Local Linear Convergence of the Alternating Direction Method of Multipliers for Quadratic Programs

Han, Deren; Yuan, Xiaoming

Published in:
SIAM Journal on Numerical Analysis

DOI:
10.1137/120886753

Published: 17/12/2013

Document Version:
Publisher's PDF, also known as Version of record

SIAM J. NUMER. ANAL.
Vol. 51, No. 6, pp. 3446–3457

LOCAL LINEAR CONVERGENCE OF THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS FOR QUADRATIC PROGRAMS*

DEREN HAN† AND XIAOMING YUAN‡

Abstract. The Douglas–Rachford alternating direction method of multipliers (ADMM) has been widely used in various areas. The global convergence of ADMM is well known, while research on its convergence rate is still in its infancy. In this paper, we show the local linear convergence rate of ADMM for a quadratic program which includes some important applications of ADMM as special cases.

Key words. alternating direction method of multipliers, linear convergence rate, quadratic program, error bound

AMS subject classifications. 90C25, 65K10, 65J22

DOI. 10.1137/120886753

1. Introduction. Since the seminal work [7, 10], the Douglas–Rachford alternating direction method of multipliers (ADMM) has been widely used in various areas such as partial differential equations, image processing, and statistical learning. In [2], ADMM was commented on as "at least comparable to very specialized algorithms (even in the serial setting), and in most cases, the simple ADM algorithm will be efficient enough to be useful." In this paper we restrict our discussion to the context of the convex programming problem, and the canonical targeted model of ADMM is

$$\min\,\{f(x)+g(y)\},\ \text{subject to (s.t.)}\ Ax+By=b,\ x\in\mathcal{X},\ y\in\mathcal{Y}, \tag{1.1}$$

where $f: \mathcal{X}\to\mathcal{R}$ and $g: \mathcal{Y}\to\mathcal{R}$ are proper convex functions, $\mathcal{X}\subseteq\mathcal{R}^n$ and $\mathcal{Y}\subseteq\mathcal{R}^m$ are closed and convex sets, $A\in\mathcal{R}^{l\times n}$ and $B\in\mathcal{R}^{l\times m}$ are matrices, and $b\in\mathcal{R}^l$ is a given vector. Some standard assumptions when discussing ADMM include the nonemptyness of the solution set of (1.1) and the full column rank of B (e.g., $B=-I$).

The iterative scheme of ADMM for solving (1.1) reads as

$$\begin{cases} x^{k+1}=\arg\min\limits_{x\in\mathcal{X}}\left\{f(x)-x^TA^T\lambda^k+\dfrac{\beta}{2}\|Ax+By^k-b\|^2\right\}, & \text{(1.2a)}\\ y^{k+1}=\arg\min\limits_{y\in\mathcal{Y}}\left\{g(y)-y^TB^T\lambda^k+\dfrac{\beta}{2}\|Ax^{k+1}+By-b\|^2\right\}, & \text{(1.2b)}\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b), & \text{(1.2c)}\end{cases}$$

where λ is the Lagrange multiplier and $\beta>0$ is a penalty parameter. In [8], the scheme (1.2a)–(1.2c) was illustrated as an application of the Douglas–Rachford splitting method (DRSM) in [17] to the dual of (1.1); and in [5] DRSM was shown to be a

*Received by the editors August 2, 2012; accepted for publication (in revised form) September 24, 2013; published electronically December 17, 2013.
http://www.siam.org/journals/sinum/51-6/88675.html

†School of Mathematical Sciences, Key Laboratory for NSLSCS of Jiangsu Province, Nanjing Normal University, Nanjing 210046, People's Republic of China (handeren@njnu.edu.cn). This author was supported by NSFC grants 11071122 and 11171159.

‡Department of Mathematics, Institute of Computational and Theoretical Studies, Hong Kong Baptist University, Hong Kong, People's Republic of China (xmyuan@hkbu.edu.hk). This author was supported by the General Research Fund from Hong Kong Research Grants Council: 203613.

special case of the proximal point algorithm in [19]. Therefore, inspired by the work [12], the so-called generalized ADMM was proposed in [5] and its iterative scheme is

$$
\begin{cases}
x^{k+1} = \arg\min_{x\in\mathcal{X}} \left\{ f(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \right\}, & (1.3a)\\
y^{k+1} = \arg\min_{y\in\mathcal{Y}} \Big\{ g(y) - y^T B^T \lambda^k & (1.3b)\\
\qquad\qquad + \frac{\beta}{2}\|\alpha Ax^{k+1} - (1-\alpha)(By^k - b) + By - b\|^2 \Big\}, & \\
\lambda^{k+1} = \lambda^k - \beta\left(\alpha Ax^{k+1} - (1-\alpha)(By^k - b) + By^{k+1} - b\right), & (1.3c)
\end{cases}
$$

where the parameter $\alpha \in (0,2)$ is an acceleration factor and it is usually suggested to take $\alpha \in (1,2)$ (see, e.g., [5]). Notice the original ADMM scheme (1.2a)–(1.2c) is a special case of the generalized ADMM scheme (1.3a)–(1.3c) with $\alpha = 1$. Our discussion of ADMM is thus focused on the scheme (1.3a)–(1.3c) from now on.

