# 从最优化方法的基本原理
# 到凸优化的分裂收缩算法

## II. 从算法的统一框架到广义的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

MISAI清华三亚国际数学论坛 2024年2月17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

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

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

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

$w^{k+1}$ 是变分不等式问题 (1.1) 的解的充分必要条件是 (1.2) 中的 $w^k = w^{k+1}$. 这个系列的《**I.** 从邻近点算法到均困的ALM和ADMM方法》中已经证明了

**定理 1** 由 PPA 算法 (1.2) 产生的迭代序列 $\{w^k\}$ 满足

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

**定理 2** 由 $\mathrm{PPA}$ 算法 (1.2) 产生的迭代序列 $\{w^k\}$ 满足

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

**证明**. 将 (1.2) 中任意的 $w$ 设为 $w^{k+2}$, 得到

$$\theta(u^{k+2}) - \theta(u^{k+1}) + (w^{k+2} - w^{k+1})^T F(w^{k+1}) \ge (w^{k+2} - w^{k+1})^T H(w^k - w^{k+1}). \tag{1.5}$$

将 (1.2) 中的 $k$ 改成 $k+1$, 就有

$$\begin{aligned} w^{k+2} \in \Omega, \quad \theta(u) - \theta(u^{k+2}) &+ (w - w^{k+2})^T F(w^{k+2}) \\ &\ge \; (w - w^{k+2})^T H(w^{k+1} - w^{k+2}), \quad \forall\, w \in \Omega. \end{aligned}$$

将上式中任意的 $w$ 设为 $w^{k+1}$, 就有

$$\theta(u^{k+1}) - \theta(u^{k+2}) + (w^{k+1} - w^{k+2})^T F(w^{k+2}) \ge (w^{k+1} - w^{k+2})^T H(w^{k+1} - w^{k+2}). \tag{1.6}$$

将 (1.5) 和 (1.6) 加在一起, 利用

$$(w^{k+1} - w^{k+2})^T (F(w^{k+1}) - F(w^{k+2})) \equiv 0$$

得到

$$(w^{k+1} - w^{k+2})^T H\{(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})\} \geq 0. \qquad (1.7)$$

设 $a = (w^{k+1} - w^{k+2}), b = (w^{k+1} - w^{k+2})$, 并利用

$$b^T H(a - b) \geq 0, \qquad \Rightarrow \qquad \|b\|_H^2 \leq \|a\|_H^2 - \|a - b\|_H^2$$

从不等式 (1.7) 直接得到 (1.4). 定理结论得证. □

因此, 邻近点算法有两个很好的性质, (1.3) 和 (1.4)

