LINEAR CONVERGENCE RATE OF THE GENERALIZED ALTERNATING DIRECTION METHOD OF MULTIPLIERS FOR A CLASS OF CONVEX MINIMIZATION PROBLEMS*

JIANWEN PENG† AND XUEQING ZHANG‡

Abstract. Rencently, the generalized aternating direction method of multipliers (GADMM) proposed by Eckstein and Bertsekas has received intensive attention from a broad spectrum of areas. In this paper, we consider the convergence rate of GADMM when applying to the convex optimization problems that the subdifferentials of the underlying functions are piecewise linear multifunctions, including LASSO, a well-known regression model in statistics, as a special case. We firstly prove some important inequalities for the sequence generated by the GADMM for solving the convex optimization problems. Secondly, we establish both the local linear convergence rate and the global linear convergence rate of GADMM for solving the convex optimization problems that the subdifferentials of the underlying functions are piecewise linear multifunctions. The maim results in this paper extend and improve some well-known ones in the literature.

Key words. local linear convergence rate, global linear convergence rate, generalized alternating direction method of multipliers, piecewise linear multifunction

AMS subject classifications. 90C25, 65K10

1. Introduction. In this paper, we consider the convex minimization model with linear constraints and an objective function which is sum of two functions without coupled variables

$$\min \quad \{f_1(x_1)+f_2(x_2)\} \quad \text{s.t.} \quad A_1x_1+A_2x_2=b, \quad x_1\in\mathcal{X}_1, x_2\in\mathcal{X}_2, \tag{1}$$

where $A_1\in\mathcal{R}^{\ell\times n}$ and $A_2\in\mathcal{R}^{\ell\times m}$ are two given matrices, $b\in\mathcal{R}^{\ell}$ is a given vector, $\mathcal{X}_1\subseteq\mathcal{R}^n$ and $\mathcal{X}_2\subseteq\mathcal{R}^m$ are two polyhedra, $f_1:\mathcal{R}^n\to\mathcal{R}$ and $f_2:\mathcal{R}^m\to\mathcal{R}$ are convex functions.

The iterative scheme of alternating direction method of multipliers (ADMM) for solving (1) reads as

$$\begin{cases} x_1^{k+1}=\operatorname{argmin}\{f_1(x_1)-x_1^\top A^\top\lambda^k+\dfrac{\beta}{2}\|A_1x_1+A_2x_2^k-b\|^2|x_1\in\mathcal{X}_1\}, \\ x_2^{k+1}=\operatorname{argmin}\{f_2(x_2)-x_2^\top B^\top\lambda^k+\dfrac{\beta}{2}\|A_1x_1^{k+1}+A_2x_2-b\|^2|x_2\in\mathcal{X}_2\}, \\ \lambda^{k+1}=\lambda^k-\beta(A_1x_1^{k+1}+A_2x_2^{k+1}-b), \end{cases} \tag{2}$$

where λ is the Lagrange multiplier and $\beta>0$ is a penalty parameter. The ADMM was proposed originally in [1] (see also [2]) which is essentially a splitting version of the augmented Lagrangian method in [3, 4]. In [5], the iterative scheme (2) was illustrated as an application of the Douglas-Rachford splitting method (DRSM) in [6] to the dual of (1); and in [7] DRSM was shown to be a special case of the proximal point algorithm in [8]. Therefore, inspired by the work [9], Eckstein and Bertsekas [7] proposed the so-called generalized aternating direction method of multipliers (GADMM) as follows

*

Funding: Research of this author was supported by the National Natural Science Foundation of China (11431004), the Chongqing Research Program of Basic Research and Frontier Technology (cstc2015jcyjBX0029), and the Education Committee Project Foundation of Chongqing for "Bayu scholar" Distinguished Professor.

†School of Mathematical Sciences, Chongqing Normal University, Chongqing 400047, P.R.China. (jwpeng6@aliyun.com)

‡School of Mathematical Sciences, Chongqing Normal University, Chongqing 400047, P.R.China. (zxqcqspb@163.com)

$$\left\{\begin{array}{ll} x_1^{k+1}= \mathrm{argmin}\{f_1(x_1) - x_1^\top A_1^\top \lambda^k + \frac{\beta}{2}\|A_1x_1 + A_2x_2^k - b\|^2 | x_1 \in \mathcal{X}_1\}, & (3)\\ x_2^{k+1}=\mathrm{argmin}\{f_2(x_2) - x_2^\top A_2^\top \lambda^k & (4)\\ \quad + \frac{\beta}{2}\|\alpha A_1x_1^{k+1} - (1-\alpha)(A_2x_2^k - b) + A_2x_2 - b\|^2 | x_2 \in \mathcal{X}_2\}\ , & \\ \lambda^{k+1} = \lambda^k - \beta(\alpha A_1x_1^{k+1} - (1-\alpha)(A_2x_2^k - b) + A_2x_2^{k+1} - b), & (5) \end{array}\right.$$

where the parameter $\alpha \in (0, 2)$ is an acceleration factor and it is usually suggested to take $\alpha \in (1, 2)$. It is easy to see that the original ADMM scheme (2) is a special case of the GADMM (3)-(5) with $\alpha = 1$.

The global convergence of ADMM and its variants were well established under very mild conditions such that the nonemptiness of the solution set of (1). However, research on ADMM's linear convergence rate is still in its infancy. Lions and Mercier [6] showed that the Douglas-Rachford operator splitting method converges linearly under the assumption that some involved monotone operator is both coercive and Lipschitz. Eckstein and Bertsekas [10] established the global linear convergence rate when ADMM was applied to solve a standard linear programming model. Boley [11] proved the local linear convergence of ADMM applying to the linear program or the quadratic program. Han and Yuan [12] proved the local linear convergence of GADMM applying to the quadratic program. Tao and Yuan [13] proved the global linear convergence of ADMM applying to the quadratic program. Yang and Han [14] proved the global linear convergence of ADMM applying to the convex optimization problem that the subdifferentials of the underlying functions are piecewise linear multifunctions. Hong and Luo [15] proved the linear convergence rate of the ADMM for minimizing the sum of any number of convex separable functions, under the requirement that a certain error bound condition holds true and the dual stepsize is sufficiently small. He, Tao and Yuan [16] proved the local linear convergence rate of the ADMM with a backward or forward substitution procedure for solving a convex minimization model with a general separable structure, under the requirement that a certain error bound condition holds true. Under the assumption that one/ some of underlying functions is/ are strongly convex, the global linear convergence of ADMM was shown in [17], [18], [19] and [20]. Han, Sun and Zhang [21] estalished the global linear rate of convergence of a semi-proximal ADMM for solving linearly constrained convex composite optimization problems under a mild calmness condition. It is well known that $F : \mathcal{R}^n \to 2^{\mathcal{R}^n}$ is a piecewise linear multifunction (see [22]) (or piecewise polyhedral multifunction; see [23]) if $G_r(F) := \{(x, y)|y \in F(x)\}$ is the union of finitely many polyhedra. The important classes of piecewise linear multifunctions include the subdifferentials of the convex piecewise linear-quadratic functions (see [24]) and the functions in the following l_1-norm regularized leaset-squares model:

$$\min_{x\in\mathcal{R}^n} \ \frac{1}{2}\|Ex - d\|^2 + \nu\|x\|_1, \tag{6}$$

where $E \in \mathcal{R}^{\ell\times n}$ is a given matrix, $d \in \mathcal{R}^l$ is a given vector, ν is a positive scalar and $\|x\|_1 := \sum_{i=1}^n |x^i|$. Popular applications of (6) include the well-known LASSO model in statistics and the basis pursuit model in compressive sensing.

It is natural to raise and to give an answer to the following question: can we establish the local linear convergence rate of GADMM or the global linear convergence rate of GADMM when the subdifferentials of the underlying functions are piecewise linear multifunctions with minor conditions? In this paper, we will give some positive answers to this question.

The rest is organized as follows. In section 2, we give some notations and definitions and describe a VI formulation of (1), which are useful for our analysis. In section 3, we prove some important inequalities for the sequence generated by the GADMM for solving the convex optimization problems. In section 4, we prove the local linear convergence rate of GADMM for the problem (1) by using the locally metric subregularity property for a piecewise linear multifunction; In section 5, we prove the global linear convergence rate of the GADMM for the convex optimization problem (1) when the subdifferentials of the underlying functions are piecewise linear multifunctions. We give some numerical results to show the global linear convergence of the GADMM in section 6. In section 7, we draw some conclusions.