The convergence of ADMM is of great interest in the literature. Under some mild conditions such as the nonemptyness of the solution set of (1.1), ADMM was proven to be globally convergent; see, e.g., [7, 8, 10]. Despite the well-known global convergence, research on ADMM's convergence rate is still in its infancy; see, e.g., [9, 11, 17]. For the generic model (1.1), it seems that the ADMM's convergence rate is at most sublinear; see [15] for a worst-case $O(1/k)$ iteration complexity of ADMM in an ergodic sense. This work shows that ADMM achieves an approximate solution of (1.1) with an accuracy of $O(1/k)$ after k iterations. See also [16] for an extended analysis to DRSM. On the other hand, for some special cases of (1.1) where f and g have particular properties, it is possible for ADMM to achieve faster convergence. A remarkable result is [4], where the global linear convergence rate was established when ADMM was applied to solve a standard linear programming model (which is of course of a very special case of (1.1)). We also refer to [1, 3, 13] for some of the most recent discussion on the linear convergence rate of ADMM.

In this paper, we aim at showing the local linear convergence rate of ADMM for solving the quadratic program

$$
\min \left\{ \frac{1}{2}x^T Q x + q^T x + \frac{1}{2}y^T R y + r^T y \right\}, \text{ s.t. } Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}, \tag{1.4}
$$

where Q and R are symmetric positive semidefinite matrices in $\mathcal{R}^{n\times n}$ and $\mathcal{R}^{m\times m}$, respectively; $A \in \mathcal{R}^{l\times n}$ and $B \in \mathcal{R}^{l\times m}$ are two given matrices; $q \in \mathcal{R}^n$, $r \in \mathcal{R}^m$, and $b \in \mathcal{R}^l$ are given vectors; $\mathcal{X} = \{x | Cx = c,\ x \in \mathcal{R}^n (\text{or } \mathcal{R}^n_+)\}$ and $\mathcal{Y} = \{y | Dy = d,\ y \in \mathcal{R}^m (\text{or } \mathcal{R}^m_+)\}$ are two polyhedral sets. As a special case of the generic model (1.1), (1.4) captures a number of important applications arising in various areas. We here only mention a few.

- Example 1 is the l_1-norm regularized least-squares model

$$
\min_{x\in\mathcal{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \nu\|x\|_1, \tag{1.5}
$$

  where $A \in \mathcal{R}^{l\times n}$ is a given matrix, $b \in \mathcal{R}^l$ is a given vector, ν is a positive scalar, and $\|x\|_1 := \sum_{i=1}^n |x_i|$. For a number of applications, in (1.5) the least-squares term reflects the fidelity of observed or fitting data, the l_1-norm term aims at inducing a sparse solution, and the parameter ν is a trade-off between these two purposes. Popular applications of (1.5) include the well-known LASSO model in statistics and the basis pursuit model in compressive sensing.

- Example 2 is the total variation image restoration model

$$\min \frac{1}{2}\|Gx - x^0\|_2^2 + \nu\||\nabla x|\|_1, \tag{1.6}$$

where $x^0 \in \mathcal{R}^n$ is the vector representation of an observed image corrupted by noise; $x \in \mathcal{R}^n$ is the image to be restored; $G : \mathcal{R}^n \to \mathcal{R}^n$ is a linear transform such as a convolution operator; ∇ is the matrix representation of the anisotropic total variation operator proposed in [20]; and $\nu > 0$ is a trade-off parameter. Image restoration models regularized by the total variation are particularly effective for restoring piecewise constant images.

- Example 3 is the standard quadratic programming problem

$$\begin{array}{ll} \min & \frac{1}{2}x^T Qx + q^T x \\ \text{s.t.} & Cx = c, \\ & x \geq 0, \end{array} \tag{1.7}$$

where the setting of Q, q, C, c, and x is the same as in (1.4).

The remaining part is organized as follows. In section 2, we summarize some necessary preliminaries for further analysis. We then prove the local linear convergence of ADMM for (1.4) in section 3, and state some conclusions in section 4.

2. Preliminaries.

In this section, we summarize some useful preliminaries for further analysis.

2.1. Notation set.

Throughout the paper, all vectors are column vectors. For any two vectors $x \in \mathcal{R}^n$ and $y \in \mathcal{R}^m$, we simply use $u = (x, y)$ to denote their adjunction, i.e., (x, y) denotes $(x^T, y^T)^T$. We use $\|x\|_p$ to denote its p-norm, where $p = 1$ or 2; and for $p = 2$, we simply denote it as $\|x\|$. For any symmetric and positive definite matrix M, we denote $\|x\|_M := \sqrt{x^T Mx}$ as its M-norm. For a given matrix A, its norm is

$$\|A\|_p := \sup_{x\neq 0}\left\{\frac{\|Ax\|_p}{\|x\|_p}\right\}.$$

Especially, for a symmetric matrix A, $\|A\|_2$ denotes its spectral norm.

2.2. Variational inequalities.

The first-order optimality condition of (1.4) can be expressed as a variational inequality (VI), and our analysis is based on the VI characterization of (1.4). Thus, some preliminaries of VI are necessary. Let $F : \mathcal{R}^n \mapsto \mathcal{R}^n$ be a monotone operator and $\mathcal{U} \subseteq \mathcal{R}^n$ be a nonempty, closed, and convex set. The VI problem denoted by VIP$(\mathcal{U}, F)$ is to find $u^* \in \mathcal{U}$ such that

$$(u - u^*)^T F(u^*) \geq 0 \quad \forall u \in \mathcal{U}. \tag{2.1}$$

The VI problem (2.1) is equivalent to solving a projection equation. More specifically, let the projection operator under the Euclidean norm $P_{\mathcal{U}}(\cdot) : \mathcal{R}^n \mapsto \mathcal{U}$ be

$$P_{\mathcal{U}}(u) = \arg\min_{v\in\mathcal{U}}\{\|v - u\|\}\ \forall u \in \mathcal{R}^n. \tag{2.2}$$

Then, we have the following lemma whose proof can be found in the literature (e.g., [6]).

