proc. (2019) pp. 461-464.

[23] W. Hager, H. Zhang. Inexact alternating direction multiplier methods for separable convex optimization, Comput. Optim. Appl. 73 (2019) 201-235.

[24] D. Kinderlehrer, G. Stampacchia. An Introduction to Variational Inequalities and Their Applications, Academic Press, New York, 1980.

[25] M. Li. A hybrid LQP-based method for structured variational inequalities, INT J. Comput. Math. 89 (2012) 1412-1425.

[26] M. Li, X. Yuan. A strictly contractive Peaceman-Rachford splitting method with logarithmic-quadratic proximal regularization for convex programming, Math. Oper. Res. 40 (2015) 842-858.

[27] Y. Liu, K. Guo, M. Yang. Convergence study on the logarithmic-quadratic proximal regularization of strictly contractive Peaceman-Rachford splitting method with larger step-size, INT J. Comput. Math. 97 (2020) 1744-1766.

[28] X. Liu, Y. Hu, G. Li, et al. Application Analysis of Optimization Methods, Science Press, Beijing, 2014.

[29] Z. Wu, F. Li, M. Li. A proximal Peaceman-Rachford splitting method for solving the multi-block separable convex minimization problems, INT J. Comput. Math. 96 (2019) 708-728.

[30] Z. Wu, M. Li. An LQP-based symmetric alternating direction method of multipliers with larger step sizes, J. Oper. Res. Soc. China, 7 (2019) 365-383.

[31] Y. Xu. Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming, SIAM J. Optim. 27 (2017) 1459-1484.

[32] Z. Xu. Compressed sensing: a survey (in Chinese), Sci. Sin. Math. 42 (2012) 865-877.