从 PPA-ADMM 到广义邻近点算法

数学之美, 不是纯数学的专利. 为应用服务的最优化方法研究, 同样可以追求简单与统一. 简单, 他人才能看懂使用; 统一, 自己才有美的享受.

发现了其中的数学之美, 研究才变得满怀激情和欲罢不能.

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

西安电子科技大学　2023 年 12 月 6 日

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

Department of Mathematics, Nanjing University,
Nanjing, 210093, China E-mail: hebma@nju.edu.cn

Current Research Areas:
Mathematical Programming, Numerical Optimization,
Variational Inequalities, Projection and contraction methods for VI,
ADMM-like splitting contraction methods for convex optimization

Education:
PhD: Applied Mathematics, The University of Wuerzburg, Germany, 1986
Thesis Advisor: Professor Dr. Josef Stoer
BSc: Computational Mathematics, Nanjing University, 1981

Work:
2015-2020 Professor, Dept. of Math, Southern University of Science and Technology (SUSTech)
2013-2015 Professor, School of Management Science and Engineering, Nanjing University
1997~2013 Professor, Department of Mathematics, Nanjing University
1992~1997 Associate Professor, Department of Mathematics,Nanjing University

我会尽量维护自己的主页, 不断修正、更新自己学术的体验.

My Thinkings: 1. 关门感想 2. 说说我的主要研究兴趣 — 兼谈华罗庚推广优选法对我的影响
3. 说说我的主要研究兴趣(续) --- 我们在ADMM类方法的主要工作 4. 古稀回首
5. 两页纸给出ADMM 收敛性证明 6. 两页纸简述我职业生涯中的主要研究工作
7. 两页纸描述古稀之后的研究进展: 从好不容易凑出一个方法到并不费力构造一簇算法
8. 两页PPT讲述求解线性约束凸优化问题预测-校正的广义邻近点(Generaized PPA)算法

《数学文化》2020 年/ 第 11 卷第 2 期刊登的我的自述: 四十年上下求索 一份珍贵的回忆材料

My Talks: 比较系统的知识建议阅读第3个报告. 也建议阅读最近的一些系列报告

For more systematic knowledge, it is recommended to read Talk 3, which is written in English.

20. 2023 年11月在 曲阜师大两次（共计四小时）课程的PPT 两次课程的内容摘要 课程I 课程II
19. 2023 年10月在天元数学东北中心八次课程的汇总讲义 前言与目录 I II III IV V VI VII VIII
18. 2023年1月在华南师大《华人数学家论坛》的报告 — 凸优化分裂收缩算法统一框架的最新进展
17. 2022年11月在西电《最优化前沿论坛》的报告 — 从好不容易凑出一个算法到并不费劲构造一簇算法
16. 2022 年7月南理工《计算机》方向暑期班六讲摘要 和六讲的PPT I II III IV V VI
15. 利用预测-校正统一框架构造凸优化的分裂收缩算法(由预测矩阵构造校正矩阵) (ArXiv: 2204.11522)
14. 2022 年元月南师大数科院系列报告B站视频辅助材料 A B C D E F G H I J K L
13. ADMM 类分裂收缩算法的一些最新进展 统一框架下Balanced-ALM 便于向多块推广的ADMM
12. 均困平衡的增广拉格朗日乘子法 — Balanced ALM (一类新的增广拉格朗日乘子法ArXiv: 2108.08554)

我的报告的 PDF 文件, 一般都可以在我的主页上查到.

报告提纲 和 算法共有的两条漂亮而又重要的性质.

1. 线性约束的凸优化问题和相应的变分不等式
2. 变分不等式邻近点算法 (PPA) 的主要性质
3. 交替方向法 (ADMM) 的主要收敛性质
4. 线性约束凸优化问题分裂收缩算法的统一框架
5. 预测-校正的广义邻近点算法 (Generalized PPA)
6. 多块可分离问题的 Gauss 型预测的变分不等式
7. 统一框架中变量替换下的一般校正方法
8. 多块可分离问题的广义邻近点算法

我们重点介绍迭代序列具备下述两条漂亮性质的算法

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

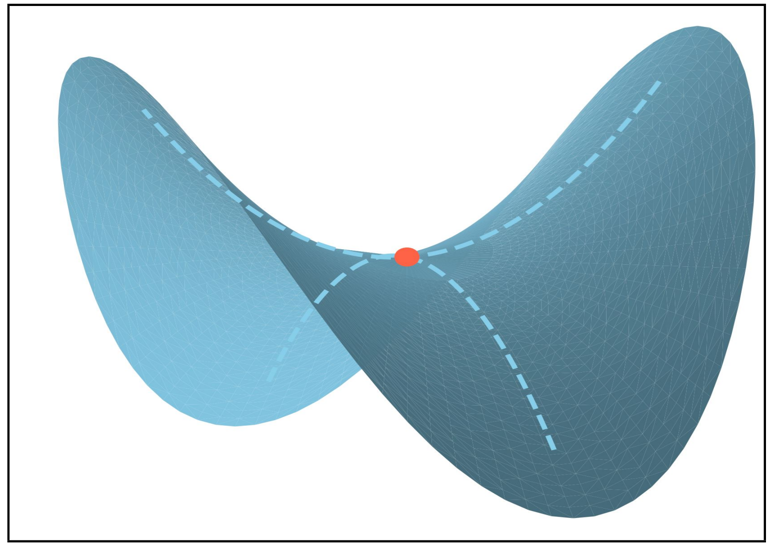
$$\|w^k - w^{k+1}\|_H^2 \le \|w^{k-1} - w^k\|_H^2.$$

PPT 包含所有的证明和必要的注释, 内容比较详细, 便于阅读理解.

1 线性约束的凸优化问题和相应的变分不等式

设$\theta(u)$为凸函数, $\mathcal{A} \in \Re^{m\times n}, b \in \Re^m$. $\mathcal{U}$为闭凸集. 考虑线性约束的凸优化问题

$$\min\{\theta(u) \mid \mathcal{A}u = b,\ u \in \mathcal{U}\}. \tag{1.1}$$



凸优化问题 (1.1) 的拉格朗日函数是定义在$\mathcal{U} \times \Re^m$上的

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b). \tag{1.2}$$

如果一对$(u^*,\lambda^*) \in \mathcal{U} \times \Re^m$ 满足

$$L(u^*,\lambda) \le L(u^*,\lambda^*) \le L(u,\lambda^*),\ \ \forall (u,\lambda) \in \mathcal{U} \times \Re^m,$$

就称其为 Lagrange 函数 (1.2) 的鞍点. 上面的两个不等式可以写成

$$\left\{ \begin{array}{ll} u^* \in \mathcal{U}, & L(u,\lambda^*) - L(u^*,\lambda^*) \ge 0,\ \ \forall\, u \in \mathcal{U}, \qquad (1.3a) \\ \lambda^* \in \Re^m, & L(u^*,\lambda^*) - L(u^*,\lambda) \ge 0,\ \ \forall\, \lambda \in \Re^m. \qquad (1.3b) \end{array} \right.$$

根据 Lagrange 函数 $L(u,\lambda)$ 的定义(见(1.2)),

$$\begin{aligned} & L(u,\lambda^*) - L(u^*,\lambda^*) \\ = \quad & [\theta(u) - (\lambda^*)^T(\mathcal{A}u - b)] - [\theta(u^*) - (\lambda^*)^T(\mathcal{A}u^* - b)] \\ = \quad & \theta(u) - \theta(u^*) + (u-u^*)^T(-\mathcal{A}^T\lambda^*), \end{aligned}$$

由 (1.3a) 得到

$$u^* \in \mathcal{U}, \quad \theta(u) - \theta(u^*) + (u-u^*)^T(-\mathcal{A}^T\lambda^*) \ge 0, \quad \forall\, u \in \mathcal{U}. \tag{1.4}$$

同样, 对 (1.3b), 因为

$$\begin{aligned} & L(u^*,\lambda^*) - L(u^*,\lambda) \\ = \quad & [\theta(u^*) - (\lambda^*)^T(\mathcal{A}u^* - b)] - [\theta(u^*) - (\lambda)^T(\mathcal{A}u^* - b)] \\ = \quad & (\lambda - \lambda^*)^T(\mathcal{A}u^* - b), \end{aligned}$$

我们有

$$\lambda^* \in \Re^m, \quad (\lambda-\lambda^*)^T(\mathcal{A}u^* - b) \ge 0, \quad \forall\, \lambda \in \Re^m. \tag{1.5}$$

注意到表达式 (1.5) (任何向量跟向量 $(\mathcal{A}u^* - b)$ 的内积都非负) 等价于

$$\mathcal{A}u^* - b = 0.$$

将 (1.4) 和 (1.5) 写在一起得到:

$$\begin{cases} u^* \in \mathcal{U}, & \theta(u) - \theta(u^*) + (u - u^*)^T(-\mathcal{A}^T\lambda^*) \geq 0, \quad \forall\, u \in \mathcal{U}, \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T(\mathcal{A}u^* - b) \geq 0, \quad \forall\, \lambda \in \Re^m. \end{cases}$$

