典型凸优化问题的分裂收缩算法讲座

VI. 多块可分离凸优化问题的Gauss型预测-校正方法
一类适用范围更广和便于推广的交替方向法

何 炳 生　　南京大学数学系

Homepage: maths.nju.edu.cn/~hebma

江苏省研究生视觉计算与可信人工智能暑期学校

2022年7月15日

1 从 ADMM 谈起

ADMM 处理的问题是

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

经典的 ADMM 求解(1.1)： From (y^k,λ^k) to (y^{k+1},λ^{k+1})

$$\left\{\begin{array}{rcl} x^{k+1} &\in& \arg\min\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k-b\|^2 \,|\, x\in\mathcal{X}\},\\ y^{k+1} &\in& \arg\min\{\theta_2(y)-y^TB^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By-b\|^2 \,|\, y\in\mathcal{Y}\},\\ \lambda^{k+1} &=& \lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). \end{array}\right. \tag{1.2}$$

ADMM 中的 x 和 y-子问题分别可以写成:

$$\begin{aligned} x^{k+1} &\in \arg\min\{\theta_1(x)-x^TA^T\lambda^k+\tfrac{\beta}{2}\|Ax+By^k-b\|^2 \,|\, x\in\mathcal{X}\}\\ &= \arg\min\{\theta_1(x)-x^TA^T\lambda^k+\tfrac{\beta}{2}\|(Ax^k+By^k-b)+A(x-x^k)\|^2 \,|\, x\in\mathcal{X}\}\\ &= \arg\min\left\{\begin{array}{c} \theta_1(x)-x^TA^T[\lambda^k-\beta(Ax^k+By^k-b)]\\ +\frac{\beta}{2}\|A(x-x^k)\|^2 \end{array}\middle|\, x\in\mathcal{X}\right\} \end{aligned}$$

和

$$
\begin{aligned}
y^{k+1} &\in \arg\min\Big\{\theta_2(y) - y^T B^T \lambda^k + \tfrac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,\Big|\, x \in \mathcal{X}\Big\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k \\ +\ \frac{\beta}{2}\|(Ax^k + By^k - b) + [A(x^{k+1} - x^k) + B(y - y^k)]\|^2 \end{array} \ \middle|\ y \in \mathcal{Y} \right\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T [\lambda^k - \beta(Ax^k + By^k - b)] \\ +\ \frac{\beta}{2}\|A(x^{k+1} - x^k) + B(y - y^k)\|^2 \end{array} \ \middle|\ y \in \mathcal{Y} \right\}
\end{aligned}
$$

因此, ADMM 也可以写成

$$
\left\{
\begin{array}{l}
\lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^k + By^k - b) \\[2mm]
x^{k+1} \in \text{argmin}\big\{\theta_1(x) - x^T A^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2}\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\big\}, \\[2mm]
y^{k+1} \in \text{argmin}\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} \\ +\ \frac{\beta}{2}\|A(x^{k+1} - x^k) + B(y - y^k)\|^2 \end{array} \ \middle|\ y \in \mathcal{Y} \right\}, \\[2mm]
\lambda^{k+1} = \lambda^k - \beta\big(Ax^{k+1} + By^{k+1} - b\big)
\end{array}
\right. \tag{1.3}
$$

其中对 $\lambda^{k+\frac{1}{2}}$ 的公式

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^k + By^k - b)$$

也可以写成

$$\lambda^{k+\frac{1}{2}} = P_{\Re^m}\big[\lambda^k - \beta(Ax^k + By^k - b)\big].$$

为了说明我们后面提出的方法和 ADMM 的关系,
我们把经典的 ADMM 改写成等价的 (1.3).

后面我们同时处理等式约束和不等式约束问题. 根据不同类型的问题, 记

$$\Lambda = \Re^m \qquad \text{或者} \qquad \Lambda = \Re^m_+.$$

对给定的向量 $a \in \Re^m$,

$$P_{\Re^m}(a) = a, \qquad [P_{\Re^m_+}(a)]_i = \max\{a_i, 0\}.$$

2 p-块可分离凸优化问题的预测校正方法

p-块可分离凸优化问题

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i)\;\Big|\;\sum_{i=1}^{p}A_i x_i = b \;(\text{or} \geq b),\;\; x_i\in\mathcal{X}_i\Big\}. \tag{2.1}$$

The Lagrangian function is

$$L(x_1,\ldots,x_p,\lambda)=\sum_{i=1}^{p}\theta_i(x_i)-\lambda^T\Big(\sum_{i=1}^{p}A_i x_i-b\Big),$$

which is defined on $\Omega=\prod_{i=1}^{p}\mathcal{X}_i\times\Lambda$, where

$$\Lambda=\begin{cases}\Re^m, & \text{if } \sum_{i=1}^{p}A_i x_i=b,\\[2ex] \Re^m_+, & \text{if } \sum_{i=1}^{p}A_i x_i\geq b.\end{cases}$$

Let $(x_1^*, \ldots, x_p^*, \lambda^*) \in \Omega$ be a saddle point of the Lagrangian function, then

$$L_{\lambda \in \Lambda}(x_1^*, \ldots, x_p^*, \lambda) \le L(x_1^*, \ldots, x_p^*, \lambda^*) \le L_{x_i \in \mathcal{X}_i}(x_1, \ldots, x_p, \lambda^*).$$

The optimality condition of (2.1) can be written as the following VI:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{2.2a}$$

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T \lambda \\ \vdots \\ -A_p^T \lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}, \tag{2.2b}$$

and

$$\theta(x) = \sum_{i=1}^p \theta_i(x_i), \qquad \Omega = \prod_{i=1}^p \mathcal{X}_i \times \Lambda.$$

Again, we denote by Ω^* the solution set of the VI (2.2).

预测 Prediction

从给定的 $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ 到预测点 $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_p^k, \tilde{\lambda}^k)$:

Prediction Step. With given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, find $\tilde{w}^k \in \Omega$:

$$
\left\{
\begin{array}{l}
\tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\
\tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^T A_2^T \lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\
\qquad\vdots \\
\tilde{x}_i^k \in \arg\min_{x_i \in \mathcal{X}_i}\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\
\qquad\vdots \\
\tilde{x}_p^k \in \arg\min_{x_p \in \mathcal{X}_p}\{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}; \\
\tilde{\lambda}^k = P_\Lambda[\lambda^k - \beta(\sum_{j=1}^{p} A_j\tilde{x}_j^k - b)].
\end{array}
\right.
\tag{2.3}
$$

预测先原始再对偶. 对可分离的原始变量子问题逐一按序求解.

