从变分不等式的邻近点算法
到广义邻近点算法

II. 从分裂收缩算法的统一框架到广义PPA 算法

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

曲阜师范大学　2023 年 11 月 23 日

1 变分不等式PPA算法的主要性质

我们对变分不等式问题

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

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

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T F(w^{k+1}) \\ &\ge\ (w - w^{k+1})^T H(w^k - w^{k+1}), \ \ \forall\, w \in \Omega. \end{aligned} \tag{1.2}$$

w^{k+1} 是变分不等式问题 (1.1) 的解的充分必要条件是 (1.2) 中的 $w^k = w^{k+1}$.
PPA算法产生的迭代序列 $\{w^k\}$ 满足

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2, \quad \forall w^* \in \Omega^*. \tag{1.3}$$

并有

$$\|w^k - w^{k+1}\|_H^2 \le \|w^{k-1} - w^k\|_H^2. \tag{1.4}$$

不等式 (1.3) 和 (1.4) 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

$$\min\{\theta_1(x)+\theta_2(y)|Ax+By=b, x\in\mathcal{X}, y\in\mathcal{Y}\} \tag{2.5}$$

转换成变分不等式(1.1), 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

$$\left\{\begin{array}{lcl} x^{k+1} & \in & \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, \\ y^{k+1} & \in & \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b) \end{array}\right. \tag{2.6}$$

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 $v = (y, \lambda)$.

Analysis 根据最优性定理, (2.6) 的 x 和 y 子问题分别满足

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \{-A^T\lambda^k + \beta A^T\big(Ax^{k+1} + By^k - b\big)\} \ge 0, \ \forall\, x \in \mathcal{X} \tag{2.7a}$$

和

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T\lambda^k + \beta B^T\big(Ax^{k+1} + By^{k+1} - b\big)\} \ge 0, \ \forall\, y \in \mathcal{Y}. \tag{2.7b}$$

以 $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)$ 代入 (2.7) (消去其中的 λ^k), 我们分别得到

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad & \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ & \{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1})\} \geq 0, \ \forall\, x \in \mathcal{X}, \end{aligned} \tag{2.8a}$$

和

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T\lambda^{k+1}\} \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{2.8b}$$

将 (2.8) 写成紧凑的形式：$u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$,

$$\begin{aligned} \theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \Bigg\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} \\ + \beta \begin{pmatrix} A^T B \\ 0 \end{pmatrix} (y^k - y^{k+1}) \Bigg\} \geq 0, \ \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \end{aligned} \tag{2.9}$$

再把上式改写成

$$\theta(u)-\theta(u^{k+1})+\begin{pmatrix} x-x^{k+1}\\ y-y^{k+1}\end{pmatrix}^T\left\{\begin{pmatrix} -A^T\lambda^{k+1}\\ -B^T\lambda^{k+1}\end{pmatrix}+\beta\begin{pmatrix} A^TB\\ B^TB\end{pmatrix}(y^k-y^{k+1})\right.$$

$$\left.+\begin{pmatrix} 0 & 0\\ 0 & \beta B^TB\end{pmatrix}\begin{pmatrix} x^{k+1}-x^k\\ y^{k+1}-y^k\end{pmatrix}\right\}\geq 0,\ \forall\,(x,y)\in\mathcal{X}\times\mathcal{Y}. \quad (2.10)$$

然后, 我们有如下的引理:

引理 1 对给定的(y^k,λ^k), 设$w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})\in\Omega$是由交替方向法 (2.6) 生成的. 我们有

$$\theta(u)-\theta(u^{k+1})+\begin{pmatrix} x-x^{k+1}\\ y-y^{k+1}\\ \lambda-\lambda^{k+1}\end{pmatrix}^T\left\{\begin{pmatrix} -A^T\lambda^{k+1}\\ -B^T\lambda^{k+1}\\ Ax^{k+1}+By^{k+1}-b\end{pmatrix}+\beta\begin{pmatrix} A^T\\ B^T\\ 0\end{pmatrix}B(y^k-y^{k+1})\right.$$

$$\left.+\begin{pmatrix} 0 & 0\\ \beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix} y^{k+1}-y^k\\ \lambda^{k+1}-\lambda^k\end{pmatrix}\right\}\geq 0,\ \forall\,w\in\Omega. \quad (2.11)$$

证明 等式 $(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)=0$ 可以改写成

$$\lambda^{k+1}\in\Re^m,\quad (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\tfrac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\ge 0,\quad \forall\lambda\in\Re^m.$$

将上式加到 (2.11), 就得到引理之结论. □

为了方便, 我们定义

$$v=\begin{pmatrix} y\\ \lambda\end{pmatrix}\quad \text{and}\quad \mathcal{V}^*=\{(y^*,\lambda^*)\,|\,(x^*,y^*,\lambda^*)\in\Omega^*\},$$

得到下面的引理:

引理 2 对给定的 (y^k,λ^k), 设 $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})\in\Omega$ 是由交替方向法 (2.6) 生成的. 我们有

$$(v^{k+1}-v^*)^T H(v^k-v^{k+1})\ge (y^k-y^{k+1})^T B^T(\lambda^k-\lambda^{k+1}),\ \forall\, w^*\in\Omega^*, \tag{2.12}$$

其中

$$H=\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}. \tag{2.13}$$

Proof. Setting $w = w^*$ in (2.11), we get

$$
\begin{array}{l}
(v^{k+1} - v^*)^T H (v^k - v^{k+1}) \\
\quad \geq \quad \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} \beta B (y^k - y^{k+1}) \\
\qquad\qquad + \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}), \ \forall\, w^* \in \Omega^*. \quad (2.14)
\end{array}
$$

