凸优化和单调变分不等式的收缩算法

第十四讲: 自变量 x-y 平等的对称型交替方向法

Symmetric ADMM

南京大学数学系　　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the manuscript [4]

1 Introduction

We consider the following structured constrained convex optimization problem

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

where $\theta_1(x):\Re^{n_1}\rightarrow\Re,\;\theta_2(y):\Re^{n_2}\rightarrow\Re$ are convex functions (but not necessary smooth), $A\in\Re^{m\times n_1}$, $B\in\Re^{m\times n_2}$ and $b\in\Re^m$, $\mathcal{X}\subset\Re^{n_1}$, $\mathcal{Y}\subset\Re^{n_2}$ are given closed convex sets.

First, as the work [18, 19] for analyzing the convergence of ADMM, we need a variational inequality (VI) reformulation of the model (1.1) and a characterization of its solution set. More specifically, solving (1.1) is equivalent to finding $w^*=(x^*,y^*,\lambda^*)\in\Omega$ such that

$$\text{VI}(\Omega,F,\theta):\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \quad (1.2a)$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad \theta(u) = \theta_1(x) + \theta_2(y), \qquad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix} \tag{1.2b}$$

$$F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m \tag{1.2c}$$

Since the mapping $F(w)$ defined in (1.2c) is affine with a skew-symmetric matrix, it is thus monotone. We denote by Ω^* the solution set of VI(Ω, F, θ), and consider the problem under the nonempty assumption onto Ω^*.

The augmented Lagrangian function of the problem (1.1) is defined by

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

Applied ADMM to the problem (1.1)

From given $v^k = (y^k, \lambda^k)$, the iteration produces $w^k = (x^{k+1}, y^{k+1}, \lambda^{k+1})$.

The iteration scheme:

1. First, for given (y^k, λ^k), x^{k+1} is the solution of the following problem

$$x^{k+1} \in \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) | x \in \mathcal{X}\}. \tag{1.3a}$$

2. Use λ^k and the obtained x^{k+1}, find y^{k+1} by

$$y^{k+1} \in \text{Argmin}\left\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \middle| y \in \mathcal{Y}\right\}. \tag{1.3b}$$

3. Set the multipliers by

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{1.3c}$$

The sequence $\{v^k\}$ generated by (1.4) satisfies

$$\|v^{k+1} - v^*\|_{H_c}^2 \le \|v^k - v^*\|_{H_c}^2 - \|v^k - v^{k+1}\|_{H_c}^2, \;\; H_c = \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}.$$

Applied ADMM-based customized PPA to the problem (1.1)

From given $v^k = (y^k, \lambda^k)$, the prediction step produces $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$.

The prediction step:

1. First, for given (y^k, λ^k), $\tilde{x}^k$ is the solution of the following problem

$$\tilde{x}^k = \in \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) | x \in \mathcal{X}\}. \tag{1.4a}$$

2. Set the multipliers by

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \tag{1.4b}$$

3. Finally, use the obtained $\tilde{x}^k$ and $\tilde{\lambda}^k$, find $\tilde{y}^k$ by

$$\tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta(\tilde{x}^k, y, \tilde{\lambda}^k) | y \in \mathcal{Y}\}. \tag{1.4c}$$

The new iterate v^{k+1} is given by

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \in (0, 2). \tag{1.5}$$

With given v^k, the predictor $\tilde{w}^k$ is generated by an "order changed ADMM". As analyzed in the last section, for the predictor $\tilde{w}^k \in \Omega$, we have

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (v-\tilde{v}^k)^T H_c(v^k-\tilde{v}^k) \ge 0, \ \forall\, w \in \Omega, \quad (1.6)$$

where

$$H_c = \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \quad (1.7)$$

Set $w = w^*$ in (1.6), we get

$$(\tilde{v}^k - v^*)^T H_c(v^k - \tilde{v}^k) \ge 0$$

and thus

$$(v^k - v^*)^T H_c(v^k - \tilde{v}^k) \ge \|v^k - \tilde{v}^k\|_{H_c}^2$$

Using the above inequality, it is easy to show that generated sequence $\{v^k\}$ by (1.4)-(1.5) satisfies

$$\|v^{k+1} - v^*\|_{H_c}^2 \le \|v^k - v^*\|_{H_c}^2 - \alpha(2-\alpha)\|v^k - \tilde{v}^k\|_{H_c}^2.$$

Symmetric ADMM for the problem (1.1)

In practice, the primal variables x and y should be treated fairly. In this paper, we propose the following iterative scheme for the problem (1.1). Again, $v = (y, \lambda)$ are the essential variables and x is the intermediate variable. The k-th iteration starts with $v^k = (y^k, \lambda^k)$, produces $v^{k+1} = (y^{k+1}, \lambda^{k+1})$ via

$$\left\{\begin{array}{l} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(Ax^{k+1} + By^{k+1} - b), \end{array}\right. \tag{1.8}$$

where $\alpha \in (0, 1)$. If $\alpha = 1$, the method is the Peaceman-Rachford splitting method (PRSM) in [20, 21] to the problem (1.1).

