变分不等式框架下结构型凸优化的分裂收缩算法

V. 三个可分离块凸优化问题的分裂收缩方法

中学的数理基础　必要的社会实践
普通的大学数学　一般的优化原理

何 炳 生　　南京大学数学系
Homepage: maths.nju.edu.cn/~hebma

天元数学东北中心　2023 年 10 月 17 – 27 日

1 Problem with three separable blocks

这一讲考虑三块可分离凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) | Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\} \tag{1.1}$$

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 (1.1) 同样可以归结为变分不等式问题

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^T F(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{1.2a}$$

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

$$w=\begin{pmatrix} x\\ y\\ z\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\\ z\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ -C^T\lambda\\ Ax+By+Cz-b\end{pmatrix}. \tag{1.2b}$$

相应的增广拉格朗日函数记为(与两个算子的符号有区别)

$$\mathcal{L}_{\beta}^{[3]}(x,y,z,\lambda) = \theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b) + \frac{\beta}{2}\|Ax+By+Cz-b\|^2. \tag{1.3}$$

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

$$\begin{cases} x^{k+1} \in \arg\min\{\mathcal{L}_{\beta}^{[3]}(x,y^k,z^k,\lambda^k) \mid x\in\mathcal{X}\}, \\ y^{k+1} \in \arg\min\{\mathcal{L}_{\beta}^{[3]}(x^{k+1},y,z^k,\lambda^k) \mid y\in\mathcal{Y}\}, \\ z^{k+1} \in \arg\min\{\mathcal{L}_{\beta}^{[3]}(x^{k+1},y^{k+1},z,\lambda^k) \mid z\in\mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b), \end{cases} \tag{1.4}$$

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 (1.4) 求解问题 (1.1) 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, 三个可分离块的实际问题中, 线性约束矩阵

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

直接推广的 ADMM 处理这种更贴近实际的三个可分离块的问题, 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

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

- 再化成三个可分离块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 (1.4) 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. 仍然是一个遗留的极具挑战性的问题！

在对直接推广的 ADMM (1.4) 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.

2 统一框架的等价表示

问题: $w^* \in \Omega, \quad \theta(u)-\theta(u^*)+(w-w^*)^\top F(w^*) \ge 0, \quad \forall\, w\in\Omega.$ (2.1)

[预测] 第 k-步迭代从给定的核心变量 v^k 开始, 求得预测点 $\tilde{w}^k$, 使得

$$\tilde{w}^k \in \Omega, \ \ \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^\top F(\tilde{w}^k) \ge (v-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k), \ \ \forall\, w\in\Omega, \tag{2.2}$$

成立. 其中矩阵 $Q^\top + Q$ 是正定的.

左端将问题 (2.1)的 w^* 换成了 $\tilde{w}^k$. 称 Q 为预测矩阵

[校正]. 根据预测得到的 $\tilde{v}^k$, 给出核心变量 v 的新迭代点 v^{k+1} 的公式为

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{2.3}$$

我们称 M 为校正矩阵. v 为核心变量, v 可以是 w, 也可以是 w 的部分分量

收敛性条件 对给定的预测矩阵 Q, 要求设计的校正矩阵 M 满足如下条件:

$$\exists \text{ 正定矩阵} \quad H \succ 0 \quad \text{使得} \quad HM = Q. \tag{2.4a}$$

此外, 能够保证

$$G = Q^\top + Q - M^\top HM \succ 0. \tag{2.4b}$$

校正 $v^{k+1} = v^k - M(v^k - \tilde{v}^k)$, 怎样给出满足收敛性条件的校正矩阵 M ?

$$\left\{\begin{array}{l} \text{预测 (2.2) 提供 } Q: Q^\top + Q \succ 0 \\ \text{收敛条件 (2.4)：选矩阵 } M \text{ 的要求:} \\ \exists\, H \succ 0, \text{ such that } HM = Q, \\ G = Q^\top + Q - M^\top HM \succ 0. \end{array}\right. \iff \left\{\begin{array}{rcl} D \succ 0, & & G \succ 0, \\ D + G & = & Q^\top + Q, \\ M^\top HM & = & D, \\ HM & = & Q. \end{array}\right.$$

$$\iff \left\{\begin{array}{rcl} D \succ 0, & & G \succ 0, \\ D + G & = & Q^\top + Q, \\ Q^\top M & = & D, \\ HM & = & Q. \end{array}\right. \iff \left\{\begin{array}{rcl} D \succ 0, & & G \succ 0, \\ D + G & = & Q^\top + Q, \\ M & = & Q^{-T}D, \\ H & = & QD^{-1}Q^\top. \end{array}\right.$$

现在的做法: 有了预测矩阵 Q, 可以选定 D, 使其满足 $0 \prec D \prec Q^\top + Q$.

