变分不等式框架下结构型凸优化的分裂收缩算法

VII. 由预测决定的广义 PPA 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 广义邻近点算法

我们介绍了变分不等式和邻近点算法的概念. 讨论了基于合格预测构造单位步长校正方法的策略, 证明了迭代序列满足收缩不等式

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall\, v^* \in \mathcal{V}^*, \tag{1.1}$$

其中 H 和 G 分别称为范数矩阵和效益矩阵. 当 $H = G$ 时, 这犹如 PPA 算法的结论. 因此, 我们把预测-校正方法的收缩不等式 (1.1) 中 $H = G$ 的方法称为广义 PPA 算法.

前面我们介绍的凸优化的分裂收缩算法基本上都在变分不等式的邻近点算法(PPA)和可分离凸优化的交替方向法(ADMM)的基础上发展起来的. 我们回顾一下这些算法的主要共有性质.

1. 变分不等式 PPA 算法的主要性质

我们对变分不等式问题

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

定义了 PPA 算法, 设 H 为对称正定矩阵, H-模下的 PPA 算法的第 k 步从已知的 w^k 出发, 求得的新迭代点 w^{k+1} 使得

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T F(w^{k+1}) \\ &\geq \; (w - w^{k+1})^T H(w^k - w^{k+1}), \;\; \forall\, w \in \Omega. \end{aligned} \tag{1.3}$$

w^{k+1} 是变分不等式问题 (1.2) 的解的充分必要条件是 (1.3) 中的 $w^k = w^{k+1}$. PPA 算法产生的迭代序列 $\{w^k\}$ 满足

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2, \quad \forall w^* \in \Omega^*, \tag{1.4}$$

并有

$$\|w^k - w^{k+1}\|_H^2 \leq \|w^{k-1} - w^k\|_H^2. \tag{1.5}$$

不等式 (1.4) 和 (1.5) 是 PPA 算法的两条重要而又漂亮的性质.

2. ADMM 算法的主要性质

把两块可分离凸优化问题

$$\min\{\theta_1(x) + \theta_2(y) | Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\} \tag{1.6}$$

转换成变分不等式(1.2), 其中

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

$$F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b \end{pmatrix}, \qquad \text{和} \qquad \Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m.$$

ADMM 的 k 次迭代从给定的 $v^k=(y^k,\lambda^k)$ 开始, 通过

$$\left\{\begin{array}{lcl} x^{k+1} & \in & \arg\min\big\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}\beta\|Ax+By^k-b\|^2 \;\big|\; x\in\mathcal{X}\big\},\\ y^{k+1} & \in & \arg\min\big\{\theta_2(y)-y^TB^T\lambda^k+\frac{1}{2}\beta\|Ax^{k+1}+By-b\|^2 \;\big|\; y\in\mathcal{Y}\big\},\\ \lambda^{k+1} & = & \lambda^k-\beta(Ax^{k+1}+By^{k+1}-b) \end{array}\right. \tag{1.7}$$

求得 $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})$. 这个方法中的核心变量是 $v=(y,\lambda)$.

在第三讲中我们证明了收缩性质

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*, \tag{1.8}$$

其中

$$H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}.$$

除此之外, 我们在 [16] 中证明了 ADMM 的迭代序列 $\{v^k\}$ 具备性质

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2. \tag{1.9}$$

不等式 (1.8) 和 (1.9) 展示了 ADMM 很好的性质. 在一些快速 ADMM 的研究 [2] 中, 都用到了 (1.9) 这条性质.

2 预测-校正的广义 PPA 算法

求解变分不等式 (1.2) 采用单位步长校正的时候, 如果预测公式

$$\tilde{w}^k \in \Omega, \quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k), \quad \forall\, w \in \Omega, \tag{2.1}$$

中的预测矩阵 Q 满足 $Q^T+Q \succ 0$, 若将 Q^T+Q 分拆成

$$D \succ 0, \qquad G \succ 0 \quad \text{和} \quad D+G=Q^T+Q, \tag{2.2}$$

再令

$$M=Q^{-T}D \qquad \text{和} \qquad H=QD^{-1}Q^T. \tag{2.3}$$

则由单位步长校正

$$v^{k+1}=v^k-M(v^k-\tilde{v}^k) \tag{2.4}$$

产生的新的迭代序列 $\{v^k\}$ 满足

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

如果我们采用一对特殊的 D 和 G, 使得

$$D = G = \tfrac{1}{2}(Q^T + Q),$$

那么, (2.5) 就变成了

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_D^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.6}$$