校正 Correction

Correction Step .

为下一次迭代提供 $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$:

Generate the new iterate $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$ with $\nu \in (0,1)$ by

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^{k} - A_1\tilde{x}_1^{k} \\ A_2x_2^{k} - A_2\tilde{x}_2^{k} \\ \vdots \\ A_px_p^{k} - A_p\tilde{x}_p^{k} \\ \lambda^{k} - \tilde{\lambda}^{k} \end{pmatrix}. \tag{2.4}$$

校正非常简单, 工作量也很小. 把校正公式分开来写就是:

$Ax_i^{k+1}, i = 1, \ldots, p$

$$\begin{pmatrix} A_1 x_1^{k+1} \\ A_2 x_2^{k+1} \\ \vdots \\ A_p x_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1 x_1^{k} \\ A_2 x_2^{k} \\ \vdots \\ A_p x_p^{k} \end{pmatrix} - \nu \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1 x_1^k - A_1 \tilde{x}_1^k \\ A_2 x_2^k - A_2 \tilde{x}_2^k \\ \vdots \\ A_p x_p^k - A_p \tilde{x}_p^k \end{pmatrix}, \tag{2.5}$$

λ^{k+1}

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - [-\nu\beta(A_1 x_1^k - A_1 \tilde{x}_1^k) + (\lambda^k - \tilde{\lambda}^k)] \\ &= \tilde{\lambda}^k + \nu\beta(A_1 x_1^k - A_1 \tilde{x}_1^k). \end{aligned} \tag{2.6}$$

3 采用 Primal-Dual 预测的预测矩阵

Analysis for the P-D Prediction 我们先看 (2.3) 中 x 子问题

$$\tilde{x}_i^k \in \arg\min\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2 | x_i \in \mathcal{X}_i\}.$$

根据最优性引理, 最优性条件是 $\tilde{x}_i^k \in \mathcal{X}_i$ 和

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{-A_i^T\lambda^k + \beta A_i^T(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k))\} \ge 0, \quad \forall x_i \in \mathcal{X}_i.$$

它可以改写成 $\tilde{x}_i^k \in \mathcal{X}_i$ 和对所有的 $x_i \in \mathcal{X}_i$ 都有

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{\underline{-A_i^T\tilde{\lambda}^k} + \beta A_i^T(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0. \tag{3.1a}$$

预测的对偶部分 $\tilde{\lambda}^k = P_\Lambda[\lambda^k - \beta(\sum_{j=1}^p A_j\tilde{x}_j^k - b)]$, 等价形式

$$\tilde{\lambda}^k = \arg\min\{\|\lambda - [\lambda^k - \beta(\sum_{j=1}^p A_j\tilde{x}_j^k - b)]\|^2 \,|\, \lambda \in \Lambda\}.$$

最优性条件是

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \big\{ \big(\underline{\textstyle\sum_{j=1}^p A_j \tilde{x}_j^k - b}\big) + \tfrac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \big\} \geq 0, \quad \forall \lambda \in \Lambda. \tag{3.1b}$$

Summating (3.1a) and (3.1b), for the predictor $\tilde{w}^k$ generated by (2.3), we have $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \geq (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \tag{3.2a}$$

where

$$Q = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{3.2b}$$

3.1 变量代换下的预测矩阵

The optimization problem (2.1) has been translated to VI (2.2), namely,

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

For the easy analysis, we need to denote the following notations:

$$P = \begin{pmatrix} \sqrt{\beta}A_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta}A_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sqrt{\beta}A_p & 0 \\ 0 & \cdots & \cdots & 0 & (1/\sqrt{\beta})I_m \end{pmatrix}, \quad \xi = Pw = \begin{pmatrix} \sqrt{\beta}A_1x_1 \\ \sqrt{\beta}A_2x_2 \\ \vdots \\ \sqrt{\beta}A_px_p \\ (1/\sqrt{\beta})\lambda \end{pmatrix}. \tag{3.3}$$

Accordingly, we define

$$\Xi = \big\{\xi \mid \xi = Pw, \; w \in \Omega\big\},$$

and

$$\Xi^* = \big\{\xi^* \mid \xi^* = Pw^*, \; w^* \in \Omega^*\big\}.$$

Using the notation P in (3.3), for the matrix Q in (3.2b), we have

$$Q = P^T \mathcal{Q} P, \qquad \text{where} \qquad \mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{3.4}$$

Thus, for the right hand side of (3.2a), we have

$$\begin{aligned} (w - \tilde{w}^k)^T Q (w^k - \tilde{w}^k) &= (w - \tilde{w}^k)^T P^T \mathcal{Q} P (w^k - \tilde{w}^k) \\ &= (\xi - \tilde{\xi}^k)^T \mathcal{Q} (\xi^k - \tilde{\xi}^k). \end{aligned}$$

Then, it follows from (3.2) that we have the following VI for the P-D prediction:

$$\begin{aligned} \tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) &+ (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ &\geq (\xi - \tilde{\xi}^k)^T \mathcal{Q} (\xi^k - \tilde{\xi}^k), \quad \forall\, w \in \Omega. \end{aligned} \tag{3.5}$$

where $\mathcal{Q}$ is given in (3.4).

3.2 变量代换下的算法统一框架

Prediction-Correction Framework for VI (2.2).

1. (Prediction Step) With given w^k and $\xi^k = Pw^k$, find $\tilde{w}^k \in \Omega$ such that

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (\xi - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k), \quad \forall\, w \in \Omega, \quad (3.6a)$$

   with $\mathcal{Q} \in \Re^{(p+1)m\times(p+1)m}$, and the matrix $\mathcal{Q}^T + \mathcal{Q}$ is positive definite.

2. (Correction Step) With the predictor $\tilde{w}^k$ by (3.6a) and $\tilde{\xi}^k = P\tilde{w}^k$, the new iterate ξ^{k+1} is updated by

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k), \quad (3.6b)$$

   where $\mathcal{M} \in \Re^{(p+1)m\times(p+1)m}$ is a non-singular matrix.