对给定的满足 $Q^\top + Q \succ 0$ 的预测, 从好不容易凑出一个方法, 到并不费劲构造一簇算法.

由于 $M = Q^{-T}D$, 校正 (2.3) 等价于 $$Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k). \qquad (2.5)$$

3 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[3], 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

$$\begin{cases} x^{k+1} \in \arg\min\{\theta_1(x) - x^TA^T\lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \,|\, x \in \mathcal{X}\}, \\ y^{k+1} \in \operatorname{argmin}\{\theta_2(y) - y^TB^T\lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 \,|\, y \in \mathcal{Y}\}, \\ z^{k+1} \in \operatorname{argmin}\{\theta_3(z) - z^TC^T\lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By^k + Cz - b\|^2 \,|\, z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b) \end{cases} \tag{3.1}$$

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λ, 都太自由, 需要校正. 校正公式是

$$v^{k+1} := v^k - \alpha(v^k - v^{k+1}), \quad \alpha \in (0, 2 - \sqrt{2}). \tag{3.2}$$

譬如说, 我们可以取$\alpha = 0.55$. 注意到(3.2)右端的 $v^{k+1} = (y^{k+1}, z^{k+1}, \lambda^{k+1})$ 是由(3.1)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(3.1)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

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b). \tag{3.3}$$

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：

$$\left\{\begin{array}{l}\tilde{x}^k \in \arg\min\left\{\mathcal{L}_\beta^{[3]}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\right\},, \quad (3.4a)\\ \tilde{y}^k \in \arg\min\left\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\right\}, \quad (3.4b)\\ \tilde{z}^k \in \arg\min\left\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k, y^k, z, \lambda^k) \mid z \in \mathcal{Z}\right\}, \quad (3.4c)\\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b). \quad (3.4d)\end{array}\right.$$

利用增广拉格朗日函数 (1.3), 子问题 (3.4a) 相当于

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

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

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

再根据 (3.4d), 就有

用 (3.4d) 定义 $\tilde{\lambda}^k$, 可以让 (3.5a) 的 $-A^T\tilde{\lambda}^k$ 后面没有 “尾巴”

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\} \geq 0, \quad \forall x \in \mathcal{X}. \quad (3.5a)$$

子问题 (3.4b) 相当于

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

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

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

再根据 (3.4d), 就有

用 $\tilde{\lambda}^k$ 的定义, (3.5b) 中 $-B^T\tilde{\lambda}^k$ 后面的 "尾巴" 是 $\beta B^T B(\tilde{y}^k - y^k)$

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{\underline{-B^T\tilde{\lambda}^k} + \beta B^T B(\tilde{y}^k - y^k)\} \ge 0, \quad \forall y \in \mathcal{Y}. \qquad (3.5b)$$

同理, 对子问题 (3.4c) 有

$$\tilde{z}^k \in \mathcal{Z}, \quad \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T\{\underline{-C^T\tilde{\lambda}^k} + \beta C^T C(\tilde{z}^k - z^k)\} \ge 0, \quad \forall z \in \mathcal{Z}. \qquad (3.5c)$$

注意到 (3.4d) 可以写成

$$\underline{(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)} - B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + (1/\beta)\,(\tilde{\lambda}^k - \lambda^k) = 0. \qquad (3.5d)$$

把 (3.5) 中的公式组合在一起, 可以写成统一框架中的预测形式:

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \forall w \in \Omega, \qquad (3.6a)$$

其中

$$Q=\begin{pmatrix} \beta B^TB & 0 & 0 \\ 0 & \beta C^TC & 0 \\ -B & -C & \frac{1}{\beta}I \end{pmatrix}. \tag{3.6b}$$

回头来看方法 (3.1)-(3.2) 在统一框架中的校正该怎么表示. 由于

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

把 (3.4) 的输出作为预测点时, 校正公式 (3.2) 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 (3.6) 这样的预测表达式, 方法 (3.1)-(3.2) 的校正公式是

$$v^{k+1}=v^k-M(v^k-\tilde{v}^k), \tag{3.7a}$$

其中

$$M = \alpha \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \tag{3.7b}$$

对这样的 Q 和 M, 设

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

就有 $HM = Q$, 说明收敛性条件满足.

