典型凸优化问题分裂收缩算法讲座

各讲的主要内容和结论简介

何 炳 生　　南京大学数学系

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

江苏省研究生视觉计算与可信人工智能暑期学校

2022 年 7 月 10 – 15 日

江苏省研究生视觉计算与可信人工智能暑期学校安排我做个系列讲座, 能向年轻朋友介绍自己的研究工作, 年过七旬的我非常珍惜这个机会. 《单调变分不等式的投影收缩算法》和《凸优化的分裂收缩算法》是我一生开展的主要研究工作. 这个系列讲座主要介绍后一类工作: 最近 10 多年来我们对典型凸优化问题求解方法开展的自成系统又颇具特色的研究。

变分不等式 (VI) 和邻近点算法 (PPA) 是我们开展凸优化方法研究的两大法宝. VI 是瞎子爬山判别是否到达了顶点的数学表达形式, PPA 是步步为营稳扎稳打的求解策略. 在约束凸优化问题拉格朗日函数的鞍点和变分不等式的解点等价的基础上, 我们提出了求解这类问题的预测–校正算法的统一框架。

这类方法包括：(1): 处理鞍点问题的原始-对偶 CP-PPA 方法和均困的增广拉格朗日乘子法；(2): 两个可分离块问题的乘子交替方向法 (ADMM) 及其线性化方法, 以及提高效率的 PPA 型算法和降低子问题难度的均困方法；(3): 求解多个可分离块凸优化问题的 Gauss 型预测-校正方法. 从算法结构上说, 就像求解线性方程组的高斯消去法（也被称之为追赶法）一样, 预测就像往前追, 校正犹如往后赶。

提供各讲的PPT 以后, 根据越读越薄的要求, 我把讲座的主要内容和结论梳理一下。

连续优化中一些代表性数学模型

1. 鞍点问题 $\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\{\Phi(x,y)=\theta_1(x)-y^TAx-\theta_2(y)\}$

2. 线性约束的凸优化问题 $\min\{\theta(x)|Ax=b\,(\text{or}\geq b),\ x\in\mathcal{X}\}$

3. 结构型凸优化 $\min\{\theta_1(x)+\theta_2(y)|Ax+By=b,\ x\in\mathcal{X},\ y\in\mathcal{Y}\}$

4. 多块可分离凸优化 $\min\{\sum_{i=1}^p\theta_i(x_i)|\sum_{i=1}^p A_ix_i=b, x_i\in\mathcal{X}_i\}$

变分不等式(VI) 是瞎子爬山的数学表达形式

邻近点算法(PPA) 是步步为营 稳扎稳打的求解方法.

变分不等式和邻近点算法是分析和设计凸优化方法的两大法宝.

1 凸优化及其在变分不等式框架下的邻近点算法

相关概念我们在[33, 34]做过介绍. 凸优化问题

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

的拉格朗日函数是定义在 $\Omega = \mathcal{U} \times \Re^m$ 上的

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

把满足条件 $(u^*, \lambda^*) \in \mathcal{U} \times \Re^m$,

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

的鞍点转换成变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{1.3}$$

的解点, 其中

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

其中的仿射算子 F, 中的矩阵的是反对称的, 满足 $(w - \tilde{w})^T\big(F(w) - F(\tilde{w})\big) \equiv 0.$

我们定义了求解变分不等式(1.3)的邻近点算法 $H \succ 0$ 是给定的正定矩阵.

从给定的 w^k 出发, 求得

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (w - w^{k+1})^T H(w^k - w^{k+1}), \quad \forall\, w \in \Omega, \tag{1.4}$$

w^{k+1} 就叫做 PPA 算法第 k 步的邻近点 **Proximal Point.**

邻近点算法的收缩性质

收缩性质 (1.5) 只要把 (1.4) 中的 w 设为 w^* 就能得到

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2, \quad \forall w^* \in \Omega^*. \tag{1.5}$$

核心变量 (essential variable) 的邻近点算法

设 v 是 w 的部分量（也可以是 $v = w$）, 有了 v^k 就可以开始一次迭代. 求得

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (v - v^{k+1})^T H(v^k - v^{k+1}), \quad \forall\, w \in \Omega, \tag{1.6}$$

w^{k+1} 同样叫做 PPA 算法的邻近点.

