凸优化和单调变分不等式的收缩算法

第十七讲: 三块可分离凸优化问题的略有改动的交替方向法

A slightly changed ADMM for convex optimization with three separable blocks

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the publication [10] and [13]

Abstract. The classical alternating direction method of multipliers (ADMM) has been well studied in the context of linearly constrained convex programming and variational inequalities where the involved operator is formed as the sum of two individual functions without crossed variables. Recently, ADMM has found many novel applications in diversified areas such as image processing and statistics. However, it is still not clear whether ADMM can be extended to the case where the operator is the sum of more than two individual functions. In this lecture, we present a little changed ADMM for solving the linearly constrained separable convex optimization whose involved operator is separable into three individual functions. The $\mathcal{O}(1/t)$ convergence rate of the proposed methods is demonstrated.

Keywords: Alternating direction method, convex programming, linear constraint, separable structure, contraction method

---

1 Introduction

An important case of structured convex optimization problem is

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

where $\theta_1 : \Re^n \to \Re$ and $\theta_2 : \Re^m \to \Re$ are closed proper convex functions (not necessarily smooth); $A \in \Re^{m\times n}$; $\mathcal{X} \subseteq \Re^n$ and $\mathcal{Y} \subseteq \Re^m$ are closed convex sets. The alternating direction method of multipliers (ADMM), which dates back to [6] and is closely related to the Douglas-Rachford operator splitting method [2], is perhaps the most popular method for solving (1.1). More specifically, for given (y^k, λ^k) in the k-th iteration, it produces the new iterate in the following order:

$$\left\{\begin{array}{l} x^{k+1} = \text{Argmin}\big\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + y^k - b\|^2 \;\big|\; x \in \mathcal{X}\big\}; \\ y^{k+1} = \text{Argmin}\big\{\theta_2(y) - (\lambda^k)^T y + \frac{\beta}{2}\|Ax^{k+1} + y - b\|^2 \big|\; y \in \mathcal{Y}\big\}; \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + y^{k+1} - b). \end{array}\right. \tag{1.2}$$

Therefore, ADMM can be viewed as a practical and structured-exploiting variant (split form or relaxed form) of ALM for solving the separable problem (1.1), with the adaption of minimizing the involved separable variables x and y separably in an alternating order. In fact, the iteration (1.2) is from (y^k, λ^k) to (y^{k+1}, λ^{k+1}), x is only an auxiliary variable in the iterative process. The sequence $\{(y^k, \lambda^k)\}$ generated by the recursion (1.2) satisfies (see Theorem 1 in [12] by setting fixed β and $\gamma \equiv 1$)

$$\begin{aligned} &\|\beta(y^{k+1} - y^*)\|^2 + \|\lambda^{k+1} - \lambda^*\|^2 \\ \le \quad & \|\beta(y^k - y^*)\|^2 + \|\lambda^k - \lambda^*\|^2 - \big(\|\beta(y^k - y^{k+1})\|^2 + \|\lambda^k - \lambda^{k+1}\|^2\big). \end{aligned}$$

Because of its efficiency and easy implementation, ADMM has attracted wide attention of many authors in various areas, see e.g. [1, 7]. In particular, some novel and attractive applications of ADMM have been discovered very recently, e.g. the total-variation problem in image processing, the covariance selection problem and semidefinite least square problem in statistics [11], the semidefinite programming problems, the sparse and low-rank recovery problem in Engineering [14], and the matrix completion problem [1].

In some practical applications [4], the model is slightly more complicated than (1.1). The mathematical form of the problem is

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) | Ax + y + z = b, \ x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}, \qquad (1.3)$$

where $\theta_1 : \Re^n \to \Re$, $\theta_2, \theta_3 : \Re^m \to \Re$ are closed proper convex functions (not necessarily smooth); $A \in \Re^{m \times n}$; $\mathcal{X} \subseteq \Re^n$, $\mathcal{Y}, \mathcal{Z} \subseteq \Re^m$ are closed convex sets. It is then natural to manage to extend ADMM to solve the problem (1.3), resulting in the following scheme:

$$\begin{cases} x^{k+1} = \text{Argmin}\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + y^k + z^k - b\|^2 \mid x \in \mathcal{X}\}; \\ y^{k+1} = \text{Argmin}\{\theta_2(y) - (\lambda^k)^T y \;+\; \frac{\beta}{2}\|Ax^{k+1} + y + z^k - b\|^2 \mid y \in \mathcal{Y}\}; \\ z^{k+1} = \text{Argmin}\{\theta_3(z) - (\lambda^k)^T z \;+\; \frac{\beta}{2}\|Ax^{k+1} + y^{k+1} + z - b\|^2 \mid z \in \mathcal{Z}\}; \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + y^{k+1} + z^{k+1} - b), \end{cases} \tag{1.4}$$

and the involved subproblems of (1.4) are solved consecutively in the ADMM manner. Unfortunately, with the $(y^{k+1}, z^{k+1}, \lambda^{k+1})$ offered by (1.4), the convergence of the extended ADMM (1.4) is still open.

In this paper, we present a little changed alternating direction method for the problem (1.3). Again, based on $(y^{k+1}, z^{k+1}, \lambda^{k+1})$ offered by (1.4), we set

$$(y^{k+1}, z^{k+1}, \lambda^{k+1}) := (y^{k+1} + (z^k - z^{k+1}), z^{k+1}, \lambda^{k+1}). \tag{1.5}$$

Note that the change of (1.5) is small. In addition, for the problem with two separable operators, by setting $z^k = 0$ for all k, the proposed method is just reduced to the algorithm (1.2) for the problem (1.1). Therefore, we call the proposed method *a little* *changed alternating direction method of multipliers for convex optimization with three separable operators*.

The outline of this paper is as follows. In Section 2, we convert the problem (1.3) to the equivalent variational inequality and characterize its solution set. Section 3 shows the contraction property of the proposed method. In Section 4, we define an auxiliary vector and derive its main associated properties, and show the $\mathcal{O}(1/t)$ convergence rate of the proposed method. Finally, some conclusions are made in Section 6.

2 The variational inequality characterization

Throughout, we assume that the solution set of (1.3) is not empty. The convergence analysis is based on the tool of variational inequality. For this purpose, we define

$$\mathcal{W} = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \Re^m.$$

It is easy to verify that the convex programming problem (1.3) is characterized by the following variational inequality: Find $w^* = (x^*, y^*, z^*, \lambda^*) \in \mathcal{W}$ such that

$$\left\{\begin{array}{l} \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0, \\ \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-\lambda^*) \geq 0, \\ \theta_3(z) - \theta_3(z^*) + (z - z^*)^T(-\lambda^*) \geq 0, \\ \qquad\qquad (\lambda - \lambda^*)^T(Ax^* + y^* + z^* - b) \geq 0, \end{array}\right. \quad \forall\, w \in \mathcal{W}, \tag{2.1}$$

or in the more compact form:

$$\text{VI}(\mathcal{W}, F, \theta) \qquad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \mathcal{W}, \tag{2.2}$$

where

$$\theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z),$$

and

$$u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \qquad w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \qquad F(w) = \begin{pmatrix} -A^T\lambda \\ -\lambda \\ -\lambda \\ Ax + y + z - b \end{pmatrix}. \qquad (2.3)$$

Note that $F(w)$ defined in (2.3) is monotone. Under the nonempty assumption on the solution set of (1.3), the solution set of (2.2)-(2.3), denoted by $\mathcal{W}^*$, is also nonempty.