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

和

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

# 2 ADMM算法的主要性质

把两块可分离凸优化问题

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

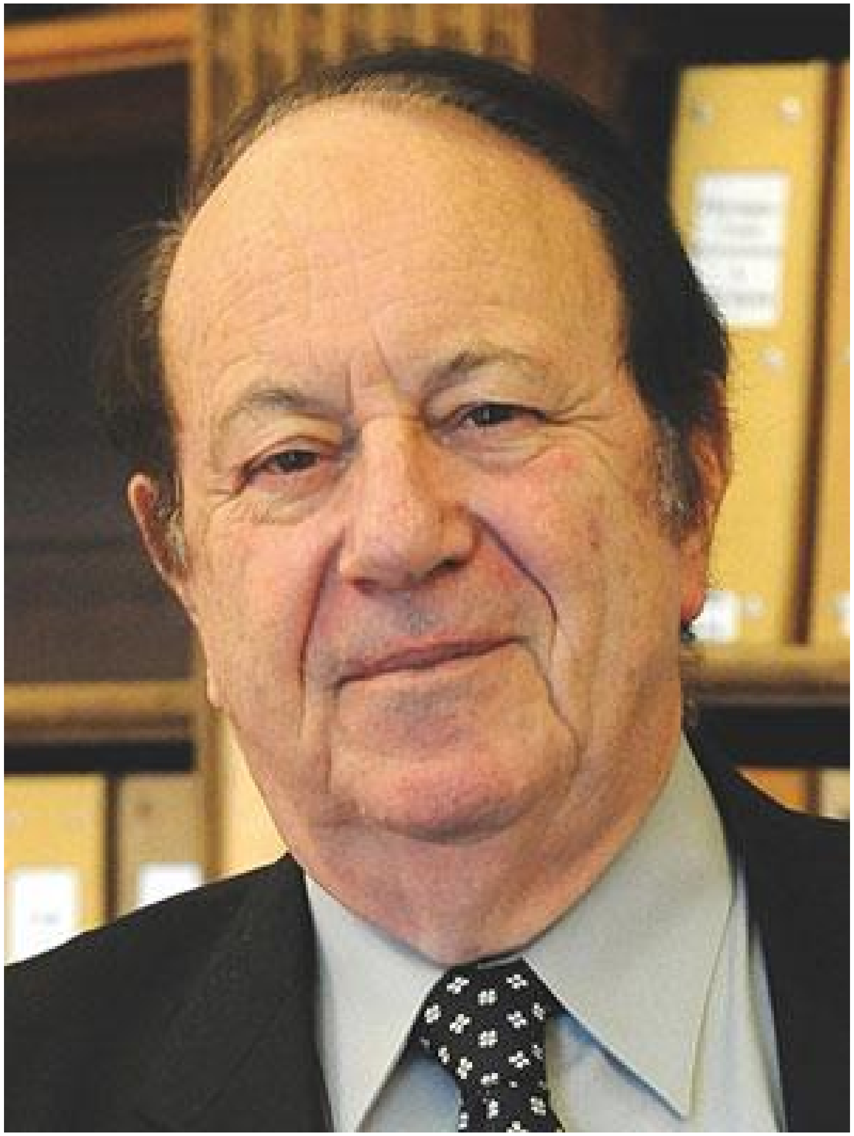
转换成变分不等式 (1.1), 其中

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

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

问题 (2.1) 的增广拉格朗日函数是

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



**Roland Glowinski**

(1937-2022)

上世纪70年代中期, Glowinski 和他的学生提出了相关算法[5], 在[3]叫做 Algorithm 2. 用来求解下面这类微分方程问题

$$\min\{f(x)+g(Mx)\}.$$

他们把它转换成问题

$$\left\{\begin{array}{ll}\min & f(x)+g(y), \\ \text{s.t} & Mx-y=0.\end{array}\right.$$

这是问题(2.1)的一个特例, 用下面的交替方向法求解。

乘子交替方向法 ADMM 的 $k$ 次迭代从给定的 $v^k = (y^k, \lambda^k)$ 开始, 通过

$$\begin{cases} x^{k+1} \in \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, \\ y^{k+1} \in \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} + By^{k+1} - b) \end{cases} \tag{2.2}$$

求得 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$. 在 [3] 中, $\gamma \in (0, \frac{\sqrt{5}+1}{2})$. 后面的讨论中, 我们默认 $\gamma = 1$. 由于 ADMM 方法的 $k$ 步迭代, $x^{k+1}$ 是根据给定的 $(y^k, \lambda^k)$ 算出来的, 我们称 $x$ 为中间变量, 称 $v = (y, \lambda)$ 为核心变量.

**定理 3** 对给定的 $(y^k, \lambda^k)$, 设 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ 是由交替方向法 (2.2) 生成的. 我们有

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \ \forall v^* \in \mathcal{V}^*. \tag{2.3}$$

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

**定理 4** 对给定的$(y^k, \lambda^k)$, 设$w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$是由交替方向法 (2.2) 生成的. 我们有

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2. \tag{2.4}$$

不等式 (2.3) 和 (2.4) 展示了 ADMM 很好的性质. 在一些快速 ADMM 的研究中 [6], 用到了 (2.4) 这条性质. 此外, 我们还发表了一些收敛更快一些的 ADMM.

**求解两可分离块问题 (2.1) 的交换 $y, \lambda$ 迭代顺序的 ADMM 算法**

第 $k$ 步迭代从给定的$(y^k, \lambda^k)$开始, 通过

(交换 $y$-$\lambda$ 迭代次序的 ADMM)

$$\begin{cases} x^{k+1} \in \operatorname{argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, & \text{(2.5a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), & \text{(2.5b)} \\ y^{k+1} \in \operatorname{argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^{k+1}) \,|\, y \in \mathcal{Y}\}. & \text{(2.5c)} \end{cases}$$

得到$w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$, 然后用

$$v^{k+1} := v^k - \alpha(v^k - v^{k+1}), \qquad \alpha \in (0, 2), \tag{2.6}$$

产生新的迭代点. 实际计算中往往取$\alpha = 1.5$. 该方法可参阅论文 [1, 8].

人们习惯于用经典的乘子交替方向法 (2.2) 求解问题 (2.1). 从问题 (2.1) 本身看, 原始变量 $x$ 和 $y$ 是平等的。

在算法设计上平等对待 $x$ 和 $y$ 子问题, 也是最自然不过的考虑. 因此我们采用对称的 (Symmetric) 交替方向法 [10]. 同年, Glowinski 在 [4] 中也提到了这个方法.

