凸优化和单调变分不等式的收缩算法

第二十讲: 单调变分不等式意义下凸优化分裂收缩算法的统一框架

A uniform framework of splitting contraction methods for convex optimization in sense of monotone variational inequality

南京大学数学系　　何炳生

hebma@nju.edu.cn

The context of this lecture is based on the publication [21]

1 Splitting Methods in a Unified Framework

We study the algorithms using the guidance of variational inequality. The optimal condition of the linearly constrained convex optimization is resulted in a variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{1.1}$$

The analysis can be fund in [17] (Sections 4 and 5 therin). In order to illustrate the unified framework, let us restudy the augmented Lagrangian method.

1.1 Extended Augmented Lagrangian Method

For the convex optimization

$$\min\{\theta(x) \,|\, Ax = b, \; x \in \mathcal{X}\}. \tag{1.2}$$

Its augmented Lagrangian function is

$$\mathcal{L}_\beta(x, \lambda) = \theta(x) - \lambda^T(Ax - b) + \frac{\beta}{2}\|Ax - b\|^2,$$

where the quadratic term is the penalty for the linear constraints $Ax = b$. The k-th iteration of the **Augmented Lagrangian Method** [28, 32] begins with a given λ^k, obtain

$w^{k+1} = (x^{k+1}, \lambda^{k+1})$ via

$$\text{(ALM)} \quad \left\{ \begin{array}{l} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, \lambda^k) \mid x \in \mathcal{X}\}, \quad (1.3a) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b). \quad (1.3b) \end{array} \right.$$

In (1.3), x^{k+1} is only a computational result of (1.3a) from given λ^k, it is called the intermediate variable. If we denote the output of (1.3) by $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$, then the optimal condition can be written as $\tilde{w}^k \in \Omega$ and

$$\left\{ \begin{array}{lcl} \theta(x) - \theta(\tilde{x}^k) & + & (x - \tilde{x}^k)^T\{-A^T\lambda^k + \beta A^T(A\tilde{x}^k - b)\} \geq 0, \;\; \forall x \in \mathcal{X}, \\ & & (\lambda - \tilde{\lambda}^k)^T\{(A\tilde{x}^k - b) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \;\; \forall \lambda \in \Re^m. \end{array} \right.$$

The above relation can be written as

$$\theta(x) - \theta(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{x}^k - b \end{pmatrix} \geq (\lambda - \tilde{\lambda}^k)^T \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k), \quad (1.4a)$$

for all $w \in \Omega$. In the classical augmented Lagrangian method, $\lambda^{k+1} = \tilde{\lambda}^k$. In practice, we can use relaxation techniques and offer the new iterate by

$$\lambda^{k+1} = \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k), \quad \alpha \in (0, 2). \quad (1.4b)$$

Setting $w = w^*$ in (1.4a), we get $(\tilde{\lambda}^k - \lambda^*)^T \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \geq 0$ and thus

$$(\lambda^k - \lambda^*)^T(\lambda^k - \tilde{\lambda}^k) \geq \|\lambda^k - \tilde{\lambda}^k\|^2. \tag{1.5}$$

By using (1.4b) and (1.5), we get

$$\|\lambda^{k+1} - \lambda^*\|^2 \leq \|\lambda^k - \lambda^*\|^2 - \alpha(2-\alpha)\|\lambda^k - \tilde{\lambda}^k\|^2.$$

In order to describe the algorithm prototype, we give the following definition.

Definition (Intermediate variables and Essential Variables)

For an iterative algorithm solving VI(Ω, F, θ), if some coordinates of w are not involved in the beginning of each iteration, then these coordinates are called intermediate variables and those required by the iteration are called essential variables (denoted by v).

- The sub-vector $w \backslash v$ is called intermediate variables.
- In some Algorithms, v is a proper sub-vector of w; however, $v = w$ is also possible.

According to the above mentioned definition, in the the augmented Lagrangian method, x is an intermediate variable and λ is the essential variable.

1.2 Algorithms in a unified framework

A Prototype Algorithm for (1.1)

[Prediction Step.] With given v^k, find a vector $\tilde{w}^k \in \Omega$ which satisfying

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k), \ \forall w \in \Omega, \quad \text{(1.6a)}$$

where the matrix Q has the property: $Q^T + Q$ is positive definite.

[Correction Step.] Determine a nonsingular matrix M and a scalar $\alpha > 0$, let

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \quad \text{(1.6b)}$$

- Usually, we do not take the output of (1.6a), $\tilde{v}^k$, as the new iterate. Thus, $\tilde{v}^k$ is called the predictor. The new iterate v^{k+1} given by (1.6b) is called the corrector. The prototype algorithm is a Prediction-Correction Method.

- We use the extended ALM in Section 1.1 as an example. In (1.4a) we have $v = \lambda$, $Q = \frac{1}{\beta} I$, while in the correction step (1.4b), $M = I$ and $\alpha \in (0, 2)$.

- When $v^k = \tilde{v}^k$, it follows from (1.6a) directly that $\tilde{w}^k$ is a solution of (1.1). Thus, one can use $\|v^k - \tilde{v}^k\| < \epsilon$ as the stopping criterion in (1.6).

Convergence Conditions

For the matrices Q and M, and the step size α determined in (1.6), the matrices

$$H = QM^{-1} \tag{1.7a}$$

and

$$G = Q^T + Q - \alpha M^T H M. \tag{1.7b}$$

are positive definite (or $H \succ 0$ and $G \succeq 0$).

