

Inexact Alternating Direction Methods of Multipliers with Logarithmic–Quadratic Proximal Regularization

Min Li · Li-Zhi Liao · Xiaoming Yuan

Received: 10 September 2012 / Accepted: 4 May 2013 / Published online: 25 May 2013
© Springer Science+Business Media New York 2013

Abstract In the literature, it was shown recently that the Douglas–Rachford alternating direction method of multipliers can be combined with the logarithmic-quadratic proximal regularization for solving a class of variational inequalities with separable structures. This paper studies the inexact version of this combination, where the resulting subproblems are allowed to be solved approximately subject to different inexactness criteria. We prove the global convergence and establish worst-case convergence rates for the derived inexact algorithms.

Keywords Alternating direction method of multipliers · Logarithmic-quadratic proximal regularization · Convergence rate · Inexact · Variational inequality

1 Introduction

Variational inequality problems (VIPs) with separable structures and linear constraints capture wide applications in some fields. For solving these VIPs, the Douglas–Rachford alternating direction method of multipliers (ADMM) proposed originally in [1] is a benchmark. The proximal point algorithm (PPA) in [2, 3] was suggested to regularize ADMM’s subproblems and accordingly an algorithmic framework of the inexact version of ADMM with proximal regularization was established.

Communicated by Masao Fukushima.

M. Li

School of Economics and Management, Southeast University, Nanjing, 210096, China
e-mail: liminju@yahoo.com

L.-Z. Liao · X. Yuan (✉)

Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
e-mail: xmyuan@hkbu.edu.hk

L.-Z. Liao

e-mail: liliao@math.hkbu.edu.hk

Moreover, in [4, 5] the ADMM subproblems were suggested to be regularized by the logarithmic-quadratic proximal (LQP) regularization proposed in [6]; thus the complementarity subproblems of applying ADMM reduce to systems of nonlinear equations. However, it is not possible to acquire the exact solutions of the complementarity subproblems of ADMM with LQP regularization. Therefore, we are interested in investigating how to solve these subproblems approximately. Our purpose is to study the inexact version of the combination of ADMM with LQP regularization, seeking appropriate inexactness criteria under which the inexact version of ADMM with LQP regularization can be still guaranteed to be convergent. We shall study several inexactness criteria with different levels of implementability. For the resulting algorithms based on the inexact version of ADMM with LQP regularization, we prove their global convergence and establish their worst-case $O(1/t)$ convergence rates.

The rest of this paper is organized as follows. We first state the model to be studied and the motivation of considering the inexact version of ADMM with LQP regularization in Sect. 2. Then, in Sect. 3, we summarize some preliminaries for further analysis. After that, a conceptual inexact version of the ADMM with LQP regularization is presented in Sect. 4; an implementable algorithm based on the inexact version of ADMM with LQP regularization is proposed in Sect. 5. For both the conceptual and the implementable algorithms, the global convergence and a worst-case convergence rate are established. Finally, some concluding remarks are drawn in Sect. 6.

2 Model and Motivation

More specifically, the VIP we consider is to find a $u^* \in \mathcal{U}$ such that

$$(u - u^*)^T F(u^*) \geq 0, \quad \forall u \in \mathcal{U}, \tag{1}$$

with

$$u := \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(u) := \begin{pmatrix} f(x) \\ g(y) \end{pmatrix} \quad \text{and} \tag{2}$$

$$\mathcal{U} := \{(x, y) \mid Ax + By = b, x \in \mathbb{R}_+^n, y \in \mathbb{R}_+^p\},$$

where $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{m \times p}$ are given matrices; $b \in \mathbb{R}^m$ is a given vector; $f : \mathbb{R}_+^n \rightarrow \mathbb{R}^n$ and $g : \mathbb{R}_+^p \rightarrow \mathbb{R}^p$ are continuous and monotone operators.

By attaching a Lagrange multiplier $\lambda \in \mathbb{R}^m$ to the linear constraint in (2), we can reformulate (1)–(2) as: Finding a $w^* \in \mathcal{W}$ such that

$$(w - w^*)^T Q(w^*) \geq 0, \quad \forall w \in \mathcal{W}, \tag{3}$$

where

$$w := \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad Q(w) := \begin{pmatrix} f(x) - A^T \lambda \\ g(y) - B^T \lambda \\ Ax + By - b \end{pmatrix}, \quad \mathcal{W} := \mathbb{R}_+^n \times \mathbb{R}_+^p \times \mathbb{R}^m. \tag{4}$$

We denote by $\text{SVI}(\mathcal{W}, Q)$ the structured VIP (3)–(4); and its solution set, denoted by \mathcal{W}^* , is assumed to be nonempty.

The ADMM in [1] for solving $\text{SVI}(\mathcal{W}, Q)$ is

$$0 \leq x^{k+1} \perp \{f(x^{k+1}) - A^T[\lambda^k - H(Ax^{k+1} + By^k - b)]\} \geq 0, \tag{5}$$

$$0 \leq y^{k+1} \perp \{g(y^{k+1}) - B^T[\lambda^k - H(Ax^{k+1} + By^{k+1} - b)]\} \geq 0, \tag{6}$$

$$\lambda^{k+1} := \lambda^k - H(Ax^{k+1} + By^{k+1} - b), \tag{7}$$

where $H \in \mathbb{R}^{m \times m}$ is a penalty matrix in the metric form. In the literature, it is often chosen as $H := \beta \cdot I$, where I is the identity matrix in $\mathbb{R}^{m \times m}$ and $\beta > 0$ is a scalar.

To truly implement the ADMM scheme (5)–(7), the resulting complementarity problems should be solved efficiently. In fact, for some particular cases, where f, g, A , and B are all special enough, the resulting subproblems of ADMM could be easy enough to have closed-form solutions or can be easily solved up to high precisions. This feature has triggered a tremendous burst of applying ADMM in the literature for solving different problems varying from semidefinite programming, image processing, statistical learning, control, computer vision to numerical linear algebra; see, e.g., [7] and references therein. On the other hand, equally importantly, for generic setting of f, g, A , and B , we can only expect to achieve approximate solutions of these subproblems subject to certain accuracy by applying certain iterative scheme internally. For this case, it is important to analyze under which criterion these subproblems can be solved approximately and how to seek truly implementable inexactness criteria. In [4, 5] the ADMM subproblems were suggested to be regularized by the LQP regularization proposed in [6]. The LQP regularization forces the solutions of ADMM subproblems to be interior points of \mathbb{R}_+^n and \mathbb{R}_+^p , respectively, thus the complementarity subproblems (5) and (6) reduce to systems of nonlinear equations. More specifically, the iterative scheme of ADMM with LQP regularization is as follows:

$$\begin{aligned} & f(x^{k+1}) - A^T[\lambda^k - H(Ax^{k+1} + By^k - b)] \\ & + R[(x^{k+1} - x^k) + \mu(x^k - X_k^2(x^{k+1})^{-1})] = 0, \end{aligned} \tag{8}$$

$$\begin{aligned} & g(y^{k+1}) - B^T[\lambda^k - H(Ax^{k+1} + By^{k+1} - b)] \\ & + S[(y^{k+1} - y^k) + \mu(y^k - Y_k^2(y^{k+1})^{-1})] = 0, \end{aligned} \tag{9}$$

$$\lambda^{k+1} := \lambda^k - H(Ax^{k+1} + By^{k+1} - b), \tag{10}$$

where $R \in \mathbb{R}^{n \times n}$ and $S \in \mathbb{R}^{p \times p}$ are symmetric positive definite diagonal matrices and they are called proximal matrices; $R(x^{k+1} - x^k)$ and $S(y^{k+1} - y^k)$ are quadratic proximal regularization terms, $\mu \in]0, 1[$ is a given constant, $X_k := \text{diag}(x_1^k, x_2^k, \dots, x_n^k)$, $(x^{k+1})^{-1} \in \mathbb{R}^n$ is a vector whose j th element is $1/x_j^{k+1}$, $Y_k := \text{diag}(y_1^k, y_2^k, \dots, y_n^k)$, $(y^{k+1})^{-1} \in \mathbb{R}^p$ is a vector whose j th element is $1/y_j^{k+1}$. In [8], this scheme was further shown to be convergent on a worst-case $O(1/t)$ rate, provided that both

the x - and the y -subproblems (i.e., (8) and (9)) are solved exactly. Here, the worst-case $O(1/t)$ convergence rate of ADMM with LQP regularization means the scheme (8)–(10) achieves a solution of SVI(\mathcal{W}, Q) with the accuracy of $O(1/t)$ after most t iterations.

To solve the complementarity subproblems (8) and (9), we shall study several inexactness criteria with different levels of implementability. When the accuracy of solving these subproblems is relaxed, it is easy to understand that some additional corrections might need to be supplemented after the iteration of ADMM with LQP regularization. Or, relaxing the accuracy of solving these subproblems might pay the price of requiring additional correction on the subproblems' solutions. We thus suggest to solve these subproblems within the framework of either under a more restricted criterion but without further correction step, or under a more relaxed criterion but with some further correction step. How to balance the accuracy of inner subproblems and further correction depends on how special a concrete application of the abstract model (3)–(4) is, and it is too inane to discuss this issue for the abstract model (3)–(4) without any specification of the involved functions and sets. Instead, in this paper, we only provide some inexact algorithms based on the framework of combining ADMM with LQP regularization. In addition to the global convergence, a worst-case $O(1/t)$ convergence rate is established for these inexact algorithms. Therefore, we show that these inexact algorithms based on the combination of ADMM with LQP regularization, which are more practical than the exact version (8)–(10), enjoy the same worst-case convergence rate as the exact version.

3 Preliminaries

In this section, we recall some basic definitions and properties, which will be frequently used in our later analysis. Some useful results proved already in the literature are also summarized.