**求解两可分离块问题 (2.1) 的对称的 ADMM 算法**

第 $k$ 步迭代从给定的 $(y^k, \lambda^k)$ 开始, 通过

(对称的 ADMM)
$$\left\{\begin{array}{ll} x^{k+1} \in \operatorname{argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, & \text{(2.7a)} \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), & \text{(2.7b)} \\ y^{k+1} \in \operatorname{argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, & \text{(2.7c)} \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b). & \text{(2.7d)} \end{array}\right.$$

得到新的迭代点 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$,

其中 $\mu \in (0,1)$ (通常取 $\mu = 0.9$).

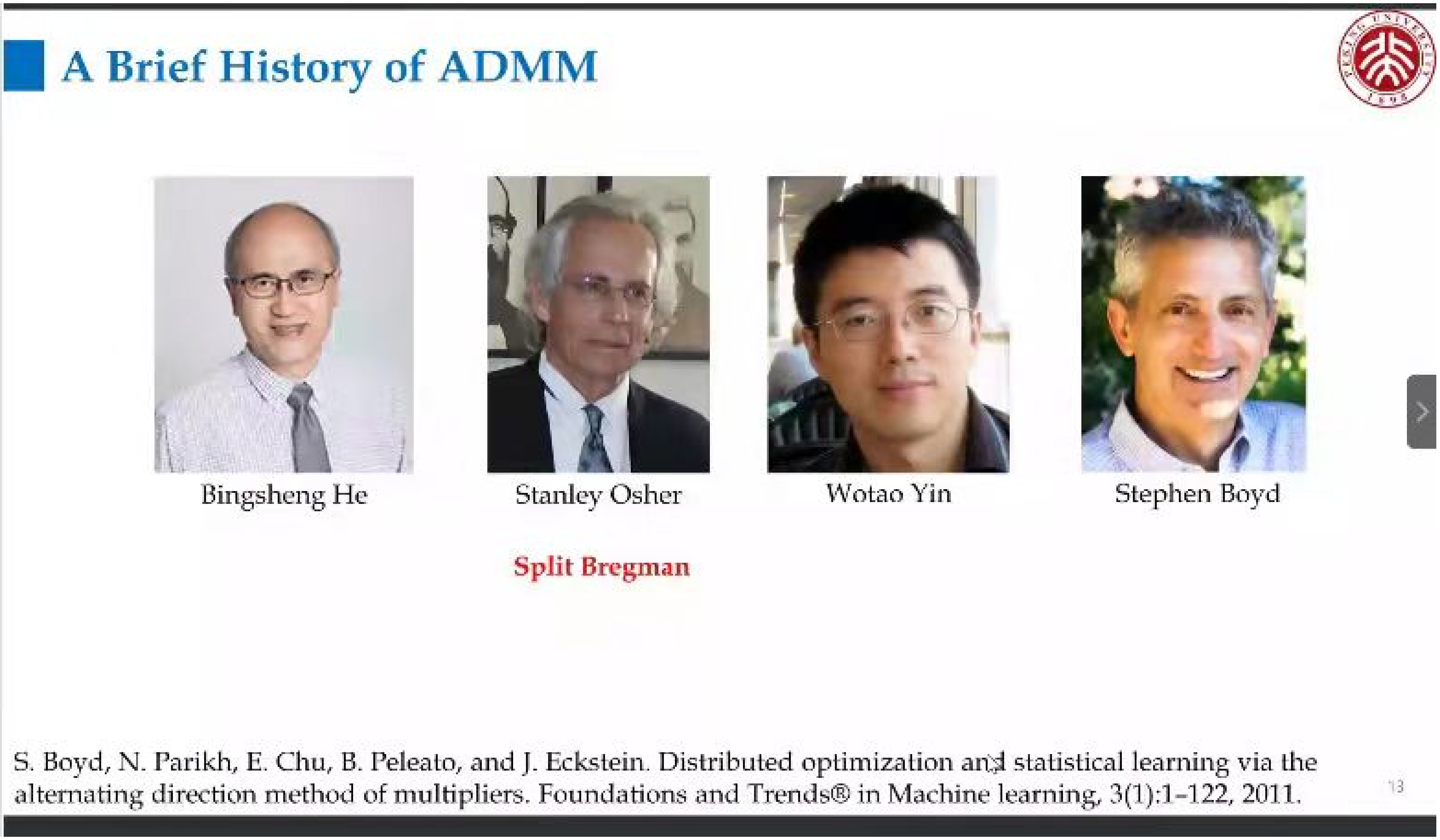
该方法可参阅论文 [10, 8].

过去十多年, ADMM 方法得到了非常广泛的重视!

我们关于ADMM的工作, 在国内也已经得到越来越多的认可.


2021 年7 月, 第三届全国大数据与人工智能科学大会在成都召开.

