Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 14, No. 3, pp. 507–525, 2018

ON THE CONVERGENCE OF ALTERNATING DIRECTION METHOD MULTIPLIERS WITH LARGER STEP SIZE THAN GLOWINSKI'S

M. Tao*

Abstract: The alternating direction method of multipliers (ADMM) is a popular approach for solving separable convex programming problems. The proximal ADMM (PADMM) is introduced for easing the solvability of each subproblem when a closed-form solution is not available. In 1982, Fortin and Glowinski proposed to enlarge the range of the step size for updating the multiplier from the value 1 to the range $(0, \frac{1+\sqrt{5}}{2})$. Meanwhile, he pointed out that a larger step size can speed up the numerical performance. In this paper, we theoretically analyze the global convergence of the PADMM with the range even larger than Glowinski's. Moreover, its worst-case convergence rate in the ergodic sense is also discussed.

Key words: *convex programming, alternating direction method of multipliers, proximal alternating direction method of multipliers, step size, convergence*

Mathematics Subject Classification: *90C25, 90C30, 65K05*

1 Introduction

We consider the following convex model

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\, x \in \mathcal{X},\, y \in \mathcal{Y}\}, \tag{1.1}$$

where $A \in \mathbb{R}^{m\times n_1}$, $B \in \mathbb{R}^{m\times n_2}$, $b \in \mathbb{R}^m$, $\mathcal{X} \subset \mathbb{R}^{n_1}$ and $\mathcal{Y} \subset \mathbb{R}^{n_2}$ are closed convex sets, $\theta_1 : \mathbb{R}^{n_1} \to (-\infty, +\infty]$ and $\theta_2 : \mathbb{R}^{n_2} \to (-\infty, +\infty]$ are convex (not necessarily smooth) functions. Throughout, we assume that the sets $\mathcal{X}$ and $\mathcal{Y}$ are assumed to be easy to calculate projections. Let $\beta > 0$ be the penalty parameter. The augmented Lagrangian function of (1.1) can be written as

$$\mathcal{L}_\beta(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^\top(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2, \tag{1.2}$$

with λ the Lagrange multiplier.

One may apply directly the augmented Lagrangian method (ALM) in [21, 22] to (1.1), the iterative scheme is

$$\begin{cases} (x^{k+1}, y^{k+1}) = \arg\min\{\mathcal{L}_\beta(x, y, \lambda^k) \,|\, x \in \mathcal{X},\ y \in \mathcal{Y}\}, & \text{(1.3a)} \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} + By^{k+1} - b), & \text{(1.3b)} \end{cases}$$

*The author is supported by the JiangSu Provincial National Natural Science Foundation of China (No. BK20181257) and the Fundamental Research Funds for the Central Universities: 14380019 and the National Science Foundation of China (No. 11301280).

where $\gamma \in (0, 2)$ guarantees the convergence. However, the nonseparability of the (x, y)-minimization problem in (1.3) may destroy the functions θ_1 and θ_2's own properites/strcutures. To overcome this difficulty, one may consider the following ADMM [16, 13] scheme to solve (1.1):

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & (1.4a) \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, & (1.4b) \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} + By^{k+1} - b), & (1.4c) \end{cases}$$

with $\gamma \in (0, \frac{1+\sqrt{5}}{2})$. Note the ADMM decouples the subproblem (1.3a) into two easier ones by fixing one variable and minimizing with the other, followed by a dual updating step. We remark here that the ADMM scheme (1.4) is different from the over-relaxed ADMM (see [2, 8, 9]). Thus, the parameter γ in (1.4c) is different from the relaxed factor in the over-relaxed ADMM. The convergence of (1.4) has long been established under a different context [7, 8, 12, 13]. We refer to, e.g., [2, 9, 10, 15], for some reviews on ADMM.

By noting the fact that the subproblems in (1.4) may still be difficult to solve even if each proximity of the functions θ_1 and θ_2 [18] preserves a closed-form solution. A remedy way is to add an appropriately chosen quadratic proximal term in the corresponding subproblem. Let us take the x-subproblem as an example. Suppose that the proximity of the function θ_1 deserves a closed-form solution, e.g., $\theta_1(x) = \|x\|_1$ [2]. In order to solve the subproblem (1.5a) relatively easily, we can linearize the augmented Lagrange term by choosing $R = \xi I - \beta A^\top A$ when $A \neq I$. Moreover, if the function θ_1 is a quadratic function, namely, $\theta_1(x) = \frac{1}{2}x^\top \Sigma_1 x + c^\top x$, then we can linearize either the function θ_1 or both θ_1 and the augmented Lagrange term simultaneously. Or equivalently, we choose $R = \xi I - \Sigma_1$ if $A = I$, or set $R = \xi I - \beta A^\top A - \Sigma_1$ when A is not identity. A similar strategy is also applicable for the y-subproblem, and more details can be found in [4, 18]. An immediate consequence is that the related subproblem is usually easy to solve.

Hence, it results in the following proximal alternating direction method of multipliers (PADMM) scheme:

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) + \frac{1}{2}\|x - x^k\|_R^2 \mid x \in \mathcal{X}\}, & (1.5a) \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_S^2 \mid y \in \mathcal{Y}\}, & (1.5b) \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} + By^{k+1} - b), & (1.5c) \end{cases}$$

where $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ and $R \succeq 0$ and $S \succeq 0$ are two symmetric positive semidefinite matrices [11]. When R and S are two positive scalar matrices (i.e., cI with some positive constant c), the above PADMM scheme with $\gamma = 1$ reduces to the splitting method in [7]. He et al. [17] established the global convergence of (1.5) with $\gamma = 1$, and both of R, S are symmetric positive definite matrices in the variational inequality context and even considered varying β, R and S at each iteration. Later, Xu [26] extended the convergence analysis for (1.5) with $\gamma \in (0, \frac{1+\sqrt{5}}{2})$, and both of R, S are symmetric positive definite matrices. Recently, Fazel et al. [11] analyzed the convergence of (1.5) with $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ and R, S are both self-adjoint positive semidefinite linear operators. Since R, S can be positive semidefinite linear operators, the PADMM (1.5) includes the classical ADMM (1.4) as a special case.

In [11], the PADMM is also called semi-Proximal ADMM.

The introduction of the relaxation factor γ in (1.5c) is attributed to Fortin and Glowinski. Both of the x- and y-subproblems in (1.5) are usually computationally expensive, while updating the Lagrange multiplier is much cheaper. Moreover, it has been numerically verified that a large step size in (1.5c) leads to a better performance, see numerical results in [14] and further in [5, 20, 25]. Furthermore, the relaxation parameter γ's range is $(0, 2)$ in the ALM (1.3) while the range is $(0, \frac{1+\sqrt{5}}{2})$ in the PADMM (1.5) (including ADMM). This gap may result from the fact that the step (1.3a) in the ALM is substituted by the componentwise minimization steps (1.4a) and (1.4b) in the ADMM, since the latter is regarded as an incomplete version of the former.

Besides [13], this question has recently attracted some efforts in the literatures [24, 6]. In [24], the authors showed the first attempt to adopt a large step size in the PADMM (1.5) for solving conic programming. Furthermore, the authors [6] simplified the conditions in [24] and emphasized the necessity of further enlarging the range $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ and commented as "the ADMM with a computationally more attractive large step-length that can even exceed the practically much preferred golden ratio of $\frac{1+\sqrt{5}}{2}$". These conditions proposed therein (see Theorem 2.2 in [24] and Theorem 4.1 in [6]) require the summability of an sequence asymptotically and some practical techniques were introduced to implement these conditions.