- We use the extended ALM in Section 1.1 as an example. Since $Q = \frac{1}{\beta} I$ in the prediction step, and $M = I$ and $\alpha \in (0, 2)$ in the correction step, it follows that

  $$H = QM^{-1} = \frac{1}{\beta} I \qquad \text{and} \qquad G = Q^T + Q - \alpha M^T H M = \frac{2-\alpha}{\beta} I.$$

  Therefore, the convergence conditions are satisfied.

- For $G \succeq 0$, it has the $O(1/t)$ convergence rate in a ergodic sense. If $G \succ 0$, the sequence $\{v^k\}$ is Fèjer monotone and converges to a $v^* \in \mathcal{V}^*$ in H-norm.

- Using the unified framework, the convergence proof is very simple. In addition, it will help us to construct more efficient splitting contraction method for convex optimization with different structures.

Given a positive definite matrix Q in (1.6a) ($Q^T + Q \succ 0$), for satisfying the convergence conditions (1.7), how to choose the matrix M and $\alpha > 0$ in the correction step (1.6b) ?

There are many possibilities, the principle is simplicity and efficiency. See an example:

- In order to ensure the symmetry and positivity of $H = QM^{-1}$, we take

$$H = QD^{-1}Q^T,$$

where D is a symmetric investable block diagonal matrix. Because

$$H = QD^{-1}Q^T \qquad \text{and} \qquad H = QM^{-1},$$

we only need to set $M^{-1} = D^{-1}Q^T$ and thus

$$M = Q^{-T}D \quad \text{satisfies the condition (1.7a).}$$

In this case, $M^THM = Q^TM = Q^TQ^{-T}D = D$.

- After choosing the matrix M, let

$$\alpha_{\max} = \arg\max\{\alpha \,|\, Q^T + Q - \alpha M^THM \succeq 0\},$$

the condition (1.7b) is satisfied for any $\alpha \in (0, \alpha_{\max})$.

1.3 Customized PPA satisfies the Convergence Condition

Recall the convex optimization problem discussed in Lecture 6, namely,

$$\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\}.$$

The related variational inequality of the saddle point of the Lagrangian function is

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega.$$

where

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \Re^m.$$

For given $v^k = w^k = (x^k, \lambda^k)$, let the output of the Customized PPA as a predictor and denote it as $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$. Then, we have

$$\left\{ \begin{array}{l} \tilde{x}^k = \arg\min\big\{\theta(x) - (\lambda^k)^T(Ax - b) + \frac{r}{2}\|x - x^k\|^2 \;\big|\; x \in \mathcal{X}\big\}, \\ \tilde{\lambda}^k = \lambda^k - \frac{1}{s}[A(2\tilde{x}^k - x^k) - b]. \end{array} \right. \tag{1.8}$$

Similar as (1.4), the output $\tilde{w}^k \in \Omega$ of the iteration (1.8) satisfies

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega.$$

It is a form of (1.6a) where

$$Q = H = \begin{pmatrix} rI & A^T \\ A & sI \end{pmatrix}.$$

This matrix is positive definite when $rs > \|A^T A\|$. We take $M = I$ in the correction (1.6b) and the new iterate is updated by

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

Then, we have and

$$H = QM^{-1} = Q \succ 0 \quad \text{and} \quad G = Q^T + Q - \alpha M^T H M = (2 - \alpha)H \succ 0.$$

The convergence conditions (1.7) are satisfied. More about customized PPA, please see

♣ G.Y. Gu, B.S. He and X.M. Yuan, Customized Proximal point algorithms for linearly constrained convex minimization and saddle-point problem: a unified Approach, Comput. Optim. Appl., 59(2014), 135-161.

1.4 Primal-Dual relaxed PPA-based Contraction Methods

For solving (1.2), with given $v^k = w^k = (x^k, \lambda^k)$, let $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$ be produced by

$$\tilde{x}^k = \arg\min\big\{\theta(x) - (\lambda^k)^T(Ax-b) + \tfrac{r}{2}\|x - x^k\|^2 \;\big|\; x \in \mathcal{X}\big\} \tag{1.9a}$$

and (according to equality constraints $Ax = b$ or inequality constraints $Ax \geq b$)

$$\tilde{\lambda}^k = \lambda^k - \frac{1}{s}(A\tilde{x}^k - b) \quad \text{or} \quad \tilde{\lambda}^k = [\lambda^k - \frac{1}{s}(A\tilde{x}^k - b)]_+. \tag{1.9b}$$

The predictor $\tilde{w}^k \in \Omega$ generated by (1.9) satisfies

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \; \forall w \in \Omega, \tag{1.10}$$

where the matrix

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}, \tag{1.11}$$

is not symmetric. However, (1.10) can be viewed as (1.6a). In this subsection, all the mentioned matrix Q is (1.11). There is example to show that the method is not necessary convergent if we directly take $w^{k+1} = \tilde{w}^k$.

