On the pointwise iteration-complexity of a dynamic regularized ADMM with over-relaxation stepsize

M.L.N. Gonçalves *

May 8, 2017

Abstract

In this paper, we extend the improved pointwise iteration-complexity result of a dynamic regularized alternating direction method of multipliers (ADMM) for a new stepsize domain. In this complexity analysis, the stepsize parameter can even be chosen in the interval $(0, 2)$ instead of interval $(0, (1+\sqrt{5})/2)$. As usual, our analysis is established by interpreting this ADMM variant as an instance of a hybrid proximal extragradient framework applied to a specific monotone inclusion problem.

2000 Mathematics Subject Classification: 47H05, 49M27, 90C25, 90C30, 90C60, 65K10.

Key words: alternating direction method of multipliers, hybrid proximal extragradient framework, pointwise iteration-complexity, convex programming.

1 Introduction

We are interested in the following linearly constrained convex problem

$$\min\{f(x) + g(y) : Ax + By = b,\ x \in \mathbb{R}^n, y \in \mathbb{R}^p\} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^p \to \mathbb{R}$ are convex functions, $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times p}$ and $b \in \mathbb{R}^m$. We assume that the solution set of (1) is nonempty. Convex optimization problems with a separable structure such as (1) appear in many applications areas such as machine learning, compressive sensing and image processing. The augmented Lagrangian method (see, e.g., [1]) attempts to solve (1) directly without taking into account its particular structure. To overcome this drawback, a variant of the augmented Lagrangian method, namely, the alternating direction method of multipliers (ADMM), was proposed and studied in [7, 9]. The ADMM takes full advantage of the special structure of the problem by considering each variable separably in an alternating form and coupling them into the Lagrange multiplier updating; for detailed reviews, see [2, 8].

Recently, several variants of the ADMM for solving (1) have been proposed in the literature; see, for example, [3, 4, 5, 11, 12, 13, 14, 15, 16, 17, 22]. A dynamic regularized ADMM (DR-ADMM) with stepsize $\theta \in (0, (1+\sqrt{5})/2)$ was proposed by Gonçalves at al. [11] whose the pointwise iteration-complexity is substantially better than ones for the ADMMs. More specifically, for given $\rho > 0$, it was

*IME/UFG- Caixa Postal 131, CEP 74001-970, Goiânia-GO, Brazil. (E-mail: maxlng@ufg.br). The work of this author was supported in part by CNPq Grants 406250/2013-8, 444134/2014-0 and 309370/2014-0.

proved in [11] that the DR-ADMM finds a ρ-approximate solution of (1) in at most $\mathcal{O}\left(\rho^{-1}\log(\rho^{-1})\right)$ iterations. Although different criteria are used, in general the ADMM and its variants need $\mathcal{O}\left(\rho^{-2}\right)$ iterations to find this same approximate solution (see, e.g., [3, 4, 5, 12, 13, 14, 15, 16, 17, 19]). The main goal of this work is to extend the improved pointwise iteration-complexity result of the DR-ADMM obtained in [11] for a new stepsize domain $\theta \in (0,(1-\alpha+\sqrt{\alpha^2+6\alpha+5})/2)$, where α is a nonnegative proximal factor associated to the proximal term added to the second subproblem of the method (see the DR-ADMM in Section 3). Since the limit of $(\sqrt{\alpha^2+6\alpha+5}-\alpha)$ as α goes to infinity is 3, the latter stepsize domain becomes $(0,2)$ (resp. $(0,(1+\sqrt{5})/2)$) when α is sufficiently large (resp. $\alpha=0$). It is worth pointing out that the ADMM with a larger stepsize parameter can substantially improve the performance of the method in many applications (see [6, 8] for more details). As in [11], our complexity analysis is done by rewriting problem (1) as a monotone inclusion problem and by analyzing the DR-ADMM in the setting of a generalized hybrid proximal extragradient (HPE). It should be mentioned that paper [10] was the first one to discuss complexity results for the ADMM with stepsize $\theta \in (0,2)$ for solving non-convex linearly constrained problems and, subsequently, paper [14] studied convergence and complexity results for the ADMM with the same stepsize domain of this paper for the convex case.

Notation: The set of real numbers is denoted by $\mathbb{R}$. The set of non-negative real numbers and the set of positive real numbers are denoted by $\mathbb{R}_+$ and $\mathbb{R}_{++}$, respectively. For $t>0$, we let $\log^+(t) := \max\{\log t, 0\}$. For a finite-dimensional real vector space $\mathcal{X}$ with inner product $\langle\cdot,\cdot\rangle$, its induced norm is denoted by $\|\cdot\|$. Denote by $\mathcal{M}_+^{\mathcal{X}}$ the space of selfadjoint positive semidefinite linear operators on $\mathcal{X}$. For each $H \in \mathcal{M}_+^{\mathcal{X}}$, the seminorm induced by H on $\mathcal{X}$ is defined by $\|\cdot\|_H := \sqrt{\langle H(\cdot),\cdot\rangle}$.

2 Preliminaries results

In this section, we present a dynamic regularized HPE framework and its pointwise iteration-complexity result. This framework is an instance of one studied in [11].

Consider the monotone inclusion problem (MIP)

$$0 \in T(z) \tag{2}$$

where $\mathcal{Z}$ is a finite-dimensional real vector space and $T:\mathcal{Z} \rightrightarrows \mathcal{Z}$ is a maximal monotone operator [1]. We assume that the solution set of (2), denoted by $T^{-1}(0)$, is nonempty.

The dynamic regularized HPE framework attempts to solve the inclusion (2) by solving approximately a sequence of regularized MIP of the following form

$$0 \in T(z) + \mu M(z-z_0) \tag{3}$$

where $z_0 \in \mathcal{Z}$, $\mu>0$ and $M \in \mathcal{M}_+^{\mathcal{Z}}$ are fixed. We also assume that the solution set of (3)

$$\bar{Z}_\mu(M) := \{z \in \mathcal{Z} : \ 0 \ \in \ T(z) + \mu M(z-z_0)\} \tag{4}$$

is nonempty for every $\mu>0$. It can be shown that if M is positive definite, then the operator $T(\cdot)+\mu M(\cdot - z_0)$ is maximal μ-strongly monotone which in turn implies that the set $\bar{Z}_\mu(M)$ is

[1] An operator $T:\mathcal{Z} \rightrightarrows \mathcal{Z}$ is said to be monotone if $\langle z-z', s-s'\rangle \geq 0$, for every $z,z' \in \mathcal{Z}$, $s \in T(z)$ and $s' \in T(z')$. Moreover, T is maximal monotone if it is monotone and, additionally, if S is a monotone operator such that $T(z) \subset S(z)$ for every $z \in \mathcal{Z}$ then $T=S$.

nonempty for every $\mu > 0$ (see, e.g., [21, Corollary 12.44 and Proposition 12.54]). Moreover, the following relation between $\bar{Z}_\mu(M)$ and $T^{-1}(0)$ holds for every $\mu > 0$:

$$\|z_0 - \bar{z}_\mu\|_M \leq \|z_0 - \bar{z}\|_M \quad \forall \bar{z}_\mu \in \bar{Z}_\mu(M),\ \forall \bar{z} \in T^{-1}(0). \tag{5}$$

The above relation follows directly from [11, Lemma 3.1] with $(dw)_z(z') = (1/2)\|z' - z\|_M^2$ for every $z, z' \in \mathcal{Z}$.