对选定的 D, 根据 (2.3), 总有

$$M^T H M = D,$$

因此, (2.6) 就成了

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

再利用 $M(v^k - \tilde{v}^k) = v^k - v^{k+1}$ (见 (2.4)), 上式就了

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.7}$$

此外, 关于统一框架中所有固定步长的方法都证明了 (见本系列讲座第四讲定理4) 都证明了

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2. \tag{2.8}$$

我们把上述分析结果写成下面的定理.

定理 1 用预测校正方法求解变分不等式 (1.2), 设预测 (2.1) 中的预测矩阵 Q 满足 $Q^T + Q \succ 0$. 若令

$$D = \tfrac{1}{2}(Q^T + Q), \qquad \text{和} \qquad M = Q^{-T}D$$

则由单位步长校正公式

$$v^{k+1} = v^k - Q^{-T}D(v^k - \tilde{v}^k) \tag{2.9}$$

产生的新的迭代点具有性质 (2.7) 和 (2.8), 其中

$$H = Q[\tfrac{1}{2}(Q^T + Q)]^{-1}Q^T.$$

求解变分不等式 (1.2), 我们把迭代序列具有性质 (2.7) 和 (2.8) 的方法, 称为广义 PPA 方法. 在实际计算中, 我们并不要求显式写出 H 的表达式,

3 变量替换下的广义 PPA 算法

仍然考虑线性约束的多块可分离凸优化问题. 这些方法的第 k-步迭代从给定

的 $(A_1x_1^k,\ldots,A_px_p^k,\lambda^k)$ 出发, 生成预测点 $\tilde{w}^k$ 满足

$$\theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(w-\tilde{w}^k)^TQ(w^k-\tilde{w}^k),\quad\forall\, w\in\Omega. \tag{3.1}$$

作为合格的预测, 其中的矩阵 Q^T+Q 往往只是本质上正定的. 利用上一讲的变换, 把预测 (3.1) 改写成 $\tilde{w}^k\in\Omega$,

$$\theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(\xi-\tilde{\xi}^k)^T\mathcal{Q}(\xi^k-\tilde{\xi}^k),\quad\forall\, w\in\Omega, \tag{3.2}$$

其中 $Q=P^T\mathcal{Q}P$,

$$\mathcal{Q}^T+\mathcal{Q}\succ 0 \tag{3.3}$$

是正定矩阵. 在 $\mathcal{Q}$ 非对称但 (3.3) 满足的时候, 必须采用必要的校正. 我们总可以选两个矩阵 $\mathcal{D}$ 和 $\mathcal{G}$, 使得

$$\mathcal{D}\succ 0,\quad \mathcal{G}\succ 0,\quad \text{和}\quad \mathcal{D}+\mathcal{G}=\mathcal{Q}^T+\mathcal{Q}. \tag{3.4}$$

根据前一讲的分析, 我们有如下的定理.

定理 2 设预测点 $\tilde{\xi}^k$ 满足条件 (3.2), 其中 $\mathcal{Q}^T+\mathcal{Q}$ 是正定矩阵. 如果由两个正定矩阵 $\mathcal{D}$ 和 $\mathcal{G}$, 使得 (3.4) 成立.

$$\mathcal{M}=\mathcal{Q}^{-T}\mathcal{D} \tag{3.5}$$

那么, 利用矩阵 (3.5) 校正

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

产生的 ξ^{k+1} 满足

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2, \quad \forall \xi^* \in \Xi^*, \tag{3.7}$$

其中矩阵 $\mathcal{H} = \mathcal{Q}\mathcal{D}^{-1}\mathcal{Q}^T$.