3.1 Some Basic Definitions and Properties

Let N be a positive definite matrix with appropriate dimensionality, the N -norm of a vector v with appropriate dimensionality is denoted by $\|v\|_N := \sqrt{v^T N v}$. Let Ω be a nonempty, closed, and convex subset of \mathbb{R}^l . The projection operator under the N -norm is defined by

$$P_{\Omega, N}(v) := \operatorname{argmin}\{\|v - u\|_N \mid u \in \Omega\}.$$

Clearly, according to this definition, we have the following property related to the projection operator under the N -norm:

$$\|P_{\Omega, N}(v) - w\|_N \leq \|v - w\|_N, \quad \forall v \in \mathbb{R}^l, w \in \Omega. \tag{11}$$

In the following, we give the definitions of monotonicity and Lipschitz continuity for a mapping.

Definition 3.1 Let $\mathcal{X} \subseteq \mathbb{R}^n$ and $f : \mathcal{X} \rightarrow \mathbb{R}^n$ be a mapping. Then,

- (i) f is said to be monotone with respect to \mathcal{X} iff the following inequality always holds:

$$(x - \tilde{x})^T (f(x) - f(\tilde{x})) \geq 0, \quad \forall x, \tilde{x} \in \mathcal{X};$$

- (ii) f is Lipschitz continuous with respect to \mathcal{X} iff there exists a constant (called Lipschitz constant) $L_f > 0$ such that

$$\|f(x) - f(\tilde{x})\| \leq L_f \|x - \tilde{x}\|, \quad \forall x, \tilde{x} \in \mathcal{X}.$$

We denote by L_f and L_g the Lipschitz constants of the mappings f and g in the model (1)–(2). Note that the mapping $Q(w)$ in (4) is monotone with respect to \mathcal{W} under the monotonicity assumption of f and g . Therefore, the solution set \mathcal{W}^* of SVI(\mathcal{W}, Q) is closed and convex; see, e.g., [9].

Throughout the rest of the paper, we define the matrix G as

$$G := \begin{pmatrix} (1 + \mu)R & 0 & 0 \\ 0 & (1 + \mu)S + B^T H B & 0 \\ 0 & 0 & H^{-1} \end{pmatrix}. \tag{12}$$

Last, we recall a characterization of the solution set \mathcal{W}^* proved in [9], see (2.3.2) in p. 159 of [9]. As shown in [8, 10], this characterization is crucial for establishing the convergence rate for ADMM or ADMM with LQP regularization. More specifically, recall that \mathcal{W}^* can be characterized as

$$\mathcal{W}^* := \bigcap_{w \in \mathcal{W}} \{ \bar{w} \in \mathcal{W} \mid (w - \bar{w})^T Q(w) \geq 0 \}.$$

Therefore, as Definition 1 in [11], $\bar{w} \in \mathcal{W}$ can be regarded as an ε -approximation solution of SVI(\mathcal{W}, Q) if it satisfies

$$\sup_{w \in \mathcal{B}_{\mathcal{W}}(\bar{w})} \{ (\bar{w} - w)^T Q(w) \} \leq \varepsilon, \quad \text{where } \mathcal{B}_{\mathcal{W}}(\bar{w}) := \{ w \in \mathcal{W} \mid \|w - \bar{w}\| \leq 1 \}. \tag{13}$$

In our later analysis, we shall establish the worst-case $O(1/t)$ convergence rate for the new algorithms to be proposed in the sense that after t iterations of these algorithms, we can find a $\bar{w} \in \mathcal{W}$ such that

$$(\bar{w} - w)^T Q(w) \leq \varepsilon, \quad \forall w \in \mathcal{B}_{\mathcal{W}}(\bar{w}),$$

with $\varepsilon = O(1/t)$.

3.2 A Key Lemma

Inspired by Proposition 1 in [6], the following lemma was proved in the literature (see, e.g., [5, 12]), and its conclusion plays an important role in analyzing the convergence of ADMM with LQP regularization and establishing its convergence rate. Our later analysis is also based on this conclusion.

Lemma 3.1 *Let $P := \text{diag}(p_1, p_2, \dots, p_t) \in \mathbb{R}^{t \times t}$ be a positive definite diagonal matrix, $q(u) \in \mathbb{R}^t$ be a monotone mapping of u with respect to \mathbb{R}^t_+ , and $\mu \in]0, 1[$. For a given $\bar{u} \in \mathbb{R}^t_{++}$, we define $\bar{U} := \text{diag}(\bar{u}_1, \bar{u}_2, \dots, \bar{u}_t)$. Then, the equation*

$$q(u) + P[(u - \bar{u}) + \mu(\bar{u} - \bar{U}^2 u^{-1})] = 0 \tag{14}$$

has the unique positive solution u . In addition, for this positive solution $u \in \mathbb{R}^t_{++}$ and any $v \in \mathbb{R}^t_+$, we have

$$(v - u)^T q(u) \geq \frac{1 + \mu}{2} (\|u - v\|_P^2 - \|\bar{u} - v\|_P^2) + \frac{1 - \mu}{2} \|\bar{u} - u\|_P^2. \tag{15}$$

Moreover, we have

$$(v - u)^T q(u) \geq (1 + \mu)(\bar{u} - u)^T P(v - u) - \mu \|\bar{u} - u\|_P^2. \tag{16}$$

3.3 Proof Framework of Convergence

Later, we shall propose several algorithms, and they differ in the inexactness criteria for solving the ADMM subproblems with LQP regularization. The convergence proofs for these different algorithms share a common framework, even though the specific techniques required are quite different in their respective proofs. For succinctness purpose, in this subsection we first show the framework of proving the convergence before these algorithms are presented. Then, when we establish the convergence for a specific algorithm later, it suffices to check the conditions presented in this framework. The proof framework presented below is thus a unified treatment on the convergence analysis for different algorithms.

We first establish an identity which will be used later for convergence. The proof of this lemma only requires straightforward manipulation.

Lemma 3.2 *Let $\bar{x}^k \in \mathbb{R}^n$, $y^k, \bar{y}^k \in \mathbb{R}^p$ and $\lambda^k \in \mathbb{R}^m$ be given. We define $\bar{\lambda}^k$ and $\hat{\lambda}^{k+1}$ as*

$$\bar{\lambda}^k := \lambda^k - H(A\bar{x}^k + B\bar{y}^k - b) \quad \text{and} \quad \hat{\lambda}^{k+1} := \lambda^k - H(A\bar{x}^k + B\bar{y}^k - b).$$

Then, for any $y \in \mathbb{R}^p$ and $\lambda \in \mathbb{R}^m$, we have

$$\begin{aligned} & 2(\lambda - \bar{\lambda}^k)^T (A\bar{x}^k + B\bar{y}^k - b) + 2(y - \bar{y}^k)^T B^T H B (y^k - \bar{y}^k) \\ &= \|y - \bar{y}^k\|_{B^T H B}^2 - \|y - y^k\|_{B^T H B}^2 + \|\lambda - \hat{\lambda}^{k+1}\|_{H^{-1}}^2 \\ & \quad - \|\lambda - \lambda^k\|_{H^{-1}}^2 + \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2. \end{aligned}$$

Proof Using the definition of $\hat{\lambda}^{k+1}$, we have

$$\begin{aligned} & 2(\lambda - \bar{\lambda}^k)^T (A\bar{x}^k + B\bar{y}^k - b) \\ &= 2(\lambda - \bar{\lambda}^k)^T H^{-1}(\lambda^k - \hat{\lambda}^{k+1}) \\ &= 2(\lambda - \hat{\lambda}^{k+1})^T H^{-1}(\lambda^k - \hat{\lambda}^{k+1}) + 2(\hat{\lambda}^{k+1} - \bar{\lambda}^k)^T H^{-1}(\lambda^k - \hat{\lambda}^{k+1}). \tag{17} \end{aligned}$$

For the first term in (17), we have the following identity:

$$2(\lambda - \hat{\lambda}^{k+1})^T H^{-1}(\lambda^k - \hat{\lambda}^{k+1}) = \|\lambda - \hat{\lambda}^{k+1}\|_{H^{-1}}^2 - \|\lambda - \lambda^k\|_{H^{-1}}^2 + \|\lambda^k - \hat{\lambda}^{k+1}\|_{H^{-1}}^2. \tag{18}$$

Then, it follows from the definitions of $\bar{\lambda}^k$ and $\hat{\lambda}^{k+1}$ that

$$\hat{\lambda}^{k+1} = \bar{\lambda}^k + HB y^k - HB \bar{y}^k,$$

and thus

$$\begin{aligned} & 2(\hat{\lambda}^{k+1} - \bar{\lambda}^k)^T H^{-1}(\lambda^k - \hat{\lambda}^{k+1}) \\ &= 2(HB y^k - HB \bar{y}^k)^T H^{-1}(\lambda^k - \hat{\lambda}^{k+1}) \\ &= \|\lambda^k - \hat{\lambda}^{k+1} + HB y^k - HB \bar{y}^k\|_{H^{-1}}^2 - \|\lambda^k - \hat{\lambda}^{k+1}\|_{H^{-1}}^2 - \|HB y^k - HB \bar{y}^k\|_{H^{-1}}^2 \\ &= \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2 - \|\lambda^k - \hat{\lambda}^{k+1}\|_{H^{-1}}^2 - \|y^k - \bar{y}^k\|_{B^T H B}^2. \end{aligned}$$

Substituting this and (18) into (17), we have

$$\begin{aligned} & 2(\lambda - \bar{\lambda}^k)^T (A\bar{x}^k + B\bar{y}^k - b) \\ &= \|\lambda - \hat{\lambda}^{k+1}\|_{H^{-1}}^2 - \|\lambda - \lambda^k\|_{H^{-1}}^2 + \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2 - \|y^k - \bar{y}^k\|_{B^T H B}^2. \end{aligned} \tag{19}$$

Note that

$$2(y - \bar{y}^k)^T B^T H B (y^k - \bar{y}^k) = \|y - \bar{y}^k\|_{B^T H B}^2 - \|y - y^k\|_{B^T H B}^2 + \|y^k - \bar{y}^k\|_{B^T H B}^2.$$

Adding the above equality and (19), the assertion is proved. □

Next, we will present a unified framework for proving the convergence of our new algorithms to be presented.