北京大学智能学院林宙辰教授在大会邀请报告中提到了我的工作。




南方科技大学2015级的一位我教过的学生, 那时已经在西安交通大学读硕士, 看到照片上有我, 赶快照了一张照片传给了我。

ADMM 可以求解两块的问题。我们已经证明：直接推广的ADMM 用来求解三块的问题是不能保证收敛的 [2]！求解三个可分离块的方法可见 [11, 12] 和中文文章 [7, 8]

# 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:

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

## 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

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

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

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

## 3.1 满足统一框架的算法

We consider the $\min - \max$ problem

前一讲曾讨论过这问题, 并用 PDHG 求解

$$\min_x \max_y \{\Phi(x,y) = \theta_1(x) - y^T Ax - \theta_2(y) \,|\, x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{3.4}$$

Let $(x^*, y^*)$ be the solution of (3.4), then we have

根据鞍点的定义

$$(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}, \;\; \Phi(x^*, y) \le \Phi(x^*, y^*) \le \Phi(x, y^*), \;\; \forall\, (x,y) \in \mathcal{X} \times \mathcal{Y}.$$

上面的两个不等式可以写成等价的

$$\left\{ \begin{array}{ll} x^* \in \mathcal{X}, \;\; \Phi(x, y^*) - \Phi(x^*, y^*) \ge 0, \;\; \forall\, x \in \mathcal{X}, & \quad (3.5a) \\ y^* \in \mathcal{Y}, \;\; \Phi(x^*, y^*) - \Phi(x^*, y) \ge 0, \;\; \forall\, y \in \mathcal{Y}. & \quad (3.5b) \end{array} \right.$$

Using the notation of $\Phi(x,y)$, it can be written as 只要把 $\Phi(x,y)$ 的形式填进去

$$\left\{\begin{array}{llll} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x-x^*)^T(-A^T y^*) \ge 0, & \forall\, x \in \mathcal{X}, & (*) \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) \;+\; (y-y^*)^T(\;Ax^*) \;\ge 0, & \forall\, y \in \mathcal{Y}. & (\diamond) \end{array}\right.$$

Furthermore, it can be written as a variational inequality in the compact form:

$$u \in \Omega, \quad \theta(u) - \theta(u^*) + (u-u^*)^T F(u^*) \ge 0, \; \forall\, u \in \Omega, \tag{3.6}$$

where 对上式中任意的 $u \in \Omega$ 分别取 $u = (x, y^*)$ 和 $u = (x^*, y)$, 就得到 $(*)$ 和 $(\diamond)$.

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad F(u) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}, \quad \Omega = \mathcal{X} \times \mathcal{Y}.$$

## The output of Original PDHG algorithm [19] as predictor

For given $(x^k, y^k)$, PDHG [19] produces a pair of $(\tilde{x}^k, \tilde{y}^k)$. First,

$$\tilde{x}^k = \text{argmin}\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, \tag{3.7a}$$

and then we obtain $\tilde{y}^k$ via

$$\tilde{y}^k = \text{argmax}\{\Phi(\tilde{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}. \tag{3.7b}$$

Ignoring the constant term in the objective function, the subproblems (3.7) are reduced to

$$\left\{ \begin{array}{ll} \tilde{x}^k = \text{argmin}\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, & \text{(3.8a)} \\ \tilde{y}^k = \text{argmin}\{\theta_2(y) + y^T A\tilde{x}^k + \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}. & \text{(3.8b)} \end{array} \right.$$

According to the basic lemma, the optimality condition of (3.8a) is $\tilde{x}^k \in \mathcal{X}$ and

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T y^k + r(\tilde{x}^k - x^k)\} \geq 0, \ \forall\, x \in \mathcal{X}. \tag{3.9}$$

Similarly, from (3.8b) we get $\tilde{y}^k \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{A\tilde{x}^k + s(\tilde{y}^k - y^k)\} \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{3.10}$$

Combining (3.9) and (3.10), we have

$$\tilde{u}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T \tilde{y}^k \\ A\tilde{x}^k \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) + A^T(\tilde{y}^k - y^k) \\ s(\tilde{y}^k - y^k) \end{pmatrix} \right\} \geq 0, \quad \forall (x, y) \in \Omega.$$

The compact form is $\tilde{u}^k \in \Omega$,

$$\theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T \{F(\tilde{u}^k) + Q(\tilde{u}^k - u^k)\} \geq 0, \ \forall u \in \Omega, \tag{3.11a}$$

where

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}. \tag{3.11b}$$

对于这样的预测, 我们考虑比较简单的校正

$$u^{k+1} = u^k - M(u^k - \tilde{u}^k) \tag{3.12}$$

校正. 其中 $M$ 为单位上三角矩阵或单位下三角矩阵. 收敛性条件 (3.3)