In this paper, we conduct a rigorous convergence analysis of the PADMM (1.5) with a step size $\gamma \in (0, \frac{1-\tau+\sqrt{\tau^2+6\tau+5}}{2})$ under some moderate and checkable conditions. This range is much larger than $(0, \frac{1+\sqrt{5}}{2})$ and the scalar τ will be specified in Assumption 2.2. Indeed, there always exists a positive scalar τ satisfying Assumption 2.2 by setting $S = \zeta I - \Sigma_2$ and $\zeta = \lambda_{\max}(\Sigma_2)$ with Σ_2 defined in (2.2). This choice of S is usually adopted when the function θ_2 is a quadratic function [4]. Note the above setting of S is positive semidefinite instead of positive definite. It implies that we obtain some improvement on the range for γ without any enlargement for the matrix S involved in the y-subproblem (1.5b). In addition to establishing the global convergence, we also analyze its ergodic iteration-complexity. Indeed, we carry out the ergodic iteration analysis in terms of the partial primal-dual gap [3] (see page 121), the feasibility violation and the decrement of the objective function.

The remainder of this paper is organized as follows. In Section 2, we summarize some notations, present the basic assumptions, and characterize the optimality condition of (1.1). In Section 3, we prepare some properties, lemmas and theorems for facilitating global convergence and convergence rate analysis. Then, the global convergence and convergence rate are studied in Section 4. Finally, some conclusions are made in Section 5.

During the second round review of this paper, we noticed that He and Ma [19] considered the following proximal-ADMM to solve (1.1):

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \frac{\tau\beta}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} + By^{k+1} - b), \end{cases} \tag{1.6}$$

with $\mathcal{L}_\beta$ defined in (1.2). They established the global convergence of (1.6) for solving (1.1) with $\Sigma_1 = \Sigma_2 = 0$ defined in (2.1)-(2.2) when the step size $\gamma \in (0, \frac{1-\tau+\sqrt{\tau^2+6\tau+5}}{2})$ and B has full column rank; and also derive a worst-case $O(1/t)$ convergence rate in the ergodic sense. Note that the scheme (1.6) is a special case of the (1.5) with setting $R = 0$ and $S = \tau\beta B^\top B$. This case has been covered by our results, see Remark 2.1 in Section 2. Nevertheless, our results differs from theirs in three aspects. First, our results are more general: We introduce semi-proximal terms both in x and y subproblems. Second, the proximal matrix S in (1.5) is not necessarily positive definite while the proximal matrix in (1.6), i.e. $\tau\beta B^\top B$ is positive definite due to B full column rank. Moreover, our analysis shows that the range of the step size γ can still be larger than $(0, \frac{1+\sqrt{5}}{2})$ when θ_2 is a quadratic or smooth function while without any enlargement in the proximal term, see Remark 2.2. Third, we carry out the ergodic iteration analysis in terms of the partial primal-dual gap, the feasibility violation and the decrement of the objective function, while the ergodic iteration analysis in [19] is measured by the partial primal-dual gap.

2 Preliminaries

In this section, we define some notations to be used, present some assumptions, and show the optimality condition of the model (1.1) in the variational inequality context.

2.1 Notations

For a set $\mathcal{D}$, $\mathrm{ri}(\mathcal{D})$ denotes the relative interior of $\mathcal{D}$. For a vector $x \in \mathbb{R}^n$, $\|x\|_2$ represents $\sqrt{\sum_{i=1}^n |x_i|^2}$. Given a symmetric positive semidefinite matrix Σ, we denote $\|x\|_\Sigma := \sqrt{x^\top \Sigma x}$. The notation $M \succeq 0$ means M is positive semidefinite and $M \succ 0$ means M is positive definite. Since θ_1 and θ_2 are closed convex functions, there exists two symmetric positive semidefinite matrices Σ_1 and Σ_2 such that

$$(x - x')^\top (g - g') \geq \|x - x'\|_{\Sigma_1}^2, \ \forall\, x,\ x' \in \mathrm{dom}(\theta_1),\ g \in \partial\theta_1(x),\ g' \in \partial\theta_1(x'), \tag{2.1}$$

and

$$(y - y')^\top (h - h') \geq \|y - y'\|_{\Sigma_2}^2, \ \forall\, y,\ y' \in \mathrm{dom}(\theta_2),\ h \in \partial\theta_2(y),\ h' \in \partial\theta_2(y'). \tag{2.2}$$

The definitions above for θ_1 and θ_2 are flexible enough to include convex functions, i.e., θ_i's $(i = 1, 2)$ are convex when $\Sigma_i = 0$.

By letting

$$u := \begin{pmatrix} x \\ y \end{pmatrix},\ s := \begin{pmatrix} g \\ h \end{pmatrix},\ \Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix},$$

the inequalities (2.1)-(2.2) can be written compactly as

$$(u - u')^\top (s - s') \geq \|u - u'\|_\Sigma^2.$$

2.2 Assumptions

Throughout this paper, we suppose that the following assumptions hold.

Assumption 2.1. *The solution set of (1.1) is nonempty. There exists*

$$u' = (x', y') \in ri(dom(\theta_1) \times dom(\theta_2)) \cap \mathcal{F},$$

where

$$\mathcal{F} := \{u = (x, y) \in \mathcal{X} \times \mathcal{Y} \mid Ax + By = b\}.$$

Assumption 2.2. *Suppose θ_1 and θ_2 in (1.1) satisfy (2.1)-(2.2) with two symmetric positive semidefinite matrices Σ_1, Σ_2. The matrices R and S defined in (1.5) are symmetric positive semidefinite. The matrices A, B in (1.1), Σ_1, Σ_2 and R, S satisfy*

$$\Sigma_1 + A^\top A + R \succ 0, \quad \Sigma_2 + B^\top B + S \succ 0.$$

For given $\beta > 0$, there exists a positive scalar $\tau > 0$ such that

$$2\Sigma_2 + S \succ \beta\tau B^\top B. \tag{2.3}$$

Remark 2.3. The condition (2.3) can be relaxed to

$$2\Sigma_2 + S \succeq \beta\tau B^\top B, \tag{2.4}$$

whenever B has full column rank.

Remark 2.4. We remark that if the function θ_2 is a quadratic or smooth function, then there always exists a positive scalar τ (not small) satisfying (2.3) for the PADMM (1.5) even with a positive semidefinite proximal term [6, 11]. For example, let $\theta_2(y) := \frac{1}{2}y^\top\Sigma_2 y + c^\top y$. We can choose $S = \zeta I - \Sigma_2$ with $\zeta = \lambda_{\max}(\Sigma_2)$ when $B = I$, and $S = \zeta I - \beta B^\top B - \Sigma_2$ with $\zeta = \lambda_{\max}(\beta B^\top B + \Sigma_2)$ when $B \neq I$ and it has full column rank. Then the maximum scalar τ satisfying (2.3) can be computed as follows:

$$\tau = \frac{\lambda_{\min}(\Sigma_2) + \zeta}{\beta\lambda_{\max}(B^\top B)} - 1.$$

Further, it reduces to

$$\tau = \frac{\lambda_{\min}(\Sigma_2) + \zeta}{\beta},$$

when $B = I$. Usually, the τ defined above is large when $\lambda_{\max}(\Sigma_2)$ is large with β given due to $\zeta \geq \lambda_{\max}(\Sigma_2)$.

Each subproblem in (1.5) is well-defined under Assumptions 2.1 and 2.2.