核心变量邻近点算法的收缩性质

只要把 (1.6) 中的 w 设为 w^* 就能得到 (1.7)

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{1.7}$$

其中

$$\mathcal{V}^* = \{v^* \text{ is a subvector of } w^* \mid w^* \in \Omega^*\}.$$

$\boxed{\textbf{延伸的邻近点算法}}$ 把 (1.6) 的输出记为 $\tilde{w}^k$, 则有

$$\begin{aligned}\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \quad \forall\, w \in \Omega.\end{aligned} \tag{1.8}$$

将任意的 w 以 w^* 代入, 就有

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \geq 0, \quad \forall v^* \in \mathcal{V}^*,$$

及等价的

$$\big\{(v^k - v^*) - (v^k - \tilde{v})\big\}^T H(v^k - \tilde{v}^k) \geq 0, \quad \forall v^* \in \mathcal{V}^*.$$

这样便有

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_H^2, \quad \forall v^* \in \mathcal{V}^*.$$

根据上式, 以平凡延伸校正

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \in (0,2) \tag{1.9}$$

得到的核心变量的新的迭代点 v^{k+1} 满足

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \alpha(2-\alpha)\|v^k - \tilde{v}^k\|_H^2 \quad \forall v^* \in \mathcal{V}^*. \tag{1.10}$$

收缩不等式 (1.5), (1.7) 和 (1.10) 是证明 PPA 算法收敛的关键不等式 [3, 9, 21, 26].

2 单块线性约束凸优化问题的 PPA 算法和均困的增广拉格朗日乘子法

鞍点问题

$$\min_x \max_y \{\Phi(x,y) = \theta_1(x) - y^T Ax - \theta_2(y) \,|\, x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{2.1}$$

相应的变分不等式

$$u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u-u^*)^T F(u^*) \geq 0, \ \forall\, u \in \Omega, \tag{2.2}$$

其中

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

PDHG 方法 [4, 32]

特点是子问题中的二次函数形式简单

$$\left\{ \begin{array}{ll} x^{k+1} = \operatorname{argmin}\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, & \text{(2.3a)} \\ y^{k+1} = \operatorname{argmin}\{\theta_2(y) + y^T A x^{k+1} + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. & \text{(2.3b)} \end{array} \right.$$

对应的变分不等式是 $u^{k+1}\in\Omega$, 并有

$$\theta(u)-\theta(u^{k+1})+(u-u^{k+1})^T\{F(u^{k+1})+Q(u^{k+1}-u^k)\}\ge 0,\quad \forall\, u\in\Omega. \tag{2.4}$$

其中

$$Q=\begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \qquad \text{is not symmetric.}$$

这样的方法不能保证收敛.

如果能实现 $u^{k+1}\in\Omega$ 和

我们首先在 [21] 中把 (2.6) 解释成变分不等式的 PPA 算法

$$\theta(u)-\theta(u^{k+1})+(u-u^{k+1})^T\{F(u^{k+1})+H(u^{k+1}-u^k)\}\ge 0,\quad \forall\, u\in\Omega. \tag{2.5}$$

其中

$$H=\begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \qquad \text{并且} \qquad rs>\|A^TA\|$$

就是个收敛的 PPA 算法 [21]. 为此, 只要对 (2.3) 略做改造, 变成：

$$\left\{\begin{array}{ll} x^{k+1}=\operatorname{argmin}\{\theta_1(x)-x^TA^Ty^k+\frac{r}{2}\|x-x^k\|^2\,|\,x\in\mathcal{X}\}, & \text{(2.6a)}\\ y^{k+1}=\operatorname{argmin}\{\theta_2(y)+y^TA[2x^{k+1}-x^k]+\frac{s}{2}\|y-y^k\|^2\,|\,y\in\mathcal{Y}\}. & \text{(2.6b)} \end{array}\right.$$

