# 一些典型凸优化问题的分裂收缩算法

## 基于变分不等式和邻近点算法的统一框架

南师大数科院系列讲座　2022 年元月 4 – 9 日

**何 炳 生**

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

## 八．　三个可分离块的凸优化问题

这一章考虑三块可分离凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)|Ax+By+Cz=b,\ x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}\} \tag{0.1}$$

的求解方法. 这个问题的拉格朗日函数是

$$L(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b).$$

问题 (0.1) 同样可以归结为变分不等式问题

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{0.2a}$$

其中

$$w=\begin{pmatrix} x\\ y\\ z\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\\ z\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ -C^T\lambda\\ Ax+By+Cz-b\end{pmatrix}, \tag{0.2b}$$

和

$$\theta(u)=\theta_1(x)+\theta_2(y)+\theta_3(z),\qquad \Omega=\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}\times\Re^m. \tag{0.2c}$$

相应的增广拉格朗日函数记为(与两个算子的符号有区别)

$$\begin{aligned}\mathcal{L}^{[3]}_\beta(x,y,z,\lambda)=&\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b)\\ &+\frac{\beta}{2}\|Ax+By+Cz-b\|^2.\end{aligned} \tag{0.3}$$

# 1 直接推广的 ADMM 求解三块可分离问题不保证收敛

对三个可分离块的凸优化问题, 采用直接推广的乘子交替方向法, 第 $k$ 步迭代是从给定的 $v^k=(y^k,z^k,\lambda^k)$ 出发, 通过

$$\left\{\begin{array}{rcl} x^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^{[3]}(x,y^k,z^k,\lambda^k)\;\middle|\; x\in\mathcal{X}\right\},\\ y^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^{[3]}(x^{k+1},y,z^k,\lambda^k)\;\middle|\; y\in\mathcal{Y}\right\},\\ z^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^{[3]}(x^{k+1},y^{k+1},z,\lambda^k)\;\middle|\; z\in\mathcal{Z}\right\},\\ \lambda^{k+1} & = & \lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b), \end{array}\right. \tag{1.1}$$

求得新的迭代点 $w^{k+1}=(x^{k+1},y^{k+1},z^{k+1},\lambda^{k+1})$. 当矩阵 $A$, $B$, $C$ 中有两个是互相正交的时候, 用方法 (1.1) 求解问题 (0.1) 是收敛的 因为这种三块的可分离问题, 实际上相当于两块可分离的问题. 对一般的三块可分离问题, 是不能保证收敛的[1].

## 值得继续研究的问题和猜想

譬如说, 三个算子的实际问题中, 线性约束矩阵

$$\mathcal{A}=[A,B,C]\quad \text{中, 往往至少有一个是单位矩阵. 即,}\quad \mathcal{A}=[A,B,I].$$

直接推广的 ADMM 处理这种更贴近实际的三个算子的问题, 既没有证明收敛, 也没有举出反例, 这仍然是一个有趣又特别有意义的问题! 举个简单的例子来说吧:

- 经典的乘子交替方向法处理问题

$$\min\{\theta_1(x)+\theta_2(y)|Ax+By=b,\ x\in\mathcal{X},y\in\mathcal{Y}\}\quad \text{是收敛的.}$$

- 将等式约束换成不等式约束, 问题就变成

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

- 再化成三个算子的等式约束问题就是

$$\min\{\theta_1(x)+\theta_2(y)+0\,|\,Ax+By+z=b,\ x\in\mathcal{X},y\in\mathcal{Y},z\ge 0\}.$$

- 直接推广的乘子交替方向法 (1.1) 处理上面这种问题, 我们猜想是收敛的, 但是至今没有证明收敛性. 仍然是一个遗留的挑战性问题！

在对直接推广的 ADMM (1.1) 证明不了收敛性的时候, 我们就着手对三块可分离的问题提出一些修正算法. 修正方法的原则是尽量对 ADMM 少做改动, 保持它原来的好品性. 特别是对问题不加(诸如目标函数强凸等)任何额外条件, 对经典 ADMM 中需要调比选取的大于零的 $\beta$, 仍然让它可以自由选取.