Next, we present the dynamic regularized HPE framework for solving (2), which will be used in order to analyze the ADMM variant of Section 3.

Dynamic regularized HPE (DR-HPE) framework.

(0) Let $z_0 \in \mathcal{Z}$, $(\eta_0, \sigma, \tau, \rho) \in \mathbb{R}_+ \times [0,1) \times (0,1) \times \mathbb{R}_{++}$ and $M \in \mathcal{M}_+^{\mathcal{Z}}$ be given, and set $\mu = 1$ and $k = 1$;

(1) find $(z_k, \tilde{z}_k, \eta_k) \in \mathcal{Z} \times \mathcal{Z} \times \mathbb{R}_+$ such that

$$M(z_{k-1} - z_k) \in \left(T(\tilde{z}_k) + \mu M(\tilde{z}_k - z_0)\right), \tag{6}$$

$$\|z_k - \tilde{z}_k\|_M^2 + \eta_k \leq \sigma\|z_{k-1} - \tilde{z}_k\|_M^2 + (1-\tau)\eta_{k-1}; \tag{7}$$

(2) if $\|z_{k-1} - z_k\|_M \leq \rho/2$, then go to step 3; otherwise, set $k \leftarrow k+1$ and go to step 1.

(3) compute $v_k := z_{k-1} - z_k - \mu(\tilde{z}_k - z_0)$; if $\|v_k\|_M \leq \rho$, then stop and output $(\tilde{z}, v) \leftarrow (\tilde{z}_k, v_k)$; else, set $\mu \leftarrow \mu/2$ and $k = 1$, and go to step 1.

end

Remarks. 1) The DR-HPE framework corresponds to the framework 3 in [11] with $\lambda_k = 1$, $\varepsilon_k = 0$ and $(dw)_z(z') = (1/2)\|z' - z\|_M^2$ for every $z, z' \in \mathcal{Z}$. Now, if M is the identity operator and $\eta_k = 0$, it becomes the DR-HPE framework in [18] with $\lambda_k = 1$ and $\varepsilon_k = 0$. 2) The scalar μ plays the role of a regularization parameter which is dinamically adapted in order to control the term $M(\tilde{z}_k - z_0)$ in (6). 3) The DR-HPE framework is a general setting which does not specify how to obtain $(z_k, \tilde{z}_k, \eta_k)$ as in step 1. Specific computation of these elements will depend on implementation of particular instances of the framework and the properties of the operators T and M. 4) If M is positive definite and $\sigma = \eta_0 = 0$, then (7) implies that $\eta_k = 0$ and $z_k = \tilde{z}_k$ for every k, and then (6) reduces to an iteration of the proximal point method (in the metric $\|\cdot\|_M$) applied to (3).

The following result gives the pointwise iteration-complexity bound for the DR-HPE framework.

Theorem 2.1. *Suppose that $1/(1-\sigma)$ and $1/\tau$ are $\mathcal{O}(1)$. Then, the DR-HPE framework finds a pair $(\tilde{z}, v)$ satisfying $Mv \in T(\tilde{z})$ and $\|v\|_M \leq \rho$, in at most*

$$\mathcal{O}\left(\left(1 + \frac{\sqrt{d^2 + \eta_0}}{\rho}\right)\left[1 + \log^+\left(\frac{\sqrt{d^2 + \eta_0}}{\rho}\right)\right]\right)$$

iterations, where $d := \inf\left\{\|z_0 - z\|_M : z \in T^{-1}(0)\right\}$.

Proof. First of all, the DR-HPE framework is a special case of framework 3 in [11] where $\lambda_k = 1$, $\varepsilon_k = 0$ and $(dw)_z(z') = (1/2)\|z' - z\|_M^2$ for every $z, z' \in \mathcal{Z}$. Moreover, it is easy to see that the distance generating function $w(\cdot) = (1/2)\|\cdot\|_M^2$ is an $(1,1)$-regular with respect to $(\mathcal{Z}, \|\cdot\|_M)$ in the sense of [11, Definition 2.2]. Hence, the proof follows directly from [11, Theorem 3.3] (see also first remark after [11, Theorem 3.3]) with $M = m = \lambda = 1$, $\varepsilon_k = 0$, $d_0 = d^2/2$, $\tilde{r} = Mv$ and by taking into account the following property of the dual semi-norm $\|M(\cdot)\|_M^* = \|\cdot\|_M$ (see [11, Proposition A1]). □

3 DR-ADMM and its pointwise iteration-complexity

In this section, we recall the DR-ADMM for solving (1) and establish its pointwise iteration-complexity result for any stepsize $\theta \in (0, (1-\alpha+\sqrt{\alpha^2+6\alpha+5})/2)$, where α is a nonnegative proximal factor associated to the proximal term added to the second subproblem of the method.

The DR-ADMM for solving (1) is described as follows:

Dynamic regularized ADMM (DR-ADMM).

(0) Let an initial point $(x_0, y_0, \gamma_0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, positive parameters β and θ, a tolerance $\rho > 0$, a proximal factor $\alpha \geq 0$, and matrices $R \in \mathcal{M}_+^{\mathbb{R}^n}$ and $S \in \mathcal{M}_+^{\mathbb{R}^p}$ be given, and set $\mu = 1$ and $k = 1$;

(1) set $\beta_1 := \beta/(\theta+\mu)$, $\beta_2 := \beta(1+\mu)$, $\hat{x}_{k-1} = (x_{k-1} + \mu x_0)/(1+\mu)$ and $\hat{\gamma}_{k-1} := (\theta\gamma_{k-1} + \mu\gamma_0)/(\theta+\mu)$ and compute $x_k \in \mathbb{R}^n$ as

$$x_k \in \operatorname{argmin}_x \left\{ f(x) - \langle \hat{\gamma}_{k-1}, Ax \rangle + \frac{\beta_1}{2}\|Ax + By_{k-1} - b\|^2 + \frac{1+\mu}{2}\|x - \hat{x}_{k-1}\|_R^2 \right\}; \tag{8}$$