We will prove the sequence $\{v^k\}$ is strictly contractiveto the solution set and establish a worst-case $O(1/t)$ convergence rate for the PRSM scheme (2.1).

2 Preliminaries

In this section, we summarize some useful preliminaries known in the literature and prove some simple conclusions for further analysis.

Let $\alpha = 1$, the iterative scheme of SPRSM (1.8) for (1.1) becomes

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b), \end{cases} \tag{2.1}$$

where $\lambda \in \Re^m$ is the Lagrange multiplier associated with the linear constraints in (1.1) and $\beta > 0$ is a penalty parameter. As analyzed in [9], the PRSM scheme (2.1) differs from ADMM "only through the addition of the intermediate update of the multipliers ($\lambda^{k+\frac{1}{2}}$); it thus offers the same set of advantages". The PRSM scheme (2.1), however, according to [9] again (see also [14]), "is less 'robust' in that it converges under more restrictive assumptions than ADMM" and "if it does converge, then its rate of convergence is faster". We refer to [1, 12] for some numerical verification of the efficiency of PRSM.

According to Theorem 2.3.5 in [7], a very useful characterization of the solution set Ω^* of VI(Ω, F, θ) can be summarized in the following theorem. Its proof can be found in [7, 18].

Theorem 2.1 *The solution set of VI(Ω, F, θ) is closed and convex; and it can be characterized as*

$$\Omega^* = \bigcap_{w \in \Omega} \big\{\tilde{w} \in \Omega : \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \geq 0\big\}. \tag{2.2}$$

Theorem 2.1 thus implies that $\tilde{w} \in \Omega$ is an approximate solution of VI(Ω, F, θ) with an accuracy of $O(1/t)$ if it satisfies

$$\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w) \leq \epsilon, \ \ \forall w \in \Omega, \tag{2.3}$$

with $\epsilon = O(1/t)$. In fact, this characterization makes it possible to analyze the convergence rate of ADMM and other splitting methods via the VI approach rather than the conventional approach based on the functional values in the literature. In the following, we shall show that either the sequence (2.1) or (1.8) provides us such an iterate satisfying (2.3) after t iterations.

As mentioned in [2] for ADMM, the variable x is an intermediate variable during the PRSM iteration since it essentially requires only (y^k, λ^k) in (2.1) or (1.8) to generate the $(k+1)$-th iterate. For this reason, we define the notations $v^k = (y^k, \lambda^k)$, $\mathcal{V} = \mathcal{Y} \times \Re^m$,

$$\mathcal{V}^* := \{v^* = (y^*, \lambda^*) \,|\, w^* = (x^*, y^*, \lambda^*) \in \Omega^*\};$$

and it suffices to analyze the convergence rate of the sequence $\{v^k\}$ to the set $\mathcal{V}^*$ in order to study the convergence rate of the sequence $\{w^k\}$ generated by (2.1) or (1.8). Note $\mathcal{V}^*$ is also closed and convex.

Finally, we define some matrices in order to present our analysis in a compact way. Let

$$M = \begin{pmatrix} I_{n_2} & 0 \\ -\alpha\beta B & 2\alpha I_m \end{pmatrix}. \tag{2.4}$$

and

$$\widetilde{Q} = \begin{pmatrix} 0 & 0 \\ \beta B^T B & -\alpha B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{2.5}$$

In (2.5), "0" is a matrix with all zero entries in appropriate dimensionality. We further define

$$Q = \begin{pmatrix} \beta B^T B & -\alpha B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{2.6}$$

as the submatrix of $\widetilde{Q}$ excluding all the first zero rows. The matrices $\widetilde{Q}$ and Q are associated with the analysis for the sequences $\{w^k\}$ and $\{v^k\}$, respectively.

Last, for $\alpha \in (0,1]$ we define a symmetric matrix

$$H = \frac{1}{2}\begin{pmatrix} (2-\alpha)\beta B^T B & -B^T \\ -B & \frac{1}{\alpha\beta} I_m \end{pmatrix}. \tag{2.7}$$

Lemma 2.1 *The matrix H defined in (2.7) is positive definite for $\alpha \in (0,1)$ and positive semi-definite for $\alpha = 1$.*

Proof We have

$$H = \frac{1}{2}\begin{pmatrix} \sqrt{\beta} B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix}\begin{pmatrix} (2-\alpha) I & -I \\ -I & \frac{1}{\alpha} I \end{pmatrix}\begin{pmatrix} \sqrt{\beta} B & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix}.$$

Note that the matrix

$$\begin{pmatrix} (2-\alpha) & -1 \\ -1 & \frac{1}{\alpha} \end{pmatrix}$$

is positive definite if $\alpha \in (0,1)$ and positive semi-definite if $\alpha = 1$. Thus assertion is thus proved. □.