**校正矩阵 $M$ 为单位上三角矩阵的凑的方法** 取

$$H=\begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix}, \qquad M=\begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix}$$

对任意的 $r,s>0$, 矩阵 $H$ 是正定的, 并且有

$$HM=\begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix}\begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix}=\begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}=Q$$

而对条件 (ii),

$$\begin{aligned} G &= Q^T+Q-M^THM=Q^T+Q-Q^TM \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix}-\begin{pmatrix} rI_n & 0 \\ A & sI_m \end{pmatrix}\begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix}-\begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n & 0 \\ 0 & sI_m-\frac{1}{r}AA^T \end{pmatrix}. \end{aligned}$$

要 $G$ 正定, 必须有 $rs>\|A^TA\|$.

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

$$M^T H M = D.$$

由矩阵方程组解得

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

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

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

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

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

矩阵 $Q^T$ 一般具有某种三角形结构, 下面的方程容易求解.

## 3.2 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*

$$\begin{array}{l} \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \\ \quad \geq \quad \frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac{1}{2}\|v^k-\tilde{v}^k\|_G^2, \quad \forall w\in\Omega. \end{array} \tag{3.13}$$

**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

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (v-\tilde{v}^k)^T H(v^k-v^{k+1}), \ \forall w\in\Omega. \tag{3.14}$$

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 (3.14) with

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

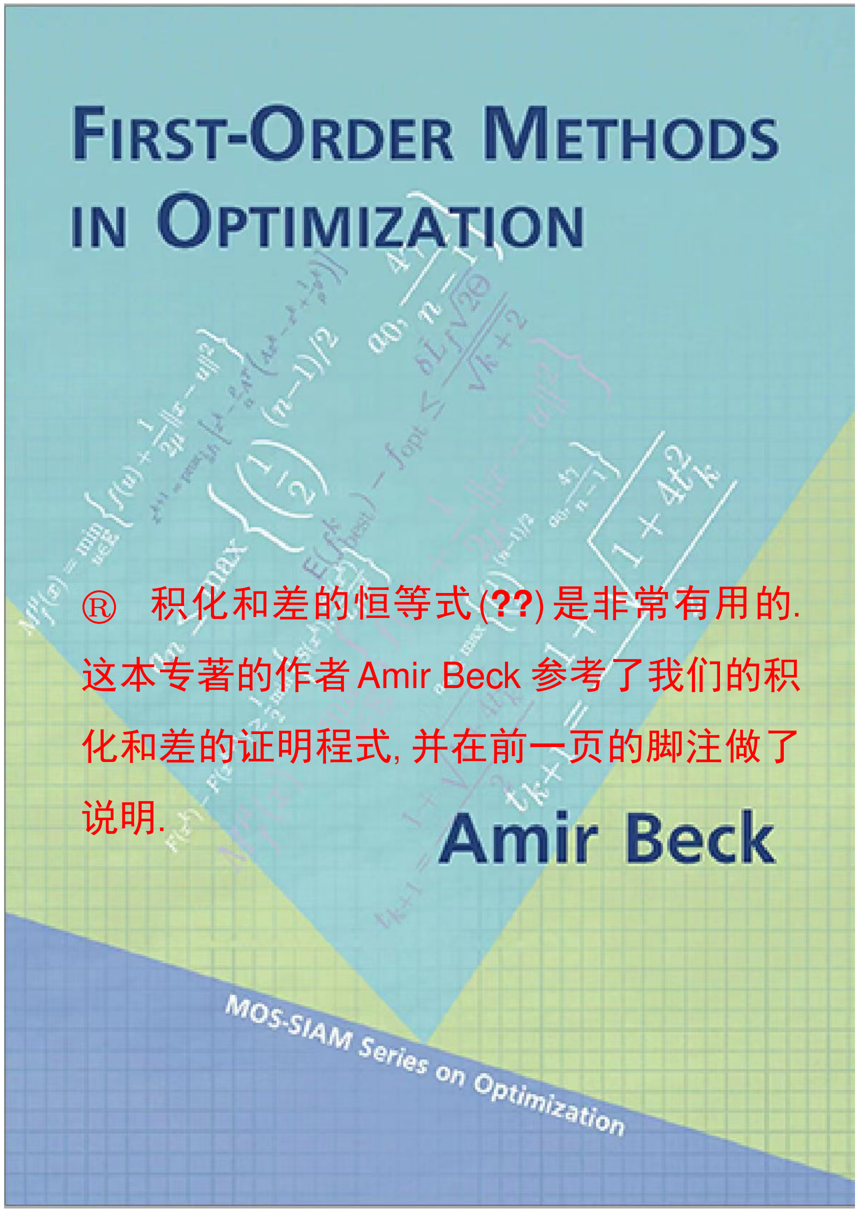
we thus obtain

$$\begin{aligned} & 2(v - \tilde{v}^k)^T H(v^k - v^{k+1}) \\ & \quad = \big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + (\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \end{aligned} \tag{3.15}$$