采用更紧凑的格式, 鞍点可以被刻画成变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{1.6a}$$

的解点, 这里

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \;\; \text{和} \;\; \Omega = \mathcal{U} \times \Re^m. \tag{1.6b}$$

分别将 $w = (u, \lambda^*)$ 和 $w = (u^*, \lambda)$ 代入 (1.6), 就得到 (1.4) 和 (1.5). 注意到 (1.6b) 中的仿射算子 $F(w)$ 为

$$F(w) = \begin{pmatrix} 0 & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} u \\ \lambda \end{pmatrix} - \begin{pmatrix} 0 \\ b \end{pmatrix}.$$

由于其中的矩阵是反对称的, 因此有

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) \equiv 0. \tag{1.7}$$

在后面的证明中, 我们常常要用到

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

其根据就是(1.7). 因此, 当 $\tilde{w} \in \Omega$ 时, 就有

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

下面最后的不等式的依据是 $\tilde{w} \in \Omega$ 和 (1.6).

$$= \theta(\tilde{u}) - \theta(u^*) + (\tilde{w} - w^*)^T F(w^*) \ \geq \ 0, \quad \forall w^* \in \Omega^*. \tag{1.8}$$

Two block separable convex optimization

两个可分离块的凸优化问题

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b, \; x \in \mathcal{X}, y \in \mathcal{Y}\} \tag{1.9}$$

是问题 (1.1) 的一个特例, 其中

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad \mathcal{U} = \mathcal{X} \times \mathcal{Y}, \quad \mathcal{A} = (A, B).$$

相应的变分不等式就是

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega. \tag{1.10a}$$

其中

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{1.10b}$$

这样就把可分离线性约束凸优化问题 (1.9), 转换成了变分不等式 (1.10).

变分不等式 (1.10) 中的仿射算子 $F(w)$ 的矩阵仍然是斜对称的, 同样有

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) \equiv 0.$$

因此在变分不等式求解的证明中也可以常常应用 (1.8) 这样的性质.

2 变分不等式 PPA 算法的主要性质

考虑一般的单调混合变分不等式问题（变分不等式 (**??**) 是 (2.1) 的特例）

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{2.1}$$

PPA 算法. 求解变分不等式 (2.1), 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

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq \ (w - w^{k+1})^T H(w^k - w^{k+1}), \ \ \forall\, w \in \Omega. \end{aligned} \tag{2.2}$$

w^{k+1} 是变分不等式问题 (2.1) 的解 $\quad \Longleftrightarrow \quad$ (2.2) 中的 $w^k = w^{k+1}$.

引理 1 设 $a, b \in \Re^n$, $H \in \Re^{n\times n}$ 是正定矩阵. 若 $b^T H(a-b) \geq 0$, 那么我们有

向量 x 的 H-模的定义：$\|x\|_H = (x^T H x)^{1/2}$.

$$\|b\|_H^2 \leq \|a\|_H^2 - \|a-b\|_H^2. \tag{2.3}$$

证明 由恒等式 $\|b\|_H^2 = \|a\|_H^2 - 2b^T H(a-b) - \|a-b\|_H^2$ 和假设条件 $b^T H(a-b) \geq 0$, 引理结论能直接得到. $\quad \Box$

定理 1 求解变分不等式(2.1), 对给定的正定矩阵 $H \succ 0$ 和向量 w^k, 如果 w^{k+1} 是由 (2.2) 提供的, 那么有

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2, \quad \forall w^* \in \Omega^*. \tag{2.4}$$

证明 将 (2.2) 中任意的 $w \in \Omega$ 设为 w^*, 就有

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

因为 w^* 是最优点, $w^{k+1} \in \Omega$, 根据 (1.8), 上式右端非负, 所以有

$$(w^{k+1} - w^*)^T H(w^k - w^{k+1}) \ge 0. \tag{2.5}$$

在引理 1 中设 $a = w^k - w^*$ 和 $b = w^{k+1} - w^*$, 便得到定理之结论. $\square$

定理 2 设 $\{w^k\}$ 是求解变分不等式 (2.1) 的 H-模下的 PPA 算法 (2.2) 产生的迭代序列. 那么我们有

$$\|w^k - w^{k+1}\|_H^2 \le \|w^{k-1} - w^k\|_H^2. \tag{2.6}$$

证明 首先, 将 (2.2) 中的 w 设为 w^k, 我们有

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

将 (2.2) 中的 k 改为 $k-1$, 就有

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

再将上面不等式中的 w 设为 w^{k+1} (相应的部分向量 u 就成了 u^{k+1}), 得到

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

将有下划线的两式加在一起并利用 $(w^k - w^{k+1})^T (F(w^k) - F(w^{k+1})) = 0$, 就得到

$$(w^k - w^{k+1})^T H\{(w^{k-1} - w^k) - (w^k - w^{k+1})\} \ge 0.$$

设 $a = (w^{k-1} - w^k)$, $b = (w^k - w^{k+1})$, 上面的不等式就是 $b^T H(a - b) \ge 0$. 由引理 1 得到

$$\|w^k - w^{k+1}\|_H^2 \le \|w^{k-1} - w^k\|_H^2 - \|(w^{k-1} - w^k) - (w^k - w^{k+1})\|_H^2.$$

从上式得到结论 (2.6) 成立. 定理得证. $\square$

定理 1 和定理 2 的结论:

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

和

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

是 PPA 算法的两条主要性质.

注意到这两条性质在 H 半正定时同样成立.

根据定理 1 的结论 (2.4), 我们有

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

对 $k = 0, 1, \ldots, t$ 累加, 就有

$$\sum_{k=0}^{\infty} \|w^k - w^{k+1}\|_H^2 \le \|w^0 - w^*\|_H, \quad \forall w^* \in \Omega^*.$$

再结合定理 1 的结论 (2.6) 便有

$$\|w^t - w^{t+1}\|_H^2 \le \frac{1}{t+1}\|w^0 - w^*\|_H^2, \quad \forall\, w^* \in \Omega^*.$$

这是证明 PPA 算法收敛速率的主要依据.

应用: 平行求解子问题的 PPA 算法

> 求解两个可分离块问题 (**??**) 相应的变分不等式 (**??**).
> 根据 PPA 算法的要求, 设计的右端矩阵为对称正定.

按 **Primal-Dual** 顺序实现的邻近点算法:

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (w - w^{k+1})^T H(w^k - w^{k+1}), \ \forall w \in \Omega, \tag{2.7a}$$

其中

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & 0 & A^T \\ 0 & \beta B^T B + \delta I_{n_2} & B^T \\ A & B & \frac{2}{\beta} I_m \end{pmatrix} \tag{2.7b}$$

正定是因为

$$\begin{pmatrix} \beta A^T A + \delta I_{n_1} & A^T \\ A & \frac{1}{\beta} I_m \end{pmatrix} \succ 0, \qquad \begin{pmatrix} \beta B^T B + \delta I_{n_2} & B^T \\ B & \frac{1}{\beta} I_m \end{pmatrix} \succ 0.$$

邻近点算法 (2.7) 的具体形式写开来就是 $w^{k+1} \in \Omega$, 对任意的 $w \in \Omega$, 有

$$\begin{cases} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ \qquad\qquad \{\underline{-A^T\lambda^{k+1}} + \boldsymbol{(\beta A^T A + \delta I_{n_1})}(x^{k+1} - x^k)\boldsymbol{+A^T}(\lambda^{k+1} - \lambda^k)\} \geq 0, \\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \\ \qquad\qquad \{\underline{-B^T\lambda^{k+1}} + \boldsymbol{(\beta B^T B + \delta I_{n_2})}(y^{k+1} - y^k)\boldsymbol{+B^T}(\lambda^{k+1} - \lambda^k)\} \geq 0, \\ \quad (\lambda - \lambda^{k+1})^T\{(\underline{Ax^{k+1} + By^{k+1} - b})\\ \qquad\qquad\qquad \boldsymbol{+A}(x^{k+1} - x^k)\boldsymbol{+B}(y^{k+1} - y^k) + \boldsymbol{(2/\beta)}\,(\lambda^{k+1} - \lambda^k) = 0. \end{cases}$$

把上面的式子归并一下, 得到 $w^{k+1} \in \Omega$, 对任意的 $w \in \Omega$, 有

$$\begin{cases} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^k + (\beta A^T A + \delta I_{n_1})(x^{k+1} - x^k)\} \geq 0, \\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^k + (\beta B^T B + \delta I_{n_2})(y^{k+1} - y^k)\} \geq 0, \\ [2(Ax^{k+1} + By^{k+1} - b) - (Ax^k + By^k - b)] + (2/\beta)(\lambda^{k+1} - \lambda^k) = 0. \end{cases} \tag{2.8}$$

(2.8) 的每一行分别是 x, y 和 λ 子问题的最优性条件.

关于最优性原理(优化问题和小变分不等式之间的关系), 我们有如下的定理.