Theorem 1 *For the matrices* $\mathcal{Q}$ *and* $\mathcal{M}$ *in the algorithm* (3.6)*, if there is a positive definite matrix* $\mathcal{H} \in \Re^{(p+1)m\times(p+1)m}$ *such that*

$$\mathcal{H}\mathcal{M} = \mathcal{Q} \tag{3.7a}$$

and

$$\mathcal{G} := \mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0, \tag{3.7b}$$

then we have

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2, \quad \forall\, \xi^* \in \Xi^*. \tag{3.8}$$

Proof. Setting w in (3.6a) as any fixed $w^* \in \Omega^*$, and using

$$(\tilde{w}^k - w^*)^T F(\tilde{w}^k) \equiv (\tilde{w}^k - w^*)^T F(w^*),$$

we get

$$(\tilde{\xi}^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*), \;\; \forall\, w^* \in \Omega^*.$$

The right-hand side of the last inequality is non-negative. Thus, we have

$$(\xi^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \geq (\xi^k - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k), \quad \forall\, \xi^* \in \Xi^*. \tag{3.9}$$

Then, by simple manipulations, we obtain

$$\begin{aligned}
&\|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \\
&\overset{(3.6b)}{=} \quad \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|(\xi^k - \xi^*) - \mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&\overset{(3.7a)}{=} \quad 2(\xi^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) - \|\mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&\overset{(3.9)}{\geq} \quad 2(\xi^k - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) - \|\mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&= \quad (\xi^k - \tilde{\xi}^k)^T [(\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T \mathcal{H} \mathcal{M}](\xi^k - \tilde{\xi}^k) \\
&\overset{(3.7b)}{=} \quad \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2.
\end{aligned}$$

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

We call (3.7) the convergence conditions for the algorithm framework (3.6).

The inequality (3.8) is the key for the convergence proofs, for details, see [13]

4 校正方法

For given $\mathcal{Q}$ which satisfies $\mathcal{Q}^T + \mathcal{Q} \succ 0$, we chose $\mathcal{D}$ and $\mathcal{G}$, such that

$$\mathcal{D} \succ 0, \qquad \mathcal{G} \succ 0, \qquad \mathcal{D} + \mathcal{G} = \mathcal{Q}^T + \mathcal{Q}.$$

Then, the correction matrix $\mathcal{M}$ in (3.6b) is given by

$$\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D}.$$

选择了想要的 $0 \prec \mathcal{D}$, 构造 $\mathcal{M}$ 不再神秘！ 下面先介绍以前在 [13] 中 “凑” 出来的 $\mathcal{M}$

First, we give some correction examples which satisfy conditions (3.7) in Theorem 1.

In order to simplify the notations to be used, we define the following $p \times p$ block matrices:

$$\mathcal{L} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix}, \quad \mathcal{I} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{4.1}$$

We also define the $1 \times p$ block matrix

$$\mathcal{E} = \begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix}. \tag{4.2}$$

Using the notations (4.1)-(4.2), the matrix $\mathcal{Q}$ in (3.4) has the form

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} \quad \text{and} \quad \mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix}. \tag{4.3}$$

In order to construct a convergent algorithm, we need only to give the matrices $\mathcal{M}$ and $\mathcal{H}$ and to verify the convergence conditions (3.7)

By setting

$$\mathcal{M} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{4.4}$$

For the above matrices $\mathcal{Q}$ and $\mathcal{M}$, the remaining tasks is to find a positive definite matrix $\mathcal{H}$, such that the convergence conditions (3.7)are satisfied.

(4.4) 中的 $\mathcal{M}$ 是我们在 [13] 中 “凑” 出来的.

How to improvise a correction matrix $\mathcal{M}$? 因为 $\mathcal{H}\mathcal{M} = \mathcal{Q}$,

$$\mathcal{H} = \mathcal{Q}\mathcal{M}^{-1}.$$

有没有一个“块下三角矩阵” $\mathcal{M}$ 满足收敛性条件呢? 因为块下三角矩阵的逆矩阵也是块下三角矩阵, 设 $\mathcal{M}$ 的逆矩阵形式为

$$\mathcal{M}^{-1} = \begin{pmatrix} X & 0 \\ Y & I_m \end{pmatrix}.$$

$\mathcal{H} = \mathcal{Q}\mathcal{M}^{-1}$ 应该是对称矩阵

$$\mathcal{H} = \mathcal{Q}\mathcal{M}^{-1} = \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} \begin{pmatrix} X & 0 \\ Y & I_m \end{pmatrix} = \begin{pmatrix} \mathcal{L}X + \mathcal{E}^T Y & \mathcal{E}^T \\ Y & I_m \end{pmatrix}. \quad (4.5)$$

因此有 $Y = \mathcal{E}$ 和 $X = S^{-1}\mathcal{L}^T$, S 是一个待定的正定矩阵. 所以

$$\mathcal{M}^{-1} = \begin{pmatrix} S^{-1}\mathcal{L}^T & 0 \\ \mathcal{E} & I_m \end{pmatrix} \quad \text{并有} \quad \mathcal{M} = \begin{pmatrix} \mathcal{L}^{-T}S & 0 \\ -\mathcal{E}\mathcal{L}^{-T}S & I_m \end{pmatrix}.$$

继续“凑”下去, 发现 $S = \nu I$ 就可以了, 我们因此也凑出了 $\mathcal{H}$.

$$\mathcal{M}^T\mathcal{H}\mathcal{M} \;=\; \mathcal{Q}^T\mathcal{M} \quad = \quad \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E} & I_m \end{pmatrix} \begin{pmatrix} \mathcal{L}^{-T}S & 0 \\ -\mathcal{E}\mathcal{L}^{-T}S & I_m \end{pmatrix}$$

$$= \quad \begin{pmatrix} S & 0 \\ 0 & I_m \end{pmatrix}.$$

因为

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{L}^T + \mathcal{L} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix} = \begin{pmatrix} \mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix}$$

取 $S = \nu\mathcal{I}$, 就能使 $\mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0$.

以 $Y = \mathcal{E}, X = S^{-1}\mathcal{L}^T$ 和 $S = \nu\mathcal{I}$ 代入(4.5), 就有

$$\mathcal{H} = \begin{pmatrix} \mathcal{L}X + \mathcal{E}^TY & \mathcal{E}^T \\ Y & I_m \end{pmatrix} = \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix}.$$

Lemma 1 *For the matrices* $\mathcal{Q}$ *and* $\mathcal{M}$ *given by* (4.3) *and* (4.4), *respectively, the matrix*

$$\mathcal{H} = \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix} \quad \text{with} \quad \nu \in (0,1) \tag{4.6}$$

is positive definite, and it satisfies $\mathcal{H}\mathcal{M} = \mathcal{Q}$.