LEMMA 2.1. *Solving VIP($\mathcal{U}, F$) is equivalent to finding a zero point of*

$$e(u, \gamma) := u - P_{\mathcal{U}}(u - \gamma F(u)), \tag{2.3}$$

where $\gamma > 0$ is an arbitrary but fixed parameter.

Thus, for a given $u \in \mathcal{U}$, the magnitude $\|e(u, \gamma)\|$ is usually used as a measurement for the solution set of VIP($\mathcal{U}, F$). Some useful properties regarding this measurement can be summarized in the following lemma. Since the proof is standard (see, e.g., [6]), we omit it.

LEMMA 2.2. *Suppose that $u \in \mathcal{U}$ is not a solution of VIP($\mathcal{U}, F$) and $\tilde{\gamma} \geq \gamma > 0$, then we have*

$$\|e(u, \tilde{\gamma})\| \geq \|e(u, \gamma)\|$$

and

$$\frac{\|e(u, \tilde{\gamma})\|}{\tilde{\gamma}} \leq \frac{\|e(u, \gamma)\|}{\gamma}. \qquad \square$$

2.3. An error bound of the VI reformulation of (1.4). It is easy to see that the model (1.4) can be characterized by the VI

$$(u - u^*)^T (Lu^* + s) \geq 0 \; \forall u \in \mathcal{U}, \tag{2.4}$$

where

$$u := \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \; L := \begin{pmatrix} Q & 0 & -A^T \\ 0 & R & -B^T \\ A & B & 0 \end{pmatrix}, \; s := \begin{pmatrix} q \\ d \\ -b \end{pmatrix}, \text{ and } \mathcal{U} := \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^l. \tag{2.5}$$

Note that (2.4) is an affine VI since the involved mapping is affine, and the set $\mathcal{U}$ is polyhedral under the setting of (1.4). We use $\mathcal{U}^*$ to denote the solution set of (2.4). Then, according to [6], $\mathcal{U}^*$ is a closed and convex set.

Using Lemma 2.1, we can further characterize the VI characterization (2.4) of (1.4) as solving the equation

$$\|e(u, \gamma)\| = 0,$$

where

$$e(u, \gamma) := \begin{pmatrix} e_{\mathcal{X}}(u, \gamma) := x - P_{\mathcal{X}}[x - \gamma(Qx + q - A^T\lambda)] \\ e_{\mathcal{Y}}(u, \gamma) := y - P_{\mathcal{Y}}[y - \gamma(Ry + d - B^T\lambda)] \\ e_{\Lambda}(u, \gamma) := \gamma(Ax + By - b) \end{pmatrix}. \tag{2.6}$$

Using the error bound established in [18], we can measure the proximity of an arbitrary $u \in \mathcal{U}$ to $\mathcal{U}^*$. This result is summarized in the following lemma.

LEMMA 2.3. *Let $\mathcal{U}^*$ be the solution set of* (2.4) *and M be a positive definite matrix in appropriate dimensionality. For any $u \in \mathcal{U}$, let $dist_M(u, \mathcal{U}^*)$ denote the distance function from u to $\mathcal{U}^*$, i.e.,*

$$dist_M(u, \mathcal{U}^*) = \min\{\|u - u^*\|_M \mid u^* \in \mathcal{U}^*\}.$$

There exist scalars $\epsilon > 0$ and $\tau > 0$ such that

$$dist_M(u, \mathcal{U}^*) \leq \tau \|e(u, 1)\|_M \tag{2.7}$$

whenever $\|e(u, 1)\|_M \leq \epsilon$.

Remark 2.1. Note that Lemma 2.2 implies that for any fixed $\gamma > 0$,

$$\|e(u,1)\|_M \le \max\{\gamma, 1/\gamma\}\|e(u,\gamma)\|_M.$$

Thus, the lemma holds for any fixed $\gamma > 0$, i.e.,

$$\mathrm{dist}_M(u, \mathcal{U}^*) \le \tau\|e(u,1)\|_M,$$

and thus we take $\gamma \equiv 1$ in the following analysis for simplicity. □

2.4. Further notation. Let $\{u^k = (x^k, y^k, \lambda^k)\}$ be the sequence generated by the ADMM scheme (1.3a)–(1.3c). Then, according to the global convergence of ADMM and the characterization in Lemma 2.1, we have $\|e(u^k,1)\| \to 0$. Recall the error bound stated in Lemma 2.3. It is thus reasonable to establish the local linear convergence of ADMM in the sense that there exists a positive constant $0 < \varrho < 1$ such that

$$\mathrm{dist}_M^2(u^{k+1}, \mathcal{U}^*) \le \varrho \cdot \mathrm{dist}_M^2(u^k, \mathcal{U}^*)$$

whenever the iterative u^k is close enough to $\mathcal{U}^*$ such that $\|e(u^k,1)\| \le \epsilon$ is satisfied.

In the following analysis, our conclusions are presented in the absence of $\{x^k\}$. This is because the variable x is only an intermediate variable and it is involved in the iteration of ADMM. See more details in [2]. Accordingly, the following notation regarding the variables y and λ will simplify our analysis:

$$v = (y, \lambda); \;\; \mathcal{V} = \mathcal{Y} \times \mathcal{R}^l; \; \Omega^* = \{(y^*, \lambda^*) \,|\, (x^*, y^*, \lambda^*) \in \mathcal{U}^*\},$$
$$v^k = (y^k, \lambda^k) \; \forall k \in \mathcal{N}.$$

We thus can establish the local linear convergence of ADMM for (1.4) if there exist a positive constant $0 < \varrho < 1$ and a positive definite matrix G such that

$$\mathrm{dist}_G^2(v^{k+1}, \Omega^*) \le \varrho \cdot \mathrm{dist}_G^2(v^k, \Omega^*) \tag{2.8}$$

whenever the iterative v^k is sufficiently close to Ω^*.