定理 3 (一阶最优性原理) 设 $\mathcal{X} \subset \Re^n$ 是闭凸集, $\theta(x)$ 和 $f(x)$ 都是凸函数, 其中 $f(x)$ 在包含 $\mathcal{X}$ 的一个开集上可微. 记 x^* 是凸优化问题 $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ 的解. 那么

$$x^* \in \operatorname{argmin}\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{2.9}$$

的充分必要条件是

优化问题和变分不等式的关系, 是方法分析的基础

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \mathcal{X}. \tag{2.10}$$

根据定理3, (2.8) 中的 x^{k+1}, y^{k+1} 和 λ^{k+1} 可以通过求解可分离的单个子问题

$$\left\{ \begin{array}{ll} x^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T \lambda^k \\ + \frac{1}{2}\beta\|A(x - x^k)\|^2 + \frac{1}{2}\delta\|x - x^k\|^2 \end{array} \middle| \, x \in \mathcal{X} \right\} & \text{(2.11a)} \\ y^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k \\ + \frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\} & \text{(2.11b)} \\ \lambda^{k+1} = \lambda^k - \frac{1}{2}\beta\big[2(Ax^{k+1} + By^{k+1} - b) - (Ax^k + By^k - b)\big] & \text{(2.11c)} \end{array} \right.$$

求得.

通过这个例子, 读者可以大概了解到论文 [1, 3, 12, 15, 17, 18] 的基本思路.

University of Colorado **Boulder** Technical Report, Department of Applied Mathematics

The Chen-Teboulle algorithm is the proximal point algorithm

Stephen Becker *

November 22, 2011; posted August 13, 2019

Abstract

We revisit the ... on the step-size p...

Recent works such as [HY12] have proposed a very simple yet powerful technique for analyzing optimization methods.

1 Background

Recent works such as [HY12] have proposed a very simple yet powerful technique for analyzing optimization methods. The idea consists simply of working with a different norm in the *product* Hilbert space. We fix an inner product $\langle x, y\rangle$ on $\mathcal{H} \times \mathcal{H}^*$. Instead of defining the norm to be the induced norm, we define the primal norm as follows (and this induces the dual norm)

$$\|x\|_V = \sqrt{\langle Vx, x\rangle} = \sqrt{\langle x, x\rangle_V}, \quad \|y\|_V^* = \|y\|_{V^{-1}} = \sqrt{\langle y, V^{-1}y\rangle} = \sqrt{\langle y, y\rangle_{V^{-1}}}$$

for any Hermitian positive definite $V \in \mathcal{B}(\mathcal{H}, \mathcal{H})$; we write this condition as $V \succ 0$. For finite dimensional spaces $\mathcal{H}$, this means that V is a positive definite matrix.

Stephen Becker, The Chen-Teboulle algorithm is the proximal point algorithm, 2011, arXiv:1908.03633[math.OC]

8年后被重提, 是凑PPA的方法应该得到重视?

构造 PPA 方法的策略是很多的. 我们把 (2.7) 中的矩阵 H 稍微变动一下, 就会得到先更新 λ, 再平行求解 x, y 子问题的 PPA 方法, 下面是简要说明.

按 **Dual-Primal** 顺序实现的邻近点算法:

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T F(w^{k+1}) \\ &\geq (w - w^{k+1})^T H(w^k - w^{k+1}), \ \forall w \in \Omega, \end{aligned} \tag{2.12a}$$

其中

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & 0 & -A^T \\ 0 & \beta B^T B + \delta I_{n_2} & -B^T \\ -A & -B & \frac{2}{\beta} I_m \end{pmatrix}. \tag{2.12b}$$

矩阵H正定是因为

$$\begin{pmatrix} \beta A^T A + \delta I_{n_1} & -A^T \\ -A & \frac{1}{\beta} I_m \end{pmatrix} \succ 0, \qquad \begin{pmatrix} \beta B^T B + \delta I_{n_2} & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \succ 0.$$

邻近点算法 (2.12) 的具体形式写开来就是 $w^{k+1} \in \Omega$, 对任意的 $w \in \Omega$, 有

$$\begin{cases} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ \qquad \{\underline{-A^T\lambda^{k+1}} + \boldsymbol{(\beta A^T A + \delta I_{n_1})}(x^{k+1} - x^k)\boldsymbol{-A^T}(\lambda^{k+1} - \lambda^k)\} \ge 0, \\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \\ \qquad \{\underline{-B^T\lambda^{k+1}} + \boldsymbol{(\beta B^T B + \delta I_{n_2})}(y^{k+1} - y^k)\boldsymbol{-B^T}(\lambda^{k+1} - \lambda^k)\} \ge 0, \\ \quad (\lambda - \lambda^{k+1})^T\{(\underline{Ax^{k+1} + By^{k+1} - b})\\ \qquad\qquad \boldsymbol{-A}(x^{k+1} - x^k)\boldsymbol{-B}(y^{k+1} - y^k) + \boldsymbol{(2/\beta)}\,(\lambda^{k+1} - \lambda^k) = 0. \end{cases}$$

把上面的式子归并一下, 得到

$$\begin{cases} x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ \qquad \{(\beta A^T A + \delta I_{n_1})(x^{k+1} - x^k) - A^T(2\lambda^{k+1} - \lambda^k)\} \ge 0, \ \forall x \in \mathcal{X} \\ y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \\ \qquad \{(\beta B^T B + \delta I_{n_2})(y^{k+1} - y^k) - B^T(2\lambda^{k+1} - \lambda^k)\} \ge 0, \forall y \in \mathcal{Y} \\ \quad (\lambda - \lambda^{k+1})^T\{(Ax^k + By^k - b) + (2/\beta)(\lambda^{k+1} - \lambda^k)\} \ge 0, \quad \forall \lambda \in \Re^m. \end{cases} \tag{2.13}$$

(2.13) 中的 λ 子问题相当于

$$(Ax^k + By^k - b) + (2/\beta)(\lambda^{k+1} - \lambda^k) = 0.$$

由已知的 (x^k, y^k, λ^k), 更新乘子

$$\lambda^{k+1} = \lambda^k - \frac{1}{2}\beta(Ax^k + By^k - b). \tag{2.14}$$

根据定理 3, (2.13) 中的 λ^{k+1}, x^{k+1} 和 y^{k+1} 可以通过求解可分离的子问题

$$\begin{cases} x^{k+1} \in \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T[2\lambda^{k+1} - \lambda^k] \\ +\frac{1}{2}\beta\|A(x - x^k)\|^2 + \frac{1}{2}\delta\|x - x^k\|^2 \end{array} \middle| x \in \mathcal{X} \right\} & \text{(2.15a)} \\[6pt] y^{k+1} \in \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T[2\lambda^{k+1} - \lambda^k] \\ +\frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \end{array} \middle| y \in \mathcal{Y} \right\}. & \text{(2.15b)} \end{cases}$$

这里介绍的 PPA 类方法, 可以从 [6] 中找到.

利用变分不等式 (VI) 和邻近点算法 (PPA), 更自由地设计 PPA 类分裂收缩算法

3 ADMM 算法的主要性质

ADMM 也有 **PPA** 的在定理 1 和定理 2 中陈述的两个漂亮的结论.

§1 中我们已经把两块可分离凸优化问题

$$\min\{\theta_1(x) + \theta_2(y) | Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\} \tag{3.1}$$

转换成的变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega. \tag{3.2}$$

其中

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

和

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

问题 (3.1) 的增广拉格朗日函数是

$$\mathcal{L}_\beta^{[2]}(x,y,\lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax+By-b) + \frac{\beta}{2}\|Ax+By-b\|^2. \tag{3.3}$$

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

$$\left\{\begin{array}{lcl} x^{k+1} & \in & \arg\min\bigl\{\mathcal{L}_\beta^{[2]}(x,y^k,\lambda^k) \bigm| x\in\mathcal{X}\bigr\}, \\ y^{k+1} & \in & \arg\min\bigl\{\mathcal{L}_\beta^{[2]}(x^{k+1},y,\lambda^k) \bigm| y\in\mathcal{Y}\bigr\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1}+By^{k+1}-b) \end{array}\right. \tag{3.4}$$

求得 $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})$. 注意到改动子问题目标函数中的常数项以后解是不变的. 因此, ADMM 的 k 次迭代可以写成

$$\left\{\begin{array}{lcl} x^{k+1} & \in & \arg\min\bigl\{\theta_1(x) - x^TA^T\lambda^k + \frac{1}{2}\beta\|Ax+By^k-b\|^2 \bigm| x\in\mathcal{X}\bigr\}, \\ y^{k+1} & \in & \arg\min\bigl\{\theta_2(y) - y^TB^T\lambda^k + \frac{1}{2}\beta\|Ax^{k+1}+By-b\|^2 \bigm| y\in\mathcal{Y}\bigr\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1}+By^{k+1}-b). \end{array}\right. \tag{3.5}$$

由于迭代是从给定的 $v^k=(y^k,\lambda^k)$ 开始的, 而 x^{k+1} 是凭 $v^k=(y^k,\lambda^k)$ 计算出来的, 我们把ADMM方法中的 x 称为中间变量, 把 $v=(y,\lambda)$ 称为核心变量.