Lemma 2.2 *The matrices* M, Q *and* H *defined respectively in* (2.4), (2.6) *and* (2.7)) *have the following relationships:*

$$HM = Q \tag{2.8}$$

and

$$G = Q^T + Q - M^T HM \succeq \frac{(1-\alpha)}{2(1+\alpha)} M^T HM. \tag{2.9}$$

Proof. Using the definition of the matrices M, Q and H, by a simple manipulation, we obtain

$$\begin{aligned} HM &= \frac{1}{2}\begin{pmatrix} (2-\alpha)\beta B^T B & -B^T \\ -B & \frac{1}{\alpha\beta} I_m \end{pmatrix} \begin{pmatrix} I & 0 \\ -\alpha\beta B & 2\alpha I_m \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} 2\beta B^T B & -2\alpha B^T \\ -2B & \frac{2}{\beta} I_m \end{pmatrix} = Q. \end{aligned}$$

The first assertion is proved. Consequently, we get

$$\begin{aligned} M^T H M &= M^T Q = \begin{pmatrix} I_{n_2} & -\alpha\beta B^T \\ 0 & 2\alpha I_m \end{pmatrix} \begin{pmatrix} \beta B^T B & -\alpha B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \\ &= \begin{pmatrix} (1+\alpha)\beta B^T B & -2\alpha B^T \\ -2\alpha B & \frac{2\alpha}{\beta} I_m \end{pmatrix}. \end{aligned} \quad (2.10)$$

Using (2.6) and the above equation, we have

$$G = (Q^T + Q) - M^T H M = (1-\alpha) \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix}. \quad (2.11)$$

Since $\begin{pmatrix} \beta & -1 \\ -1 & \frac{2}{\beta} \end{pmatrix}$ is positive definite, we say G is essentially positive definite .

Note that

$$2(1+\alpha)\begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix} - M^T H M = \begin{pmatrix} (1+\alpha)\beta B^T B & -2B^T \\ -2B & \frac{4+2\alpha}{\beta} I_m \end{pmatrix}. \tag{2.12}$$

Because

$$\begin{pmatrix} (1+\alpha) & -2 \\ -2 & 4+2\alpha \end{pmatrix} \succeq 0, \quad \forall\, \alpha \geq 0,$$

the right-hand side of (2.12) is positive semidefinite. Thus, it follows that

$$\begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix} \succeq \frac{1}{2(1+\alpha)} M^T H M. \tag{2.13}$$

Substituting (2.13) in (2.11), we obtain (2.9) and the lemma is proved. □

Remark 2.1 *When $\alpha = 1$, the matrices H defined in* (2.7) *and $Q^T + Q - M^T H M$ are both positive semi-definite. But, in the following analysis we still use $\|v - \tilde{v}\|_H$ and $\|v - \tilde{v}\|_{Q^T+Q-M^T HM}$ to denote* *respectively*

$$\|v-\tilde{v}\|_H = \big((v-\tilde{v})^T H (v-\tilde{v})\big)^{1/2}$$

and

$$\|v-\tilde{v}\|_{Q^T+Q-M^THM} = \big((v-\tilde{v})^T (Q^T+Q-M^THM)(v-\tilde{v})\big)^{1/2},$$

for v, $\tilde{v} \in \mathcal{Y} \times \Re^m$*. This slight abuse of notation will alleviate the notation in our analysis greatly.*

We also have the following lemma.

Lemma 2.3 *Let* $\{w^k\}$ *be the sequence generated by the strictly contractive PRSM (1.8), and* H *be defined in (2.7). Then,* w^{k+1} *is a solution of VI*(Ω, F, θ) *if* $\|v^k - v^{k+1}\|_H^2 = 0$*.*

Proof. See Lemma 3.2 in [19]. □

3 Contraction Analysis

In this section, we analyze the contraction property for the sequence $\{v^k\}$ generated by the PRSM scheme (2.1) or the strictly contractive PRSM scheme (1.8) with respect to the set $\mathcal{V}^*$. The convergence rate analysis for (2.1) and (1.8) to be presented is based on this analysis of contraction property. Since (2.1) can be included by the strictly contractive PRSM scheme (1.8) if we extend the value of $\alpha = 1$ and the algebra of convergence analysis for these two schemes are of the same framework, below we only present the contraction analysis for (1.8) and the analysis for (2.1) is readily obtained by taking $\alpha = 1$ in our analysis.