(2) set $\tilde{\gamma}_k := \hat{\gamma}_{k-1} - \beta_1(Ax_k + By_{k-1} - b)$, $\hat{y}_{k-1} := (y_{k-1} + \mu y_0)/(1+\mu)$ and $u_k := \tilde{\gamma}_k + \beta_2(Ax_k + B\hat{y}_{k-1} - b)$, and compute $(y_k, \gamma_k) \in \mathbb{R}^p \times \mathbb{R}^m$ as

$$y_k \in \operatorname{argmin}_y \left\{ g(y) - \langle u_k, By \rangle + \frac{\beta_2}{2}\left[\|Ax_k + By - b\|^2 + \alpha\|B(y - \hat{y}_{k-1})\|^2 + \frac{1}{\beta}\|y - \hat{y}_{k-1}\|_S^2 \right] \right\}, \tag{9}$$

$$\gamma_k := \gamma_{k-1} - \theta\beta\left[Ax_k + By_k - b + \mu(\tilde{\gamma}_k - \gamma_0)/(\beta\theta)\right]; \tag{10}$$

(3) If

$$\left(\|\Delta x_k\|_R^2 + (1+\alpha)\beta\|B\Delta y_k\|^2 + \|\Delta y_k\|_S^2 + (1/(\beta\theta))\|\Delta\gamma_k\|^2\right)^{1/2} \leq \rho/2, \tag{11}$$

where

$$\Delta x_k := x_{k-1} - x_k, \quad \Delta y_k := y_{k-1} - y_k, \quad \Delta\gamma_k := \gamma_{k-1} - \gamma_k, \tag{12}$$

then go to step 4; else set $k \leftarrow k+1$ and go to step 1;

(4) set $v_k^x := \Delta x_k - \mu(x_k - x_0)$, $v_k^y := \Delta y_k - \mu(y_k - y_0)$ and $v_k^\gamma := \Delta\gamma_k - \mu(\tilde{\gamma}_k - \gamma_0)$; if

$$\left(\|v_k^x\|_R^2 + (1+\alpha)\beta\|Bv_k^y\|^2 + \|v_k^y\|_S^2 + (1/(\beta\theta))\|v_k^\gamma\|^2\right)^{1/2} \leq \rho, \tag{13}$$

then stop and output $(x, y, \tilde{\gamma}, v^x, v^y, v^\gamma) \leftarrow (x_k, y_k, \tilde{\gamma}_k, v_k^x, v_k^y, v_k^\gamma)$; otherwise, set $\mu \leftarrow \mu/2$ and $k = 1$, and go to step 1.

end

Remarks. 1) The DR-ADMM is equivalent to the DR-ADMM in [11] with an appropriate choice of linear operator G. It should be noted, however, that the complexity result presented there does not establish any relationship between the stepsize θ and proximal term defined by G. 2) As in the DR-HPE framework, the scalar μ in the DR-ADMM can be seen as a regularization parameter. 3) Suitable choices of R and S may becomes the subproblems (8) and (9) easier to solve or even have a closed-form solutions (see [15, 23, 24] for more details). 4) For convenience, the term "cycle" will be used to refer to an execution of steps 1-3 of the DR-ADMM with a fixed μ.

In what follows, we show that the DR-ADMM with $\theta \in (0,(1-\alpha+\sqrt{\alpha^2+6\alpha+5})/2)$ is still a special case of the DR-HPE framework applied to a specific monotone inclusion problem. As a consequence, its pointwise iteration-complexity result will follows from Theorem 2.1.

Let us first deduce the aforementioned monotone inclusion problem. It is well known that a pair $(\bar{x},\bar{y})$ is a solution of (1) and $\bar{\gamma}$ is an associated Lagrange multiplier if and only if $(\bar{x},\bar{y},\bar{\gamma})$ satisfies

$$0 \in \partial f(\bar{x}) - A^*\bar{\gamma}, \quad 0 \in \partial g(\bar{y}) - B^*\bar{\gamma}, \quad A\bar{x} + B\bar{y} = b.$$

Since it is assumed that the solution set of (1) is nonempty, the existence of the Lagrange multipliers for problem (1) is guaranteed; see, for example, [20, Corollary 28.2.2]. Hence, we may solve (1) by means of obtaining a triple $(\bar{x},\bar{y},\bar{\gamma})$ satisfying the following monotone inclusion problem

$$0 \in T(x,y,\gamma) := \begin{bmatrix} \partial f(x) - A^*\gamma \\ \partial g(y) - B^*\gamma \\ Ax + By - b \end{bmatrix}. \tag{14}$$

In order to analyze the DR-ADMM in the setting of Section 2, consider the vector space $\mathcal{Z} := \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ and the following linear operator

$$Q := \begin{pmatrix} R & 0 & 0 \\ 0 & (1+\alpha)\beta B^*B + S & 0 \\ 0 & 0 & (\theta\beta)^{-1} I \end{pmatrix} : \mathcal{Z} \to \mathcal{Z} \tag{15}$$

where I is the $m \times m$ identity operator. We assume that the set $\bar{Z}_\mu(Q)$ as defined in (4) with $z_0 = (x_0, y_0, \lambda_0)$, T and Q as in (14) and (15), respectively, is nonempty for every $\mu > 0$. We mention that this assumption is not restrictive. Indeed, it is easy to see that a triple $(x,y,\gamma) \in \bar{Z}_\mu(Q)$ if and only if (x,y,γ) satisfies the inclusions

$$0 \in \partial f(x) - A^*\gamma + \mu R(x - x_0), \qquad 0 \in \partial g(y) - B^*\gamma + \mu[(1+\alpha)\beta B^*B(y-y_0) + S(y-y_0)],$$
$$0 = Ax + By - b + \mu(\gamma - \gamma_0)(\beta\theta)^{-1},$$

which is equivalent to the pair (x,y) be a solution and γ an associated Lagrange multiplier of the following optimization problem

$$\min_{(x,y,u)} \left\{ f(x) + g(y) + \frac{\mu}{2} \| (x - x_0, y - y_0, u(\theta\beta/\mu) + \gamma_0) \|_Q^2 : Ax + By + u = b \right\}.$$

Therefore, any classical condition guaranteeing solution of the above problem implies that $\bar{Z}_\mu(Q)$ is nonempty. For instance, coerciviness of f and g, or positive definiteness of R and S and injectiveness of B (which is equivalent to Q be definite positive).

The next result shows that the DR-ADMM generates a suitable pair $(z_k, \tilde{z}_k)$ satisfying the inclusion (6) with T as in (14) and $M = Q$, where Q is as in (15).