The Theorem 2.3.5 in [5] provides an insightful characterization for the solution set of a generic VI. This characterization actually provides us a novel and simple approach which enables us to derive the $O(1/t)$ convergence rate for the original ADMM in [13]. In the following theorem, we specify this result for the derived VI$(\mathcal{W}, F, \theta)$. Note that the proof of the next theorem is an incremental extension of Theorem 2.3.5 in [5] and also Theorem 2.1 in [13]. But, we include all the details because of its crucial importance in our analysis.

Theorem 2.1 *The solution set of VI*$(\mathcal{W}, F, \theta)$ *is convex and it can be characterized as*

$$\mathcal{W}^* = \bigcap_{w\in\mathcal{W}} \big\{\bar{w} \in \mathcal{W} : \big(\theta(u) - \theta(\bar{u})\big) + (w - \bar{w})^T F(w) \geq 0\big\}. \tag{2.4}$$

Proof. Indeed, if $\bar{w} \in \mathcal{W}^*$, according to (2.2) we have

$$\theta(u) - \theta(\bar{u}) + (w - \bar{w})^T F(\bar{w}) \geq 0, \ \ \forall\, w \in \mathcal{W}.$$

By using the monotonicity of F on $\mathcal{W}$, this implies

$$\theta(u) - \theta(\bar{u}) + (w - \bar{w})^T F(w) \geq 0, \ \ \forall\, w \in \mathcal{W}.$$

Thus, $\bar{w}$ belongs to the right-hand set in (2.4).

Conversely, suppose $\bar{w}$ belongs to the latter set. Let $w \in \mathcal{W}$ be arbitrary. The vector

$$\tilde{w} = \tau\bar{w} + (1-\tau)w$$

belongs to $\mathcal{W}$ for all $\tau \in (0, 1)$. Thus we have

$$\theta(\tilde{u}) - \theta(\bar{u}) + (\tilde{w} - \bar{w})^T F(\tilde{w}) \geq 0. \tag{2.5}$$

Because $\theta(\cdot)$ is convex and $\tilde{u} = \tau\bar{u} + (1-\tau)u$, we have

$$\theta(\tilde{u}) \le \tau\theta(\bar{u}) + (1-\tau)\theta(u).$$

Substituting it in (2.5), we get

$$(\theta(u) - \theta(\bar{u})) + (w - \bar{w})^T F(\tau\bar{w} + (1-\tau)w) \ge 0$$

for all $\tau \in (0,1)$. Letting $\tau \to 1$ yields

$$(\theta(u) - \theta(\bar{u})) + (w - \bar{w})^T F(\bar{w}) \ge 0.$$

Thus $\bar{w} \in \mathcal{W}^*$. Now, we turn to prove the convexity of $\mathcal{W}^*$. For each fixed but arbitrary $w \in \mathcal{W}$, the set

$$\{\bar{w} \in \mathcal{W} :\ \theta(\bar{u}) + \bar{w}^T F(w) \le \theta(u) + w^T F(w)\}$$

is convex and so is the equivalent set

$$\{\bar{w} \in \mathcal{W} :\ \big(\theta(u) - \theta(\bar{u})\big) + (w - \bar{w})^T F(w) \ge 0\}.$$

Since the intersection of any number of convex sets is convex, it follows that the solution set of VI$(\mathcal{W}, F, \theta)$ is convex. $\Box$

Theorem 2.1 thus implies that $\bar{w} \in \mathcal{W}$ is an approximate solution of VI$(\mathcal{W}, F, \theta)$ with the accuracy $\epsilon > 0$ if it satisfies

$$\theta(u) - \theta(\bar{u}) + F(w)^T(w - \bar{w}) \geq -\epsilon, \ \ \forall \, w \in \mathcal{W}.$$

In this paper, we show that, for given $\epsilon > 0$ and a substantial compact set $\mathcal{D} \subset \mathcal{W}$, after t iterations of the proposed methods, we can find a $\bar{w} \in \mathcal{W}$ such that

$$\hat{w} \in \mathcal{W} \quad \text{and} \quad \sup_{w \in \mathcal{D}} \big\{ \theta(\hat{u}) - \theta(u) + (\hat{w} - w)^T F(w) \big\} \leq \epsilon. \tag{2.6}$$

The convergence rate $\mathcal{O}(1/t)$ of the proposed methods is thus established.

For convenience of coming analysis, we define the following matrices:

$$P = \begin{pmatrix} \beta I_m & 0 & 0 \\ \beta I_m & \beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}, \qquad D = \begin{pmatrix} \beta I_m & 0 & 0 \\ 0 & \beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{2.7}$$

Note that for the above defined matrices M and D, we have

$$D^{-1}P = \begin{pmatrix} I_m & 0 & 0 \\ I_m & I_m & 0 \\ 0 & 0 & I_m \end{pmatrix}, \quad P^{-T}D = \begin{pmatrix} I_m & -I_m & 0 \\ 0 & I_m & 0 \\ 0 & 0 & I_m \end{pmatrix}. \tag{2.8}$$

3 Contraction property of the proposed method

In the alternating direction method, x is only the auxiliary variable in the iteration process. For convenience of analysis, we use the notation

$$v = (y, z, \lambda),$$

which is a sub-vector of w. For $w^* \in \mathcal{W}^*$, we also define

$$\mathcal{V}^* := \{v^* = (y^*, z^*, \lambda^*) \,|\, (x^*, y^*, z^*, \lambda^*) \in \mathcal{W}^*\}.$$

In addition, we divide each iteration of the proposed method in two steps-the prediction step and the correction step. From a given $v^k = (y^k, z^k, \lambda^k)$, we use

$\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k) \in \mathcal{W}$ to denote the solution of (1.4) and call it as the prediction step.

The prediction step in the k-th iteration (use ADMM manner):

Begin with given $v^k = (y^k, z^k, \lambda^k)$, generate $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ in the following order:

$$\bar{x}^k = \text{Argmin}\big\{\theta_1(x) + \frac{\beta}{2}\|(Ax + y^k + z^k - b) - \frac{1}{\beta}\lambda^k\|^2 \mid x \in \mathcal{X}\big\}, \quad \text{(3.1a)}$$

$$\bar{y}^k = \text{Argmin}\big\{\theta_2(y) + \frac{\beta}{2}\|(A\bar{x}^k + y + z^k - b) - \frac{1}{\beta}\lambda^k\|^2 \mid y \in \mathcal{Y}\big\}, \quad \text{(3.1b)}$$

$$\bar{z}^k = \text{Argmin}\big\{\theta_3(z) + \frac{\beta}{2}\|(A\bar{x}^k + \bar{y}^k + z - b) - \frac{1}{\beta}\lambda^k\|^2 \mid z \in \mathcal{Z}\big\}, \quad \text{(3.1c)}$$

$$\bar{\lambda}^k = \lambda^k - \beta(A\bar{x}^k + \bar{y}^k + \bar{z}^k - b). \quad \text{(3.1d)}$$

For the new iterative loop, we need only to produce $v^{k+1} = (y^{k+1}, z^{k+1}, \lambda^{k+1})$. Use the notation of $\bar{w}^k$, the update form (1.5) can be written as

The correction step: Update the new iterate $v^{k+1} = (y^{k+1}, z^{k+1}, \lambda^{k+1})$ by

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ z^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_m & -I_m & 0 \\ 0 & I_m & 0 \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} y^k - \bar{y}^k \\ z^k - \bar{z}^k \\ \lambda^k - \bar{\lambda}^k \end{pmatrix}. \tag{3.2}$$

Using the notation $P^{-T}D$, the correction step can be written as

$$v^{k+1} = v^k - P^{-T}D(v^k - \bar{v}^k).$$

We consider the general correction update form

$$v^{k+1} = v^k - \alpha P^{-T}D(v^k - \bar{v}^k), \quad \alpha \in (0, 1]. \tag{3.3}$$

In other words, the update form (3.2) is a special case of (3.3) with $\alpha = 1$. Taking $\alpha \in (0, 1)$, the method is a special case of the method proposed in [10].

