变分不等式框架下结构型凸优化的分裂收缩算法

VI. 多个可分离块凸优化问题的ADMM类秩一校正 & 秩二校正方法

中学的数理基础　必要的社会实践
普通的大学数学　一般的优化原理

何 炳 生　　南京大学数学系

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

天元数学东北中心　2023 年 10 月 17 – 27 日

1 p-块可分离凸优化问题的变分不等式

p-块可分离凸优化问题

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i)\;\Big|\;\sum_{i=1}^{p}A_ix_i=b\ (\text{or}\ \geq b),\quad x_i\in\mathcal{X}_i\Big\}. \tag{1.1}$$

The Lagrangian function is

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

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

$$\Lambda=\begin{cases}\Re^m, & \text{if } \sum_{i=1}^{p}A_ix_i=b,\\ \Re^m_+, & \text{if } \sum_{i=1}^{p}A_ix_i\geq b.\end{cases}$$

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 (1.1) can be written as the following VI:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{1.2a}$$

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T \lambda \\ \vdots \\ -A_p^T \lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}, \tag{1.2b}$$

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 (1.2).

2 从交替方向法得到的启示

Let us consider the general separable convex optimization model

$$\min\left\{\theta_1(x)+\theta_2(y)\mid Ax+By=b, x\in\mathcal{X}, y\in\mathcal{Y}\right\}. \tag{2.1}$$

The augmented Lagrangian function is

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

Applied the classical ADMM [1, 2] to the problem (2.1):

ADMM for (2.1) From (y^k,λ^k) to (y^{k+1},λ^{k+1})

$$\left\{\begin{array}{lcl} x^{k+1} & \in & \arg\min\{\mathcal{L}_\beta(x,y^k,\lambda^k)\,|\,x\in\mathcal{X}\}, \\ y^{k+1} & \in & \arg\min\{\mathcal{L}_\beta(x^{k+1},y,\lambda^k)\,|\,y\in\mathcal{Y}\}, \\ \lambda^{k+1} & = & \lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). \end{array}\right. \tag{2.2}$$

Ignoring some constant terms in the objective functions of the corresponding subproblems, we can rewrite the ADMM (2.2) as

$$\left\{\begin{array}{rcl} x^{k+1} & \in & \text{argmin}\{\theta_1(x) - x^TA^T\lambda^k + \frac{\beta}{2}\|Ax + By^k - b)\|^2 \mid x\in\mathcal{X}\} \\ & = & \text{argmin}\left\{ \begin{array}{c} \theta_1(x) - x^TA^T\lambda^k \\ +\frac{\beta}{2}\|A(x-x^k) + (Ax^k + By^k - b)\|^2 \end{array} \middle| \, x\in\mathcal{X} \right\} \\ & = & \text{argmin}\left\{ \begin{array}{c} \theta_1(x) - x^TA^T[\lambda^k - \beta(Ax^k + By^k - b)] \\ +\frac{\beta}{2}\|A(x-x^k)\|^2 \end{array} \middle| \, x\in\mathcal{X} \right\} \\ y^{k+1} & \in & \text{argmin}\{\theta_2(y) - y^TB^T\lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By - b)\|^2 \mid y\in\mathcal{Y}\} \\ & = & \text{argmin}\left\{ \begin{array}{c} \theta_2(y) - y^TB^T\lambda^k \\ +\frac{\beta}{2}\|B(y-y^k) + (Ax^{k+1} + By^k - b)\|^2 \end{array} \middle| \, y\in\mathcal{Y} \right\} \\ & = & \text{argmin}\left\{ \begin{array}{c} \theta_2(y) - y^TB^T[\lambda^k - \beta(Ax^k + By^k - b)] \\ +\frac{\beta}{2}\|A(x^{k+1}-x^k) + B(y-y^k)\|^2 \end{array} \middle| \, y\in\mathcal{Y} \right\} \\ \lambda^{k+1} & = & \lambda^k - \beta\big(Ax^{k+1} + By^{k+1} - b\big) \end{array}\right. \tag{2.3}$$

如果记

$$\lambda^{k+\frac{1}{2}} := \lambda^k - \beta(Ax^k + By^k - b). \tag{2.4a}$$

ADMM 迭代式 (2.3) 就可以写成

$$\left\{\begin{array}{rcl} x^{k+1} & \in & \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b)\|^2 \mid x \in \mathcal{X}\} \\ & = & \text{argmin}\left\{\begin{array}{c} \theta_1(x) - x^T A^T \lambda^{k+\frac{1}{2}} \\ +\frac{\beta}{2}\|A(x - x^k)\|^2 \end{array} \middle| \, x \in \mathcal{X}\right\} \\ y^{k+1} & \in & \text{argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By - b)\|^2 \mid y \in \mathcal{Y}\} \\ & = & \text{argmin}\left\{\begin{array}{c} \theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} \\ +\frac{\beta}{2}\|A(x^{k+1} - x^k) + B(y - y^k)\|^2 \end{array} \middle| \, y \in \mathcal{Y}\right\} \\ \lambda^{k+1} & = & \lambda^k - \beta\big(Ax^{k+1} + By^{k+1} - b\big) \\ & = & \lambda^k - \beta\big(Ax^k + By^k - b\big) + \beta A(x^k - x^{k+1}) + \beta B(y^k - y^{k+1}) \end{array}\right. \tag{2.4b}$$