既然如此, 我们也可以通过合理预设正定矩阵　构造 PPA 算法

例如, 对可以转换成 $\min\max$ 问题的凸优化问题

$$\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\}$$

相应的变分不等式是

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, u \in \Omega, \tag{2.7a}$$

其中

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \ \text{and} \quad \Omega = \mathcal{X} \times \Re^m. \tag{2.7b}$$

譬如说, 可以采取

- x 子问题中二次函数形式简单, 但有 $rs > \|A^TA\|$ 的要求.

$$\boxed{H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}, \qquad H = \begin{pmatrix} rI_n & -A^T \\ -A & sI_m \end{pmatrix}, \quad rs > \|A^TA\|.}$$

- x子问题中二次函数形式简单, 对参数也只有 $r > 0$ 和 $\delta > 0$的要求. 但是求新的 λ, 整个迭代过程中对正定矩阵 $(\frac{1}{r}AA^T + \delta I_m)$ 要做一次 Cholesky 分解.

$$H = \begin{pmatrix} rI_n & A^T \\ A & \frac{1}{r}AA^T + \delta I_m \end{pmatrix}, \qquad H = \begin{pmatrix} rI_n & -A^T \\ -A & \frac{1}{r}AA^T + \delta I_m \end{pmatrix}$$

- 对参数只有 $r > 0$ 和 $\delta > 0$的要求, 求新的λ也容易. 只是 x子问题中二次函数形式是 $\|A(x - x^k)\|^2$.

$$H = \begin{pmatrix} \beta A^TA + \delta I_n & A^T \\ A & \frac{1}{\beta}I_m \end{pmatrix}, \qquad H = \begin{pmatrix} \beta A^TA + \delta I_n & -A^T \\ -A & \frac{1}{\beta}I_m \end{pmatrix}$$

- 也可以设定下述正定矩阵, 再去反推子问题如何安排.

$$H = \begin{pmatrix} \delta I_n & 0 \\ 0 & \dfrac{1}{\beta}I_m \end{pmatrix}, \qquad H = \begin{pmatrix} I_n & 0 \\ 0 & I_m \end{pmatrix}$$

可以根据问题的实际需要 [37], 选择不同的正定矩阵 H, 构造不同的方法

3 交替方向法及其收敛性证明

我们利用 VI 和 PPA 来证明 ADMM 的收敛性, 采用 [20, 22] 中的程式.

交替方向法处理的是两个可分离块的凸优化问题

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

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

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

其中

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}\quad \Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m. \tag{3.2b}$$

理解 ADMM 的收敛性证明, 读明白这一节的几页 PPT 就可以了.

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

$$\left\{\begin{array}{rcl} x^{k+1} &=& \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, \quad (3.3a) \\ y^{k+1} &=& \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\}, \quad (3.3b) \\ \lambda^{k+1} &=& \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \quad (3.3c) \end{array}\right.$$

根据最优性引理 1, ADMM k-步迭代满足

$$\left\{\begin{array}{l} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^k + \beta A^T(Ax^{k+1} + By^k - b)\} \ge 0, \quad \forall x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^k + \beta B^T(Ax^{k+1} + By^{k+1} - b)\} \ge 0, \quad \forall y \in \mathcal{Y}, \\ \qquad (\lambda - \lambda^{k+1})^T\{(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \ge 0, \quad \forall \lambda \in \Re^m. \end{array}\right.$$

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

$$\left\{\begin{array}{l} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1})\} \ge 0, \quad \forall x \in \mathcal{X}, \quad (3.4a) \\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\} \ge 0, \ \forall y \in \mathcal{Y}, \quad (3.4b) \\ \qquad (\lambda - \lambda^{k+1})^T\{(Ax^{k+1} + By^{k+1} - b) \qquad + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \ge 0, \quad \forall \lambda \in \Re^m. \quad (3.4c) \end{array}\right.$$