Theorem 3.1 *Let $c_0 \in]0, +\infty[$ and $\mu \in]0, 1[$ be positive constants; $\{\eta_k\}$ be a non-negative sequence with $\sum_{k=0}^\infty \eta_k < +\infty$; and G be defined by (12). For any $w^* \in \mathcal{W}^*$, if the sequences $\{w^k = (x^k, y^k, \lambda^k)\}$ and $\{\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{\lambda}^k)\}$ satisfy*

$$\begin{aligned} \|w^{k+1} - w^*\|_G^2 &\leq \|w^k - w^*\|_G^2 + \eta_k - c_0(\|x^k - \bar{x}^k\|_R^2 \\ &\quad + \|y^k - \bar{y}^k\|_S^2 + \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2), \quad \forall k \geq 0, \end{aligned} \tag{20}$$

then $\{w^k\}$ is bounded and

$$\lim_{k \rightarrow \infty} \|w^k - \bar{w}^k\|_G = 0. \tag{21}$$

Furthermore, if the mapping Q is continuous and

$$\liminf_{k \rightarrow \infty} (w - \bar{w}^k)^T Q(\bar{w}^k) \geq 0, \quad \forall w \in \mathcal{W}, \tag{22}$$

then the sequence $\{w^k\}$ converges to a point in \mathcal{W}^* .

Proof It follows from (20) that for any $l \leq k$ and $w^* \in \mathcal{W}^*$, we have

$$\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 + \eta_k \leq \|w^l - w^*\|_G^2 + \sum_{i=l}^k \eta_i. \tag{23}$$

Thus the sequence $\{w^k\}$ is bounded, since $\sum_{i=0}^\infty \eta_i < +\infty$. Summing the inequality (20) over $k = 0, 1, \dots$, we get

$$c_0 \sum_{k=0}^\infty (\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2 + \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2) \leq \|w^0 - w^*\|_G^2 + \sum_{k=0}^\infty \eta_k < +\infty.$$

Therefore, we have

$$\lim_{k \rightarrow \infty} (\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2 + \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2) = 0.$$

It follows that

$$\lim_{k \rightarrow \infty} \|x^k - \bar{x}^k\|_R = 0, \quad \lim_{k \rightarrow \infty} \|y^k - \bar{y}^k\|_S = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}} = 0.$$

The first assertion (21) is proved.

Since $\{w^k\}$ is bounded and $\lim_{k \rightarrow \infty} \|w^k - \bar{w}^k\|_G = 0$, we find that $\{\bar{w}^k\}$ is also bounded and then it has at least one cluster point. Let w^∞ be a cluster point of $\{\bar{w}^k\}$ and the subsequences $\{\bar{w}^{k_j}\}$ and $\{w^{k_j}\}$ both converge to w^∞ . It follows from (22) that

$$\liminf_{j \rightarrow \infty} (w - \bar{w}^{k_j})^T Q(\bar{w}^{k_j}) \geq 0, \quad \forall w \in \mathcal{W},$$

and consequently

$$(w - w^\infty)^T Q(w^\infty) \geq 0, \quad \forall w \in \mathcal{W},$$

since the mapping Q is continuous. This means that w^∞ is a solution of $\text{SVI}(\mathcal{W}, Q)$. Note that the inequality (23) is true for all solution points of $\text{SVI}(\mathcal{W}, Q)$, hence we have

$$\|w^{k+1} - w^\infty\|_G^2 \leq \|w^l - w^\infty\|_G^2 + \sum_{i=l}^\infty \eta_i, \quad \forall k \geq 0, \forall l \leq k. \tag{24}$$

Since $w^{k_j} \rightarrow w^\infty$ ($j \rightarrow \infty$) and $\sum_{i=0}^\infty \eta_i < +\infty$, for any given $\varepsilon > 0$, there exists a $j_0 > 0$ such that

$$\|w^{k_{j_0}} - w^\infty\|_G^2 \leq \frac{\varepsilon^2}{2} \quad \text{and} \quad \sum_{i=k_{j_0}}^\infty \eta_i \leq \frac{\varepsilon^2}{2}. \tag{25}$$

Therefore, for any $k \geq k_{j_0}$, it follows from (24) and (25) that

$$\|w^{k+1} - w^\infty\|_G \leq \sqrt{\|w^{k_{j_0}} - w^\infty\|_G^2 + \sum_{i=k_{j_0}}^\infty \eta_i} \leq \varepsilon.$$

This implies that the sequence $\{w^k\}$ converges to a point w^∞ in \mathcal{W}^* . □

According to Theorem 3.1, it suffices to check conditions (20) and (22) when we establish the convergence for the algorithms to be proposed. As the new algorithms are based on the combination of ADMM with LQP regularization, where the subproblems are solved inexactly, we abbreviate these algorithms as I-ADMM-LQP X, where “X” is an arabian number representing a concrete algorithm.

4 The First Algorithm

We propose the first algorithm based on the combination of ADMM with LQP regularization where the subproblems are solved approximately. The inexactness criterion required by this algorithm involves the exact solutions of the resulting subproblems, thus is not applicable directly in implementation. We still present this algorithm, however, because its convergence analysis provides a similar sketch for the other algorithms to be analyzed later, and is quite simple. Thus, by this simpler analysis, it becomes easier to understand the latter analysis for other implementable algorithms. In addition, although the inexactness criterion itself is not checkable directly, it provides the possibility to seek implementable criteria based on some error bounds; see, e.g., [9, 13]. Thus, it deserves at least theoretical interest. Note this criterion is widely used in PPA and ADMM literature.

4.1 Algorithm 1

In this algorithm, we seek approximate solutions of the subproblems (8)–(9) and the accuracy is controlled by a summable sequence of positive scales.

Algorithm 1 (I-ADMM-LQP 1)

Step 0. Let $\varepsilon > 0$; $\mu \in]0, 1[$; $w^0 := (x^0, y^0, \lambda^0) \in \mathbb{R}_{++}^n \times \mathbb{R}_{++}^p \times \mathbb{R}^m$; $R \in \mathbb{R}^{n \times n}$, $S \in \mathbb{R}^{p \times p}$ and $H \in \mathbb{R}^{m \times m}$ be positive definite diagonal matrices; and $\{\nu_k\}$ be a nonnegative sequence satisfying $\sum_{k=0}^\infty \nu_k < +\infty$. Set $k = 0$.

Step 1. Find $x^{k+1} \in \mathbb{R}_{++}^n$ such that

$$\|x^{k+1} - x_*^{k+1}\| \leq \nu_k, \tag{26}$$

where x_*^{k+1} is the exact solution of the system

$$f(x) - A^T[\lambda^k - H(Ax + By^k - b)] + R[(x - x^k) + \mu(x^k - X_k^2 x^{-1})] = 0. \tag{27}$$

Step 2. Find $y^{k+1} \in \mathbb{R}_{++}^p$ such that

$$\|y^{k+1} - y_*^{k+1}\| \leq \nu_k, \tag{28}$$

where y_*^{k+1} is the exact solution of the system

$$g(y) - B^T[\lambda^k - H(Ax_*^{k+1} + By - b)] + S[(y - y^k) + \mu(y^k - Y_k^2 y^{-1})] = 0. \tag{29}$$

Step 3. Update the Lagrange multiplier

$$\lambda^{k+1} := \lambda^k - H(Ax^{k+1} + By^{k+1} - b). \tag{30}$$

Step 4. Set $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$. If $\|w^{k+1} - w^k\| \leq \varepsilon$, stop; otherwise set $k = k + 1$ and goto Step 1.

Remark 4.1 It follows from Lemma 3.1 that there exist unique $x_*^{k+1} \in \mathbb{R}_{++}^n$ and $y_*^{k+1} \in \mathbb{R}_{++}^p$ satisfying (27) and (29), respectively. This guarantees that we can find $x^{k+1} \in \mathbb{R}_{++}^n$ and $y^{k+1} \in \mathbb{R}_{++}^p$ satisfying (26) and (28), respectively.

4.2 Convergence

In this subsection, we prove the convergence of Algorithm 1. First, we define

$$\lambda_*^{k+1} := \lambda^k - H(Ax_*^{k+1} + By_*^{k+1} - b), \tag{31}$$

and set

$$w_*^{k+1} := \begin{pmatrix} x_*^{k+1} \\ y_*^{k+1} \\ \lambda_*^{k+1} \end{pmatrix} \quad \text{and} \quad \bar{w}^k := \begin{pmatrix} \bar{x}^k \\ \bar{y}^k \\ \bar{\lambda}^k \end{pmatrix} = \begin{pmatrix} x_*^{k+1} \\ y_*^{k+1} \\ \lambda^k - H(Ax_*^{k+1} + By^k - b) \end{pmatrix} \tag{32}$$

to simplify our notation in the following analysis.

To prove the convergence of Algorithm 1, we first present some lemmas.

Lemma 4.1 *Let $\{w_*^{k+1}\}$ be defined by (32) and $\{w^k\}$ be generated by Algorithm 1. Then, there exists a positive constant ρ such that*

$$\|w_*^{k+1} - w^{k+1}\|_G \leq \rho v_k, \quad \forall k \geq 0, \tag{33}$$

where G is defined by (12).

Proof It follows from (30) and (31) that

$$\lambda_*^{k+1} - \lambda^{k+1} = HA(x^{k+1} - x_*^{k+1}) + HB(y^{k+1} - y_*^{k+1}).$$

Together with (26), (28), and (12), the above equation implies (33) immediately. \square

Lemma 4.2 *Let the sequence $\{w^k\}$ be generated by Algorithm 1, and the accompanying sequences $\{w_*^{k+1}\}$ and $\{\bar{w}^k\}$ be defined by (32). Then, for any $w := (x, y, \lambda) \in \mathcal{W}$, we have*

$$(w - \bar{w}^k)^T Q(\bar{w}^k) \geq \frac{1}{2}(\|w_*^{k+1} - w\|_G^2 - \|w^k - w\|_G^2) + \frac{1 - \mu}{2}(\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2) + \frac{1}{2}\|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2, \tag{34}$$

where G is defined by (12).