Corrector—the new iterate For given v^k and the predictor $\tilde{v}^k$ by (1.9), we use

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{1.12}$$

to produce the new iterate, where

$$M = \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix}$$

is a upper triangular block matrix whose diagonal part is unit matrix. Note that

$$H = QM^{-1} = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & -\frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix} \succ 0.$$

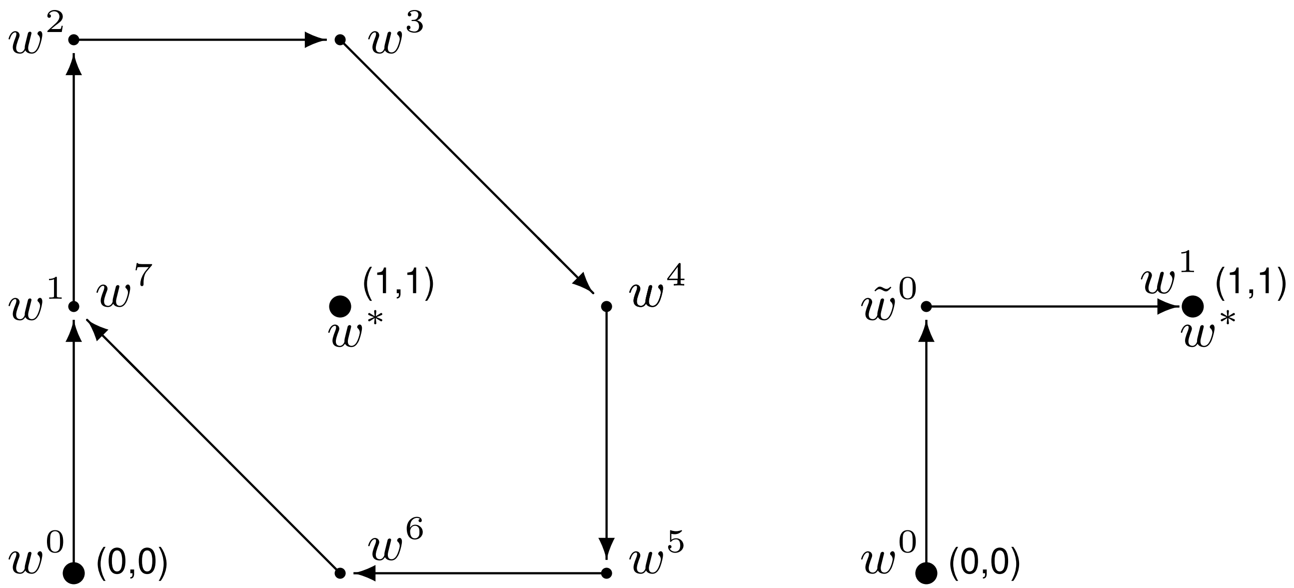
In addition,

$$\begin{aligned} G &= Q^T + Q - M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} rI_n & 0 \\ 0 & sI_m - \frac{1}{r}AA^T \end{pmatrix}. \end{aligned}$$

G is positive definite when $rs > \|A^T A\|$. The convergence conditions (1.7) are satisfied.

Convergence behaviors for LP

A Toy Example: $\min\{x \mid x = 1,\ x \geq 0\}$, $(x^*, y^*) = (1, 1)$.



Original PDHG **PDHG + Correction**

This example shows, sometimes the correction has surprising effectiveness.

In the correction step (1.12), the matrix M is a upper-triangular matrix. We can also use the lower-triangular matrix

$$M = \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix}$$

According to (1.7a), $H = QM^{-1}$, by a simple computation, we have

$$H = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ \frac{1}{s}A & I_m \end{pmatrix} = \begin{pmatrix} rI_n + \frac{1}{s}A^TA & A^T \\ A & sI_m \end{pmatrix}.$$

H is positive definite for any $r, s > 0$. In addition,

$$\begin{aligned} G &= Q^T + Q - M^THM = Q^T + Q - Q^TM \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix} = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}. \end{aligned}$$

G is positive definite when $rs > \|A^TA\|$. The convergence conditions (1.7) are satisfied.

For a given prediction, there are different corrections which satisfy the convergence conditions (1.7). For example, we can take a convex combination of the above mentioned matrices. Namely, for $\tau \in [0,1]$

$$
\begin{aligned}
M &= (1-\tau)\begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} + \tau \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix} \\
&= \begin{pmatrix} I_n & \frac{1-\tau}{r}A^T \\ -\frac{\tau}{s}A & I_m \end{pmatrix}.
\end{aligned}
$$

For this matrix M, we denote

$$\Pi = I + \frac{\tau(1-\tau)}{rs}AA^T.$$

Clearly, Π is positive definite. Let

$$H = \begin{pmatrix} rI_n + \frac{\tau^2}{s}A^T\Pi^{-1}A & \tau A^T\Pi^{-1} \\ \tau\Pi^{-1}A & s\Pi^{-1} \end{pmatrix}.$$

It is easy to verify that H is positive definite for any $r, s > 0$ and

$$HM = Q.$$

Now, we turn to observe the matrix G, it leads that

$$\begin{aligned} G &= Q^T + Q - M^T HM = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ A & sI_m \end{pmatrix} \begin{pmatrix} I_n & \frac{1-\tau}{r} A^T \\ \frac{-\tau}{s} A & I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n & \tau A^T \\ \tau A & s(I_m - \frac{1-\tau}{rs} AA^T) \end{pmatrix} \\ &= \begin{pmatrix} rI_n & \tau A^T \\ \tau A & \tau^2 sI \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & s(1-\tau)[(1+\tau)I_m - \frac{1}{rs} AA^T] \end{pmatrix}. \end{aligned}$$

For $\tau \in [0,1]$, G is positive definite when $rs > \|A^T A\|$. The convergence conditions (1.7) are satisfied. Especially, in the case $\tau = 1/2$, when $rs > \frac{3}{4}\|A^T A\|$,

$$G = \begin{pmatrix} rI_n & \frac{1}{2} A^T \\ \frac{1}{2} A & s(I_m - \frac{1}{2rs} AA^T) \end{pmatrix} \succ 0.$$

We do not need to calculate H and G, only verifying their positivity is necessary.