在 (3.4b) 的后半部加上和为零的两项, 得到

$$
\begin{cases}
\underline{\theta_1(x)-\theta_1(x^{k+1})} + (x-x^{k+1})^T\{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1}) \\[2ex]
\underline{\theta_2(y)-\theta_2(y^{k+1})} + (y-y^{k+1})^T\{-B^T\lambda^{k+1} + \underbrace{\beta B^T B(y^k - y^{k+1}) + \beta B^T B(y^{k+1}-y^k)}\} \ge 0, \\[2ex]
\qquad (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b) \qquad\qquad + \frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\} \ge 0.
\end{cases}
$$

利用变分不等式 (3.2), 进行合理整合, 得到

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

将上式中那个任意的 w, 设成解点 w^* 便有

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

经转换, 得到

$$\begin{aligned}&\begin{pmatrix}y^{k+1}-y^*\\\lambda^{k+1}-\lambda^*\end{pmatrix}^T\begin{pmatrix}\beta B^TB&0\\0&\frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix}y^k-y^{k+1}\\\lambda^k-\lambda^{k+1}\end{pmatrix}\\&\quad\ge\begin{pmatrix}x^{k+1}-x^*\\y^{k+1}-y^*\end{pmatrix}^T\beta\begin{pmatrix}A^T\\B^T\end{pmatrix}B(y^k-y^{k+1})\\&\qquad+\underbrace{[\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^{k+1})]}.\end{aligned}\tag{3.5}$$

假如 (3.5) 式右端非负, 证明就基本上完成了. 由于

$$\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^{k+1})=\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^*)\ge 0.$$

这说明 (3.5) 式右端下划线部分非负. 从而由 (3.5) 式得到

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \ \geq \ \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}). \tag{3.6}$$

其中

$$v = \begin{pmatrix} y \\ \lambda \end{pmatrix}, \quad H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}$$

只要 (3.6) 式的右端非负, 就有 $(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0.$

对 (3.6) 式的右端进行处理, 有

$$\begin{aligned}
&\begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) = (y^k - y^{k+1})^T B^T \beta (A, B) \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix} \\
&= \quad (y^k - y^{k+1})^T B^T \beta \big(Ax^{k+1} + By^{k+1} - (Ax^* + By^*)\big) \qquad \text{利用} (Ax^* + By^* = b) \\
&= \quad (y^k - y^{k+1}) B^T \beta \big(Ax^{k+1} + By^{k+1} - b\big) = (y^k - y^{k+1}) B^T (\lambda^k - \lambda^{k+1}).
\end{aligned} \tag{3.7}$$

利用 (3.4b) 有

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

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

(将任意的 y 分别设成 y^k 和 y^{k+1})

$$\theta_2(y^k) - \theta_2(y^{k+1}) + (y^k - y^{k+1})^T\{-B^T\lambda^{k+1}\} \geq 0.$$

$$\theta_2(y^{k+1}) - \theta_2(y^k) + (y^{k+1} - y^k)^T\{-B^T\lambda^k\} \geq 0.$$

(将上面两式相加, 就有) $(y^k - y^{k+1})B^T(\lambda^k - \lambda^{k+1}) \geq 0.$ ((3.7) 式右端非负)

证明了 (3.7) 式右端非负, 进而得到 (3.6) 式右端非负. 所以

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0. \tag{3.8}$$

第一讲的 Lemma 2 告诉我们:

$$b^T H(a-b) \geq 0 \quad \Rightarrow \quad \|b\|_H^2 \leq \|a\|_H^2 - \|a-b\|_H^2. \tag{3.9}$$

在 (3.9) 中置 $a = (v^k - v^*)$ 和 $b = (v^{k+1} - v^*)$, 根据 (3.8) 就得到收敛的关键不等式

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

PPA 意义下的交替方向法

经典的 ADMM 处理的如下的两个可分离块的凸优化问题

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

我们已经把它转化成相应的变分不等式

Problem: $w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\ \forall\, w\in\Omega.$

其中 $\theta(u)=\theta_1(x)+\theta_2(y)$,

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}, \tag{3.2}$$

$\Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m$.

既然 ADMM 是在 VI 的框架下借助 PPA 证明了收敛性, 我们就可以根据不同的要求, 直接构造不同的分块正定矩阵, 设计相应的 PPA 算法 [37].