For the last term of (3.15), using $HM = Q$ and $2v^T Qv = v^T(Q^T + Q)v$, we have

$$\begin{aligned} & \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ & \quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ & \quad \overset{(3.3a)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ & \quad = \quad 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ & \quad = \quad (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\ & \quad \overset{(3.3b)}{=} \quad \|v^k - \tilde{v}^k\|_G^2. \end{aligned} \tag{3.16}$$

Substituting (3.15), (3.16) in (3.14), the assertion of this theorem is proved. □




® 积化和差的恒等式 (**??**) 是非常有用的. 这本专著的作者 Amir Beck 参考了我们的积化和差的证明程式, 并在前一页的脚注做了说明.



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 $\mathbf{x} \in \text{dom}(h_1)$ and $\mathbf{z} \in \text{dom}(h_2)$,

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

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 $\mathbf{x} \in \text{dom}(h_1)$, $\mathbf{z} \in \text{dom}(h_2)$, 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

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

## 3.3 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*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{3.17}$$

**Proof**. Setting $w = w^*$ in (3.13), we get

$$\begin{aligned} & \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ & \quad \ge \quad \|v^k - \tilde{v}^k\|_G^2 + 2\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \end{aligned} \tag{3.18}$$

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

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \|v^k - \tilde{v}^k\|_G^2. \tag{3.19}$$

The assertion (3.17) 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*

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2. \tag{3.20}$$

**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

$$(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} \geq \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \tag{3.21}$$

Finally, by using $\|a\|^2_H - \|b\|^2_H = 2a^T H(a - b) - \|a - b\|^2_H$ and (3.21), we get

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

This is the equivalent form of (3.20) and the proof is complete. □

因此, 统一框架的算法有很好的性质, (3.17) 和 (3.20)

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

和

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

统一框架中的算法是由预测矩阵 $Q$ 和选择的 $D \prec Q^T + Q$ 确定的.

## 3.4 广义 PPA 算法 — 统一框架中的一个特例

求解变分不等式 (3.1) 采用单位步长校正的时候, 如果预测公式

$$\tilde{w}^k \in \Omega, \;\; \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \;\; \forall\, w \in \Omega, \tag{3.22}$$

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

$$D \succ 0, \qquad G \succ 0 \qquad \text{和} \qquad D + G = Q^T + Q, \tag{3.23}$$

再令

$$M = Q^{-T}D \qquad 和 \qquad H = QD^{-1}Q^T. \tag{3.24}$$

则由单位步长校正

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k) \tag{3.25}$$

产生的新的迭代序列$\{v^k\}$满足

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{3.26}$$

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, (3.26) 就变成了

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_D^2, \quad \forall v^* \in \mathcal{V}^*. \tag{3.27}$$

对选定的$D$, 由于

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

因此, (3.27) 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 (3.25)), 上式就了

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{3.28}$$

因此, 统一框架的算法有很好的性质, (3.28) 和 (3.20)

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

和

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

其中

$$H = 2Q[Q^T + Q]^{-1}Q^T, \qquad Q \text{ 是 (3.2a) 中的预测矩阵.}$$

换句话说, 广义 PPA 算法是由预测 (3.2a) 唯一确定的.

# 4 应用：$p$-块可分离凸优化问题的变分不等式

We consider the $p$-block separable convex optimization problem

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i) \;\Big|\; \sum_{i=1}^{p}A_i x_i = b \;(\text{or} \ge b),\;\; x_i\in\mathcal{X}_i\Big\}. \tag{4.1}$$

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

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

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 (4.1) can be written as the following VI:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \qquad \text{(4.2a)}$$

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_p^T\lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}, \qquad \text{(4.2b)}$$

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 $\Omega^*$ the solution set of the VI (4.2).

# 多块问题 (4.2) 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 $\tilde{w}^k \in \Omega$:

$$\left\{\begin{array}{l} \tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\ \tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^T A_2^T \lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\ \qquad\vdots \\ \tilde{x}_i^k \in \arg\min_{x_i \in \mathcal{X}_i}\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\ \qquad\vdots \\ \tilde{x}_p^k \in \arg\min_{x_p \in \mathcal{X}_p}\{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}; \\ \tilde{\lambda}^k = P_\Lambda[\lambda^k - \beta(\sum_{j=1}^{p} A_j\tilde{x}_j^k - b)]. \end{array}\right. \tag{4.3}$$

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.