如果选

$$\mathcal{D} = \mathcal{G} = \frac{1}{2}(\mathcal{Q}^T + \mathcal{Q}) \tag{3.8}$$

那么, (3.7) 就变成了

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

对选定的 $\mathcal{D}$, 根据 $\mathcal{D} = \mathcal{M}^T\mathcal{H}\mathcal{M}$, 并利用 (3.6), 上式就成了

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \xi^{k+1}\|_{\mathcal{H}}^2, \quad \forall \xi^* \in \Xi^*. \tag{3.9}$$

下面证明收敛性的另一条重要性质: 序列 $\{\|\xi^k - \xi^{k+1}\|_{\mathcal{H}}\}$ 是单调不增的.

定理 3 如果预测点 $\tilde{\xi}^k$ 满足条件 (3.2), 那么, 由校正 (3.6) 产生的新的迭代

点 ξ^{k+1} 满足

$$\|\xi^{k+1} - \xi^{k+2}\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^{k+1}\|_{\mathcal{H}}^2. \tag{3.10}$$

证明. 首先, 我们证明迭代序列满足

$$\begin{aligned} &(\xi^k - \xi^{k+1})^T \mathcal{H}[(\xi^k - \xi^{k+1}) - (\xi^{k+1} - \xi^{k+2})] \\ \ge\quad & \frac{1}{2}\|(\xi^k - \tilde{\xi}^k) - (\xi^{k+1} - \tilde{\xi}^{k+1})\|_{(\mathcal{Q}^T+\mathcal{Q})}^2. \end{aligned} \tag{3.11}$$

将预测 (3.2) 中的 k 改为 $k+1$, 我们有

$$\theta(x) - \theta(\tilde{x}^{k+1}) + (w - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \ge (\xi - \tilde{\xi}^{k+1})^T \mathcal{Q}(\xi^{k+1} - \tilde{\xi}^{k+1}), \quad \forall\, w \in \Omega,$$

将上式中任意的 w 设为 $\tilde{w}^k$, 得到

$$\theta(\tilde{x}^k) - \theta(\tilde{x}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \ge (\tilde{\xi}^k - \tilde{\xi}^{k+1})^T \mathcal{Q}(\xi^{k+1} - \tilde{\xi}^{k+1}). \tag{3.12}$$

将预测 (3.2) 式中任意的 w 设为 $\tilde{w}^{k+1}$, 就有

$$\theta(\tilde{x}^{k+1}) - \theta(\tilde{x}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \ge (\tilde{\xi}^{k+1} - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k). \tag{3.13}$$

将 (3.12), (3.13) 加在一起, 利用 $(\tilde{w}^k - \tilde{w}^{k+1})^T (F(\tilde{w}^k) - F(\tilde{w}^{k+1})) \equiv 0$, 得到

$$(\tilde{\xi}^k - \tilde{\xi}^{k+1})^T \mathcal{Q}\{(\xi^k - \tilde{\xi}^k) - (\xi^{k+1} - \tilde{\xi}^{k+1})\} \geq 0.$$

对上式两边加上

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

并利用 $\xi^T \mathcal{Q}\xi = \frac{1}{2}\xi^T(\mathcal{Q}^T + \mathcal{Q})\xi$, 我们得到

$$\begin{array}{rl} & (\xi^k - \xi^{k+1})^T \mathcal{Q}\{(\xi^k - \tilde{\xi}^k) - (\xi^{k+1} - \tilde{\xi}^{k+1})\} \\ \geq & \dfrac{1}{2}\|(\xi^k - \tilde{\xi}^k) - (\xi^{k+1} - \tilde{\xi}^{k+1})\|^2_{(\mathcal{Q}^T + \mathcal{Q})}. \end{array}$$

在上式左端利用 $\mathcal{Q} = \mathcal{H}\mathcal{M}$ 和校正公式 (3.6), 就得到 (3.11).

下面, 我们在恒等式 $\|a\|^2_{\mathcal{H}} - \|b\|^2_{\mathcal{H}} = 2a^T\mathcal{H}(a-b) - \|a-b\|^2_{\mathcal{H}}$ 中置 $a = (\xi^k - \xi^{k+1})$ 和 $b = (\xi^{k+1} - \xi^{k+2})$, 得到

$$\begin{array}{rl} & \|\xi^k - \xi^{k+1}\|^2_{\mathcal{H}} - \|\xi^{k+1} - \xi^{k+2}\|^2_{\mathcal{H}} \\ = & 2(\xi^k - \xi^{k+1})^T \mathcal{H}\{(\xi^k - \xi^{k+1}) - (\xi^{k+1} - \xi^{k+2})\} \\ & -\|(\xi^k - \xi^{k+1}) - (\xi^{k+1} - \xi^{k+2})\|^2_{\mathcal{H}}. \end{array}$$