Observe the first part of the right hand side of (2.14),

$$
\begin{array}{rl}
\multicolumn{2}{l}{\begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} \beta B (y^k - y^{k+1})} \\
= & (y^k - y^{k+1})^T B^T \beta \big(A, B\big) \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix} \\
= & (y^k - y^{k+1})^T B^T \beta \big(Ax^{k+1} + By^{k+1} - (Ax^* + By^*)\big) \\
= & (y^k - y^{k+1}) B^T \underline{\beta \big(Ax^{k+1} + By^{k+1} - b\big)} \\
= & (y^k - y^{k+1}) B^T \underline{(\lambda^k - \lambda^{k+1})}. \qquad (2.15)
\end{array}
$$

To the second part, since $(w^{k+1}-w^*)^TF(w^{k+1}) = (w^{k+1}-w^*)^TF(w^*)$ and w^* is the optimal solution, it follows that

$$\begin{aligned} & \theta(u^{k+1}) - \theta(u^*) + (w^{k+1}-w^*)^TF(w^{k+1}) \\ = \quad & \theta(u^{k+1}) - \theta(u^*) + (w^{k+1}-w^*)^TF(w^*) \ge 0. \end{aligned} \tag{2.16}$$

The assertion (2.14) immediately. □

引理 3 对给定的 (y^k, λ^k), 设 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ 是由交替方向法 (2.6) 生成的. 我们有

$$(y^k - y^{k+1})^T B^T (\lambda^k - \lambda^{k+1}) \ge 0. \tag{2.17}$$

Proof. Because (2.8b) is true for the k-th iteration and the previous iteration, we have

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \ge 0, \ \forall\, y \in \mathcal{Y}, \tag{2.18}$$

and

$$\theta_2(y) - \theta_2(y^k) + (y - y^k)^T\{-B^T\lambda^k\} \ge 0, \ \forall\, y \in \mathcal{Y}, \tag{2.19}$$

Setting $y = y^k$ in (2.18) and $y = y^{k+1}$ in (2.19), respectively, and then adding the two resulting inequalities, we get the assertion (2.17) immediately. □

将 (2.17) 代入 (2.12), 我们得到

$$(v^{k+1}-v^*)^T H(v^k - v^{k+1}) \ge 0, \quad \forall\, v^* \in \mathcal{V}^*. \tag{2.20}$$

在上一讲中我们已经说明： $b^T H(a-b) \ge 0 \quad \Rightarrow \|b\|_H^2 \le \|a\|_H^2 - \|a-b\|_H^2.$
设 $a = v^k - v^*$, $b = v^{k+1} - v^*$, 就有下面的定理.

定理 1 对给定的 (y^k, λ^k), 设 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ 是由交替方向法 (2.6) 生成的. 我们有

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \ \forall\, v^* \in \mathcal{V}^*. \tag{2.21}$$

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

$$\|v^{k+1}-v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2. \tag{2.22}$$

不等式 (2.21) 和 (2.22) 展示了 ADMM 很好的性质. 在一些快速 ADMM 的研究 [2] 中, 都用到了 (2.22) 这条性质.

交替方向法收敛性证明的 再阐述

交替方向法处理的是两个可分离块的凸优化问题

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\}. \tag{2.23}$$

将其拉格朗日函数 $L(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)$ 的鞍点归结为等价的变分不等式的解点：

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{2.24a}$$

其中

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix},\quad \Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m. \tag{2.24b}$$

ADMM 的 k 步迭代从给定的核心变量 $v^k=(y^k,\lambda^k)$ 出发

$$\left\{\begin{array}{rcl} x^{k+1} &=& \arg\min\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k-b\|^2 \mid x\in\mathcal{X}\}, \quad (2.25a)\\ y^{k+1} &=& \arg\min\{\theta_2(y)-y^TB^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By-b\|^2 \mid y\in\mathcal{Y}\}, \quad (2.25b)\\ \lambda^{k+1} &=& \lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). \quad (2.25c)\end{array}\right.$$

根据最优性引理 1, ADMM k-步迭代满足

$$\begin{cases} x^{k+1} \in \mathcal{X}, & \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^k + \beta A^T(Ax^{k+1} + By^k - b)\} \geq 0, \ \forall x \in \mathcal{X}, \\ y^{k+1} \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^k + \beta B^T(Ax^{k+1} + By^{k+1} - b)\} \geq 0, \ \forall y \in \mathcal{Y}, \\ \lambda^{k+1} \in \Re^m, & (\lambda - \lambda^{k+1})^T\{(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \geq 0, \ \forall \lambda \in \Re^m. \end{cases}$$

利用 $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)$ 上面的式子可以整理改写成

$$x^{k+1}\in\mathcal{X}, \ \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^{k+1} + \beta A^TB(y^k - y^{k+1})\} \geq 0, \ \forall x \in \mathcal{X}, \quad (2.26a)$$

$$y^{k+1}\in\mathcal{Y}, \ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1} \} \geq 0, \ \forall y \in \mathcal{Y}, \quad (2.26b)$$

$$\lambda^{k+1} \in \Re^m, \ (\lambda - \lambda^{k+1})^T\{(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \geq 0, \ \forall \lambda \in \Re^m. \quad (2.26c)$$

在 (2.26b) 的后半部加上和为零的两项, 得到

$$\begin{cases} \underline{\theta_1(x) - \theta_1(x^{k+1})} + (x - x^{k+1})^T\{-A^T\lambda^{k+1} + \beta A^TB(y^k - y^{k+1}) \} \geq 0, \\ \underline{\theta_2(y) - \theta_2(y^{k+1})} + (y - y^{k+1})^T\{-B^T\lambda^{k+1} + \underbrace{\beta B^TB(y^k - y^{k+1}) + \beta B^TB(y^{k+1} - y^k)} \} \geq 0, \\ \qquad (\lambda - \lambda^{k+1})^T\{(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \geq 0. \end{cases}$$