We also define some matrices which will be used to simplify our notation in the following analysis. First, let

$$H_\alpha := \begin{pmatrix} \frac{\beta}{\alpha}B^TB & \frac{1-\alpha}{\alpha}B^T \\ \frac{1-\alpha}{\alpha}B & \frac{1}{\alpha\beta}I_m \end{pmatrix}. \tag{2.9}$$

In particular, when $\alpha = 1$, we have

$$H_1 := \begin{pmatrix} \beta B^TB & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix}. \tag{2.10}$$

Moreover, let

$$G_1 := \begin{pmatrix} \beta^2 B^TB & \beta B^T \\ \beta B & I_m \end{pmatrix} \tag{2.11}$$

and

$$G_2 := \begin{pmatrix} B^T(\beta^2AA^T + (1-\alpha)^2 I_l)B & (1-\alpha)(\beta B^TA + \frac{1}{\beta}B^T) \\ (1-\alpha)(\beta A^TB + \frac{1}{\beta}B) & (\frac{1}{\beta^2} + (1-\alpha)^2)I_l \end{pmatrix}. \tag{2.12}$$

Note the positive definiteness of H_α and positive semidefinitness of G_2 defined above are guaranteed when $\alpha \in (0,2)$ and B is assumed to be full column rank. In our analysis, we consider the case where $G = H_\alpha$ in (2.8) and $M = \begin{pmatrix} I_n & 0 \\ 0 & G \end{pmatrix}$. That is, in the next section we shall prove

$$\text{dist}^2_{H_\alpha}(v^{k+1}, \Omega^*) \le \varrho \cdot \text{dist}^2_{H_\alpha}(v^k, \Omega^*)$$

for some $\varrho \in (0,1)$ whenever v^k is sufficiently close to Ω^*.

In the following, for a matrix, its minimal and maximal eigenvalues are denoted by $\lambda_{\min}$ and $\lambda_{\max}$, respectively.

3. Local linear convergence of ADMM for (1.4). In this section, we establish the local linear convergence of ADMM for solving (1.4). We first compare the distance to an arbitrary point in Ω^* for two consecutive iterates of the sequence $\{v^k\}$ in Theorem 3.1. Before that, we prove a useful inequality in the following lemma.

LEMMA 3.1. *Let $\{u^k\}$ be the sequence generated by the ADMM scheme* (1.3a)–(1.3c) *and the sequence $\{\bar\lambda^k\}$ be defined as*

$$\bar\lambda^k := \lambda^k - \beta(Ax^{k+1} + By^k - b). \tag{3.1}$$

Let H_α be defined in (2.9). *Then, we have*

$$\begin{pmatrix} y^{k+1} - y^* \\ \bar\lambda^k - \lambda^* \end{pmatrix}^T H_\alpha \begin{pmatrix} y^k - y^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix} \ge 0 \quad \forall (y^*, \lambda^*) \in \Omega^*. \tag{3.2}$$

Proof. Invoking the first-order optimality condition for (1.3a) and (1.3c), respectively, we have

$$(x' - x^{k+1})^T (Qx^{k+1} + q - A^T\lambda^k + \beta A^T(Ax^{k+1} + By^k - b)) \ge 0 \; \forall x' \in \mathcal{X}, \tag{3.3}$$

and

$$(y' - y^{k+1})^T (Ry^{k+1} + r - B^T\lambda^k + \beta B^T(By^{k+1} - By^k) - (1-\alpha)(\bar\lambda^k - \lambda^k)) \ge 0 \; \forall y' \in \mathcal{Y}. \tag{3.4}$$

Recall the definition (3.1). It is easy to see that

$$\lambda^{k+1} = \bar\lambda^k - \beta(By^{k+1} - By^k) - (1-\alpha)(\bar\lambda^k - \lambda^k). \tag{3.5}$$

Setting $x' := x^*$ in (3.3) and $y' := y^*$ in (3.4) and using (3.1)–(3.5), we obtain

$$(x^* - x^{k+1})^T (Qx^{k+1} + q - A^T\bar\lambda^k) \ge 0 \tag{3.6}$$

and

$$(y^* - y^{k+1})^T (Ry^{k+1} + r - B^T\lambda^{k+1}) \ge 0. \tag{3.7}$$

Recall that

$$\begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}^T \begin{pmatrix} Qx^* + q - A^T\lambda^* \\ Ry^* + r - B^T\lambda^* \end{pmatrix} \ge 0 \; \forall x \in \mathcal{X}, \; \forall y \in Y. \tag{3.8}$$

Setting $x := x^{k+1}$ and $y := y^{k+1}$ in (3.8), and adding the resulting inequality to (3.6)–(3.7),

$$\begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} (Qx^* - Qx^{k+1}) - A^T(\lambda^* - \bar\lambda^k) \\ (Ry^* - Ry^{k+1}) - B^T(\lambda^* - \lambda^{k+1}) \end{pmatrix} \ge 0.$$

From the positive semidefiniteness of Q and R, the above inequality yields

$$\begin{pmatrix} x^{k+1}-x^* \\ y^{k+1}-y^* \end{pmatrix}^T \begin{pmatrix} A^T(\bar{\lambda}^k-\lambda^*) \\ B^T(\lambda^{k+1}-\lambda^*) \end{pmatrix} \geq 0.$$

It then follows from the relation (3.5) and $Ax^*+By^*=b$ that

$$(3.9)\quad (\bar{\lambda}^k-\lambda^*)^T(Ax^{k+1}+By^{k+1}-b) \\ -(By^{k+1}-By^*)^T(\beta(By^{k+1}-By^k)+(1-\alpha)(\bar{\lambda}^k-\lambda^k)) \geq 0.$$

Since

$$(3.10)\qquad (1-\alpha)(By^{k+1}-By^k)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k) = -\alpha(Ax^{k+1}+By^{k+1}-b),$$

the assertion (3.2) is trivial by using (2.9) and (3.9). ▯

With Lemma 3.1, we now can analyze the difference of the distance to Ω^* for two consecutive iterates of the sequence $\{v^k\}$.