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

2.1. Notation set and some definitions. Throughout this paper, all vectors are column vectors. For any two vectors $x_1 \in \mathcal{R}^n$ and $x_2 \in \mathcal{R}^m$, we simply use $u = (x_1, x_2)$ to denote their adjunction, i.e., (x_1, x_2) denotes $(x_1^\top, x_2^\top)^\top$. We use $\|x_1\|_p$ denotes the p-norm of vector x_1, where $p = 1$ or 2; and for $p = 2$, we simply denote it as $\|x_1\|$. For any symmetric and positive definite matrix M, we denote $\|x_1\|_M = \sqrt{x_1^\top M x_1}$ as its M-norm. For a given matrix A_1, its minimal and maximal eigenvalues are denoted by $\lambda_{\min}$ and $\lambda_{\max}$, respectively, its norm is

$$\|A_1\|_p := \sup_{x_1 \neq 0} \left\{ \frac{\|A_1 x_1\|_p}{\|x_1\|_p} \right\}. \tag{7}$$

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

Further more, for a multifunction $F : \mathcal{R}^n \to 2^{\mathcal{R}^n}$, we say that F is monotone if for any $x_1, x_2 \in \mathcal{R}^n$, one has

$$(x_1 - x_2)^\top (\xi - \zeta) \geq 0 \quad \forall \xi \in F(x_1), \forall \zeta \in F(x_2).$$

It is well known that the subdifferential of a convex function f denoted by ∂f is a monotone multifunction (see [25]).

Let $\mathcal{K} \subseteq \mathcal{R}^n$ be a nonempty, closed, and convex set and the projection operator under the Euclidean norm $P_{\mathcal{K}}(\cdot) : \mathcal{R}^n \mapsto \mathcal{K}$ be defined by

$$P_{\mathcal{K}}(u) = \operatorname{argmin}_{v \in \mathcal{K}} \{\|v - u\|\}, \quad \forall u \in \mathcal{R}^n.$$

A key property is that $P_{\mathcal{K}}$ is a nonexpansive map, i.e.,

$$\|P_{\mathcal{K}}(x_1) - P_{\mathcal{K}}(x_2)\| \leq \|x_1 - x_2\|, \quad \forall x_1 \in \mathcal{R}^n, \quad \forall x_2 \in \mathcal{R}^n \tag{8}$$

The normal cone of $\mathcal{K}$ at z denoted by $N_{\mathcal{K}}(z)$ (see [26]) is

$$N_{\mathcal{K}}(z) = \{v \in \mathcal{R}^n : v^T (x_2 - z) \leq 0, \forall x_2 \in \mathcal{K}\}.$$

2.2. Variational inequalities. The first-order optimality condition of (1) can be expressed as a variational inequality (VI), and our analysis is based on the VI characterization of (1). Throughout this paper, we need the following assumptions being used by Yang and Han [14]:

Assumption 1: The optimal solution set of problem (1) is nonempty.

Assumption 2: Let $\mathcal{S} = \{(x_1, x_2) : A_1x_1 + A_2x_2 = b\}$ and $\mathcal{Z} = \mathcal{X}_1 \times \mathcal{X}_2$. The pair $\{\mathcal{S}, \mathcal{Z}\}$ has the strong conical hull intersection property, that is, for any $z \in \mathcal{S} \cap \mathcal{Z}$, $N_{\mathcal{S}\cap\mathcal{Z}} = N_{\mathcal{S}}(z) + N_{\mathcal{Z}}(z)$.

Under the above assumptions, (x_1^*, x_2^*) is an optimal solution of (1) iff there exists $\xi \in \partial f(x_1^*), \zeta \in \partial g(x_2^*)$, and $\lambda^* \in \mathcal{R}^\ell$ such that

$$\left\{ \begin{array}{c} \xi - A_1^\top \lambda^* \in -N_{\mathcal{X}_1}(x_1^*), \\ \zeta - A_2^\top \lambda^* \in -N_{\mathcal{X}_2}(x_2^*), \\ A_1x_1^* + A_2x_2^* - b = 0. \end{array} \right. \tag{9}$$

The KKT condition (9) can be expressed as a set-valued VI. Let $F : \mathcal{R}^n \to 2^{\mathcal{R}^n}$ be a monotone operator and $\mathcal{K} \subseteq \mathcal{R}^n$ be a nonempty, closed, and convex set. The set-valued VI problem denoted by $SVI(\mathcal{K}, F)$ is to find $u^* \in \mathcal{K}$ such that

$$(u - u^*)^\top \theta \geq 0, \ \forall u \in \mathcal{K},$$

for some $\theta \in F(u^*)$.

For $u \in \mathcal{K}$ and $\gamma > 0$, define the error function $e(u, \gamma)$ as follows:

$$e(u, \gamma) := u - P_{\mathcal{K}}(u - \gamma F(u)).$$

Then $e(u, \gamma)$ is a multifunction and the following lemma holds.

Lemma 2.1[14]. Solving $SVI(\mathcal{K}, F)$ amounts to finding u^* such that $0 \in e(u^*, \gamma)$, or equivalently,

$$\mathrm{dist}(0, e(u^*, \gamma)) = 0. \tag{10}$$

Let $f_1, f_2, A_1, A_2, \mathcal{X}_1$, and $\mathcal{X}_2$ be as in problem (1). Let

$$u := \begin{pmatrix} x_1 \\ x_2 \\ \lambda \end{pmatrix}, \qquad L(u) := \begin{pmatrix} \partial f_1(x_1) - A_1^\top \lambda \\ \partial f_2(x_2) - A_2^\top \lambda \\ A_1x_1 + A_2x_2 - b \end{pmatrix}, \qquad \mathcal{U} := \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{R}^\ell.$$

Then $(x_1^*, x_2^*, \lambda^*)$ satisfies (9) iff $(x_1^*, x_2^*, \lambda^*)$ is a solution of set-valued VI problem SVI$(\mathcal{U}, L)$; equivalently there exists $\xi \in \partial f_1(x_1^*), \zeta \in \partial f_2(x_2^*)$ such that

$$(u - u^*)^\top \theta(u^*) \geq 0, \qquad \forall u \in \mathcal{U}. \tag{11}$$

where

$$u^* = \begin{pmatrix} x_1^* \\ x_2^* \\ \lambda^* \end{pmatrix}; \qquad \theta(u^*) = \begin{pmatrix} \xi - A_1^\top \lambda^* \\ \zeta - A_2^\top \lambda^* \\ A_1x_1^* + A_2x_2^* - b \end{pmatrix}.$$

The error function $e(u, \gamma)$ associated with SVI$(\mathcal{U}, L)$ becomes

$$e(u, \gamma) := \begin{pmatrix} e_{\mathcal{X}_1}(u, \gamma) := x_1 - P_{\mathcal{X}_1}[x_1 - \gamma(\partial f_1(x_1) - A_1^\top \lambda)] \\ e_{\mathcal{X}_2}(u, \gamma) := x_2 - P_{\mathcal{X}_2}[x_2 - \gamma(\partial f_2(x_2) - A_2^\top \lambda)] \\ e_\Lambda(u, \gamma) := \gamma(A_1x_1 + A_2x_2 - b) \end{pmatrix}. \tag{12}$$

We use $\mathcal{U}^*$ to denote the solution of SVI($\mathcal{U}, L$) in (11). From Lemma 2.1, it follows that $\mathcal{U}^* = \{u | \text{dist}(0, e(u, \gamma)) = 0\}$. Then, we give some notations for further analysis:

$$v = (x_2, \lambda); \mathcal{V} = \mathcal{X}_2 \times R^\ell; v^k = (x_2^k, \lambda^k), v^* = (x_2^*, \lambda^*);$$
$$\Omega^* = \{(x_2^*, \lambda^*) | (x_1^*, x_2^*, \lambda^*) \in \mathcal{U}^*\}.$$

From Theorem 3.3 (i) in [22], we can easily obtain the following locally metric subregularity property for a piecewise linear multifunction:

Lemma 2.2. Let $F : \mathcal{R}^n \to 2^{\mathcal{R}^n}$ be a piecewise linear multifunction. Then, there exists $\kappa, \tau > 0$ such that

$$\text{dist}(x, F^{-1}(0)) \leq \tau \text{dist}(0, F(x)),$$

whenever $x \in \mathcal{R}^n$ satisfying $\text{dist}(0, F(x)) < \kappa$.