3.1 Properties of the vector $\bar{w}^k$ by the prediction step

We establish the following lemma.

Lemma 3.1 *Let $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ be generated by the alternating direction-prediction step* (3.1a)–(3.1d) *from the given vector $v^k = (y^k, z^k, \lambda^k)$. Then, we have $\bar{w}^k \in \mathcal{W}$ and*

$$\theta(u) - \theta(\bar{u}^k) + (w - \bar{w}^k)^T d(v^k, \bar{w}^k) \geq (v - \bar{v}^k)^T P(v^k - \bar{v}^k), \ \forall\, w \in \mathcal{W}, \quad (3.4)$$

where

$$d(v^k, \bar{w}^k) = F(\bar{w}^k) + \eta(v^k, \bar{v}^k), \quad (3.5)$$

$$\eta(v^k, \bar{v}^k) = \begin{pmatrix} A^T \\ I_m \\ I_m \\ 0 \end{pmatrix} \beta\big((y^k - \bar{y}^k) + (z^k - \bar{z}^k)\big), \quad (3.6)$$

and the matrix P is defined in (2.7).

Proof. The proof consists of some manipulations. Recall (3.1a)–(3.1c). We have $\bar{u}^k \in \mathcal{U}$ and

$$\begin{cases} \theta_1(x) - \theta_1(\bar{x}^k) + (x - \bar{x}^k)^T (A^T[\beta(A\bar{x}^k + y^k + z^k - b) - \lambda^k]) \geq 0, \\ \theta_2(y) - \theta_2(\bar{y}^k) + (y - \bar{y}^k)^T (\beta(A\bar{x}^k + \bar{y}^k + z^k - b) - \lambda^k) \geq 0, \\ \theta_3(z) - \theta_3(\bar{z}^k) + (z - \bar{z}^k)^T (\beta(A\bar{x}^k + \bar{y}^k + \bar{z}^k - b) - \lambda^k) \geq 0, \end{cases}$$

for all $u \in \mathcal{U}$. Using (3.1d), the above inequality can be written as

$$\theta(u) - \theta(\bar{u}^k) + \begin{pmatrix} x - \bar{x}^k \\ y - \bar{y}^k \\ z - \bar{z}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\bar{\lambda}^k \\ -\bar{\lambda}^k \\ -\bar{\lambda}^k \end{pmatrix} + \begin{pmatrix} \beta A^T(y^k - \bar{y}^k) + \beta A^T(z^k - \bar{z}^k) \\ \beta(z^k - \bar{z}^k) \\ 0 \end{pmatrix} \right\} \geq 0. \tag{3.7}$$

Adding the following term

$$\begin{pmatrix} x - \bar{x}^k \\ y - \bar{y}^k \\ z - \bar{z}^k \end{pmatrix}^T \begin{pmatrix} 0 \\ \beta(y^k - \bar{y}^k) \\ \beta(y^k - \bar{y}^k) + \beta(z^k - \bar{z}^k) \end{pmatrix}$$

to the both sides of the last inequality, we get $\bar{w}^k \in \mathcal{W}$ and

$$\theta(u)-\theta(\bar{u}^k)+\begin{pmatrix} x-\bar{x}^k \\ y-\bar{y}^k \\ z-\bar{z}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\bar{\lambda}^k \\ -\bar{\lambda}^k \\ -\bar{\lambda}^k \end{pmatrix} + \begin{pmatrix} A^T\beta\big((y^k-\bar{y}^k)+(z^k-\bar{z}^k)\big) \\ \beta\big((y^k-\bar{y}^k)+(z^k-\bar{z}^k)\big) \\ \beta\big((y^k-\bar{y}^k)+(z^k-\bar{z}^k)\big) \end{pmatrix} \right\}$$

$$\geq \quad \begin{pmatrix} x-\bar{x}^k \\ y-\bar{y}^k \\ z-\bar{z}^k \end{pmatrix}^T \begin{pmatrix} 0 \\ \beta(y^k-\bar{y}^k) \\ \beta(y^k-\bar{y}^k)+\beta(z^k-\bar{z}^k) \end{pmatrix}, \quad \forall\, u \in \mathcal{U}. \quad (3.8)$$

Because

$$A\bar{x}^k+\bar{y}^k+\bar{z}^k-b=\frac{1}{\beta}(\lambda^k-\bar{\lambda}^k),$$

adding the equal terms

$$(\lambda-\bar{\lambda}^k)^T(A\bar{x}^k+\bar{y}^k+\bar{z}^k-b) \quad \text{and} \quad (\lambda-\bar{\lambda}^k)^T\frac{1}{\beta}(\lambda^k-\bar{\lambda}^k)$$

to the left side and right side of (3.8), respectively, we get $\bar{w}^k \in \mathcal{W}$ and

$$\theta(u)-\theta(\bar{u}^k)+\begin{pmatrix} x-\bar{x}^k \\ y-\bar{y}^k \\ z-\bar{z}^k \\ \lambda-\bar{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\bar{\lambda}^k \\ -\bar{\lambda}^k \\ -\bar{\lambda}^k \\ A\bar{x}^k+\bar{y}^k+\bar{z}^k-b \end{pmatrix} + \begin{pmatrix} A^T\beta\big((y^k-\bar{y}^k)+(z^k-\bar{z}^k)\big) \\ \beta\big((y^k-\bar{y}^k)+(z^k-\bar{z}^k)\big) \\ \beta\big((y^k-\bar{y}^k)+(z^k-\bar{z}^k)\big) \\ 0 \end{pmatrix} \right\}$$

$$\geq \begin{pmatrix} y-\bar{y}^k \\ z-\bar{z}^k \\ \lambda-\bar{\lambda}^k \end{pmatrix}^T \begin{pmatrix} \beta(y^k-\bar{y}^k) \\ \beta(y^k-\bar{y}^k)+\beta(z^k-\bar{z}^k) \\ \frac{1}{\beta}(\lambda^k-\bar{\lambda}^k) \end{pmatrix}, \quad \forall\, w \in \mathcal{W}.$$

Using the notations of $F(w)$, $d(v^k, \bar{w}^k)$ and P, the assertion follows immediately and the lemma is proved. $\square$

Lemma 3.2 *Let* $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ *be generated by the alternating direction prediction step (3.1a)–(3.1d) from the given vector* $v^k = (y^k, z^k, \lambda^k)$*. Then for all*

$v^* \in \mathcal{V}^*$*, we have*

$$
\begin{array}{l}
(v^k - v^*)^T P(v^k - \bar{v}^k) \\
\quad \geq \quad \frac{1}{2}\|v^k - \bar{v}^k\|_D^2 + \frac{1}{2}\beta\|(y^k - \bar{y}^k) + (z^k - \bar{z}^k) + \frac{1}{\beta}(\lambda^k - \bar{\lambda}^k)\|^2, \qquad (3.9)
\end{array}
$$

where the matrices P *is defined in (2.7).*

Proof. Setting $w = w^*$ in (3.4) (notice that v^* is a sub-vector of w^*), it yields that

$$
\begin{aligned}
(\bar{v}^k - v^*)^T P(v^k - \bar{v}^k) \quad \geq \quad & \big(\theta(\bar{u}^k) - \theta(u^*)\big) + (\bar{w}^k - w^*)^T F(\bar{w}^k) \\
& + (\bar{w}^k - w^*)^T \eta(v^k, \bar{v}^k). \qquad (3.10)
\end{aligned}
$$