Proof. It is easy to check the positive definiteness of $\mathcal{H}$. In addition, for the block matrix $\mathcal{Q}$ in (3.4), we have

$$\begin{aligned} \mathcal{H}\mathcal{M} &= \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix} \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} = \mathcal{Q}. \end{aligned}$$

The assertions of this lemma are proved. □

这样凑出来的 $\mathcal{M}$ 和 $\mathcal{H}$, 能否满足 $\mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0$? 还需要检查一下.

Lemma 2 *Let* $\mathcal{Q}$, $\mathcal{M}$ *and* $\mathcal{H}$ *be defined in* (3.4), (4.4) *and* (4.6), *respectively. Then the* *matrix*

$$\mathcal{G} := (\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T\mathcal{H}\mathcal{M} \tag{4.7}$$

is positive definite.

Proof. By elementary matrix multiplications, we know that

$$\mathcal{M}^T\mathcal{H}\mathcal{M} = \mathcal{Q}^T\mathcal{M} = \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E} & I_m \end{pmatrix}\begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} = \mathcal{D}.$$

Then, it follows from $\mathcal{L}^T + \mathcal{L} = \mathcal{I} + \mathcal{E}^T\mathcal{E}$ (see (4.1)-(4.2)) that

$$\begin{aligned} \mathcal{G} &= (\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T\mathcal{H}\mathcal{M} \\ &= \begin{pmatrix} \mathcal{L}^T + \mathcal{L} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix} - \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix}. \end{aligned}$$

Thus, the matrix $\mathcal{G}$ is positive definite for any $\nu \in (0, 1)$. $\Box$

Finally, correction step can be written

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k). \tag{4.8}$$

Lemma 1 and Lemma 2 have verified the convergence conditions (3.7) and thus the key convergence inequality (3.8) holds. The algorithm (2.3) $\&$ (4.8) is convergent.

Recall the respective definitions $\mathcal{L}$ and $\mathcal{E}$ in (4.1) and (4.2). We have

$$\mathcal{L}^{-T} = \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix}$$

and

$$\mathcal{E}\mathcal{L}^{-T} = \begin{pmatrix} I_m & 0 & \cdots & 0 \end{pmatrix}.$$

Thus

$$\mathcal{M} = \begin{pmatrix} \nu \mathcal{L}^{-T} & 0 \\ -\nu \mathcal{E} \mathcal{L}^{-T} & I_m \end{pmatrix} = \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu I_m & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{4.9}$$

By a manipulation, we have

$$\begin{pmatrix} A_1 x_1^{k+1} \\ A_2 x_2^{k+1} \\ \vdots \\ A_p x_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1 x_1^{k} \\ A_2 x_2^{k} \\ \vdots \\ A_p x_p^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & 0 \\ 0 & \nu I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -\nu I_m \\ 0 & \cdots & 0 & \nu I_m \end{pmatrix} \begin{pmatrix} A_1 x_1^k - A_1 \tilde{x}_1^k \\ A_2 x_2^k - A_2 \tilde{x}_2^k \\ \vdots \\ A_p x_p^k - A_p \tilde{x}_p^k \end{pmatrix}, \tag{4.10}$$

and

$$\lambda^{k+1} = \tilde{\lambda}^k + \nu\beta(A_1 x_1^k - A_1 \tilde{x}_1^k). \tag{4.11}$$

5 More Choices based on the predictions

只要 $\mathcal{Q}^{-T}$ 结构简单, 构造校正矩阵 $\mathcal{M}$ 的方法并不神秘！是非常容易的.

The matrix $\mathcal{Q}$ in (3.4) has the form

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} \quad \text{and thus} \quad \mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix}.$$

To further analyze the correction steps associated with the correction matrix $\mathcal{M}$, let us take a closer look at the matrix $\mathcal{Q}^{-T}$.

According to the primal-dual prediction (2.3), the matrix $\mathcal{Q}$ in (3.4), we have

$$\mathcal{Q}^{-T} = \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E} & I_m \end{pmatrix}^{-1} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{5.1}$$

and

$$\begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix} = \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix} .$$

The calculation $\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D}$ is essentially very easy for different $\mathcal{D}$!

Since

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix},$$

it can be decomposed as

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix}.$$

The both matrices in the right hand side are positive definite. If we chose

$$\mathcal{D} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} \quad \text{and thus} \quad \mathcal{G} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix},$$

it is just the correction in Section §4.

Conversely, we can also choose

$$\mathcal{D} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix} \quad \text{and thus} \quad \mathcal{G} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}$$

and thus get the another correction method.

There are many positive definite decompositions of $\mathcal{Q}^T + \mathcal{Q}$, for example,

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} (1-\nu)\mathcal{I} & 0 \\ 0 & (1-\nu)I_m \end{pmatrix} + \begin{pmatrix} \nu\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & (1+\nu)I_m \end{pmatrix}.$$

and

$$\mathcal{Q}^T + \mathcal{Q} = \mathcal{D} + \mathcal{G} = \alpha(\mathcal{Q}^T + \mathcal{Q}) + (1-\alpha)(\mathcal{Q}^T + \mathcal{Q}), \quad \alpha \in (0,1).$$

6 平行处理子问题的方法

6.1 PPA 方法

我们先设定 (3.6a) 中的预测矩阵 Q 是对称正定的

$$\mathcal{H}=\begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ 0 & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & I_m & I_m \\ I_m & I_m & \cdots & I_m & (p+\delta)I_m \end{pmatrix} \tag{6.1}$$

因为 $\mathcal{H}=H_0\otimes I_m$, 其中

$$H_0=\begin{pmatrix} I_p & e \\ e^T & p+\delta \end{pmatrix} \quad (e \;\text{为}\; p\text{-维全 1 列向量})$$

是正定矩阵. 这可以通过合同变换

$$\begin{pmatrix} I_p & 0 \\ -e^T & 1 \end{pmatrix}\begin{pmatrix} I_p & e \\ e^T & p+\delta \end{pmatrix}\begin{pmatrix} I_p & -e \\ 0 & 1 \end{pmatrix}=\begin{pmatrix} I_p & 0 \\ 0 & \delta \end{pmatrix} \quad (\text{来验证}).$$

由于 (3.6a) 中的预测矩阵是 (6.1) 中对称正定的 $\mathcal{H}$, 结合 (3.3) 中的变换, 我们只要设计下面的预测.