It is worthy to note that Lemma 2.2 plays an important role when we can establish the local linear convergence rate of GADMM scheme (3)-(5) for (1).

Now we define some matrices which will be used to simplify our notation in the following analysis. Let

$$\mathcal{T}_\alpha = \frac{1}{\alpha} \begin{pmatrix} \beta A_2^\top A_2 & (1-\alpha) A_2^\top \\ (1-\alpha) A_2 & \frac{1}{\beta} I_\ell \end{pmatrix}, \tag{13}$$

$$\mathcal{T}_1 = \begin{pmatrix} \beta A_2^\top A_2 & 0 \\ 0 & \frac{1}{\beta} I_\ell \end{pmatrix}, \tag{14}$$

$$M_\alpha = \begin{pmatrix} I_n & 0 \\ 0 & \mathcal{T}_\alpha \end{pmatrix}, \tag{15}$$

and

$$G_\alpha = \begin{pmatrix} A_2^\top (\beta^2 A_1 A_1^\top + (1-\alpha)^2 I_\ell) A_2 & (1-\alpha)(\beta A_2^\top A_1 A_1^\top + \frac{1}{\beta} A_2^\top) \\ (1-\alpha)(\beta A_1 A_1^\top A_2 + \frac{1}{\beta} A_2) & (\frac{1}{\beta^2} + (1-\alpha)^2 A_1 A_1^\top) I_\ell \end{pmatrix}. \tag{16}$$

Note the positive definiteness of $\mathcal{T}_\alpha$, M_α and the positive semidefiniteness G_α are guaranteed when $\alpha \in (0, 2)$ and A_2 is assumed to be full column rank.

3. Global convergence of the GADMM. In this section, we will prove two important inequalities (see Theorem 3.1) which will play a fundamental role in the analysis of both the local linear convergence rate and global linear convergence rate of GADMM.

Lemma 3.1. Let $\{u^k\}$ be the sequence generated by the GADMM scheme (3)-(5) and the sequence $\{\bar{\lambda}^k\}$ be defined as

$$\bar{\lambda}^k := \lambda^k - \beta(A_1 x_1^{k+1} + A_2 x_2^k - b). \tag{17}$$

Then, we have

$$\begin{pmatrix} x_2^{k+1} - x_2^* \\ \bar{\lambda}^k - \lambda^* \end{pmatrix}^\top \mathcal{T}_\alpha \begin{pmatrix} x_2^k - x_2^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix} \geq 0 \ \ \forall\ (x_2^*, \lambda^*) \in \Omega^*. \tag{18}$$

Proof. Using the first-order optimality condition for (3), there exists $\xi \in \partial f_1(x_1^{k+1})$ such that

$$(x_1' - x_1^{k+1})^\top (\xi - A_1^\top \lambda^k + \beta A_1^\top (A_1 x_1^{k+1} + A_2 x_2^k - b)) \geq 0 \ \forall x_1' \in \mathcal{X}_1, \tag{19}$$

It follows from (17) that

$$(x_1' - x_1^{k+1})^\top (\xi - A_1^\top \bar{\lambda}^k) \geq 0 \ \forall x_1' \in \mathcal{X}_1, \tag{20}$$

By the first-order optimality condition for (4), there exists $\zeta \in \partial f_2(x_2^{k+1})$ such that

$$(x_2' - x_2^{k+1})^\top (\zeta - A_2^\top \lambda^k + \beta A_2^\top (A_2 x_2^{k+1} - A_2 x_2^k) + \alpha A_2^\top (\lambda^k - \bar{\lambda}^k)) \geq 0, \forall x_2' \in \mathcal{X}_2. \tag{21}$$

It follows from (5) and (17), we have

$$\lambda^{k+1} = \bar{\lambda}^k - \beta(A_2 x_2^{k+1} - A_2 x_2^k) - (1-\alpha)(\bar{\lambda}^k - \lambda^k) \tag{22}$$

Then, (21) can be rewritten as

$$(x_2' - x_2^{k+1})^\top (\zeta - A_2^\top \lambda^{k+1}) \geq 0, \forall x_2' \in \mathcal{X}_2. \tag{23}$$

Setting $x_1' := x_1^*$ in (20) and $x_2' := x_2^*$ in (23), we obtain

$$(x_1^* - x_1^{k+1})^\top (\xi - A_1^\top \bar{\lambda}^k) \geq 0, \tag{24}$$

and

$$(x_2^* - x_2^{k+1})^\top (\zeta - A_2^\top \lambda^{k+1}) \geq 0. \tag{25}$$

Clearly, there exist $\xi^* \in \partial f_2(x^*)$ and $\zeta^* \in \partial f_2(x_2^*)$ such that

$$\begin{pmatrix} x_1 - x_1^* \\ x_2 - x_2^* \end{pmatrix}^\top \begin{pmatrix} \xi^* - A_1^\top \lambda^* \\ \zeta^* - A_2^\top \lambda^* \end{pmatrix} \geq 0 \quad \forall\, x_1 \in \mathcal{X}_1, \forall\, x_2 \in \mathcal{X}_2. \tag{26}$$

Setting $x_1 := x_1^{k+1}$ and $x_2 := x_2^{k+1}$ in (26), and adding the resulting inequality to (24)-(25),

$$\begin{pmatrix} x_1^{k+1} - x_1^* \\ x_2^{k+1} - x_2^* \end{pmatrix}^\top \begin{pmatrix} (\xi^* - \xi) - A_1^\top (\lambda^* - \bar{\lambda}^k) \\ (\zeta^* - \zeta) - A_2^\top (\lambda^* - \lambda^{k+1}) \end{pmatrix} \geq 0.$$

From the monotonicity of ∂f_1 and ∂f_2, we have

$$\begin{cases} (\xi^* - \xi)^\top (x_1^* - x_1^{k+1}) \geq 0, \\ (\zeta^* - \zeta)^\top (x_2^* - x_2^{k+1}) \geq 0. \end{cases}$$

the above inequalities yield

$$\begin{pmatrix} x_1^{k+1} - x_1^* \\ x_2^{k+1} - x_2^* \end{pmatrix}^\top \begin{pmatrix} A_1^\top (\bar{\lambda}^k - \lambda^*) \\ A_2^\top (\lambda^{k+1} - \lambda^*) \end{pmatrix} \geq 0.$$

It then follows from (22) and $A_1 x_1^* + A_2 x_2^* = b$ that

$$\begin{aligned} &(\bar{\lambda}^k - \lambda^*)^\top (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b) \\ &\qquad - (A_2 x_2^{k+1} - A_2 x_2^*)^\top (\beta(A_2 x_2^{k+1} - A_2 x_2^k) + (1-\alpha)(\bar{\lambda}^k - \lambda^k)) \geq 0. \end{aligned} \tag{27}$$

since

$$(1-\alpha)(A_2x_2^{k+1}-A_2x_2^k)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)=-\alpha(A_1x_1^{k+1}+A_2x_2^{k+1}-b), \tag{28}$$

the assertion (18) is trivial by using (14) and (27).

Remark 3.1.

Theorem 3.1. If $\alpha\in(0,2)$, A_2 is full column rank, and the sequence $\{u^k\}$ is generated by the GADMM scheme (3)-(5), then the following inequalities hold:

$$\left\|v^{k+1}-v^*\right\|_{\mathcal{T}_\alpha}^2\le\left\|v^k-v^*\right\|_{\mathcal{T}_\alpha}^2-\frac{2-\alpha}{\alpha^2}\left\|v^k-v^{k+1}\right\|_{\mathcal{T}_1}^2,\ \forall\ \ v^*\in\Omega^*, \tag{29}$$

and

$$\left\|v^{k+1}-v^*\right\|_{\mathcal{T}_\alpha}^2\le\left\|v^k-v^*\right\|_{\mathcal{T}_\alpha}^2-\frac{2-\alpha}{\alpha}\left\|v^k-v^{k+1}\right\|_{\mathcal{T}_\alpha}^2,\ \forall\ \ v^*\in\Omega^*, \tag{30}$$

where $v=\begin{pmatrix}x_2\\ \lambda\end{pmatrix}$.