Analysis 根据一阶最优性原理（也就是定理3）, ADMM (3.5) 的 x, y 子问题分别满足

$$\begin{aligned} x^{k+1}\in\mathcal{X}, \quad & \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T \\ & \{-A^T\lambda^k+\beta A^T\big(Ax^{k+1}+By^k-b\big)\}\geq 0,\ \forall\, x\in\mathcal{X} \end{aligned} \tag{3.6a}$$

和

$$\begin{aligned} y^{k+1}\in\mathcal{Y}, \quad & \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T \\ & \{-B^T\lambda^k+\beta B^T\big(Ax^{k+1}+By^{k+1}-b\big)\}\geq 0,\ \forall\, y\in\mathcal{Y}. \end{aligned} \tag{3.6b}$$

以 ADMM 的乘子 λ 的更新公式

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

代入 (3.6) (消去其中的 λ^k), 我们分别得到

$$\begin{aligned} x^{k+1}\in\mathcal{X}, \quad & \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T \\ & \{-A^T\lambda^{k+1}+\beta A^TB(y^k-y^{k+1})\}\geq 0,\ \forall\, x\in\mathcal{X}, \end{aligned} \tag{3.7a}$$

和

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{3.7b}$$

将 (3.7) 写成紧凑的形式：$u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$,

$$\begin{aligned}\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \Bigg\{ &\begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} \\ &+ \beta \begin{pmatrix} A^T B \\ 0 \end{pmatrix} (y^k - y^{k+1}) \Bigg\} \geq 0, \ \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \end{aligned} \tag{3.8}$$

再把上式改写成

$$\begin{aligned}\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \Bigg\{ &\begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} + \beta \begin{pmatrix} A^T B \\ B^T B \end{pmatrix} (y^k - y^{k+1}) \\ &+ \begin{pmatrix} 0 & 0 \\ 0 & \beta B^T B \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \Bigg\} \geq 0, \ \forall\, (x, y) \in \mathcal{X} \times \mathcal{Y}. \end{aligned} \tag{3.9}$$

然后, 我们有如下的引理:

引理 2 求解*VI*(3.2), 设对给定的(y^k,λ^k), $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})\in\Omega$是由交替方向法(3.5)生成的. 我们有

$$\theta(u)-\theta(u^{k+1})+\begin{pmatrix} x-x^{k+1}\\ y-y^{k+1}\\ \lambda-\lambda^{k+1}\end{pmatrix}^T\left\{\begin{pmatrix} -A^T\lambda^{k+1}\\ -B^T\lambda^{k+1}\\ Ax^{k+1}+By^{k+1}-b\end{pmatrix}+\beta\begin{pmatrix} A^T\\ B^T\\ 0\end{pmatrix}B(y^k-y^{k+1})\right.$$
$$\left.+\begin{pmatrix} 0 & 0 & 0\\ 0 & \beta B^TB & 0\\ 0 & 0 & \frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix} x^{k+1}-x^k\\ y^{k+1}-y^k\\ \lambda^{k+1}-\lambda^k\end{pmatrix}\right\}\geq 0,\ \forall w\in\Omega. \quad (3.10)$$

证明 ADMM的乘子更新公式$\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b)$改写成

$$(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)=0,$$

进而有

$$(\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\geq 0,\quad \forall\lambda\in\Re^m. \quad (3.11)$$

将上式加到(3.9), 就得到引理之结论(3.10). □

注意到(3.10)中黑色部分是(3.9). 蓝色部分就是(3.11).

为了方便, 我们定义

$$v=\begin{pmatrix} y\\ \lambda \end{pmatrix},\quad \mathcal{V}^*=\{(y^*,\lambda^*)\,|\,w^*\in\Omega^*\}\ \ \text{和}\ \ H=\begin{pmatrix} \beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m \end{pmatrix}. \tag{3.12}$$

这样, 我们可以把 (3.10) 写成

$$\begin{aligned}&\theta(u)-\theta(u^{k+1})+(w-w^{k+1})^TF(w^{k+1})\\ &\quad+\begin{pmatrix} x-x^{k+1}\\ y-y^{k+1} \end{pmatrix}^T\begin{pmatrix} A^T\\ B^T \end{pmatrix}B(y^k-y^{k+1})+(v-v^{k+1})^TH(v^{k+1}-v^k)\ \geq\ 0.\end{aligned} \tag{3.13}$$

从而得到下面的引理:

引理 3 求解 *VI*(3.2), 设对给定的 (y^k,λ^k), $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})\in\Omega$ 是由交替方向法 (3.5) 生成的. 那么, 我们有

$$(v^{k+1}-v^*)^TH(v^k-v^{k+1})\geq(y^k-y^{k+1})^TB^T(\lambda^k-\lambda^{k+1}),\ \forall\,w^*\in\Omega^*, \tag{3.14}$$

其中矩阵 H 由 (3.12) 给出.

证明. 将 (3.13) 中的 w 设为 w^*, 利用 $F(w)$ 的表达式, 我们得到

$$
\begin{aligned}
&(v^{k+1}-v^*)^T H(v^k-v^{k+1}) \\
\geq\;& \begin{pmatrix} x^{k+1}-x^* \\ y^{k+1}-y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} \beta B(y^k-y^{k+1}) \\
&+\underline{\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^T F(w^{k+1})},\ \forall\, w^* \in \Omega^*. \quad (3.15)
\end{aligned}
$$

根据 (1.8), (3.15) 式右端加下划线的部分非负. (3.15) 式右端的第一部分

$$
\begin{aligned}
&\begin{pmatrix} x^{k+1}-x^* \\ y^{k+1}-y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} \beta B(y^k-y^{k+1}) \\
=\;& (y^k-y^{k+1})^T B^T \beta(A,B) \begin{pmatrix} x^{k+1}-x^* \\ y^{k+1}-y^* \end{pmatrix} \\
=\;& (y^k-y^{k+1})^T B^T \beta\big(Ax^{k+1}+By^{k+1}-(Ax^*+By^*)\big) \\
=\;& (y^k-y^{k+1}) B^T \underline{\beta\big(Ax^{k+1}+By^{k+1}-b\big)} \\
=\;& (y^k-y^{k+1}) B^T \underline{(\lambda^k-\lambda^{k+1})}. \qquad \boxed{\text{有下划线的两部分相等}} \quad (3.16)
\end{aligned}
$$

引理的结论直接从 (3.15) 和 (3.16) 得到. □

引理 4 求解 *VI*(3.2), 设对给定的 (y^k, λ^k), $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ 是由交替方向法 (3.5) 生成的. 那么, 我们有

$$(y^k - y^{k+1})^T B^T (\lambda^k - \lambda^{k+1}) \geq 0. \tag{3.17}$$

证明. 因为 (3.7b) 对 k-次迭代有

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{3.18}$$

同理, 对前一次迭代就有

$$\theta_2(y) - \theta_2(y^k) + (y - y^k)^T \{-B^T \lambda^k\} \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{3.19}$$

分别在 (3.18) 中设 $y = y^k$ 和 (3.19) 中设 $y = y^{k+1}$, 然后将两式相加, 就得不等式 (3.17), 引理得证. □

将 (3.17) 代入 (3.14), 我们得到

$$(v^{k+1} - v^*)^T H (v^k - v^{k+1}) \geq 0, \quad \forall\, v^* \in \mathcal{V}^*. \tag{3.20}$$

前面我们已经说明，由

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

设 $a = v^k - v^*$, $b = v^{k+1} - v^*$, 就有下面的定理.

定理 4 求解 *VI* (3.2), 设对给定的 (y^k, λ^k), $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ 是由交替方向法 (3.5) 生成的. 我们有

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \ \forall\, v^* \in \mathcal{V}^*. \tag{3.21}$$

除此之外, 我们在 [14] 中证明了 ADMM 的迭代序列 $\{v^k\}$ 具备性质

$$\|v^{k+1} - v^{k+2}\|_H^2 \leq \|v^k - v^{k+1}\|_H^2. \tag{3.22}$$

这条性质可以从下一节的分析中直接得到, 所以我们在这里不做证明.

不等式 (3.21) 和 (3.22) 展示了 ADMM 跟 PPA 有一样很好的性质. 我们注意到, 在一些快速 ADMM 的研究 [2] 中, 都用到了 (3.22) 这条性质.

4 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:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{4.1}$$

Algorithms in a unified framework

A unified Algorithmic Framework for VI (4.1)

统一框架由预测-校正两部分组成

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{4.2a}$$

其中 Q 不一定对称, 但是 $Q^T + Q$ 半正定.

[Correction Step.] 给一个满足收敛条件 (4.3) 的矩阵 M, 通过校正

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k) \tag{4.2b}$$

给出新的 v^{k+1}.

Q 和 M 分别叫做预测矩阵和校正矩阵

Convergence Conditions

对算法框架 (4.2) 中的预测矩阵 Q 和校正矩阵 M, 存在半正定矩阵 H, 使得

$$HM = Q, \tag{4.3a}$$

并且

$$G = Q^T + Q - M^T HM \succeq 0. \tag{4.3b}$$

如果预测矩阵 Q 非奇异, 构造满足收敛条件 (4.3) 的矩阵 H, M, G 是容易的. 由于

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

我们总可以取

$$D \succ 0, \qquad G \succeq 0 \qquad \text{并且} \qquad D + G = Q^T + Q. \tag{4.4}$$

然后令

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

在 Q 非奇异、$D \succ 0$ 的假设下, 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矩阵 M, H 和 G.