Proof Note that $\bar{x}^k = x_*^{k+1}$ and $\bar{y}^k = y_*^{k+1}$. Applying Lemma 3.1 to (27) by setting $P = R$, $\bar{u} = x^k$, $u = \bar{x}^k = x_*^{k+1}$, $q(u) = f(\bar{x}^k) - A^T[\lambda^k - H(A\bar{x}^k + By^k - b)] \stackrel{(32)}{=} f(\bar{x}^k) - A^T\bar{\lambda}^k$ and $v = x$ in (15), we have

$$(x - \bar{x}^k)^T [f(\bar{x}^k) - A^T\bar{\lambda}^k] \geq \frac{1 + \mu}{2} (\|x_*^{k+1} - x\|_R^2 - \|x^k - x\|_R^2) + \frac{1 - \mu}{2} \|x^k - \bar{x}^k\|_R^2, \quad \forall x \in \mathbb{R}_+^n. \tag{35}$$

Similarly, applying Lemma 3.1 to (29) and using (32), we get

$$(y - \bar{y}^k)^T [g(\bar{y}^k) - B^T\bar{\lambda}^k + B^T HB(\bar{y}^k - y^k)] \geq \frac{1 + \mu}{2} (\|y_*^{k+1} - y\|_S^2 - \|y^k - y\|_S^2) + \frac{1 - \mu}{2} \|y^k - \bar{y}^k\|_S^2, \quad \forall y \in \mathbb{R}_+^p.$$

It follows from the above inequality that

$$(y - \bar{y}^k)^T [g(\bar{y}^k) - B^T\bar{\lambda}^k] \geq (y - \bar{y}^k)^T B^T HB(y^k - \bar{y}^k) + \frac{1 + \mu}{2} (\|y_*^{k+1} - y\|_S^2 - \|y^k - y\|_S^2) + \frac{1 - \mu}{2} \|y^k - \bar{y}^k\|_S^2. \tag{36}$$

Setting $\hat{\lambda}^{k+1} = \lambda_*^{k+1}$ in Lemma 3.2 and using $\bar{y}^k = y_*^{k+1}$, we have

$$(\lambda - \bar{\lambda}^k)^T (A\bar{x}^k + B\bar{y}^k - b) = (y - \bar{y}^k)^T B^T HB(\bar{y}^k - y^k) + \frac{1}{2} (\|y - y_*^{k+1}\|_{B^T HB}^2 - \|y - y^k\|_{B^T HB}^2) + \frac{1}{2} (\|\lambda - \lambda_*^{k+1}\|_{H^{-1}}^2 - \|\lambda - \lambda^k\|_{H^{-1}}^2) + \frac{1}{2} \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2.$$

Combining (35), (36) and the above equation together and by simple manipulations, we can get (34) immediately. The proof is completed. □

The following result shows the contraction of the sequence generated by Algorithm 1, based on which the convergence of Algorithm 1 can be established easily.

Lemma 4.3 *Let the sequence $\{w^k\}$ be generated by Algorithm 1. Then $\{w^k\}$ is bounded, i.e., for any $w^* \in \mathcal{W}^*$, there are positive constants C_{w^*} and ρ , such that*

$$\|w^k - w^*\|_G \leq C_{w^*}, \quad \forall k \geq 0,$$

and

$$\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 + 2\rho C_{w^*} v_k + \rho^2 v_k^2 - (1 - \mu) (\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2 + \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2),$$

where G is defined by (12).

Proof Setting $w = w^*$ in (34), we get

$$2(w^* - \bar{w}^k)^T Q(\bar{w}^k) \geq \|w_*^{k+1} - w^*\|_G^2 - \|w^k - w^*\|_G^2 + (1 - \mu)(\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2) + \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2.$$

On the other hand, since Q is monotone, $\bar{w}^k \in \mathcal{W}$, and $w^* \in \mathcal{W}^*$, we have

$$0 \geq (w^* - \bar{w}^k)^T Q(w^*) \geq (w^* - \bar{w}^k)^T Q(\bar{w}^k).$$

From the above two inequalities and using $\mu \in]0, 1[$, we obtain

$$\|w_*^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 - (1 - \mu)(\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2 + \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2). \tag{37}$$

It follows from (37) and Lemma 4.1 that there is a positive constant ρ such that

$$\|w^{k+1} - w^*\|_G \leq \|w_*^{k+1} - w^*\|_G + \|w^{k+1} - w_*^{k+1}\|_G \leq \|w^k - w^*\|_G + \rho v_k.$$

Consequently, for any $k \geq 0$ we get

$$\|w^{k+1} - w^*\|_G \leq \|w^0 - w^*\|_G + \rho \sum_{i=0}^k v_i \leq \|w^0 - w^*\|_G + \rho \sum_{i=0}^{\infty} v_i := C_{w^*} < +\infty. \tag{38}$$

Therefore, the sequence $\{w^k\}$ generated by Algorithm 1 is bounded. It follows from (33), (37) and (38) that

$$\begin{aligned} & \|w^{k+1} - w^*\|_G^2 \\ &= \|(w_*^{k+1} - w^*) + (w^{k+1} - w_*^{k+1})\|_G^2 \\ &\leq \|w_*^{k+1} - w^*\|_G^2 + 2\|w_*^{k+1} - w^*\|_G \cdot \|w^{k+1} - w_*^{k+1}\|_G + \|w^{k+1} - w_*^{k+1}\|_G^2 \\ &\leq \|w^k - w^*\|_G^2 - (1 - \mu)(\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2 + \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2) \\ &\quad + 2\rho v_k \|w^k - w^*\|_G + \rho^2 v_k^2 \\ &\leq \|w^k - w^*\|_G^2 + 2\rho C_{w^*} v_k + \rho^2 v_k^2 \\ &\quad - (1 - \mu)(\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2 + \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2). \end{aligned}$$

The proof is completed. □

Now, we are ready to prove the convergence of Algorithm 1.

Theorem 4.1 *The sequence $\{w^k\}$ generated by Algorithm 1 converges to some w^∞ which is a solution of $\text{SVI}(\mathcal{W}, Q)$.*

Proof From $\sum_{k=0}^{\infty} v_k < +\infty$ and $v_k \geq 0$, it follows that

$$\sum_{k=0}^{\infty} (2\rho C_{w^*} v_k + \rho^2 v_k^2) < +\infty.$$

Setting $\eta_k = 2\rho C_{w^*} v_k + \rho^2 v_k^2$, $c_0 = 1 - \mu$ in (20), from Lemma 4.3 and Theorem 3.1 we have

$$\lim_{k \rightarrow \infty} \|w^k - \bar{w}^k\|_G = 0. \tag{39}$$

Then it follows from (12), (31), and (32) that

$$\lim_{k \rightarrow \infty} \|w_*^{k+1} - \bar{w}^k\|_G = \lim_{k \rightarrow \infty} \|\lambda_*^{k+1} - \bar{\lambda}^k\|_{H^{-1}} = \lim_{k \rightarrow \infty} \|H(By^k - B\bar{y}^k)\|_{H^{-1}} = 0.$$

From (34) and the above two formulas, we get

$$\liminf_{k \rightarrow \infty} (w - \bar{w}^k)^T Q(\bar{w}^k) \geq 0, \quad \forall w \in \mathcal{W}.$$

The convergence of Algorithm 1 is then obtained immediately from Theorem 3.1. \square

4.3 Convergence Rate

Now, we show the worst-case $O(1/t)$ convergence rate for Algorithm 1.

Theorem 4.2 *For any integer $t > 0$, there is a $\bar{w}_t \in \mathcal{W}$ which is a convex combination of the iterates $\bar{w}^0, \bar{w}^1, \dots, \bar{w}^t$ defined by (32). Then, for any $w \in \mathcal{W}$, we have*

$$(\bar{w}_t - w)^T Q(w) \leq \frac{1}{t+1} \left(\frac{1}{2} \|w^0 - w\|_G^2 + \rho \sum_{k=0}^t v_k \|w^{k+1} - w\|_G \right), \tag{40}$$

where $\bar{w}_t := (\sum_{k=0}^t \bar{w}^k) / (t+1)$.

Proof From (34), we have

$$(w - \bar{w}^k)^T Q(\bar{w}^k) + \frac{1}{2} \|w^k - w\|_G^2 \geq \frac{1}{2} \|w_*^{k+1} - w\|_G^2, \quad \forall w \in \mathcal{W}.$$

It follows from (33) that

$$\begin{aligned} \|w_*^{k+1} - w\|_G^2 &\geq (\|w^{k+1} - w\|_G - \|w^{k+1} - w_*^{k+1}\|_G)^2 \\ &= \|w^{k+1} - w\|_G^2 - 2\|w^{k+1} - w\|_G \cdot \|w^{k+1} - w_*^{k+1}\|_G \\ &\quad + \|w^{k+1} - w_*^{k+1}\|_G^2 \\ &\geq \|w^{k+1} - w\|_G^2 - 2\rho v_k \|w^{k+1} - w\|_G, \quad \forall w \in \mathcal{W}. \end{aligned}$$

Since Q is monotone, from the above two inequalities, we have

$$\begin{aligned} & (w - \bar{w}^k)^T Q(w) + \frac{1}{2} \|w^k - w\|_G^2 \\ & \geq \frac{1}{2} \|w^{k+1} - w\|_G^2 - \rho v_k \|w^{k+1} - w\|_G, \quad \forall w \in \mathcal{W}. \end{aligned} \tag{41}$$