Proof. Recall that for any two vectors g and h with the same dimension and for a suitable positive definite matrix H, we have the identity

$$\|g-h\|_{\mathcal{T}}^2=\|g\|_{\mathcal{T}}^2-\|h\|_{\mathcal{T}}^2-2(g-h)^\top\mathcal{T}h. \tag{31}$$

Setting

$$\mathcal{T}:=\mathcal{T}_\alpha,\quad g:=v^k-v^*\ \ \text{and}\ \ h:=v^k-v^{k+1},$$

we have

$$\begin{aligned}\left\|v^{k+1}-v^*\right\|_{\mathcal{T}_\alpha}^2=&\left\|v^k-v^*\right\|_{\mathcal{T}_\alpha}^2-\left\|v^k-v^{k+1}\right\|_{\mathcal{T}_\alpha}^2\\ &-2\left(v^{k+1}-v^*\right)^\top\mathcal{T}_\alpha\left(v^k-v^{k+1}\right).\end{aligned} \tag{32}$$

Then, setting

$$g:=\begin{pmatrix}x_2^k-x_2^{k+1}\\ \lambda^k-\bar{\lambda}^k\end{pmatrix}\ \ \text{and}\ \ h:=\begin{pmatrix}0\\ \lambda^{k+1}-\bar{\lambda}^k\end{pmatrix},$$

yields

$$\begin{aligned}\left\|\begin{matrix}x_2^k-x_2^{k+1}\\ \lambda^k-\lambda^{k+1}\end{matrix}\right\|_{\mathcal{T}_\alpha}^2=&\left\|\begin{matrix}x_2^k-x_2^{k+1}\\ \lambda^k-\bar{\lambda}^k\end{matrix}\right\|_{\mathcal{T}_\alpha}^2-\left\|\begin{matrix}0\\ \lambda^{k+1}-\bar{\lambda}^k\end{matrix}\right\|_{\mathcal{T}_\alpha}^2\\ &-2\begin{pmatrix}x_2^k-x_2^{k+1}\\ \lambda^k-\lambda^{k+1}\end{pmatrix}^\top\mathcal{T}_\alpha\begin{pmatrix}0\\ \lambda^{k+1}-\bar{\lambda}^k\end{pmatrix}.\end{aligned}$$

Substituting the above equality into (32), we can obtain

$$\begin{aligned}\left\|v^{k+1}-v^*\right\|_{\mathcal{T}_\alpha}^2=&\left\|\begin{matrix}x_2^k-x_2^*\\ \lambda^k-\lambda^*\end{matrix}\right\|_{\mathcal{T}_\alpha}^2-\left\|\begin{matrix}x_2^k-x_2^{k+1}\\ \lambda^k-\bar{\lambda}^k\end{matrix}\right\|_{\mathcal{T}_\alpha}^2+\left\|\begin{matrix}0\\ \lambda^{k+1}-\bar{\lambda}^k\end{matrix}\right\|_{\mathcal{T}_\alpha}^2\\ &-2\begin{pmatrix}x_2^k-x_2^{k+1}\\ \lambda^k-\lambda^{k+1}\end{pmatrix}^\top\mathcal{T}_\alpha\begin{pmatrix}x_2^{k+1}-x_2^*\\ \bar{\lambda}^k-\lambda^*\end{pmatrix}\\ \le&\left\|\begin{matrix}x_2^k-x_2^*\\ \lambda^k-\lambda^*\end{matrix}\right\|_{\mathcal{T}_\alpha}^2-\left\|\begin{matrix}x_2^k-x_2^{k+1}\\ \lambda^k-\bar{\lambda}^k\end{matrix}\right\|_{\mathcal{T}_\alpha}^2+\left\|\begin{matrix}0\\ \lambda^{k+1}-\bar{\lambda}^k\end{matrix}\right\|_{\mathcal{T}_\alpha}^2,\end{aligned} \tag{33}$$

where the second inequality follows from (18). By using (17) and (5), it is easy to see that

(34)
$$\left\|\begin{array}{c} x_2^k - x_2^{k+1} \\ \lambda^k - \bar{\lambda}^k \end{array}\right\|_{\mathcal{T}_\alpha}^2 - \left\|\begin{array}{c} 0 \\ \lambda^{k+1} - \bar{\lambda}^k \end{array}\right\|_{\mathcal{T}_\alpha}^2 = \frac{2-\alpha}{\alpha^2\beta}\|\beta A_2(x_2^k - x_2^{k+1}) + (\lambda^k - \lambda^{k+1})\|^2$$
$$= \frac{2-\alpha}{\alpha^2\beta}\left\|\begin{array}{c} x_2^k - x_2^{k+1} \\ \lambda^k - \lambda^{k+1} \end{array}\right\|_{\bar{G}}^2$$
$$= \frac{2-\alpha}{\alpha^2\beta}\left\|v^k - v^{k+1}\right\|_{\bar{G}}^2,$$

where

$$\bar{G} = \begin{pmatrix} \beta^2 A_2^\top A_2 & \beta A_2^\top \\ \beta A_2 & I_\ell \end{pmatrix}$$

And substituting the last equality (34) into (33), we get

$$\left\|v^{k+1} - v^*\right\|_{\mathcal{T}_\alpha}^2 \le \left\|v^k - v^*\right\|_{\mathcal{T}_\alpha}^2 - \frac{2-\alpha}{\alpha^2\beta}\left\|v^k - v^{k+1}\right\|_{\bar{G}}^2. \tag{35}$$

On the other hand, seting $x_2' := x_2^k$ in (23), there exists $\zeta \in \partial f_2(x_2^{k+1})$ such that

$$(x_2^k - x_2^{k+1})^\top(\zeta - A_2^\top\lambda^{k+1}) \ge 0. \tag{36}$$

Moreover, since x_2^k is the solution of (4) at the $(k-1)$th iteration and $x_2^{k+1} \in \mathcal{X}_2$, there exists $\hat{\zeta} \in \partial f_2(x_2^k)$ such that

$$(x_2^{k+1} - x_2^k)^\top(\hat{\zeta} - A_2^\top\lambda^k) \ge 0. \tag{37}$$

Adding (36) and (37) and using the monotonicity of ∂f_2 , we have

$$(A_2x_2^k - A_2x_2^{k+1})^\top(\lambda^k - \lambda^{k+1}) \ge 0. \tag{38}$$

whic implies

$$\begin{aligned} \left\|v^k - v^{k+1}\right\|_{\bar{G}}^2 &= \|\beta A_2(x_2^k - x_2^{k+1}) + (\lambda^k - \lambda^{k+1})\|^2 \\ &\ge \beta^2\|A_2x_2^k - A_2x_2^{k+1}\|^2 + \|\lambda^k - \lambda^{k+1}\|^2 \\ &= \beta\left\|v^k - v^{k+1}\right\|_{\mathcal{T}_1}^2, \end{aligned} \tag{39}$$

where $\mathcal{T}_1$ is defined in (14). By using (35) and the above equality, the assertion (29) is proved.

From inequality (34) and (38), we have

$$\begin{aligned}
&\left\| \begin{array}{c} x_2^k - x_2^{k+1} \\ \lambda^k - \bar{\lambda}^k \end{array} \right\|_{\mathcal{T}_\alpha}^2 - \left\| \begin{array}{c} 0 \\ \lambda^{k+1} - \bar{\lambda}^k \end{array} \right\|_{\mathcal{T}_\alpha}^2 \\
&= \frac{2-\alpha}{\alpha^2\beta}\|\beta A_2(x_2^k - x_2^{k+1}) + (\lambda^k - \lambda^{k+1})\|^2 \\
&= \frac{2-\alpha}{\alpha}(\frac{\beta}{\alpha}\|A_2x_2^k - A_2x_2^{k+1}\|^2 + \frac{1}{\alpha\beta}\|\lambda^k - \lambda^{k+1}\|^2) \\
&\quad + \frac{2-\alpha}{\alpha}\cdot\frac{2(1-\alpha)}{\alpha}(\lambda^k - \lambda^{k+1})^\top A_2(x_2^k - x_2^{k+1}) \\
&\quad + \frac{2(2-\alpha)}{\alpha}(\lambda^k - \lambda^{k+1})^\top A_2(x_2^k - x_2^{k+1}) \\
&\geq \frac{2-\alpha}{\alpha}(\frac{\beta}{\alpha}\|A_2x_2^k - A_2x_2^{k+1}\|^2 + \frac{1}{\alpha\beta}\|\lambda^k - \lambda^{k+1}\|^2) \\
&\quad + \frac{2-\alpha}{\alpha}\cdot\frac{2(1-\alpha)}{\alpha}(\lambda^k - \lambda^{k+1})^\top A_2(x_2^k - x_2^{k+1}) \\
&= \frac{2-\alpha}{\alpha}\left\|v^k - v^{k+1}\right\|_{\mathcal{T}_\alpha}^2 .
\end{aligned}$$

By substituting the above equality into (33), then the assertion (30) is proved.