假如记 $\tilde{\lambda}^k = \lambda^{k+\frac{1}{2}}, \quad \tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1},$ 则有 **预测**

$$\left\{ \begin{array}{lcl} \tilde{\lambda}^k & = & \lambda^k - \beta(Ax^k + By^k - b) \\ \tilde{x}^k & \in & \operatorname{argmin}\left\{ \theta_1(x) - x^T A^T \tilde{\lambda}^k + \frac{\beta}{2}\|A(x - x^k)\|^2 \;\middle|\; x \in \mathcal{X} \right\} \\ \tilde{y}^k & \in & \operatorname{argmin}\left\{ \begin{array}{l} \theta_2(y) - y^T B^T \tilde{\lambda}^k \\ + \frac{\beta}{2}\|A(\tilde{x}^k - x^k) + B(y - y^k)\|^2 \end{array} \;\middle|\; y \in \mathcal{Y} \right\} \end{array} \right. \tag{2.5}$$

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

$$\begin{array}{rcl} \lambda^{k+1} & = & \lambda^k - \beta\big(Ax^k + By^k - b\big) + \beta A(x^k - x^{k+1}) + \beta B(y^k - y^{k+1}) \\ & = & \lambda^k - [(\lambda^k - \tilde{\lambda}^k) - \beta A(x^k - \tilde{x}^k) - \beta B(y^k - \tilde{y}^k) \end{array}$$

校正

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta A & -\beta B & I_m \end{pmatrix} \begin{pmatrix} x^k - \tilde{x}^k \\ y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}$$

2.1 ADMM with wider applications

Let us consider the general two-block separable convex optimization model

$$\min\big\{\theta_1(x)+\theta_2(y)\mid Ax+By=b\ (\text{or} \geq b), x\in\mathcal{X}, y\in\mathcal{Y}\big\}. \tag{2.6}$$

The linear constraints can be a system of linear equations or linear inequalities.

We define

$$\Lambda=\begin{cases}\Re^m, & \text{if } Ax+By=b,\\ \Re^m_+, & \text{if } Ax+By\geq b,\end{cases}$$

and denote the projection on Λ by $P_\Lambda[\cdot]$. For such special Λ, the projection on Λ is clear !

The only difference： $P_{\Re^m}(\lambda)=\lambda, \quad P_{\Re^m_+}(\lambda)=\max\{\lambda,0\}.$

A Dual-Primal Extension of the ADMM for (2.6).

From (Ax^k, By^k, λ^k) to $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$: Find $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via

$$\left\{ \begin{array}{l} \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(Ax^k + By^k - b\big)\big], \\ \tilde{x}^k \in \text{argmin}\big\{\theta_1(x) - x^T A^T \tilde{\lambda}^k + \frac{1}{2}\beta\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\big\}, \\ \tilde{y}^k \in \text{argmin}\big\{\theta_2(y) - y^T B^T \tilde{\lambda}^k + \frac{1}{2}\beta\|A(\tilde{x}^k - x^k) + B(y - y^k)\|^2 \mid y \in \mathcal{Y}\big\}. \end{array} \right. \tag{2.7}$$

预测先做 Primal 部分, 再做 Dual 部分, 顺序也可以倒过来.

A Primal-Dual Extension of the ADMM for (2.6).

From (Ax^k, By^k, λ^k) to $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$: Find $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via

$$\left\{ \begin{array}{l} \tilde{x}^k \in \text{argmin}\big\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\big\}, \\ \tilde{y}^k \in \text{argmin}\big\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|A(\tilde{x}^k - x^k) + B(y - y^k)\|^2 \mid y \in \mathcal{Y}\big\}, \\ \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(A\tilde{x}^k + B\tilde{y}^k - b\big)\big]. \end{array} \right. \tag{2.8}$$

无论是 dual-primal, 还是 primal-dual 方法, 都可以向多块问题直接推广.

多块问题 (1.2) 的 DUAL-PRIMAL 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 $\tilde{w}^k \in \Omega$:

$$\left\{\begin{array}{l}
\tilde{\lambda}^k = P_{\Lambda}\big[\lambda^k - \beta\big(\sum_{j=1}^p A_jx_j^k - b\big)\big] \\
\tilde{x}_1^k \in \arg\min\big\{\theta_1(x_1) - x_1^TA_1^T\tilde{\lambda}^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\big\}; \\
\tilde{x}_2^k \in \arg\min\big\{\theta_2(x_2) - x_2^TA_2^T\tilde{\lambda}^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\big\}; \\
\qquad\qquad \vdots \\
\tilde{x}_i^k \in \arg\min_{x_i\in\mathcal{X}_i}\{\theta_i(x_i) - x_i^TA_i^T\tilde{\lambda}^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1}A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\
\qquad\qquad \vdots \\
\tilde{x}_p^k \in \mathrm{argmin}_{x_p\in\mathcal{X}_p}\{\theta_p(x_p) - x_p^TA_p^T\tilde{\lambda}^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1}A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}.
\end{array}\right. \tag{2.9}$$

预测先对偶再原始. 对可分离的原始变量子问题逐一按序求解.