Summing the inequality (41) over $k = 0, 1, \dots, t$, we obtain

$$\begin{aligned} & \left[(t + 1)w - \left(\sum_{k=0}^t \bar{w}^k \right) \right]^T Q(w) + \frac{1}{2} \|w^0 - w\|_G^2 \\ & \geq \frac{1}{2} \|w^{t+1} - w\|_G^2 - \rho \sum_{k=0}^t v_k \|w^{k+1} - w\|_G \\ & \geq -\rho \sum_{k=0}^t v_k \|w^{k+1} - w\|_G, \quad \forall w \in \mathcal{W}. \end{aligned}$$

Since $\sum_{k=0}^t 1/(t + 1) = 1$, \bar{w}_t is a convex combination of $\bar{w}^0, \bar{w}^1, \dots, \bar{w}^t$ and thus $\bar{w}_t \in \mathcal{W}$. Using the notation of \bar{w}_t , we derive

$$(w - \bar{w}_t)^T Q(w) + \frac{1}{2(t + 1)} \|w^0 - w\|_G^2 \geq -\frac{\rho}{t + 1} \sum_{k=0}^t v_k \|w^{k+1} - w\|_G, \quad \forall w \in \mathcal{W}.$$

The assertion (40) follows from the above inequality immediately. □

It follows from Lemma 4.3 that the sequence $\{w^k\}$ generated by Algorithm 1 is bounded. According to (39), the sequence $\{\bar{w}^k\}$ defined by (32) is also bounded. Therefore, there exists a constant $D > 0$ such that

$$\|w^k\|_G \leq D \quad \text{and} \quad \|\bar{w}^k\|_G \leq D, \quad \forall k \geq 0.$$

Recall that \bar{w}_t is the average of $\{\bar{w}^0, \bar{w}^1, \dots, \bar{w}^t\}$. Thus, we have $\|\bar{w}_t\|_G \leq D$. Denote $E_1 := \sum_{k=0}^\infty v_k < +\infty$. For any $w \in \mathcal{B}_{\mathcal{W}}(\bar{w}_t) := \{w \in \mathcal{W} \mid \|w - \bar{w}_t\|_G \leq 1\}$, we get

$$\begin{aligned} & (\bar{w}_t - w)^T Q(w) \\ & \leq \frac{1}{t + 1} \left(\frac{1}{2} \|w^0 - w\|_G^2 + \rho \sum_{k=0}^t v_k \|w^{k+1} - w\|_G \right) \\ & \leq \frac{1}{t + 1} \left[\frac{1}{2} (\|w^0 - \bar{w}_t\|_G + \|\bar{w}_t - w\|_G)^2 \right. \\ & \quad \left. + \rho \sum_{k=0}^t v_k (\|w^{k+1} - \bar{w}_t\|_G + \|\bar{w}_t - w\|_G) \right] \end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{t+1} \left[\frac{1}{2} (\|w^0\|_G + \|\bar{w}_t\|_G + \|\bar{w}_t - w\|_G)^2 \right. \\ &\quad \left. + \rho \sum_{k=0}^t v_k (\|w^{k+1}\|_G + \|\bar{w}_t\|_G + \|\bar{w}_t - w\|_G) \right] \\ &\leq \frac{1}{t+1} \left[\frac{1}{2} (2D+1)^2 + \rho E_1 (2D+1) \right]. \end{aligned}$$

Thus, for any given $\varepsilon > 0$, after at most $t := \lceil \frac{(2D+1)(2D+1+2\rho E_1)}{2\varepsilon} - 1 \rceil$ iterations, we have

$$(\bar{w}_t - w)^T Q(w) \leq \varepsilon, \quad \forall w \in \mathcal{B}_{\mathcal{W}}(\bar{w}_t),$$

which means \bar{w}_t is an approximate solution of $\text{SVI}(\mathcal{W}, Q)$ with an accuracy of $O(1/t)$. That is, a worst-case $O(1/t)$ convergence rate of Algorithm 1 in ergodic sense is established.

5 The Second Algorithm

As we have mentioned, the inexactness criterion in Algorithm 1 (i.e., (26) and (28)) is not applicable directly due to the lack of x_*^{k+1} and y_*^{k+1} . Now we propose a new inexactness criterion in the absence of these exact solutions and embed it into the subproblems (8) and (9). A new implementable ADMM with LQP regularization is thus proposed. Note the new inexactness criterion allows the relative error of solving the subproblems (8) and (9) to be fixed as a constant.

5.1 Algorithm 2

To ensure the convergence when the relative error of solving the subproblems (8) and (9) is fixed as a constant, the approximate solutions of (8) and (9) should be corrected to generate a new iterate $(x^{k+1}, y^{k+1}, \lambda^{k+1})$. In this algorithm, we thus use $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \mathbb{R}_{++}^n \times \mathbb{R}_{++}^p \times \mathbb{R}^m$ to denote the approximate solutions of (8)–(10). Below, for notational simplicity we choose $R := rI$ and $S := sI$ with $r, s > 0$ as the proximal matrices for this algorithm.

Algorithm 2 (I-ADMM-LQP 2)

Step 0. Let $\varepsilon > 0; \mu, \nu, \sigma \in]0, 1[; \gamma \in]0, 2[; r, s > 0; w^0 := (x^0, y^0, \lambda^0) \in \mathbb{R}_{++}^n \times \mathbb{R}_{++}^p \times \mathbb{R}^m$ and H be a positive definite diagonal matrix. Set $k = 0$.

Step 1. Find $\tilde{x}^k \in \mathbb{R}_{++}^n$ and $\xi_x^k \in \mathbb{R}^n$ such that

$$\begin{aligned} &f(\tilde{x}^k) - A^T [\lambda^k - H(A\tilde{x}^k + By^k - b)] \\ &\quad + R \{ (\tilde{x}^k - x^k) + \mu [x^k - X_k^2(\tilde{x}^k)^{-1}] \} + \xi_x^k = 0, \end{aligned} \tag{42}$$

where ξ_x^k satisfies the following inexactness criterion:

$$\|\xi_x^k\| \leq \nu r(1 - \mu)\|x^k - \tilde{x}^k\|. \tag{43}$$

Step 2. Find $\tilde{y}^k \in \mathbb{R}_{++}^p$ and $\xi_y^k \in \mathbb{R}^p$ such that

$$\begin{aligned} &g(\tilde{y}^k) - B^T[\lambda^k - H(A\tilde{x}^k + B\tilde{y}^k - b)] \\ &+ S\{(\tilde{y}^k - y^k) + \mu[y^k - Y_k^2(\tilde{y}^k)^{-1}]\} + \xi_y^k = 0, \end{aligned} \tag{44}$$

where ξ_y^k satisfies the following inexactness criterion:

$$\|\xi_y^k\| \leq \nu s(1 - \mu)\|y^k - \tilde{y}^k\|. \tag{45}$$

Step 3. Update $\tilde{\lambda}^k$ via

$$\tilde{\lambda}^k := \lambda^k - H(A\tilde{x}^k + B\tilde{y}^k - b). \tag{46}$$

Step 4. Set $\tilde{w}^k := (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$, and then generate the new iterate w^{k+1} by

$$w^{k+1} := (1 - \sigma)w^k + \sigma P_{\mathcal{W},G}[w^k - \alpha_k d(w^k, \tilde{w}^k, \xi^k)] \tag{47}$$

with the step-size

$$\alpha_k := \gamma \alpha_k^*, \tag{48}$$

where

$$\begin{aligned} \alpha_k^* &:= \frac{\varphi(w^k, \tilde{w}^k, \xi^k)}{\|d(w^k, \tilde{w}^k, \xi^k)\|_G^2}, & \xi^k &:= \begin{pmatrix} \xi_x^k \\ \xi_y^k \\ 0 \end{pmatrix}, \\ d(w^k, \tilde{w}^k, \xi^k) &:= (w^k - \tilde{w}^k) - G^{-1}\xi^k, \\ \varphi(w^k, \tilde{w}^k, \xi^k) &:= (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k) + (w^k - \tilde{w}^k)^T Gd(w^k, \tilde{w}^k, \xi^k) \\ &\quad - \mu(\|x^k - \tilde{x}^k\|_R^2 + \|y^k - \tilde{y}^k\|_S^2), \end{aligned} \tag{49}$$

and G is defined by (12).

Step 5. If $\|w^{k+1} - w^k\| \leq \varepsilon$, stop; otherwise set $k = k + 1$ and goto Step 1.

Remark 5.1 Note that Lemma 3.1 guarantees that there exists a unique solution $\tilde{x}^k \in \mathbb{R}_{++}^n$ for a given $\xi_x^k \in \mathbb{R}^n$ and a unique solution $\tilde{y}^k \in \mathbb{R}_{++}^p$ for a given $\xi_y^k \in \mathbb{R}^p$. To implement Algorithm 2, it is easy to determine the error terms ξ_x^k and ξ_y^k subject to the inexactness criteria (43) and (45). For example, in practice within finite iterations, we can choose

$$\xi_x^k := f(x^k) - f(\tilde{x}^k) + A^T H A(x^k - \tilde{x}^k). \tag{51}$$

Then, the system of equations (42) reduces to

$$f(x^k) - A^T[\lambda^k - H(Ax^k + By^k - b)] + r\{(\tilde{x}^k - x^k) + \mu[x^k - X_k^2(\tilde{x}^k)^{-1}]\} = 0, \tag{52}$$

whose solution can be given explicitly by

$$(\tilde{x}^k)_i := \frac{q_i^k + \sqrt{(q_i^k)^2 + 4\mu(x_i^k)^2}}{2}, \quad i = 1, \dots, n$$

with

$$q^k := (1 - \mu)x^k - \frac{1}{r}\{f(x^k) - A^T[\lambda^k - H(Ax^k + By^k - b)]\}.$$

It is easy to verify that $\tilde{x}^k \in \mathbb{R}_{++}^n$ whenever $x^k \in \mathbb{R}_{++}^n$. Similarly, we can choose

$$\xi_y^k := g(y^k) - g(\tilde{y}^k) + B^T H B(y^k - \tilde{y}^k). \tag{53}$$

Then the positive solution of (44) can be obtained explicitly by

$$(\tilde{y}^k)_i := \frac{\tau_i^k + \sqrt{(\tau_i^k)^2 + 4\mu(y_i^k)^2}}{2}, \quad i = 1, \dots, p$$

with

$$\tau^k := (1 - \mu)y^k - \frac{1}{s}\{g(y^k) - B^T[\lambda^k - H(A\tilde{x}^k + By^k - b)]\}.$$

Furthermore, $\tilde{y}^k \in \mathbb{R}_{++}^p$ whenever $y^k \in \mathbb{R}_{++}^p$.