这一章的后面几节, 除了对已经发表的一些算法分别用统一框架去验证收敛性, 也介绍一些最近根据统一框架构造的方法.

## 2 部分平行分裂的 ADMM 预测校正方法

这一节的方法源自 2009 年发表的[5], 还是把 $x$ 当成中间变量, 迭代从 $v^k = (y^k, z^k, \lambda^k)$ 到 $v^{k+1} = (y^{k+1}, z^{k+1}, \lambda^{k+1})$, 只是平行处理 $y$ 和 $z$-子问题, 再更新 $\lambda$. 换句话说, 把

$$\left\{\begin{array}{lcl} x^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^{[3]}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\right\}, \\ y^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^{[3]}(x^{k+1}, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\right\}, \\ z^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^{[3]}(x^{k+1}, y^k, z, \lambda^k) \mid z \in \mathcal{Z}\right\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b) \end{array}\right. \tag{2.1}$$

生成的点 $(x^{k+1}, y^{k+1}, z^{k+1}, \lambda^{k+1})$ 当成预测点. 再把核心变量往回拉一点. 原因是 $y$, $z$ 子问题平行处理, 包括据此更新的 $\lambda$, 都太自由, 需要校正. 校正公式是

$$v^{k+1} := v^k - \alpha(v^k - v^{k+1}), \quad \alpha \in (0, 2 - \sqrt{2}). \tag{2.2}$$

譬如说, 我们可以取 $\alpha = 0.55$. 注意到 (2.2) 右端的 $v^{k+1} = (y^{k+1}, z^{k+1}, \lambda^{k+1})$ 是由 (2.1) 提供的.

我们用统一框架来验证这个部分平行分裂的预测校正方法的收敛性. 先把由 (2.1) 生成的 $(x^{k+1}, y^{k+1}, z^{k+1})$ 视为 $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$, 并定义

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b). \tag{2.3}$$

这样, 预测点 $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k)$ 就可以看成由下式生成:

$$\left\{\begin{array}{ll} \tilde{x}^k = \arg\min\left\{\mathcal{L}_\beta^{[3]}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\right\}, & \text{(2.4a)} \\ \tilde{y}^k = \arg\min\left\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\right\}, & \text{(2.4b)} \\ \tilde{z}^k = \arg\min\left\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k, y^k, z, \lambda^k) \mid z \in \mathcal{Z}\right\},, & \text{(2.4c)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b). & \text{(2.4d)} \end{array}\right.$$

利用增广拉格朗日函数 (0.3), 子问题 (2.4a) 相当于

$$\tilde{x}^k = \text{argmin}\{\theta_1(x) - x^TA^T\lambda^k + \tfrac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \,|\, x \in \mathcal{X}\},$$

根据最优性引理, $\tilde{x}^k \in \mathcal{X}$,

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

再根据 (2.4d), 就有

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\} \geq 0, \;\; \forall x \in \mathcal{X}. \tag{2.5a}$$

子问题 (2.4b) 相当于

$$\tilde{y}^k = \text{argmin}\{\theta_2(y) - y^TB^T\lambda^k + \tfrac{1}{2}\beta\|A\tilde{x}^k + By + Cz^k - b\|^2 \,|\, y \in \mathcal{Y}\},$$

同样根据最优性条件引理, 有 $\tilde{y}^k \in \mathcal{Y}$,

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

再根据 (2.4d), 就有

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\big\{\underline{-B^T\tilde{\lambda}^k} \\ + \beta B^T B(\tilde{y}^k - y^k)\big\} \geq 0, \quad \forall y \in \mathcal{Y}. \tag{2.5b}$$

同理, 对子问题 (2.4c) 有

$$\tilde{z}^k \in \mathcal{Z}, \quad \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T\big\{\underline{-C^T\tilde{\lambda}^k} \\ + \beta C^T C(\tilde{z}^k - z^k)\big\} \geq 0, \quad \forall\, z \in \mathcal{Z}. \tag{2.5c}$$