满足 (4.4) 的 D 和 G 可以有无穷多的选择

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.

4.1 ADMM 可以转换成统一框架中的预测-校正算法

我们先说明 ADMM 算法 (3.5) 可以转换成统一框架 (4.2) 的一种具体形式, 并满足收敛条件 (4.3). 为此, 我们首先通过

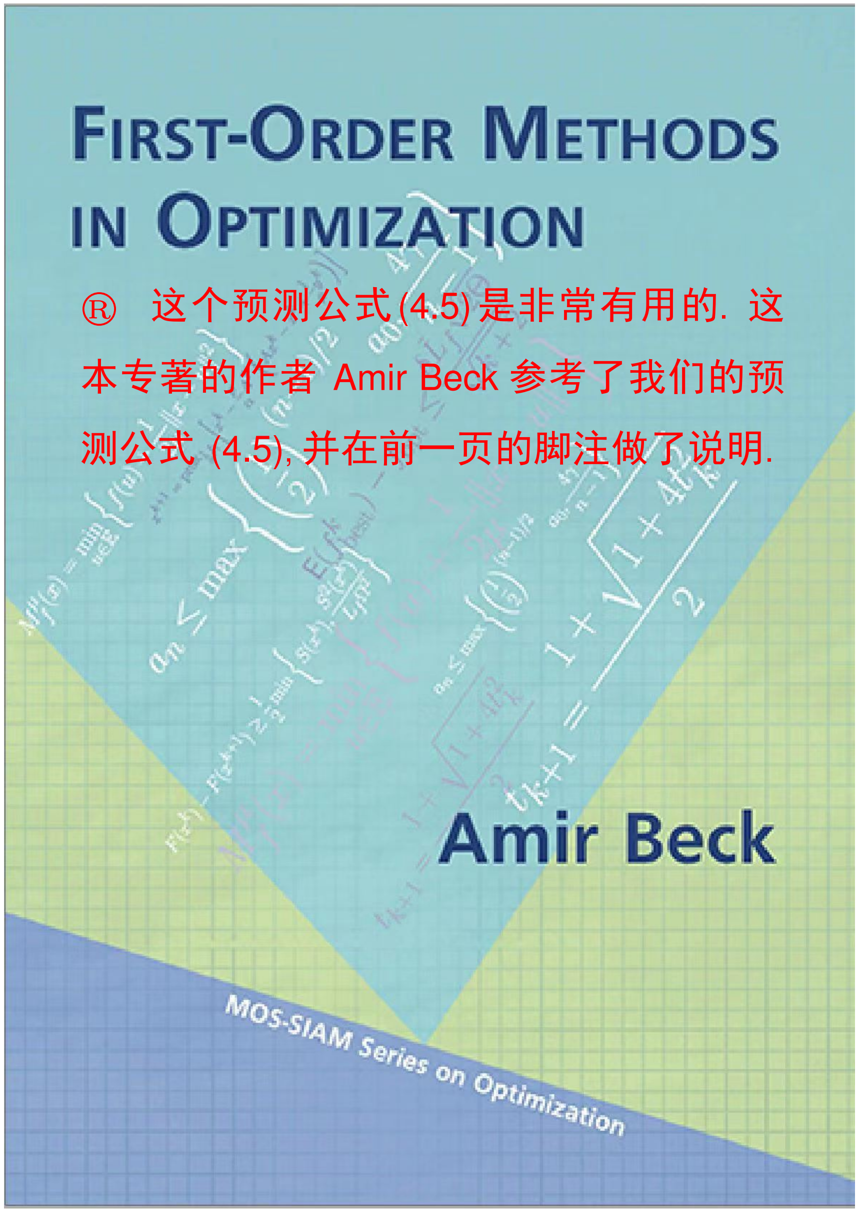
$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}. \tag{4.5}$$

定义预测向量 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$, 其中 x^{k+1} 和 y^{k+1} 是由 ADMM (3.5) 生成的.

$\boxed{\text{我们关于ADMM 几个重要理论结果 [11, 14] 的证明都用到了这种变换}}$

预测 据此, 我们可以把预测写成

$$\text{(预测)} \quad \begin{cases} \tilde{x}^k \in \arg\min\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & \text{(4.6a)} \\ \tilde{y}^k \in \arg\min\{\mathcal{L}_\beta^{[2]}(\tilde{x}^k, y, \lambda^k) \mid y \in \mathcal{Y}\}, & \text{(4.6b)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). & \text{(4.6c)} \end{cases}$$



® 这个预测公式 (4.5) 是非常有用的. 这本专著的作者 Amir Beck 参考了我们的预测公式 (4.5), 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 $\mathbf{x} \in \mathrm{dom}(h_1)$ 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

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

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

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

按照统一框架中 (4.2a) 的形式, 将预测 (4.6) 的结果表述成下面的引理.

引理 5 求解变分不等式 (3.2), 对给定的 v^k, 由 (4.6) 生成的 $\tilde{w}^k$ 满足

$$\tilde{w}^k \in \Omega,\ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k),\ \forall w \in \Omega, \tag{4.7a}$$

其中

$$Q = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix}. \tag{4.7b}$$

证明 根据一阶最优性原理的定理 3, (4.6a) 的最优性条件是

$$\tilde{x}^k \in \mathcal{X},\ \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T \lambda^k + \beta A^T (A\tilde{x}^k + By^k - b)\} \geq 0,\ \forall\, x \in \mathcal{X}.$$

利用 (4.6c) 中的 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, 上式就是 $\boxed{x\text{-子问题后面没有尾巴}}$

$$\tilde{x}^k \in \mathcal{X},\ \ \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{\underline{-A^T \tilde{\lambda}^k}\} \geq 0,\ \ \forall\, x \in \mathcal{X} \tag{4.8a}$$

同理, y-子问题 (4.6b) 的最优性条件可以写成 $\boxed{y\text{-子问题的尾巴是}\,[\beta B^T B(\tilde{y}^k - y^k)]}$

$$\tilde{y}^k \in \mathcal{Y},\ \ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{\underline{-B^T \tilde{\lambda}^k} + \beta B^T B(\tilde{y}^k - y^k)\} \geq 0,\ \forall\, y \in \mathcal{Y}. \tag{4.8b}$$

把生成 $\tilde{\lambda}^k$ 的等式 (4.6c) 写成

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0.$$

这等同于

$\boxed{\lambda\text{-子问题的尾巴是 } [-B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k)]}$

$$\tilde{\lambda}^k \in \Re^m, \quad (\lambda - \tilde{\lambda}^k)^T \{\underline{(A\tilde{x}^k + B\tilde{y}^k - b)} - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \tag{4.8c}$$

将 (4.8a), (4.8b) 和 (4.8c) 写在一起, 利用变分不等式 (3.2) 的表达式, 注意到下划线部分放在一起就是 $F(\tilde{w}^k)$, 预测点 $\tilde{w}^k$ 满足 (4.7), 引理得证. $\square$

校正 交替方向法 (3.5) 是故意分拆成预测-校正的, 我们要回头讨论校正.

引理 6 交替方向法 (3.5) 中的 $v^{k+1} = (y^{k+1}, \lambda^{k+1})$ 和通过 (4.5) 定义的 $\tilde{v}^k = (\tilde{y}^k, \tilde{\lambda}^k)$ 的关系可以写成

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{4.9a}$$

其中

$$M = \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix}. \tag{4.9b}$$

证明 首先, $y^{k+1}=\tilde{y}^k$. 由 (3.5) 给出的 λ^{k+1}. 利用 (4.5) 中 $\tilde{w}^k$ 的定义, 可以表示成

$$\begin{aligned}\lambda^{k+1} &= \lambda^k-\beta(A\tilde{x}^k+B\tilde{y}^k-b)\\ &= \lambda^k-\left[-\beta B(y^k-\tilde{y}^k)+\beta(A\tilde{x}^k+By^k-b)\right]\\ &= \lambda^k-\left[-\beta B(y^k-\tilde{y}^k)+(\lambda^k-\tilde{\lambda}^k)\right].\end{aligned}$$

这样, ADMM 算法 (3.4) 中的 $v^{k+1}=(y^{k+1},\lambda^{k+1})$ 和通过 (4.5) 定义的 $\tilde{v}^k=(\tilde{y}^k,\tilde{\lambda}^k)$ 的关系可以写成

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

这就是 (4.9), 引理得证. $\square$

下面我们只要用 (4.3) 去验证方法的收敛性条件. 首先, 设

$$H=\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I\end{pmatrix}. \qquad H\ \text{只是半正定} \tag{4.10}$$

对于 (4.9b) 中的矩阵 M 和 (4.7b) 中的矩阵 Q, 有

$$HM = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix} = Q, \tag{4.11}$$

条件 (4.3a) 满足. 另外,

$$\begin{aligned} G &= Q^T + Q - M^T HM = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} \beta B^T B & -B^T \\ 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \\ &= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \end{aligned} \tag{4.12}$$

因此, 矩阵 G 是半正定的, 条件 (4.3b) 也满足.