Remark 5.2 The inexactness criteria (43) and (45) can be met with proper values of r and s , respectively. For example, when both f and g in (2) are Lipschitz continuous, we can choose ξ_x^k as (51) and

$$r \geq \frac{L_f + \|A^T H A\|}{\nu(1 - \mu)}, \tag{54}$$

where L_f is f 's Lipschitz constant. Then, it follows that

$$\|\xi_x^k\| \stackrel{(51)}{\leq} (L_f + \|A^T H A\|)\|x^k - \tilde{x}^k\| \stackrel{(54)}{\leq} \nu r(1 - \mu)\|x^k - \tilde{x}^k\|,$$

which means that the inexactness criterion (43) is satisfied. Similarly, we can choose ξ_y^k as (53) and

$$s \geq \frac{L_g + \|B^T H B\|}{\nu(1 - \mu)}, \tag{55}$$

where L_g is g 's Lipschitz constant. Then, we have

$$\|\xi_y^k\| \stackrel{(53)}{\leq} (L_g + \|B^T H B\|)\|y^k - \tilde{y}^k\| \stackrel{(55)}{\leq} \nu s(1 - \mu)\|y^k - \tilde{y}^k\|,$$

and thus the inexactness criterion (45) is satisfied.

Remark 5.3 As we have mentioned, since the ADMM subproblems with LQP regularization are allowed to be solved subject to a constant accuracy, Algorithm 2 thus requires certain correction step to further correct the approximate solutions to ensure the convergence. More specifically, the additional correction required by Algorithm 2 in (47) can be specified as

$$\begin{cases} x^{k+1} := (1 - \sigma)x^k + \sigma P_{\mathbb{R}_+^n} \{x^k - \alpha_k[x^k - \tilde{x}^k - \frac{1}{(1+\mu)r} \xi_x^k]\}, \\ y^{k+1} := (1 - \sigma)y^k + \sigma P_{\mathbb{R}_+^p, M} \{y^k - \alpha_k[y^k - \tilde{y}^k - M^{-1} \xi_y^k]\}, \\ \lambda^{k+1} := (1 - \sigma)\lambda^k + \sigma[\lambda^k - \alpha_k(\lambda^k - \tilde{\lambda}^k)] = \lambda^k - \sigma \alpha_k(\lambda^k - \tilde{\lambda}^k), \end{cases}$$

where $M := (1 + \mu)S + B^T H B$. This additional computation is easy and inexpensive computationally.

5.2 Convergence

Recall we let $w^k := (x^k, y^k, \lambda^k) \in \mathbb{R}_{++}^n \times \mathbb{R}_{++}^p \times \mathbb{R}^m$ be a given vector, $\tilde{w}^k := (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \mathbb{R}_{++}^n \times \mathbb{R}_{++}^p \times \mathbb{R}^m$ be generated by Algorithm 2, w^k, \tilde{w}^k and ξ^k satisfy (43) and (45). In this subsection, we prove the convergence of Algorithm 2.

For convenience of the following analysis, we define

$$\bar{w}^k := \begin{pmatrix} \bar{x}^k \\ \bar{y}^k \\ \bar{\lambda}^k \end{pmatrix} = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \lambda^k - H(A\tilde{x}^k + B y^k - b) \end{pmatrix}. \tag{56}$$

Lemma 5.1 *Let the sequence $\{w^k\}$ be generated by Algorithm 2, and the accompanying sequence $\{\bar{w}^k\}$ be defined by (56). Then, for any $w := (x, y, \lambda) \in \mathcal{W}$, we have*

$$(w - \bar{w}^k)^T Q(\bar{w}^k) \geq \frac{1}{2\sigma\gamma\alpha_k^*} (\|w^{k+1} - w\|_G^2 - \|w^k - w\|_G^2) + \frac{2-\gamma}{2} \varphi(w^k, \tilde{w}^k, \xi^k), \tag{57}$$

where $\varphi(w^k, \tilde{w}^k, \xi^k)$ is defined by (50).

Proof Recall that $\bar{x}^k = \tilde{x}^k$ and $\bar{y}^k = \tilde{y}^k$. It follows from (42), (44) and (56) that

$$\begin{aligned} f(\bar{x}^k) - A^T \bar{\lambda}^k + \xi_x^k + R\{(\bar{x}^k - x^k) + \mu[x^k - X_k^2(\bar{x}^k)^{-1}]\} &= 0, \\ g(\bar{y}^k) - B^T \bar{\lambda}^k + B^T H B(\bar{y}^k - y^k) + \xi_y^k + S\{(\bar{y}^k - y^k) + \mu[y^k - Y_k^2(\bar{y}^k)^{-1}]\} &= 0. \end{aligned}$$

Applying Lemma 3.1 to the above equations, respectively, the inequality (16) can be written as

$$\begin{aligned} (x - \bar{x}^k)^T [f(\bar{x}^k) - A^T \bar{\lambda}^k + \xi_x^k] &\geq (1 + \mu)(x^k - \bar{x}^k)^T R(x - \bar{x}^k) - \mu \|x^k - \bar{x}^k\|_R^2, \quad \forall x \in \mathbb{R}_+^n, \\ (y - \bar{y}^k)^T [g(\bar{y}^k) - B^T \bar{\lambda}^k + B^T H B(\bar{y}^k - y^k) + \xi_y^k] &\geq (1 + \mu)(y^k - \bar{y}^k)^T S(y^k - \bar{y}^k) - \mu \|y^k - \bar{y}^k\|_S^2, \quad \forall y \in \mathbb{R}_+^p. \end{aligned}$$

Thus we have

$$\begin{aligned}
 &(x - \bar{x}^k)^T [f(\bar{x}^k) - A^T \bar{\lambda}^k] \\
 &\geq (x - \bar{x}^k)^T \{(1 + \mu)R(x^k - \bar{x}^k) - \xi_x^k\} - \mu \|x^k - \bar{x}^k\|_R^2, \quad \forall x \in \mathbb{R}_+^n, \\
 &(y - \bar{y}^k)^T [g(\bar{y}^k) - B^T \bar{\lambda}^k] \\
 &\geq (y - \bar{y}^k)^T \{(1 + \mu)S + B^T HB\}(y^k - \bar{y}^k) - \xi_y^k\} - \mu \|y^k - \bar{y}^k\|_S^2, \quad \forall y \in \mathbb{R}_+^p.
 \end{aligned}$$

Note that

$$(\lambda - \bar{\lambda}^k)^T (A\bar{x}^k + B\bar{y}^k - b) = (\lambda - \bar{\lambda}^k)^T H^{-1}(\lambda^k - \tilde{\lambda}^k), \quad \forall \lambda \in \mathbb{R}^m.$$

Combining the above three formulas together and using (49) and the notation of G and ξ^k , we get

$$\begin{aligned}
 (w - \bar{w}^k)^T Q(\bar{w}^k) &\geq (w - \bar{w}^k)^T Gd(w^k, \tilde{w}^k, \xi^k) - \mu(\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2) \\
 &= (w - w^k)^T Gd(w^k, \tilde{w}^k, \xi^k) + (w^k - \bar{w}^k)^T Gd(w^k, \tilde{w}^k, \xi^k) \\
 &\quad - \mu(\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2), \quad \forall w \in \mathcal{W}.
 \end{aligned} \tag{58}$$

For any $w \in \mathcal{W}$, we have

$$\begin{aligned}
 \|w^{k+1} - w\|_G^2 &= \|(1 - \sigma)w^k + \sigma P_{\mathcal{W},G}[w^k - \gamma\alpha_k^*d(w^k, \tilde{w}^k, \xi^k)] - w\|_G^2 \\
 &\leq [(1 - \sigma)\|w^k - w\|_G \\
 &\quad + \sigma\|P_{\mathcal{W},G}[w^k - \gamma\alpha_k^*d(w^k, \tilde{w}^k, \xi^k)] - w\|_G]^2 \\
 &\stackrel{(11)}{\leq} (1 - \sigma)\|w^k - w\|_G^2 + \sigma\|w^k - \gamma\alpha_k^*d(w^k, \tilde{w}^k, \xi^k) - w\|_G^2 \\
 &\stackrel{(49)}{=} \|w^k - w\|_G^2 - 2\sigma\gamma\alpha_k^*(w^k - w)^T Gd(w^k, \tilde{w}^k, \xi^k) \\
 &\quad + \sigma\gamma^2\alpha_k^*\varphi(w^k, \tilde{w}^k, \xi^k).
 \end{aligned} \tag{60}$$

It follows from (60) that

$$\begin{aligned}
 &(w - w^k)^T Gd(w^k, \tilde{w}^k, \xi^k) \\
 &\geq \frac{1}{2\sigma\gamma\alpha_k^*}(\|w^{k+1} - w\|_G^2 - \|w^k - w\|_G^2) - \frac{\gamma}{2}\varphi(w^k, \tilde{w}^k, \xi^k).
 \end{aligned} \tag{61}$$