利用 (3.11) 替换上面等式右端的第一项, 得到

$$\begin{aligned} &\|\xi^k-\xi^{k+1}\|_{\mathcal{H}}^2-\|\xi^{k+1}-\xi^{k+2}\|_{\mathcal{H}}^2 \\ &\quad\geq \|(\xi^k-\tilde{\xi}^k)-(\xi^{k+1}-\tilde{\xi}^{k+1})\|_{(\mathcal{Q}^T+\mathcal{Q})}^2 \\ &\qquad -\|(\xi^k-\xi^{k+1})-(\xi^{k+1}-\xi^{k+2})\|_{\mathcal{H}}^2. \end{aligned} \tag{3.14}$$

用校正公式 (3.6) 处理上式右端得到

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

由于 $(\mathcal{Q}^T+\mathcal{Q})-\mathcal{M}^T\mathcal{H}\mathcal{M}=\mathcal{G}\succeq 0$, (3.14) 右端非负, 定理结论得证. $\square$

不等式 (3.9) 和 (3.10) 说明, 变量替换下的广义 PPA 算法同样具备和 PPA 算法的性质 (1.4) 和 (1.5).

在广义邻近点算法(Generalized PPA)中, 校正矩阵 $\mathcal{M}$ 是由 (3.2) 中的预测矩阵 $\mathcal{Q}$ 唯一确定的. 如果 (3.2) 中的 $\mathcal{Q}$ 是对称的, 根据相关的定义, 校正矩阵为

$$M=\tfrac{1}{2}(I+Q^{-T}Q) \qquad \text{或} \qquad \mathcal{M}=\tfrac{1}{2}\big(\mathcal{I}+\mathcal{Q}^{-T}\mathcal{Q}\big), \tag{3.15}$$

就是单位矩阵. 我们将校正矩阵并非单位矩阵, 迭代序列又具备 (1.4)-(1.5) 这类性质的算法, 称为广义邻近点算法.

4 基于秩一校正的广义 PPA 算法

前一讲介绍的方法, 预测产生的 Q 矩阵是一个容易求逆的矩阵与一个广义秩一矩阵的和. 我们对这样的串型预测, 给出广义邻近点算法的校正公式.

4.1 Primal-Dual 预测的广义 PPA 算法

设预测是由前一讲中的 Primal-Dual 预测给出的, 我们得到形如 (3.2) 的变分不

等式, 其中

$$\mathcal{Q}_{PD}=\begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{4.1}$$

利用前一讲给出的 $\mathcal{L}$, $\mathcal{E}$, 那么

$$\mathcal{Q}_{PD}=\begin{pmatrix} \mathcal{L} & \mathcal{E} \\ 0 & I_m \end{pmatrix}. \tag{4.2}$$

由于

$$\mathcal{M}_{PD}=\frac{1}{2}\big(\mathcal{I}_{p+1}+\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}\big).$$

我们先来考察一下 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$. 注意到

$$\mathcal{Q}_{PD}^{T}=\begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E}^T & I_m \end{pmatrix} \quad \text{和} \quad \mathcal{Q}_{PD}^{-T}=\begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix}.$$

所以

$$
\begin{aligned}
\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD} &= \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix} \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ 0 & I_m \end{pmatrix} \\
&= \begin{pmatrix} \mathcal{L}^{-T}\mathcal{L} & \mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} & I_m - \mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} \end{pmatrix}
\end{aligned} \tag{4.3}
$$

分别计算 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$ 的四块. 因为

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

矩阵 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$ 左上角块,

$$
\begin{aligned}
\mathcal{L}^{-T}\mathcal{L} &= \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix} \\
&= \begin{pmatrix} 0 & -I_m & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -I_m \\ I_m & \cdots & I_m & I_m \end{pmatrix}. \qquad (4.4)
\end{aligned}
$$