2 Convergence proof in the unified framework

In this section, assuming the conditions (1.7) in the unified framework are satisfied, we prove some convergence properties.

Theorem 2.1 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (1.1) *and* $\tilde{w}^k$ *is obtained in the* k*-th iteration. If* v^k, v^{k+1} *and* $\tilde{w}^k$ *satisfy the conditions in the unified framework, then we have*

$$\begin{aligned} &\alpha\big(\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)\big) \\ &\quad \geq \quad \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{\alpha}{2}\|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega. \end{aligned} \quad (2.1)$$

Proof. Using $Q = HM$ (see (1.7a)) and the relation (1.6b), the right hand side of (1.7a) can be written as $(v - \tilde{v}^k)^T \frac{1}{\alpha} H(v^k - v^{k+1})$ and hence

$$\alpha\{\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)\} \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \ \forall w \in \Omega. \quad (2.2)$$

Applying the identity

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

to the right hand side of (2.2) with

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

we thus obtain

$$\begin{aligned}&(v - \tilde{v}^k)^T H(v^k - v^{k+1}) \\ &\quad = \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2).\end{aligned} \tag{2.3}$$

For the last term of (2.3), we have

$$\begin{aligned}&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ &\quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ &\quad \overset{(1.6b)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - \alpha M(v^k - \tilde{v}^k)\|_H^2 \\ &\quad = \quad 2\alpha(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - \alpha^2(v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ &\quad = \quad \alpha(v^k - \tilde{v}^k)^T(Q^T + Q - \alpha M^T HM)(v^k - \tilde{v}^k) \\ &\quad \overset{(1.7b)}{=} \quad \alpha\|v^k - \tilde{v}^k\|_G^2.\end{aligned} \tag{2.4}$$

Substituting (2.3), (2.4) in (2.2), the assertion of this theorem is proved. $\square$

2.1 Convergence in a strictly contraction sense

For the convergence in a strictly contraction sense, the matrix G should be positive definite.

Theorem 2.2 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (1.1) *and* $\tilde{w}^k$ *is obtained in the* k*-th iteration. If* v^k*,* v^{k+1} *and* $\tilde{w}^k$ *satisfy the conditions in the unified framework, then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha\|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.5}$$

Proof. Setting $w = w^*$ in (2.1), we get

$$\begin{aligned} & \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ & \quad \ge \quad \alpha\|v^k - \tilde{v}^k\|_G^2 + 2\alpha\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \end{aligned} \tag{2.6}$$

By using the optimality of w^* and the monotonicity of $F(w)$, we have

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \ge 0$$

and thus

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \alpha\|v^k - \tilde{v}^k\|_G^2. \tag{2.7}$$

The assertion (2.5) follows directly. □

2.2 Convergence rate in an ergodic sense

Equivalent Characterization of the Solution Set of VI

For the convergence rate analysis, we need another characterization of the solution set of VI (1.1). It can be described the following theorem and the proof can be found in [9] (Theorem 2.3.5) or [25] (Theorem 2.1).

Theorem 2.3 *The solution set of VI*(Ω, F, θ) *is convex and it can be characterized as*

$$\Omega^* = \bigcap_{w\in\Omega} \big\{\tilde{w} \in \Omega : \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \geq 0\big\}. \tag{2.8}$$

Proof. Indeed, if $\tilde{w} \in \Omega^*$, we have

$$\theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(\tilde{w}) \geq 0, \quad \forall\, w \in \Omega.$$

By using the monotonicity of F on Ω, this implies that

$$\theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq 0, \quad \forall\, w \in \Omega.$$

Thus, $\tilde{w}$ belongs to the right-hand set in (2.8). Conversely, suppose $\tilde{w}$ belongs to the latter set of (2.8). Let $w \in \Omega$ be arbitrary. The vector

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

belongs to Ω for all $\alpha \in (0, 1)$. Thus we have

$$\theta(\bar{u}) - \theta(\tilde{u}) + (\bar{w} - \tilde{w})^T F(\bar{w}) \geq 0. \tag{2.9}$$

Because $\theta(\cdot)$ is convex, we have

$$\theta(\bar{u}) \leq \alpha\theta(\tilde{u}) + (1 - \alpha)\theta(u) \quad \Rightarrow \quad (1 - \alpha)(\theta(u) - \theta(\tilde{u})) \geq \theta(u) - \theta(\tilde{u}).$$

Substituting it in (2.9) and using $\bar{w} - \tilde{w} = (1 - \alpha)(w - \tilde{w})$, we get

$$(\theta(u) - \theta(\tilde{u})) + (w - \tilde{w})^T F(\alpha\tilde{w} + (1 - \alpha)w) \geq 0$$

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

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

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

$$\{\tilde{w} \in \Omega : \; \theta(\tilde{u}) + \tilde{w}^T F(w) \le \theta(u) + w^T F(w)\}$$

and its equivalent expression

$$\{\tilde{w} \in \Omega : \; \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \ge 0\}$$

is convex. Since the intersection of any number of convex sets is convex, it follows that the solution set of VI(Ω, F, θ) is convex. $\square$

In Theorem 2.3, we have proved the equivalence of

$$\tilde{w} \in \Omega, \;\; \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(\tilde{w}) \ge 0, \;\; \forall\, w \in \Omega,$$

and

$$\tilde{w} \in \Omega, \;\; \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \ge 0, \;\; \forall\, w \in \Omega.$$