上面的两个引理告诉我们, 求解变分不等式 (3.2) 的经典的交替方向法 (3.5), 可以拆解成按 (4.7) 预测和 (4.9) 校正的方法. 它们分别相当于统一框架中的预测 (4.2a) 和校正 (4.2b).

ADMM 可以拆解成统一框架中的预测-校正方法, 也具有 §4.2 中将要证明的定理的性质

4.2 统一框架算法的收敛性质

我们证明统一框架算法 (4.2) 在收敛性条件 (4.3), 即

$$\boxed{Q^T + Q \succeq 0, \quad H \succeq 0, \quad HM = Q, \quad G = Q^T + Q - M^T HM \succeq 0.}$$

满足下的一些性质.

定理 5 为求解变分不等式 (4.1), 设 $\{v^k\}$, $\{\tilde{w}^k\}$ 是由统一框架的算法 (4.2) 产生的. 假如收敛性条件 (4.3) 满足, 则有

$$\begin{aligned} &\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ \geq \quad & \frac{1}{2}\left(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\right) + \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega. \end{aligned} \tag{4.13}$$

证明. 利用 $Q = HM$ (见(4.3a)) 和校正公式 (4.2b), 对预测 (4.2a) 的右端进行处理: $Q(v^k - \tilde{v}^k) = HM(v^k - \tilde{v}^k) = H(v^k - v^{k+1})$, 因此有

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \quad \forall w \in \Omega. \tag{4.14}$$

对 (4.14) 的右端使用恒等式

$$(a-b)^T H(c-d) = \frac{1}{2}\{\|a-d\|_H^2 - \|a-c\|_H^2\} + \frac{1}{2}\{\|c-b\|_H^2 - \|d-b\|_H^2\}, \tag{4.15}$$

并在其中令

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

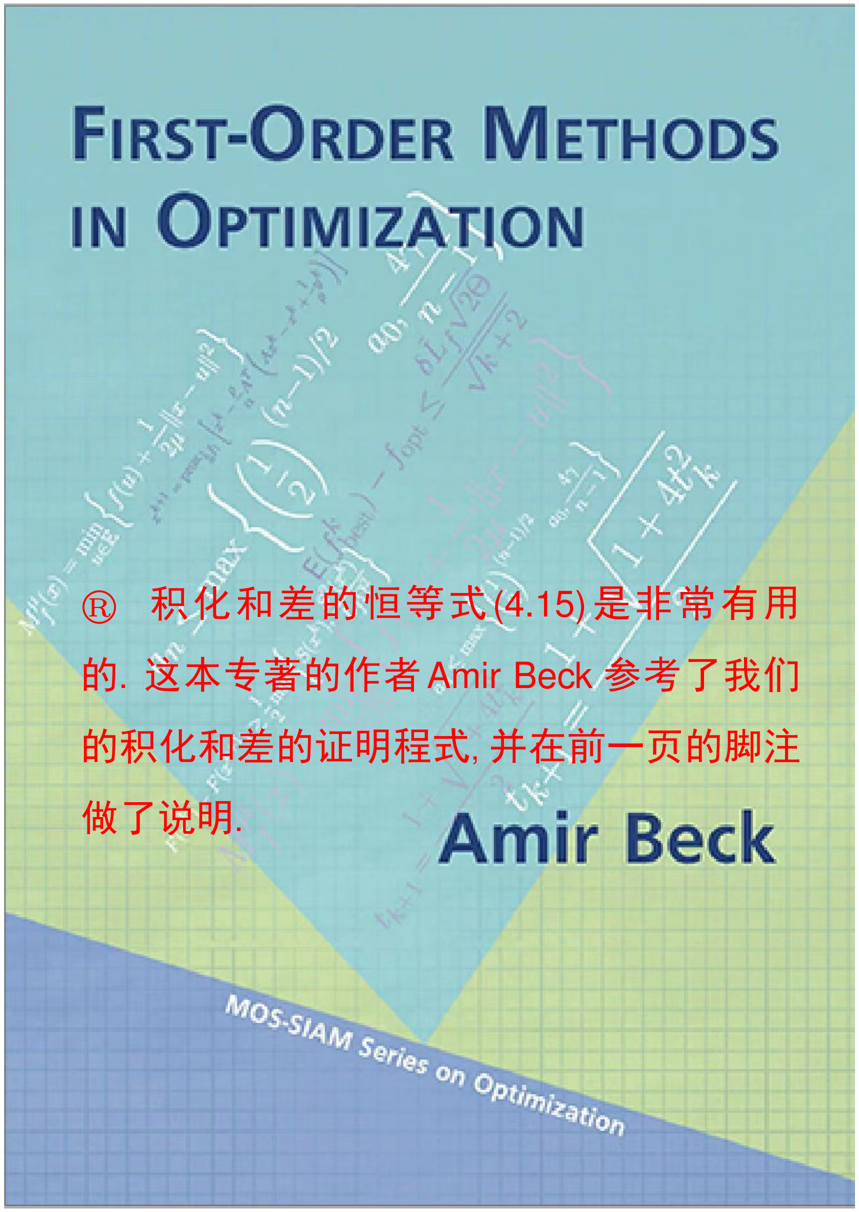
我们因此得到

$$\begin{aligned} & 2(v - \tilde{v}^k)^T H(v^k - v^{k+1}) \\ & \quad = \left(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\right) + (\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \end{aligned} \tag{4.16}$$

对 (4.16) 右端的后一部分, 利用 $HM = Q$ 和 $2v^TQv = v^T(Q^T + Q)v$, 得到

$$\begin{aligned} & \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ & \quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ & \quad \overset{(4.3a)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ & \quad = \quad 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ & \quad = \quad (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\ & \quad \overset{(4.3b)}{=} \quad \|v^k - \tilde{v}^k\|_G^2. \end{aligned} \tag{4.17}$$

将 (4.16), (4.17) 代入 (4.14), 定理的结论得证. $\square$



® 积化和差的恒等式(4.15)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.

15.3. Convergence Analysis of AD-PMM 429

We will use the following notation:

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

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 $\mathbf{y} \in \mathbb{R}^m$,

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

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

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

Using the above identity, we can conclude that

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

as well as

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

and

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

Convergence in a strictly contraction sense

定理 6 为求解变分不等式 (4.1), 设 $\{v^k\}$, $\{\tilde{w}^k\}$ 是由统一框架的算法 (4.2) 产生的. 假如收敛性条件 (4.3) 满足, 则有

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{4.18}$$

证明. 将 (4.13) 中的 w 设为任意确定的 w^*, 我们得到

$$\begin{aligned} &\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ &\quad \ge \quad \|v^k - \tilde{v}^k\|_G^2 + \underline{2\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}}. \end{aligned} \tag{4.19}$$

根据 (1.8), (4.19) 式右端加下划线的部分非负. 因此

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \|v^k - \tilde{v}^k\|_G^2. \tag{4.20}$$

直接得到了 (4.18), 定理得证. $\square$

定理 7 为求解变分不等式(4.1), 设$\{v^k\}$, $\{\tilde{w}^k\}$是由统一框架的算法(4.2)产生的. 假如收敛性条件(4.3)满足, 则有

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2. \tag{4.21}$$

证明 首先, 我们有

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

和

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

分别将上面两式中的w设成$\tilde{w}^{k+1}$和$\tilde{w}^k$, 得到

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

和

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

将上述两个不等式相加, 就有

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

再将 $\{(v^k-\tilde v^k)-(v^{k+1}-\tilde v^{k+1})\}^TQ\{(v^k-\tilde v^k)-(v^{k+1}-\tilde v^{k+1})\}$ 加到上式两边, 得到

$$(v^k-v^{k+1})^TQ\{(v^k-\tilde v^k)-(v^{k+1}-\tilde v^{k+1})\}\ge \tfrac12\|(v^k-\tilde v^k)-(v^{k+1}-\tilde v^{k+1})\|^2_{(Q^T+Q)}.$$

并因此

$$\begin{aligned}&(v^k-v^{k+1})^TH\{(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\}\\ \ge\ &\tfrac12\|(v^k-\tilde v^k)-(v^{k+1}-\tilde v^{k+1})\|^2_{(Q^T+Q)}.\end{aligned}\tag{4.22}$$

最后, 利用 $\|a\|_H^2-\|b\|_H^2=2a^TH(a-b)-\|a-b\|_H^2$ 和 (4.22), 我们得到

$$\begin{aligned}&\|v^k-v^{k+1}\|_H^2-\|v^{k+1}-v^{k+2}\|_H^2\\ =\ &2(v^k-v^{k+1})^TH\{(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\}\\ &-\|(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\|_H^2\\ \overset{(4.22)}{\ge}\ &\|(v^k-\tilde v^k)-(v^{k+1}-\tilde v^{k+1})\|^2_{(Q^T+Q)}-\|(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\|_H^2\\ =\ &\|(v^k-\tilde v^k)-(v^{k+1}-\tilde v^{k+1})\|^2_{(Q^T+Q-M^THM)}\\ =\ &\|(v^k-\tilde v^k)-(v^{k+1}-\tilde v^{k+1})\|^2_{G}.\end{aligned}$$

由上式 (4.21) 成立, 证毕. □

由§4.1的论证, ADMM算法也有定理7的性质

5 预测-校正的广义 PPA 算法

上一节, 为求解变分不等式 (4.1), 我们设计了预测-校正算法 (4.2), 其预测是

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

校正为

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{5.2}$$

在收敛性条件

$$\boxed{Q^T + Q \succeq 0, \quad H \succeq 0, \quad HM = Q, \quad G = Q^T + Q - M^T HM \succeq 0}$$

满足的前提下, 证明了（见定理6）

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.3}$$

和（见定理7）

$$\|v^{k+1} - v^{k+2}\|_H^2 \leq \|v^k - v^{k+1}\|_H^2. \tag{5.4}$$

如果 (5.3) 中右端的 $\|v^k - \tilde{v}^k\|_G^2$ 成了 $\|v^k - v^{k+1}\|_H^2$, 那就是 PPA 的性质了.