PPA 方法中的预测 Prediction

从给定的 $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ 到预测点 $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_p^k, \tilde{\lambda}^k)$:

Prediction Step. With given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, find $\tilde{w}^k \in \Omega$:

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\
\tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^T A_2^T \lambda^k + \frac{\beta}{2}\|A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\
\qquad \vdots \\
\tilde{x}_i^k \in \arg\min\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|A_i(x_i - x_i^k)\|^2 \mid x_i \in \mathcal{X}_i\}; \\
\qquad \vdots \\
\tilde{x}_p^k \in \arg\min\{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{\beta}{2}\|A_p(x_p - x_p^k)\|^2 \mid x_p \in \mathcal{X}_p\}; \\
\tilde{\lambda}^k = P_\Lambda\Big\{\lambda^k - \frac{1}{p+\delta}\beta\big(\sum_{j=1}^p A_j[2\tilde{x}_j^k - x_j^k] - b\big)\Big\}.
\end{array}\right. \tag{6.2}$$

Analysis for the PPA Prediction 我们先看 (6.2) 中 x 子问题

$$\tilde{x}_i^k \in \arg\min\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|A_i(x_i - x_i^k)\|^2 | x_i \in \mathcal{X}_i\}.$$

根据最优性引理, 最优性条件是 $\tilde{x}_i^k \in \mathcal{X}_i$ 和

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{-A_i^T\lambda^k + \beta A_i^T A_i(\tilde{x}_i^k - x_i^k)\} \ge 0, \quad \forall x_i \in \mathcal{X}_i.$$

它可以改写成 $\tilde{x}_i^k \in \mathcal{X}_i$ 和对所有的 $x_i \in \mathcal{X}_i$ 都有

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{\underline{-A_i^T\tilde{\lambda}^k} + \beta A_i^T A_i(\tilde{x}_i^k - x_i^k) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0. \quad \text{(6.3a)}$$

预测的对偶部分 $\tilde{\lambda}^k = P_\Lambda\Big\{\lambda^k - \frac{1}{p+\delta}\beta\big(\sum_{j=1}^p A_j[2\tilde{x}_j^k - x_j^k] - b\big)\Big\}$, 等价形式

$$\tilde{\lambda}^k = \arg\min\Big\{\Big\|\lambda - \Big[\lambda^k - \frac{1}{p+\delta}\beta\big(\sum_{j=1}^p A_j[2\tilde{x}_j^k - x_j^k] - b\big)\Big]\Big\|^2 \,\Big|\, \lambda \in \Lambda\Big\}.$$

最优性条件是

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T\Big\{\tilde{\lambda}^k - \Big[\lambda^k - \frac{1}{p+\delta}\beta\big(\sum_{j=1}^p A_j[2\tilde{x}_j^k - x_j^k] - b\big)\Big]\Big\} \ge 0, \quad \forall \lambda \in \Lambda.$$

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \Big\{ (\tilde{\lambda}^k - \lambda^k) + \tfrac{1}{p+\delta} \beta \Big(\sum_{j=1}^{p} A_j [2\tilde{x}_j^k - x_j^k] - b \Big) \Big] \Big\} \ge 0, \quad \forall \lambda \in \Lambda.$$

也就是

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \Big\{ \big(\sum_{j=1}^{p} A_j \tilde{x}_j^k - b \big) + \sum_{j=1}^{p} A_j (\tilde{x}_j^k - x_j^k) + \tfrac{p+\delta}{\beta} (\tilde{\lambda}^k - \lambda^k) \Big\} \ge 0, \quad \forall \lambda \in \Lambda. \tag{6.3b}$$

Summating (6.3a) and (6.3b), for the predictor $\tilde{w}^k$ generated by (6.2), we have $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \ge (w - \tilde{w}^k)^T H (w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \tag{6.4a}$$

where

$$H = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ 0 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \beta A_p^T A_p & A_p^T \\ A_1 & A_2 & \cdots & A_p & \frac{p+\delta}{\beta} I_m \end{pmatrix}. \tag{6.4b}$$

利用 (3.3) 中的变换, (6.1) 中的矩阵 $\mathcal{H}$ 和 (6.4b) 中的矩阵 H 满足 $H = P^T \mathcal{H} P$.

6.2 秩二校正的方法

这一小节介绍 [9] 中的秩二校正方法. 设定 (3.6a) 中的预测矩阵 $\mathcal{Q}$ 为

$$\mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ 0 & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & I_m & I_m \\ -I_m & -I_m & \cdots & -I_m & I_m \end{pmatrix} \tag{6.5}$$

因为 $\mathcal{Q} = Q_0 \otimes I_m$, 其中

$$Q_0 = \begin{pmatrix} I_p & e \\ -e^T & 1 \end{pmatrix}, \qquad Q_0^T + Q_0 = \begin{pmatrix} 2I_p & 0 \\ 0 & 2 \end{pmatrix}$$

其中 e 是 p-维全 1 列向量. 这里说的“秩二”, 是指

$$Q_0 = \begin{pmatrix} I_p & e \\ -e^T & 1 \end{pmatrix} = \begin{pmatrix} I_p & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & e \\ -e^T & 0 \end{pmatrix}.$$

中后者是个“秩二”矩阵.

由于 (3.6a) 中的预测矩阵是 (6.5) 中的非对称正定的 Q, 结合 (3.3) 中的变换, 我们只要设计下面的预测.

“秩二”校正方法中的预测 Prediction

从给定的 $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ 到预测点 $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_p^k, \tilde{\lambda}^k)$:

Prediction Step. With given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, find $\tilde{w}^k \in \Omega$:

$$
\left\{
\begin{array}{l}
\tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\
\tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^T A_2^T \lambda^k + \frac{\beta}{2}\|A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\
\qquad \vdots \\
\tilde{x}_i^k \in \arg\min\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|A_i(x_i - x_i^k)\|^2 \mid x_i \in \mathcal{X}_i\}; \\
\qquad \vdots \\
\tilde{x}_p^k \in \arg\min\{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{\beta}{2}\|A_p(x_p - x_p^k)\|^2 \mid x_p \in \mathcal{X}_p\}; \\
\tilde{\lambda}^k = P_\Lambda\{\lambda^k - \beta(\sum_{j=1}^p A_j x_j^k - b)\}.
\end{array}
\right. \tag{6.6}
$$