Proposition 3.1. *Let* $\{(x_k, y_k, \gamma_k, \tilde{\gamma}_k)\}$ *be the kth iterate of a cycle of the DR-ADMM and let* $\{(\Delta x_k, \Delta y_k, \Delta\gamma_k)\}$ *be as in* (12)*. Then,*

$$Q\begin{pmatrix}\Delta x_k\\ \Delta y_k\\ \Delta\gamma_k\end{pmatrix} \in \begin{pmatrix}\partial f(x_k) - A^*\tilde{\gamma}_k\\ \partial g(y_k) - B^*\tilde{\gamma}_k\\ Ax_k + By_k - b\end{pmatrix} + \mu Q\begin{pmatrix}x_k - x_0\\ y_k - y_0\\ \tilde{\gamma}_k - \gamma_0\end{pmatrix} \tag{16}$$

where Q *is as in* (15)*. As a consequence,* $z_k = (x_k, y_k, \gamma_k)$ *and* $\tilde{z}_k = (x_k, y_k, \tilde{\gamma}_k)$ *satisfy the inclusion* (6) *with* $M = Q$ *and* T *as in* (14)*.*

Proof. From the optimality condition for (8) and definitions of $\tilde{\gamma}_k$ and $\hat{x}_{k-1}$, we have

$$\begin{aligned}0 &\in \partial f(x_k) - A^*\left(\hat{\gamma}_{k-1} - \beta_1(Ax_k + By_{k-1} - b)\right) + (1+\mu)R(x_k - \hat{x}_{k-1})\\ &= \partial f(x_k) - A^*\tilde{\gamma}_k + R(x_k - x_{k-1}) + \mu R(x_k - x_0).\end{aligned} \tag{17}$$

Now, from the optimality condition for (9) and definition of u_k, we obtain

$$\begin{aligned}0 &\in \partial g(y_k) - B^*(u_k - \beta_2(Ax_k + By_k - b)) + (1+\mu)\alpha\beta B^*B(y_k - \hat{y}_{k-1}) + (\beta_2/\beta)S(y_k - \hat{y}_{k-1})\\ &= \partial g(y_k) - B^*\tilde{\gamma}_k + [(1+\mu)\alpha\beta + \beta_2]B^*B(y_k - \hat{y}_{k-1}) + (\beta_2/\beta)S(y_k - \hat{y}_{k-1})\\ &= [(1+\alpha)\beta B^*B + S](y_k - y_{k-1}) + \partial g(y_k) - B^*\tilde{\gamma}_k + \mu[(1+\alpha)\beta B^*B + S](y_k - y_0)\end{aligned} \tag{18}$$

where the last equality is due to definitions of β_2 and $\hat{y}_{k-1}$. On the other hand, definition of γ_k in (10) implies that

$$0 = (\gamma_k - \gamma_{k-1})/(\beta\theta) + Ax_k + By_k - b + \mu(\tilde{\gamma}_k - \gamma_0)/(\beta\theta).$$

Hence, the inclusion (16) follows from the last equality, (17), (18) and definitions in (12) and (15).

The second part of the proposition follows immediately from (16) and definitions of z_k, $\tilde{z}_k$, M and T. □

The following lemma describes some important properties of the sequences generated during a cycle of the DR-ADMM.

Lemma 3.2. *Let* $\{(x_k, y_k, \gamma_k, \tilde{\gamma}_k)\}$ *be the kth iterate of a cycle of the DR-ADMM and let* $\{(\Delta x_k, \Delta y_k, \Delta\gamma_k)\}$ *be as in* (12)*. Then, the following statements hold:*

(a) $\tilde{\gamma}_k - \gamma_{k-1} = -\beta B\Delta y_k - \Delta\gamma_k/\theta$*;*

(b) if $k = 1$ *and* $\theta \in [1, 2)$*, then*

$$\frac{1}{\theta}\langle B\Delta y_1, \Delta\gamma_1\rangle \geq \frac{1}{2}\|\Delta y_1\|^2_{\alpha\beta B^*B+S} - \frac{2\theta d_0}{2-\theta}$$

where $d_0 := \inf\{\|(x_0, y_0, \gamma_0) - (x, y, \gamma)\|_Q : (x, y, \gamma)$ *is solution of* (14)$\}$*;*

(c) if $k \geq 2$*, then*

$$2\langle B\Delta y_k, \Delta\gamma_k\rangle \geq 2(1-\theta)\langle B\Delta y_k, \Delta\gamma_{k-1}\rangle + \theta\|\Delta y_k\|^2_{\alpha\beta B^*B+S} - \theta\|\Delta y_{k-1}\|^2_{\alpha\beta B^*B+S}.$$

Proof. (a) Definitions of γ_k, $\tilde{\gamma}_{k-1}$ and β_1 in the DR-ADMM imply that

$$\begin{aligned}\gamma_k &= \gamma_{k-1} - \mu(\tilde{\gamma}_k - \gamma_0) - \theta\beta(Ax_k + By_{k-1} - b) - \theta\beta B(y_k - y_{k-1})\\ &= \gamma_{k-1} - \mu(\tilde{\gamma}_k - \gamma_0) + (\theta + \mu)(\tilde{\gamma}_k - \hat{\gamma}_{k-1}) - \theta\beta B(y_k - y_{k-1})\\ &= (1-\theta)\gamma_{k-1} + \theta\tilde{\gamma}_k - \theta\beta B(y_k - y_{k-1})\end{aligned}$$

where the last equality is due to definition of $\hat{\gamma}_k$. Hence, item (a) follows by simple calculus and (12).

(b) Let a point $\bar{z}_\mu := (\bar{x}_\mu, \bar{y}_\mu, \bar{\gamma}_\mu) \in \bar{Z}_\mu(Q)$ (see the assumption following (15)) and define

$$\tilde{z}_1 = (x_1, y_1, \tilde{\gamma}_1) \quad \text{and} \quad z_k = (x_k, y_k, \gamma_k),\ k = 0, 1. \tag{19}$$