多块问题 (1.2) 的 PRIMAL-DUAL 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 $\tilde{w}^k \in \Omega$:

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\
\tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^T A_2^T \lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\
\qquad\qquad \vdots \\
\tilde{x}_i^k \in \arg\min_{x_i\in\mathcal{X}_i}\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\
\qquad\qquad \vdots \\
\tilde{x}_p^k \in \arg\min_{x_p\in\mathcal{X}_p}\{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}; \\
\tilde{\lambda}^k = P_\Lambda[\lambda^k - \beta(\sum_{j=1}^p A_j\tilde{x}_j^k - b)].
\end{array}\right. \tag{2.10}$$

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.

3 采用 Primal-Dual 预测的预测矩阵

Analysis for the P-D Prediction 我们先看 (2.10) 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

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

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 (2.10), 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 (1.1) has been translated to VI (1.2), 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:

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

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 (1.2).

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

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

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

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

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

定理 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

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

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

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

Then, by simple manipulations, we obtain

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

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 [5]

4 基于 Primal-Dual 预测的校正方法

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 [5] 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

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

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

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

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

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

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 [5] 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

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

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

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

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

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

因为

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

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

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

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

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

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

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

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

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.

引理 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

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

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

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

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

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 (2.10) & (4.8) is convergent.

Recall the respective definitions $\mathcal{L}$ and $\mathcal{E}^T$ 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

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

Thus

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

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

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

校正非常简单, 工作量也很小. 把校正公式分开来写就是:

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

$$\begin{pmatrix} A_1 x_1^{k+1} \\ A_2 x_2^{k+1} \\ \vdots \\ A_p x_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1 x_1^{k} \\ A_2 x_2^{k} \\ \vdots \\ A_p x_p^{k} \end{pmatrix} - \nu \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1 x_1^k - A_1 \tilde{x}_1^k \\ A_2 x_2^k - A_2 \tilde{x}_2^k \\ \vdots \\ A_p x_p^k - A_p \tilde{x}_p^k \end{pmatrix}, \tag{4.12}$$

$\boldsymbol{\lambda^{k+1}}$

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - [-\nu\beta(A_1 x_1^k - A_1 \tilde{x}_1^k) + (\lambda^k - \tilde{\lambda}^k)] \\ &= \tilde{\lambda}^k + \nu\beta(A_1 x_1^k - A_1 \tilde{x}_1^k). \end{aligned} \tag{4.13}$$

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

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

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 (2.10), the matrix $\mathcal{Q}$ in (3.4), we have

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

and

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

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

Since

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

it can be decomposed as

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

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

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

it is just the correction in Section §4.

Conversely, we can also choose

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

and thus get the another correction method.

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

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

and

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

对基于 Dual-Primal 预测的方法, 建议读者自己去构造校正矩阵 $\mathcal{M}$.

这一讲求解多块可分离线性约束的凸优化问题的方法, 仍然是变分不等式框架下的预测-校正方法. 采用 Primal-Dual 和 Dual-Primal 的 Gauss 型预测, 分别得到预测矩阵

$$
Q_{PD} = \begin{pmatrix}
\beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\
\beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\
\vdots & \ddots & \ddots & 0 & \vdots \\
\beta A_p^T A_1 & \cdots & \beta A_p^T A_{p-1} & \beta A_p^T A_p & A_p^T \\
0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m
\end{pmatrix}
$$

和

$$
Q_{DP} = \begin{pmatrix}
\beta A_1^T A_1 & 0 & \cdots & 0 & 0 \\
\beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & 0 \\
\vdots & \ddots & \ddots & 0 & \vdots \\
\beta A_p^T A_1 & \cdots & \beta A_p^T A_{p-1} & \beta A_p^T A_p & 0 \\
-A_1 & \cdots & -A_{p-1} & -A_p & \frac{1}{\beta} I_m
\end{pmatrix}.
$$

利用 (3.3) 做了替换以后, 得到特殊结构的预测矩阵 $\mathcal{Q}$, 它们分别是

$$\mathcal{Q}_{PD} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & \ddots & \ddots & 0 & \vdots \\ I_m & \cdots & I_m & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}, \quad \mathcal{Q}_{DP} = \begin{pmatrix} I_m & 0 & \cdots & 0 & 0 \\ I_m & I_m & \ddots & \vdots & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ I_m & \cdots & I_m & I_m & 0 \\ -I_m & \cdots & -I_m & -I_m & I_m \end{pmatrix}.$$

而 $\mathcal{Q}^T + \mathcal{Q}$ 的形式分别为

$$\begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & 2I_m \end{pmatrix} \quad \text{和} \quad \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & -\mathcal{E} \\ -\mathcal{E}^T & 2I_m \end{pmatrix}.$$