注意到 (2.4d) 可以写成

$$\underline{(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)} - B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + (1/\beta)\,(\tilde{\lambda}^k - \lambda^k) = 0. \tag{2.5d}$$

把 (2.6) 中的公式组合在一起, 可以写成统一框架中的预测形式:

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \forall w \in \Omega, \tag{2.6a}$$

其中

$$Q = \begin{pmatrix} \beta B^T B & 0 & 0 \\ 0 & \beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix}. \tag{2.6b}$$

回头来看方法 (2.1)-(2.2) 在统一框架中的校正该怎么表示. 由于

$$y^{k+1} = \tilde{y}^k, \quad z^{k+1} = \tilde{z}^k, \quad \text{和} \quad \lambda^{k+1} = \tilde{\lambda}^k + \beta B(y^k - \tilde{y}^k) + \beta C(y^k - \tilde{y}^k).$$

把 (2.4) 的输出作为预测点时, 校正公式 (2.2) 就可以表示成

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

也就是说, 利用了统一框架中 (2.6)这样的预测表达式, 方法 (2.1)-(2.2) 的校正公式是

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \tag{2.7a}$$

其中

$$M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \tag{2.7b}$$

对这样的 $Q$ 和 $M$, 设

$$H = \begin{pmatrix} \beta B^T B & 0 & 0 \\ 0 & \beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix},$$

就有 $HM = Q$, 说明收敛性条件满足.

根据统一框架, 要找出一个 $\alpha > 0$, 使得条件

$$G = (Q^T + Q) - \alpha M^T H M \succ 0$$

满足. 简单的矩阵运算得到

$$\begin{aligned} G &= (Q^T + Q) - \alpha M^T H M = (Q^T + Q) - \alpha M^T Q \\ &= \begin{pmatrix} \sqrt{\beta} B^T & 0 & 0 \\ 0 & \sqrt{\beta} C^T & 0 \\ 0 & 0 & \frac{1}{\sqrt{\beta}} I \end{pmatrix} \begin{pmatrix} 2(1-\alpha) I & -\alpha I & -(1-\alpha) I \\ -\alpha I & 2(1-\alpha) I & -(1-\alpha) I \\ -(1-\alpha) I & -(1-\alpha) I & (2-\alpha) I \end{pmatrix} \\ &\quad \begin{pmatrix} \sqrt{\beta} B & 0 & 0 \\ 0 & \sqrt{\beta} C & 0 \\ 0 & 0 & \frac{1}{\sqrt{\beta}} I \end{pmatrix}. \end{aligned}$$

容易验证, 对所有的 $\alpha \in (0, 2-\sqrt{2})$, 矩阵

$$\begin{pmatrix} 2(1-\alpha) & -\alpha & -(1-\alpha) \\ -\alpha & 2(1-\alpha) & -(1-\alpha) \\ -(1-\alpha) & -(1-\alpha) & (2-\alpha) \end{pmatrix} \succ 0.$$

收敛性条件满足.

# 3 带高斯回代的 ADMM 方法

带高斯回代的ADMM方法 [6] 是2012年发表的. 直接推广的乘子交替方向法 (1.1) 对三个算子的问题不能保证收敛, 是因为它们处理有关核心变量的 $y$ 和 $z$-子问题不公平. 采取补救的办法是将 (1.1) 提供的 $(y^{k+1}, z^{k+1}, \lambda^{k+1})$ 当成预测点, , 校正公式为

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \lambda^{k+1} \end{pmatrix} := \begin{pmatrix} y^k \\ z^k \\ \lambda^k \end{pmatrix} - \nu \begin{pmatrix} I & -(B^T B)^{-1} B^T C & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix}. \tag{3.1}$$

其中 $\nu \in (0,1)$, 右端的 $(y^{k+1}, z^{k+1}, \lambda^{k+1})$ 是由 (1.1) 提供的. 这个方法发表在 [6]. 想法是不公平, 就要做找补, 调整. 事实上, 也可以就用 (1.1) 提供的 $\lambda^{k+1}$, 只通过