Using (12), the fact that $-2\langle a, b\rangle \leq \|a\|^2 + \|b\|^2\ \forall a, b \in \mathbb{R}^m$, and $\theta \geq 1$, we obtain

$$\begin{aligned}\frac{1}{2}\|\Delta y_1\|^2_{\alpha\beta B^*B+S} - \frac{1}{\theta}\langle B\Delta y_1, \Delta\gamma_1\rangle &\leq \frac{1}{2}\left((1+\alpha)\beta\|B(y_1 - y_0)\|^2 + \|y_1 - y_0\|^2_S + \frac{1}{\beta\theta}\|\gamma_1 - \gamma_0\|^2\right)\\ &\leq (1+\alpha)\beta\left(\|B(y_1 - \bar{y}_\mu)\|^2 + \|B(y_0 - \bar{y}_\mu)\|^2\right) + \|y_1 - \bar{y}_\mu)\|^2_S\\ &\quad + \|y_0 - \bar{y}_\mu)\|^2_S + \frac{1}{\beta\theta}\|\gamma_1 - \bar{\gamma}_\mu\|^2 + \frac{1}{\beta\theta}\|\gamma_0 - \bar{\gamma}_\mu\|^2\end{aligned}$$

which, combined with (15), yields

$$\frac{1}{2}\|\Delta y_1\|^2_{\alpha\beta B^*B+S} - \frac{1}{\theta}\langle B\Delta y_1, \Delta\gamma_1\rangle \leq \|z_1 - \bar{z}_\mu\|^2_Q + \|z_0 - \bar{z}_\mu\|^2_Q. \tag{20}$$

On the other hand, note that

$$\|z_1 - \bar{z}_\mu\|^2_Q = \|z_0 - \bar{z}_\mu\|^2_Q + \|z_1 - \tilde{z}_1\|^2_Q - \|z_0 - \tilde{z}_1\|^2_Q + 2\langle Q(z_1 - z_0), \tilde{z}_1 - \bar{z}_\mu\rangle. \tag{21}$$

As $0 \in T(\bar{z}_\mu) + \mu Q(\bar{z}_\mu - z_0)$ and $Q(z_0 - z_1) \in (T(\tilde{z}_1) + \mu Q(\tilde{z}_1 - z_0))$ (see Proposition 3.1 with $k = 1$), we have $\langle Q(z_1 - z_0), \tilde{z}_1 - \bar{z}_\mu\rangle \leq 0$. This inequality together with (21) imply that

$$\|z_1 - \bar{z}_\mu\|^2_Q \leq \|z_0 - \bar{z}_\mu\|^2_Q + \|z_1 - \tilde{z}_1\|^2_Q - \|z_0 - \tilde{z}_1\|^2_Q. \tag{22}$$

Now, using the definitions in (15) and (19), we have

$$\begin{aligned}\|z_1 - \tilde{z}_1\|^2_Q - \|z_0 - \tilde{z}_1\|^2_Q &\leq \frac{1}{\beta\theta}\|\gamma_1 - \tilde{\gamma}_1\|^2 - \beta\|B(y_1 - y_0)\|^2 - \frac{1}{\beta\theta}\|\tilde{\gamma}_1 - \gamma_0\|^2\\ &= \frac{(\theta-2)}{\beta\theta^2}\|\gamma_1 - \gamma_0\|^2 - \frac{2}{\theta}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle - \beta\|B(y_1 - y_0)\|^2\\ &= \frac{(\theta-1)}{\beta\theta^2}\|\gamma_1 - \gamma_0\|^2 - \left\|B(y_1 - y_0) + \frac{\gamma_1 - \gamma_0}{\theta}\right\|^2,\end{aligned}$$

where the first equality is due to item (a) with $k = 1$. Therefore,

$$\begin{aligned}\|z_1 - \tilde{z}_1\|^2_Q - \|z_0 - \tilde{z}_1\|^2_Q &\leq \frac{(\theta-1)}{\beta\theta^2}\|\gamma_1 - \gamma_0\|^2 \leq \frac{2(\theta-1)}{\theta}\left(\frac{\|\gamma_1 - \bar{\gamma}_\mu\|^2}{\beta\theta} + \frac{\|\gamma_0 - \bar{\gamma}_\mu\|^2}{\beta\theta}\right)\\ &\leq \frac{2(\theta-1)}{\theta}\left(\|z_0 - \bar{z}_\mu\|^2_Q + \|z_1 - \bar{z}_\mu\|^2_Q\right)\end{aligned}$$

where the second inequality is due to the fact that $2\langle a,b\rangle \le \|a\|^2+\|b\|^2$ for all $a,b\in\mathbb{R}^m$, and the last inequality is due to (15) and definitions of z_0, z_1 and $\bar{z}_\mu$. Hence, combining the last estimative with (22), we obtain

$$\|z_1-\bar{z}_\mu\|_Q^2 \le \frac{\theta}{2-\theta}\left(1+\frac{2(\theta-1)}{\theta}\right)\|z_0-\bar{z}_\mu\|_Q^2 = \frac{3\theta-2}{2-\theta}\|z_0-\bar{z}_\mu\|_Q^2.$$

Therefore, statement (b) follows from (20), the last inequality, (5) with $M=Q$, and the definition of d_0.

(c) From (16) and definitions in (12) and (15), we obtain

$$B^*(\tilde{\gamma}_j-(1+\alpha)\beta B(y_j-y_{j-1}))-S(y_j-y_{j-1})\in\partial g_{\mu,\beta}(y_j) \qquad \forall j\ge 1,$$

where $g_{\mu,\beta}(y):=g(y)+(\mu/2)\|y-y_0\|^2_{(1+\alpha)\beta B^*B+S}$ for every $y\in\mathbb{R}^p$. Hence, using item (a), we have

$$(1/\theta)B^*(\gamma_j-(1-\theta)\gamma_{j-1})-(\alpha\beta B^*B+S)(y_j-y_{j-1})\in\partial g_{\mu,\beta}(y_j) \qquad \forall j\ge 1.$$

Using (12) and the previous inclusion for $j=k-1$ and $j=k$, it follows from the monotonicity of the subdifferential of $g_{\mu,\beta}$ that

$$0\le\langle B^*\Delta\gamma_k,\Delta y_k\rangle-(1-\theta)\langle B^*\Delta\gamma_{k-1},\Delta y_k\rangle-\theta\|\Delta y_k\|^2_{\alpha\beta B^*B+S}+\theta\langle(\alpha\beta B^*B+S)\Delta y_{k-1},\Delta y_k\rangle$$

which, combined with the fact that $2\langle(\alpha\beta B^*B+S)\Delta y_{k-1},\Delta y_k\rangle\le\|\Delta y_k\|^2_{\alpha\beta B^*B+S}+\|\Delta y_{k-1}\|^2_{\alpha\beta B^*B+S}$, yields item (c). □

In the next lemma, we establish a technical result which will be used in order to prove that the DR-ADMM with $\theta\in[1,(1-\alpha+\sqrt{\alpha^2+6\alpha+5})/2)$ is a special case of the DR-HPE framework.