矩阵 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$ 右上角块,

$$\mathcal{L}^{-T}\mathcal{E} = \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} I_m \\ I_m \\ \vdots \\ I_m \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix}. \tag{4.5}$$

矩阵 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$ 左下角块, 利用 (4.4), 得到

$$\begin{aligned} -\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} &= -\begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix} \begin{pmatrix} 0 & -I_m & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -I_m \\ I_m & I_m & \cdots & I_m \end{pmatrix} \\ &= \begin{pmatrix} -I_m & 0 & \cdots & 0 \end{pmatrix}. \end{aligned} \tag{4.6}$$

矩阵 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$ 右下角块,

$$I_m - \mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} = I_m - \begin{pmatrix} I_m & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} I_m \\ I_m \\ \vdots \\ I_m \end{pmatrix} = 0. \tag{4.7}$$

组装在一起就是

$$\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD} = \begin{pmatrix} 0 & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -I_m & 0 \\ I_m & \cdots & I_m & I_m & I_m \\ -I_m & \cdots & 0 & 0 & 0 \end{pmatrix}. \tag{4.8}$$

最后得到

$$\mathcal{M}_{PD} = \frac{1}{2}\big(\mathcal{I}_{p+1} + \mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}\big) = \frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & I_m \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}. \quad (4.9)$$

利用前一讲的变换, 采用 (4.9) 中的矩阵 $\mathcal{M}_{PD}$ 的校正 (3.6) 可以写成等价的

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & \frac{1}{\beta}I_m \\ -\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix}\begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \quad (4.10)$$

4.2 Dual-Primal 预测的广义 PPA 算法

设预测时由前一讲中的 Dual-Primall 预测给出的. 我们得到形如 (3.2) 的变分不等式, 其中

$$\mathcal{Q}_{DP} = \begin{pmatrix} I_m & 0 & \cdots & 0 & 0 \\ I_m & I_m & \ddots & \vdots & 0 \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & 0 \\ -I_m & -I_m & \cdots & -I_m & I_m \end{pmatrix}. \tag{4.11}$$

利用记号 $\mathcal{L}$, $\mathcal{E}$, 那么

$$\mathcal{Q}_{DP} = \begin{pmatrix} \mathcal{L} & 0 \\ -\mathcal{E}^T & I_m \end{pmatrix}. \tag{4.12}$$

由于

$$\mathcal{M}_{DP} = \frac{1}{2}\big(\mathcal{I} + \mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP}\big).$$

我们先来考察一下 $\mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP}$. 注意到

$$\mathcal{Q}_{DP}^{T} = \begin{pmatrix} \mathcal{L}^T & -\mathcal{E} \\ 0 & I_m \end{pmatrix} \qquad \text{和} \qquad \mathcal{Q}_{DP}^{-T} = \begin{pmatrix} \mathcal{L}^{-T} & \mathcal{L}^{-T}\mathcal{E} \\ 0 & I_m \end{pmatrix}.$$

则有

$$\mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP} = \begin{pmatrix} \mathcal{L}^{-T} & \mathcal{L}^{-T}\mathcal{E} \\ 0 & I_m \end{pmatrix}\begin{pmatrix} \mathcal{L} & 0 \\ -\mathcal{E}^T & I_m \end{pmatrix} = \begin{pmatrix} \mathcal{L}^{-T}\mathcal{L} - \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T & \mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^T & I_m \end{pmatrix}. \tag{4.13}$$

分别计算分块矩阵 $\mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP}$ 中的四块. 从 (4.4) 和 (4.5) 我们已经有了

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

因此 (4.13) 的左上角部分

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

矩阵 $\mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP}$ 的右上角部分, $\mathcal{L}^{-T}\mathcal{E}$ 在 (4.5) 中已经有了交代. 所以

$$\begin{aligned}\mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP} &= \begin{pmatrix}\mathcal{L}^{-T}\mathcal{L}-\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T & \mathcal{L}^{-T}\mathcal{E}\\ -\mathcal{E}^T & I_m\end{pmatrix}\\ &= \begin{pmatrix}0 & -I_m & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & -I_m & 0\\ 0 & \cdots & 0 & 0 & I_m\\ -I_m & \cdots & \cdots & -I_m & I_m\end{pmatrix}. \qquad (4.14)\end{aligned}$$