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} := \begin{pmatrix} y^k \\ z^k \end{pmatrix} - \nu \begin{pmatrix} I & -(B^T B)^{-1} B^T C \\ 0 & I \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}. \tag{3.2}$$

校正 $y$ 和 $z$ (无需校正$\lambda$). 由于为下一步迭代只需要准备 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$, 我们只要做比(3.2)更简单的

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \end{pmatrix} := \begin{pmatrix} By^{k} \\ Cz^{k} \end{pmatrix} - \nu \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} By^{k} - By^{k+1} \\ Cz^{k} - Cz^{k+1} \end{pmatrix}. \tag{3.3}$$

上式中这个从下到上的的过程我们把它叫做高斯回代.

求解三个可分离块凸优化 (0.1), [6] 中介绍了带高斯回代的 ADMM 方法. 把由直接推广的 (1.1) 生成的 $x^{k+1}, y^{k+1}, z^{k+1}$ 分别视为 $\tilde{x}^k$, $\tilde{y}^k$, $\tilde{z}^k$, 并定义

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

把 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k)$ 看做预测点, 它由下面的公式

$$\left\{ \begin{array}{lcl} \tilde{x}^k & = & \operatorname{argmin}\big\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \big| x \in \mathcal{X}\big\}, \\ \tilde{y}^k & = & \operatorname{argmin}\big\{\theta_2(y) - (\lambda^k)^T By + \frac{\beta}{2}\|A\tilde{x}^k + By + Cz^k - b\|^2 \big| y \in \mathcal{Y}\big\}, \\ \tilde{z}^k & = & \operatorname{argmin}\big\{\theta_3(z) - (\lambda^k)^T Cz + \frac{\beta}{2}\|A\tilde{x}^k + B\tilde{y}^k + Cz - b\|^2 \big| z \in \mathcal{Z}\big\}, \\ \tilde{\lambda}^k & = & \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b). \end{array} \right.$$

这样, 利用最优性引理把最优性条件写出来, 利用变分不等式 (0.2) 的形式, 就可以写成统一框架中预测的形式:

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \forall w \in \Omega, \tag{3.4a}$$

其中

$$Q = \begin{pmatrix} \beta B^T B & 0 & 0 \\ \beta C^T B & \beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix}. \tag{3.4b}$$

利用这样的预测点, 只校正 $y$ 和 $z$ 的公式 (3.2) (注意 $\lambda^{k+1}$ 和 $\tilde{\lambda}^k$ 的关系) 就可以写成

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^{k} \\ z^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I & -\nu(B^T B)^{-1} B^T C & 0 \\ 0 & \nu I & 0 \\ -\beta B & -\beta C & I \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ z^k - \tilde{z}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

也就是说, 在统一框架的校正公式中

$$M = \begin{pmatrix} \nu I & -\nu(B^T B)^{-1} B^T C & 0 \\ 0 & \nu I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \tag{3.5}$$

对于矩阵

$$H = \begin{pmatrix} \frac{1}{\nu}\beta B^T B & \frac{1}{\nu}\beta B^T C & 0 \\ \frac{1}{\nu}\beta C^T B & \frac{1}{\nu}\beta[C^T C + C^T B(B^T B)^{-1} B^T C] & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix} \tag{3.6}$$

可以验证 $H$ 正定并有 $HM = Q$. 此外,

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

由于 $\nu \in (0,1)$, 矩阵 $G$ 正定, 收敛性条件满足.