Lemma 3.3. *Assume that* $\theta\in[1,(1-\alpha+\sqrt{\alpha^2+6\alpha+5})/2)$. *Then, there exists a parameter* $\bar{\tau}\in(0,1/2)$ *such that*

$$\bar{\sigma}:=\left(\frac{b+\sqrt{b^2-4ac}}{2a}\right)\in(0,1), \tag{23}$$

where $a:=(1-\bar{\tau})(1+\alpha)(1+\theta)-\alpha-(1-\theta)^2$, $c:=\left[1-\bar{\tau}-\alpha\bar{\tau}(1-\theta)-(1-\theta)^2\right](1-\theta)^2$ *and* $b:=[(1-\bar{\tau})(1+\alpha)(1+\theta)-\alpha-2(1-\theta)](1-\theta)^2-\alpha\bar{\tau}(1-\theta)+1-\bar{\tau}$. *Moreover,*

$$\max\left\{(1-\theta)^2,\frac{\bar{\tau}(\theta-1)}{(1-\bar{\tau})\theta-\bar{\tau}},\frac{1-\bar{\tau}[1+\alpha(1-\theta)]}{(1-\tau)(1+\alpha)(1+\theta)-\alpha}\right\}\le\bar{\sigma}, \tag{24}$$

and the matrix

$$G(\sigma)=\begin{bmatrix}(1-\bar{\tau})[\sigma(1+\theta)-1]+\alpha[\theta\sigma-\bar{\tau}(\sigma+\theta+\sigma\theta-1)] & (\sigma+\theta-1)(1-\theta)\\ (\sigma+\theta-1)(1-\theta) & \sigma-(1-\theta)^2\end{bmatrix} \tag{25}$$

is positive semidefinite for $\sigma=\bar{\sigma}$.

Proof. First of all, if $\theta=1$, then $\bar{\sigma}\in(0,1)$ for any $\bar{\tau}\in(0,1/2)$. Let us now assume that $\theta\in(1,(1-\alpha+\sqrt{\alpha^2+6\alpha+5})/2)$. Note that, if $\bar{\tau}=0$, then

$a=\theta[3-\theta+\alpha]>0,\ b=\theta[(3+\alpha)(1-\theta)^2+2-\theta]>0,\ a-b+c=\theta^2[1+2\alpha+(1-\alpha)\theta-\theta^2]>0,$

where the last inequality is due to the fact that $\theta \in (1, (1-\alpha+\sqrt{\alpha^2+6\alpha+5})/2)$. Moreover,

$$b^2 - 4ac = h(\alpha) := (5+6\alpha+\alpha^2)(1-\theta)^4 + 2(3+\alpha)(1-\theta)^3 - (1+2\alpha)(1-\theta)^2 - 2(1-\theta)^2 + 1 > 0,$$

where the above inequality follows from the fact that the minimum value of h is greater than zero for any $\theta \in (1,2)$. Therefore, we conclude that there exists $\bar{\tau} \in (0, 1/2)$ close to 0 such that

$$a > 0, \quad b > 0, \quad a - b + c > 0, \quad b^2 - 4ac \geq 0, \tag{26}$$

which in turn implies $\bar{\sigma} \in (0,1)$, concluding the proof of the first part of the lemma.

It is a simple algebraic computation to see that $\bar{\sigma}$ is the largest root of the second-order equation $\det(G(\sigma)) = 0$ and $\det(G(\sigma)) > 0$ for every $\sigma > \bar{\sigma}$. Moreover, since $\det(G(\sigma)) \leq 0$ for σ equal to $(1-\theta)^2$ and $[1-\bar{\tau}(1+\alpha(1-\theta))]/[(1-\tau)(1+\alpha)(1+\theta)-\alpha]$, and

$$\bar{\tau}(\theta-1)/[(1-\bar{\tau})\theta - \bar{\tau}] \leq [1-\bar{\tau}(1+\alpha(1-\theta))]/[(1-\tau)(1+\alpha)(1+\theta)-\alpha]$$

we obtain (24) holds. Therefore, since $\det(G(\bar{\sigma})) = 0$, the diagonal entries of $G(\bar{\sigma})$ are positive, and $G(\bar{\sigma})$ is symmetric, we conclude that $G(\bar{\sigma})$ is positive semidefinite. □

In next proposition, we will prove that the sequences $\{z_k\}$ and $\{\tilde{z}_k\}$ as in proposition 3.1 satisfy the error condition (7) with $M = Q$ and appropriate choices of τ, σ and $\{\eta_k\}$.

Proposition 3.4. *Assume that $\theta \in (0, (1-\alpha+\sqrt{\alpha^2+6\alpha+5})/2)$. Let $\{(x_k, y_k, \gamma_k, \tilde{\gamma}_k)\}$ be the kth iterate of a cycle of the DR-ADMM and let $\{(\Delta x_k, \Delta y_k, \Delta\gamma_k)\}$ be as in* (12). *Consider Q and d_0 as in* (15) *and Lemma 3.2(b), respectively. Let τ, σ and $\{\eta_k\}$ as*

(i) any $\tau \in (0,1)$, $\sigma = \theta + (\theta-1)^2$, and $\eta_k = 0$ for all $k \geq 0$, if $\theta \in (0,1)$;

(ii) $\tau = \bar{\tau}$ and $\sigma = \bar{\sigma}$, where $\bar{\tau}$ and $\bar{\sigma}$ are given by Lemma 3.3, and

$$\eta_0 = \frac{4(\bar{\sigma}+\theta-1)d_0}{(2-\theta)(1-\bar{\tau})}, \quad \eta_k = \frac{[\bar{\sigma}-(\theta-1)^2]}{\beta\theta^3}\|\Delta\gamma_k\|^2 + \frac{[\bar{\sigma}+\theta-1]}{\theta(1-\bar{\tau})}\|\Delta y_k\|^2_{\alpha\beta B^*B+S}, \ \forall k \geq 1, \tag{27}$$

if $\theta \in [1, (1-\alpha+\sqrt{\alpha^2+6\alpha+5})/2)$.