We use the late one to define the approximate solution of VI (1.1). Namely, for given $\epsilon > 0$, $\tilde{w} \in \Omega$ is called an ϵ-approximate solution of VI(Ω, F, θ), if it satisfies

$$\tilde{w} \in \Omega, \;\; \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \ge -\epsilon, \;\; \forall\, w \in \mathcal{D}_{(\tilde{w})},$$

where

$$\mathcal{D}_{(\tilde{w})} = \{w \in \Omega \,|\, \|w - \tilde{w}\| \le 1\}.$$

We need to show that for given $\epsilon > 0$, after t iterations, it can offer a $\tilde{w} \in \mathcal{W}$, such that

$$\tilde{w} \in \mathcal{W} \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \big\{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\big\} \le \epsilon. \tag{2.10}$$

Theorem 2.1 is also the base for the convergence rate proof. Using the monotonicity of F, we have

$$(w - \tilde{w}^k)^T F(w) \ge (w - \tilde{w}^k)^T F(\tilde{w}^k).$$

Substituting it in (2.1), we obtain

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2\alpha}\|v - v^k\|_H^2 \ge \frac{1}{2\alpha}\|v - v^{k+1}\|_H^2, \ \forall w \in \Omega. \tag{2.11}$$

Note that the above assertion is hold for $G \succeq 0$.

Theorem 2.4 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (1.1) *and* $\tilde{w}^k$ *is obtained in the* k*-th iteration. Assume that* v^k, v^{k+1} *and* $\tilde{w}^k$ *satisfy the conditions* *in the unified framework and let $\tilde{w}_t$ be defined by*

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k. \tag{2.12}$$

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

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2\alpha(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{2.13}$$

Proof. First, it holds that $\tilde{w}^k \in \Omega$ for all $k \ge 0$. Together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (2.12) implies that $\tilde{w}_t \in \Omega$. Summing the inequality (2.11) over $k = 0, 1, \dots, t$, we obtain

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

Use the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2\alpha(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{2.14}$$

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 (2.14), the assertion of this theorem follows directly. $\square$

Recall (2.10). The conclusion (2.13) thus indicates obviously that the method is able to generate an approximate solution (i.e., $\tilde{w}_t$) with the accuracy $O(1/t)$ after t iterations. That is, in the case $G \succeq 0$, the convergence rate $O(1/t)$ of the method is established.

- **For the unified framework and the convergence proof, the reader can consult: B.S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman-Rachford splitting method for convex programming, *SIAM Journal on Optimization* 24(2014), 1011-1040.**
- **B. S. He and X. M. Yuan, On the $O(1/n)$ convergence rate of the alternating direction method, *SIAM J. Numerical Analysis* 50(2012), 700-709.**

2.3 Convergence rate in pointwise iteration-complexity

In this subsection, we show that if the matrix G defined in (1.7b) is positive definite, a worst-case $O(1/t)$ convergence rate in a nonergodic sense can also be established for the prototype algorithm (1.6). Note in general a nonergodic convergence rate is stronger than the ergodic convergence rate.

We first need to prove the following lemma.

Lemma 2.1 *For the sequence generated by the prototype algorithm (1.6) where the Convergence Condition is satisfied, we have*

$$\begin{aligned} (v^k - \tilde{v}^k)^T M^T H M\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \\ \geq \quad \frac{1}{2\alpha}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \end{aligned} \tag{2.15}$$

Proof. First, set $w = \tilde{w}^{k+1}$ in (1.6a), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{2.16}$$

Note that (1.6a) is also true for $k := k+1$ and thus we have

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

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

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}). \tag{2.17}$$

Combining (2.16) and (2.17) and using the monotonicity of F, we get

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq 0. \tag{2.18}$$

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 (2.18), and using $v^T Q v = \frac{1}{2} v^T (Q^T + Q) v$, we obtain

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

Substituting $(v^k - v^{k+1}) = \alpha M(v^k - \tilde{v}^k)$ in the left-hand side of the last inequality and using $Q = HM$, we obtain (2.15) and the lemma is proved. □

Now, we are ready to prove (2.19), the key inequality in this section.

Theorem 2.5 *For the sequence generated by the prototype algorithm (1.6) where the Convergence Condition is satisfied, we have*

$$\|M(v^{k+1}-\tilde{v}^{k+1})\|_H \le \|M(v^k-\tilde{v}^k)\|_H, \quad \forall\, k>0. \tag{2.19}$$

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

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

Inserting (2.15) into the first term of the right-hand side of the last equality, we obtain

$$\begin{aligned}
&\|M(v^k - \tilde{v}^k)\|_H^2 - \|M(v^{k+1} - \tilde{v}^{k+1})\|_H^2 \\
&\quad \geq \quad \frac{1}{\alpha}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2 - \|M[(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})]\|_H^2 \\
&\quad = \quad \frac{1}{\alpha}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2 \geq 0,
\end{aligned}$$

where the last inequality is because of the positive definiteness of the matrix $(Q^T + Q) - \alpha M^T HM \succeq 0$. The assertion (2.19) follows immediately. □

Note that it follows from $G \succ 0$ and Theorem 2.2 there is a constant $c_0 > 0$ such that

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - c_0\|M(v^k - \tilde{v}^k)\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.20}$$

Now, with (2.20) and (2.19), we can establish the worst-case $O(1/t)$ convergence rate in a nonergodic sense for the prototype algorithm (1.6).