实际上都采用延伸的邻近点算法　预测矩阵是正定的, 校正则是平凡的延伸.

Prediction

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

Correction $$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \in (0, 2).$$

收敛性质

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha(2-\alpha)\|v^k - \tilde{v}^k\|_H^2.$$

通过设定不同的正定矩阵 H, 就有不同的 PPA 算法和延伸的 PPA 算法.

<u>方案 I</u>: 核心变量是 $v = (y, \lambda)$ 的 PPA 算法

$$H = \begin{pmatrix} \beta B^T B + \delta I_{n_2} & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad \text{(a small } \delta > 0\text{)}. \tag{3.3}$$

<u>方案 II</u>: 核心变量是 $v = (y, \lambda)$ 的线性化 ADMM

$$H = \begin{pmatrix} sI_{n_2} & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad s > \beta \|B^T B\|. \tag{3.4}$$

<u>方案 III</u>:: 平行处理 x 和 y-子问题, 核心变量是 $w = (x, y, \lambda)$ 的 PPA 型 ADMM.

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & 0 & A^T \\ 0 & \beta B^T B + \delta I_{n_2} & B^T \\ A & B & \frac{2}{\beta} I_m \end{pmatrix}, \tag{3.5}$$

或者

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & 0 & -A^T \\ 0 & \beta B^T B + \delta I_{n_2} & -B^T \\ -A & -B & \frac{2}{\beta} I_m \end{pmatrix}. \tag{3.6}$$

上面提到的 PPA 算法散见于 [37] 和我的报告.

4 凸优化问题分裂收缩算法的统一框架

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

A unified Algorithmic Framework for (4.1)

[Prediction Step.] 从给定的 w 的核心变量 v^k 开始, 找到一个$\tilde{w}^k \in \Omega$, 使其满足

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

其中预测矩阵 Q 满足 $Q^T + Q \succ 0$. (不一定收敛的方法都可以暂时看作预测)

[Correction Step.] 选取非奇异校正矩阵 M, 由

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

给出新的核心变量 v^{k+1}

Convergence Conditions

For the matrices Q and M, there is a positive definite matrix H such that

$$HM = Q. \qquad (4.3a)$$

and

$$G := Q^T + Q - M^T HM \succ 0. \qquad (4.3b)$$

其实, 只要预测 (4.2a) 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

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

就得到满足收敛条件的校正矩阵 M. 实际计算中, 我们只要校正矩阵 M.

H 和 G 只是用来验证收敛条件的, 计算过程中不需要.

统一框架的工作可参阅[33, 34, 37], 最新文献是今年四月的[25].

$\boxed{\text{按照统一框架, 算法收敛性证明是简单的.}}$

Theorem 1 设 $\{v^k\}$ 是用统一框架 (5.2) 求解变分不等式 (4.1) 产生的序列. 如果收敛性条件 (4.3) 满足, 则有

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{4.4}$$

Proof 在预测 (4.2a) 中令 $w = w^*$, 就有

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

所以

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

由于 $(\tilde{w}^k - w^*)^T F(\tilde{w}^k) = (\tilde{w}^k - w^*)^T F(w^*)$, 上式右端非负, 进而得到

$$(v^k - v^{k+1})^T H(\tilde{v}^k - v^*) \ge 0. \tag{4.5}$$

将恒等式

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

用于 (4.5) 的左端, 令 $a = v^k,\ b = v^{k+1},\ c = \tilde{v}^k$ 和 $d = v^*$, 我们得到

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

根据 (4.5) 就有

$$
\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \geq \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2. \tag{4.6}
$$

再把上式的右端化简一下,

$$
\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&\quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\
&\quad \overset{(4.2b)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\
&\quad = \quad 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\
&\quad = \quad (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\
&\quad \overset{(4.3b)}{=} \quad \|v^k - \tilde{v}^k\|_G^2.
\end{aligned} \tag{4.7}
$$

将 (4.7) 代入 (4.6) 就得到定理的结论. $\square$

5 根据统一框架设计预测-校正算法的一般原则

我们处理的问题是以下的变分不等式

$$\text{(VI)}\qquad w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{5.1}$$

求解变分不等式 VI (5.1) 的预测-校正框架

[预测.] 根据给定的 v^k, 求得预测点 $\tilde{w}^k \in \Omega$ 使得

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{5.2a}$$

其中矩阵 Q 不必对称, 但 $Q^T + Q$ 为正定矩阵. (我们着眼 Q 非对称).