THEOREM 3.1. *The sequence* $\{u^k\}$ *generated by the ADMM scheme* (1.3a)–(1.3c) *satisfies*

$$(3.11)\quad \left\| \begin{matrix} y^{k+1}-y^* \\ \lambda^{k+1}-\lambda^* \end{matrix} \right\|_{H_\alpha}^2 \leq \left\| \begin{matrix} y^k-y^* \\ \lambda^k-\lambda^* \end{matrix} \right\|_{H_\alpha}^2 - \frac{2-\alpha}{\alpha^2\beta} \left\| \begin{matrix} y^k-y^{k+1} \\ \lambda^k-\lambda^{k+1} \end{matrix} \right\|_{H_1}^2 \quad \forall (y^*,\lambda^*)\in\Omega^*,$$

where H_α *and* H_1 *are defined in* (2.9) *and* (2.10), *respectively.*

Proof. Recall that for any two vectors a and b with the same dimensions, we have the identity

$$(3.12)\qquad \|a-b\|_{H_\mu}^2 = \|a\|_{H_\mu}^2 - \|b\|_{H_\mu}^2 - 2(a-b)^T H_\mu b.$$

Setting

$$a := \begin{pmatrix} y^k-y \\ \lambda^k-\lambda \end{pmatrix} \quad\text{and}\quad b := \begin{pmatrix} y^k-y^{k+1} \\ \lambda^k-\lambda^{k+1} \end{pmatrix},$$

we have

$$\left\| \begin{matrix} y^{k+1}-y^* \\ \lambda^{k+1}-\lambda^* \end{matrix} \right\|_{H_\alpha}^2 = \left\| \begin{matrix} y^k-y^* \\ \lambda^k-\lambda^* \end{matrix} \right\|_{H_\alpha}^2 \\ - \left\| \begin{matrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{matrix} \right\|_{H_\alpha}^2 - 2\begin{pmatrix} y^{k+1}-y^* \\ \lambda^{k+1}-\lambda^* \end{pmatrix}^T H_\alpha \begin{pmatrix} y^k-y^{k+1} \\ \lambda^k-\lambda^{k+1} \end{pmatrix}.$$

Then, setting

$$a := \begin{pmatrix} y^{k+1}-y^k \\ \lambda^k-\lambda^{k+\frac{1}{2}} \end{pmatrix} \quad\text{and}\quad b := \begin{pmatrix} 0 \\ \lambda^{k+1}-\lambda^{k+\frac{1}{2}} \end{pmatrix}$$

yields

$$\left\| \begin{matrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{matrix} \right\|_{H_\alpha}^2 = \left\| \begin{matrix} y^{k+1}-y^k \\ \lambda^k-\bar{\lambda}^k \end{matrix} \right\|_{H_\alpha}^2 - \left\| \begin{matrix} 0 \\ \lambda^{k+1}-\bar{\lambda}^k \end{matrix} \right\|_{H_\alpha}^2 \\ -2\begin{pmatrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix}^T H_\alpha \begin{pmatrix} 0 \\ \lambda^{k+1}-\bar{\lambda}^k \end{pmatrix}.$$

Adding the above two inequalities,

$$\left\| \begin{array}{c} y^{k+1} - y^* \\ \lambda^{k+1} - \lambda^* \end{array} \right\|_{H_\alpha}^2 = \left\| \begin{array}{c} y^{k} - y^* \\ \lambda^{k} - \lambda^* \end{array} \right\|_{H_\alpha}^2 - \left\| \begin{array}{c} y^{k+1} - y^k \\ \lambda^{k} - \bar\lambda^k \end{array} \right\|_{H_\alpha}^2 + \left\| \begin{array}{c} 0 \\ \lambda^{k+1} - \bar\lambda^k \end{array} \right\|_{H_\alpha}^2$$
$$- 2 \begin{pmatrix} y^{k+1} - y^* \\ \bar\lambda^{k} - \lambda^* \end{pmatrix}^T H_\alpha \begin{pmatrix} y^{k} - y^{k+1} \\ \lambda^{k} - \lambda^{k+1} \end{pmatrix}$$
$$\leq \left\| \begin{array}{c} y^{k} - y^* \\ \lambda^{k} - \lambda^* \end{array} \right\|_{H_\alpha}^2 - \left\| \begin{array}{c} y^{k+1} - y^k \\ \lambda^{k} - \bar\lambda^k \end{array} \right\|_{H_\alpha}^2 + \left\| \begin{array}{c} 0 \\ \lambda^{k+1} - \bar\lambda^k \end{array} \right\|_{H_\alpha}^2, \tag{3.13}$$

where the inequality follows from (3.2). Using (3.1) and (1.3c), it is easy to check that