根据统一框架, 要对 (3.7b) 中的 M 找出一个 $\alpha > 0$, 使得条件

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

满足. 简单的矩阵运算得到

$$Q^T + Q = \begin{pmatrix} 2\beta B^T B & 0 & -B^T \\ 0 & 2\beta C^T C & -C^T \\ -B & -C & \frac{2}{\beta} I \end{pmatrix}$$

和

$$
\begin{aligned}
M^TQ &= \alpha \begin{pmatrix} I & 0 & -\beta B^T \\ 0 & I & -\beta C^T \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} \beta B^T B & 0 & 0 \\ 0 & \beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix} \\
&= \alpha \begin{pmatrix} 2\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & 2\beta C^T C & -C^T \\ -B & -C & \frac{1}{\beta} I \end{pmatrix}.
\end{aligned}
$$

所以有

$$
G = Q^T + Q - M^T Q = \begin{pmatrix} 2(1-\alpha)\beta B^T B & -\alpha\beta B^T C & -(1-\alpha)B^T \\ -\alpha C^T B & 2(1-\alpha)\beta C^T C & -(1-\alpha)C^T \\ -(1-\alpha)B & -(1-\alpha)C & (2-\alpha)\frac{1}{\beta} I_m \end{pmatrix}.
$$

由于

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

只要验证, 对什么样的 $\alpha > 0$, 矩阵

$$
\begin{pmatrix} 2(1-\alpha) & -\alpha & -(1-\alpha) \\ -\alpha & 2(1-\alpha) & -(1-\alpha) \\ -(1-\alpha) & -(1-\alpha) & (2-\alpha) \end{pmatrix} \succ 0. \tag{3.8}
$$

经过计算, 对所有的 $\alpha \in (0, 2-\sqrt{2})$, (3.8) 中的矩阵正定, 收敛性条件满足.

4 带高斯回代的 ADMM 方法

Direct extension of ADMM

$$\begin{cases} x^{k+1} \in \arg\min\{\mathcal{L}_\beta^{[3]}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} \in \arg\min\{\mathcal{L}_\beta^{[3]}(x^{k+1}, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\}, \\ z^{k+1} \in \arg\min\{\mathcal{L}_\beta^{[3]}(x^{k+1}, y^{k+1}, z, \lambda^k) \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta\big(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b\big). \end{cases} \tag{4.1}$$

我们在 [1] 中证明, 对三个可分离块的凸优化问题, 直接推广的(4.1) 并不保证收敛.

在此之前, 我们好不容易凑成一些求解三个可分离块凸优化问题的方法 [5, 6]

直接推广的乘子交替方向法 (4.1) 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 (4.1) 提供的

$(y^{k+1}, z^{k+1}, \lambda^{k+1})$ 当成预测点, 校正公式为

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \lambda^{k+1} \end{pmatrix} := \begin{pmatrix} y^k \\ z^k \\ \lambda^k \end{pmatrix} - \nu \begin{pmatrix} I & -(B^TB)^{-1}B^TC & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix}. \tag{4.2}$$

其中 $\nu \in (0,1)$, 右端的 $(y^{k+1}, z^{k+1}, \lambda^{k+1})$ 是由 (4.1) 提供的. 这个方法发表在[5]. 想法是不公平, 就要做找补, 调整. 事实上, 也可以就用 (4.1) 提供的 λ^{k+1}, 只通过

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} := \begin{pmatrix} y^k \\ z^k \end{pmatrix} - \nu \begin{pmatrix} I & -(B^TB)^{-1}B^TC \\ 0 & I \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}. \tag{4.3}$$