First, to further alleviate the notation in our analysis, we need to define an auxiliary sequence $\{\tilde{w}^k\}$ as

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}, \qquad (3.1)$$

where (x^{k+1}, y^{k+1}) is generated by (2.1) or (1.8). Note with the notation of $\tilde{w}^k$, the strictly contractive PRSM (1.8) can be written as

$$\left\{ \begin{array}{l} \tilde{x}^k = \arg\min\Big\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + By^k - b\|^2 \;\Big|\; x \in \mathcal{X}\Big\}, \\[2ex] \tilde{y}^k = \arg\min\left\{ \begin{array}{c} \theta_2(y) - [\lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k)]^T By \\ + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \end{array} \;\middle|\; y \in \mathcal{Y} \right\}. \end{array} \right. \tag{3.2}$$

Then, based on (1.8) and (3.1), we immediately have

$$x^{k+1} = \tilde{x}^k, \qquad y^{k+1} = \tilde{y}^k \qquad \text{and} \qquad \lambda^{k+\frac{1}{2}} = \tilde{\lambda}^k$$

and

$$\begin{aligned} \lambda^{k+1} &= \lambda^{k+1/2} - \alpha\beta(A\tilde{x}^k + B\tilde{y}^k - b) \\ &= \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\big[\beta(A\tilde{x}^k + By^k - b) - \beta B(y^k - \tilde{y}^k)\big] \\ &= \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\big[(\lambda^k - \tilde{\lambda}^k) - \beta B(y^k - \tilde{y}^k)\big] \\ &= \lambda^k - \big[2\alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\beta B(y^k - \tilde{y}^k)\big]. \end{aligned} \tag{3.3}$$

Furthermore, together with $y^{k+1} = \tilde{y}^k$, we have the following relationship

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_{n_2} & 0 \\ -\alpha\beta B & (1+\alpha)I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix};$$

which can be rewritten into a compact form by using the notations of v^k and $\tilde{v}^k$:

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{3.4}$$

where M is defined in (2.4).

Now, we start to prove some properties for the sequence $\{\tilde{w}^k\}$ defined in (3.1). Recall our primary purpose is to analyze the convergence rate for the sequences (2.1) and (1.8) based on the solution characterization (2.2), and the accuracy of an approximate solution $\tilde{w} \in \Omega$ is measured by a upper bound of the quantity of $\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)$ for all $w \in \Omega$ (see (2.3)). Hence, we are interested in estimating how accurate the point $\tilde{w}^k$ defined in (3.1) is to a solution point of VI(Ω, F, θ). The main result is proved in Theorem 3.1. But before that, we first show two lemmas. The first lemma presents a upper bound of

$\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)$ for all $w \in \Omega$ in term of a quadratic term involving the matrix Q.

Lemma 3.1 *For given $v^k \in \mathcal{Y} \times \Re^m$, let w^{k+1} be generated by the strictly contractive PRSM scheme (1.8) and $\tilde{w}^k$ be defined in* (3.1)*. Then, we have $\tilde{w}^k \in \Omega$ and*

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall\, w \in \Omega, \tag{3.5}$$

where the matrix Q is defined in (2.6)*.*

Proof. Since $x^{k+1} = \tilde{x}^k$, by deriving the first-order optimality condition of the x-minimization problem in (3.2), we have

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{A^T[\beta(A\tilde{x}^k + By^k - b) - \lambda^k]\} \geq 0, \ \forall\, x \in \mathcal{X}. \tag{3.6}$$

According to the definition (3.1), we have

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \tag{3.7}$$

Using (3.7), the inequality (3.6) can be written as

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\tilde{\lambda}^k\} \ge 0, \quad \forall\, x \in \mathcal{X}. \tag{3.8}$$

Similarly, by deriving the first-order optimality condition of the y-minimization problem in (3.2), we get

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T B^T \left\{ \begin{array}{c} \beta(A\tilde{x}^k + B\tilde{y}^k - b) \\ -[\lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k)] \end{array} \right\} \ge 0, \ \forall\, y \in \mathcal{Y}. \tag{3.9}$$

Again, using (3.7), we have

$$\begin{aligned} &\beta(A\tilde{x}^k + B\tilde{y}^k - b) - [\lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k)] \\ &\quad = \quad -(\tilde{\lambda}^k - \lambda^k) + \beta B(\tilde{y}^k - y^k) - [\lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k)] \\ &\quad = \quad -\tilde{\lambda}^k + \beta B(\tilde{y}^k - y^k) - \alpha(\tilde{\lambda}^k - \lambda^k). \end{aligned}$$