$$\left\| \begin{array}{c} y^{k+1} - y^k \\ \lambda^{k} - \bar\lambda^k \end{array} \right\|_{H_\alpha}^2 - \left\| \begin{array}{c} 0 \\ \lambda^{k+1} - \bar\lambda^k \end{array} \right\|_{H_\alpha}^2 = \frac{2-\alpha}{\alpha^2\beta} \|\beta B(y^k - y^{k+1}) + (\lambda^k - \lambda^{k+1})\|^2$$
$$= \frac{2-\alpha}{\alpha^2\beta} \left\| \begin{array}{c} y^{k} - y^{k+1} \\ \lambda^{k} - \lambda^{k+1} \end{array} \right\|_{G_1}^2,$$

where G_1 is defined in (2.11). Substituting the last equality into (3.13), we get

$$\left\| \begin{array}{c} y^{k+1} - y^* \\ \lambda^{k+1} - \lambda^* \end{array} \right\|_{H_\alpha}^2 \leq \left\| \begin{array}{c} y^{k} - y^* \\ \lambda^{k} - \lambda^* \end{array} \right\|_{H_\alpha}^2 - \frac{2-\alpha}{\alpha^2\beta} \left\| \begin{array}{c} y^{k} - y^{k+1} \\ \lambda^{k} - \lambda^{k+1} \end{array} \right\|_{G_1}^2. \tag{3.14}$$

On the other hand, setting $y' := y^k$ in (3.4), we have

$$(y^k - y^{k+1})^T (Ry^{k+1} + r - B^T\lambda^{k+1}) \geq 0. \tag{3.15}$$

Moreover, since y^k is the solution of (1.3b) at the $(k-1)$th iteration and $y^{k+1} \in \mathcal{Y}$, it holds that

$$(y^{k+1} - y^k)^T (Ry^k + r - B^T\lambda^k) \geq 0. \tag{3.16}$$

Adding (3.15) and (3.16) and using the positive semidefiniteness of R again, we obtain

$$(\lambda^k - \lambda^{k+1})^T (By^k - By^{k+1}) \geq 0,$$

which implies

$$\|\beta B(y^k - y^{k+1}) + (\lambda^k - \lambda^{k+1})\|^2 \geq \beta^2 \|By^k - By^{k+1}\|^2 + \|\lambda^k - \lambda^{k+1}\|^2,$$

and the assertion (3.11) follows immediately. ◻

To estimate the rate of the decrease of $\operatorname{dist}^2_{H_\alpha}(u^k, \Omega^*)$, we are interested in comparing $\|e(u^{k+1}, 1)\|$ and the distance of two consecutive iterates of ADMM. This is completed in the following lemma.

LEMMA 3.2. *Let* $\{u^k\}$ *be the sequence generated by the ADMM scheme* (1.3a)–(1.3c) *and* H_1 *and* G_2 *be defined in* (2.10) *and* (2.12), *respectively. Then we have*

$$\left\| \begin{array}{c} y^{k} - y^{k+1} \\ \lambda^{k} - \lambda^{k+1} \end{array} \right\|_{H_1}^2 \geq \frac{\lambda_{\min}(H_1)\alpha^2}{\lambda_{\max}(G_2)} \|e(u^{k+1}, 1)\|^2. \tag{3.17}$$

Proof. It follows from (3.3) that

$$x^{k+1} = P_{\mathcal{X}}(x^{k+1} - (Qx^{k+1} + q - A^T\bar\lambda^k)).$$

Thus,

$$\begin{aligned}\|e_{\mathcal{X}}(u^{k+1},1)\| &= \|x^{k+1} - P_{\mathcal{X}}(x^{k+1} - (Qx^{k+1} + q - A^T\lambda^{k+1})\| \\ &= \|P_{\mathcal{X}}(x^{k+1} - (Qx^{k+1} + q - A^T\bar{\lambda}^k)) \\ &\quad - P_{\mathcal{X}}(x^{k+1} - (Qx^{k+1} + q - A^T\lambda^{k+1})\| \\ &\le \|A^T(\bar{\lambda}^k - \lambda^{k+1})\|, \end{aligned} \tag{3.18}$$

where the inequality follows from the fact that the projection is nonexpansive (see [6]). From (3.1) and (3.5), we have

$$\begin{aligned}\bar{\lambda}^k - \lambda^{k+1} &= \beta(By^{k+1} - By^k) + (1-\alpha)(\bar{\lambda}^k - \lambda^k) \\ &= \beta((2-\alpha)(By^{k+1} - By^k) - (1-\alpha)(Ax^{k+1} + By^{k+1} - b)) \\ &= \frac{\beta}{\alpha}(By^{k+1} - By^k) + \frac{1-\alpha}{\alpha}(\lambda^{k+1} - \lambda^k), \end{aligned} \tag{3.19}$$

where the last equality follows from (3.10). Similarly, (3.4) implies

$$y^{k+1} = P_{\mathcal{Y}}(y^{k+1} - \beta(Ry^{k+1} + d - B^T\lambda^{k+1})).$$

Hence

$$\|e_{\mathcal{Y}}(u^{k+1},1)\| = \|y^{k+1} - P_{\mathcal{Y}}(y^{k+1} - (Ry^{k+1} + d - B^T\lambda^{k+1}))\| = 0. \tag{3.20}$$

Finally,

$$\|e_{\Lambda}(u^{k+1},1)\| = \|Ax^{k+1} + By^{k+1} - b\| = \frac{1}{\alpha}\left\|(1-\alpha)(By^{k+1} - By^k) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\right\|. \tag{3.21}$$