Then, $z_k = (x_k, y_k, \gamma_k)$, $\tilde{z}_k = (x_k, y_k, \tilde{\gamma}_k)$, η_{k-1} and η_k satisfy the error condition (7) *with $M = Q$.*

Proof. Using definitions of z_k, $\tilde{z}_k$ and Δy_k, and the fact that $M = Q$, we have

$$\sigma\|z_{k-1}-\tilde{z}_k\|^2_M - \|z_k - \tilde{z}_k\|^2_M \geq (1+\alpha)\sigma\beta\|B\Delta y_k\|^2 + \sigma\|\Delta y_k\|^2_S + \frac{\sigma}{\beta\theta}\|\gamma_{k-1}-\tilde{\gamma}_k\|^2 - \frac{1}{\beta\theta}\|\tilde{\gamma}_k - \gamma_k\|^2,$$

which, combined with (12) and Lemma 3.2(a), yields

$$\begin{aligned}
&\sigma\|z_{k-1}-\tilde{z}_k\|^2_M - \|z_k-\tilde{z}_k\|^2_M \\
&\geq (1+\alpha)\sigma\beta\|B\Delta y_k\|^2 + \sigma\|\Delta y_k\|^2_S + \frac{\sigma}{\beta\theta}\left\|\beta B\Delta y_k + \frac{\Delta\gamma_k}{\theta}\right\|^2 - \frac{1}{\beta\theta}\left\|\beta B\Delta y_k + \frac{(1-\theta)\Delta\gamma_k}{\theta}\right\|^2 \\
&= [(1+\alpha)\theta\sigma + \sigma - 1]\frac{\beta\|B\Delta y_k\|^2}{\theta} + \sigma\|\Delta y_k\|^2_S + [\sigma-(1-\theta)^2]\frac{\|\Delta\gamma_k\|^2}{\beta\theta^3} + \frac{2(\sigma+\theta-1)}{\theta^2}\langle\Delta\gamma_k, B\Delta y_k\rangle.
\end{aligned} \tag{28}$$

If $\theta \in (0,1)$, then the last inequality and $\sigma = \theta + (\theta-1)^2$ imply that

$$\sigma\|z_{k-1}-\tilde{z}_k\|_M^2 - \|z_k-\tilde{z}_k\|_M^2 \geq [\theta+(\theta-1)^2]\|\Delta y_k\|^2_{\alpha\beta B^*B+S} + \left\|\theta\sqrt{\beta}B\Delta y_k + \frac{\Delta\gamma_k}{\theta\sqrt{\beta}}\right\|^2 \geq 0,$$

which, combined with definition of $\{\eta_k\}$, proves the desired inequality.

Assume now that $\theta \in [1,(1-\alpha+\sqrt{\alpha^2+6\alpha+5})/2)$. Let us consider two case: $k=1$ and $k>1$. Case 1 ($k=1$): It follows from Lemma 3.2(b), definition of η_0 in (27), and $\theta \geq 1$ that

$$\frac{2(\bar{\sigma}+\theta-1)}{\theta^2}\langle B\Delta y_1, \Delta\gamma_1\rangle \geq \frac{(\bar{\sigma}+\theta-1)}{\theta}\left(\alpha\beta\|B\Delta y_1\|^2 + \|\Delta y_1\|_S^2\right) - (1-\bar{\tau})\eta_0$$

which, combined with (28) with $k=1$ and definitions σ, τ and η_1, yields

$$\begin{aligned}&\sigma\|z_0-\tilde{z}_1\|_M^2 - \|z_1-\tilde{z}_1\|_M^2 + (1-\tau)\eta_0 - \eta_1 \\ &\geq \left[(1+\alpha)\theta\bar{\sigma}+\bar{\sigma}-1-\frac{\alpha\bar{\tau}(\bar{\sigma}+\theta-1)}{1-\bar{\tau}}\right]\frac{\beta}{\theta}\|B\Delta y_1\|^2 + \left[\theta\bar{\sigma}-\frac{\bar{\tau}(\bar{\sigma}+\theta-1)}{1-\bar{\tau}}\right]\frac{1}{\theta}\|\Delta y_1\|_S^2 \geq 0\end{aligned}$$

where the last inequality is due to inequality (24). Thus, the error condition (7) holds for $k=1$. Case 2 ($k>1$): Combining estimate (28) with Lemma 3.2(c), we have

$$\begin{aligned}\sigma\|z_{k-1}-\tilde{z}_k\|_M^2 - \|z_k-\tilde{z}_k\|_M^2 &\geq [(1+\alpha)(\theta\bar{\sigma}+\bar{\sigma}+\theta-1)-\theta]\frac{\beta\|B\Delta y_k\|^2}{\theta} + [\bar{\theta}\sigma+\bar{\sigma}+\theta-1]\frac{\|\Delta y_k\|_S^2}{\theta} \\ &+ [\bar{\sigma}-(1-\theta)^2]\frac{\|\Delta\gamma_k\|^2}{\beta\theta^3} + \frac{2(1-\theta)(\bar{\sigma}+\theta-1)}{\theta^2}\langle B\Delta y_k, \Delta\gamma_{k-1}\rangle - \frac{(\bar{\sigma}+\theta-1)}{\theta}\|\Delta y_{k-1}\|^2_{\alpha\beta B^*B+S}.\end{aligned}$$

From the last inequality and definition of $\{\eta_k\}$ in (27), we obtain

$$\begin{aligned}&\sigma\|z_{k-1}-\tilde{z}_k\|_M^2 - \|z_k-\tilde{z}_k\|_M^2 + (1-\tau)\eta_{k-1} - \eta_k \\ &\geq [(1-\bar{\tau})(\bar{\sigma}(1+\theta)-1)+\alpha(\theta\bar{\sigma}-\bar{\tau}(\bar{\sigma}+\theta+\bar{\sigma}\theta-1))]\frac{\beta\|B\Delta y_k\|^2}{(1-\bar{\tau})\theta} + \left[\bar{\theta}\bar{\sigma}-\frac{\bar{\tau}(\bar{\sigma}+\theta-1)}{1-\bar{\tau}}\right]\frac{\|\Delta y_k\|_S^2}{\theta} \\ &+ (1-\bar{\tau})[\bar{\sigma}-(1-\theta)^2]\frac{\|\Delta\gamma_{k-1}\|^2}{\beta\theta^3} + \frac{2(1-\theta)(\bar{\sigma}+\theta-1)}{\theta^2}\langle B\Delta y_k, \gamma_{k-1}-\Delta\gamma_{k-1}\rangle \\ &= \left[\bar{\theta}\bar{\sigma}-\frac{\bar{\tau}(\bar{\sigma}+\theta-1)}{1-\bar{\tau}}\right]\frac{\|\Delta y_k\|_S^2}{\theta} + (w_1,w_2)G(\bar{\sigma})(w_1,w_2)^*,\end{aligned}$$

where $G(\bar{\sigma})$ is as in (25), $w_1 = (\sqrt{\beta\theta/(1-\tau)})B\Delta y_k$ and $w_2 = (\sqrt{(1-\tau)/(\beta\theta)})\Delta\gamma_{k-1}$. Hence, the error condition (7) for $k>1$ now follows from Lemma 3.3. □

We are now ready to prove the main result of this section.