其逆矩阵 $\mathcal{Q}^{-T}$ 的形式分别是

$$\mathcal{Q}_{PD}^{-T}=\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0\\ 0 & I_m & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & -I_m & 0\\ 0 & \cdots & 0 & I_m & 0\\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix},\quad \mathcal{Q}_{DP}^{-T}=\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0\\ 0 & I_m & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & -I_m & 0\\ 0 & \cdots & 0 & I_m & I_m\\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}.$$

选取满足条件

$$\mathcal{D}+\mathcal{G}=\mathcal{Q}^T+\mathcal{Q}$$

的正定矩阵 $\mathcal{D}$ 和 $\mathcal{G}$, 策略是很多的. 然后, 令

$$\mathcal{M}=\mathcal{Q}^{-T}\mathcal{D}\quad \text{和}\quad \xi^{k+1}=\xi^k-\mathcal{M}(\xi^k-\tilde{\xi}^k),$$

则有

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

其中 $\mathcal{H}=\mathcal{Q}\mathcal{D}^{-1}\mathcal{Q}^T$. 由于 $\mathcal{Q}^{-T}$ 的结构相当简单, 校正是容易实现的.

6 为什么说是秩一校正

这一讲讨论的方法, 由串行预测生成的矩阵 $\mathcal{Q}_{PD}$ 和 $\mathcal{Q}_{DP}$, 都是一个容易求逆的矩阵和一个广义秩一矩阵的和. 譬如说,

$$\mathcal{Q}_{PD}^T = \mathcal{Q}_{0PD}^T \otimes I_m, \tag{6.1}$$

其中

$$\mathcal{Q}_{0PD}^T = \begin{pmatrix} 1 & 1 & \cdots & 1 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 1 & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}, \qquad \otimes \text{表示 Kronecker 积.}$$

把 $\mathcal{Q}_{0PD}^T$ 中的 1 改成 I_m, 就得到了 $\mathcal{Q}_{PD}^T$. 注意到

$$\mathcal{Q}_{0PD}^T = Q_{1PD}^T + Q_{2PD}^T,$$

其中

$$Q_{1PD}^T = \begin{pmatrix} 1 & 1 & \cdots & 1 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 1 & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \qquad Q_{2PD}^T = \begin{pmatrix} 0 & \cdots & \cdots & 0 & 0 \\ \vdots & & & \vdots & \vdots \\ \vdots & & & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 \\ 1 & \cdots & \cdots & 1 & 0 \end{pmatrix}.$$

由于 Q_{1PD} 容易求逆, Q_{2PD} 是秩一矩阵

$$Q_{1PD}^{-T} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 & 0 \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix}, \qquad Q_{2PD}^T = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & \cdots & \cdots & 1 & 0 \end{pmatrix}.$$

利用线性代数中的秩一校正求逆公式

$$(A+uv^T)^{-1} = A^{-1} - \frac{1}{1+v^TA^{-1}u}A^{-1}uv^TA^{-1},$$

容易得到

$$\mathcal{Q}_{0PD}^{-T} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 & 0 \\ 0 & \cdots & 0 & 1 & 0 \\ -1 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

由 (6.1) 和

$$\mathcal{Q}_{PD}^{-T} = \mathcal{Q}_{0PD}^{-T} \otimes I_m,$$

我们得到

$$\mathcal{Q}_{PD}^{-T} = \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}.$$

秩一校正是这一章用到的关键技术.

这一讲的方法主要取材于尚未正式发表的 arXiv 上的文章 [9].

7 根据统一框架设计用秩二校正的预测方法

我们仍然考虑线性约束的多块可分离凸优化问题. 其相应的变分不等式在前一讲也已经做了介绍. 采用统一框架中的方法求解变分不等式. 这些方法的第 k-步迭代从给定的 $(A_1x_1^k,\dots,A_px_p^k,\lambda^k)$ 出发, 生成预测点 $\tilde{w}^k\in\Omega$, 满足

$$\theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(w-\tilde{w}^k)^TQ(w^k-\tilde{w}^k),\quad\forall\, w\in\Omega, \tag{7.1}$$

其中 Q^T+Q 是本质上正定的. 这一讲前面的方法采用的是串行预测, 然后用广义秩一校正. 这一节介绍的方法, 我们仍然采用的是串行预测, 然后进行广义秩二校正.