Analysis for the Prediction (6.6) 我们先看 (6.6) 中 x 子问题

$$\tilde{x}_i^k \in \arg\min\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|A_i(x_i - x_i^k)\|^2 | x_i \in \mathcal{X}_i\}.$$

根据最优性引理, 最优性条件是 $\tilde{x}_i^k \in \mathcal{X}_i$ 和

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{-A_i^T\lambda^k + \beta A_i^T A_i(\tilde{x}_i^k - x_i^k)\} \geq 0, \quad \forall x_i \in \mathcal{X}_i.$$

它可以改写成 $\tilde{x}_i^k \in \mathcal{X}_i$ 和对所有的 $x_i \in \mathcal{X}_i$ 都有

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{\underline{-A_i^T\tilde{\lambda}^k} + \beta A_i^T A_i(\tilde{x}_i^k - x_i^k) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0. \quad \text{(6.7a)}$$

预测的对偶部分 $\tilde{\lambda}^k = P_\Lambda\Big\{\lambda^k - \beta\big(\sum_{j=1}^p A_j x_j^k - b\big)\Big\}$, 等价形式

$$\tilde{\lambda}^k = \arg\min\Big\{\Big\|\lambda - \Big[\lambda^k - \beta\big(\sum_{j=1}^p A_j x_j^k - b\big)\Big]\Big\|^2 \,|\, \lambda \in \Lambda\Big\}.$$

最优性条件是

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T\Big\{\tilde{\lambda}^k - \Big[\lambda^k - \beta\big(\sum_{j=1}^p A_j x_j^k - b\big)\Big]\Big\} \geq 0, \quad \forall \lambda \in \Lambda.$$

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \Big\{ (\tilde{\lambda}^k - \lambda^k) + \beta\big(\sum_{j=1}^{p} A_j x_j^k - b\big)\Big] \Big\} \ge 0, \quad \forall \lambda \in \Lambda.$$

也就是

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \Big\{ \big(\sum_{j=1}^{p} A_j \tilde{x}_j^k - b\big) - \sum_{j=1}^{p} A_j(\tilde{x}_j^k - x_j^k) + \tfrac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \Big\} \ge 0, \quad \forall \lambda \in \Lambda. \tag{6.7b}$$

Summating (6.7a) and (6.7b), for the predictor $\tilde{w}^k$ generated by (6.6), we have $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \ge (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \tag{6.8a}$$

where

$$Q = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ 0 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \beta A_p^T A_p & A_p^T \\ -A_1 & -A_2 & \cdots & -A_p & \frac{1}{\beta} I_m \end{pmatrix}. \tag{6.8b}$$

采用 (3.3) 中的变换

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\xi - \tilde{\xi}^k)^T Q(\xi^k - \tilde{\xi}^k), \;\; \forall\, w \in \Omega, \tag{6.9a}$$

where

$$\mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ 0 & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & I_m & I_m \\ -I_m & -I_m & \cdots & -I_m & I_m \end{pmatrix}. \tag{6.9b}$$

注意到

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} 2\mathcal{I} & 0 \\ 0 & 2I_m \end{pmatrix}, \quad \text{我们可以取} \quad \mathcal{D} = \alpha \begin{pmatrix} \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}, \; \alpha \in (0, 2).$$

由于 $\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D}$, 校正公式

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k) \tag{6.10}$$

也变成了

$$\xi_\alpha^{k+1} = \xi^k - \alpha\mathcal{Q}^{-T}(\xi^k - \tilde{\xi}^k). \tag{6.11}$$

先讨论一下校正公式中 $\mathcal{Q}^{-T}$ 具体形式是什么.

Lemma 3 对 (6.9b) 中定义的矩阵 $\mathcal{Q}$, 我们有

$$\mathcal{Q}^{-T} = \begin{pmatrix} I_m & 0 & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix} - \frac{1}{p+1}\begin{pmatrix} I_m & I_m & \cdots & I_m & -I_m \\ I_m & I_m & \cdots & I_m & -I_m \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ I_m & I_m & \cdots & I_m & -I_m \\ I_m & I_m & \cdots & I_m & pI_m \end{pmatrix}. \tag{6.12}$$

证明 注意到 $\mathcal{Q}$ 是一个单位矩阵与斜对称 (Skew-symmetric) 矩阵的和, 并且有

$$\mathcal{Q} = \mathcal{Q}_0 \otimes I_m,$$

其中

$$\mathcal{Q}_0 = \begin{pmatrix} 1 & & & 1 \\ & \ddots & & \vdots \\ & & 1 & 1 \\ -1 & \cdots & -1 & 1 \end{pmatrix}_{(p+1)\times(p+1)} = \begin{pmatrix} I_p & e \\ -e^T & 1 \end{pmatrix},$$

$e \in \Re^p$ 是p-维全1向量, $\otimes$ 表示 Kronecker 积. 注意到 $\mathcal{Q}_0$ 是单位矩阵与一个秩二矩阵的和.

$$\mathcal{Q}_0^T = \begin{pmatrix} I_p & -e \\ e^T & 1 \end{pmatrix} = \begin{pmatrix} I_p & \\ & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ \vdots & \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & \cdots & 0 & -1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}.$$

设

$$U = \begin{pmatrix} e & 0 \\ 0 & 1 \end{pmatrix}, \quad V = \begin{pmatrix} 0 & e \\ -1 & 0 \end{pmatrix}, \quad \text{则} \quad \mathcal{Q}_0^T = I_{p+1} + UV^T.$$

用线性代数中的 Sherman-Morrison-Woodbury 公式,

$$\begin{aligned}
\mathcal{Q}_0^{-T} &= (I_{p+1} + UV^T)^{-1} \\
&= I_{p+1} - U(I_2 + V^TU)^{-1}V^T \\
&= I_{p+1} - \begin{pmatrix} e_p & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ p & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -1 \\ e_p^T & 0 \end{pmatrix} \\
&= \begin{pmatrix} I_p & \\ & 1 \end{pmatrix} - \frac{1}{p+1} \begin{pmatrix} e_p e_p^T & -e_p \\ e_p^T & p \end{pmatrix}.
\end{aligned} \tag{6.13}$$

利用 Kronecker 积的基本性质, 我们有

$$\mathcal{Q}^{-T} = (\mathcal{Q}_0 \otimes I_m)^{-T} = \mathcal{Q}_0^{-T} \otimes I_m^{-T} = \mathcal{Q}_0^{-T} \otimes I_m,$$

这与 (6.12) 中表示的一致. 引理的证. □

由于 $\mathcal{Q}_0$ 是单位矩阵与一个秩二矩阵的和, 校正公式 (6.11) 中的

$$\mathcal{Q}^{-T} = \mathcal{Q}_0^{-T} \otimes I_m.$$

所以我们把方法叫做带广义秩二校正的方法.