Consequently, it follows from (3.9) that

$$\theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T\left\{\begin{array}{c}-B^T\tilde{\lambda}^k+\beta B^TB(\tilde{y}^k-y^k)\\ -\alpha B^T(\tilde{\lambda}^k-\lambda^k)\end{array}\right\}\ge 0,\ \forall\, y\in\mathcal{Y}. \tag{3.10}$$

In addition, based on (3.1) we have

$$(A\tilde{x}^k+B\tilde{y}^k-b)-B(\tilde{y}^k-y^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k)=0. \tag{3.11}$$

Combining (3.8), (3.10) and (3.11) together, we get $\tilde{w}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)\in\Omega$; and for any $w=(x,y,\lambda)\in\Omega$, it holds

$$\theta(u)-\theta(\tilde{u}^k)+\left(\begin{array}{c}x-\tilde{x}^k\\ y-\tilde{y}^k\\ \lambda-\tilde{\lambda}^k\end{array}\right)^T\left(\begin{array}{c}-A^T\tilde{\lambda}^k\\ -B\tilde{\lambda}^k\\ A\tilde{x}^k+B\tilde{y}^k-b\end{array}\right)$$

$$+\left(\begin{array}{c}x-\tilde{x}^k\\ y-\tilde{y}^k\\ \lambda-\tilde{\lambda}^k\end{array}\right)^T\left(\begin{array}{c}0\\ \beta B^TB(\tilde{y}^k-y^k)-\alpha B^T(\tilde{\lambda}^k-\lambda^k)\\ -B(\tilde{y}^k-y^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k)\end{array}\right)\ge 0.$$

The assertion (3.5) is only a compact form of the above inequality by using the notations of Q in (2.6), w and F in (1.2b) and v. The proof is complete. $\square$

The second lemma aims at expressing the bound $(v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k)$ found in Lemma 3.1 by the difference of two quadratic terms involving two consecutive iterates of the sequence $\{v^k\}$ and a quadratic term involving v^k and the auxiliary iterate $\tilde{v}^k$. This equivalent expression is convenient for the manipulation over the whole sequence $\{v^k\}$ recursively and thus for establishing the convergence rate of $\{v^k\}$ in either ergodic or nonergodic sense.

Lemma 3.2 *Let $\{w^k\}$ be generated by the strictly contractive PRSM scheme* (1.8) *and $\{\tilde{w}^k\}$ be defined in (3.1); M, Q and H be defined in (2.4), (2.6) and (2.7), respectively. Then we have*

$$(v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k) = \frac{1}{2}\big(\|v-v^{k+1}\|_H^2 - \|v-v^k\|_H^2\big) + \frac{(1-\alpha)}{4(1+\alpha)}\|M(v^k-\tilde{v}^k)\|_H^2, \;\forall v\in\mathcal{V}. \quad (3.12)$$

Proof. By using $Q=HM$ and $M(v^k-\tilde{v}^k)=(v^k-v^{k+1})$ (see (3.4)), it follows that

$$\begin{aligned}(v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k) &= (v-\tilde{v}^k)^T HM(v^k-\tilde{v}^k) \\ &= (v-\tilde{v}^k)^T H(v^k-v^{k+1}). \qquad (3.13)\end{aligned}$$

For the vectors a, b, c, d in the same space and a matrix H with appropriate dimensionality, we have the identity

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

In this identity, we take

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

and submit it to the right-hand side of (3.13). The resulting equation is

$$\begin{aligned}(v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k) &= \frac{1}{2}\big(\|v-v^k\|_H^2 - \|v-v^{k+1}\|_H^2\big) \\ &\quad + \frac{1}{2}(\|v^k-\tilde{v}^k\|_H^2 - \|v^{k+1}-\tilde{v}^k\|_H^2). \qquad (3.14)\end{aligned}$$

Now, we deal with the last term of the right-hand side of (3.14). By using (3.4) and (2.8), we get

$$\begin{array}{ll}
\multicolumn{2}{l}{\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2} \\
= & \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\
\overset{(3.4)}{=} & \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\
= & 2(v^k - \tilde{v}^k)HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)M^T HM(v^k - \tilde{v}^k) \\
\overset{(2.8)}{=} & (v^k - \tilde{v}^k)^T(Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\
\overset{(2.9)}{\geq} & \dfrac{(1-\alpha)}{2(1+\alpha)}(v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\
= & \dfrac{(1-\alpha)}{2(1+\alpha)}\|M(v^k - \tilde{v}^k)\|_H^2.
\end{array}$$

Substituting it in (3.14), we obtain the assertion (3.12). The proof is complete. □

Now we are ready to present an inequality where a upper bound of $\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}^k - w)^T F(w)$ is found for all $w \in \Omega$. This inequality is also crucial for analyzing the contraction property and the convergence rate for the iterative sequence generated by either (2.1) or (1.8).