Combining (3.18)–(3.21), we have

$$\begin{aligned}\|e(u^{k+1},1)\|^2 &= \|e_{\mathcal{X}}(u^{k+1},1)\|^2 + \|e_{\mathcal{Y}}(u^{k+1},1)\|^2 + \|e_{\Lambda}(u^{k+1},1)\|^2 \\ &\le \frac{1}{\alpha^2}\Big\|A^T(\beta(By^{k+1} - By^k) + (1-\alpha)(\lambda^{k+1} - \lambda^k))\|^2 \\ &\qquad + \frac{1}{\alpha^2}\|(1-\alpha)(By^{k+1} - By^k) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\Big\|^2 \\ &= \frac{1}{\alpha^2}\left\| \begin{array}{c} y^k - y^{k+1} \\ \lambda^k - \lambda^{k+1} \end{array} \right\|_{G_2}^2. \end{aligned}$$

Recall the definitions of H_1 and G_2. We thus have

$$\left\| \begin{array}{c} y^k - y^{k+1} \\ \lambda^k - \lambda^{k+1} \end{array} \right\|_{H_1}^2 \ge \frac{\lambda_{\min}(H_1)}{\lambda_{\max}(G_2)} \left\| \begin{array}{c} y^k - y^{k+1} \\ \lambda^k - \lambda^{k+1} \end{array} \right\|_{G_2}^2 \ge \frac{\alpha^2\lambda_{\min}(H_1)}{\lambda_{\max}(G_2)}\|e(u^{k+1},1)\|^2,$$

which is assertion (3.17). The proof is completed. ∎

Now, combining the conclusions in Theorem 3.1 and Lemma 3.2, we have the following corollary which is crucial for establishing the local linear convergence of ADMM.

COROLLARY 3.1. *Let* $\{u^k\}$ *be the sequence generated by the ADMM scheme* (1.3a)–(1.3c) *and* H_α, H_1, *and* G_2 *be defined in* (2.9), (2.10), *and* (2.12), *respectively.*

Then for any $v^* \in \Omega^*$, we have

$$\left\| \begin{array}{c} y^{k+1} - y^* \\ \lambda^{k+1} - \lambda^* \end{array} \right\|_{H_\alpha}^2 \le \left\| \begin{array}{c} y^k - y^* \\ \lambda^k - \lambda^* \end{array} \right\|_{H_\alpha}^2 - \frac{(2-\alpha)\lambda_{\min}(H_1)}{\beta\tau^2\lambda_{\max}(M)\lambda_{\max}(G_2)} \cdot dist_{H_\alpha}^2(v^{k+1}, \Omega^*). \tag{3.22}$$

Proof. It follows from Lemma 2.3 that

$$\begin{aligned} \|e(u^{k+1},1)\|^2 &\ge \frac{1}{\lambda_{\max}(M)}\|e(u^{k+1},1)\|_M^2 \ge \frac{1}{\tau^2\lambda_{\max}(M)} \cdot \mathrm{dist}_M^2(u^{k+1},\mathcal{U}^*) \\ &\ge \frac{1}{\tau^2\lambda_{\max}(M)} \cdot \mathrm{dist}_{H_\alpha}^2(v^{k+1},\Omega^*). \end{aligned}$$

Combining this inequality with (3.17), we obtain

$$\left\| \begin{array}{c} y^k - y^{k+1} \\ \lambda^k - \lambda^{k+1} \end{array} \right\|_{H_1}^2 \ge \frac{\alpha^2\lambda_{\min}(H_1)}{\tau^2\lambda_{\max}(M)\lambda_{\max}(G_2)} \cdot \mathrm{dist}_{H_\alpha}^2(v^{k+1},\Omega^*),$$

which, together with (3.11), implies the assertion (3.22) immediately. ∎

Now, we are ready to present the main result of the local linear convergence of ADMM for solving (1.4).

THEOREM 3.2. *Let $\{u^k\}$ be the sequence generated by the ADMM scheme* (1.3a)–(1.3c). *When the iterative w^k is close enough to Ω^* such that $\|e(u^k,1)\| \le \epsilon$ is satisfied, we have*

$$dist_{H_\alpha}^2(v^{k+1},\Omega^*) \le \frac{1}{1+\xi} \cdot dist_{H_\alpha}^2(v^k,\Omega^*), \tag{3.23}$$

where

$$\xi := \frac{(2-\alpha)\lambda_{\min}(H_1)}{\beta\tau^2\lambda_{\max}(M)\lambda_{\max}(G_2)} > 0.$$

Proof. Obviously, for any $(y^*, \lambda^*) \in \Omega^*$ we have

$$\begin{aligned} \mathrm{dist}_{H_\alpha}^2(v^{k+1},\Omega^*) &\le \left\| \begin{array}{c} y^{k+1} - y^* \\ \lambda^{k+1} - \lambda^* \end{array} \right\|_{H_\alpha}^2 \\ &\le \left\| \begin{array}{c} y^k - y^* \\ \lambda^k - \lambda^* \end{array} \right\|_{H_\alpha}^2 - \frac{(2-\alpha)\lambda_{\min}(H_1)}{\beta\tau^2\lambda_{\max}(M)\lambda_{\max}(G_2)} \cdot \mathrm{dist}_{H_\alpha}^2(v^{k+1},\Omega^*), \end{aligned}$$

where the first inequality comes from the definition of $\mathrm{dist}_{H_\alpha}^2(v^{k+1},\Omega^*)$ and the second inequality is because of (3.22). Since the above inequality holds for any $(y^*,\lambda^*) \in \Omega^*$, we have

$$\mathrm{dist}_{H_\alpha}^2(v^{k+1},\Omega^*) \le \mathrm{dist}_{H_\alpha}^2(v^k,\Omega^*) - \frac{(2-\alpha)\lambda_{\min}(H_1)}{\beta\tau^2\lambda_{\max}(M)\lambda_{\max}(G_2)} \cdot \mathrm{dist}_{H_\alpha}^2(v^{k+1},\Omega^*),$$

and this inequality implies the assertion (3.23) immediately. This completes the proof. ∎

Theorem 3.2 indicates the local linear convergence of the ADMM scheme (1.3a)–(1.3c) for solving the model (1.4).