利用 ξ (见(3.3)) 和 $\mathcal{Q}^{-T}$ (见(6.12)) 的表达式, 校正公式 (6.11) 的具体形式是

$$\begin{pmatrix} \sqrt{\beta}A_1x_1^{k+1} \\ \sqrt{\beta}A_2x_2^{k+1} \\ \vdots \\ \sqrt{\beta}A_px_p^{k+1} \\ \frac{1}{\sqrt{\beta}}\lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \sqrt{\beta}A_1x_1^{k} \\ \sqrt{\beta}A_2x_2^{k} \\ \vdots \\ \sqrt{\beta}A_px_p^{k} \\ \frac{1}{\sqrt{\beta}}\lambda^{k} \end{pmatrix} - \alpha \begin{pmatrix} \sqrt{\beta}(A_1x_1^k - A_1\tilde{x}_1^k) \\ \sqrt{\beta}(A_2x_2^k - A_2\tilde{x}_2^k) \\ \vdots \\ \sqrt{\beta}(A_px_p^k - A_p\tilde{x}_p^k) \\ \frac{1}{\sqrt{\beta}}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}$$

$$+\frac{\alpha}{p+1}\begin{pmatrix} I_m & I_m & \cdots & I_m & -I_m \\ I_m & I_m & \cdots & I_m & -I_m \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ I_m & I_m & \cdots & I_m & -I_m \\ I_m & I_m & \cdots & I_m & pI_m \end{pmatrix} \begin{pmatrix} \sqrt{\beta}(A_1x_1^k - A_1\tilde{x}_1^k) \\ \sqrt{\beta}(A_2x_2^k - A_2\tilde{x}_2^k) \\ \vdots \\ \sqrt{\beta}(A_px_p^k - A_p\tilde{x}_p^k) \\ \frac{1}{\sqrt{\beta}}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}.$$

利用记号

$$\mathcal{A} = (A_1, A_2, \ldots, A_p),$$

它的等价形式可以写成

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k+1} \end{pmatrix} - \alpha \begin{pmatrix} A_1x_1^{k} - A_1\tilde{x}_1^k \\ A_2x_2^{k} - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^{k} - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}$$

$$+ \frac{\alpha}{p+1} \begin{pmatrix} \mathcal{A}(x^k - \tilde{x}^k) - \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \\ \mathcal{A}(x^k - \tilde{x}^k) - \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \\ \vdots \\ \mathcal{A}(x^k - \tilde{x}^k) - \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \\ \beta\mathcal{A}(x^k - \tilde{x}^k) + p(\lambda^k - \tilde{\lambda}^k) \end{pmatrix} \quad (6.14)$$

从 (6.14) 公式中可以得到

$$\begin{aligned} A_i x_i^{k+1} &= A_i x_i^k - \alpha(A_i x_i^k - A_i \tilde{x}_i^k) + \frac{\alpha}{p+1}\Big\{\mathcal{A}(x^k - \tilde{x}^k) - \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k)\Big\} \\ &= A_i x_i^k - \alpha(A_i x_i^k - A_i \tilde{x}_i^k) + \frac{\alpha}{p+1}\Big\{\mathcal{A}(x^k - \tilde{x}^k) - (\mathcal{A}x^k - b)\Big\} \\ &= A_i x_i^k - \alpha(A_i x_i^k - A_i \tilde{x}_i^k) - \frac{\alpha}{p+1}(\mathcal{A}\tilde{x}^k - b), \end{aligned}$$

和

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) + \frac{\alpha}{p+1}\Big\{\beta\mathcal{A}(x^k - \tilde{x}^k) + p(\lambda^k - \tilde{\lambda}^k)\Big\} \\ &= \lambda^k - \alpha\beta(\mathcal{A}x^k - b) + \frac{\alpha}{p+1}\Big\{\beta\mathcal{A}(x^k - \tilde{x}^k) + p\beta(\mathcal{A}x^k - b)\Big\} \\ &= \lambda^k - \frac{\alpha\beta}{p+1}(\mathcal{A}\tilde{x}^k - b). \end{aligned}$$

因此, 校正公式可以简单地写成

$$\left\{ \begin{aligned} A_i x_i^{k+1} &= A_i x_i^k - \alpha(A_i x_i^k - A_i \tilde{x}_i^k) - \frac{\alpha}{p+1}(\mathcal{A}\tilde{x}^k - b), \ i = 1, \ldots, p, \\ \lambda^{k+1} &= \lambda^k - \frac{\alpha\beta}{p+1}(\mathcal{A}\tilde{x}^k - b). \end{aligned} \right. \tag{6.15}$$

7 Conclusions

- 我的学术报告中常用的一个题目是“构造凸优化的分裂收缩算法– 用好 VI 和 PPA 两大法宝”, 是指构造变分不等式意义下的 PPA 算法, 文章首先发表在 [15]. 后来又做了一些人为地将预测矩阵设计成对称正定矩阵的方法 [2, 3], 包括我们 2021 年才提出的均困平衡的增广拉格朗日乘子法 [18]. 有时我们也称这样的方法为按需定制的 PPA - (Customized PPA).
- 对预测矩阵 Q 为非对称的预测-校正方法, 利用统一框架的套路证明收敛性, 最初出现在我和袁晓明 (Xiaoming Yuan) 2012 年 SIAM 数值分析的文章 [14] 中, 后面我们发表的一些论文 [1, 8, 10, 11, 17], 都用这个套路证明收敛性. 把它归结为统一框架, 是在南京大学讨论班上, 那是在我 2013 年即将退休之前, 以后便常常出现在我的“讲习班”讲义和报告的 PPT 中.
- 第一次在正式出版物里提到这个统一框架, 是在 2016 年《高校计算数学学报》的我的中文文章 [4] 中. 2018 年我在《运筹学学报》的综述文章 “我和乘子交替方向法 20 年”[5] 中指出, 我们发表的方法都可以用这个框架非常简单地证明收敛性. 英文出版物中首次出现统一框架的是我和袁晓明 2018 年在 COAP 的文章 [16].
- 从 2018 年开始, 我在自己的报告和论文 [7] 中, 经常讲用统一框架去构造算法主要还是按收敛条件去“凑”. 如何根据确定的预测矩阵 $\boldsymbol{Q}$ 凑出满足收敛条件的校正矩阵 $\boldsymbol{M}$. 似乎给人一种难以效仿的神秘感觉.