Theorem 3.5. *Assume that $\theta \in (0,(1-\alpha+\sqrt{\alpha^2+6\alpha+5})/2)$ and let Q be as in* (15)*. Then, the DR-ADMM is an instance of the DR-HPE framework for solving problem* (14) *with inputs $z_0 = (x_0,y_0,\gamma_0)$, $M=Q$, and parameters τ, σ and η_0 as defined in Proposition 3.4. As a consequence, it terminates in at most*

$$\mathcal{O}\left(\left(1+\frac{d_0}{\rho}\right)\left[1+\log^+\left(\frac{d_0}{\rho}\right)\right]\right) \tag{29}$$

iterations with $(x, y, \tilde{\gamma}, v^x, v^y, v^\gamma)$ *satisfying*

$$Q \begin{pmatrix} v^x \\ v^y \\ v^\gamma \end{pmatrix} \in \begin{pmatrix} \partial f(x) - A^*\tilde{\gamma} \\ \partial g(y) - B^*\tilde{\gamma} \\ Ax + By - b \end{pmatrix} \quad \text{and} \quad \|(v^x, v^y, v^\gamma)\|_Q \leq \rho, \tag{30}$$

where d_0 *is as in Lemma 3.2(b).*

Proof. Let $\{(x_k, y_k, \gamma_k, \tilde{\gamma}_k)\}$ be the sequence generated by a cycle of the DR-ADMM and consider the sequences $\{z_k\}$ and $\{\tilde{z}_k\}$ defined by

$$z_{k-1} = (x_{k-1}, y_{k-1}, \gamma_{k-1}), \quad \tilde{z}_k = (x_k, y_k, \tilde{\gamma}_k), \quad \forall k \geq 1. \tag{31}$$

It follows from Propositions 3.1 and 3.4 that the sequences $\{z_k\}$ and $\{\tilde{z}_k\}$ satisfy inclusion (6) and the error condition (7) with T as in (14), $M = Q$, and τ, σ and $\{\eta_k\}$ as defined in Proposition 3.4. Moreover, using $M = Q$ and (31), it is easy to see that steps 3 and 4 of the DR-ADMM correspond to steps 2 and 3 of the DR-HPE framework, respectively. Therefore, the first statement of the theorem is proved.

Now, since $\eta_0 = 0$ or $\eta_0 = \mathcal{O}(d_0^2)$, the second part of the theorem follows from the first one and Theorem 2.1 with $M = Q$, T as in (14), $v = (v^x, v^y, v^\gamma)$, $\tilde{z} = (x, y, \tilde{\gamma})$ and $d = d_0$. $\square$

We end this section by making two remarks. 1) As already mentioned in Section 1, if α is sufficiently large (resp. $\alpha = 0$), then the stepsize θ belong to the interval $(0, 2)$ (resp. $(0, (1+\sqrt{5})/2)$). 2) Note that (30) can be seen as an optimality/feasibility measure of (1). Indeed, since Q is symmetric semidefinite positive, if $\|(v^x, v^y, v^\gamma)\|_Q = 0$, then the left-hand side of the inclusion in (30) is zero, and hence the pair (x, y) is a solution of (1) and $\tilde{\gamma}$ is an associated Lagrange multiplier.

References

[1] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods.* Academic Press, New York, 1982.

[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(1):1–122, 2011.

[3] Y. Cui, X. Li, D. Sun, and K. C. Toh. On the convergence properties of a majorized ADMM for linearly constrained convex optimization problems with coupled objective functions. *J. Optim. Theory Appl.*, 169(3):1013–1041, 2016.

[4] W. Deng and W. Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. *J. Sci. Comput.*, pages 1–28, 2015.

[5] E. X. Fang, B. He, H. Liu, and X. Yuan. Generalized alternating direction method of multipliers: new theoretical insights and applications. *Math. Prog. Comp.*, 7(2):149–187, 2015.

[6] M. Fazel, T. K. Pong, D. Sun, and P. Tseng. Hankel matrix rank minimization with applications to system identification and realization. *SIAM J. Matrix Anal. Appl.*, 34(3):946–977, 2013.

[7] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.*, 2:17–40, 1976.

[8] R. Glowinski. *Numerical Methods for Nonlinear Variational Problems.* Springer Series in Computational Physics. Springer-Verlag, 1984.

[9] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par penalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. *RAIRO Anal. Numér.*, 9:41–76, 1975.

[10] M. L. N. Gonçalves, J. G. Melo, and R. D. C. Monteiro. Convergence rate bounds for a proximal ADMM with over-relaxation stepsize parameter for solving nonconvex linearly constrained problems. *Avaliable on https://arxiv.org/abs/1702.01850.*

[11] M. L. N. Gonçalves, J. G. Melo, and R. D. C. Monteiro. Improved pointwise iteration-complexity of a regularized ADMM and of a regularized non-euclidean HPE framework. *SIAM J. Optim.*, 27(1):379–407, 2017.

[12] Y. Gu, B. Jiang, and H. Deren. A semi-proximal-based strictly contractive Peaceman-Rachford splitting method. *Avaliable on https://arxiv.org/abs/1506.02221.*

[13] W. W. Hager, M. Yashtini, and H. Zhang. An $O(1/k)$ convergence rate for the variable stepsize Bregman operator splitting algorithm. *SIAM J. Numer. Anal.*, 54(3):1535–1556, 2016.

[14] B. He and F. Ma. Convergence study on the proximal alternating direction method with larger step size. *Avaliable on http://www.optimization-online.org/DB_FILE/2017/02/5856.pdf.*

[15] B. He and X. Yuan. On the $\mathcal{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. *SIAM Journal on Numer. Anal.*, 50(2):700–709, 2012.

[16] B. He and X. Yuan. On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. *Numer. Math.*, 130(3):567–577, 2015.

[17] T. Lin, S. Ma, and S. Zhang. An extragradient-based alternating direction method for convex minimization. *Found. Comput. Math.*, pages 1–25, 2015.

[18] M. Marques Alves, R. D. C. Monteiro, and B. F. Svaiter. Regularized HPE-type methods for solving monotone inclusions with improved pointwise iteration-complexity bounds. *SIAM J. Optim.*, 26(4):2730–2743, 2016.

[19] R. D. C. Monteiro and B. F Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM J. Optim.*, 23(1):475–507, 2013.

[20] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, 1970.

[21] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis.* Springer, Berlin, 1998.

[22] O. Sarmiento, E.A. Papa Quiroz, and P.R. Oliveira. A proximal multiplier method for separable convex minimization. *Optimization*, 65(2):501–537, 2016.

[23] X. Wang and X Yuan. The linearized alternating direction method of multipliers for dantzig selector. *SIAM J. Sci. Comput.*, 34(5):2792–2811, 2012.

[24] J. Yang and X. Yuan. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. *Math. Comput.*, 82(281):301–329, 2013.