Theorem 3.1 *For given $v^k \in \mathcal{Y} \times \Re^m$, let w^{k+1} be generated by the strictly contractive PRSM scheme (1.8) and $\tilde{w}^k$ be defined in* (3.1)*; M and H be defined in (2.4) and (2.7), respectively. Then, we have $\tilde{w}^k \in \Omega$ and*

$$\begin{aligned} &\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}^k - w)^T F(w) \\ \leq\ & \frac{1}{2}\big(\|v - v^k\|_H^2 - \|v - v^{k+1}\|_H^2\big) \\ & - \frac{1-\alpha}{4(1+\alpha)}\|M(v^k - \tilde{v}^k)\|_H^2, \ \forall w \in \Omega. \end{aligned} \tag{3.15}$$

Proof. First, by using the structure of $F(w)$ (see (1.2c)), we have

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) = 0, \quad \forall w, \ \tilde{w} \in \Omega.$$

Then, using the above equality and replacing the right-hand side term in (3.5) with the identity (3.12), we obtain the assertion (3.15). The proof is complete. □

The assertion (3.15) also enables us to study the contraction property of the sequence $\{v^k\}$ generated by (2.1) or (1.8). In fact, setting $w = w^*$ in (3.15) where w^* being an arbitrary solution point in Ω^*, we get

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2$$
$$\geq \quad \frac{1-\alpha}{2(1+\alpha)}\|M(v^k - \tilde{v}^k)\|_H^2 + 2\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*)\}.$$

which implies that

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \frac{1-\alpha}{2(1+\alpha)}\|M(v^k - \tilde{v}^k)\|_H^2. \quad (3.16)$$

Recall the identity (3.4). We thus have

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \frac{1-\alpha}{2(1+\alpha)}\|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \quad (3.17)$$

Therefore, when $\alpha = 1$, i.e., for the PRSM scheme (2.1), we have

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2, \quad (3.18)$$

which means the sequence $\{v^k\}$ generated by (2.1) is contractive, but not strictly, to the set $\mathcal{V}^*$. In fact, it is possible that the sequence $\{v^k\}$ stays away from the solution set with a constant distance, hence no convergence is guaranteed by (3.18). In [5], such an example is constructed. On the other hand, when $\alpha \in (0,1)$, the inequality (3.17) ensures a reduction of $\frac{1-\alpha}{2(1+\alpha)}\|v^k - v^{k+1}\|_H^2$ to the set $\mathcal{V}^*$ at the $(k+1)$-th iteration, i.e., the strict contraction of $\{v^k\}$ is guaranteed for the sequence generated by (1.8). We can thus expect that the strictly contractive PRSM (1.8) converges faster to $\mathcal{V}^*$ than the original PRSM (2.1). As we have mentioned, the difference of contraction between (3.17) and (3.18) is also the reason why we can establish a nonergodic convergence rate for the strictly contractive PRSM (1.8) while only an ergodic convergence rate can be established for the original PRSM (2.1).

4 Convergence rate of (2.1) in ergodic sense

In this section, we show that although the original PRSM (2.1) might not be convergent to a solution point of the model (1.1), it is still possible to find an approximate solution of VI(Ω, F, θ) with an accuracy of $O(1/t)$ based on the first t iterations of the PRSM scheme (2.1). This estimate helps us better understand the convergence property of the original PRSM (2.1).

Theorem 4.1 *Let* $\{w^k\}$ *be generated by* (2.1) *and* $\{\tilde{w}^k\}$ *be defined by (3.1). Let* $\tilde{w}_t$ *be defined as*

$$\tilde{w}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{w}^k. \qquad (4.1)$$

Then, for any integer number $t > 0$, $\tilde{w}_t \in \Omega$ *and*

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2(t+1)} \|v - v^0\|_H^2, \quad \forall w \in \Omega, \qquad (4.2)$$

where H is defined in (2.7).

Proof. First, because of (3.1), it holds that $\tilde{w}^k \in \Omega$ for all $k \geq 0$. Together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (4.1) implies that $\tilde{w}_t \in \Omega$. Second, by taking $\alpha = 1$ in (3.15) (and using $(w-\tilde{w})^T F(\tilde{w}) = (w-\tilde{w})^T F(w)$) we have

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(w)+\frac{1}{2}\|v-v^k\|_H^2 \geq \frac{1}{2}\|v-v^{k+1}\|_H^2, \ \forall w \in \Omega. \tag{4.3}$$

Summing the inequality (4.3) over $k = 0, 1, \ldots, t$, we obtain

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

for all $w \in \Omega$. Use the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k)-\theta(u)+(\tilde{w}_t-w)^T F(w) \leq \frac{1}{2(t+1)}\|v-v^0\|_H^2, \quad \forall w \in \Omega. \tag{4.4}$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1}\sum_{k=0}^{t}\tilde{u}^k,$$

we have that

$$\theta(\tilde{u}_t) \le \frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k).$$