Remark 3.1. In ADMM literature, it has been commonly observed that the value of β may effect ADMM's numerical performance significantly. Although it is still unclear how to determine this parameter optimally for different applications, many articles have mentioned the fact that we should avoid extreme values of β in computations and it is usually preferred to modify it in a moderate manner; see, e.g., [2, 14]. Our result (3.23) provides a justification for the reason why we prefer to avoid extreme values of β. Let's use the original ADMM scheme (1.2a)–(1.2c) as an illustrative example. For this case, since $\alpha = 1$, $H_\alpha = H_1$ and G_2 reduces to

$$G_2 = \begin{pmatrix} \beta^2 B^T AA^T B & 0 \\ 0 & \frac{1}{\beta^2} I_l \end{pmatrix}.$$

Then, both H_1 and G_2 are block diagonal matrices with two diagonal blocks. For H_1 (resp., G_2), one diagonal block involves β (resp., β^2), while the other involves $\frac{1}{\beta}$ (resp., $\frac{1}{\beta^2}$). Therefore, extreme values of β, either too large or too small, shall increase the factor $(1+\xi)^{-1}$ in (3.23), which equivalently means that the convergence of ADMM will be slowed down. This also explains why some self-adaptive strategies in [2, 14] to adjust β dynamically during iterations are empirically useful.

4. Conclusions. We establish the local linear convergence rate of the ADMM for solving a class of quadratic programming problems. Since the model under our consideration already includes some important applications, the linear convergence rate derived in this paper is a strong theoretical support of the empirical efficiency of ADMM which has been widely witnessed in various areas. Our approach to deriving this linear convergence rate is based on the affine variational inequality (LVI) reformulation of the model under consideration and a profound error bound in LVI literature. It is interesting to extend our technique to other models with more general settings, and thus derive the linear convergence rate of ADMM for more general models. In particular, the extension to the convex quadratically constrained quadratic semidefinite programs considered in [21] and the convex nonlinear semidefinite programming problem in [22] is of great interest.

Acknowledgments. The authors are grateful to two anonymous referees for their valuable comments on an earlier version of this paper. This work was also partially supported by the Institute of Computational and Theoretical Studies of Hong Kong Baptist University while the first author was a visiting research fellow of this institute.

REFERENCES

[1] D. Boley, *Local Linear Convergence of ADMM on Quadratic or Linear Programs*, manuscript, 2012.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2010), pp. 1–122.

[3] W. Deng and W. Yin, *On the Global and Linear Convergence of the Generalized Alternating Direction Method of Multipliers*, manuscript, 2012.

[4] J. Eckstein and D. P. Bertsekas, *An Alternating Direction Method for Linear Programming*, Technical report LIDS-P-1967, Laboratory for Information and Decision Systems, Machachusetts Institute of Technology, Cambridge, MA, 1990.

[5] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318.

[6] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vol. I, Berlin, Springer Verlag, 2003.
[7] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximations*, Comput. Math. Appl., 2 (1976), pp. 17–40.
[8] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., North-Holland, Amsterdam, 1983, pp. 299–331.
[9] R. Glowinski, T. Kärkkäinen, and K. Majava, *On the convergence of operator-splitting methods*, in Numerical Methods for Scientific Computing, Variational Problems and Applications, Barcelona, 2003, Y. Kuznetsov, P. Neittanmäki, and O. Pironneau, eds.
[10] R. Glowinski and A. Marrocco, *Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problèmes non linéaires*, Rev. Franc. Autom. Inf. Rech. Oper. Anal., R2 (1975), pp. 41–76.
[11] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, Stud. Appl. Numer. Math., SIAM, Philadelphia, 1989.
[12] E. G. Gol'shtein and N. V. Tret'yakov, *Modified Lagrangian in convex programming and their generalizations*, Math. Program. Study, 10 (1979), pp. 86–97.
[13] M. Hong and Z. Luo, *On the Linear Convergence of the Alternating Direction Method of Multipliers*, manuscript, 2012.
[14] B. S. He, H. Yang, and S. L. Wang, *Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities*, J. Optim. Theory Appl., 106 (2000), pp. 337–356.
[15] B. He and X. Yuan, *On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.
[16] B. S. He and X. M. Yuan, *On convergence rate of the Douglas-Rachford operator splitting method*, Math. Program., submitted.
[17] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979.
[18] Z.-Q. Luo and P. Tseng, *Error bound and convergence analysis of matrix splitting algorithms for the affine variational inquality problem*, SIAM J. Optim., 2 (1992), pp. 43–54.
[19] B. Martinet, *Regularision d'inéquations variationnelles par approximations successive*, Rev. Franc. Autom. Inf. Rech. Oper., 126 (1970), pp. 154–159.
[20] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D., 60 (1992), pp. 259–268.
[21] J. Sun and S. Zhang, *A modified alternating direction method for convex quadratically constrained quadratic semidefinite programs*, Eur. J. Oper. Res., 207 (2010), pp. 1210–1220.
[22] S. Zhang, J. Ang, and J. Sun, *An alternating direction method for solving convex nonlinear semidefinite programming problem*, Optimization, 62 (2013), pp. 527–543.