利用变分不等式 (2.24), 进行合理整合, 得到

$$\underline{\theta(u)-\theta(u^{k+1})}+(w-w^{k+1})^T F(w^{k+1})$$
$$+\begin{pmatrix} x-x^{k+1} \\ y-y^{k+1}\end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T\end{pmatrix} B(y^k-y^{k+1})+\begin{pmatrix} y-y^{k+1} \\ \lambda-\lambda^{k+1}\end{pmatrix}^T \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k\end{pmatrix} \ge 0.$$

将上式中那个任意的 w, 设成解点 w^* 便有

$$\underline{\theta(u^*)-\theta(u^{k+1})}+(w^*-w^{k+1})^T F(w^{k+1})$$
$$+\begin{pmatrix} x^*-x^{k+1} \\ y^*-y^{k+1}\end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T\end{pmatrix} B(y^k-y^{k+1})+\begin{pmatrix} y^*-y^{k+1} \\ \lambda^*-\lambda^{k+1}\end{pmatrix}^T \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k\end{pmatrix} \ge 0.$$

经转换, 得到

$$\begin{pmatrix} y^{k+1}-y^* \\ \lambda^{k+1}-\lambda^*\end{pmatrix}^T \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix} y^k-y^{k+1} \\ \lambda^k-\lambda^{k+1}\end{pmatrix} \quad \text{后面记}\ v=\begin{pmatrix} y \\ \lambda\end{pmatrix},\quad H=\begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta}I_m\end{pmatrix}$$
$$\ge \begin{pmatrix} x^{k+1}-x^* \\ y^{k+1}-y^*\end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T\end{pmatrix} B(y^k-y^{k+1})+\underbrace{[\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^T F(w^{k+1})]}. \quad (2.27)$$

假如 (2.27) 式右端非负, 证明就基本上完成了. 由于

$$\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}) = \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \geq 0.$$

(2.27) 式右端下划线部分非负. 因此从 (2.27) 式得到

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \;\geq\; \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}). \tag{2.28}$$

对 (2.28) 式的右端进行处理, 有

$$\begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) = (y^k - y^{k+1})^T B^T \beta (A, B) \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}$$

$$\begin{aligned} &= (y^k - y^{k+1})^T B^T \beta\big(Ax^{k+1} + By^{k+1} - (Ax^* + By^*)\big) \qquad \text{利用}(Ax^* + By^* = b) \\ &= (y^k - y^{k+1})^T B^T \beta\big(Ax^{k+1} + By^{k+1} - b\big) \\ &= (y^k - y^{k+1})^T B^T (\lambda^k - \lambda^{k+1}). \end{aligned} \tag{2.29}$$

后面我们要证明 $(y^k - y^{k+1})^T B^T (\lambda^k - \lambda^{k+1}) \geq 0$.

利用 (2.26b) 有

$$\begin{array}{llcll} & \theta_2(y) - \theta_2(y^{k+1}) & + & (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} & \geq & 0, \ \forall\, y \in \mathcal{Y}, \\ \text{和} & \theta_2(y) - \theta_2(y^{k}) & + & (y - y^{k})^T\{-B^T\lambda^{k}\} & \geq & 0, \ \forall\, y \in \mathcal{Y}. \end{array}$$

（将任意的 y 分别设成 y^k 和 y^{k+1}）

$$\begin{array}{lcll} \theta_2(y^{k}) - \theta_2(y^{k+1}) & + & (y^{k} - y^{k+1})^T\{-B^T\lambda^{k+1}\} & \geq 0. \\ \theta_2(y^{k+1}) - \theta_2(y^{k}) & + & (y^{k+1} - y^{k})^T\{-B^T\lambda^{k}\} & \geq 0. \end{array}$$

(将上面两式相加, 就有) $(y^k - y^{k+1})^T B^T(\lambda^k - \lambda^{k+1}) \geq 0.$ ((2.29) 式右端非负)

证明了(2.29) 式右端非负, 进而得到 (2.28) 式右端非负. 所以

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0. \tag{2.30}$$

Lemma 2 告诉我们:

$$b^T H(a-b) \geq 0 \quad \Rightarrow \quad \|b\|_H^2 \leq \|a\|_H^2 - \|a-b\|_H^2. \tag{2.31}$$

在 (2.31) 中置 $a = (v^k - v^*)$ 和 $b = (v^{k+1} - v^*)$, 根据 (2.30) 就得到收敛的关键不等式

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

由 $\|v^k - v^{k+1}\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2$ 得 $\sum_{k=0}^{\infty} \|v^k - v^{k+1}\|_H^2 \leq \|v^0 - v^*\|_H^2.$

3 凸优化分裂收缩算法的统一框架

我们总是用变分不等式 (VI) 指导算法设计, 把线性约束的凸优化问题归结为下面的变分不等式:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{3.1}$$

Algorithms in a unified framework

A unified Algorithmic Framework for (3.1)

统一框架由预测-校正两部分组成

[Prediction Step.] 从给定的 v^k 出发, 求得预测点 $\tilde{w}^k \in \Omega$ 使其满足

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

其中 Q 不一定对称, 但是 $Q^T + Q$ 正定.