定理 8 为求解变分不等式 (4.1), 设 $\{v^k\}$, $\{\tilde{w}^k\}$ 是由统一框架的算法 (4.2) 产生的. 假如收敛性条件 (4.3) 满足, 并且有

$$M^T HM = G, \tag{5.5}$$

那么, 我们有

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.6}$$

证明. 由额外的条件 (5.5), 不等式 (5.3) 可以改写成

$$\begin{aligned}\|v^{k+1} - v^*\|_H^2 \quad &\le \quad \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_{M^THM}^2 \\ &= \quad \|v^k - v^*\|_H^2 - \|\underline{M(v^k - \tilde{v}^k)}\|_H^2, \quad \forall v^* \in \mathcal{V}^*.\end{aligned}$$

再根据校正公式 (5.2), 马上得到定理结论 (5.6). □

换句话说, 有了附加条件 (5.5), 统一框架的算法也具备定理 1 和定理 2 的结论.

定理 8 是在 条件 (4.3) 附加条件 (5.5) 的情况下证明的. 能否构造性地给出呢？

$\boxed{\text{能! 当预测矩阵 } Q \text{ 满足 } Q^T+Q\succ 0 \text{ 的时候.}}$ 我们可以将 Q^T+Q 分拆成

$$D\succ 0, \qquad G\succ 0 \qquad \text{和} \qquad D+G=Q^T+Q, \tag{5.7}$$

再令

$$M=Q^{-T}D \qquad \text{和} \qquad H=QD^{-1}Q^T. \tag{5.8}$$

则有

$$H\succ 0, \qquad HM=(QD^{-1}Q^T)(Q^{-T}D)=Q$$

和

$$M^THM=(DQ^{-1})(QD^{-1}Q^T)(Q^{-T}D)=D. \tag{5.9}$$

因此

$$G=Q^T+Q-M^THM=Q^T+Q-D=G\succ 0.$$

条件 (4.3) 满足, (5.3) 和 (5.4) 都成立.

我们采用一对特殊的 D 和 G, 使得 $D=G=\frac{1}{2}(Q^T+Q)$, 那么, (5.3) 就成了

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-\tilde{v}^k\|_D^2, \quad \forall v^*\in\mathcal{V}^*. \tag{5.10}$$

根据 $D = M^T HM$ (见(5.9)) 和 $M(v^k - \tilde{v}^k) = v^k - v^{k+1}$ (见 (5.2)), 就有

$$\|v^k - \tilde{v}^k\|_D^2 = \|M(v^k - \tilde{v}^k)\|_H^2 = \|v^k - v^{k+1}\|_H^2.$$

这时, (5.10) 就成了

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.11}$$

我们把上述分析结果写成下面的定理.

定理 9 用预测校正方法求解变分不等式(2.1), 设预测(5.1)中的预测矩阵 Q 满足 $Q^T + Q \succ 0$. 若令

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

则由校正公式(5.2) 产生的新的迭代点具有性质(5.11)和(5.4), 其中 $H = Q[\frac{1}{2}(Q^T + Q)]^{-1}Q^T$. 性质 (5.11) 和 (5.4) 是 *PPA* 算法所具有的好性质.

求解变分不等式 (2.1), 我们把迭代序列具有性质 (5.11) 和 (5.4) 的预测-校正方法, 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.

6 多块可分离凸优化问题预测的变分不等式

我们考虑用统一框架的算法和广义 PPA 算法求解 p-块可分离凸优化问题

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i)\ \Big|\ \sum_{i=1}^{p}A_ix_i=b\ (\text{or}\geq b),\quad x_i\in\mathcal{X}_i\Big\}. \tag{6.1}$$

凸优化问题 (6.1) 的拉格朗日函数是定义在 $\Omega=\prod_{i=1}^{p}\mathcal{X}_i\times\Lambda$ 上的

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

其中

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

设 $(x_1^*,\ldots,x_p^*,\lambda^*)\in\Omega$ 是 Lagrange 函数的鞍点, 则有

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

凸优化问题 (6.1) 的鞍点等价于下面变分不等式的解:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{6.2a}$$

其中

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_p^T\lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}, \tag{6.2b}$$

并有

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

我们仍然用 Ω^* 表示变分不等式 (6.2) 的解集. 同样, 这里的仿射算子有 $(w - \tilde{w})^T (F(w) - F(\tilde{w})) \equiv 0$. 因此, 当 $\tilde{w} \in \Omega$ 时, 就有

$$\theta(\tilde{x}) - \theta(x^*) + (\tilde{w} - w^*)^T F(\tilde{w}) \ge 0, \quad \forall w^* \in \Omega^*. \tag{6.3}$$

多块问题 (6.2) 的 PRIMAL-DUAL 预测

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 $\tilde{w}^k \in \Omega$:

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^T A_1^T\lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\[2mm]
\tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^T A_2^T\lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\[2mm]
\qquad\qquad \vdots \\[2mm]
\tilde{x}_i^k \in \arg\min_{x_i\in\mathcal{X}_i}\{\theta_i(x_i) - x_i^T A_i^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\[2mm]
\qquad\qquad \vdots \\[2mm]
\tilde{x}_p^k \in \arg\min_{x_p\in\mathcal{X}_p}\{\theta_p(x_p) - x_p^T A_p^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}; \\[2mm]
\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^{p} A_j\tilde{x}_j^k - b\big)\big].
\end{array}\right. \tag{6.4}$$

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.

6.1 采用 Primal-Dual 预测的预测变分不等式

Analysis for the P-D Prediction 我们先看 (6.4) 中 x 子问题

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

根据最优性原理(定理3, 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

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{\underline{-A_i^T\tilde{\lambda}^k} + \beta A_i^T\Big(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)\Big) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0. \tag{6.5a}$$

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

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

最优性条件是

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \Big\{ \big(\underline{\textstyle\sum_{j=1}^p A_j \tilde{x}_j^k - b} \big) + \tfrac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \Big\} \geq 0, \quad \forall \lambda \in \Lambda. \quad \text{(6.5b)}$$

将 (6.5a) 和 (6.5b) 组装在一起, 我们有

引理 7 为求解变分不等式(6.2), 从给定的 $(A_1 x_1^k, A_2 x_2^k, \cdots, A_p x_p^k, \lambda^k)$ 出发, 由 (6.4) 生成的 $\tilde{w}^k \in \Omega$ 满足

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \geq (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \quad \text{(6.6a)}$$

其中

$$Q = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \quad \text{(6.6b)}$$

换句话说: 通过预测, 我们得到一个以 $\tilde{w}^k$ 为解的变分不等式 (6.6).

6.2 变量代换下的预测矩阵

我们已经把凸优化问题 (6.1) 转换成变分不等式 (6.2), 即

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

为方便分析, 我们定义:

$$P = \begin{pmatrix} \sqrt{\beta}A_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta}A_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sqrt{\beta}A_p & 0 \\ 0 & \cdots & \cdots & 0 & (1/\sqrt{\beta})I_m \end{pmatrix}, \quad \xi = Pw = \begin{pmatrix} \sqrt{\beta}A_1x_1 \\ \sqrt{\beta}A_2x_2 \\ \vdots \\ \sqrt{\beta}A_px_p \\ (1/\sqrt{\beta})\lambda \end{pmatrix}. \tag{6.7}$$

同时定义

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

和

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

利用 (6.7) 中定义的矩阵 P, 对 (6.6b) 中的矩阵 Q, 我们有

$$Q = P^T \mathcal{Q} P, \qquad \text{其中} \qquad \mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{6.8}$$

因此, (6.6a) 的右端可以写成

矩阵 $\mathcal{Q}$ 是一个容易求逆的矩阵和一个广义秩一矩阵的和

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

利用变换 (6.7), 刻画预测 (6.4) 的变分不等式 (6.6) 可以改写成

$$\begin{aligned} \tilde{w}^k \in \Omega, \quad & \theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \\ & \geq (\xi-\tilde{\xi}^k)^T \mathcal{Q}(\xi^k-\tilde{\xi}^k), \quad \forall\, w \in \Omega. \end{aligned} \tag{6.9}$$

其中矩阵 $\mathcal{Q}$ 由 (6.8) 给出.