利用前一讲定义的变换, 可以把预测 (7.1) 改写成

$$\tilde{w}^k\in\Omega,\quad\theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(\xi-\tilde{\xi}^k)^T\mathcal{Q}(\xi^k-\tilde{\xi}^k),\quad\forall\, w\in\Omega, \tag{7.2}$$

其中

$$Q=P^T\mathcal{Q}P,\qquad\text{并且}\qquad\mathcal{Q}^T+\mathcal{Q}\succ 0. \tag{7.3}$$

由 (7.2) 得到

$$(\tilde{\xi}^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \geq 0, \quad \forall \xi \in \Xi^*. \tag{7.4}$$

接着, 我们就可以选择正定矩阵 $\mathcal{D}$ 和 $\mathcal{G}$, 使得

$$\mathcal{D} \succ 0, \quad \mathcal{G} \succ 0, \quad \mathcal{D} + \mathcal{G} = \mathcal{Q}^T + \mathcal{Q}. \tag{7.5}$$

最后, 用

$$\xi^{k+1} = \xi^k - \mathcal{Q}^{-T}\mathcal{D}(\xi^k - \tilde{\xi}^k) \tag{7.6}$$

得到 ξ^{k+1}. 算法产生的序列 $\{\xi^k\}$ 满足收敛性质

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \leq \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2, \quad \forall \xi^* \in \Xi^*. \tag{7.7}$$

讨论根据统一框架构造算法, 实际上就是预先设定矩阵 $\mathcal{Q}$, 使得

1. $\mathcal{Q}^T + \mathcal{Q} \succ 0$.
2. 对 $Q = P^T \mathcal{Q} P$ 的预测矩阵 Q, 相应的预测 (7.1) 可以实施.
3. $\mathcal{Q}^{-T}$ (或者 $\mathcal{Q}^{-1}$) 的表达式简单, 使得校正 (7.6) 容易实现.

求解多块可分离凸优化问题, 预测按串行逐渐向前推进, 如果将矩阵 $\mathcal{Q}$ 写成 2×2 的分块形式, 其左上角是下三角矩阵 $\mathcal{L}$.

7.1 Primal-Dual 预测后再秩二校正的方法

设计一个可以执行 Primal-Dual 预测的矩阵

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ \mathcal{E}^T & \frac{5}{2} I_m \end{pmatrix}. \tag{7.8}$$

由于

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & 2\mathcal{E} \\ 2\mathcal{E}^T & 5I_m \end{pmatrix} = \begin{pmatrix} \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} \mathcal{E} \\ 2I_m \end{pmatrix} \big(\mathcal{E}^T, \ 2I_m\big), \tag{7.9}$$

$\mathcal{Q}^T + \mathcal{Q}$ 是单位矩阵与一个半正定矩阵的和, 所以是正定的. 注意到如果将 (7.8) 中 $\mathcal{Q}$ 矩阵左上角的 $\mathcal{L}$ 换成 $\mathcal{I}$, $\mathcal{Q}$ 就成了对称矩阵, 但对 $p \geq 3$, 这样的矩阵就不再是正定的. 利用变换前一讲的记号, 对应于 (7.8) 中的 $\mathcal{Q}$, 相应

的 $Q = P^T \mathcal{Q} P$, 所以

$$Q = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\ A_1 & A_2 & \cdots & A_p & \frac{5}{2\beta} I_m \end{pmatrix}. \tag{7.10}$$

要实现预测

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \tag{7.11}$$

其中矩阵 Q 由 (7.10) 给出. 根据多块问题变分不等式的形式, 预测 (7.11) 的原始和对偶部分可以分别通过

$$\left\{\begin{array}{l} \tilde{x}_1^k \in \arg\min\Big\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{1}{2}\beta\|A_1(x_1 - x_1^k)\|^2 \;\Big|\; x_1 \in \mathcal{X}_1\Big\}; \\ \qquad\qquad\vdots \\ \tilde{x}_i^k \in \arg\min\left\{\begin{array}{c} \theta_i(x_i) - x_i^T A_i^T \lambda^k + \\ \frac{1}{2}\beta\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2 \end{array}\middle| x_i \in \mathcal{X}_i\right\}; \\ \qquad\qquad\vdots \\ \tilde{x}_p^k \in \arg\min\left\{\begin{array}{c} \theta_p(x_p) - x_p^T A_p^T \lambda^k + \\ \frac{1}{2}\beta\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2 \end{array}\middle| x_p \in \mathcal{X}_p\right\} \end{array}\right. \tag{7.12a}$$

和

$$\tilde{\lambda}^k = P_\Lambda\Big\{\lambda^k - \frac{2}{5}\beta\Big[\big(\sum_{i=1}^p A_i\tilde{x}_i^k - b\big) + \sum_{i=1}^p A_i(\tilde{x}_i^k - x_i^k)\Big]\Big\} \tag{7.12b}$$

完成. 根据最优性定理, (7.12a) 中 x_i-子问题的最优性条件是

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

这可以写成

$$\begin{aligned}\tilde{x}_i^k\in\mathcal{X}_i, \ \ \theta_i(x_i)-\theta_i(\tilde{x}_i^k)+(x_i-\tilde{x}_i^k)^T\{-A_i^T\tilde{\lambda}^k\}\\ \ge (x_i-\tilde{x}_i^k)^T\{A_i^T\beta[\sum_{j=1}^{i}A_j(x_j^k-\tilde{x}_j^k)]+A_i^T(\lambda^k-\tilde{\lambda}^k)\}, \ \ \forall x_i\in\mathcal{X}_i.\end{aligned} \tag{7.13a}$$