2.3 Optimality condition

In this section, we characterize the optimality condition of the problem (1.1) in the variational inequality context. Denote $\Omega := \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$. Define the Lagrangian function of the problem (1.1) by

$$\mathcal{L}(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^\top(Ax + By - b). \tag{2.5}$$

Under Assumption 2.1, it follows from [23, Corollary 28.2.2] and [23, Corollary 28.3.1] that $(x^*, y^*) \in \mathrm{dom}\theta_1 \times \mathrm{dom}\theta_2$ is a solution of (1.1) if and only if there exists a Lagrangian multiplier $\lambda^* \in \mathbb{R}^m$ such that (x^*, y^*, λ^*) is a saddle point to the Lagrangian function (2.5). Thus, under Assumptions 2.1 and 2.2, we can characterize the optimality condition of (1.1) as

$$\begin{cases} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x - x^*)^\top(-A^\top\lambda^*) \geq \frac{1}{2}\|x - x^*\|^2_{\Sigma_1}, & \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) + (y - y^*)^\top(-B^\top\lambda^*) \geq \frac{1}{2}\|y - y^*\|^2_{\Sigma_2}, & \forall\, y \in \mathcal{Y}, \\ \lambda^* \in \mathbb{R}^m, & (\lambda - \lambda^*)^\top(Ax^* + By^* - b) \geq 0, & \forall\, \lambda \in \mathbb{R}^m. \end{cases} \tag{2.6}$$

More compactly, the system (2.6) can be written as the following variational inequality (VI):

$$\mathrm{VI}(\Omega, F, \theta):\ w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^\top F(w^*) \geq \frac{1}{2}\|u - u^*\|_\Sigma^2,\ \forall\, w \in \Omega, \tag{2.7a}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix},\ w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix},\ \ \theta(u) = \theta_1(x) + \theta_2(y),\ F(w) = \begin{pmatrix} -A^\top\lambda \\ -B^\top\lambda \\ Ax + By - b \end{pmatrix}. \tag{2.7b}$$

The solution set of $\mathrm{VI}(\Omega, F, \theta)$ is denoted by Ω^* which is nonempty under Assumption 2.1.

$\boxed{3}$ Preparation for Convergence

In this section, we present several lemmas and theorems for facilitating global convergence and convergence rate analysis. The following three lemmas are elementary, and the proof is trivial and thus omitted.

Lemma 3.1. *The mapping $F(w)$ defined in (2.7b) satisfies*

$$(w' - w)^\top [F(w') - F(w)] = 0, \quad \forall\, w',\ w \in \mathbb{R}^{n_1+n_2+m}.$$

Lemma 3.2. *For a matrix $H \in \mathbb{R}^{n\times n}$ with $H \succeq 0$ and vectors $a,\ b,\ c,\ d \in \mathbb{R}^n$, we have*

$$(a-b)^\top H(c-d) = \frac{1}{2}\left[\|a-d\|_H^2 - \|a-c\|_H^2\right] + \frac{1}{2}\left[\|c-b\|_H^2 - \|d-b\|_H^2\right].$$

Lemma 3.3. *Suppose that Σ_1, Σ_2 and R, S are symmetric positive semidefinite. The matrices A and B are given in (1.1). Then, for any $\beta > 0$, we have $\Sigma_1 + \beta A^\top A + R \succ 0$ whenever $\Sigma_1 + A^\top A + R \succ 0$. Analogously, $\Sigma_2 + \beta B^\top B + S \succ 0$ whenever $\Sigma_2 + B^\top B + S \succ 0$.*

Next, we introduce an auxiliary vector $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$, defined as

$$\tilde{x}^k = x^{k+1}, \qquad \tilde{y}^k = y^{k+1}, \tag{3.1a}$$

and

$$\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b), \tag{3.1b}$$

where the sequence $\{w^k\}$ is generated by the PADMM (1.5).

The coming lemma reveals the relationship among w^k, $\tilde{w}^k$ and w^{k+1}.

Lemma 3.4. *Let $\{w^{k+1}\}$ be generated by* (1.5) *and $\tilde{w}^k$ be defined by* (3.1). *Then, we have*

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

where

$$M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -\gamma\beta B & \gamma I_m \end{pmatrix}. \tag{3.2}$$

In the following, we characterize the optimality condition of the PADMM (1.5) with variational inequality.

Lemma 3.5. *Let $\{w^{k+1}\}$ be generated by the PADMM* (1.5) *and $\tilde{w}^k$ be defined by* (3.1). *Then, we have*

$$\tilde{w}^k \in \Omega,\ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top \left[F(\tilde{w}^k) - Q(w^k - \tilde{w}^k)\right] \geq 0,\ \ \forall\, w \in \Omega, \tag{3.3a}$$

where $\theta(u)$ and $F(w)$ are defined in (2.7b) and the matrix Q is defined by

$$Q = \begin{pmatrix} R & 0 & 0 \\ 0 & S + \beta B^\top B & 0 \\ 0 & -B & \frac{1}{\beta} I \end{pmatrix}. \tag{3.3b}$$

Proof. First, according to the optimality condition of the x-subproblem in (1.5a), we get

$$\tilde{x}^k \in \mathcal{X},\ \ \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^\top \left[-A^\top \tilde{\lambda}^k + R(\tilde{x}^k - x^k)\right] \geq \frac{1}{2}\|x - \tilde{x}^k\|_{\Sigma_1}^2,\quad \forall\, x \in \mathcal{X}. \tag{3.4}$$

Analogously, using the optimality condition of the y-subproblem in (1.5b), we have

$$\begin{aligned} \tilde{y}^k \in \mathcal{Y},\ \ &\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^\top \left[-B^\top \tilde{\lambda}^k + (S + \beta B^\top B)(\tilde{y}^k - y^k)\right] \\ &\geq \frac{1}{2}\|y - \tilde{y}^k\|_{\Sigma_2}^2,\ \forall\, y \in \mathcal{Y}. \end{aligned} \tag{3.5}$$

Finally, according to the definition of $\tilde{\lambda}^k$, we obtain

$$\tilde{\lambda}^k \in \mathbb{R}^m,\ \ (\lambda - \tilde{\lambda}^k)^\top \left[(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k)\right] \geq 0,\quad \forall\, \lambda \in \mathbb{R}^m. \tag{3.6}$$

Combining (3.4), (3.5) and (3.6), and using the notations in (2.7b) and the definition of Q in (3.3b), the assertion follows immediately. ☐

Next, we further introduce a positive semidefinite matrix

$$H = \begin{pmatrix} R & 0 & 0 \\ 0 & S + \beta B^\top B & 0 \\ 0 & 0 & \frac{1}{\gamma\beta} I \end{pmatrix}. \tag{3.7}$$

According to the definitions of Q in (3.3b) and M in (3.2), it holds that $Q = HM$. Consequently, we can rewrite

$$(w - \tilde{w}^k)^\top Q(w^k - \tilde{w}^k) = (w - \tilde{w}^k)^\top H(w^k - w^{k+1}), \tag{3.8}$$

where the equality follows from Lemma 3.4. Based on Lemma 3.5 and (3.8), we show the discrepancy between the auxiliary vector $\tilde{w}^k$ and any point in Ω.

Theorem 3.6. *For the sequence $\{w^k\}$ generated by the PADMM (1.5), we have*

$$\begin{aligned} &\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \\ &\quad\geq \frac{1}{2}\left(\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2\right) + \frac{1}{2}\left(\|w^k - \tilde{w}^k\|_H^2 - \|w^{k+1} - \tilde{w}^k\|_H^2\right),\ \ \forall\, w \in \Omega, \end{aligned} \tag{3.9}$$

where H is defined in (3.7).

Proof. First, since $Q = HM$, substituting (3.8) into (3.3a) yields

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (w - \tilde{w}^k)^\top H(w^k - w^{k+1}), \quad \forall\, w \in \Omega. \tag{3.10}$$

Due to Lemma 3.2, we have

$$\begin{aligned}(w - \tilde{w}^k)^\top H(w^k - w^{k+1}) &= \frac{1}{2}\big(\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2\big) \\ &+\frac{1}{2}(\|w^k - \tilde{w}^k\|_H^2 - \|w^{k+1} - \tilde{w}^k\|_H^2).\end{aligned} \tag{3.11}$$

Substituting (3.11) into the right-hand side of (3.10), the assertion follows directly. □

Consequently, we take a further analysis for the second term of the right-hand side of (3.9). Thus, we define it as:

$$\Delta(w^k, w^{k+1}) = \frac{1}{2}\big(\|w^k - \tilde{w}^k\|_H^2 - \|w^{k+1} - \tilde{w}^k\|_H^2\big). \tag{3.12}$$