$(\mathcal{Q}^T + \mathcal{Q})$ 是正定的.

7 变量替换下的预测-校正方法

我们仍然考虑§6 中线性约束的多块可分离凸优化问题 (6.1). 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, 生成的预测点 $\tilde{w}^k\in\Omega$ 满足

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

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. 利用上一节的变换 (6.7), 把预测 (7.1) 改写成 $\tilde{w}^k\in\Omega$,

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

其中 $Q=P^T\mathcal{Q}P$. 利用记号

$$\mathcal{L}=\begin{pmatrix} I_m & 0 & \cdots & 0\\ I_m & I_m & \ddots & \vdots\\ \vdots & & \ddots & 0\\ I_m & I_m & \cdots & I_m\end{pmatrix},\quad \mathcal{I}=\begin{pmatrix} I_m & 0 & \cdots & 0\\ 0 & I_m & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & I_m\end{pmatrix}, \tag{7.3}$$

和 $1 \times p$ 块状矩阵

$$\mathcal{E}^T = \begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix}. \tag{7.4}$$

可以把 (6.8) 中的矩阵 $\mathcal{Q}$ 写成

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ 0 & I_m \end{pmatrix} \quad \text{因此} \quad \mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & 2I_m \end{pmatrix} \tag{7.5}$$

并且

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} \mathcal{E} \\ I_m \end{pmatrix} \begin{pmatrix} \mathcal{E}^T & I_m \end{pmatrix} \succ 0. \tag{7.6}$$

在 $\mathcal{Q}$ 非对称但 (7.6) 满足的时候, 必须采用必要的校正. 总可以选两个矩阵 $\mathcal{D}$ 和 $\mathcal{G}$, 使得

$$\mathcal{D} \succ 0, \quad \mathcal{G} \succ 0, \quad \text{和} \quad \mathcal{D} + \mathcal{G} = \mathcal{Q}^T + \mathcal{Q}. \tag{7.7}$$

注意到 $\mathcal{Q}^T$ 是容易求逆的.

$$\mathcal{Q}^{-T} = \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E}^T & I_m \end{pmatrix}^{-1} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{7.8}$$

根据前一讲的分析, 我们有如下的定理.

定理 10 设预测点 $\tilde{\xi}^k$ 满足条件 (7.2), 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 (7.7) 成立.

$$\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D} \tag{7.9}$$

那么, 利用矩阵 (7.9) 校正

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k), \tag{7.10}$$

产生的 ξ^{k+1} 满足

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2, \quad \forall \xi^* \in \Xi^*, \tag{7.11}$$

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

算法具有的另一条重要性质: 序列 $\{\|\xi^k - \xi^{k+1}\|_{\mathcal{H}}\}$ 是单调不增的.

下面的定理和定理 7 类似, 证明方法也完全相同.

定理 11 如果预测点 $\tilde{\xi}^k$ 满足条件 (7.2), 那么, 由校正 (7.10) 产生的新的迭代点 ξ^{k+1} 满足

$$\|\xi^{k+1} - \xi^{k+2}\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^{k+1}\|_{\mathcal{H}}^2. \tag{7.12}$$

证明. 首先, 我们证明迭代序列满足

$$\begin{aligned}
&(\xi^k - \xi^{k+1})^T \mathcal{H}[(\xi^k - \xi^{k+1}) - (\xi^{k+1} - \xi^{k+2})] \\
\ge\ & \frac{1}{2}\|(\xi^k - \tilde{\xi}^k) - (\xi^{k+1} - \tilde{\xi}^{k+1})\|_{(\mathcal{Q}^T+\mathcal{Q})}^2.
\end{aligned} \tag{7.13}$$

将预测 (7.2) 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

$$\theta(\tilde{x}^k) - \theta(\tilde{x}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \ge (\tilde{\xi}^k - \tilde{\xi}^{k+1})^T \mathcal{Q}(\xi^{k+1} - \tilde{\xi}^{k+1}). \tag{7.14}$$

将预测 (7.2) 式中任意的 w 设为 $\tilde{w}^{k+1}$, 就有

$$\theta(\tilde{x}^{k+1}) - \theta(\tilde{x}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \ge (\tilde{\xi}^{k+1} - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k). \tag{7.15}$$

将 (7.14), (7.15) 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

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

在上式左端利用 $\mathcal{Q} = \mathcal{H}\mathcal{M}$ 和校正公式 (7.10), 就得到 (7.13).

然后, 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

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

利用 (7.13) 替换上面等式右端的第一项, 得到

$$\begin{aligned}\|\xi^k-\xi^{k+1}\|_{\mathcal{H}}^2-\|\xi^{k+1}-\xi^{k+2}\|_{\mathcal{H}}^2 \geq \|(\xi^k-\tilde{\xi}^k)-(\xi^{k+1}-\tilde{\xi}^{k+1})\|_{(\mathcal{Q}^T+\mathcal{Q})}^2 \\ -\|(\xi^k-\xi^{k+1})-(\xi^{k+1}-\xi^{k+2})\|_{\mathcal{H}}^2. \quad \text{(7.16)}\end{aligned}$$

用校正公式 (7.10) 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 $(\mathcal{Q}^T+\mathcal{Q})-\mathcal{M}^T\mathcal{H}\mathcal{M}=\mathcal{G}\succeq 0$, (7.16) 右端非负, 定理结论得证. □

由于 (见 (7.5))

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

可以用多种办法将其分解成两个正定矩阵. 例如

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

可以任取其中之一为$\mathcal{D}$. 由于

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

RR

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

校正 (7.10) 通过

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

是特别容易实现的. 文献 [9, 16] 给出了一些具体方法.

8 变量代换下的广义 PPA 算法

我们对由 (7.2) 决定的串型预测, 给出广义邻近点算法的校正公式.

广义邻近点算法就是选

$$\mathcal{D} = \mathcal{G} = \tfrac{1}{2}(\mathcal{Q}^T + \mathcal{Q}) \tag{8.1}$$

那么, (7.11) 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 (7.10), 上式就成了

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \xi^{k+1}\|_{\mathcal{H}}^2, \quad \forall \xi^* \in \Xi^*. \tag{8.2}$$

在广义邻近点算法(Generalized PPA)中, 校正矩阵 $\mathcal{M}$ 是由 (7.2) 中的预测矩阵 $\mathcal{Q}$ 唯一确定的. 不等式 (8.2) 和 (7.12)

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

说明, 变量替换下的广义 PPA 算法同样具备和 PPA 算法的性质 (2.4) 和 (2.6).

我们将这些经过预测-校正, 迭代序列又具有 PPA 性质 (2.4) 和 (2.6) 的算法, 称为广义邻近点算法.

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

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} \mathcal{L} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} 0 & \mathcal{E} \\ 0 & 0 \end{pmatrix}. \tag{8.3}$$

由于

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

和

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

容易计算得到

$$\mathcal{M}=\frac{1}{2}(\mathcal{I}_{p+1}+\mathcal{Q}^{-T}\mathcal{Q})=\frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & I_m \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{8.4}$$

在 (7.10) 中采用这个 $\mathcal{M}$ 做校正, 结合前一讲的变换, 校正可以写成等价的

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix}=\begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix}-\frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & \frac{1}{\beta}I_m \\ -\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix}\begin{pmatrix} A_1x_1^{k}-A_1\tilde{x}_1^{k} \\ A_2x_2^{k}-A_2\tilde{x}_2^{k} \\ \vdots \\ A_px_p^{k}-A_p\tilde{x}_p^{k} \\ \lambda^{k}-\tilde{\lambda}^{k} \end{pmatrix}. \tag{8.5}$$

得到 $(A_1x_1^{k+1},A_2x_2^{k+1},\cdots,A_px_p^{k+1},\lambda^{k+1})$, 就可以开始一次新的预测 (6.4).

由 (6.4) 预测, (8.5) 校正的方法是一个 **Generalized Proximal Point Algorithm** .

9 Conclusions

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. 我们的工作有自己的特色.

- 首先, 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.
- 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.
- 根据统一框架, 我们已经从好不容易凑出一个方法, 到了并不费劲构造一簇算法.
- 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 本身是个有竞争力的好方法.
- 对非凸问题有什么想法？ 怎样设计更好的预测？ 怎样提高一阶方法的收敛速度？

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

[5] 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

[6] 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.

[7] 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.

[8] 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.

[9] B. S. He, S. J. Xu and X. M. Yuan, Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints, Handbook of Numerical Analysis, 24

(2023) 511-557. or arXiv:2107.01897v2[math.OC].

[10] 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.

[11] 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.

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

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

[14] 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.

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

[16] 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]

[17] B. S. He, X. M. Yuan and W. X. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, Comput. Optim. Appl. 56 (2013), 559 – 572.

[18] S. J. Xu, A dual-primal balanced augmented Lagrangian method for lineraly constrained convex programming, J. Appl. Math. and Computing 69 (2023), 1015-1035