[Correction Step.] 给一个合适的非奇异矩阵 M, 由下式确定新的迭代点

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{3.2b}$$

Q 和 M 分别叫做预测矩阵和校正矩阵

Convergence Conditions

对算法框架 (3.2) 中的预测矩阵 Q 和校正矩阵 M, 存在正定矩阵 H, 使得

$$HM = Q, \tag{3.3a}$$

并且

$$G = Q^T + Q - M^T HM \succ 0. \tag{3.3b}$$

其实, 只要预测 (3.2a) 中的预测矩阵 Q 满足

$$Q^T + Q \succ 0,$$

我们总可以取

$$0 \prec G \prec Q^T + Q.$$

然后记

$$D = (Q^T + Q) - G,$$

则 $D \succ 0$. 令

$$M^T HM = D.$$

由矩阵方程组解得

$$\left\{ \begin{array}{r} HM = Q, \\ M^T HM = D. \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{r} HM = Q, \\ Q^T M = D. \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{l} H = QD^{-1}Q^T, \\ M = Q^{-T}D. \end{array} \right.$$

就得到满足收敛条件的校正矩阵 M.

实际计算中, 我们只要校正矩阵 M.

H 和 G 只是用来验证收敛条件的.

换句话说, 只要

$$Q^T + Q \succ 0.$$

我们就可以选两个正定矩阵 $D \succ 0$ 和 $G \succ 0$, 使得

这里可以有无穷多的选择

$$D + G = Q^T + Q.$$

将 (3.2b) 中的校正矩阵 M 取成

$$M = Q^{-T} D$$

条件 (3.3) 自然满足.

校正公式 (3.2b) 就是

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

可以通过

$$Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k) \qquad \text{来实现.}$$

3.1 Convergence proof in the unified framework

In this section, assuming the conditions (3.3) in the unified framework are satisfied, we prove some convergence properties.

定理 1 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (3.1) *and* $\tilde{w}^k$ *is obtained in the* k*-th iteration. If* v^k, v^{k+1} *and* $\tilde{w}^k$ *satisfy the conditions in the unified framework, then we have*

$$\begin{array}{l} \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ \quad \geq \quad \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega. \end{array} \tag{3.4}$$

Proof. Using $Q = HM$ (see (3.3a)) and the relation (3.2b), the right hand side of (3.3a) can be written as $(v - \tilde{v}^k)^T H(v^k - v^{k+1})$ and hence

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \ \forall w \in \Omega. \tag{3.5}$$

Applying the identity

$$Q(v^k - \tilde{v}^k) = HM(v^k - \tilde{v}^k) = H(v^k - v^{k+1}).$$

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

to the right hand side of (3.5) with

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

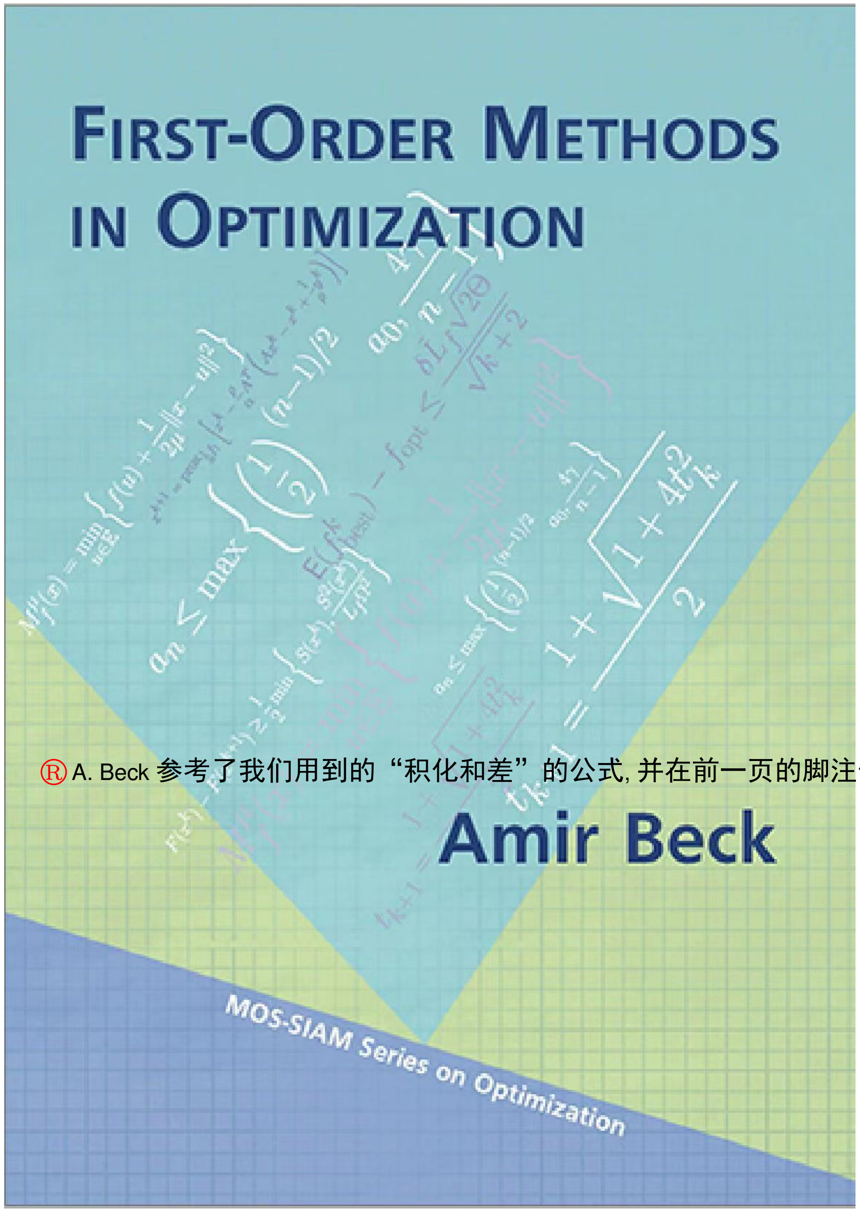
we thus obtain

$$\begin{aligned} & 2(v - \tilde{v}^k)^T H(v^k - v^{k+1}) \\ & \quad = \big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + (\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \end{aligned} \tag{3.6}$$