Now, we deal with the last term in the right hand side of the inequality (3.10). By using the notation of $\eta(v^k, \bar{v}^k)$, $Ax^* + y^* + z^* = b$ and (3.1d), we obtain

$$
\begin{aligned}
& (\bar{w}^k - w^*)^T \eta(v^k, \bar{v}^k) \\
& \quad = \quad \beta(A\bar{x}^k + \bar{y}^k + \bar{z}^k - Ax^* - y^* - z^*)^T \{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\} \\
& \quad = \quad \beta(A\bar{x}^k + \bar{y}^k + \bar{z}^k - b)^T \{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\} \\
& \quad = \quad (\lambda^k - \bar{\lambda}^k)^T \{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\}.
\end{aligned}
$$

Substituting it in (3.10), we get

$$
\begin{aligned}
(\bar{v}^k - v^*)^T P(v^k - \bar{v}^k) \;\; &\geq \;\; \big(\theta(\bar{u}^k) - \theta(u^*)\big) + (\bar{w}^k - w^*)^T F(\bar{w}^k) \\
&\quad +(\lambda^k - \bar{\lambda}^k)^T \{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\}. \quad (3.11)
\end{aligned}
$$

Since F is monotone, we have

$$
\theta(\bar{u}^k) - \theta(u^*) + (\bar{w}^k - w^*)^T F(\bar{w}^k) \geq \theta(\bar{u}^k) - \theta(u^*) + (\bar{w}^k - w^*)^T F(w^*) \geq 0.
$$

Substituting it in the right hand side of (3.11), we obtain

$$
(\bar{v}^k - v^*)^T P(v^k - \bar{v}^k) \geq (\lambda^k - \bar{\lambda}^k)^T \{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\}.
$$

It follows from the last equality that

$$
\begin{aligned}
&(v^k - v^*)^T P(v^k - \bar{v}^k) \\
&\quad \geq (v^k - \bar{v}^k)^T P(v^k - \bar{v}^k) + (\lambda^k - \bar{\lambda}^k)^T \{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\}. \quad (3.12)
\end{aligned}
$$

Observe the matrices P and D (see (2.7)), by a manipulation, the right hand side of (3.12) can be written as

$$
\begin{aligned}
&(v^k-\bar{v}^k)^T P(v^k-\bar{v}^k)+(\lambda^k-\bar{\lambda}^k)^T\{(y^k-\bar{y}^k)+(z^k-\bar{z}^k)\}\\
= \quad & \begin{pmatrix} y^k-\bar{y}^k \\ z^k-\bar{z}^k \\ \lambda^k-\bar{\lambda}^k \end{pmatrix}^T \begin{pmatrix} \beta I_m & \frac{1}{2}\beta I_m & \frac{1}{2}I_m \\ \frac{1}{2}\beta I_m & \beta I_m & \frac{1}{2}I_m \\ \frac{1}{2}I_m & \frac{1}{2}I_m & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} y^k-\bar{y}^k \\ z^k-\bar{z}^k \\ \lambda^k-\bar{\lambda}^k \end{pmatrix}\\
= \quad & \frac{1}{2}\|v^k-\bar{v}^k\|_D^2+\frac{1}{2}\beta\|(y^k-\bar{y}^k)+(z^k-\bar{z}^k)+\frac{1}{\beta}(\lambda^k-\bar{\lambda}^k)\|^2.
\end{aligned}
$$

Substituting it in the right hand side of (3.12), the assertion of this lemma is proved. □

Whenever $v^k \neq \bar{v}^k$, the right hand side of (3.9) is positive. For any positive definite matrix H, (3.9) implies that

$$
\begin{aligned}
&\langle H(v^k-v^*), H^{-1}P(v^k-\bar{v}^k)\rangle\\
&\quad \geq \quad \frac{1}{2}\|v^k-\bar{v}^k\|_D^2+\frac{1}{2}\beta\|(y^k-\bar{y}^k)+(z^k-\bar{z}^k)+\frac{1}{\beta}(\lambda^k-\bar{\lambda}^k)\|^2,
\end{aligned}
$$

and $H^{-1}P(v^k - \bar{v}^k)$ is an ascent direction of the distance function $\frac{1}{2}\|v - v^*\|_H^2$ at the point v^k. By choosing

$$H = PD^{-1}P^T = \begin{pmatrix} \beta I_m & \beta I_m & 0 \\ \beta I_m & 2\beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}, \tag{3.13}$$

and using the matrix $P^{-T}D$ (see (2.8)), we have

$$H^{-1}P = P^{-T}D = \begin{pmatrix} I_m & -I_m & 0 \\ 0 & I_m & 0 \\ 0 & 0 & I_m \end{pmatrix}.$$

3.2 Correction and the contractive property

The alternating direction-prediction step begins with given $v^k = (y^k, z^k, \lambda^k)$. In fact x is only the auxiliary variable in the iteration process. For the new loop, we need only to produce $v^{k+1} = (y^{k+1}, z^{k+1}, \lambda^{k+1})$ and we call this step as the **correction step**.

Theorem 3.1 *Let* $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ *be generated by the alternating* *direction-prediction step* (3.1a)–(3.1d) *from the given vector* $v^k = (y^k, z^k, \lambda^k)$ *and* v^{k+1} *be given by* (3.3)*. Then we have*

$$\begin{array}{rcl} \|v^{k+1} - v^*\|_H^2 & \leq & \|v^k - v^*\|_H^2 - \alpha(1-\alpha)\|v^k - \bar{v}^k\|_D^2 \\ & & -\alpha\beta\|(y^k - \bar{y}^k) + (z^k - \bar{z}^k) + \frac{1}{\beta}(\lambda^k - \bar{\lambda}^k)\|^2, \quad \forall v^* \in \mathcal{V}^*, \end{array} \quad (3.14)$$

where H *is defined in* (3.13)*.*

Proof. By using (3.3) and the definition of the matrix H, we have

$$\begin{array}{l} \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ \quad = \quad \|v^k - v^*\|_H^2 - \|(v^k - v^*) - (v^k - v^{k+1})\|_H^2 \\ \quad = \quad \|v^k - v^*\|_H^2 - \|(v^k - v^*) - \alpha P^{-T} D(v^k - \bar{v}^k)\|_H^2 \\ \quad = \quad 2\alpha(v^k - v^*)^T P(v^k - \bar{v}^k) - \alpha^2\|v^k - \bar{v}^k\|_D^2. \end{array}$$

Substituting (3.9) in the last equality, we get

$$\begin{array}{l} \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ \quad \geq \quad \alpha(1-\alpha)\|v^k - \bar{v}^k\|_D^2 + \alpha\beta\|(y^k - \bar{y}^k) + (z^k - \bar{z}^k) + \frac{1}{\beta}(\lambda^k - \bar{\lambda}^k)\|^2. \end{array}$$

The assertion of this theorem is proved. $\square$

For any $\alpha \in (0,1)$, it follows (3.14) that

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha(1-\alpha)\|v^k - \bar{v}^k\|_D^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{3.15}$$

The above inequality is essential for the global convergence of the method using update form (3.3) with $\alpha \in (0,1)$, see [10]. In fact, based on (3.15), the sequence $\{v^k\}$ is bounded, we have

$$\lim_{k\to\infty} \|v^k - \bar{v}^k\|_D^2 = 0. \tag{3.16}$$

Consequently the sequence $\{\bar{v}^k\}$ is also bounded and it converges to a limit point v^∞. On the other hand, due to Lemma 3.1, we have

$$\begin{aligned} &\theta(u) - \theta(\bar{u}^k) + (w - \bar{w}^k)^T F(\bar{w}^k) \\ &\quad \ge \quad -(w - \bar{w}^k)^T \eta(v^k, \bar{v}^k) + (v - \bar{v}^k)^T P(v^k - \bar{v}^k), \;\; \forall\, w \in \mathcal{W}. \end{aligned} \tag{3.17}$$

From (3.16) and (3.17) we can derive $v^\infty \in \mathcal{V}^*$ and the induced w^∞ is a solution of VI$(\mathcal{W}, F, \theta)$.