最后, 我们得到

$$\mathcal{M}_{DP} = \frac{1}{2}\big(\mathcal{I} + \mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP}\big) = \frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & I_m \\ I_m & \cdots & I_m & I_m & 2I_m \end{pmatrix}. \quad (4.15)$$

由相应的变换, 采用 (4.15) 中的矩阵 $\mathcal{M}_{DP}$ 的校正 (3.6) 可以写成等价的

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^k \\ A_2x_2^k \\ \vdots \\ A_px_p^k \\ \lambda^k \end{pmatrix} - \frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & \frac{1}{\beta}I_m \\ -\beta I_m & -\beta I_m & \cdots & -\beta I_m & 2I_m \end{pmatrix}\begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \quad (4.16)$$

5 基于秩二校正的广义 PPA 算法

第六讲 $\S 7$ 讲述的方法, 预测产生的 $\mathcal{Q}$ 矩阵是一个容易求逆的矩阵与一个广义秩二矩阵的和. 广义 PPA 的校正完全是由预测决定的. 因此, 广义邻近点算法只需根据预测矩阵 $\mathcal{Q}$ 给出相应的校正矩阵 $\mathcal{M}$.

5.1 Primal-Dual 预测的广义 PPA 算法

设预测是由第六讲 $\S 7.1$ 讲述的方法中的 Primal-Dual 预测给出的, 我们得到形如 (3.2) 的变分不等式, 其中的 $\mathcal{Q}$ 我们记为 $\mathcal{Q}_{PD}$.

$$\mathcal{Q}_{PD} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ I_m & I_m & \cdots & I_m & \frac{5}{2}I_m \end{pmatrix} = \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ \mathcal{E}^T & \frac{5}{2}I_m \end{pmatrix}. \tag{5.1}$$

我们要给出

$$\mathcal{M} = \frac{1}{2}\mathcal{Q}_{PD}^{-T}(\mathcal{Q}_{PD}^{T} + \mathcal{Q}_{PD}),$$

首先看给出 $\mathcal{Q}_{PD}^{-T}$ 的形式. 由第六讲的 $\S 7.1$ 节的 (7.14) 式

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

得到

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

因此校正矩阵

$$\mathcal{M}_{PD} = \frac{1}{2}\mathcal{Q}_{PD}^{-T}(\mathcal{Q}_{PD}^{T} + \mathcal{Q}_{PD}) = \frac{1}{3}\mathcal{B}_{PD} + \frac{1}{2}\mathcal{C}_{PD}, \tag{5.2}$$

其中

$$\mathcal{B}_{PD} = \begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^{T}\mathcal{L}^{-T}\mathcal{L} - \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^{T} & \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^{T}\mathcal{L}^{-T}\mathcal{E} - \frac{5}{2}\mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^{T}\mathcal{L}^{-T}\mathcal{L} + \mathcal{E}^{T} & -\mathcal{E}^{T}\mathcal{L}^{-T}\mathcal{E} + \frac{5}{2}I_m \end{pmatrix} \tag{5.3}$$

和

$$\mathcal{C}_{PD} = \begin{pmatrix} \mathcal{I} + \mathcal{L}^{-T}\mathcal{L} & \mathcal{L}^{-T}\mathcal{E} \\ 0 & I_m \end{pmatrix}. \tag{5.4}$$

我们先计算矩阵 $\mathcal{B}_{PD}$ 的四块. 利用 (参见第六讲 §7.1节)

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

得到

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

和

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

因此矩阵 $\mathcal{B}_{PD}$ 的 (1,1) 块

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

矩阵 $\mathcal{B}_{PD}$ 的 (1,2) 块

$$\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} - \frac{5}{2}\mathcal{L}^{-T}\mathcal{E} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix} - \frac{5}{2}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix} = -\frac{3}{2}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix}.$$

矩阵 $\mathcal{B}_{PD}$ 的 (2,1) 块

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

矩阵 $\mathcal{B}_{PD}$ 的 (2,2) 块

$$
\frac{5}{2}I_m - \mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} = \frac{3}{2}I_m.
$$