[校正.] 用一个非奇异矩阵 M 对核心变量 v 进行校正

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

统一框架中有 $\boxed{\text{预测矩阵 } Q}$ 和 $\boxed{\text{校正矩阵 } M.}$

收敛性条件

对统一框架里的预测矩阵 Q 和校正矩阵 M, 存在正定矩阵 H, 使得

$$HM = Q, \tag{5.3a}$$

和

$$G := Q^T + Q - M^T HM \succ 0. \tag{5.3b}$$

在 [**?**] 中, 我们进一步明确提出, 收敛性条件 (5.3) 可以写成等价的

$$D \succ 0, \quad G \succ 0, \quad \text{and} \quad D + G = Q^T + Q \tag{5.4a}$$

然后再令

$$HM = Q \qquad \text{和} \qquad M^T HM = D. \tag{5.4b}$$

由矩阵方程组

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

解得

$$M = Q^{-T} D \tag{5.5}$$

就是满足收敛条件的校正矩阵.

选择满足条件 (5.4) 的矩阵 D 和 G 的方案许多, 譬如说

$$D = \alpha[Q^T + Q] \quad \text{and} \quad G = (1-\alpha)[Q^T + Q], \qquad \alpha \in (0,1),$$

也是一种选择.

由于校正公式是

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

这可以写成

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

两边左乘 Q^T 利用$Q^T M = D$, 得到

$$Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k), \tag{5.6}$$

预测矩阵 Q 往往是下三角矩阵, 人们可以比较容易地通过 (5.6) 求得 v^{k+1}.

以求解三个可分离问题为例, 对原始变量进行 Gauss 型预测

$$\begin{cases} \tilde{x}^k \in \arg\min\left\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\right\}, \\ \tilde{y}^k \in \arg\min\left\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|A\tilde{x}^k + By + Cz^k - b\|^2 \mid y \in \mathcal{Y}\right\}, \\ \tilde{z}^k \in \arg\min\left\{\theta_3(z) - z^T C^T \lambda^k + \frac{\beta}{2}\|A\tilde{x}^k + B\tilde{y}^k + Cz - b\|^2 \mid z \in \mathcal{Z}\right\}. \end{cases} \tag{5.7}$$

并定义

$$\tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big). \tag{5.8}$$

得到预测公式

$$\tilde{w}^k \in \Omega,\ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k),\ \forall w \in \Omega, \tag{5.9}$$

其中

$$Q = \begin{pmatrix} \beta B^T B & 0 & 0 \\ \beta C^T B & \beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I_m \end{pmatrix}. \tag{5.10}$$

由于

$$Q^T + Q = \begin{pmatrix} 2\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & 2\beta C^T C & -C^T \\ -B & -C & \frac{2}{\beta} I_m \end{pmatrix}, \tag{5.11}$$

可以取

$$D = \begin{pmatrix} \nu\beta B^T B & 0 & 0 \\ 0 & \nu\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \tag{5.12}$$

或者

$$G = \begin{pmatrix} (1-\nu)\beta B^T B & 0 & 0 \\ 0 & (1-\nu)\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}, \tag{5.13}$$

再或者

$$D = G = \frac{1}{2}\left[Q^T + Q\right] = \begin{pmatrix} \beta B^T B & \frac{1}{2}\beta B^T C & -\frac{1}{2} B^T \\ \frac{1}{2}\beta C^T B & \beta C^T C & -\frac{1}{2} C^T \\ -\frac{1}{2} B & -\frac{1}{2} C & \frac{1}{\beta} I_m \end{pmatrix}. \tag{5.14}$$

再注意到进行预测 (5.7) 只需要 (By^k, Cz^k, λ^k) 在手, 我们可以通过

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

容易地求得开始下一次迭代必须的

$$(By^{k+1}, Cz^{k+1}, \lambda^{k+1}).$$

6 一类适用范围更广和便于推广的交替方向法

这一讲以 [19] 中一个代表性算法介绍应用范围更广和便于推广的交替方向法.