Lemma 3.7. *For the sequence $\{w^k\}$ generated by the PADMM (1.5), we have*

$$\begin{aligned}\Delta(w^k, w^{k+1}) = \frac{1}{2}\Big(&\|x^k - x^{k+1}\|_R^2 + \|y^k - y^{k+1}\|_{S+\beta B^\top B}^2 \\ &+(2-\gamma)\beta\|r^{k+1}\|^2 + 2\beta(By^k - By^{k+1})^\top r^{k+1}\Big),\end{aligned} \tag{3.13}$$

with

$$r^{k+1} := Ax^{k+1} + By^{k+1} - b. \tag{3.14}$$

Proof. First, due to (3.1), we have

$$\lambda^k - \tilde{\lambda}^k = \beta r^{k+1} + \beta B(y^k - y^{k+1}). \tag{3.15}$$

Consequently, combining with (1.5c), we get

$$\begin{aligned}\lambda^{k+1} - \tilde{\lambda}^k &= \lambda^{k+1} - \lambda^k + \lambda^k - \tilde{\lambda}^k \\ &= -\gamma\beta r^{k+1} + \beta r^{k+1} + \beta B(y^k - y^{k+1}) \\ &= (1-\gamma)\beta r^{k+1} + \beta B(y^k - y^{k+1}).\end{aligned} \tag{3.16}$$

Combining the above two equalities, we further obtain

$$\begin{aligned}&\frac{1}{\gamma\beta}\Big(\|\lambda^k - \tilde{\lambda}^k\|^2 - \|\lambda^{k+1} - \tilde{\lambda}^k\|^2\Big) \\ &= \frac{1}{\gamma\beta}\big(\|\beta r^{k+1} + \beta B(y^k - y^{k+1})\|^2 - \|(1-\gamma)\beta r^{k+1} + \beta B(y^k - y^{k+1})\|^2\big) \\ &= \beta\left[(2-\gamma)\|r^{k+1}\|^2 + 2(By^k - By^{k+1})^\top r^{k+1}\right].\end{aligned} \tag{3.17}$$

On the other hand, invoking (3.1) and the definition of the matrix H in (3.7), we get

$$\begin{aligned}&\frac{1}{2}\big(\|w^k - \tilde{w}^k\|_H^2 - \|w^{k+1} - \tilde{w}^k\|_H^2\big) \\ &= \frac{1}{2}\left[\|x^k - x^{k+1}\|_R^2 + \|y^k - y^{k+1}\|_{S+\beta B^\top B}^2 + \frac{1}{\gamma\beta}\Big(\|\lambda^k - \tilde{\lambda}^k\|^2 - \|\lambda^{k+1} - \tilde{\lambda}^k\|^2\Big)\right].\end{aligned}$$

Finally, substituting (3.17) into the above equality, and recalling the definition of $\Delta(w^k, w^{k+1})$ in (3.12), the assertion follows immediately. □

Then, we give a further analysis for the crossing term in the right-hand side of (3.13).

Lemma 3.8. *Let $\{w^{k+1}\}$ be generated by the PADMM (1.5) and $\tilde{w}^k$ be defined by (3.1). Then, we have*

$$\begin{aligned}\beta(By^k - By^{k+1})^\top r^{k+1} \geq (1-\gamma)\beta(By^k - By^{k+1})^\top r^k + \|y^k - y^{k+1}\|^2_{\Sigma_2} \\ +\frac{1}{2}\|y^k - y^{k+1}\|^2_S - \frac{1}{2}\|y^k - y^{k-1}\|^2_S,\end{aligned} \tag{3.18}$$

where r^{k+1} is defined in (3.14).

Proof. Setting $y = y^k$ in (3.5) and $y = y^{k+1}$ in (3.5) with $k := k-1$, recalling $\tilde{y}^k = y^{k+1}$ (see (3.1a)), and then adding them together, we obtain

$$\begin{aligned}&(y^k - y^{k+1})^\top \left\{-B^\top(\tilde{\lambda}^k - \tilde{\lambda}^{k-1}) + (S + \beta B^\top B)[(y^{k+1} - y^k) - (y^k - y^{k-1})]\right\} \\ &\geq \|y^k - y^{k+1}\|^2_{\Sigma_2}.\end{aligned} \tag{3.19}$$

On the other hand, due to (3.15), we get

$$\begin{aligned}\tilde{\lambda}^k - \tilde{\lambda}^{k-1} &= [\lambda^k - \beta r^{k+1} + \beta B(y^{k+1} - y^k)] - [\lambda^{k-1} - \beta r^k + \beta B(y^k - y^{k-1})] \\ &= \quad \lambda^k - \lambda^{k-1} - \beta r^{k+1} + \beta r^k + \beta B(y^{k+1} - y^k) - \beta B(y^k - y^{k-1}) \\ &\overset{(1.5c)}{=} \quad -\gamma\beta r^k + \beta r^k - \beta r^{k+1} + \beta B(y^{k+1} - y^k) - \beta B(y^k - y^{k-1}) \\ &= \quad (1-\gamma)\beta r^k - \beta r^{k+1} + \beta B(y^{k+1} - y^k) - \beta B(y^k - y^{k-1}).\end{aligned}$$

Then, substituting the above equality into (3.19), we obtain

$$\begin{aligned}\beta(By^k - By^{k+1})^\top r^{k+1} &\geq (1-\gamma)\beta(By^k - By^{k+1})^\top r^k \\ &\quad +\|y^k - y^{k+1}\|^2_S + (y^k - y^{k+1})^\top S(y^k - y^{k-1}) + \|y^k - y^{k+1}\|^2_{\Sigma_2}.\end{aligned}$$

Finally, using Cauchy-Schwarz inequality, it yields

$$(y^k - y^{k+1})^\top S(y^k - y^{k-1}) \geq -\frac{1}{2}\|y^k - y^{k+1}\|^2_S - \frac{1}{2}\|y^{k-1} - y^k\|^2_S.$$

Combining the last two inequalities, the assertion follows directly. ☐

In the following lemma, we further treat the crossing term in (3.18).

Lemma 3.9. *Let $\{w^{k+1}\}$ be generated by (1.5). Suppose that Assumptions 2.1 and 2.2 hold and $\gamma \in (0, 2)$. Then, there exists a sufficiently small positive scalar σ such that*

$$\begin{aligned}&(1-\gamma)\beta(By^k - By^{k+1})^\top r^k + \frac{1}{2}\|y^k - y^{k+1}\|^2_{S+2\Sigma_2+\beta B^\top B} + \frac{2-\gamma}{2}\beta\|r^{k+1}\|^2 \\ &\geq \frac{1}{2}C_1\beta\|By^k - By^{k+1}\|^2 + \frac{(1-\gamma)^2}{2\kappa}\beta\left(\|r^{k+1}\|^2 - \|r^k\|^2\right) \\ &\quad +\frac{1}{2}C_2\beta\|r^{k+1}\|^2 + \frac{1}{2}\sigma\|y^k - y^{k+1}\|^2_S,\end{aligned} \tag{3.20}$$

where r^k is defined in (3.14) and

$$\kappa := \frac{1}{2}\left(1 + \tau + \frac{(1-\gamma)^2}{2-\gamma}\right), \tag{3.21}$$

$$C_1 := 1 + \tau - \kappa, \tag{3.22}$$

$$C_2 := 2 - \gamma - \frac{(1-\gamma)^2}{\kappa}. \tag{3.23}$$

Proof. First, invoking (2.3) in Assumption 2.2, there exists a sufficiently small positive scalar σ $(0 < \sigma < 1)$ such that

$$2\Sigma_2 + (1-\sigma)S \succeq \beta\tau B^\top B.$$

Consequently, we get

$$\begin{aligned}
&(1-\gamma)\beta(By^k - By^{k+1})^\top r^k + \frac{1}{2}\|y^k - y^{k+1}\|^2_{S+2\Sigma_2+\beta B^\top B} + \frac{2-\gamma}{2}\beta\|r^{k+1}\|^2 \\
&\geq (1-\gamma)\beta(By^k - By^{k+1})^\top r^k + \frac{\beta}{2}(1+\tau)\|By^k - By^{k+1}\|^2 \\
&\quad + \frac{2-\gamma}{2}\beta\|r^{k+1}\|^2 + \frac{1}{2}\sigma\|y^k - y^{k+1}\|^2_S.
\end{aligned} \tag{3.24}$$

Then, using Cauchy-Schwarz inequality, we have

$$(1-\gamma)\beta(By^k - By^{k+1})^\top r^k \geq -\frac{\kappa}{2}\beta\|By^k - By^{k+1}\|^2 - \frac{(1-\gamma)^2}{2\kappa}\beta\|r^k\|^2.$$

Consequently, substituting the last inequality into (3.24), and combining the definitions of C_1 in (3.22) and C_2 in (3.23), the conclusion follows directly. $\square$

The lemma below shows that both of the coefficients C_1, C_2 are positive when γ belongs to a certain range larger than $(0, \frac{1+\sqrt{5}}{2})$.