4 Convergence rate in an ergodic sense

The convergence in the last section is only for $\alpha \in (0, 1)$ with update form (3.3). Since (1.5), namely, the little changed alternating direction method of multipliers, is equivalent to (3.2) with $\alpha = 1$, it follows from Theorem 3.1 that

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \beta\|(y^k - \bar{y}^k) + (z^k - \bar{z}^k) + \frac{1}{\beta}(\lambda^k - \bar{\lambda}^k)\|^2, \ \forall v^* \in \mathcal{V}^*.$$

The sequence $\{\|v^k - v^*\|_H\}$ is monotonically non-increasing. However, we have not obtained the global convergence in the contraction sense. This section, however, shows that the method using update form (3.2) has convergence rate $O(1/t)$ for all $\alpha \in (0, 1]$ in an ergodic sense.

According to Theorem 2.1, for given $\epsilon > 0$ and a substantial compact set $\mathcal{D} \subset \mathcal{W}$, our task is to find a $\tilde{w}$ such that (see (2.6))

$$\tilde{w} \in \mathcal{W} \quad \text{and} \quad \sup_{w \in \mathcal{D}} \big\{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\big\} \leq \epsilon,$$

in $O(1/\epsilon)$ iterations. Generally, our complexity analysis follows the line of [3, 13], but instead of using $\bar{w}^k$ directly, we need first to introduce an auxiliary vector.

The additional auxiliary vector

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{z}^k \\ \tilde{\lambda}^k \end{pmatrix}, \quad \text{where} \quad \tilde{u}^k = \bar{u}^k \tag{4.1a}$$

and

$$\tilde{\lambda}^k = \bar{\lambda}^k - \beta\big((y^k - \bar{y}^k) + (z^k - \bar{z}^k)\big). \tag{4.1b}$$

In order to rewrite the assertion in Lemma 3.1 in form of $\tilde{w}$, we need the following lemmas to express the terms $d(v^k, \bar{w}^k)$ and $P(v^k - \bar{v}^k)$ in form of w^k and $\tilde{w}^k$.

Lemma 4.1 *For the $\tilde{w}^k$ defined in* (4.1) *and the $\bar{w}^k$ generated by* (3.1)*, we have*

$$d(v^k, \bar{w}^k) = F(\tilde{w}^k), \tag{4.2}$$

where $d(v^k, \bar{w}^k)$ defined in (3.5)*. In addition, it holds that*

$$P(v^k - \bar{v}^k) = Q(v^k - \tilde{v}^k), \tag{4.3}$$

where

$$Q = \begin{pmatrix} \beta I_m & 0 & 0 \\ \beta I_m & \beta I_m & 0 \\ -I_m & -I_m & \frac{1}{\beta} I_m \end{pmatrix}. \tag{4.4}$$

Proof. Since $\tilde{u}^k = \bar{u}^k$ and $\tilde{\lambda}^k = \bar{\lambda}^k - \beta\big((y^k - \bar{y}^k) + (z^k - \bar{z}^k)\big)$, we have

$$\beta\big((y^k - \bar{y}^k) + (z^k - \bar{z}^k)\big) = \bar{\lambda}^k - \tilde{\lambda}^k. \tag{4.5}$$

Substituting it in the notation of $\eta(v^k, \bar{v}^k)$ (see (3.6)), we get

$$\begin{aligned} d(v^k, \bar{w}^k) &= F(\bar{w}^k) + \eta(v^k, \bar{v}^k) \\ &= \begin{pmatrix} -A^T\bar{\lambda}^k \\ -\bar{\lambda}^k \\ -\bar{\lambda}^k \\ A\bar{x}^k + \bar{y}^k + \bar{z}^k - b \end{pmatrix} + \begin{pmatrix} A^T \\ I \\ I \\ 0 \end{pmatrix} (\bar{\lambda}^k - \tilde{\lambda}^k) = F(\tilde{w}^k). \end{aligned}$$

The last equation is due to $(\bar{x}^k, \bar{y}^k, \bar{z}^k) = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$. The equation (4.1b) implies

$$\begin{aligned}\frac{1}{\beta}(\lambda^k - \bar{\lambda}^k) &= \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) - \big((y^k - \bar{y}^k) + (z^k - \bar{z}^k)\big) \\ &= \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) - \big((y^k - \tilde{y}^k) + (z^k - \tilde{z}^k)\big).\end{aligned}$$

Using the matrix P (see (2.7)), (4.1a) and the last equation, we obtain

$$\begin{aligned}P(v^k - \bar{v}^k) &= \begin{pmatrix} \beta I_m & 0 & 0 \\ \beta I_m & \beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^k - \bar{y}^k \\ z^k - \bar{z}^k \\ \lambda^k - \bar{\lambda}^k \end{pmatrix} \\ &= \begin{pmatrix} \beta I_m & 0 & 0 \\ \beta I_m & \beta I_m & 0 \\ -I_m & -I_m & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ z^k - \tilde{z}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} = Q(v^k - \tilde{v}^k).\end{aligned}$$

Thus (4.3) holds and the lemma is proved. $\square$

By using (4.2), the assertion in Lemma 3.1 can be rewritten accordingly in the following lemma.

Lemma 4.2 *Let* $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ *be generated by the alternating direction-prediction step* (3.1a)–(3.1d) *from the given vector* $v^k = (y^k, z^k, \lambda^k)$ *and* $\tilde{w}^k$ *be defined by* (4.1)*. Then we have* $\tilde{w}^k \in \mathcal{W}$ *and*

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T P(v^k - \bar{v}^k), \ \forall w \in \mathcal{W}. \quad (4.6)$$

Proof. The assertion follows from (3.4), (4.2) and (4.1). □

Now, we are ready to prove the key inequalities for the convergence rate of the proposed method, which are given in the following lemmas.

Lemma 4.3 *Let* $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ *be generated by the alternating direction-prediction step* (3.1a)–(3.1d) *from the given vector* $v^k = (y^k, z^k, \lambda^k)$ *and* $\tilde{w}^k$ *be defined by* (4.1)*. If the new iterate* v^{k+1} *is updated by* (3.3)*, then we have*

$$\begin{aligned} &\big(\theta(u) - \theta(\tilde{u}^k)\big) + (w - \tilde{w}^k)^T F(\tilde{w}^k) + \frac{1}{2}(\|v - v^k\|_H^2 - \|v - v^{k+1}\|_H^2) \\ \geq\ & \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2), \quad \forall w \in \mathcal{W}, \end{aligned} \quad (4.7)$$

where matrix H *is defined in* (3.13)*.*

Proof. According to the update form (3.3), we have

$$v^k - \bar{v}^k = D^{-1}P^T(v^k - v^{k+1}).$$

Substituting it into the right hand side of (4.6) and using $PD^{-1}P^T = H$, we obtain

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (v-\tilde{v}^k)^T H(v^k - v^{k+1}), \quad \forall\, w \in \mathcal{W}. \quad (4.8)$$

By setting

$$a = v, \qquad b = \tilde{v}^k, \qquad c = v^k, \qquad \text{and} \qquad d = v^{k+1},$$

in the identity

$$(a-b)^T H(c-d) = \frac{1}{2}(\|a-d\|_H^2 - \|a-c\|_H^2) + \frac{1}{2}(\|c-b\|_H^2 - \|d-b\|_H^2),$$

we obtain

$$\begin{aligned} &(v-\tilde{v}^k)^T H(v^k - v^{k+1}) \\ = \quad & \frac{1}{2}(\|v-v^{k+1}\|_H^2 - \|v-v^k\|_H^2) + \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \end{aligned}$$