## 采用 Primal-Dual 预测的预测矩阵

**Analysis for the P-D Prediction** 我们先看 (4.3) 中 $x$ 子问题

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

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

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

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

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{\underline{-A_i^T\tilde{\lambda}^k} + \beta A_i^T(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0. \tag{4.4a}$$

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

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

最优性条件是

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \{ \big( \underline{\textstyle\sum_{j=1}^p A_j \tilde{x}_j^k - b} \big) + \tfrac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \} \geq 0, \quad \forall \lambda \in \Lambda. \quad \text{(4.4b)}$$

Summating (4.4a) and (4.4b), for the predictor $\tilde{w}^k$ generated by (4.3), we have $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \geq (w - \tilde{w}^k)^T Q (w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \quad \text{(4.5a)}$$

where

$$Q = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \quad \text{(4.5b)}$$

# 变量代换后的预测矩阵

The optimization problem (4.1) has been translated to VI (4.2), 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:

$$P = \begin{pmatrix} \sqrt{\beta}A_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta}A_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sqrt{\beta}A_p & 0 \\ 0 & \cdots & \cdots & 0 & (1/\sqrt{\beta})I_m \end{pmatrix}, \quad z = Pw = \begin{pmatrix} \sqrt{\beta}A_1x_1 \\ \sqrt{\beta}A_2x_2 \\ \vdots \\ \sqrt{\beta}A_px_p \\ (1/\sqrt{\beta})\lambda \end{pmatrix}. \tag{4.6}$$

Accordingly, we define

$$\mathcal{Z} = \{z \mid z = Pw\},$$

and

$$\mathcal{Z}^* = \{z^* \mid z^* = Pw^*,\ w^* \in \Omega^*\}.$$

Using the notation $P$ in (4.6), for the matrix $Q$ in (4.5b), we have

$$Q = P^T \mathcal{Q} P, \qquad \text{where} \qquad \mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{4.7}$$

Thus, for the right hand side of (4.5a), we have

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

Then, it follows from (4.5) that we have the following VI for the P-D prediction:

$$\begin{aligned} \tilde{w}^k \in \Omega, \;\; \theta(x) - \theta(\tilde{x}^k) &+ (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ &\geq (z - \tilde{z}^k)^T \mathcal{Q} (z^k - \tilde{z}^k), \;\; \forall\, w \in \Omega. \end{aligned} \tag{4.8}$$

where $\mathcal{Q}$ is given in (4.9).

Using the notation $P$ in (4.6), for the matrix $Q$ in (4.5b), we have

$$Q = P^T \mathcal{Q} P, \qquad \text{where} \qquad \mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{4.9}$$

Thus, for the right hand side of (4.5a), 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 (4.5) that we have the following VI for the P-D prediction:

$$\begin{aligned} \tilde{w}^k \in \Omega, \;\; \theta(x) - \theta(\tilde{x}^k) &+ (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ &\geq (z - \tilde{z}^k)^T \mathcal{Q} (z^k - \tilde{z}^k), \;\; \forall\, w \in \Omega. \end{aligned} \tag{4.10}$$

where $\mathcal{Q}$ is given in (4.9).

# 5 基于预测 (4.10) 的校正方法

**Prediction-Correction Framework for VI (4.2).**

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

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (z - \tilde{z}^k)^T \mathcal{Q}(z^k - \tilde{z}^k), \quad \forall\, w \in \Omega, \tag{5.1}$$

   where the matrix $\mathcal{Q}^T + \mathcal{Q}$ is positive definite.

2. (Correction Step) With the predictor $\tilde{w}^k$ by (5.1) and $\tilde{z}^k = P\tilde{w}^k$, the new iterate $z^{k+1}$ is updated by

$$z^{k+1} = z^k - \mathcal{M}(z^k - \tilde{z}^k), \tag{5.2}$$

   where $\mathcal{M}$ is a non-singular matrix.

## 跟定理 2 和定理 3 和类似, 相应的我们有下面的两个定理.

**定理 4** *For the matrices* $\mathcal{Q}$ (5.1) *and* $\mathcal{M}$ *in* (5.2), *if there is a positive definite matrix* $\mathcal{H}$ *such that*

$$\mathcal{H}\mathcal{M} = \mathcal{Q} \tag{5.3a}$$

*and*

$$\mathcal{G} := \mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0, \tag{5.3b}$$

*then we have*

$$\|z^{k+1} - z^*\|_{\mathcal{H}}^2 \le \|z^k - z^*\|_{\mathcal{H}}^2 - \|z^k - \tilde{z}^k\|_{\mathcal{G}}^2, \quad \forall\, z^* \in \mathcal{Z}^*. \tag{5.4}$$

**定理 5** 如果预测条件 (5.1) 条件满足, 新迭代点由校正 (5.2) 产生, 那么序列 $\{\|z^k - z^{k+1}\|_{\mathcal{H}}\}$ 是单调不增的, 即

$$\|z^{k+1} - z^{k+2}\|_{\mathcal{H}}^2 \le \|z^k - z^{k+1}\|_{\mathcal{H}}^2. \tag{5.5}$$

因此, 变量变换下统一框架的算法有性质(5.4) 和 (5.5), 即

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

和

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

统一框架中的算法是由预测矩阵 $\mathcal{Q}$ 和 选择的 $\mathcal{D} \prec \mathcal{Q}^T + \mathcal{Q}$ 确定的.