事实上, 从 (3.4) 得到了

$$\begin{pmatrix} B\tilde{y}^k - By^* \\ C\tilde{z}^k - Cz^* \\ \tilde{\lambda}^k - \lambda^* \end{pmatrix}^T \begin{pmatrix} \beta I & 0 & 0 \\ \beta I & \beta I & 0 \\ -I & -I & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} B(y^k - \tilde{y}^k) \\ C(z^k - \tilde{z}^k) \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \geq 0.$$

若令

$$\xi = \begin{pmatrix} By \\ Cz \\ \lambda \end{pmatrix} \qquad \text{和} \qquad \mathcal{Q} = \begin{pmatrix} \beta I & 0 & 0 \\ \beta I & \beta I & 0 \\ -I & -I & \frac{1}{\beta} I \end{pmatrix}, \tag{3.7}$$

就有

$$(\tilde{\xi}^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \geq 0,$$

和因此得到的

$$(\xi^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \geq (\xi^k - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k). \tag{3.8}$$

回代公式 (3.3) 就可以写成

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k) \qquad \text{其中} \qquad \mathcal{M} = \begin{pmatrix} \nu I & -\nu I & 0 \\ 0 & \nu I & 0 \\ -\beta I & -\beta I & I \end{pmatrix}. \tag{3.9}$$

对于

$$\mathcal{H} = \begin{pmatrix} \frac{1}{\nu}\beta I & \frac{1}{\nu}\beta I & 0 \\ \frac{1}{\nu}\beta I & \frac{2}{\nu}\beta I & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}, \qquad \text{我们有} \qquad \mathcal{HM} = \mathcal{Q}.$$

利用 (3.8) 和 $\mathcal{HM} = \mathcal{Q}$, 我们得到

$$\begin{aligned} &\|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \\ &\quad = \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|(\xi^k - \xi^*) - \mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\ &\quad = 2(\xi^k - \xi^*)^T \mathcal{HM}(\xi^k - \tilde{\xi}^k) - \|\mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\ &\quad \geq 2(\xi^k - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) - -\|\xi^k - \tilde{\xi}^k\|_{(\mathcal{M}^T \mathcal{HM})}^2 \\ &\quad = (\xi^k - \tilde{\xi}^k)[\mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T \mathcal{HM}](\xi^\| - \tilde{\xi}^\|). \end{aligned} \tag{3.10}$$

由于

$$
\begin{aligned}
\mathcal{G} &= \mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} = \mathcal{Q}^T + \mathcal{Q} - \mathcal{Q}^T\mathcal{M} \\
&= \begin{pmatrix} 2\beta I & \beta I & -I \\ \beta I & 2\beta I & -I \\ -I & -I & \frac{2}{\beta}I \end{pmatrix} - \begin{pmatrix} \beta I & \beta I & -I \\ 0 & \beta I & -I \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix} \begin{pmatrix} \nu I & -\nu I & 0 \\ 0 & \nu I & 0 \\ -\beta I & -\beta I & I \end{pmatrix} \\
&= \begin{pmatrix} (1-\nu)\beta I & 0 & 0 \\ 0 & (1-\nu)\beta I & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix} \succ 0,
\end{aligned}
$$

所以

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

从另一个角度证明了方法的收敛性.

## 4 部分平行并加正则项的 ADMM 方法

下面的方法与 §2 中方法相同的是平行求解 $y$, $z$-子问题, 不同的是不做后处理, 而是给这两个子问题预先都加个正则项. 方法写起来就是

$$
\begin{cases}
x^{k+1} = \arg\min\left\{\mathcal{L}_\beta^{[3]}(x, y^k, z^k, \lambda^k) \,\middle|\, x \in \mathcal{X}\right\}, \\
y^{k+1} = \arg\min\left\{\mathcal{L}_\beta^{[3]}(x^{k+1}, y, z^k, \lambda^k) + \frac{\nu}{2}\beta\|B(y-y^k)\|^2 \middle| y \in \mathcal{Y}\right\}, \\
z^{k+1} = \arg\min\left\{\mathcal{L}_\beta^{[3]}(x^{k+1}, y^k, z, \lambda^k) + \frac{\nu}{2}\beta\|C(z-z^k)\|^2 \middle| z \in \mathcal{Z}\right\}, \\
\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b),
\end{cases}
$$