For the last term of (3.6), using $HM = Q$ and $2v^TQv = v^T(Q^T + Q)v$, we have

$$\begin{aligned} & \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ & \quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ & \quad \overset{(3.3a)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ & \quad = \quad 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ & \quad = \quad (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\ & \quad \overset{(3.3b)}{=} \quad \|v^k - \tilde{v}^k\|_G^2. \end{aligned} \tag{3.7}$$

Substituting (3.6), (3.7) in (3.5), the assertion of this theorem is proved. $\Box$



Ⓡ A. Beck 参考了我们用到的“积化和差”的公式, 并在前一页的脚注做了说明

15.3. Convergence Analysis of AD-PMM 429

We will use the following notation:

$$\tilde{\mathbf{x}}^k = \mathbf{x}^{k+1},$$
$$\tilde{\mathbf{z}}^k = \mathbf{z}^{k+1},$$
$$\tilde{\mathbf{y}}^k = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^k - \mathbf{c}).$$

Using (15.15), (15.16), the subgradient inequality, and the above notation, we obtain that for any $\mathbf{x} \in \mathrm{dom}(h_1)$ and $\mathbf{z} \in \mathrm{dom}(h_2)$,

$$h_1(\mathbf{x}) - h_1(\tilde{\mathbf{x}}^k) + \left\langle \rho\mathbf{A}^T\left(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right) + \mathbf{G}(\tilde{\mathbf{x}}^k - \mathbf{x}^k), \mathbf{x} - \tilde{\mathbf{x}}^k \right\rangle \geq 0,$$
$$h_2(\mathbf{z}) - h_2(\tilde{\mathbf{z}}^k) + \left\langle \rho\mathbf{B}^T\left(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right) + \mathbf{Q}(\tilde{\mathbf{z}}^k - \mathbf{z}^k), \mathbf{z} - \tilde{\mathbf{z}}^k \right\rangle \geq 0.$$

Using the definition of $\tilde{\mathbf{y}}^k$, the above two inequalities can be rewritten as

$$h_1(\mathbf{x}) - h_1(\tilde{\mathbf{x}}^k) + \langle \mathbf{A}^T\tilde{\mathbf{y}}^k + \mathbf{G}(\tilde{\mathbf{x}}^k - \mathbf{x}^k), \mathbf{x} - \tilde{\mathbf{x}}^k \rangle \geq 0,$$
$$h_2(\mathbf{z}) - h_2(\tilde{\mathbf{z}}^k) + \langle \mathbf{B}^T\tilde{\mathbf{y}}^k + (\rho\mathbf{B}^T\mathbf{B} + \mathbf{Q})(\tilde{\mathbf{z}}^k - \mathbf{z}^k), \mathbf{z} - \tilde{\mathbf{z}}^k \rangle \geq 0.$$

Adding the above two inequalities and using the identity

$$\mathbf{y}^{k+1} - \mathbf{y}^k = \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c}),$$

we can conclude that for any $\mathbf{x} \in \mathrm{dom}(h_1)$, $\mathbf{z} \in \mathrm{dom}(h_2)$, and $\mathbf{y} \in \mathbb{R}^m$,

$$H(\mathbf{x},\mathbf{z}) - H(\tilde{\mathbf{x}}^k, \tilde{\mathbf{z}}^k) + \left\langle \begin{pmatrix} \mathbf{x} - \tilde{\mathbf{x}}^k \\ \mathbf{z} - \tilde{\mathbf{z}}^k \\ \mathbf{y} - \tilde{\mathbf{y}}^k \end{pmatrix}, \begin{pmatrix} \mathbf{A}^T\tilde{\mathbf{y}}^k \\ \mathbf{B}^T\tilde{\mathbf{y}}^k \\ -\mathbf{A}\tilde{\mathbf{x}}^k - \mathbf{B}\tilde{\mathbf{z}}^k + \mathbf{c} \end{pmatrix} - \begin{pmatrix} \mathbf{G}(\mathbf{x}^k - \tilde{\mathbf{x}}^k) \\ \mathbf{C}(\mathbf{z}^k - \tilde{\mathbf{z}}^k) \\ \frac{1}{\rho}(\mathbf{y}^k - \mathbf{y}^{k+1}) \end{pmatrix} \right\rangle \geq 0, \quad (15.17)$$

where $\mathbf{C} = \rho\mathbf{B}^T\mathbf{B} + \mathbf{Q}$. We will use the following identity that holds for any positive semidefinite matrix $\mathbf{P}$:

$$(\mathbf{a} - \mathbf{b})^T\mathbf{P}(\mathbf{c} - \mathbf{d}) = \frac{1}{2}\left(\|\mathbf{a} - \mathbf{d}\|_{\mathbf{P}}^2 - \|\mathbf{a} - \mathbf{c}\|_{\mathbf{P}}^2 + \|\mathbf{b} - \mathbf{c}\|_{\mathbf{P}}^2 - \|\mathbf{b} - \mathbf{d}\|_{\mathbf{P}}^2\right).$$