Substituting it in (4.4), the assertion of this theorem follows directly. $\Box$

Let $v^0 = (y^0, \lambda^0)$ be the initial iterate. For a given compact set $\mathcal{D} \subset \mathcal{Y} \times \Re^m$, let $d = \sup\{\|v - v^0\|_H \,|\, v \in \mathcal{D}\}$. Then, after t iterations of the PRSM (2.1), the point $\tilde{w}_t \in \Omega$ defined in (4.1) satisfies

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

which means $\tilde{w}_t$ is an approximate solution of VI(Ω, F, θ) with an accuracy of $O(1/t)$ (recall (2.3)).

Remark 4.2 *In the proof of Theorem 4.1, we take* $\alpha = 1$ *in* (4.3)*. Obviously, the* *proof is still valid if we take $\alpha \in (0,1)$. Thus, a worst-case $O(1/t)$ convergence rate in ergodic sense can be established easily for the strictly contractive PRSM (1.8). As we shall show in Section 5, this is less interesting because a nonergodic worst-case $O(1/t)$ convergence rate can be established for (1.8). We thus omit the detail.*

5 Convergence rate of (1.8) in nonergodic sense

In this section, we show that the sequence $\{v^k\}$ generated by the strictly contractive PRSM scheme (1.8) is convergent to a point in $\mathcal{V}^*$, and its worst-case convergence rate is $O(1/t)$ in nonergodice sense. Our starting point for the analysis is the inequality (3.17), and a crucial property is the monotonicity of the sequence $\{\|v^k - v^{k+1}\|_H^2\}$. That is, we will prove that

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2, \quad \forall\, k \ge 0.$$

We first take a closer look at the assertion (3.5) in Lemma 3.1.

Lemma 5.1 *Let* $\{w^k\}$ *be the sequence generated by the strictly contractive PRSM* (1.8) *and* $\tilde{w}^k$ *be defined in (3.1); the matrix* Q *be defined in* (2.6). *Then, we have*

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \geq 0. \tag{5.1}$$

Proof. Set $w = \tilde{w}^{k+1}$ in (3.5), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{5.2}$$

Note that (3.5) is also true for $k := k+1$ and thus

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

for all $w \in \Omega$. Set $w = \tilde{w}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}). \tag{5.3}$$

Adding (5.2) and (5.3) and using the monotonicity of F, we get (5.1) immediately.
$\square$

Lemma 5.2 *Let* $\{w^k\}$ *be the sequence generated by the strictly contractive PRSM* (1.8) *and* $\tilde{w}^k$ *be defined in (3.1); the matrices* M*,* Q *and* H *be defined in* (2.4), (2.6) *and* (2.7)*, respectively. Then, we have*

$$\begin{aligned}
(v^k-\tilde{v}^k)^T M^T H M\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\} \\
\geq \quad \frac{1}{2}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}.
\end{aligned} \tag{5.4}$$

Proof. Adding the equation

$$\begin{aligned}
\{(v^k-v^{k+1})-(\tilde{v}^k-\tilde{v}^{k+1})\}^T Q\{(v^k-v^{k+1})-(\tilde{v}^k-\tilde{v}^{k+1})\} \\
= \quad \frac{1}{2}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}
\end{aligned}$$

to both sides of (5.1), we get

$$\begin{aligned} & (v^k - v^{k+1})^T Q\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \\ \geq \quad & \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \end{aligned} \tag{5.5}$$

By using (see (3.4) and (2.8))

$$v^k - v^{k+1} = M(v^k - \tilde{v}^k) \qquad \text{and} \qquad Q = HM,$$

to the term $v^k - v^{k+1}$ in the left hand side of (5.5), we obtain

$$\begin{aligned} & (v^k - \tilde{v}^k)^T M^T HM\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \\ \geq \quad & \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \end{aligned}$$

and the lemma is proved. □

Now, we are ready to prove the monotonicity of the sequence $\{\|v^k - v^{k+1}\|_H^2\}$.

Theorem 5.1 *Let* $\{w^k\}$ *be the sequence generated by the strictly contractive PRSM* (1.8) *and* $\tilde{w}^k$ *be defined in (3.1); the matrix* H *be defined in* (2.7)*. Then,* *we have*

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2. \tag{5.6}$$

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

$$\|a\|_H^2 - \|b\|_H^2 = 2a^T H(a-b) - \|a-b\|_H^2,$$

we obtain

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

Inserting (5.4) into the first term of the right-hand side of the last equality, we obtain

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

The assertion (5.6) follows from the above inequality, the relationship (3.4) and

(2.9) immediately. The proof is complete. $\square$

Now, we can establish a worst-case $O(1/t)$ convergence rate in nonergodic sense for the strictly contractive PRSM scheme (1.8).