Theorem 2.6 *Let $\{v^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the prototype algorithm (1.6) under the Convergence Condition. For any integer $t > 0$, we have*

$$\|M(v^t - \tilde{v}^t)\|_H^2 \leq \frac{1}{(t+1)c_0}\|v^0 - v^*\|_H^2. \tag{2.21}$$

Proof. First, it follows from (2.20) that

$$\sum_{k=0}^{\infty} c_0 \|M(v^k - \tilde{v}^k)\|_H^2 \le \|v^0 - v^*\|_H^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{2.22}$$

According to Theorem 2.5, the sequence $\{\|M(v^k - \tilde{v}^k)\|_H^2\}$ is monotonically non-increasing. Therefore, we have

$$(t+1)\|M(v^t - \tilde{v}^t)\|_H^2 \le \sum_{k=0}^{t} \|M(v^k - \tilde{v}^k)\|_H^2. \tag{2.23}$$

The assertion (2.21) follows from (2.22) and (2.23) immediately. □

Let $d := \inf\{\|v^0 - v^*\|_H \mid v^* \in \mathcal{V}^*\}$. Then, for any given $\epsilon > 0$, Theorem 2.6 shows that the it needs at most $\lfloor d^2/c_0\epsilon \rfloor$ iterations to ensure that $\|M(v^k - \tilde{v}^k)\|_H^2 \le \epsilon$. Recall that v^k is a solution of VI(Ω, F, θ) if $\|M(v^k - \tilde{v}^k)\|_H^2 = 0$ (see (1.6a) and due to $Q = HM$). A worst-case $O(1/t)$ convergence rate in pointwise iteration-complexity is thus established for the prototype algorithm (1.6).

Notice that, for a differentiable unconstrained convex optimization $\min f(x)$, it holds that

$$f(x) - f(x^*) = \nabla f(x^*)^T (x - x^*) + O(\|x - x^*\|^2) = O(\|x - x^*\|^2).$$

3 Conclusions and Remarks

3.1 ADMM vs AMA

We consider the following convex optimization problem

$$\min \{\theta(u) \,|\, \mathcal{A}u = b,\ u \in \mathcal{U}\} \tag{3.1}$$

Solving (3.1) **by using the penalty function method**

$$u^{k+1} = \text{Argmin}\Big\{\theta(u) + \frac{\beta_k}{2}\|\mathcal{A}u - b\|^2 \Big| u \in \mathcal{U}\Big\}$$

Solving (3.1) **by using the augmented Lagrangian method**

Begin with a given λ^k,

$$u^{k+1} = \text{Argmin}\Big\{\theta(u) - (\lambda^k)^T(\mathcal{A}u - b) + \tfrac{\beta}{2}\|\mathcal{A}u - b\|^2 \Big| x \in \mathcal{X}\Big\}$$

$$\lambda^{k+1} = \lambda^k - \beta(\mathcal{A}u^{k+1} - b).$$

我们现在要求解的是

$$\min\{\theta_1(x)+\theta_2(y)\,|\,Ax+By=b,\ x\in\mathcal{X},\ y\in\mathcal{Y}\} \tag{3.2}$$

求解问题 (3.2) 的罚函数方法

$$(x^{k+1},y^{k+1})=\text{Argmin}\Big\{\theta_1(x)+\theta_2(y)+\frac{\beta_k}{2}\|Ax+By-b\|^2\Big|x\in\mathcal{X},\ y\in\mathcal{Y}\Big\}$$

求解问题 (3.2) 的增广 Lagrange 乘子法 从给定的 λ^k 开始

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

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

两类不同方法, 子问题难能度完全一样.
增广 Lagrange 乘子法优于罚函数方法。
J. Nocedal, S. J. Wright: Numerical Optimization

处理具有可分离结构问题 (3.2) 时共同的缺点：
没有利用 x 和 y 的可分离结构！ 求解会无从着手。

- 两个可分离算子问题罚函数方法松弛后就是交替极小化算法 AMA。
- 两个可分离算子问题的增广 Lagrange 乘子法松弛就是乘子交替方向法 ADMM。

求解问题 (3.2) 的松弛的罚函数方法 — 交替极小化方法(AMA)

从给定的 y^k 开始

$$x^{k+1} = \text{Argmin}\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k - b\|^2 | x \in \mathcal{X}\},$$
$$y^{k+1} = \text{Argmin}\{\theta_2(y) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 | y \in \mathcal{Y}\}.$$

求解问题 (3.2) 的松弛的增广 Lagrange 乘子法 — ADMM

从给定的 (y^k, λ^k) 开始

$$x^{k+1} = \text{Argmin}\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + By^k - b\|^2 | x \in \mathcal{X}\},$$
$$y^{k+1} = \text{Argmin}\{\theta_2(y) - (\lambda^k)^T By + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 | y \in \mathcal{Y}\},$$
$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b).$$

都松弛, 乘子交替方向法 (ADMM) 应该优于交替极小化方法 (AMA)

3.2 两个算子问题 ADMM 方法的 (主要) 改进

1. ADMM in sense of PPA 交换顺序并外延 从 (y^k, λ^k) 出发.

$$\left\{ \begin{array}{l} \left\{ \begin{array}{lcl} x^{k+1} & = & \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} & = & \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+1}) \,|\, y \in \mathcal{Y}\}, \end{array} \right. \quad (3.3a) \\ \left\{ \begin{array}{lcl} y^{k+1} & := & y^k - \gamma(y^k - y^{k+1}), \\ \lambda^{k+1} & := & \lambda^k - \gamma(\lambda^k - \lambda^{k+1}). \end{array} \right. \quad \text{（松弛延拓）} \quad (3.3b) \end{array} \right.$$

这里 $\gamma \in (0, 2)$. 赋值号 $:=$ 表示 (3.3b) 右端的 (y^{k+1}, λ^{k+1}) 是由算法的前半部分 (3.3a) 产生的. 对多数问题, 这样往往能加快收敛速度.