Remark 3.2. (i) If $f_1(x_1) = \frac{1}{2}x_1^\top Qx_1 + q^\top x_1$ and $f_2(x_2) = \frac{1}{2}x_2^\top Rx_2 + r^\top x_2$, then convex optimization problem (1) becomes the quadrtic program (1.4) in [12] and by Theorem 3.1, we recover Theorem 3.1 in [12] with the coefficient $\frac{2-\alpha}{\alpha^2\beta}$ of H_1 being replaced by $\frac{2-\alpha}{\alpha^2}$.

(ii) If $\alpha = 1$, then by Theorem 3.1, we can obtain the following Corollary 3.1 which is slightly diffirent with Theorem 8 in [14]:

Corollary 3.1. If A_2 is full column rank, and the sequence $\{u^k\}$ is generated by the ADMM scheme (2), then we have:

$$\left\|v^{k+1} - v^*\right\|_{\mathcal{T}_1}^2 \leq \left\|v^k - v^*\right\|_{\mathcal{T}_1}^2 - \left\|v^k - v^{k+1}\right\|_{\mathcal{T}_1}^2, \ \forall \ \ v^* \in \Omega^*.$$

Yang and Han [14] proved the the global convergence of ADMM scheme (2). By combine the proof of Theorem 8 in [14] with Theorem 3.1, we can easily obtain the following global linear convergence of the GADMM and its proof is omitted here.

Theorem 3.2. Assume that $\alpha \in (0,2)$, A and B are full column rank, then the sequence $\{u^k\}$ generated by the GADMM scheme (3)-(5) converges to a solution point $u^* \in \mathcal{U}^*$.

4. Local linear convergence of the GADMM. Han and Yuan [12] proved the local linear convergence rate of GADMM scheme (3)-(5) for a quadrtic program. In this section, we will establish the local linear convergence rate of GADMM for a convex optimization problem (1) under the assumption that ∂f_1 and ∂f_2 are piecewise linear multifunctions.

Lemma 4.1. Let $\alpha \in (0,2)$, A_2 be full column rank, and $\{u^k\}$ be the sequence generated by the GADMM scheme (3)-(5) for (1). Then we have

$$\|v^{k+1} - v^k\|_{\mathcal{T}_1}^2 \geq \frac{\lambda_{\min}(\mathcal{T}_1)\alpha^2}{\lambda_{\max}(G_\alpha)} \mathrm{dist}^2(0, e(u^{k+1}, 1)). \tag{40}$$

Proof. Let $\bar{\lambda}^k$ be defined (17). From (20), we know that there exists $\xi \in \partial f(x^{k+1})$ such that

$$x_1^{k+1} = P_{\mathcal{X}_1}(x_1^{k+1} - (\xi - A_1^\top \bar{\lambda}^k)). \tag{41}$$

Thus, we have

$$\begin{aligned} &\mathrm{dist}(0, e_{\mathcal{X}}(u^{k+1}, 1)) \\ =&\mathrm{dist}(x_1^{k+1}, P_{\mathcal{X}_1}(x_1^{k+1} - (\partial f_1(x_1^{k+1}) - A_1^\top \lambda^{k+1}))) \\ \leq&\|P_{\mathcal{X}_1}(x_1^{k+1} - (\xi - A_1^\top \bar{\lambda}^k)) - P_{\mathcal{X}_1}(x_1^{k+1} - (\xi - A_1^\top \lambda^{k+1}))\| \\ \leq&\|A_1^\top(\bar{\lambda}^k - \lambda^{k+1})\|. \end{aligned} \tag{42}$$

Using (22) and (17), we get

$$\lambda^{k+1} - \bar{\lambda}^k = \beta(A_2 x_2^k - A_2 x_2^{k+1}) + \beta(1-\alpha)(A_1 x_2^{k+1} + A_2 x_2^k - b). \tag{43}$$

It follows from (5) that

$$A_1 x_1^{k+1} + A_2 x_2^k - b = \frac{1}{\alpha\beta}(\lambda^k - \lambda^{k+1}) - \frac{1}{\alpha}(A_2 x_2^{k+1} - A_2 x_2^k). \tag{44}$$

By (43) and (44), we get

$$\bar{\lambda} - \lambda^{k+1} = \frac{\beta}{\alpha}(A_2 x_2^{k+1} - A_2 x_2^k) - \frac{1-\alpha}{\alpha}(\lambda^k - \lambda^{k+1}). \tag{45}$$

From (23), we know that there exists $\zeta \in \partial f_2(x_2^{k+1})$ such that

$$x_2^{k+1} = P_{\mathcal{X}_2}(x_2^{k+1} - (\zeta - A_2^\top \lambda^{k+1})).$$

Thus,

$$\mathrm{dist}(0, e_{\mathcal{X}_2}(u^{k+1}, 1)) = \mathrm{dist}(x_2^{k+1}, P_{\mathcal{X}_2}(x_2^{k+1} - (\zeta - B^\top \lambda^{k+1}))) = 0. \tag{46}$$

It is easy to see that

$$\begin{aligned} \mathrm{dist}(0, e_\Lambda(u^{k+1}, 1)) =&\|e_\Lambda(u^{k+1}, 1)\| \\ =&\|A_1 x_1^{k+1} + A_2 x_2^{k+1} - b\| \\ =&\|\frac{1}{\alpha\beta}(\lambda^k - \lambda^{k+1}) - \frac{1-\alpha}{\alpha}(A_2 x_2^{k+1} - A_2 x_2^k)\|. \end{aligned} \tag{47}$$

By (42), and (45)-(47), we have

$$
\begin{aligned}
&\mathrm{dist}^2(0, e(u^{k+1}, 1))\\
=&\mathrm{dist}^2(0, e_{\mathcal{X}_1}(u^{k+1}, 1)) + \mathrm{dist}^2(0, e_{\mathcal{X}_2}(u^{k+1}, 1)) + \mathrm{dist}^2(0, e_{\Lambda}(u^{k+1}, 1))\\
\leq&\frac{1}{\alpha^2}\|A_1^\top[\beta(A_2x_2^{k+1} - A_2x_2^k) + (1-\alpha)(\lambda^{k+1} - \lambda^k)]\|^2\\
&+\frac{1}{\alpha^2}\|(1-\alpha)(A_2x_2^{k+1} - A_2x_2^k) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\|^2\\
=&\frac{1}{\alpha^2}\left\|v^k - v^{k+1}\right\|_{G_\alpha}^2 .
\end{aligned} \tag{48}
$$

Recall the definitions of $\mathcal{T}_1$ and G_α. We thus have

$$
\left\|v^k - v^{k+1}\right\|_{\mathcal{T}_1}^2 \geq \frac{\lambda_{\min(\mathcal{T}_1)}}{\lambda_{\max(G_\alpha)}}\left\|v^k - v^{k+1}\right\|_{G_\alpha}^2 \geq \frac{\alpha^2\lambda_{\min(\mathcal{T}_1)}}{\lambda_{\max(G_\alpha)}}\mathrm{dist}^2(0, e(u^{k+1}, 1)), \tag{49}
$$

which is assertion (40). The proof is completed.

It is worthy noting that if the subdifferentials ∂f_1 and ∂f_2 are piecewise linear multifunctions, then $e(u,\gamma)$ defined in (12) is also a piecewise linear multifunction of u. From Lemma 2.2, we can obtain the following local error bound, which play an key role in the proof of local linear convergence rate of GADMM.

Lemma 4.2. Let the subdifferentials ∂f_1 and ∂f_2 be piecewise linear multifunctions, $\mathcal{U}^*$ be the solution set of SVI$(\mathcal{U}, L)$ in (11). Then there exist scalars $\kappa, \tau > 0$ such that

$$
\mathrm{dist}(u, \mathcal{U}^*) \leq \tau \mathrm{dist}(0, e(u, 1)). \tag{50}
$$

whenever $u \in \mathcal{U}$ and $\mathrm{dist}(0, e(u, 1)) \leq \kappa$.