其中 $\nu > 1$. 上述做法相当于

$$
\begin{cases}
x^{k+1} = \mathrm{argmin}\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b - \frac{1}{\beta}\lambda^k\|^2 \mid x \in \mathcal{X}\}, \\
\qquad \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\
y^{k+1} = \mathrm{argmin}\{\theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\mu\beta}{2}\|B(y-y^k)\|^2 \mid y \in \mathcal{Y}\}, \\
z^{k+1} = \mathrm{argmin}\{\theta_3(z) - z^T C^T \lambda^{k+\frac{1}{2}} + \frac{\mu\beta}{2}\|C(z-z^k)\|^2 \mid z \in \mathcal{Z}\}, \\
\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b),
\end{cases}
\tag{4.1}
$$

其中 $\mu = \nu + 1$. 例如, 可以取 $\mu = 2.01$. 这类发表在 [7, 8] 的算法思想是: 让 $y$ 和 $z$ 各自独立, 又不准备校正, 那就预先加正则项让它们不致走得太远. [7] 中的方法被 UCLA Osher 教授的课题组成功用来求解图像降维问题 [2].

把由 (4.1) 生成的 $(x^{k+1}, y^{k+1}, z^{k+1}, \lambda^{k+\frac{1}{2}})$ 视为预测点$(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k)$, 这个预测公式就成为

$$
\begin{cases}
\tilde{x}^k = \mathrm{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, \\
\tilde{y}^k = \mathrm{argmin}\{\theta_2(y) - y^T B^T \tilde{\lambda}^k + \frac{\mu\beta}{2}\|B(y-y^k)\|^2 \mid y \in \mathcal{Y}\}, \\
\tilde{z}^k = \mathrm{argmin}\{\theta_3(z) - z^T C^T \tilde{\lambda}^k + \frac{\mu\beta}{2}\|C(z-z^k)\|^2 \mid z \in \mathcal{Z}\}, \\
\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b).
\end{cases}
\tag{4.2}
$$

这样, 利用最优性引理和变分不等式 (0.2) 的形式, 预测就可以写成统一框架中的形式:

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \forall w \in \Omega, \tag{4.3a}$$

其中

$$Q = \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix}. \tag{4.3b}$$

利用这样的预测点, 校正 $y$ 和 $z$ 的公式 (注意 $\lambda^{k+1}$ 和 $\tilde{\lambda}^k$ 的关系) 就可以写成

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

也就是说, 在统一框架的校正公式中

$$M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \tag{4.4}$$

对于矩阵

$$H = \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix},$$

可以验证 $H$ 正定并有 $HM = Q$. 此外,

$$\begin{aligned} G &= (Q^T + Q) - M^T H M = (Q^T + Q) - M^T Q \\ &= \begin{pmatrix} 2\mu\beta B^T B & 0 & -B^T \\ 0 & 2\mu\beta C^T C & -C^T \\ -B & -C & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} (1+\mu)\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & (1+\mu)\beta C^T C & -C^T \\ -B & -C & \frac{1}{\beta} I \end{pmatrix} \\ &= \begin{pmatrix} (\mu-1)\beta B^T B & -\beta B^T C & 0 \\ -\beta C^T B & (\mu-1)\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}. \end{aligned}$$

由于 $\mu > 2$, 矩阵 $G$ 正定, 收敛性条件满足. 方法的收敛性得到证明.

# 参考文献

[1] 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.

[2] 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.

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

[4] B. S. He, PPA-like contraction methods for convex optimization: a framework using variational inequality approach. J. Oper. Res. Soc. China **3** (2015) 391 – 420.

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

[6] 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.

[7] 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.

[8] Tao M and Yuan X M. Recovering low-rank and sparse components of matrices from incomplete and noisy observations *SIAM Journal on Optimization*, 2011, 21: 57-81.

[9] B.S. He, My 20 years research on alternating directions method of multipliers (in Chinese). Oper. Res. Trans., 2018, 22: 1-31.

[10] 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.

[11] 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 http://maths.nju.edu. cn/˜hebma 中的《My Talk》.

[12] 何炳生. 凸优化和单调变分不等式的收缩算法, 见 http://maths.nju.edu.cn/˜hebma 中的系列讲义.