Theorem 5.2 *Let $\{w^t\}$ be the sequence generated by the strictly contractive PRSM scheme* (1.8)*. For any $v^* \in \mathcal{V}^*$, we have*

$$\|v^t - v^{t+1}\|_H^2 \le \frac{2(1+\alpha)}{(t+1)(1-\alpha)} \|v^0 - v^*\|_H^2. \tag{5.7}$$

Proof. First, it follows from (3.17) that

$$\frac{1-\alpha}{2(1+\alpha)} \sum_{k=0}^{\infty} \|v^k - v^{k+1}\|_H^2 \le \|v^0 - v^*\|_H^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{5.8}$$

According to Theorem 5.1, the sequence $\{\|v^k - v^{k+1}\|_H^2\}$ is monotonically non-increasing. Therefore, we have

$$(t+1)\|v^t - v^{t+1}\|_H^2 \le \sum_{k=0}^{t} \|v^k - v^{k+1}\|_H^2. \tag{5.9}$$

The assertion (5.7) follows from (5.8) and (5.9) immediately. The proof is complete. □

Notice that $\mathcal{V}^*$ is convex and closed. Let $v^0 = (y^0, \lambda^0)$ be the initial iterate and $d := \inf\{\|v^0 - v^*\|_H \mid v^* \in \mathcal{V}^*\}$. Then, for any given $\epsilon > 0$, Theorem 5.2 shows that the strictly contractive PRSM scheme (1.8) needs at most $\lfloor d^2/\epsilon \rfloor$ iterations to ensure that $\|v^k - v^{k+1}\|_H^2 \le \epsilon$. It follows from Lemma 2.3 that w^{k+1} is a solution of VI(Ω, F, θ) if $\|v^k - v^{k+1}\|_H^2 = 0$. A worst-case $O(1/t)$ convergence rate in nonergodic sense for the strictly contractive PRSM scheme (1.8) is thus established in Theorem 5.2.

References

[1] D.P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods, Academic Press, 1982.

Ökonometrie und Unternehmensforschung, Springer-Verlag, Berlin-Heidelberg-New York, 1975.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Nov. 2010.

[3] T. F. Chan and R. Glowinski, *Finite element approximation and iterative solution of a class of mildly non-linear elliptic equations*, technical report, Stanford University, 1978.

[4] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A relaxed customized proximal point algorithm for separable convex programming. http://www.optimization-online.org/DB_HTML/2011/08/3141.html

[5] E. Corman and X. M. Yuan, *A generalized proximal point algorithm and its convergence rate*, manuscript, 2012.

[6] J. Douglas and H. H. Rachford, *On the numerical solution of the heat conduction problem in 2 and 3 space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421-439.

[7] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vol. I. Springer Series in Operations Research, Springer Verlag, New York, 2003.

[8] M. Fortin and R. Glowinski, Augmented Lagrangian Methods: Applications to the Numerical Solutions of Boundary Value Problems, Stud. Math. Appl. 15, North $-$ Holland, Amsterdam, 1983.

[9] 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.

[10] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems viafinite element approximations, Computers and Mathematics with Applications, 2:17-40, 1976.

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

[12] R. Glowinski and P. Le Tallec, Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics, SIAM Studies in Applied Mathematics, Philadelphia, PA, 1989.

[13] R. Glowinski, On alternating direction methods of multipliers: A historical perspective, 2012.

[14] R. Glowinski, T. Kärkkäinen and K. Majava, *On the convergence of operator-splitting methods*, in Numerical Methods for Scienfic computing, Variational Problems and Applications, edited by Y. Kuznetsov, P. Neittanmaki and O. Pironneau, Barcelona, 2003.

[15] E. G. Gol'shtein and N. V. Tret'yakov, *Modified Lagrangian in convex programming and their generalizations*, Math. Program. Studies, 10 (1979), pp. 86-97.

[16] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, pp. 69-76, 1997.

[17] B.S. He, "Inexact implicit methods for monotone general variational inequalities", Math. Program. Vol. 86, Series A, pp. 199-217, 1999

[18] B. S. He and X. M. Yuan, *On the O(1/n) convergence rate of Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal. 50 (2012), pp. 700-709.

[19] B. S. He and X. M. Yuan, *On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers*, Numerische Mathematik, 130 (2015) 567-577.

[20] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Nmuer. Anal., 16(1979), pp. 964-979.

[21] D. H. Peaceman and H. H. Rachford, *The numerical solution of parabolic elliptic differential equations*, SIAM J. Appl. Math. 3 (1955), pp. 28-41.