Substituting it in the right hand side of (4.8) and by a manipulation, we get (4.7) and the lemma is proved. □

Lemma 4.4 *Let* $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ *be generated by the alternating direction prediction step* (3.1a)–(3.1d) *from the given vector* $v^k = (y^k, z^k, \lambda^k)$ *and* $\tilde{w}^k$ *be defined by* (4.1)*. If the new iterate* v^{k+1} *is updated by* (3.3)*, then we have*

$$\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 = \frac{\alpha}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 + \alpha(1-\alpha)\|v^k - \bar{v}^k\|_D^2. \quad (4.9)$$

where matrix H *is defined in* (3.13)*.*

Proof. In view of the update form (3.3) and $H = PD^{-1}P^T$, we have

$$\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&\quad = \|v^k - \tilde{v}^k\|_H^2 - \|v^k - \tilde{v}^k + (v^{k+1} - v^k)\|_H^2 \\
&\quad = \|v^k - \tilde{v}^k\|_H^2 - \|v^k - \tilde{v}^k - \alpha P^{-T} D(v^k - \bar{v}^k)\|_H^2 \\
&\quad = 2\alpha(v^k - \tilde{v}^k)^T P(v^k - \bar{v}^k) - \alpha^2\|P^{-T} D(v^k - \bar{v}^k)\|_H^2 \\
&\quad = 2\alpha(v^k - \tilde{v}^k)^T P(v^k - \bar{v}^k) - \alpha^2\|v^k - \bar{v}^k\|_D^2. \qquad (4.10)
\end{aligned}$$

By using the relation in the equation (4.3) and the matrix Q (see (4.4)), we obtain

$$\begin{array}{rl} & 2(v^k-\tilde{v}^k)^T P(v^k-\bar{v}^k) \\ = & 2(v^k-\tilde{v}^k)^T Q(v^k-\tilde{v}^k) \\ = & (v^k-\tilde{v}^k)^T (Q+Q^T)(v^k-\tilde{v}^k) \\ = & \left(\begin{array}{c} y^k-\tilde{y}^k \\ z^k-\tilde{z}^k \\ \lambda^k-\tilde{\lambda}^k \end{array}\right)^T \left(\begin{array}{ccc} 2\beta I_m & \beta I_m & -I_m \\ \beta I_m & 2\beta I_m & -I_m \\ -I_m & -I_m & \frac{2}{\beta} I_m \end{array}\right) \left(\begin{array}{c} y^k-\tilde{y}^k \\ z^k-\tilde{z}^k \\ \lambda^k-\tilde{\lambda}^k \end{array}\right) \\ = & \beta\|(y^k-\tilde{y}^k)+(z^k-\tilde{z}^k)-\frac{1}{\beta}(\lambda^k-\tilde{\lambda}^k)\|^2+\|v^k-\tilde{v}^k\|_D^2. \end{array} \tag{4.11}$$

Because $\tilde{y}^k=\bar{y}^k$ and $\tilde{z}^k=\bar{z}^k$, it follows from and (4.5) that

$$\beta\|(y^k-\tilde{y}^k)+(z^k-\tilde{z}^k)-\frac{1}{\beta}(\lambda^k-\tilde{\lambda}^k)\|^2=\frac{1}{\beta}\|\lambda^k-\bar{\lambda}^k\|^2. \tag{4.12}$$

Combining (4.10), (4.11) and (4.12) together, we obtain

$$\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2=\alpha\Big(\frac{1}{\beta}\|\lambda^k-\bar{\lambda}^k\|^2+\|v^k-\tilde{v}^k\|_D^2\Big)-\alpha^2\|v^k-\bar{v}^k\|_D^2. \tag{4.13}$$

Again, because $\tilde{y}^k = \bar{y}^k$ and $\tilde{z}^k = \bar{z}^k$, we have

$$\frac{1}{\beta}\|\lambda^k - \bar{\lambda}^k\|^2 + \|v^k - \tilde{v}^k\|_D^2 = \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 + \|v^k - \bar{v}^k\|_D^2.$$

Substituting it in the right hand side of (4.13), we get

$$\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 = \frac{\alpha}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 + \alpha(1-\alpha)\|v^k - \bar{v}^k\|_D^2,$$

and thus the proof is complete. $\square$

Combining the assertions in Lemma 4.3 and Lemma 4.4, we have proved the key inequality for the proposed method, namely,

$$\big(\theta(u)-\theta(\tilde{u}^k)\big)+(w-\tilde{w}^k)^T F(\tilde{w}^k)+\frac{1}{2}(\|v-v^k\|_H^2-\|v-v^{k+1}\|_H^2) \geq 0, \ \forall w \in \mathcal{W}. \tag{4.14}$$

Note that the above inequality is true for any $\alpha \in (0,1]$. Having the key inequalities in the above lemmas, the $\mathcal{O}(1/t)$ rate of convergence (in an ergodic sense) can be obtained easily.

Theorem 4.1 *For any integer $t > 0$, we have a $\tilde{w}_t \in \mathcal{W}$ which satisfies*

$$\big(\theta(\tilde{u}_t) - \theta(u)\big) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2(t+1)} \|v - v^0\|_H^2, \ \forall w \in \mathcal{W},$$

where

$$\tilde{w}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{w}^k.$$

Proof. Since $F(\cdot)$ is monotone, it follows (4.14) that

$$\big(\theta(u) - \theta(\tilde{u}^k)\big) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}(\|v - v^k\|_H^2 - \|v - v^{k+1}\|_H^2) \ge 0, \ \forall w \in \mathcal{W},$$

or, equivalently,

$$\big(\theta(\tilde{u}^k) - \theta(u)\big) + (\tilde{w}^k - w)^T F(w) + \frac{1}{2}(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) \le 0, \ \forall w \in \mathcal{W}.$$

Summing the above inequality over $k = 0, \ldots, t$, we obtain

$$\sum_{k=0}^{t}\big(\theta(\tilde{u}^k) - \theta(u)\big) + \Big(\sum_{k=0}^{t} \tilde{w}^k - \sum_{k=0}^{t} w\Big)^T F(w) + \frac{1}{2}(\|v - v^{t+1}\|_H^2 - \|v - v^0\|_H^2) \le 0.$$

Dropping the term $\|v^{t+1} - v^0\|_G^2$, we get

$$\Big(\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k) - \theta(u)\Big) + \Big(\frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k - w\Big)^T F(w) \le \frac{\|v - v^0\|_H^2}{2(t+1)}, \quad \forall w \in \mathcal{W}. \tag{4.15}$$

By incorporating the notation of $\tilde{w}_t$ and using

$$\theta(\tilde{u}_t^k) \le \frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k) \qquad \text{(due to the convexity of } \theta(u))$$

it follows from (4.15) that

$$\big(\theta(\tilde{u}_t) - \theta(u)\big) + (\tilde{w}_t - w)^T F(w) \le \frac{\|v - v^0\|_H^2}{2(t+1)}, \quad \forall w \in \mathcal{W}.$$

Hence, the proof is complete. $\square$

For given substantial compact set $\mathcal{D} \subset \mathcal{W}$, we define

$$d = \sup\{\|v - v^0\|_H \mid w \in \mathcal{D}\},$$

where $v^0 = (y^0, z^0, \lambda^0)$ is the initial point. Because $\varrho_k \ge \frac{1}{2}$, it follows that $\Upsilon_t \ge \frac{t+1}{2}$.

After t iterations of the proposed method, we can find a $\tilde{w} \in \mathcal{W}$ such that

$$\sup_{w\in\mathcal{D}} \big\{\theta(\tilde{u}) - \theta(u)\big) + (\tilde{w} - w)^T F(w)\big\} \leq \frac{d^2}{2(t+1)}.$$

The convergence rate $\mathcal{O}(1/t)$ of the proposed method is thus proved.