Lemma 3.10. *Assume that $\tau \in (0, +\infty)$, and the step size γ in (1.5) satisfies*

$$\gamma \in \left(0, \frac{1 - \tau + \sqrt{\tau^2 + 6\tau + 5}}{2}\right). \tag{3.25}$$

Then, we have

i) $(1+\tau) - \frac{(1-\gamma)^2}{2-\gamma} > 0$*;*

ii) $C_1 > 0$ *and* $C_2 > 0$*, where* C_1*,* C_2 *are defined in (3.22), (3.23), respectively.*

Proof. For assertion i), we first observe that

$$1 + \tau - \frac{(1-\gamma)^2}{2-\gamma} > 0 \Leftrightarrow \gamma^2 + (\tau - 1)\gamma - (1 + 2\tau) < 0.$$

Then, it is easy to show that the above last inequality holds when γ satisfies (3.25). For assertion ii), note that the upper bound for γ in (3.25) is an increasing function with respect to the parameter τ. Moreover, the limit value of the upper bound is 2 when τ tends to positive infinity. Thus, we have $\gamma \in (0, \frac{1-\tau+\sqrt{\tau^2+6\tau+5}}{2}) \subset (0, 2)$. Consequently, we have

$$\gamma \text{ satisfies (3.25)} \overset{(a)}{\Rightarrow} \frac{(1-\gamma)^2}{2-\gamma} < \kappa < 1 + \tau \overset{(b)}{\Rightarrow} C_1 > 0,\ C_2 > 0.$$

In fact, for implication (a), (3.25) guarantees the inequality $\frac{(1-\gamma)^2}{2-\gamma} < 1+\tau$ holds. Note that κ defined in (3.21) is the middle point of the interval $\left(\frac{(1-\gamma)^2}{2-\gamma}, 1+\tau\right)$. Thus, the implication (a) is obviously true. The implication (b) follows directly from the definitions of C_1 in (3.22) and C_2 in (3.23). $\square$

The coming theorem is the basis for establishing the global convergence and estimating the convergence rate for the sequence generated by the PADMM (1.5).

Theorem 3.11. *Let $\{w^{k+1}\}$ be generated by the PADMM (1.5) and $\tilde{w}^k$ be defined in (3.1). Suppose that Assumptions 2.1 and 2.2 hold. Then, for any $w \in \Omega$, we have*

$$\begin{aligned}\Phi(w^{k+1}, w^k, w) &\le \Phi(w^k, w^{k-1}, w) + \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(w)\\ &\quad - \left\{\frac{1}{2}C_1\beta\|By^k - By^{k+1}\|^2 + \frac{1}{2}C_2\beta\|r^{k+1}\|^2 + \frac{1}{2}\|x^k - x^{k+1}\|_R^2 + \frac{1}{2}\sigma\|y^k - y^{k+1}\|_S^2\right\},\end{aligned} \tag{3.26}$$

with

$$\Phi(w^{k+1}, w^k, w) := \frac{1}{2}\|w^{k+1} - w\|_H^2 + \frac{1}{2}\|y^k - y^{k+1}\|_S^2 + \frac{(1-\gamma)^2}{2\kappa}\beta\|r^{k+1}\|^2, \tag{3.27}$$

where κ is defined in (3.21).

Proof. First, substituting (3.18) into (3.13), we get

$$\begin{aligned}&\Delta(w^k, w^{k+1})\\ &\ge \frac{1}{2}\Big(\|x^k - x^{k+1}\|_R^2 + \|y^k - y^{k+1}\|_{S+2\Sigma_2+\beta B^\top B}^2 + (2-\gamma)\beta\|r^{k+1}\|^2\\ &\quad + 2(1-\gamma)\beta(By^k - By^{k+1})^\top r^k\Big) + \frac{1}{2}\|y^k - y^{k+1}\|_S^2 - \frac{1}{2}\|y^k - y^{k-1}\|_S^2.\end{aligned}$$

Then, using (3.20) and the definitions of C_1 and C_2 in (3.22), (3.23), respectively, we have

$$\begin{aligned}&\Delta(w^k, w^{k+1})\\ &\ge \frac{1}{2}C_1\beta\|By^k - By^{k+1}\|^2 + \frac{(1-\gamma)^2}{2\kappa}\beta\left(\|r^{k+1}\|^2 - \|r^k\|^2\right) + \frac{1}{2}C_2\beta\|r^{k+1}\|^2\\ &\quad + \frac{1}{2}\|x^k - x^{k+1}\|_R^2 + \frac{1}{2}\|y^k - y^{k+1}\|_S^2 - \frac{1}{2}\|y^k - y^{k-1}\|_S^2 + \frac{1}{2}\sigma\|y^k - y^{k+1}\|_S^2.\end{aligned}$$

Consequently, invoking Theorem 3.6 and the definition $\Delta(w^k, w^{k+1})$ in (3.12), and then combining the above inequality, we obtain

$$\begin{aligned}&\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k)\\ &\ge \frac{1}{2}\left(\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2\right) + \frac{1}{2}\|x^k - x^{k+1}\|_R^2\\ &\quad + \frac{1}{2}C_1\beta\|By^k - By^{k+1}\|^2 + \frac{(1-\gamma)^2}{2\kappa}\beta\left(\|r^{k+1}\|^2 - \|r^k\|^2\right) + \frac{1}{2}C_2\beta\|r^{k+1}\|^2\\ &\quad + \frac{1}{2}\|y^k - y^{k+1}\|_S^2 - \frac{1}{2}\|y^k - y^{k-1}\|_S^2 + \frac{1}{2}\sigma\|y^k - y^{k+1}\|_S^2.\end{aligned}$$

Finally, invoking Lemma 3.1 and using the definition $\Phi(w^{k+1}, w^k, w)$ in (3.27), the assertion (3.26) follows directly. $\square$

Based on Theorem 3.11, we can show the following contractive property of the sequence $\{w^k\}$ generated by the PADMM (1.5).

Theorem 3.12. *Let the sequence $\{w^k\}$ be generated by the PADMM* (1.5)*. Suppose that Assumptions 2.1 and 2.2 hold. Then, for any $w^* \in \Omega^*$, we have*

$$\begin{aligned}&\Phi(w^{k+1}, w^k, w^*) \le \Phi(w^k, w^{k-1}, w^*)\\ &-\left\{\frac{1}{2}C_1\beta\|By^k - By^{k+1}\|^2 + \frac{1}{2}C_2\beta\|r^{k+1}\|^2 + \frac{1}{2}\|x^k - x^{k+1}\|_R^2\right.\\ &\quad\left.+\frac{1}{2}\|u^{k+1} - u^*\|_\Sigma^2 + \frac{1}{2}\sigma\|y^k - y^{k+1}\|_S^2\right\}.\end{aligned} \tag{3.28}$$

Proof. First, using the definition of $\tilde{w}^k$ in (3.1) and combining (2.7) with setting $w = w^{k+1}$, we have

$$\begin{aligned}&\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^\top F(w^*)\\ &= \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^\top F(w^*)\\ &\ge \frac{1}{2}\|u^{k+1} - u^*\|_\Sigma^2.\end{aligned}$$

Setting $w = w^*$ in (3.26)-(3.27), and combining the above inequality, the assertion (3.28) follows directly. □

4 Convergence analysis

In this section, we show the global convergence and estimate the convergence rate in terms of the iteration complexity for the PADMM (1.5).