- X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A proximal point algorithms revisit on the alternating direction method of multipliers, Science China Math., 56 (2013), 2179-2186.

2. Symmetric ADMM　对称的交替方向法

原始变量 x 和 y 本质上是平等的. 所以建议采用对称的交替方向法.

Symmetric Alternating Direction Method of Multipliers is described as

$$\left\{\begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\[2ex] \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), \\[2ex] y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, \\[2ex] \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b). \end{array}\right. \tag{3.4}$$

wehre $\mu \in (0,1)$ (usually $\mu = 0.9$).

- B.S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman- Rachford splitting method for convex programming, *SIAM Journal on Optimization* **24**(2014), 1011-1040.

3.3 多个可分离算子的凸优化问题

我们以 3 个可分离算子的问题为例

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)|Ax+By+Cz=b,\ x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}\}. \quad (3.5)$$

这个问题的增广 Lagrange 函数是

$$\mathcal{L}_\beta^3(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b)+\frac{\beta}{2}\|Ax+By+Cz-b\|^2.$$

$$\left\{\begin{array}{rcl} x^{k+1} & = & \arg\min\{\mathcal{L}_\beta^3(x,y^k,z^k,\lambda^k)\ |\ x\in\mathcal{X}\}, \\ y^{k+1} & = & \arg\min\{\mathcal{L}_\beta^3(x^{k+1},y,z^k,\lambda^k)\ |\ y\in\mathcal{Y}\}, \\ z^{k+1} & = & \arg\min\{\mathcal{L}_\beta^3(x^{k+1},y^{k+1},z,\lambda^k)\ |\ z\in\mathcal{Z}\}, \\ \lambda^{k+1} & = & \lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b). \end{array}\right. \quad (3.6)$$

对 $m\geq 3$, 直接推广的交替方向法不能保证收敛.

- C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, Mathematical Programming, 155 (2016) 57-79.

直接推广 ADMM： 我们发表在 2016 Math.Progr. 的三个算子问题

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

的第一个例子中， $\theta_1(x)=\theta_2(y)=\theta_3(z)=0,\mathcal{X}=\mathcal{Y}=\mathcal{Z}=\Re,$

$\mathcal{A}=[A,B,C]\in\Re^{3\times3}$ 是个非奇异矩阵， $b=0\in\Re^3.$

还有一些据此延伸的例子, 证明了直接推广的 ADMM 并不收敛.
这些例子更多的是在理论方面的意义.

值得继续研究的问题： 三个算子的实际问题中, 线性约束矩阵

$\mathcal{A}=[A,B,C]$ 往往至少有一个是单位矩阵, 即， $\mathcal{A}=[A,B,I].$

直接推广的 ADMM 处理这种更贴近实际的三个算子的问题,
既没有证明收敛，也没有举出反例，至今我们于心不甘！！

举个简单的例子来说：

- 乘子交替方向法 (ADMM) 处理问题

$$\min\{\theta_1(x)+\theta_2(y)|Ax+By=b,\ x\in\mathcal{X},y\in\mathcal{Y}\}$$ 是收敛的.

- 将等式约束换成不等式约束, 问题就变成

$$\min\{\theta_1(x)+\theta_2(y)|Ax+By\le b,\ x\in\mathcal{X},y\in\mathcal{Y}\}.$$

- 再化成三个算子的等式约束问题

$$\min\{\theta_1(x)+\theta_2(y)+0\,|\,Ax+By+z=b,\ x\in\mathcal{X},y\in\mathcal{Y},z\ge 0\}$$

- 直接推广的 ADMM 处理上面这种问题, 不少人做过尝试, 但是至今既没有证明收敛性, 也没有举出反例！

基于上述认知, 我们对三个算子的问题提出了一些修正算法. 注意：我们的方法对问题不加任何条件！ 对 β 不加限制, 只对方法动手术！

处理办法一：带高斯回代的 ADMM 方法

以 (3.6) 提供的 (y^{k+1}, z^{k+1}) 为预测, 取 $\alpha \in (0,1)$, 校正公式为

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} := \begin{pmatrix} y^k \\ z^k \end{pmatrix} - \alpha \begin{pmatrix} I & -(B^TB)^{-1}B^TC \\ 0 & I \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}. \quad (3.7)$$

由于为下一步迭代只要准备 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$, 我们只要做

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \end{pmatrix} := \begin{pmatrix} By^k \\ Cz^k \end{pmatrix} - \alpha \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} B(y^k - y^{k+1}) \\ C(z^k - z^{k+1}) \end{pmatrix}.$$

- 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.

对 y 和 z, 有先后, 不公平, 那就要做找补, 调整.

处理办法二： ADMM + Prox-Parallel Splitting ALM

y, z 子问题平行, 如果不想做后处理, 就给它们俩预先都加个正则项

$$\begin{cases} x^{k+1} = \arg\min\left\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\right\}, \\ y^{k+1} = \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\beta\|B(y-y^k)\|^2 \middle| y \in \mathcal{Y}\right\}, \\ z^{k+1} = \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\beta\|C(z-z^k)\|^2 \middle| z \in \mathcal{Z}\right\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{cases}$$

- B. He, M. Tao and X. Yuan, A splitting method for separable convex programming. IMA J. Numerical Analysis, 31(2015), 394-426.