5 Convergence rate in the non-ergodic sense

If we use (3.3) with $\alpha \in (0,1)$ to update the new iterate, it follows from (3.15) that

$$\sum_{k=0}^{\infty} \|v^k - \bar{v}^k\|_D^2 \leq \frac{1}{\alpha(1-\alpha)} \|v^0 - v^*\|_H^2 \quad \forall\, v^* \in \mathcal{V}^*. \tag{5.1}$$

This section will show that the sequence $\{\|v^k - v^{k+1}\|_D^2\}$ is monotonically non-increasing, *i. e.*,

$$\|v^{k+1} - \bar{v}^{k+1}\|_D^2 \leq \|v^k - \bar{v}^k\|_D^2, \quad \forall\, k \geq 0. \tag{5.2}$$

Based on (5.1) and (5.2), we drive

$$\|v^k - \bar{v}^k\|_D^2 \le \frac{1}{(k+1)\alpha(1-\alpha)} \|v^0 - v^*\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.3}$$

Since $\|v^k - \bar{v}^k\|_D^2$ is viewed as the stopping criterion, we obtain the worst-case $O(1/t)$ convergence rate in a non-ergodic sense. An important relation in the coming proof is (see (3.3))

$$P^T(v^k - v^{k+1}) = \alpha D(v^k - \bar{v}^k), \tag{5.4}$$

where the matrices P and D are given in (2.7). Lemma 4.2 enables us to establish an important inequality in the following lemma.

Lemma 5.1 *Let* $\{v^k\}$ *be the sequence generated by* (3.3)*, the associated sequence* $\{\tilde{w}^k\}$ *be defined by (4.1). Then, we have*

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \ge 0, \tag{5.5}$$

where Q *is given defined in (4.4).*

Proof. Set $w = \tilde{w}^{k+1}$ in (4.6) and use (4.3), we obtain

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \ge (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{5.6}$$

Note that (4.6) is also true for $k := k+1$ and thus we have

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (v - \tilde{v}^{k+1})^T P(v^{k+1} - \bar{v}^{k+1}),$$

for all $w \in \Omega$. Set $w = \tilde{w}^k$ in the above inequality and use (4.3), we get

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\bar{w}^{k+1}) \geq (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}).$$

Adding this inequality with (5.6), we get assertion (5.5) due to the monotonicity of F. □

Lemma 5.2 *Let the sequence* $\{\bar{w}^k\}$ *be generated by* (3.1), *and* $\{v^k\}$ *be the sequence updated by* (3.3). *Then, we have*

$$\begin{aligned} &\alpha(v^k - \bar{v}^k)^T D\{(v^k - \bar{v}^k) - (v^{k+1} - \bar{v}^{k+1})\} \\ &\geq \quad \frac{1}{2}\|P(v^k - \bar{v}^k) - P(v^{k+1} - \bar{v}^{k+1})\|^2_{(Q^{-T}+Q^{-1})}, \end{aligned} \tag{5.7}$$

where the matrices P *and* D *are given in (2.7) and* Q *is given defined in (4.4).*

Proof. Adding the term

$$\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}$$

to the both sides of (5.5), we get

$$(v^k - v^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \tag{5.8}$$

By using (5.4) and (4.3), we have

$$(v^k - v^{k+1})^T = \alpha(v^k - \bar{v}^k)^T D P^{-1},$$

and

$$\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} = Q^{-1} P\{(v^k - \bar{v}^k) - (v^{k+1} - \bar{v}^{k+1})\}.$$

Substituting them in (5.8), we obtain

$$\begin{aligned} &\alpha(v^k - \bar{v}^k)^T D\{(v^k - \bar{v}^k) - (v^{k+1} - \bar{v}^{k+1})\} \\ \geq \quad &\frac{1}{2}\|Q^{-1}P(v^k - \bar{v}^k) - Q^{-1}P(v^{k+1} - \bar{v}^{k+1})\|^2_{(Q^T+Q)}. \end{aligned} \tag{5.9}$$

Because

$$Q^{-T}(Q^T + Q)Q^{-1} = Q^{-T} + Q^{-1},$$

the both right hand sides of (5.7) and (5.9) are equal. The assertion is proved. □

Finally, we are ready to show the assertion (5.2) in the following theorem.

Theorem 5.1 *Let* $\{\bar{v}^k\}$ (3.1), $\{v^k\}$ *be the sequence generated by* (3.3). *The sequence* $\{\|v^k - v^{k+1}\|_D^2\}$ *is monotonically non-increasing.*

Proof. Setting $a = (v^k - \bar{v}^k)$ and $b = (v^{k+1} - \bar{v}^{k+1})$ in the identity

$$\|a\|_D^2 - \|b\|_D^2 = 2a^T D(a-b) - \|a-b\|_D^2,$$

we obtain

$$\begin{aligned}
&\|v^k - \bar{v}^k\|_D^2 - \|v^{k+1} - \bar{v}^{k+1}\|_D^2 \\
&\quad = 2(v^k - \bar{v}^k)^T D\{(v^k - \bar{v}^k) - (v^{k+1} - \bar{v}^{k+1})\} - \|(v^k - \bar{v}^k) - (v^{k+1} - \bar{v}^{k+1})\|_D^2.
\end{aligned}$$

Inserting (5.7) into the first term of the right-hand side of the last equality, we obtain

$$\begin{aligned}
&\|v^k - \bar{v}^k\|_D^2 - \|v^{k+1} - \bar{v}^{k+1}\|_D^2 \\
&\quad \geq \quad \frac{1}{\alpha}\|P(v^k - \bar{v}^k) - P(v^{k+1} - \bar{v}^{k+1})\|_{(Q^{-T}+Q^{-1})}^2 \\
&\qquad\quad -\|P(v^k - \bar{v}^k) - P(v^{k+1} - \bar{v}^{k+1})\|_{(P^{-T}DP^{-1})}^2 \\
&\quad \geq \quad \|P(v^k - \bar{v}^k) - P(v^{k+1} - \bar{v}^{k+1})\|_{\{(Q^{-T}+Q^{-1})-(P^{-T}DP^{-1})\}}^2.
\end{aligned}$$

Notice that (see (4.4) and (2.7))

$$(Q^{-T} + Q^{-1}) - (P^{-T} D P^{-1}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{\beta} I_m & I_m \\ 0 & I_m & \beta I_m \end{pmatrix}$$

is a positive semidefinite matrix, the assertion (5.2) follows immediately and the lemma is proved. $\square$

With (5.1) and (5.2), we derived (5.3). The worst-case $O(1/t)$ convergence rate in a non-ergodic sense for the proposed method with $\alpha \in (0, 1)$ is proved.

By using (5.4), we can also use $\|v^k - v^{k+1}\|_H \le \epsilon$ as the stop criterion. Since

$$H = P D^{-1} P^T,$$

we have

$$\|v^k - \bar{v}^k\|_D^2 = \frac{1}{\alpha^2} \|D^{-1} P^T (v^k - v^{k+1})\|_D^2 = \frac{1}{\alpha^2} \|v^k - v^{k+1}\|_H^2.$$

Therefore, the assertions (5.1) and (5.3) can be rewritten as

$$\sum_{k=0}^{\infty} \|v^k - v^{k+1}\|_H^2 \le \frac{\alpha}{(1-\alpha)} \|v^0 - v^*\|_H^2 \quad \forall\, v^* \in \mathcal{V}^*, \tag{5.10}$$

and

$$\|v^k - v^{k+1}\|_H^2 \le \frac{\alpha}{(k+1)(1-\alpha)} \|v^0 - v^*\|_H^2, \quad \forall v^* \in \mathcal{V}^*, \tag{5.11}$$

respectively.