如果我们采用一对特殊的 $\mathcal{D}$ 和 $\mathcal{G}$, 使得

$$\mathcal{D} = \mathcal{G} = \frac{1}{2}(\mathcal{Q}^T + \mathcal{Q}) \tag{5.6}$$

那么(5.4) 就变成了

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

对选定的 $\mathcal{D}$, 根据 $\mathcal{D} = \mathcal{M}^T\mathcal{H}\mathcal{M}$, 并利用 (5.2), 上式就成了

$$\|z^{k+1} - z^*\|_{\mathcal{H}}^2 \le \|z^k - z^*\|_{\mathcal{H}}^2 - \|z^k - z^{k+1}\|_{\mathcal{H}}^2, \quad \forall z^* \in \mathcal{Z}^*.$$

上面的不等式和 (5.5) 说明算法就是一个广义的 PPA.

# 6 基于预测 (4.10) 的校正的具体实施

设预测是由 Primal-Dual 预测给出的, 我们得到形如 (4.10) 的变分不等式, 其中

$$\mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{6.1}$$

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

$$\mathcal{L} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix}, \quad \mathcal{I} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{6.2}$$

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

$$\mathcal{E}^T = \begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix}. \tag{6.3}$$

The matrix $\mathcal{Q}$ in (6.1) has the form

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

According to the prediction (4.3), the matrix $\mathcal{Q}$ in (4.9), we have

$$\mathcal{Q}^T = \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E}^T & I_m \end{pmatrix}, \qquad \mathcal{Q}^{-T} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{6.4}$$

and in detail, 请注意 $\mathcal{E}^T\mathcal{L}^{-T} = (I_m, 0, \ldots, 0)$

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

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

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

Conversely, we can also choose

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

and thus get the another correction method.

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

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

we can set

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

or vice versa.

---

Another example,

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

Especially, we can choose $\mathcal{D}=\frac{1}{2}(\mathcal{Q}^T+\mathcal{Q})$ (Generalized PPA). Thus

$$\mathcal{Q}^{-T}\mathcal{D}=\frac{1}{2}\begin{pmatrix}2\mathcal{I} & \mathcal{L}^{-T}\mathcal{E}\\ -\mathcal{E}^T & I_m\end{pmatrix}. \qquad \mathcal{L}^{-T}\mathcal{E}=\begin{pmatrix}0\\ \vdots\\ 0\\ I_m\end{pmatrix}$$

计算用到了 $\mathcal{E}^T\mathcal{L}^{-T}=(I_m,0,\ldots,0)$, 和 $\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E}=I_m$.

把预测-校正的 G-PPA 写在一起：根据提供的 $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ 进行**预测**:

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

再为下一次迭代开始给出新的 $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$ 而进行**校正**:

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

# 7 Conclusions

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

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

[4] R. Glowinski, On alternating direction methods of multipliers: a historical perspective, *Modeling, Simulation and Optimization for Science and Technology*, edited by W. Fitzgibbon, Y. A. Kuznetsov, P. Neittaanmäki and O. Pironneau, Eccomas, Springer, 2014, pp. 59-82.

[5] 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)

[6] 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.

[7] 何炳生, 修正乘子交替方向法求解三个可分离算子的凸优化, 运筹学学报 Vol.19 No.3 (2015) 57-70.

[8] 何炳生, 我和乘子交替方向法20年, 运筹学学报 22 (2018), 1-31.

[9] 何炳生, 利用统一框架设计凸优化的分裂收缩算法, 高等学校计算数学学报, 2022, 44: 1-35.

[10] 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.

[11] 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.

[12] 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.

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

Handbook of of Numerical Analysis ISSN 1570-8659, https://doi.org/10.1016/bs.hna.2022.08.02

[14] B. S. He and X. M. Yuan, On the $O(1/n)$ convergence rate of the alternating direction method, SIAM J. Numerical Analysis, 50 (2012), 700-709.

[15] B. S. He and X. M. Yuan, 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]

[19] M. Zhu and T. F. Chan, An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration, CAM Report 08-34, UCLA, Los Angeles, CA, 2008.