所以

$$\mathcal{B}_{PD} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & -I_m & \cdots & -I_m & -\frac{3}{2}I_m \\ 0 & I_m & \cdots & I_m & \frac{3}{2}I_m \end{pmatrix}. \tag{5.5}$$

利用

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

得到

$$\mathcal{C}_{PD} = \begin{pmatrix} \mathcal{I} + \mathcal{L}^{-T}\mathcal{L} & \mathcal{L}^{-T}\mathcal{E} \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} I_m & -I_m & 0 & 0 & 0 \\ 0 & \ddots & \ddots & 0 & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & I_m \\ 0 & \cdots & \cdots & 0 & I_m \end{pmatrix}. \tag{5.6}$$

有了 (5.5) 和 (5.6), 校正矩阵

$$
\begin{aligned}
\mathcal{M}_{PD} &= \frac{1}{3}\mathcal{B}_{PD} + \frac{1}{2}\mathcal{C}_{PD} \\
&= \frac{1}{3}\begin{pmatrix}
0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 \\
0 & -I_m & \cdots & -I_m & -\frac{3}{2}I_m \\
0 & I_m & \cdots & I_m & \frac{3}{2}I_m
\end{pmatrix} \\
&\quad + \frac{1}{2}\begin{pmatrix}
I_m & -I_m & 0 & 0 & 0 \\
0 & \ddots & \ddots & 0 & \vdots \\
0 & 0 & I_m & -I_m & 0 \\
I_m & \cdots & I_m & 2I_m & I_m \\
0 & \cdots & \cdots & 0 & I_m
\end{pmatrix}
\end{aligned}
$$

的形式是非常简单的.

5.2 Dual-Primal 预测的广义 PPA 算法

由第六讲 §7.2 讲述的方法 中的Dual-Primal 预测给出的, 我们得到形如 (3.2) 的变分不等式, 其中的 $\mathcal{Q}$ 我们记为 $\mathcal{Q}_{DP}$.

$$\mathcal{Q}_{DP} = \begin{pmatrix} I_m & 0 & \cdots & 0 & -I_m \\ I_m & I_m & \ddots & \vdots & -I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & -I_m \\ -I_m & -I_m & \cdots & -I_m & \frac{5}{2}I_m \end{pmatrix} = \begin{pmatrix} \mathcal{L} & -\mathcal{E} \\ -\mathcal{E}^T & \frac{5}{2}I_m \end{pmatrix}. \tag{5.7}$$

我们要给出

$$\mathcal{M} = \frac{1}{2}\mathcal{Q}_{PD}^{-T}(\mathcal{Q}_{PD}^T + \mathcal{Q}_{PD}),$$

首先看给出 $\mathcal{Q}_{DP}^{-T}$ 的形式. 由第六讲§7.2节中的(7.21) 式

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

得到

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

因此校正矩阵

$$\mathcal{M}_{DP}=\frac{1}{2}\mathcal{Q}_{DP}^{-T}(\mathcal{Q}_{DP}^T+\mathcal{Q}_{DP})=\frac{1}{3}\mathcal{B}_{DP}+\frac{1}{2}\mathcal{C}_{DP}, \tag{5.8}$$

其中

$$\mathcal{B}_{DP}=\begin{pmatrix}\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L}-\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T & -\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E}+\frac{5}{2}\mathcal{L}^{-T}\mathcal{E}\\ \mathcal{E}^T\mathcal{L}^{-T}\mathcal{L}-\mathcal{E}^T & -\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E}+\frac{5}{2}I_m\end{pmatrix} \tag{5.9}$$

和

$$\mathcal{C}_{DP} = \begin{pmatrix} \mathcal{I} + \mathcal{L}^{-T}\mathcal{L} & -\mathcal{L}^{-T}\mathcal{E} \\ 0 & I_m \end{pmatrix}. \tag{5.10}$$

将 (5.9) 跟 (5.3) 比较, 利用 (5.5), 有

$$\mathcal{B}_{DP} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & -I_m & \cdots & -I_m & \frac{3}{2}I_m \\ 0 & -I_m & \cdots & -I_m & \frac{3}{2}I_m \end{pmatrix}. \tag{5.11}$$