6 Conclusions

Because of the attractive efficiency of the well-known alternating direction method (ADM), it is of strong desire to extend the ADMM to the linearly constrained convex programming problem with three separable operators. The convergence of the direct extension of the ADMM to the problem with 3 separable parts, however, is still open. The method proposed in this lecture (with update form (3.2)) is convergent and its variety to the direct extension of ADMM is tiny. We proved its $O(1/t)$ convergence rate in an ergodic sense. The

$O(1/t)$ non-ergodic convergence rate is also proved for the method using update form (3.3) with $\alpha \in (0, 1)$.

Appendix. Why the direct extension of ADMM performs well in practice ?

In this appendix, we try to explain why the direct extension of ADMM performs well in practice. If we use the direct extension of ADMM, then $w^{k+1} = \bar{w}^k$ and thus the relation (3.11) can be written as

$$\begin{aligned} & (v^{k+1} - v^*)^T P(v^k - v^{k+1}) \\ \geq \quad & \big(\theta(u^{k+1}) - \theta(u^*)\big) + (w^{k+1} - w^*)^T F(w^{k+1}) \\ & + (\lambda^k - \lambda^{k+1})^T \{(y^k - y^{k+1}) + (z^k - z^{k+1})\}. \end{aligned} \tag{A.1}$$

Note that M is not symmetric, but $M^T + M$ is positive definite. Again, using $\bar{w}^k = w^{k+1}$, the third part of (3.7) is

$$z^{k+1} \in \mathcal{Z}, \;\; \theta_3(z) - \theta_3(z^{k+1}) + (z - z^{k+1})^T(-\lambda^{k+1}) \geq 0, \;\; \forall\, z \in \mathcal{Z}.$$

It holds also for the previous iteration, thus we have

$$z^k \in \mathcal{Z}, \;\; \theta_3(z) - \theta_3(z^k) + (z - z^k)^T(-\lambda^k) \geq 0, \;\; \forall\, z \in \mathcal{Z}.$$

Setting $z = z^k$ and $z = z^{k+1}$ in the above two sub-VIs, respectively, and then add the two resulting inequalities, we obtain

$$(\lambda^k - \lambda^{k+1})^T (z^k - z^{k+1}) \geq 0. \tag{A.2}$$

Since $(w^{k+1} - w^*)^T F(w^{k+1}) = (w^{k+1} - w^*)^T F(w^*)$ and

$$P = D + \begin{pmatrix} 0 & 0 & 0 \\ \beta I_m & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \text{(see (2.7))}$$

it follows from (A.1) and (A.2) that

$$(v^{k+1} - v^*)^T D(v^k - v^{k+1}) \geq (\lambda^k - \lambda^{k+1})^T (y^k - y^{k+1}) + \Delta_k \tag{A.3}$$

where

$$\Delta_k = \big(\theta(u^{k+1}) - \theta(u^*)\big) + (w^{k+1} - w^*)^T F(w^*) + (z^{k+1} - z^*)^T \beta(y^{k+1} - y^k). \tag{A.4}$$

Therefore, by using (A.3), we obtain

$$
\begin{aligned}
&\|v^k - v^*\|_D^2 - \|v^{k+1} - v^*\|_D^2 \\
&\quad = \quad \|(v^{k+1} - v^*) + (v^k - v^{k+1})\|_D^2 - \|v^{k+1} - v^*\|_D^2 \\
&\quad = \quad 2(v^{k+1} - v^*)^T D(v^k - v^{k+1}) + \|v^k - v^{k+1}\|_D^2 \\
&\quad \geq \quad \|v^k - v^{k+1}\|_D^2 + 2(\lambda^k - \lambda^{k+1})^T (y^k - y^{k+1}) + 2\Delta_k. \qquad \text{(A.5)}
\end{aligned}
$$

Because

$$
D = \begin{pmatrix} \beta I_m & 0 & 0 \\ 0 & \beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \quad \text{and} \quad v = \begin{pmatrix} y \\ z \\ \lambda \end{pmatrix}
$$

it follows that

$$
\begin{aligned}
&\|v^k - v^{k+1}\|_D^2 + 2(\lambda^k - \lambda^{k+1})^T (y^k - y^{k+1}) \\
&\quad = \quad \beta\|z^k - z^{k+1}\|^2 + \beta\|(y^k - y^{k+1}) + \frac{1}{\beta}(\lambda^k - \lambda^{k+1})\|^2. \qquad \text{(A.6)}
\end{aligned}
$$

Substituting (A.6) and (A.3) in the right hand side of (A.5), we obtain

$$
\begin{aligned}
&\|v^k - v^*\|_D^2 - \|v^{k+1} - v^*\|_D^2 \\
&\quad = \quad \beta\|z^k - z^{k+1}\|^2 + \beta\|(y^k - y^{k+1}) + \frac{1}{\beta}(\lambda^k - \lambda^{k+1})\|^2 \\
&\qquad\quad +2\{\big(\theta(u^{k+1}) - \theta(u^*)\big) + (w^{k+1} - w^*)^T F(w^*)\} \\
&\qquad\quad +2(z^{k+1} - z^*)^T \beta(y^{k+1} - y^k).
\end{aligned} \tag{A.7}
$$

In the right hand side of (A.7), the terms

$$\beta\|z^k - z^{k+1}\|^2 + \beta\|(y^k - y^{k+1}) + \frac{1}{\beta}(\lambda^k - \lambda^{k+1})\|^2,$$

and

$$2(\theta(u^{k+1}) - \theta(u^*)\big) + 2(w^{k+1} - w^*)^T F(w^*)$$

are non-negative. However, we do not know whether the last term of the right hand side of (A.7), *i.e.*

$$2(z^{k+1} - z^*)^T \beta(y^{k+1} - y^k)$$

is non-negative. It is pity that we can not show that the right hand side of (A.7) is positive. It seems that the direct extension of ADMM performs well because the right hand side of

(A.7) is positive in practice.

If the right hand side of (A.7) is positive, then the sequence $\{\|v^k - v^*\|_D\}$ is Fej$\grave{e}$r monotone and has the contractive property.

References

[1] 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.

[2] J. Douglas and H. H. Rachford, On the numerical solution of the heat conduction problem in 2 and 3 space variables, *Transactions of the American Mathematical Society* **82** (1956), 421–439.

[3] X.J. Cai, G.Y. Gu and B. S. He *On the $O(1/t)$ convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators*, Comput. Optim. Appl., 57(2014), 339-363.

[4] E. Esser, M. Möller, S. Osher G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, 2011, arXiv: 1102.0844v1 [stat.ML] 4 Feb 2011.

[5] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity problems, Volume I*, Springer Series in Operations Research, Springer-Verlag, 2003.

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

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

[8] B. S. He, Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities, *Computational Optimization and Applications* **42**(2009), 195–212.

[9] 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.

[10] B. S. He, M. Tao and X.M. Yuan, Alternating direction method with Gaussian back substitution for separable convex programming, *SIAM Journal on Optimization* **22**(2012), 313-340.

[11] 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.

[12] B. S. He and H. Yang, Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities, *Operations Research Letters* **23**(1998), 151–161.

[13] B. S. He and X. M. Yuan, On the $O(1/t)$ convergence rate of the alternating direction method, *SIAM J. Numerical Analysis* **50**(2012), 700-709.

[14] Z. C. Lin, M. M. Chen, L. Q. Wu, and Y. Ma, The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices, *manuscript*, 2009.

[15] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.* **15** (2004), 229–251.

[16] P. Tseng, On accelerated proximal gradient methods for convex-concave optimization, manuscript, 2008.