对偶预测 (7.12b) 的最优性条件是 $\tilde{\lambda}^k\in\Lambda$,

$$(\lambda-\tilde{\lambda}^k)^T\Big\{\tilde{\lambda}^k-\Big[\lambda^k-\frac{2}{5}\beta\big[\big(\sum_{i=1}^{p}A_i\tilde{x}_i^k-b\big)+\sum_{i=1}^{p}A_i(\tilde{x}_i^k-x_i^k)\big]\Big]\Big\}\ge 0, \ \ \forall\,\lambda\in\Lambda.$$

这可以改写成等价的 $\tilde{\lambda}^k \in \Lambda$,

$$(\lambda-\tilde{\lambda}^k)^T\Big\{\Big[(\sum_{i=1}^p A_i\tilde{x}_i^k - b) + \sum_{i=1}^p A_i(\tilde{x}_i^k - x_i^k)\Big] + \frac{5}{2\beta}(\tilde{\lambda}^k - \lambda^k)\Big\} \geq 0, \quad \forall\, \lambda \in \Lambda$$

并进一步有

$$\begin{aligned} \tilde{\lambda}^k \in \Lambda, \quad (\lambda-\tilde{\lambda}^k)^T\{\sum_{i=1}^p A_i\tilde{x}_i^k - b\} \\ \geq (\lambda - \tilde{\lambda}^k)^T\Big\{\sum_{i=1}^p A_i(x_i^k - \tilde{x}_i^k) + \frac{5}{2\beta}(\lambda^k - \tilde{\lambda}^k)\Big\}, \quad \forall\, \lambda \in \Lambda. \end{aligned} \tag{7.13b}$$

把 (7.13a) 和 (7.13b) 放在一起, 就是预测 (7.11), 其中的矩阵 Q 由 (7.10) 给出. 得到了满足 (7.11) 的 $\tilde{w}^k$, 也得到了相应的 $\tilde{\xi}^k = P\tilde{w}^k$.

还需要关心的是, 对 (7.8) 中的 $\mathcal{Q}$, $\mathcal{Q}^{-T}$ 的形式是否简单. 对这里的 $\mathcal{Q}$, 为了防止混淆, 我们记其为 $\mathcal{Q}_{PD}$, 有

$$\mathcal{Q}_{PD}^T = \mathcal{Q}_1^T + \mathcal{Q}_2^T,$$

其中

$$\mathcal{Q}_1^T = \begin{pmatrix} \mathcal{L}^T & 0 \\ 0 & \frac{5}{2} I_m \end{pmatrix}, \qquad \mathcal{Q}_2^T = \begin{pmatrix} 0 & \mathcal{E} \\ \mathcal{E}^T & 0 \end{pmatrix}.$$

$\mathcal{Q}_1$ 是个容易求逆的矩阵, 而

$$\mathcal{Q}_2^T = \begin{pmatrix} I_m & 0 \\ \vdots & \vdots \\ I_m & 0 \\ 0 & I_m \end{pmatrix} \begin{pmatrix} 0 & \cdots & 0 & I_m \\ I_m & \cdots & I_m & 0 \end{pmatrix} = \begin{pmatrix} \mathcal{E} & 0 \\ 0 & I_m \end{pmatrix} \begin{pmatrix} 0 & I_m \\ \mathcal{E}^T & 0 \end{pmatrix}$$

是个广义秩二矩阵. 利用线性代数中的求逆公式

$$(A + UV)^{-1} = A^{-1} - A^{-1}U(I + VA^{-1}U)^{-1}VA^{-1},$$

经过演算可得

$$\mathcal{Q}_{PD}^{-T} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ 0 & \frac{2}{5} I_m \end{pmatrix} + \frac{2}{3} \begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T} & -\mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^T\mathcal{L}^{-T} & \frac{2}{5} I_m \end{pmatrix}.$$

上式也可以写成

$$\mathcal{Q}_{PD}^{-T} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ 0 & 0 \end{pmatrix} + \frac{2}{3} \begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T} & -\mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{7.14}$$

由于

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

和

$$\mathcal{E}^T\mathcal{L}^{-T} = \big(I_m,\, I_m,\, \ldots,\, I_m\big)\begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} = \big(I_m,\, 0,\, \ldots,\, 0\big),$$

我们有

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

所以, (7.14) 中的 $\mathcal{Q}_{PD}^{-T}$ 形式是相当简单的. 写开来就是

$$\mathcal{Q}_{PD}^{-T} = \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ 0 & \cdots & 0 & 0 & 0 \end{pmatrix} + \frac{2}{3}\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ I_m & 0 & \cdots & 0 & -I_m \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix},$$

校正容易实现. 由 (7.9) 知道

$$\mathcal{Q}_{PD}^{T} + \mathcal{Q}_{PD} = \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & 2\mathcal{E} \\ 2\mathcal{E}^T & 5I_m \end{pmatrix} = \begin{pmatrix} \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} \mathcal{E} \\ 2I_m \end{pmatrix} \left(\mathcal{E}^T,\ 2I_m\right).$$

选择符合条件 (7.5) 的矩阵 $\mathcal{D}$ 有许多选法. 例如, 若取

$$\mathcal{D} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}, \quad \nu \in (0,1),$$