Since $(\tilde{w}^k - \bar{w}^k)^T Gd(w^k, \tilde{w}^k, \xi^k) = (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\bar{y}^k)$ (see (12), (49) and (56)), it follows from the notation of $\varphi(w^k, \tilde{w}^k, \xi^k)$ that

$$\begin{aligned}
 &(w^k - \bar{w}^k)^T Gd(w^k, \tilde{w}^k, \xi^k) - \mu(\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2) \\
 &= (w^k - \tilde{w}^k)^T Gd(w^k, \tilde{w}^k, \xi^k) - \mu(\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2) \\
 &\quad + (\tilde{w}^k - \bar{w}^k)^T Gd(w^k, \tilde{w}^k, \xi^k)
 \end{aligned}$$

$$\begin{aligned}
 &= (w^k - \tilde{w}^k)^T Gd(w^k, \tilde{w}^k, \xi^k) - \mu(\|x^k - \tilde{x}^k\|_R^2 + \|y^k - \tilde{y}^k\|_S^2) \\
 &\quad + (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k) \\
 &= \varphi(w^k, \tilde{w}^k, \xi^k).
 \end{aligned} \tag{62}$$

Substituting (61) and (62) into (59), the assertion (57) is proved. □

Using the notation of M , the matrix G can be rewritten as

$$G := \begin{pmatrix} (1 + \mu)R & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & H^{-1} \end{pmatrix}.$$

Proposition 5.1 *For given $w^k := (x^k, y^k, \lambda^k) \in \mathbb{R}^{n_+} \times \mathbb{R}^{p_+} \times \mathbb{R}^m$, let $\tilde{w}^k := (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ be generated by (42)–(46). Then we have*

$$\varphi(w^k, \tilde{w}^k, \xi^k) \geq \frac{1}{4} \|d(w^k, \tilde{w}^k, \xi^k)\|_G^2. \tag{63}$$

Proof It follows from (49), (50) and the notation of G and x^k that

$$\begin{aligned}
 \varphi(w^k, \tilde{w}^k, \xi^k) &= (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k) + \|w^k - \tilde{w}^k\|_G^2 - (w^k - \tilde{w}^k)^T \xi^k \\
 &\quad - \mu(\|x^k - \tilde{x}^k\|_R^2 + \|y^k - \tilde{y}^k\|_S^2) \\
 &= (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k) + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 \\
 &\quad + \|By^k - B\tilde{y}^k\|_H^2 + \|x^k - \tilde{x}^k\|_R^2 \\
 &\quad - (x^k - \tilde{x}^k)^T \xi_x^k + \|y^k - \tilde{y}^k\|_S^2 - (y^k - \tilde{y}^k)^T \xi_y^k.
 \end{aligned} \tag{64}$$

Using $\bar{\lambda}^k = \tilde{\lambda}^k - H(By^k - B\tilde{y}^k)$ (see (46) and (56)), we have

$$\begin{aligned}
 &(\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k) + \frac{1}{2}(\|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 + \|By^k - B\tilde{y}^k\|_H^2) \\
 &= \frac{1}{2} \|\lambda^k - \tilde{\lambda}^k + H(By^k - B\tilde{y}^k)\|_{H^{-1}}^2 \\
 &= \frac{1}{2} \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2.
 \end{aligned}$$

Substituting this into (64), it follows that

$$\begin{aligned}
 \varphi(w^k, \tilde{w}^k, \xi^k) &= \frac{1}{2}(\|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 + \|By^k - B\tilde{y}^k\|_H^2) \\
 &\quad + \|x^k - \tilde{x}^k\|_R^2 - (x^k - \tilde{x}^k)^T \xi_x^k + \|y^k - \tilde{y}^k\|_S^2 - (y^k - \tilde{y}^k)^T \xi_y^k.
 \end{aligned} \tag{65}$$

Since $R := rI$, it follows from (43) that

$$\|R^{-1}\xi_x^k\|_R \leq \nu(1 - \mu)\|x^k - \tilde{x}^k\|_R. \tag{66}$$

Using the Cauchy–Schwarz inequality and (66), we have

$$\begin{aligned} \frac{1 - \mu}{2}\|x^k - \tilde{x}^k\|_R^2 - \frac{1}{2}(x^k - \tilde{x}^k)^T \xi_x^k &\geq \frac{1 - \mu}{2}\|x^k - \tilde{x}^k\|_R^2 - \frac{1}{2}\|x^k - \tilde{x}^k\|_R \|R^{-1}\xi_x^k\|_R \\ &\geq \frac{(1 - \mu)(1 - \nu)}{2}\|x^k - \tilde{x}^k\|_R^2 \geq 0. \end{aligned}$$

From the above inequality, we obtain

$$\|x^k - \tilde{x}^k\|_R^2 - (x^k - \tilde{x}^k)^T \xi_x^k \geq \frac{1}{2}[\|x^k - \tilde{x}^k\|_{(1+\mu)R}^2 - (x^k - \tilde{x}^k)^T \xi_x^k]. \tag{67}$$

Using (66), we have

$$\begin{aligned} \|[(1 + \mu)R]^{-1}\xi_x^k\|_{(1+\mu)R}^2 &= \frac{1}{(1 + \mu)^2}\|R^{-1}\xi_x^k\|_{(1+\mu)R}^2 \\ &\leq \frac{\nu^2(1 - \mu)^2}{(1 + \mu)^2}\|x^k - \tilde{x}^k\|_{(1+\mu)R}^2 \\ &\leq \|x^k - \tilde{x}^k\|_{(1+\mu)R}^2. \end{aligned}$$

It follows from the above inequality that

$$\begin{aligned} &\|x^k - \tilde{x}^k\|_{(1+\mu)R}^2 - (x^k - \tilde{x}^k)^T \xi_x^k \\ &\geq \frac{1}{2}\|x^k - \tilde{x}^k\|_{(1+\mu)R}^2 - (x^k - \tilde{x}^k)^T \xi_x^k + \frac{1}{2}\|[(1 + \mu)R]^{-1}\xi_x^k\|_{(1+\mu)R}^2 \\ &= \frac{1}{2}\|x^k - \tilde{x}^k - [(1 + \mu)R]^{-1}\xi_x^k\|_{(1+\mu)R}^2. \end{aligned}$$

Substituting this into (67), we obtain

$$\|x^k - \tilde{x}^k\|_R^2 - (x^k - \tilde{x}^k)^T \xi_x^k \geq \frac{1}{4}\|x^k - \tilde{x}^k - [(1 + \mu)R]^{-1}\xi_x^k\|_{(1+\mu)R}^2. \tag{68}$$

Similarly, using (45), we obtain

$$\|y^k - \tilde{y}^k\|_S^2 - (y^k - \tilde{y}^k)^T \xi_y^k \geq \frac{1}{4}\|y^k - \tilde{y}^k - [(1 + \mu)S]^{-1}\xi_y^k\|_{(1+\mu)S}^2.$$

Therefore, it follows from (65) and the above two inequalities that

$$\begin{aligned} \varphi(w^k, \tilde{w}^k, \xi^k) &\geq \frac{1}{4}(\|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 + \|By^k - B\tilde{y}^k\|_H^2) + \|x^k - \tilde{x}^k\|_R^2 - (x^k - \tilde{x}^k)^T \xi_x^k \\ &\quad + \|y^k - \tilde{y}^k\|_S^2 - (y^k - \tilde{y}^k)^T \xi_y^k \end{aligned}$$

$$\begin{aligned} &\geq \frac{1}{4} [\|x^k - \tilde{x}^k - [(1 + \mu)R]^{-1} \xi_x^k\|_{(1+\mu)R}^2 + \|By^k - B\tilde{y}^k\|_H^2 \\ &\quad + \|y^k - \tilde{y}^k - [(1 + \mu)S]^{-1} \xi_y^k\|_{(1+\mu)S}^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2]. \end{aligned} \tag{69}$$

Using the notation of M , we get

$$(\xi_y^k)^T [(1 + \mu)S]^{-1} \xi_y^k \geq (\xi_y^k)^T M^{-1} \xi_y^k,$$

and thus

$$\|[(1 + \mu)S]^{-1} \xi_y^k\|_{(1+\mu)S}^2 \geq \|M^{-1} \xi_y^k\|_M^2. \tag{70}$$

By a simple manipulation, we have

$$\begin{aligned} &\|By^k - B\tilde{y}^k\|_H^2 + \|y^k - \tilde{y}^k - [(1 + \mu)S]^{-1} \xi_y^k\|_{(1+\mu)S}^2 \\ &= \|y^k - \tilde{y}^k\|_{(1+\mu)S+B^T H B}^2 - 2(y^k - \tilde{y}^k)^T \xi_y^k + \|[(1 + \mu)S]^{-1} \xi_y^k\|_{(1+\mu)S}^2 \\ &\stackrel{(70)}{\geq} \|y^k - \tilde{y}^k\|_M^2 - 2(y^k - \tilde{y}^k)^T \xi_y^k + \|M^{-1} \xi_y^k\|_M^2 \\ &= \|y^k - \tilde{y}^k - M^{-1} \xi_y^k\|_M^2. \end{aligned} \tag{71}$$

Substituting (71) into (69) and using the notation of $d(w^k, \tilde{w}^k, \xi^k)$ and G , the assertion of this proposition is proved. \square

The following corollary follows from the definition of α_*^k and (63) directly, and we omit its proof.

Corollary 5.1 *The step-size α_k^* defined in Step 4 of Algorithm 2 satisfies $\alpha_k^* \geq 1/4$ for all $k \geq 0$.*

Next, we will show the contraction of the sequence generated by Algorithm 2, based on which the convergence of Algorithm 2 can be established easily.