其中的“更广”, 体现在线性约束可以是等式的, 也可以是不等式的(无需增加松弛变量); “便于推广”, 是指像求解线性方程组的 Gauss 消去法那样, 不管自变量有多少个可分离块, 按部就班, 处理方法都一样.

p-块可分离凸优化问题

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

的最优性条件能够写成下面的变分不等式:

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

其中

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

和 $\Omega = \prod_{i=1}^{p} \mathcal{X}_i \times \Lambda$. 对应于等式和不等式约束, 分别有 $\Lambda = \Re^m$ 和 $\Lambda = \Re^m_+$.

首先, 经典的ADMM 可以写成

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

对 p 块可分离的问题, 我们采用下述方法预测

Prediction Step. With given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, find $\tilde{w}^k \in \Omega$:

$$\begin{cases} \tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\ \qquad\vdots \\ \tilde{x}_i^k \in \arg\min_{x_i \in \mathcal{X}_i} \{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\| \sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\ \qquad\vdots \\ \tilde{x}_p^k \in \arg\min_{x_p \in \mathcal{X}_p} \{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{\beta}{2}\| \sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}; \\ \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^{p} A_j\tilde{x}_j^k - b\big)\big]. \end{cases} \tag{6.3}$$

统一框架仍然是构造这类方法的指导原则: 上面的预测就像 Gauss 消去法那样往前"追". 利用最优性条件的引理得到预测的变分不等式形式是

$$\tilde{w}^k \in \Omega,\ \theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (w-\tilde{w}^k)^T Q(w^k-\tilde{w}^k),\ \forall w \in \Omega, \tag{6.4a}$$

其中

$$Q = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{6.4b}$$

校正也像 Gauss 消去法那样往后"赶", 将自变量从后向前依次算出来.

$$\begin{pmatrix} A_1 x_1^{k+1} \\ A_2 x_2^{k+1} \\ \vdots \\ A_p x_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1 x_1^k \\ A_2 x_2^k \\ \vdots \\ A_p x_p^k \end{pmatrix} - \nu \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1 x_1^k - A_1 \tilde{x}_1^k \\ A_2 x_2^k - A_2 \tilde{x}_2^k \\ \vdots \\ A_p x_p^k - A_p \tilde{x}_p^k \end{pmatrix}, \tag{6.5a}$$

$$\lambda^{k+1} = \tilde{\lambda}^k + \nu\beta(A_1 x_1^k - A_1 \tilde{x}_1^k). \tag{6.5b}$$

事实上, 采用变换

$$P = \begin{pmatrix} \sqrt{\beta}A_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta}A_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sqrt{\beta}A_p & 0 \\ 0 & \cdots & \cdots & 0 & (1/\sqrt{\beta})I_m \end{pmatrix}, \quad \xi = Pw = \begin{pmatrix} \sqrt{\beta}A_1 x_1 \\ \sqrt{\beta}A_2 x_2 \\ \vdots \\ \sqrt{\beta}A_p x_p \\ (1/\sqrt{\beta})\lambda \end{pmatrix}. \tag{6.6}$$

对 (6.4b) 中的矩阵 Q, 有

$$Q = P^T \mathcal{Q} P, \qquad \text{where} \qquad \mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{6.7}$$

将不等式 (6.4a) 的右端改写成

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

预测公式 (6.4) 能够写成:

$$\tilde{w}^k \in \Omega, \ \ \theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (\xi-\tilde{\xi}^k)^T \mathcal{Q}(\xi^k-\tilde{\xi}^k), \ \forall w \in \Omega. \tag{6.8}$$

其中 $\mathcal{Q}$ 是由 (6.7) 给出的. 记

$$\mathcal{L} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix}, \quad \mathcal{I} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{6.9}$$