4.1 Global convergence

We summarize the global convergence of (1.5) in the following theorem.

Theorem 4.1. *Suppose Assumptions 2.1 and 2.2 hold. Assume that the step size γ satisfies (3.25). Let the sequence $\{w^k\}$ be generated by the PADMM* (1.5)*. Then, the sequence $\{w^k\}$ converges to a solution point $w^\infty \in \Omega^*$.*

Proof. Let (x^0, y^0, λ^0) be the initial iterate. According to Theorem 3.12, we have

$$\begin{aligned}&\Phi(w^{k+1}, w^k, w^*) \le \Phi(w^k, w^{k-1}, w^*)\\ &-\left(\frac{1}{2}C_1\beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{2}C_2\beta\|r^{k+1}\|^2\right.\\ &\left.+\frac{1}{2}\|x^{k+1} - x^k\|_R^2 + \frac{1}{2}\|u^{k+1} - u^*\|_\Sigma^2 + \frac{1}{2}\sigma\|y^k - y^{k+1}\|_S^2\right), \qquad \forall\, w^* \in \Omega^*.\end{aligned} \tag{4.1}$$

It follows from the above inequality that

$$\sum_{k=1}^{\infty}\left(\frac{1}{2}C_1\beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{2}C_2\beta\|r^{k+1}\|^2 + \frac{1}{2}\|x^{k+1} - x^k\|_R^2\right.$$

$$+\frac{1}{2}\sigma\|y^k-y^{k+1}\|_S^2+\frac{1}{2}\|u^{k+1}-u^*\|_\Sigma^2\Big)\leq \Phi(w^1,w^0,w^*).$$

Invoking Lemma 3.10, it yields $C_1, C_2 > 0$, and hence we have

$$\begin{aligned}&\lim_{k\to\infty}\|B(y^k-y^{k+1})\|=0, \quad \lim_{k\to\infty}\|r^{k+1}\|=0,\\ &\lim_{k\to\infty}\|x^{k+1}-x^k\|_R=0, \ \lim_{k\to\infty}\|y^k-y^{k+1}\|_S=0,\end{aligned} \tag{4.2}$$

and

$$\lim_{k\to\infty}\|x^{k+1}-x^*\|_{\Sigma_1}=0, \ \lim_{k\to\infty}\|y^{k+1}-y^*\|_{\Sigma_2}=0. \tag{4.3}$$

Then, it follows from the last equations that the sequences $\{\Sigma_1 x^{k+1}\}$ and $\{\Sigma_2 y^{k+1}\}$ are bounded. It follows from (3.28) and the definitions of $\Phi(w^{k+1}, w^k, w)$ in (3.27) , H in (3.7) that the sequences $\{Rx^{k+1}\}$, $\{(S+\beta B^\top B)y^{k+1}\}$ and $\{\lambda^k\}$ are all bounded. Combining the boundedness of the sequences $\{\Sigma_2 y^{k+1}\}$ and $\{(S+\beta B^\top B)y^{k+1}\}$, the sequence $\{(S+\Sigma_2+\beta B^\top B)y^{k+1}\}$ is bounded. Together with the positive definiteness of $(S+\Sigma_2+\beta B^\top B)$, this implies the boundedness of $\{y^{k+1}\}$. Noting $Ax^*+By^*=b$, and using the triangle inequality, we have

$$\|A(x^{k+1}-x^*)\|\leq\|r^{k+1}\|+\|By^{k+1}-By^*\|.$$

We further obtain the boundedness of the sequence $\{Ax^{k+1}\}$. Combining with the boundedness of the sequences $\{Rx^{k+1}\}$, $\{\Sigma_1 x^{k+1}\}$, and the positive definiteness of $(R+\Sigma_1+\beta A^\top A)$, the sequence $\{x^{k+1}\}$ is bounded. Thus, the subsequence $\{w^k\}$ is bounded. Hence, there exists a point $w^\infty\in\Omega$ and a subsequence $\{w^{k_t}\}$ such that

$$\lim_{t\to\infty}\|w^{k_t}-w^\infty\|=0. \tag{4.4}$$

According to the optimality condition of (1.5), we have

$$\begin{cases}\theta_1(x)-\theta_1(x^{k+1})\\ \quad+(x-x^{k+1})^\top\{-A^\top\left[\lambda^{k+1}-(1-\gamma)\beta r^{k+1}-\beta B(y^k-y^{k+1})\right]+R(x^{k+1}-x^k)\}\geq 0,\\ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^\top\{-B^\top\left[\lambda^{k+1}+(\gamma-1)\beta r^{k+1}\right]+S(y^{k+1}-y^k)\}\geq 0,\\ (\lambda-\lambda^{k+1})^\top[Ax^{k+1}+By^{k+1}-b-\frac{1}{\beta\gamma}(\lambda^k-\lambda^{k+1})]\geq 0,\end{cases}$$
$$\forall\, w\in\Omega.$$

Setting $k=k_t-1$ in the above inequality system, and then letting $t\to\infty$ and using (4.2), (4.4) and the lower-semicontinuity of θ_1 and θ_2, we have

$$\begin{cases}\theta_1(x)-\theta_1(x^\infty)+(x-x^\infty)^\top\{-A^\top\lambda^\infty\}\geq 0,\\ \theta_2(y)-\theta_2(y^\infty)+(y-y^\infty)^\top\{-B^\top\lambda^\infty\}\geq 0,\\ (\lambda-\lambda^\infty)^\top(Ax^\infty+By^\infty-b)\geq 0, \qquad \forall\, w\in\Omega.\end{cases}$$

It implies that w^∞ is a solution point. Hence, (4.1) is also valid if w^* is replaced by w^∞, i.e.,

$$\Phi(w^{k+1},w^k,w^\infty)\leq\Phi(w^k,w^{k-1},w^\infty).$$

Since $\Phi(w^{k+1}, w^k, w^\infty) \geq 0$, it implies that $\lim_{k\to\infty} \Phi(w^{k+1}, w^k, w^\infty)$ exists. Thus, we get

$$\lim_{k\to\infty} \Phi(w^{k+1}, w^k, w^\infty) = \lim_{t\to\infty} \Phi(w^{k_t}, w^{k_t-1}, w^\infty) = 0, \tag{4.5}$$

where the last equality follows from the definition of $\Phi(w^{k+1}, w^k, w)$ in (3.27) and (4.4) and (4.2). Then, it follows from (4.5) that

$$\lim_{k\to\infty} \|x^{k+1} - x^\infty\|_R = 0,\ \lim_{k\to\infty} \|y^{k+1} - y^\infty\|_{S+\beta B^\top B} = 0,\ \lim_{k\to\infty} \|\lambda^{k+1} - \lambda^\infty\| = 0. \tag{4.6}$$

Moreover, (4.3) is also valid if w^* is replaced by w^∞, i.e.,

$$\lim_{k\to\infty} \|x^{k+1} - x^\infty\|_{\Sigma_1} = 0,\ \lim_{k\to\infty} \|y^{k+1} - y^\infty\|_{\Sigma_2} = 0. \tag{4.7}$$

Hence, we can show

$$\lim_{k\to\infty} \|y^{k+1} - y^\infty\|_{\Sigma_2+\beta B^\top B+S} = 0.$$

Using the positive definiteness of $(S + \Sigma_2 + \beta B^\top B)$, we get

$$\lim_{k\to\infty} \|y^{k+1} - y^\infty\| = 0.$$

Then, using the triangle inequality, we have

$$\|Ax^{k+1} - Ax^\infty\| \leq \|r^{k+1}\| + \|By^{k+1} - By^\infty\|.$$

Thus, it follows that

$$\lim_{k\to\infty} \|Ax^{k+1} - Ax^\infty\| = 0.$$

Then, combining (4.6) and (4.7) with the above equality, we obtain

$$\lim_{k\to\infty} \|x^{k+1} - x^\infty\|_{\Sigma_1+\beta A^\top A+R} = 0.$$

Thus, due to the positive definiteness of $(R + \Sigma_1 + \beta A^\top A)$, we get

$$\lim_{k\to\infty} \|x^{k+1} - x^\infty\| = 0.$$

Therefore, we have shown that the whole sequence $\{w^k\}$ converges to w^∞, which is a solution point. This completes the proof. $\square$

Remark 4.2. From the proof, we see that Theorem 4.1 is also valid even if (2.4) holds and B has full column rank.