- 2022年初我在南师大做报告时有人问过. 最近我又在中科大和南航做线上报告, 教学相长, 得到一些新的看法, 觉得有必要将回答整理成下面的材料与听众共享.

- 我们从预测矩阵满足 $Q^T + Q \succ 0$ 出发. 根据条件 $HM = Q$, 我们有

$$H = QM^{-1}.$$

因为 H 是正定矩阵, 必须对称. 从上式又看到, H 有个左因子 Q, 那它必须有个右因子 Q^T, 中间夹一个"待定的"正定矩阵. 我们设这个正定矩阵为 D^{-1}, 则有

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

比较上面两式, 我们得到 $M^{-1} = D^{-1}Q^T$, 因此

$$\boldsymbol{M = Q^{-T}D}.$$

这个我们大概在10年前就知道. 当时往往考虑选择的 D 应该是个块对角矩阵.

- 至此, 我们还不知道矩阵 D 具体形式是什么. 计算一下收敛性条件中的 M^THM,

$$M^THM = (DQ^{-1})(QD^{-1}Q^T)(Q^{-T}D) = D.$$

上式已经出现在我2018的暑期讲习班的讲义中, 没有向前再迈一步.

- 利用上式和 $G = Q^T + Q - M^THM \succ 0$, 这个待定的正定矩阵 D 只需要满足

$$\boldsymbol{0 \prec D \prec Q^T + Q} \qquad （\text{因此}, 0 \prec G = Q^T + Q - D）$$

就可以了. 明确这一条, 得益于为 2022 年以来在 南师大, 南航 和 中科大 讲课, 迫使我深入思考, 把方法讲明白.

- 在选了满足上述条件的矩阵 D 以后, 根据确定的 Q 和 D, 找未知矩阵 H 和 M 使得

$$HM = Q \qquad \text{和} \qquad M^T HM = D,$$

我们的目的就达到了.

- 这样的 M 和 H: 可以通过求解下面的矩阵方程组得到.

$$\begin{cases} HM &= Q, \\ M^T HM &= D. \end{cases} \Leftrightarrow \begin{cases} HM &= Q, \\ Q^T M &= D. \end{cases} \Leftrightarrow \begin{cases} H &= QD^{-1}Q^T, \\ M &= Q^{-T}D. \end{cases} .$$

- 选择不同的满足条件的矩阵 D (这非常容易), 就有不同的校正方法. 譬如说,

$$D = \alpha[Q^T + Q], \quad \alpha \in (0, 1).$$

- 报告的第 2 节开始, 对一般线性约束凸优化问题, 采用 primal-dual 预测, 子问题的求解方式是 ADMM 类型的逐个向前. 我们需要的 $\boldsymbol{Q^{-T}}$ 形式非常简单. 是的, 它需要额外的校正. 可喜的是, 校正花费很少, 又特别容易实现!

- 我们特别推崇“预测-校正”, 尤其是那种代价很小的校正. 生机勃勃的果树, 修剪就是校正. 社会治理也是一种校正, 当然也考虑成本! 交替按序预测, 降低了问题难度; 全局整体校正, 把握了收敛方向.

- 预测-校正方法既可以用来求解等式约束的问题, 又可以用来求解不等式约束的问题. 适用从一块到任意多块的可分离问题, 算法结构和收敛性证明完全统一.

- 适用范围广的算法会不会影响效率？对经典 ADMM 擅长的两块可分离的等式约束凸优化问题, 我们用 2- 节本文提到的带校正的交替方向法去求解, 与网上他人提供的 ADMM 代码比较, 发现这种担心是多余的.

- **Question A.** In the prediction step, how to arrange a “good” prediction matrix whose matrix $\mathcal{Q}$ satisfies

$$\mathcal{Q}^T + \mathcal{Q} \succeq \mathcal{I}.$$

- **Question B** For the given prediction matrix $\mathcal{Q}$, what are the criteria for choosing matrix $\mathcal{D}$ which satisfies

$$0 \prec \mathcal{D} \prec \mathcal{Q}^T + \mathcal{Q}.$$

希望各位以质疑的态度审视我的观点, 对的就相信, 不对的请批评指正.

References

[1] 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, Series A* 2016.

[2] 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.

[3] G. Y. Gu, B. S. He and X. M. Yuan, Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach, Comput. Optim. Appl., 59 (2014), 135-161.

[4] B. S. He, From the projection and contraction methods for variatioonal inequality to the splittingcontraction methods fro convex optimization (in Chinese), Numerical Mathematics, A Journal of Chinese Universities, 38 (2016), 74-96.

[5] B. S. He, My 20 years research on alternating directions method of multipliers (in Chinese), Operations Research Trasactions, 22 (2018), 1-31.

[6] B. S. He, A uniform framework of contraction methods for convex optimization and monotone variational inequality (in Chinese). Sci Sin Math, 48 (2018), 255-272, doi: 10.1360/N012017-00034

[7] B. S. He, Using a unified framework to design the splitting and contraction methods for convex optimization (in Chinese), Numerical Mathematics, A Journal of Chinese Universities, 44 (2022), 1-35.

[8] B. S. He, H. Liu, Z. R. Wang and X. M. Yuan, A strictly Peaceman-Rachford splitting method for convex programming, SIAM J. Optim. 24 (2014), 1011-1040.

[9] B.S. He, F. Ma, S. J. Xu and X. M. Yuan, A rank-two relaxed parallel splitting version of the augmented Lagrangian method with step size in (0, 2) for separable convex programming, arXiv:2205.02723[math.OC].

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

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

[13] B. S. He, S. J. Xu and X. M. Yuan, Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints, arXiv:2107.01897v2[math.OC].

[14] 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.

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

[16] B. S. He and X. M. Yuan, A class of ADMM-based algorithms for three-block separable convex programming, Comput. Optim. Appl. 70 (2018), 791-826.

[17] 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.

[18] B. S. He and X. M. Yuan, Balanced Augmented Lagrangian Method for Convex Programming, arXiv 2108.08554 [math.OC]

[19] B. S. He and X. M. Yuan, On construction of splitting contraction algorithms in a prediction-correction framework for separable convex optimization, arXiv 2204.11522 [math.OC]