Using the above identity, we can conclude that

$$(\mathbf{x} - \tilde{\mathbf{x}}^k)^T\mathbf{G}(\mathbf{x}^k - \tilde{\mathbf{x}}^k) = \frac{1}{2}\left(\|\mathbf{x} - \tilde{\mathbf{x}}^k\|_{\mathbf{G}}^2 - \|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2 + \|\tilde{\mathbf{x}}^k - \mathbf{x}^k\|_{\mathbf{G}}^2\right)$$
$$\geq \frac{1}{2}\|\mathbf{x} - \tilde{\mathbf{x}}^k\|_{\mathbf{G}}^2 - \frac{1}{2}\|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2, \quad (15.18)$$

as well as

$$(\mathbf{z} - \tilde{\mathbf{z}}^k)^T\mathbf{C}(\mathbf{z}^k - \tilde{\mathbf{z}}^k) = \frac{1}{2}\|\mathbf{z} - \tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 - \frac{1}{2}\|\mathbf{z} - \mathbf{z}^k\|_{\mathbf{C}}^2 + \frac{1}{2}\|\mathbf{z}^k - \tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 \quad (15.19)$$

and

$$\begin{aligned}
&2(\mathbf{y} - \tilde{\mathbf{y}}^k)^T(\mathbf{y}^k - \mathbf{y}^{k+1}) \\
&= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \|\tilde{\mathbf{y}}^k - \mathbf{y}^k\|^2 - \|\tilde{\mathbf{y}}^k - \mathbf{y}^{k+1}\|^2 \\
&= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \rho^2\|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}\|^2 \\
&- \|\mathbf{y}^k + \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}) - \mathbf{y}^k - \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c})\|^2 \\
&= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \rho^2\|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}\|^2 - \rho^2\|\mathbf{B}(\mathbf{z}^k - \tilde{\mathbf{z}}^k)\|^2.
\end{aligned}$$

3.2 Convergence in a strictly contraction sense

定理 2 *Let $\{v^k\}$ be the sequence generated by a method for the problem* (3.1) *and $\tilde{w}^k$ is obtained in the k-th iteration. If v^k, v^{k+1} and $\tilde{w}^k$ satisfy the conditions in the unified framework, then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{3.8}$$

Proof. Setting $w = w^*$ in (3.4), we get

$$\begin{aligned} &\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ &\quad \ge \quad \|v^k - \tilde{v}^k\|_G^2 + 2\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \end{aligned} \tag{3.9}$$

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

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

and thus

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \|v^k - \tilde{v}^k\|_G^2. \tag{3.10}$$

The assertion (3.8) follows directly. $\square$

定理 3 *For solving the variational inequality* (3.1)*, let* $\{w^k\}, \{\tilde{w}^k\}$ *be the sequence generated by* (3.2)*. If the conditions* (3.3) *are satisfied, then we have*

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2 - \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2. \quad (3.11)$$

Proof Note that we have

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

and

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

Set the vector w in the above two inequalities by $\tilde{w}^{k+1}$ and $\tilde{w}^k$, respectively, we get

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

and

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

Adding the above two inequalities, it follows that

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

Adding $\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}$ to the both sides of the last inequality, we get

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

and thus

$$(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} \geq \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \tag{3.12}$$

Finally, by using $\|a\|^2_H - \|b\|^2_H = 2a^T H(a-b) - \|a-b\|^2_H$ and (3.12), we get

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

This is the equivalent form of (3.11) and the proof is complete. □

4 预测-校正的广义 PPA 算法

求解变分不等式 (1.1) 采用单位步长校正的时候, 如果预测公式

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall\, w \in \Omega, \tag{4.1}$$

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

$$D \succ 0, \qquad G \succ 0 \qquad \text{和} \qquad D + G = Q^T + Q, \tag{4.2}$$

再令

$$M = Q^{-T} D \qquad \text{和} \qquad H = Q D^{-1} Q^T. \tag{4.3}$$

则由单位步长校正

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k) \tag{4.4}$$

产生的新的迭代序列 $\{v^k\}$ 满足

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{4.5}$$

如果我们采用一对特殊的 D 和 G, 使得

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

那么, (4.5) 就变成了

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_D^2, \quad \forall v^* \in \mathcal{V}^*. \tag{4.6}$$

对选定的 D, 根据 (4.3), 总有

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

因此, (4.6) 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 (4.4)), 上式就了

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{4.7}$$

此外, 关于统一框架中所有固定步长的方法 (见这一讲的定理 3) 都证明了

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2. \tag{4.8}$$

我们把上述分析结果写成下面的定理.

定理 4 用预测校正方法求解变分不等式 (1.1), 设预测 (4.1) 中的预测矩阵 Q 满足 $Q^T + Q \succ 0$. 若令

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

则由单位步长校正公式

$$v^{k+1} = v^k - Q^{-T} D(v^k - \tilde{v}^k) \tag{4.9}$$

产生的新的迭代点具有性质 (4.7) 和 (4.8), 其中

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

求解变分不等式 (1.1), 我们把迭代序列具有性质 (4.7) 和 (4.8) 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.

5 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

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i)\;\Big|\;\sum_{i=1}^{p}A_i x_i = b \text{ (or } \geq b), \quad x_i\in\mathcal{X}_i\Big\}. \tag{5.1}$$

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 (5.1) can be written as the following VI:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{5.2a}$$

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T \lambda \\ \vdots \\ -A_p^T \lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}, \tag{5.2b}$$

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 (5.2).

多块问题 (5.2) 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 $\tilde{w}^k \in \Omega$:

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^TA_1^T\lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\};\\
\tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^TA_2^T\lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\};\\
\qquad\vdots\\
\tilde{x}_i^k \in \arg\min_{x_i\in\mathcal{X}_i}\{\theta_i(x_i) - x_i^TA_i^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1}A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\};\\
\qquad\vdots\\
\tilde{x}_p^k \in \arg\min_{x_p\in\mathcal{X}_p}\{\theta_p(x_p) - x_p^TA_p^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1}A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\};\\
\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^{p}A_j\tilde{x}_j^k - b\big)\big].
\end{array}\right. \tag{5.3}$$

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.