Corollary 5.2 *For any $w^* \in \mathcal{W}^*$, there is a positive constant c_1 such that the sequence $\{w^k\}$ generated by Algorithm 2 satisfies*

$$\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 - c_1 (\|x^k - \tilde{x}^k\|_R^2 + \|y^k - \tilde{y}^k\|_S^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2). \tag{72}$$

Proof Since $w^* \in \mathcal{W}^*$, $\tilde{w}^k \in \mathcal{W}$ and Q is monotone, we have

$$0 \geq (w^* - \tilde{w}^k)^T Q(w^*) \geq (w^* - \tilde{w}^k)^T Q(\tilde{w}^k).$$

Then, setting $w = w^*$ in (57), we have

$$\begin{aligned} &0 \geq (w^* - \tilde{w}^k)^T Q(\tilde{w}^k) \\ &\geq \frac{1}{2\sigma\gamma\alpha_k^*} (\|w^{k+1} - w^*\|_G^2 - \|w^k - w^*\|_G^2) + \frac{2-\gamma}{2} \varphi(w^k, \tilde{w}^k, \xi^k). \end{aligned}$$

It follows from the above equation and Corollary 5.1 that

$$\begin{aligned} \|w^{k+1} - w^*\|_G^2 &\leq \|w^k - w^*\|_G^2 - \sigma\gamma(2 - \gamma)\alpha_k^* \varphi(w^k, \tilde{w}^k, \xi^k) \\ &\leq \|w^k - w^*\|_G^2 - \frac{\sigma\gamma(2 - \gamma)}{4} \varphi(w^k, \tilde{w}^k, \xi^k). \end{aligned} \tag{73}$$

Using (65), we have

$$\begin{aligned} \varphi(w^k, \tilde{w}^k, \xi^k) &\geq \frac{1}{2} \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2 + \|x^k - \bar{x}^k\|_R^2 - (x^k - \bar{x}^k)^T \xi_x^k \\ &\quad + \|y^k - \bar{y}^k\|_S^2 - (y^k - \bar{y}^k)^T \xi_y^k. \end{aligned}$$

It follows from (43) and (45) that

$$\begin{aligned} &\frac{1}{2} \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2 + \|x^k - \bar{x}^k\|_R^2 - (x^k - \bar{x}^k)^T \xi_x^k + \|y^k - \bar{y}^k\|_S^2 - (y^k - \bar{y}^k)^T \xi_y^k \\ &\geq \frac{1}{2} \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2 + \|x^k - \bar{x}^k\|_R^2 - \|x^k - \bar{x}^k\| \cdot \|\xi_x^k\| \\ &\quad + \|y^k - \bar{y}^k\|_S^2 - \|y^k - \bar{y}^k\| \cdot \|\xi_y^k\| \\ &\geq \frac{1}{2} \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2 + \|x^k - \bar{x}^k\|_R^2 - \nu(1 - \mu) \|x^k - \bar{x}^k\|_R^2 \\ &\quad + \|y^k - \bar{y}^k\|_S^2 - \nu(1 - \mu) \|y^k - \bar{y}^k\|_S^2 \\ &\geq \frac{(1 - \mu)(1 - \nu)}{2} (\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2 + \|\lambda^k - \bar{\lambda}^k\|_{H^{-1}}^2). \end{aligned}$$

Setting $c_1 := \sigma\gamma(2 - \gamma)(1 - \mu)(1 - \nu)/8$, the assertion (72) is proved from the above three inequalities. □

Theorem 5.1 *The sequence $\{w^k\}$ generated by Algorithm 2 converges to some w^∞ which is a solution of SVI(\mathcal{W}, Q).*

Proof Setting $\eta_k \equiv 0, c_0 = c_1$ in (20), from (72) and Theorem 3.1 we have

$$\lim_{k \rightarrow \infty} \|w^k - \bar{w}^k\|_G = 0, \tag{74}$$

and $\{w^k\}$ is bounded. It follows from (73) and (63) that

$$\lim_{k \rightarrow \infty} \|d(w^k, \tilde{w}^k, \xi^k)\|_G = 0. \tag{75}$$

From (58), we obtain

$$\begin{aligned} (w - \bar{w}^k)^T Q(\bar{w}^k) &\geq -\|w - \bar{w}^k\|_G \|d(w^k, \tilde{w}^k, \xi^k)\|_G \\ &\quad - \mu (\|x^k - \bar{x}^k\|_R^2 + \|y^k - \bar{y}^k\|_S^2), \quad \forall w \in \mathcal{W}. \end{aligned}$$

Together with (74) and (75), we get

$$\liminf_{k \rightarrow \infty} (w - \bar{w}^k)^T Q(\bar{w}^k) \geq 0, \quad \forall w \in \mathcal{W}.$$

The convergence of Algorithm 2 is then obtained immediately from Theorem 3.1. \square

5.3 Convergence Rate

Now, we show the worst-case $O(1/t)$ convergence rate for Algorithm 2.

Theorem 5.2 *For any integer $t > 0$, there is a $\bar{w}_t \in \mathcal{W}$, which is a convex combination of the iterates $\bar{w}^0, \bar{w}^1, \dots, \bar{w}^t$ defined in (56). For any $w \in \mathcal{W}$, we have*

$$(\bar{w}_t - w)^T Q(w) \leq \frac{1}{2\sigma\gamma\Upsilon_t} \|w^0 - w\|_G^2, \tag{76}$$

where

$$\Upsilon_t := \sum_{k=0}^t \alpha_k^* \quad \text{and} \quad \bar{w}_t := \frac{1}{\Upsilon_t} \sum_{k=0}^t \alpha_k^* \bar{w}^k.$$

Proof It follows from (57) and (63) that

$$(w - \bar{w}^k)^T \alpha_k^* Q(\bar{w}^k) + \frac{1}{2\sigma\gamma} \|w^k - w\|_G^2 \geq \frac{1}{2\sigma\gamma} \|w^{k+1} - w\|_G^2, \quad \forall w \in \mathcal{W}.$$

Since Q is monotone, from the above inequality, we have

$$(w - \bar{w}^k)^T \alpha_k^* Q(w) + \frac{1}{2\sigma\gamma} \|w^k - w\|_G^2 \geq \frac{1}{2\sigma\gamma} \|w^{k+1} - w\|_G^2, \quad \forall w \in \mathcal{W}. \tag{77}$$

Summing the inequality (77) over $k = 0, 1, \dots, t$, we obtain

$$\begin{aligned} \left[\left(\sum_{k=0}^t \alpha_k^* \right) w - \left(\sum_{k=0}^t \alpha_k^* \bar{w}^k \right) \right]^T Q(w) + \frac{1}{2\sigma\gamma} \|w^0 - w\|_G^2 &\geq \frac{1}{2\sigma\gamma} \|w^{t+1} - w\|_G^2 \\ &\geq 0, \quad \forall w \in \mathcal{W}. \end{aligned}$$

Since $(\sum_{k=0}^t \alpha_k^*)/\Upsilon_t = 1$, \bar{w}_t is a convex combination of $\bar{w}^0, \bar{w}^1, \dots, \bar{w}^t$ and thus $\bar{w}_t \in \mathcal{W}$. Using the notation of Υ_t and \bar{w}_t , we derive

$$(w - \bar{w}_t)^T Q(w) + \frac{1}{2\sigma\gamma\Upsilon_t} \|w^0 - w\|_G^2 \geq 0, \quad \forall w \in \mathcal{W}.$$

The assertion (76) follows from the above inequality immediately. \square

Using Theorem 5.2, we can prove a worst-case $O(1/t)$ convergence rate of Algorithm 2 in ergodic sense as Algorithm 1.

6 Conclusions

This paper studies how to combine the alternating direction method of multipliers and the logarithmic-quadratic proximal method, allowing the resulting subproblems to be solved approximately subject to different inexactness criteria. We demonstrate how to develop inexact versions of ADMM with LQP regularization by applying the most popular and fundamental choices of inexactness criteria in the literature. It deserves further research on embedding other inexactness criteria into the subproblems of ADMM with LQP regularization and thus developing other algorithms of the same kind as those in this paper. We expect that certain correction steps are required if some relaxed inexactness criteria are chosen for solving the resulting subproblems of the combination of ADMM and LQP method.

Acknowledgements The first author was supported by National Natural Science Foundation of China grant 11001053, Program for New Century Excellent Talents in University grant NCET-12-0111 and Natural Science Foundation of Jiangsu Province grant BK2012662. The second author was supported in part by Hong Kong General Research Fund grants HKBU 201409 and HKBU 201611. The third author was supported by the grant FRG2/11-12/120 from Hong Kong Baptist University and the General Research Fund HKBU 203311 from Hong Kong Research Grants Council.

References

1. Glowinski, R., Marrocco, A.: Sur l'approximation par éléments finis d'ordre un et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *Rev. Fr. Autom. Inform. Rech. Opér., Anal. Numér.* **2**, 41–76 (1975)
2. He, B.S., Liao, L.-Z., Han, D.R., Yang, H.: A new inexact alternating directions method for monotone variational inequalities. *Math. Program.* **92**(1), 103–118 (2002)
3. Martinet, B.: Regularisation d'inéquations variationnelles par approximations successives. *Rev. Fr. Autom. Inform. Rech. Opér.* **126**, 154–159 (1970)
4. Auslender, A., Teboulle, M.: Entropic proximal decomposition method for convex programs and variational inequalities. *Math. Program.* **91**(1), 33–47 (2001)
5. Yuan, X.M., Li, M.: An LQP-based decomposition method for solving a class of variational inequalities. *SIAM J. Optim.* **21**(4), 1309–1318 (2011)
6. Auslender, A., Teboulle, M., Ben-Tiba, S.: A logarithmic-quadratic proximal method for variational inequalities. *Comput. Optim. Appl.* **12**, 31–40 (1999)
7. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011)
8. Tao, M., Yuan, X.M.: On the $O(1/t)$ convergence rate of alternating direction method with logarithmic-quadratic proximal regularization. *SIAM J. Optim.* **22**(4), 1431–1448 (2012)
9. Facchinei, F., Pang, J.S.: *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vols. I and II. Springer, New York (2003)
10. He, B.S., Yuan, X.M.: On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method. *SIAM J. Numer. Anal.* **50**(2), 700–709 (2012)
11. Nesterov, Y.: Gradient methods for minimizing composite objective function pp. 1–30. *Core Discussion Paper, 2007/76* (2007)
12. He, B.S., Liao, L.-Z., Yuan, X.M.: A LQP-based interior prediction-correction method for nonlinear complementarity problems. *J. Comput. Math.* **24**(1), 33–44 (2006)
13. Yamashita, N., Fukushima, M.: Equivalent unconstrained minimization and global error bounds for variational inequality problems. *SIAM J. Control Optim.* **35**(1), 273–284 (1997)