Lemma 4.3. Let $\alpha \in (0, 2)$, A_2 be full column rank, and $\{u^k\}$ be the sequence generated by the GADMM scheme (3)-(5). Then there exist scalars $\kappa, \tau > 0$ such that for any $v^* \in \Omega^*$, we have

$$
\left\|v^{k+1} - v^*\right\|_{\mathcal{T}_\alpha}^2 \leq \left\|v^k - v^*\right\|_{\mathcal{T}_\alpha}^2 - \frac{(2-\alpha)\lambda_{\min}(\mathcal{T}_1)}{\tau^2\lambda_{\max}(M_\alpha)\lambda_{\max}(G_\alpha)} \cdot \mathrm{dist}_{\mathcal{T}_\alpha}^2(v^{k+1}, \Omega^*), \tag{51}
$$

whenever $\mathrm{dist}(0, e(u^{k+1}, 1)) \leq \kappa$.

Proof. It follows from Lemma 4.2 that there exist scalars $\kappa, \tau > 0$ such that

$$
\begin{aligned}
\mathrm{dist}^2(0, e(u^{k+1}, 1) &\geq \frac{1}{\tau^2}\mathrm{dist}^2(u^{k+1}, \mathcal{U}^*)\\
&\geq \frac{1}{\tau^2\lambda_{\max}(M_\alpha)} \cdot \mathrm{dist}_{M_\alpha}^2(u^{k+1}, \mathcal{U}^*)\\
&\geq \frac{1}{\tau^2\lambda_{\max}(M_\alpha)} \cdot \mathrm{dist}_{\mathcal{T}_\alpha}^2(v^{k+1}, \Omega^*).
\end{aligned} \tag{52}
$$

whenever $\mathrm{dist}(0, e(u^{k+1}, 1)) \leq \kappa$.

Combining the above inequality with (40), we can obtain

$$
\left\|v^k - v^{k+1}\right\|_{\mathcal{T}_1}^2 \geq \frac{\alpha^2\lambda_{\min}(\mathcal{T}_1)}{\tau^2\lambda_{\max}(M_\alpha)\lambda_{\max}(G_\alpha)} \cdot \mathrm{dist}_{H_\alpha}^2(v^{k+1}, \Omega^*), \tag{53}
$$

which, together with (29) in Theorem 3.1, implies the assertion (50) immediately.

Now, we are ready to present the main result of the local linear convergence rate of GADMM scheme (3)-(5) for solving (1).

Theorem 4.1. If $\alpha \in (0,2)$, A_2 is full column rank, the subdifferentials ∂f_1 and ∂f_2 are piecewise linear multifunctions and $\{u^k\}$ is the sequence generated by the GADMM scheme (3)-(5), then there exist scalars $\kappa, \tau > 0$ such that

$$\mathrm{dist}^2_{\mathcal{T}_\alpha}(v^{k+1}, \Omega^*) \le \frac{1}{1+\delta} \cdot \mathrm{dist}^2_{\mathcal{T}_\alpha}(v^k, \Omega^*), \tag{54}$$

whenever $\mathrm{dist}(0, e(u^{k+1}, 1)) \le \kappa$, where

$$\delta := \frac{(2-\alpha)\lambda_{\min}(\mathcal{T}_1)}{\tau^2 \lambda_{\max}(M_\alpha)\lambda_{\max}(G_\alpha)} > 0.$$

Proof. Obviously, for any $(y^*, \lambda^*) \in \Omega^*$, by using the definition of $\mathrm{dist}^2_{H_\alpha}(v^{k+1}, \Omega^*)$ and Lemma 4.3, there exist scalars $\kappa, \tau > 0$ such that

$$\begin{aligned} \mathrm{dist}^2_{\mathcal{T}_\alpha}(v^{k+1}, \Omega^*) &\le \left\| v^{k+1} - v^* \right\|^2_{\mathcal{T}_\alpha} \\ &\le \left\| v^k - v^* \right\|^2_{\mathcal{T}_\alpha} - \frac{(2-\alpha)\lambda_{\min}(\mathcal{T}_1)}{\tau^2 \lambda_{\max}(M_\alpha)\lambda_{\max}(G_\alpha)} \cdot \mathrm{dist}^2_{\mathcal{T}_\alpha}(v^{k+1}, \Omega^*), \end{aligned} \tag{55}$$

whenever $\mathrm{dist}(0, e(u^{k+1}, 1)) \le \kappa$.

Since the above inequality holds for any $(x_2^*, \lambda^*) \in \Omega^*$ and $\alpha \in (0,2)$, we get

$$\mathrm{dist}^2_{\mathcal{T}_\alpha}(v^{k+1}, \Omega^*) \le \mathrm{dist}^2_{\mathcal{T}_\alpha}(v^k, \Omega^*) - \frac{(2-\alpha)\lambda_{\min}(\mathcal{T}_1)}{\tau^2 \lambda_{\max}(M_\alpha)\lambda_{\max}(G_\alpha)} \cdot \mathrm{dist}^2_{\mathcal{T}_\alpha}(v^{k+1}, \Omega^*),$$

and the above inequality implies the assertion (53) immediately. This completes the proof.

5. global linear convergence of GADMM. Yang and Han [14] proved the global linear convergence rate of classical ADMM scheme (2) for a convex optimization problem (1) under the assumption that ∂f_1 and ∂f_2 are piecewise linear multifunctions. In this section we will establish the global linear convergence rate of GADMM scheme (3)-(5) for the same convex minimization problem in [14].

Lemma 5.1. If $\alpha \in (0,2)$, A_2 is full column rank and $\{u^k\}$ is the sequence generated by the GADMM scheme (3)-(5), then we have

$$\left\| v^{k+1} - v^k \right\|_{\mathcal{T}_\alpha} \ge \varrho \mathrm{dist}(0, e(u^{k+1}, 1)), \tag{56}$$

where

$$\varrho = \sqrt{\frac{\alpha\beta}{1 + \beta^2 \|A_1^\top\|^2}} > 0. \tag{57}$$

Proof. From (48), we have

$$
\begin{aligned}
&\operatorname{dist}^2(0, e(u^{k+1}, 1)) \\
\leq&\frac{1}{\alpha^2}\|A_1^\top[\beta(A_2x_2^{k+1} - A_2x_2^k) + (1-\alpha)(\lambda^{k+1} - \lambda^k)]\|^2 \\
&+ \frac{1}{\alpha^2}\|(1-\alpha)(A_2x_2^{k+1} - A_2x_2^k) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\|^2 \\
=&\frac{\|A_1^\top\|^2}{\alpha^2}[\beta^2\|A_2x_2^k - A_2x_2^{k+1}\|^2 + (1-\alpha)^2\|\lambda^k - \lambda^{k+1}\|^2 \\
&+ 2\beta(1-\alpha)A_2(x_2^k - x_2^{k+1})(\lambda^k - \lambda^{k+1})] \\
&+ \frac{1}{\alpha^2}[(1-\alpha)^2\|A_2x_2^k - A_2x_2^{k+1}\|^2 + \frac{1}{\beta^2}\|\lambda^k - \lambda^{k+1}\|^2 \\
&+ \frac{2(1-\alpha)}{\beta}(\lambda^k - \lambda^{k+1})^\top A_2(x_2^k - x_2^{k+1})] \\
=&\frac{\beta^2\|A_1^\top\|^2 + (1-\alpha)^2}{\alpha\beta}\cdot\frac{\beta}{\alpha}\|A_2x_2^k - A_2x_2^{k+1}\|^2 \\
&+ \frac{\beta^2(1-\alpha)^2\|A_1^\top\|^2 + 1}{\alpha\beta}\cdot\frac{1}{\alpha\beta}\|\lambda^k - \lambda^{k+1}\|^2 \\
&+ \frac{\beta^2\|A_1^\top\|^2 + 1}{\alpha\beta}\cdot\frac{2(1-\alpha)}{\alpha}(\lambda^k - \lambda^{k+1})^\top A_2(x_2^k - x_2^{k+1}) \\
=&(\frac{\beta^2\|A_1^\top\|^2 + (1-\alpha)^2}{\beta^2\|A_1^\top\|^2 + 1}\cdot\frac{\beta^2\|A_1^\top\|^2 + 1}{\alpha\beta})\cdot\frac{\beta}{\alpha}\|A_2x_2^k - A_2x_2^{k+1}\|^2 \\
&+ (\frac{\beta^2(1-\alpha)^2\|A_1^\top\|^2 + 1}{\beta^2\|A_1^\top\|^2 + 1}\cdot\frac{\beta^2\|A_1^\top\|^2 + 1}{\alpha\beta})\cdot\frac{1}{\alpha\beta}\|\lambda^k - \lambda^{k+1}\|^2 \\
&+ \frac{\beta^2\|A_1^\top\|^2 + 1}{\alpha\beta}\cdot\frac{2(1-\alpha)}{\alpha}(\lambda^k - \lambda^{k+1})^\top A_2(x_2^k - x_2^{k+1}).
\end{aligned}
\tag{58}
$$