条件 (7.5) 满足. 由

$$\xi^{k+1} = \xi^k - \mathcal{Q}_{PD}^{-T}\mathcal{D}(\xi^k - \tilde{\xi}^k)$$

产生的迭代序列 $\{\xi^k\}$ 满足关键收缩不等式 (7.7). 与上式等价的校正公式是

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

$$- \frac{2}{3}\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ \nu I_m & 0 & \cdots & 0 & -\frac{1}{\beta}I_m \\ -\nu\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{7.15}$$

7.2 Dual-Primal 预测后再校正的方法

同样, 可以设计一个 Dual-Primal 的预测矩阵

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L}^T & -\mathcal{E} \\ -\mathcal{E}^T & \frac{5}{2} I_m \end{pmatrix}. \tag{7.16}$$

其中的 $\mathcal{L}$, $\mathcal{E}$ 如前一讲给出. 由于

$$\begin{aligned} \mathcal{Q}^T + \mathcal{Q} &= \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & -2\mathcal{E} \\ -2\mathcal{E}^T & 5I_m \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} \mathcal{E} \\ -2I_m \end{pmatrix} \big(\mathcal{E}^T, \ -2I_m\big). \end{aligned} \tag{7.17}$$

$\mathcal{Q}^T + \mathcal{Q}$ 是单位矩阵与一个半正定矩阵的和, 所以是正定的. 同样, 如果将 (7.16) 中的 $\mathcal{Q}$ 矩阵左上角的 $\mathcal{L}$ 换成 $\mathcal{I}$, $\mathcal{Q}$ 就成了对称矩阵, 但对 $p \geq 3$, 这样的矩阵就不再是正定的. 利用相应变换的记号, 对应于 (7.16) 的 $\mathcal{Q}$, 相应

的 $Q = P^T \mathcal{Q} P$, 所以

$$Q = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & -A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & -A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & -A_p^T \\ -A_1 & -A_2 & \cdots & -A_p & \frac{5}{2\beta} I_m \end{pmatrix}, \tag{7.18}$$

要实现预测

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \tag{7.19}$$

其中矩阵 Q 由 (7.18) 给出. 根据多块问题变分不等式的形式, (7.19) 的最后一行是

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \Big\{ \big(\sum_{i=1}^p A_i \tilde{x}_i^k - b \big) \\ \ge (\lambda - \tilde{\lambda}^k)^T \Big\{ -\sum_{i=1}^p A_i (x_i^k - \tilde{x}_i^k) + \tfrac{5}{2\beta} (\lambda^k - \tilde{\lambda}^k) \Big\}, \quad \forall \lambda \in \Lambda.$$

也就是

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \Big\{ \big(\sum_{i=1}^p A_i x_i^k - b \big) + \tfrac{5}{2\beta} (\tilde{\lambda}^k - \lambda^k) \Big\} \ge 0, \quad \forall \lambda \in \Lambda.$$

上面的 $\tilde{\lambda}^k$ 可以通过

$$\tilde{\lambda}^k = P_\Lambda \big[\lambda^k - \tfrac{2}{5} \beta \big(\sum_{i=1}^p A_i x_i^k - b \big) \big] \tag{7.20a}$$

得到. 有了对偶变量的预测, 串行迭代的 x_i-子问题需要满足的最优性条件是

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

其中第一个 $-A_i^T \tilde{\lambda}^k$ 对应的是 $F(\tilde{w}^k)$ 中相应的那部分. 上式归并以后得到

$$\tilde{x}_i^k \in \mathcal{X}_i, \;\; \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \Big\{ -A_i^T (2\tilde{\lambda}^k - \lambda^k) \\ + A_i^T \beta \big[\sum_{j=1}^{i} A_j (\tilde{x}_j^k - x_j^k) \big] \Big\} \ge 0, \;\; \forall x_i \in \mathcal{X}_i.$$

根据最优性条件的定理, 它是优化问题

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

的最优性条件. 因此, 原始变量 x 的预测是

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\left\{\begin{array}{c}\theta_1(x_1) - x_1^T A_1^T (2\tilde{\lambda}^k - \lambda^k) \\ +\frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2\end{array} \middle| \, x_1 \in \mathcal{X}_1\right\}; \\
\qquad\qquad\qquad \vdots \\
\tilde{x}_i^k \in \arg\min\left\{\begin{array}{c}\theta_i(x_i) - x_i^T A_i^T (2\tilde{\lambda}^k - \lambda^k) \\ +\frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\end{array} \middle| \, x_i \in \mathcal{X}_i\right\}; \\
\qquad \vdots \\
\tilde{x}_p^k \in \arg\min\left\{\begin{array}{c}\theta_p(x_p) - x_p^T A_p^T (2\tilde{\lambda}^k - \lambda^k) \\ +\frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\end{array} \middle| \, x_p \in \mathcal{X}_p\right\}.
\end{array}\right. \tag{7.20b}$$