5.1 采用 Primal-Dual 预测的预测矩阵

Analysis for the P-D Prediction 我们先看 (5.3) 中 x 子问题

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

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

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

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

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{\underline{-A_i^T\tilde{\lambda}^k} + \beta A_i^T\Big(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)\Big) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0. \tag{5.4a}$$

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

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

最优性条件是

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \big\{ \big(\underline{\textstyle\sum_{j=1}^p A_j \tilde{x}_j^k - b} \big) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \big\} \geq 0, \quad \forall \lambda \in \Lambda. \qquad \text{(5.4b)}$$

Summating (5.4a) and (5.4b), for the predictor $\tilde{w}^k$ generated by (5.3), we have $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \geq (w - \tilde{w}^k)^T Q (w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \qquad \text{(5.5a)}$$

where

$$Q = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \qquad \text{(5.5b)}$$

5.2 变量代换下的预测矩阵

The optimization problem (5.1) has been translated to VI (5.2), namely,

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

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

$$P = \begin{pmatrix} \sqrt{\beta}A_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta}A_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sqrt{\beta}A_p & 0 \\ 0 & \cdots & \cdots & 0 & (1/\sqrt{\beta})I_m \end{pmatrix}, \quad \xi = Pw = \begin{pmatrix} \sqrt{\beta}A_1x_1 \\ \sqrt{\beta}A_2x_2 \\ \vdots \\ \sqrt{\beta}A_px_p \\ (1/\sqrt{\beta})\lambda \end{pmatrix}. \tag{5.6}$$

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 (5.6), for the matrix Q in (5.5b), we have

$$Q = P^T \mathcal{Q} P, \qquad \text{where} \qquad \mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{5.7}$$

Thus, for the right hand side of (5.5a), 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 (5.5) that we have the following VI for the P-D prediction:

$$\begin{aligned} \tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) &+ (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ &\geq (\xi - \tilde{\xi}^k)^T \mathcal{Q} (\xi^k - \tilde{\xi}^k), \quad \forall\, w \in \Omega. \end{aligned} \tag{5.8}$$

where $\mathcal{Q}$ is given in (5.7).

6 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

$$\theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(w-\tilde{w}^k)^TQ(w^k-\tilde{w}^k),\quad\forall\, w\in\Omega. \tag{6.1}$$

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 (6.1) 改写成 $\tilde{w}^k\in\Omega$,

$$\theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(\xi-\tilde{\xi}^k)^T\mathcal{Q}(\xi^k-\tilde{\xi}^k),\quad\forall\, w\in\Omega, \tag{6.2}$$

其中 $Q=P^T\mathcal{Q}P$,

$$\mathcal{Q}^T+\mathcal{Q}\succ 0 \tag{6.3}$$

是正定矩阵. 在 $\mathcal{Q}$ 非对称但 (6.3) 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

$$\mathcal{D}\succ 0,\quad \mathcal{G}\succ 0,\quad \text{和}\quad \mathcal{D}+\mathcal{G}=\mathcal{Q}^T+\mathcal{Q}. \tag{6.4}$$

根据前一讲的分析, 我们有如下的定理.

定理 5 设预测点$\tilde{\xi}^k$满足条件 (6.2), 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 (6.4) 成立.

$$\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D} \tag{6.5}$$

那么, 利用矩阵 (6.5) 校正

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k), \tag{6.6}$$

产生的 ξ^{k+1} 满足

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2, \quad \forall \xi^* \in \Xi^*, \tag{6.7}$$

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

如果选

$$\mathcal{D} = \mathcal{G} = \frac{1}{2}(\mathcal{Q}^T + \mathcal{Q}) \tag{6.8}$$

那么, (6.7) 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 (6.6), 上式就成了

$$\|\xi^{k+1}-\xi^*\|_{\mathcal{H}}^2 \le \|\xi^k-\xi^*\|_{\mathcal{H}}^2 - \|\xi^k-\xi^{k+1}\|_{\mathcal{H}}^2, \quad \forall \xi^* \in \Xi^*. \tag{6.9}$$

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 6 如果预测点 $\tilde{\xi}^k$ 满足条件 (6.2), 那么, 由校正 (6.6) 产生的新的迭代点 ξ^{k+1} 满足

$$\|\xi^{k+1}-\xi^{k+2}\|_{\mathcal{H}}^2 \le \|\xi^k-\xi^{k+1}\|_{\mathcal{H}}^2. \tag{6.10}$$

证明. 首先, 我们证明迭代序列满足

$$\begin{aligned}(\xi^k-\xi^{k+1})^T\mathcal{H}[(\xi^k-\xi^{k+1})-(\xi^{k+1}-\xi^{k+2})] \\ \ge \quad \frac{1}{2}\|(\xi^k-\tilde{\xi}^k)-(\xi^{k+1}-\tilde{\xi}^{k+1})\|_{(\mathcal{Q}^T+\mathcal{Q})}^2.\end{aligned} \tag{6.11}$$

将预测 (6.2) 中的 k 改为 $k+1$, 我们有

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

将上式中任意的 w 设为 $\tilde{w}^k$, 得到

$$\theta(\tilde{x}^k) - \theta(\tilde{x}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \ge (\tilde{\xi}^k - \tilde{\xi}^{k+1})^T \mathcal{Q}(\xi^{k+1} - \tilde{\xi}^{k+1}). \tag{6.12}$$

将预测 (6.2) 式中任意的 w 设为 $\tilde{w}^{k+1}$, 就有

$$\theta(\tilde{x}^{k+1}) - \theta(\tilde{x}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \ge (\tilde{\xi}^{k+1} - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k). \tag{6.13}$$

将 (6.12), (6.13) 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

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

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

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

在上式左端利用 $\mathcal{Q} = \mathcal{H}\mathcal{M}$ 和校正公式 (6.6), 就得到 (6.11).