Remark 4.3. As we mentioned before, the PADMM (1.5) is reduced to the classical ADMM (1.4) when $R = S = 0$. Therefore, the convergence of the ADMM (1.4) with γ satisfying (3.25) can also be established under Assumptions 2.1 and 2.2 with $R = S = 0$.

$\boxed{4.2}$ Convergence rate

To estimate the convergence rate in terms of the iteration complexity, we first recall the concept the partial primal-dual gap in [3] which is defined as:

$$\mathcal{G}_{\mathcal{D}}(\tilde{w}) := \sup_{w\in\mathcal{D}}\big\{\mathcal{L}(\tilde{u},\lambda) - \mathcal{L}(u,\tilde{\lambda})\big\},\ \forall\, \tilde{w}\in\Omega,$$

where the set $\mathcal{D}$ is a subset of Ω.

With simple calculations, we have

$$\mathcal{G}_{\mathcal{D}}(\tilde{w}) = \sup_{w\in\mathcal{D}}\big\{\theta(\tilde{u}) - \theta(u) + (\tilde{w}-w)^\top F(w)\big\}.$$

As shown in [3], if $\mathcal{D}$ contains a solution of the VI (2.7), we have $\mathcal{G}_{\mathcal{D}}(\tilde{w}) \geq 0$ for any $\tilde{w}\in\Omega$. If $\tilde{w}$ lies in the interior of $\mathcal{D}$ and $\mathcal{G}_{\mathcal{D}}(\tilde{w}) = 0$, then $\tilde{w}$ is a saddle point of $\mathcal{L}$ defined in (2.5).

Next, we define the average of the past t iterations as follows:

$$\tilde{x}_t := \frac{1}{t}\sum_{k=1}^{t}\tilde{x}^k,\quad \tilde{y}_t := \frac{1}{t}\sum_{k=1}^{t}\tilde{y}^k,\quad \tilde{u}_t := \frac{1}{t}\sum_{k=1}^{t}\tilde{u}^k,\quad \text{and}\quad \tilde{w}_t := \frac{1}{t}\sum_{k=1}^{t}\tilde{w}^k. \tag{4.8}$$

Obviously, $\tilde{w}_t \in \Omega$ because of the convexity of Ω. In the following, we characterize the convergence rate of $\tilde{w}_t$ defined in (4.8) in terms of the partial primal-dual gap, and the feasibility violation and the decrement of the objective function.

Theorem 4.4. *Suppose Assumptions 2.1 and 2.2 hold. Assume that the step size γ satisfies (3.25). Let the sequence $\{w^k\}$ be generated by the PADMM* (1.5)*, and $\tilde{x}_t$, $\tilde{y}_t$, $\tilde{u}_t$ $\tilde{w}_t$ be defined in (4.8). Then, the following assertions hold.*

1) *Let $\tilde{w}_t$ be defined in (4.8). There exists a point $w^\infty\in\Omega^*$ such that*

$$\lim_{t\to\infty}\tilde{w}_t = w^\infty.$$

[2)] *For $\bar{C}_1 := \frac{1}{2}\|y^0-y^1\|_S^2 + \frac{(1-\gamma)^2}{2\kappa}\beta\|r^1\|^2$ with κ defined in (3.21), we have*

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^\top F(w) \leq \frac{1}{t}\left[\frac{1}{2}\|w^1-w\|_H^2 + \bar{C}_1\right]. \tag{4.9}$$

3) *There exists a constant $\bar{C}_2 > 0$ such that*

$$\|A\tilde{x}_t + B\tilde{y}_t - b\|^2 \leq \frac{\bar{C}_2}{t^2}. \tag{4.10}$$

4) *There exists a constant $\bar{C}_3 > 0$ such that*

$$|\theta(\tilde{u}_t) - \theta(u^\infty)| \leq \frac{\bar{C}_3}{t}. \tag{4.11}$$

Proof. 1) First, it follows from Theorem 4.1 that there exists $w^\infty\in\Omega^*$ such that w^k converges to it. Then, from (4.8), we have

$$\lim_{t\to\infty}\tilde{w}_t = \lim_{t\to\infty}\frac{1}{t}\sum_{k=1}^{t}\tilde{w}^k = \lim_{k\to\infty}\tilde{w}^k = \lim_{k\to\infty}w^{k+1} = w^\infty,$$

where the second equality is due to Stolz-Cesàro Theorem (see, e.g. [1]), and the third follows from (3.1), (3.16) and Theorem 4.1.

2) It follows from Theorem 3.11 that

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \geq \Phi(w^{k+1}, w^k, w) - \Phi(w^k, w^{k-1}, w), \ \ \forall\, w \in \Omega. \tag{4.12}$$

Then, summarizing the inequalities (4.12) over $k = 1, \ldots, t$, we obtain

$$t\theta(u) - \sum_{k=1}^{t} \theta(\tilde{u}^k) + \Big(tw - \sum_{k=1}^{t} \tilde{w}^k\Big)^\top F(w) + \Phi(w^1, w^0, w) \geq 0, \quad \forall\, w \in \Omega.$$

Then, using the notation of $\tilde{w}_t$ in (4.8), the last inequality can be written as

$$\frac{1}{t}\sum_{k=0}^{t} \theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^\top F(w) \leq \frac{1}{t}\Phi(w^1, w^0, w), \quad \forall\, w \in \Omega. \tag{4.13}$$

It follows from the definition of $\tilde{u}_t$ in (4.8) and the convexity of $\theta(u)$, it follows that

$$\theta(\tilde{u}_t) \leq \frac{1}{t}\sum_{k=1}^{t} \theta(\tilde{u}^k).$$

Substituting them into (4.13), the assertion (4.9) follows directly.

3) By invoking Theorem 4.1, we see that the sequence $\{w^k\}$ converges to a solution point $w^\infty \in \Omega^*$. Thus, the sequence $\{w^k\}$ is bounded. Let us define

$$\bar{C}_2 := \frac{2}{\beta^2\gamma^2}\left(\|\lambda^1 - \lambda^\infty\|^2 + \sup_k[\|\lambda^{k+1} - \lambda^\infty\|^2]\right)$$

which is a constant independent of k. Then, we have

$$\begin{aligned} \|A\tilde{x}_t + B\tilde{y}_t - b\|^2 &= \left\|\frac{1}{t}\sum_{k=1}^{t}\left[A\tilde{x}^k + B\tilde{y}^k - b\right]\right\|^2 = \left\|\frac{1}{t}\sum_{k=1}^{t}\left[\frac{1}{\beta\gamma}(\lambda^k - \lambda^{k+1})\right]\right\|^2 \\ &= \left\|\frac{1}{t\beta\gamma}\left(\lambda^1 - \lambda^{t+1}\right)\right\|^2 \leq \frac{2}{\beta^2\gamma^2 t^2}\left(\|\lambda^1 - \lambda^\infty\|^2 + \|\lambda^{k+1} - \lambda^\infty\|^2\right) = \frac{\bar{C}_2}{t^2}, \end{aligned}$$

where the first equality follows from (4.8), the second follows from (1.5c) and the fourth follows from Cauchy-Schwarz inequality. The assertion (4.10) is proved immediately.