和

$$\mathcal{E} = \begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix}. \tag{6.10}$$

就有

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} \quad \text{和} \quad \mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I}+\mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix}. \tag{6.11}$$

选择 $\mathcal{D} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}$, 就满足 $0 \prec \mathcal{D} \prec \mathcal{Q}^T + \mathcal{Q}$. 这样, 校正矩阵

$$\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix}\begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{6.12}$$

再利用 (6.6) 中的关系以及

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

校正

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k) \tag{6.13}$$

就是通过 (6.5) 实现的. (6.5a) 的过程, 就像线性方程组求解的 Gauss 回代过程.

我们有理由把 [19] 中的这个方法叫做适用范围广又便于推广的 ADMM 类预测-校正方法.

References

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

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Foundations and Trends in Machine Learning Vol. 3, No. 1 (2010) 1 – 122.

[3] 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.

[4] A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vison, 40, 120-145, 2011.

[5] A. Chambolle and T Pock, On the ergodic convergence rates of a first-order primal – dual algorithm， Math. Program., A 159 (2016) 253-287.

[6] 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.

[7] E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

[8] R. Fletcher, Practical mathods of Optimization, Second Edition, JOHN WILEY & SONS, 1987.

[9] 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.

[10] D. Gabay, Applications of the method of multipliers to variational inequalities, *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, edited by M. Fortin and R. Glowinski, North Holland, Amsterdam, The Netherlands, 1983, pp. 299–331.

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

[12] B.S. He, A new method for a class of linear variational inequalities, Math. Progr., **66**, 137-144, 1994.

[13] B.S. He, Solving a class of linear projection equations, Numerische Mathematik, **68**, 71-80, 1994.

[14] B. S. He, Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities, *Computational Optimization and Applications* **42**(2009), 195–212.

[15] B. S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman-Rachford splitting method for convex programming, *SIAM Journal on Optimization* **24**(2014), 1011-1040.

[16] 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.

[17] B.S. He, M. Tao and X.M. Yuan, A splitting method for separable convex programming, *IMA*

Journal of Numerical Analysis, **31**, 394-426, 2015.

[18] Bingsheng He, Shengjie Xu and Jing Yuan, Indefinite linearized augmented Lagrangian method for convex programming with linear inequality constraints, arXiv 2105.02425 [MathOC]

[19] 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].

[20] B. S. He and H. Yang, Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities, *Operations Research Letters* **23**(1998), 151–161.

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

[22] 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.

[23] B.S. He and X.M. Yuan, On the optimal proximal parameter of an ADMM-like splitting method for separable convex programming. Mathematical methods in image processing and inverse problems, 139 – 163, Springer Proc. Math. Stat., 360, Springer, Singapore, 2021, and Optimization-Online 6235, 2017.

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

[25] 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]

[26] 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.

[27] M. R. Hestenes, Multiplier and gradient methods, *JOTA* **4**, 303-320, 1969.

[28] Feng Ma, Yiming Bi and Bin Ga, A prediction – correction-based primal – dual hybrid gradient method for linearly constrained convex minimization, Numerical Algorithms (2019) 82:641 – 662

[29] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, *Rev. Francaise d'Inform. Recherche Oper.*, **4**, 154-159, 1970.

[30] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in Optimization, R. Fletcher, ed., Academic Press, New York, NY, pp. 283-298, 1969.

[31] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Cont. Optim.*, **14**, 877-898, 1976.

[32] 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.

[33] 何炳生. 从变分不等式的统一收缩算法到凸优化的分裂收缩算法, 高等学校计算数学学报, 2016, 38: 74–96.

[34] 何炳生. 我和乘子交替方向法 20 年. 运筹学学报, 2018, 22: 1–31.

[35] 何炳生. 凸优化和单调变分不等式收缩算法的统一框架. 中国科学：数学, 2018, 48: 255–272.

[36] 何炳生. 凸优化的一阶分裂算法–变分不等式为工具的统一框架, 见作者主页:

maths.nju.edu.cn/~hebma 中的《My Talk》.

[37] 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.

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

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.