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

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

利用 (6.11) 替换上面等式右端的第一项, 得到

$$\begin{aligned}
&\|\xi^k - \xi^{k+1}\|_{\mathcal{H}}^2 - \|\xi^{k+1} - \xi^{k+2}\|_{\mathcal{H}}^2 \\
&\quad \geq \|(\xi^k - \tilde{\xi}^k) - (\xi^{k+1} - \tilde{\xi}^{k+1})\|_{(\mathcal{Q}^T+\mathcal{Q})}^2 \\
&\qquad - \|(\xi^k - \xi^{k+1}) - (\xi^{k+1} - \xi^{k+2})\|_{\mathcal{H}}^2.
\end{aligned} \tag{6.14}$$

用校正公式 (6.6) 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 $(\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T\mathcal{H}\mathcal{M} = \mathcal{G} \succeq 0$, (6.14) 右端非负, 定理结论得证. $\square$

不等式 (6.9) 和 (6.10) 说明, 变量替换下的广义 PPA 算法同样具备和 PPA 算法的性质 (1.3) 和 (1.4).

在广义邻近点算法(Generalized PPA)中, 校正矩阵 $\mathcal{M}$ 是由 (6.2) 中的预测矩阵 $\mathcal{Q}$ 唯一确定的. 如果 (6.2) 中的 $\mathcal{Q}$ 是对称的, 根据相关的定义, 校正矩阵为

$$M = \frac{1}{2}(I + Q^{-T}Q) \quad \text{或} \quad \mathcal{M} = \frac{1}{2}\big(\mathcal{I} + \mathcal{Q}^{-T}\mathcal{Q}\big), \tag{6.15}$$

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 (1.3)-(1.4) 这类性质的算法, 称为广义邻近点算法.

7 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 $\mathcal{Q}$ 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.

设预测是由 Primal-Dual 预测给出的, 我们得到形如 (6.2) 的变分不等式, 其中

$$\mathcal{Q}_{PD} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{7.1}$$

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

$$\mathcal{Q}_{PD} = \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ 0 & I_m \end{pmatrix}. \tag{7.2}$$

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

$$\begin{aligned} \mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD} &= \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}^{T}\mathcal{L}^{-T} & I_m \end{pmatrix} \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ 0 & I_m \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{L}^{-T}\mathcal{L} & \mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^{T}\mathcal{L}^{-T}\mathcal{L} & I_m - \mathcal{E}^{T}\mathcal{L}^{-T}\mathcal{E} \end{pmatrix} \end{aligned} \tag{7.3}$$

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,

$$
\begin{aligned}
\mathcal{L}^{-T}\mathcal{L} &= \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix} \\
&= \begin{pmatrix} 0 & -I_m & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -I_m \\ I_m & \cdots & I_m & I_m \end{pmatrix}. \qquad (7.4)
\end{aligned}
$$

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

$$\mathcal{L}^{-T}\mathcal{E} = \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} I_m \\ I_m \\ \vdots \\ I_m \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix}. \tag{7.5}$$

矩阵 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$ 左下角块, 利用 (7.4), 得到

$$\begin{aligned} -\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} &= -\begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix} \begin{pmatrix} 0 & -I_m & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -I_m \\ I_m & I_m & \cdots & I_m \end{pmatrix} \\ &= \begin{pmatrix} -I_m & 0 & \cdots & 0 \end{pmatrix}. \end{aligned} \tag{7.6}$$

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

$$I_m - \mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} = I_m - \begin{pmatrix} I_m & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} I_m \\ I_m \\ \vdots \\ I_m \end{pmatrix} = 0. \tag{7.7}$$

组装在一起就是

$$\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD} = \begin{pmatrix} 0 & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -I_m & 0 \\ I_m & \cdots & I_m & I_m & I_m \\ -I_m & \cdots & 0 & 0 & 0 \end{pmatrix}. \tag{7.8}$$

最后得到

$$\mathcal{M}_{PD}=\frac{1}{2}\big(\mathcal{I}_{p+1}+\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}\big)=\frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \vdots\\ 0 & 0 & I_m & -I_m & 0\\ I_m & \cdots & I_m & 2I_m & I_m\\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{7.9}$$

利用前一讲的变换, 采用 (7.9) 中的矩阵 $\mathcal{M}_{PD}$ 的校正 (6.6) 可以写成等价的

$$\begin{pmatrix} A_1x_1^{k+1}\\ A_2x_2^{k+1}\\ \vdots\\ A_px_p^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix} A_1x_1^{k}\\ A_2x_2^{k}\\ \vdots\\ A_px_p^{k}\\ \lambda^{k}\end{pmatrix}-\frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \vdots\\ 0 & 0 & I_m & -I_m & 0\\ I_m & \cdots & I_m & 2I_m & \frac{1}{\beta}I_m\\ -\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix}\begin{pmatrix} A_1x_1^{k}-A_1\tilde{x}_1^{k}\\ A_2x_2^{k}-A_2\tilde{x}_2^{k}\\ \vdots\\ A_px_p^{k}-A_p\tilde{x}_p^{k}\\ \lambda^{k}-\tilde{\lambda}^{k}\end{pmatrix} \tag{7.10}$$

8 Conclusions

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]