4) It follows from $\mathcal{L}(\tilde{u}_t, \lambda^\infty) \geq \mathcal{L}(u^\infty, \lambda^\infty)$ that

$$\begin{aligned} \theta(\tilde{u}_t) - \theta(u^\infty) &\geq \langle \lambda^\infty, A\tilde{x}_t + B\tilde{y}_t - b\rangle \geq -\frac{1}{2}\left(\frac{1}{t}\|\lambda^\infty\|^2 + t\|A\tilde{x}_t + B\tilde{y}_t - b\|^2\right) \\ &\geq -\frac{1}{2t}(\|\lambda^\infty\|^2 + \bar{C}_2), \end{aligned} \tag{4.14}$$

where the second inequality is because of the Cauchy-Schwarz inequality and the last is due to (4.10). On the other hand, setting $w := w^\infty$ in (4.9), we obtain

$$\theta(\tilde{u}_t) - \theta(u^\infty) + (\tilde{w}_t - w^\infty)^\top F(w^\infty) \leq \frac{1}{t}\left[\frac{1}{2}\|w^1 - w^\infty\|_H^2 + \bar{C}_1\right].$$

Invoking the definition of F in (2.7b), we have

$$(\tilde{w}_t - w^\infty)^\top F(w^\infty) = -\langle \lambda^\infty, A\tilde{x}_t + B\tilde{y}_t - b\rangle \geq -\frac{1}{2t}(\|\lambda^\infty\|^2 + \bar{C}_2),$$

where the last inequality is similar to (4.14). Combining these two inequalities above, we get

$$\theta(\tilde{u}_t) - \theta(u^\infty) \leq \frac{1}{t}\left[\frac{1}{2}\|w^1 - w^\infty\|_H^2 + \bar{C}_1\right] + \frac{1}{2t}(\|\lambda^\infty\|^2 + \bar{C}_2). \tag{4.15}$$

The inequalities (4.14) and (4.15) indicate that the assertion (4.11) holds by setting $\bar{C}_3 := \left[\frac{1}{2}\|w^1 - w^\infty\|_H^2 + \bar{C}_1\right] + \frac{1}{2}(\|\lambda^\infty\|^2 + \bar{C}_2)$. □

From the proof, we can see that the assertions 3) and 4) of Theorem 4.4 follow directly from the first two assertions. Suppose that the set $\mathcal{D}$ is compact and contains the sequence $\{w^k\}$ and its limit w^∞, then we define

$$\tilde{\Lambda} := \sup_{w\in\mathcal{D}}\{\frac{1}{2}\|w^1 - w\|_H^2 + \bar{C}_1\}.$$

Thus, after t iterations of the PADMM (1.5), the point $\tilde{w}_t$ defined in (4.8) satisfies

$$\mathcal{G}_{\mathcal{D}}(\tilde{w}_t) \leq \frac{\tilde{\Lambda}}{t}.$$

Therefore, we establish the $O(1/t)$ ergodic convergence rate measured by the partial primal-dual gap, the primal feasibility and the objective function, respectively.

5 Conclusions

The proximal alternating direction method of multipliers (PADMM) (including ADMM as a special case) is a popular and efficient method for separable convex programming. Moreover, Glowinski proposed a large dual step size belonging $(0, \frac{1+\sqrt{5}}{2})$ to accelerate the numerical performance of ADMM. However, it was unknown whether or not the step size in the Glowinski' ADMM can be further enlarged; a case with potential advantages in numerics. We carry out a rigorous convergence analysis in a more general framework, i.e. PADMM under some moderate and checkable conditions. Furthermore, we also establish the worst-case $O(1/t)$ convergence rate in the ergodic sense, measured by the partial primal-dual gap, the feasibility violation and the decrement of the objective function.

References

[1] J.M. Ash, A. Berele and S. Catoiu, Plausible and genuine extensions of L'Hospital's rule, *Mathematics Magazine* 85 (2012) 52–60.

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

[3] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *Journal of Mathematical Imaging and Vision*, 40 (2011) 120–145.

[4] R.H. Chan, M. Tao and X.M. Yuan, Linearized alternating direction method of multipliers for constrained linear least-squares problem, *East Asian Journal on Applied Mathematics* 2 (2012). 326–341.

[5] C.H. Chen, B.S. He and X.M. Yuan, Matrix completion via alternating direction method, *IMA Journal of Numerical Analysis* 32 (2012) 227–245.

[6] L. Chen, D.F. Sun and K.C. Toh, A note on the convergence of ADMM for linearly constrained convex optimization problems, *Computational Optimization and Applications* 66 (2017) 327–343.

[7] J. Eckstein, Some saddle-function splitting methods for convex programming, *Optimization Methods and Software* 4 (1994) 75–83.

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

[9] J. Eckstein and M. Fukushima, Reformulations and applications of the alternating direction method of multipliers, *Large Scale Optimization*, W.W. Hager, D.W. Hearn and P.M. Pardalos (eds.), Springer, Boston, MA, 1994, pp. 115–134.

[10] J. Eckstein and W. Yao, Understanding the convergence of the alternating direction method of multipliers: theoretical and computational perspectives, *Pacific Journal of Optimization* 11 (2015) 619–644.

[11] M. Fazel, T.K. Pong, D.F. Sun and P. Tseng, Hankel matrix rank minimization with applications to system identification and realization, *SIAM Journal on Matrix Analysis and Applications* 34 (2013) 946–977.

[12] D. Gabay, Applications of the method of multipliers to variational inequalities, *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, M. Fortin and R. Glowinski (eds)., North Holland, Amsterdam, The Netherlands, 1983, pp. 299–331.

[13] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite-element approximations, *Computers* & *Mathematics with Applications* 2 (1976) 17–40.

[14] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[15] R. Glowinski, On alternating direction methods of multipliers: A historical perspective, *In Modeling, Simulation and Optimization for Science and Technology*, W. Fitzgibbon, Y. A. Kuznetsov, P. Neittaanmaki & O. Pironneau, eds., Computational Methods in Applied Sciences, Vol. 34, Springer, Dordrecht, 2014, pp. 59–82.

[16] R. Glowinski and A. Marrocco, Sur l'approximation par éléments finis d'ordre un et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires, *Revue francaise d'automatique, informatique, recherche opacuteerationnelle., Analyse Numérique* 2 (1975) 41–76.

[17] B.S. He, L.Z. Liao, D. Han, and H. Yang, A new inexact alternating directions method for monontone variational inequalities, *Mathematical Programming* 92 (2002) 103–118.

[18] B.S. He, F. Ma and X.M. Yuan, Linearized alternating direction method of multipliers via positive-indefinite proximal regularization for convex programming, available at http://www.optimization-online.org/DB_HTML/2016/07/5569.html.

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

[20] B.S. He, M.H. Xu and X.M. Yuan, Solving large-scale least squares covariance matrix problems by alternating direction methods, *SIAM Journal on Matrix Analysis and Applications* 32 (2011). 136–152.

[21] M.R. Hestenes, Multiplier and gradient methods, *Journal of Optimization Theory and Applications* 4 (1969) 303–320.

[22] M.J.D. Powell, A method for nonlinear constraints in minimization problems, *In Optimization edited by R. Fletcher*, Academic Press, New York, 1969, pp. 283–298.

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

[24] D.F. Sun, K.C. Toh and L.Q. Yang, A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type constraints, *SIAM Journal on Optimization* 25 (2015) 882–915.

[25] Z.W. Wen, D. Goldfarb and W.T. Yin, Alternating direction augmented Lagrangian methods for semidefinite programming, *Mathematical Programming Computation* 2 (2010) 203–230.

[26] M.H. Xu, Proximal alternating directions method for structured variational inequalities, *Journal of Optimization Theory and Applications* 134 (2007) 107–117.

Manuscript received 3 September 2016
revised 21 Dcember 2016, 6 March 2017
accepted for publication 11 April 2017

M. Tao
Department of Mathematics, National Key Laboratory for Novel Software Technolog
Nanjing University, Nanjing, 210093, China
E-mail address: taom@nju.edu.cn