这样, 我们就得到了满足 (7.4) 的 $\tilde{w}^k$, 也得到了相应的 $\tilde{\xi}^k = P\tilde{w}^k$. 同样需要关心的是, 对 (7.16) 中的 $\mathcal{Q}$, $\mathcal{Q}^{-T}$ 的形式是否简单. 对这里的 $\mathcal{Q}$, 为了防止混淆,

我们记其为 $\mathcal{Q}_{DP}$, 又因为

$$\mathcal{Q}_{DP}{}^{T} = \mathcal{Q}_1^T + \mathcal{Q}_2^T,$$

其中

$$\mathcal{Q}_1^T = \begin{pmatrix} \mathcal{L}^T & 0 \\ 0 & \frac{5}{2} I_m \end{pmatrix}, \qquad \mathcal{Q}_2 = \begin{pmatrix} 0 & -\mathcal{E} \\ -\mathcal{E}^T & 0 \end{pmatrix}.$$

$\mathcal{Q}_1^T$ 是个容易求逆的矩阵, 而

$$\mathcal{Q}_2^T = \begin{pmatrix} I_m & 0 \\ \vdots & \vdots \\ I_m & 0 \\ 0 & -I_m \end{pmatrix} \begin{pmatrix} 0 & \cdots & 0 & -I_m \\ I_m & \cdots & I_m & 0 \end{pmatrix} = \begin{pmatrix} \mathcal{E} & 0 \\ 0 & -I_m \end{pmatrix} \begin{pmatrix} 0 & -I_m \\ \mathcal{E}^T & 0 \end{pmatrix}$$

是个广义秩二矩阵. 利用 $\mathcal{Q}_1^T$ 求 $\mathcal{Q}^T$ 是个秩二校正的过程. 经过简单演算可得

$$\mathcal{Q}_{DP}^{-T} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ 0 & 0 \end{pmatrix} + \frac{2}{3} \begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T} & \mathcal{L}^{-T}\mathcal{E} \\ \mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{7.21}$$

读者将上式和 (7.14) 做比较, 就能得到 (7.21) 中的 $\mathcal{Q}^{-T}$ 的具体形式

$$\mathcal{Q}_{DP}^{-T} = \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ 0 & \cdots & 0 & 0 & 0 \end{pmatrix} + \frac{2}{3} \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ I_m & 0 & \cdots & 0 & I_m \\ I_m & 0 & \cdots & 0 & I_m \end{pmatrix},$$

校正容易实现. 由于

$$\mathcal{Q}_{DP}^{T} + \mathcal{Q}_{DP} = \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & -2\mathcal{E} \\ -2\mathcal{E}^T & 5I_m \end{pmatrix} = \begin{pmatrix} \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} \mathcal{E} \\ -2I_m \end{pmatrix} \begin{pmatrix} \mathcal{E}^T & -2I_m \end{pmatrix}.$$

同样, 若取

$$\mathcal{D} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}, \quad \nu \in (0,1),$$

条件 (7.5) 满足. 由

$$\xi^{k+1} = \xi^k - \mathcal{Q}_{DP}^{-T}\mathcal{D}(\xi^k - \tilde{\xi}^k)$$

产生的迭代序列 $\{\xi^k\}$ 满足关键收缩不等式 (7.7). 与上式等价的校正公式是

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

$$- \frac{2}{3} \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ \nu I_m & 0 & \cdots & 0 & \frac{1}{\beta} I_m \\ \nu\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \quad (7.22)$$

8 从秩一校正到秩二校正的预测方法

这一讲, 我们讨论了多块可分离问题的求解方法, 包括等式线性约束问题和不等式约束问题, 建立了求解方法的统一的框架 [5, 9].

- 秩一校正的方法, 起源于对 ADMM 的等价改写, 得到了渐进式预测, 然后进一步扩展到多块的可分离问题的求解上. 无论是 Primal-Dual 还是 Dual -Primal 预测, 生成的预测矩阵都是一个容易求逆的矩阵和一个广义秩一矩阵的和.

- §7 则是根据设定的预测矩阵, 再去构造预测方法. 其中的预测矩阵都是一个容易求逆的矩阵和一个广义秩二矩阵的和. 这个迭代预测的变分不等式可以分解, 然后通过求解相应的分裂后简单的凸优化问题去实现.

- 有了满足 $\mathcal{Q}^T + \mathcal{Q} \succ 0$ 的预测矩阵 $\mathcal{Q}$, 校正的方法是千变万化的. 只要选择

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

  采用

  $$\xi^{k+1} = \xi^k - \mathcal{Q}^{-T}\mathcal{D}(\xi^k - \tilde{\xi}^k)$$

  就能实现, 其中 ξ 和 w 的关系是由变换 (3.3) 确定的.

- 这里介绍的满足统一框架的算法, 都有 [6, 8] 中提到的类似的收敛性质.

References

[1] R. Glowinski and A. Marrocco, Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problémes non linéaires. RAIRO Anal. Numer. R2, 41–76 (1975)

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

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

[4] B. S. He, Using a unified framework to design the splitting and contraction methods for convex optimization (in Chinese), 高等学校计算数学学报 44 (2022), 1-35.

[5] 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].

[6] 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.

[7] 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.

[8] 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.

[9] 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]