将 (5.10) 跟 (5.4) 比较, 利用 (5.6), 有

$$
\begin{aligned}
\mathcal{C}_{DP} &= \begin{pmatrix} \mathcal{I} + \mathcal{L}^{-T}\mathcal{L} & -\mathcal{L}^{-T}\mathcal{E} \\ 0 & I_m \end{pmatrix} \\
&= \begin{pmatrix} I_m & -I_m & 0 & 0 & 0 \\ 0 & \ddots & \ddots & 0 & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & -I_m \\ 0 & \cdots & \cdots & 0 & I_m \end{pmatrix}.
\end{aligned} \tag{5.12}
$$

有了 (5.11) 和 (5.12), 校正矩阵

$$
\begin{aligned}
\mathcal{M}_{DP} &= \frac{1}{3}\mathcal{B}_{DP} + \frac{1}{2}\mathcal{C}_{DP} \\
&= \frac{1}{3}\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & -I_m & \cdots & -I_m & \frac{3}{2}I_m \\ 0 & -I_m & \cdots & -I_m & \frac{3}{2}I_m \end{pmatrix} \\
&\quad + \frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & 0 & 0 \\ 0 & \ddots & \ddots & 0 & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & -I_m \\ 0 & \cdots & \cdots & 0 & I_m \end{pmatrix}
\end{aligned}
$$

的形式是非常简单的.

6 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 [14]. 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 [1, 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 [17]. 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 [13] 中, 后面我们发表的一些论文 [8, 10, 11, 16], 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 [15].
- 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这样的问题, 后来我又在中科大和南航做线上报告, 教学相长, 得到一些新的, 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,

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

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

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

就可以了. 明确这一条, 得益于为 2022 年以来在南师大(线下), 南航 和 中科大(线上)讲课, 促使我深入思考, 努力想办法讲清楚.

- 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.

$$\left\{\begin{array}{rcl} HM &=& Q, \\ M^THM &=& D. \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} HM &=& Q, \\ Q^TM &=& D. \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} H &=& QD^{-1}Q^T, \\ M &=& Q^{-T}D. \end{array}\right. .$$

- 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,

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

- 前一讲的方法, 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, 我们用前一讲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

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

- 这一讲介绍的广义 PPA 算法, 是由预测唯一确定的. 对给定的预测矩阵 Q, 取

$$D = \frac{1}{2}(Q^T + Q), \qquad \text{和} \qquad v^{k+1} = v^k - Q^{-T} D(v^k - \tilde{v}^k),$$

迭代序列 $\{v^k\}$ 满足（其中 $H = QD^{-1}Q^T$）

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2$$

和

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

为什么说我们在数值优化方面做出了一些颇有特色又自成系统的工作呢？

首先, 变分不等式和邻近点算法是我们的主要工具. 任何一本关于数值优化的书, 都没有专门提及变分不等式 (VI), 也不会刻意介绍邻近点算法 (PPA), 尽管线性约束的凸优化问题的增广拉格朗日乘子法 (ALM) 是乘子 λ 的 PPA 算法.

- 我们把线性约束的凸优化问题转换成一个等价的结构型单调变分不等式，然后说明什么是变分不等式的 PPA 算法, 讨论了 PPA 算法的收敛性质.
- 变分不等式的 PPA 算法迭代的每一步, 都利用其可分离结构, 分解成一些简单的“小微”变分不等式, 求解这些小微变分不等式, 又可以通过求解相应的凸优化问题实现.
- 后来我们又有了基于 VI 的预测-校正方法的统一框架, 既可以用它来验证算法的收敛性, 又可以用它“按需设计”求解可分离凸优化问题的算法, 这就是我们与众不同的逻辑.
- 我们又应该保持清醒的头脑, 即使是 ADMM, 它也是松弛了的 ALM, 是关于乘子 λ 的 PPA 算法. 同时也可以强调, 求解线性约束凸优化问题, ALM 是个有竞争力的好方法.

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] 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.

[2] T. Goldstein, B. O’ Donoghue, S. Setzer and R. Baraniuk, Fast Alternating Direction Optimization Methods, SIAM J. Imaging Science, Vol. 7, No. 3, pp. 1588 – 1623, 2014.

[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, 高等学校计算数学学报 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), 高等学校计算数学学报 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, Mathematics of Computation, 92(2023), 1633-1663.

[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, 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].

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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]