太自由, 又不校正, 就加正则项, 不忘自己昨天的承诺.

求解线性约束凸优化，增广拉格朗日乘子法 (ALM) 优于罚函数方法，有点优化基础知识的人都知道。

对两个可分离算子的线性约束凸优化问题，增广拉格朗日乘子法 (ALM) 和罚函数方法, 松弛后分别成了乘子交替方向法 (ADMM) 和交替极小化方法 (AMA)。

人们因此有理由对 ADMM 格外关心。

ADMM 不是我们提出来的。有了 10 年投影收缩算法的研究的基础, 使得我对 ADMM 类方法格外感兴趣。带领学生对ADMM 方法做一些有价值的改进和证明一些重要的理论结果，便顺理成章。

方法上，交换了原始变量 y 和对偶变量 λ 次序，进而得到因需定制的 PPA 意义下的 ADMM (Science in China, Mathematics, 2013)；

平等对待原始变量 x 和 y，两次校正对偶变量 λ，就得到对称型的 ADMM (SIAM Optimization, 2014)。

这些方法，道理上能站住脚，计算表现也不俗。

理论上，我们证明了 ADMM 在遍历意义下 (SIAM Numerical Analysis，2012) 和点列意义下(Numer. Mathematik, 2015) 的 $O(1/t)$ 的收敛速率. 证明都不复杂.

ADMM 的广泛应用，人们自然想到向三个算子的问题推广。

我们在不能证明“直接推广的方法”收敛的时候，提出了一些处理多个算子问题的ADMM 类方法

(Computational Optimization and Applications, 2009).

(SIAM Optimization, 2012;   IMA Numerical Analysis, 2015).

这些方法的共同特点是不需要对问题加任何条件！对 β 不加限制，只对方法动手术！

后来我们又给出“直接推广的ADMM 方法处理三个算子问题不保证收敛”的例子(Math. Progr., 2016),说明:
以前提出的一些策略，手段上是必须的，机制上也是合理的。

- The analysis is guided by variational inequality.
- The most methods mentioned fall in a unified prediction-correction framework, in which the convergence analysis is quite simple.
- All the discussed methods are closed related to Proximal Point Algorithms.
- All the discussed ADMM-like splitting methods are rooted from Augmented Lagrangian Method.
- A through reading will acquaint you with the ADMM, while a more carefully reading may make you familiar with the tricks on constructing splitting methods according to the problem you met.
- The discussed first order splitting contraction methods are only appropriate for some structured convex optimization in some practical applications.

References

[1] 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 Vol. 3, No. 1 (2010) 1 – 122.

[2] J. F. Cai, E. J. Candès and Z. W. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. **20**, 1956-1982, 2010.

[3] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A proximal point algorithms revisit on the alternating direction method of multipliers, Science China Mathematics, 56 (2013), 2179-2186.

[4] A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vison, 40, 120-145, 2011.

[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] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, to appear in *Mathematical Programming, Series A*.

[7] 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.

[8] 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, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

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

[10] 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.

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

[12] D. Hallac, Ch. Wong, S. Diamond, A. Sharang, R. Sosič, S. Boyd and J. Leskovec, SnapVX: A Network-Based Convex Optimization Solver, Journal of Machine Learning Research 18 (2017) 1-5.

[13] B.S. He, *A new method for a class of linear variational inequalities*, Math. Progr., **66**, pp. 137-144, 1994.

[14] B.S. He, *Solving a class of linear projection equations*, Numerische Mathematik, **68**, pp. 71-80, 1994.

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

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

[17] B. S. He, PPA-like contraction methods for convex optimization: a framework using variational inequality approach. J. Oper. Res. Soc. China 3 (2015) 391 – 420.

[18] 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.

[19] B. S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman-Rachford splitting method for convex programming, *SIAM Journal on Optimization* **24**(2014), 1011-1040.

[20] 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.

[21] B.S. He, M. Tao and X.M. Yuan, A splitting method for separable convex programming, *IMA J. Numerical Analysis* **31**(2015), 394-426.

[22] 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.

[23] 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.

[24] B.S. He, H. Yang, and S.L. Wang, Alternating directions method with self-adaptive penalty parameters for monotone variational inequalities, *JOTA* **23**(2000), 349–368.

[25] 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.

[26] B.S. He and X.M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, *SIAM J. Imag. Science* **5**(2012), 119-149.

[27] 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.

[28] M. R. Hestenes, Multiplier and gradient methods, *JOTA* **4**, 303-320, 1969.

[29] Z. Lin, R. Liu and H. Li, *Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning*, Machine Learning, 99 (2) (2015), pp. 287-325.

[30] 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.

[31] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, *Rev. Francaise d'Inform. Recherche Oper.*, **4**, 154-159, 1970.

[32] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in Optimization, R. Fletcher, ed., Academic Press, New York, NY, pp. 283-298, 1969.

[33] X. F. Wang and X. M. Yuan, *The linearized alternating direction method of multipliers for Dantzig selector*, SIAM J. Sci. Comput., 34 (2012), pp. A2792–A2811.

[34] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Cont. Optim.*, **14**, 877-898, 1976.

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

[36] J. F. Yang and X. M. Yuan, *Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization*, Math. Comp., 82 (2013), pp. 301-329.

[37] X. Q. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (2010), pp. 20-46.