And for any $\alpha \in (0, 2)$, we have

$$
\frac{\beta^2\|A_1^\top\|^2 + (1-\alpha)^2}{\beta^2\|A_1^\top\|^2 + 1} \leq 1, \tag{59}
$$

and

$$
\frac{\beta^2(1-\alpha)^2\|A_1^\top\|^2 + 1}{\beta^2\|A_1^\top\|^2 + 1} \leq 1. \tag{60}
$$

By (58)-(60) and the definition of $\mathcal{T}_\alpha$ in (13), we have

$$
\operatorname{dist}^2(0, e(u^{k+1}, 1)) \leq \frac{\beta^2\|A_1^\top\|^2 + 1}{\alpha\beta}\cdot\left\|v^k - v^{k+1}\right\|_{\mathcal{T}_\alpha}^2,
$$

which implies that (56) holds.

According to Theorem 3.2, we know the sequence $\{u^k\}$ generated by the GADM-M scheme (3)-(5) is bounded. Then there exists a bounded set U such that $\{u^k\} \subset U$.

Lemma 5.2.[14] Let U be a bound set in $\mathcal{R}^n$. Then there exists $\eta > 0$ such that

$$\operatorname{dist}(u, \mathcal{U}^*) \leq \eta \operatorname{dist}(0, e(u,1)), \qquad \forall u \in U. \tag{61}$$

Now, we show the global linear convergence rate of the GADMM as follows:

Theorem 5.1. Assume that $\alpha \in (0,2)$, A_1 and A_2 are full column rank, the subdifferentials ∂f_1 and ∂f_2 are piecewise linear multifunctions and $\{u^k\}$ is the sequence generated by the GADMM scheme (3)-(5). Then

$$\operatorname{dist}^2_{\mathcal{T}_\alpha}(v^{k+1}, \Omega^*) \leq \frac{1}{1+\varepsilon^2} \operatorname{dist}^2_{\mathcal{T}_\alpha}(v^k, \Omega^*), \tag{62}$$

where $\varepsilon := \frac{\varrho \sqrt{\frac{2-\alpha}{\alpha}}}{\lambda_{\max}(\mathcal{T}_\alpha)\eta}$ and ϱ is defined in (57).

Proof. By Theorem 3.2, we know the sequence $\{u^k\}$ is bounded. Then there exists a bounded set U such that $\{u^k\} \subseteq U$. According to Lemmas 5.1 and 5.2, there exists $\eta > 0$ such that

$$\begin{aligned} & \left\|v^k - v^{k+1}\right\|_{\mathcal{T}_\alpha} \\ \geq & \varrho \operatorname{dist}(0, e(u^{k+1}, 1)) \\ \geq & \frac{\varrho}{\eta} \operatorname{dist}(u^{k+1}, \mathcal{U}^*) \\ \geq & \frac{\varrho}{\eta} \operatorname{dist}(v^{k+1}, \Omega^*) \\ \geq & \frac{\varrho}{\lambda_{\max}(\mathcal{T}_\alpha)\eta} \operatorname{dist}_{\mathcal{T}_\alpha}(v^{k+1}, \Omega^*), \end{aligned} \tag{63}$$

By selecting $v^* \in \Omega^*$, such that

$$\operatorname{dist}_{\mathcal{T}_\alpha}(v^k, \Omega^*) = \|v^k - v^*\|_{\mathcal{T}_\alpha}.$$

From (63) and the above equality, we have

$$\begin{aligned} (1+\varepsilon^2)\operatorname{dist}^2_{\mathcal{T}_\alpha}(v^{k+1}, \Omega^*) &= \|v^{k+1} - v^*\|^2_{\mathcal{T}_\alpha} + \varepsilon^2 \operatorname{dist}^2_{H_\alpha}(v^{k+1}, \Omega^*) \\ &\leq \|v^{k+1} - v^*\|^2_{\mathcal{T}_\alpha} + \frac{2-\alpha}{\alpha}\|v^{k+1} - v^k\|^2_{\mathcal{T}_\alpha}, \end{aligned} \tag{64}$$

where $\varepsilon := \frac{\varrho \sqrt{\frac{2-\alpha}{\alpha}}}{\lambda_{\max}(\mathcal{T}_\alpha)\eta}$ and $\mathcal{T}_\alpha$ and ϱ is defined in (13) and (57), respectively.
By (30) in Theorem 3.1, we have

$$\|v^{k+1} - v^*\|^2_{\mathcal{T}_\alpha} + \frac{2-\alpha}{\alpha}\|v^{k+1} - v^k\|^2_{\mathcal{T}_\alpha} \leq \|v^k - v^*\|^2_{\mathcal{T}_\alpha} = \operatorname{dist}^2_{\mathcal{T}_\alpha}(v^k, \Omega^*).$$

By (64) and the abpve inequality, we get

$$(1+\varepsilon^2)\operatorname{dist}^2_{\mathcal{T}_\alpha}(v^{k+1}, \Omega^*) \leq \operatorname{dist}^2_{\mathcal{T}_\alpha}(v^k, \Omega^*) \tag{65}$$

The assertion (62) is prove by using the above inequality.

6. Numerical experiments. In this section, we implement the GADMM (3)-(5) on a popular LASSO model to show its linear convergence and we also compare its numerical performance with the ADMM (2). The problem under consideration is of the following form:

$$\begin{array}{ll} \min_{x_1,x_2\in\mathcal{R}^n} & \frac{1}{2}\|Ex_1-d\|^2+\nu\|x_2\|_1, \\ s.t. & x_1=x_2, \end{array} \tag{66}$$

where $E\in\mathcal{R}^{\ell\times n}$ is a given matrix, $d\in\mathcal{R}^l$ is a given vector, ν is a positive scalar and $\|x_2\|_1:=\sum_{i=1}^n|x_2^i|$. Clearly, the above optimization model (66) is the equivalent form of (6). We now numerically verify the global linear convergence of the GADMM. In our experiments, we set $\nu=1$ and generate 100×1500 matrices E and corresponding vectors $d\in\mathcal{R}^{100}$, where each element of E and d is independently sampled from the standard Gaussian distribution.

All experiments are implemented in MATLAB R2010b on a hp-notebook with an Intel Core i5-3340M CPU at 2.70 GHz and 8 GB memory.

In Figure 1 and Figure 2, the iteration number k is given in the x-axis, the y-axis shows the the value of $\mathrm{dist}_{H_\alpha}(v^k,\Omega^*)$, where v^k is the iterate generated by the ADMM (2) or GADMM (3)-(5), $\mathcal{U}^*$ is the KKT solution set of (66) and $\Omega^*=\{(x_2^*,\lambda^*)|(x_1^*,x_2^*,\lambda^*)\in\mathcal{U}^*\}$. We first compare the GADMM (3)-(5) with ADMM (2) for their averaged performances. To further observe the linear convergence of the two tested algorithms, in Figure 1, we plot the evolutions of the error when the two methods are applied to solve (66). Here, we set $\beta=1$, $\alpha=1.5$ for GADMM (3)-(5).

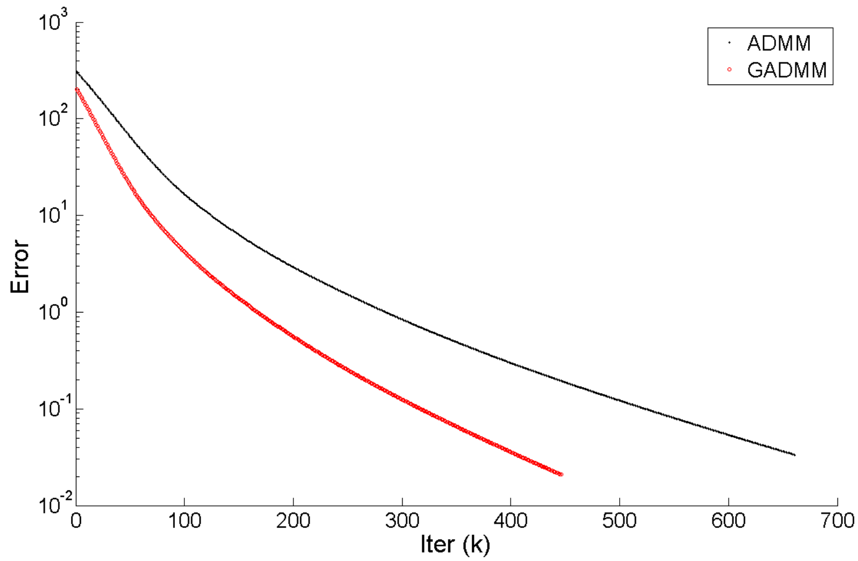

FIG. 1. The numerical comparison of GADMM and ADMM

From Figure 1, we observe that the GADMM scheme (3)-(5) provides faster convergence than that provided by ADMM scheme (2) when $\alpha=1.5$.

Then, we test the sensitivities of β and α for the GADMM (3)-(5). First of all we fix $\alpha=1.5$ and choose different values of β. Then we set $\beta=1$ and choose different values of α in the interval $(0, 2)$. We repeat each scheme ten times and report some numerical results for the above experiment, and we plot them in Figure 2.

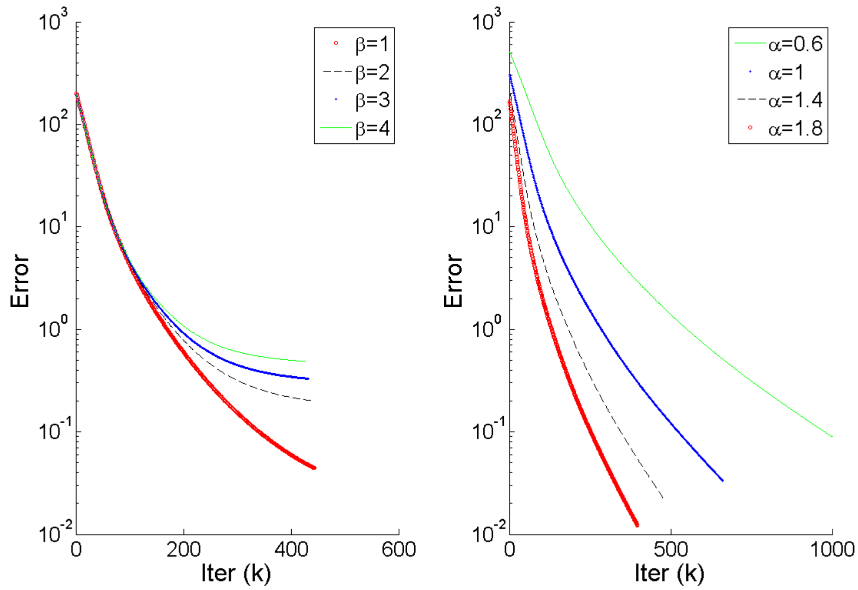

FIG. 2. Comparison results of GADMM on (66) for different β and α

From Figure 2, we can also observe that the sequence $\{v^k\}$ generated by the GADMM scheme (3)-(5) converges to Ω^* linearly. On the one hand, the selection of β can affect the convergence rate of the GADMM significantly when α is fixed. In this case, we observe that If β is very large, the constant $\frac{1}{1+\varepsilon^2}$ in (62) is close to 1, which results in the low accuracy of the solution. On the other hand, the selection of α can also affect the convergence rate of the GADMM significantly when β is fixed. In this case, From our experiments, we know that for the relaxation factor $\alpha \in (0, 2)$, the acceleration performance of (GADMM) is better if α is larger.

7. conclusions. In this paper, we considered the linear convergence rate of the generalized alternating direction method of multipliers (GADMM) for solvinging the sum of two separable convex functions with linear constraints. Firstly, under the assumption that the subdifferentials of the underlying functions are piecewise linear multifunctions and A_2 is full colume rank, we establish the local linear convergence rate of the GADMM which refines and extends the main results in [12]. Secondly, under the condition that the subdifferentials of the underlying functions are piecewise linear multifunctions, A_1 and A_2 are full colume rank, we also show the global linear convergence rate of the GADMM.

REFERENCES

[1] R. Glowinski and A. Marrocco, Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problémes non linéaires, M2AN Math. Model. Numer. Anal., (1975), pp. 41-76.

[2] D. Gabay, B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximations, Comput. Math. Appl. 2(1976) 17-40.
[3] Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. 4, 302-320 (1969).
[4] Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization. Academic Press (1969).
[5] 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.
[6] P. L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal., 16 (1979), pp. 964-979.
[7] 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.
[8] B. Martinet, Regularision d'inéquations variationnelles par approximations successive, Rev. Franc. Autom. Inf. Rech. Oper., 126 (1970), pp. 154-159.
[9] 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.
[10] J. Eckstein and D. P. Bertsekas, An Alternating Direction Method for Linear Programming, Report LIDS-P-1967, MIT, Cambridge, MA, 1990.
[11] D. Boley, Local linear convergence of ADMM on quadratic or linear programs, SIAM J. Optim., 23 (2013), pp. 2183-2207.
[12] D. Han and X. Yuan, Local linear convergence of the alternating direction method of multipliers for quadratic programs, SIAM J. Numer. Anal., 51 (2013), pp. 3446-3457.
[13] M. Tao and X. Yuan, On Glowinski's Open Question of Alternating Direction Method of Multipliers, optimization-online, http://www.optimization-online.org/DB_FILE/2017/06/6084.
[14] W. H. Yang and D. Han, linear convergence of the alternating direction method of multipliers for a class of convex optimization priblems, SIAM J. NUMER. ANAL., 54 (2016), pp. 625-640.
[15] M. Hong and Z. Q. Luo, On the Linear Convergence of Alternating Direction Method of Multipliers, Math. Program., Ser. A 162 (2017), pp. 165-199.
[16] B. He, M. Tao and X. Yuan, Convergence Rate Analysis for the Alternating Direction Method of Multipliers with a Substitution Procedure for Separable Convex Programming, Math. Oper. Res., 42(3)2017, pp. 662¨C691.
[17] W. Deng and W. T. Yin, On the Global and Linear Convergence of the Generalized Alternating Direction Method of Multipliers, J Sci Comput 66 (2016), pp. 889-916.
[18] R. Nishihara, L. Lessart, B. Recht, A. Packard, and M. I. Jordan, A general analysis of the convergence of ADMM, in Proceedings of the 32nd International Conference on Machine Learning (ICML), Lille, France, F. Bach and D. Blei, eds., 2015.
[19] T. Lin, S. Ma, and S. Zhang, On the Global Linear Convergence of the ADMM with Multiblock Variables, SIAM J. Optim., 25 (2015), pp. 1478-1497.
[20] X. Cai, D. Han, X. Yuan, On the convergence of the direct extension of ADMM for three-block separable convex minimization models with one strongly convex function, Comput Optim Appl 66(2017), pp. 39-73.
[21] D. Han, D, Sun and L. Zhang, Linear Rate Convergence of the Alternating Direction Method of Multipliers for Convex Composite Programming, Math. Oper. Res., Preprint.
[22] X. Y. Zheng and K. F. Ng, Metric subregularity of piecewise linear multifunctions and applications to piecewise linear multiobjective optimization, SIAM J. Optim., 24 (2014), pp. 154-174.
[23] S. M. Robinson, Some continuity properties of polyhedral multifunctions, Math. Program. Stud., 14 (1981), pp. 206-214.
[24] J. Sun, On Monotropic Piecewise Qudratic Programming, PhD Thesis, Department of Mathematics, University of Washington, Seattle, 1986.
[25] R. T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
[26] F. Facchinei and J. S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, Vol. I, Berlin, Springer Verlag, 2003.