校正 y 和 z (无需校正λ). 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(4.3)更简单的

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \end{pmatrix} := \begin{pmatrix} By^k \\ Cz^k \end{pmatrix} - \nu \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} By^k - By^{k+1} \\ Cz^k - Cz^{k+1} \end{pmatrix}. \tag{4.4}$$

4.1 The prediction matrix Q — Triangular Matrix

我们把直接推广 (4.1) 中的 $u^{k+1}=(x^{k+1},y^{k+1},z^{k+1})$ 写成 $\tilde{u}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{z}^k)$,

这样就有

$$\left\{\begin{array}{l} \tilde{x}^k \in \arg\min\big\{\mathcal{L}_\beta^{[3]}(x,y^k,z^k,\lambda^k)\mid x\in\mathcal{X}\big\},\\[2mm] \tilde{y}^k \in \arg\min\big\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k,y,z^k,\lambda^k)\mid y\in\mathcal{Y}\big\},\\[2mm] \tilde{z}^k \in \arg\min\big\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k,\tilde{y}^k,z,\lambda^k)\mid z\in\mathcal{Z}\big\}.\end{array}\right. \tag{4.5}$$

x,y,z 子问题的形式是

$$\left\{\begin{array}{l} \tilde{x}^k \in \arg\min\big\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}\beta\|Ax+By^k+Cz^k-b\|^2\mid x\in\mathcal{X}\big\},\\[2mm] \tilde{y}^k \in \arg\min\big\{\theta_2(y)-y^TB^T\lambda^k+\frac{1}{2}\beta\|A\tilde{x}^k+By+Cz^k-b\|^2\mid y\in\mathcal{Y}\big\},\\[2mm] \tilde{z}^k \in \arg\min\big\{\theta_3(z)-z^TC^T\lambda^k+\frac{1}{2}\beta\|A\tilde{x}^k+B\tilde{y}^k+Cz-b\|^2\mid z\in\mathcal{Z}\big\}.\end{array}\right.$$

利用优化问题和变分不等式之间等价关系的引理 1, 得到 $\tilde{u}^k \in \mathcal{U}$,

$$\left\{\begin{array}{l}\theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T\big\{-A^T\lambda^k \\ \qquad\qquad\qquad +\beta A^T\big(A\tilde{x}^k+By^k+Cz^k-b\big)\big\}\ge 0, \quad \forall\, x\in\mathcal{X}, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T\big\{-B^T\lambda^k \\ \qquad\qquad\qquad +\beta B^T\big(A\tilde{x}^k+B\tilde{y}^k+Cz^k-b\big)\big\}\ge 0, \quad \forall\, y\in\mathcal{Y}, \\ \theta_3(z)-\theta_3(\tilde{z}^k)+(z-\tilde{z}^k)^T\big\{-C^T\lambda^k \\ \qquad\qquad\qquad +\beta C^T\big(A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b\big)\big\}\ge 0, \quad \forall\, z\in\mathcal{Z}.\end{array}\right. \tag{4.6}$$

定义

$$\tilde{\lambda}^k=\lambda^k-\beta\big(A\tilde{x}^k+By^k+Cz^k-b\big), \tag{4.7}$$

上式可以写成等价的等式

$$(A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b)-B(\tilde{y}^k-y^k)-C(\tilde{z}^k-z^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k)=0. \tag{4.8}$$

对于给定的 $\tilde{\lambda}^k\in\Re^m$ 和 0 向量 p, 相应的关系式也可以写成

$$\tilde{\lambda}^k\in\Re^m, \quad (\lambda-\tilde{\lambda}^k)^Tp\ge 0, \quad \forall\lambda\in\Re^m.$$

将 (4.6) 和 (4.8) 加在一起, 利用变分不等式形式 (1.2), 我们得到 $\tilde{w}^k \in \Omega$,

$$\left\{\begin{array}{l} \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\}\ge 0, \quad \forall\, x\in\mathcal{X}, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T\{\underline{-B^T\tilde{\lambda}^k}+\beta B^TB(\tilde{y}^k-y^k)\}\ge 0, \quad \forall\, y\in\mathcal{Y}, \\ \theta_3(z)-\theta_3(\tilde{z}^k)+(z-\tilde{z}^k)^T\left\{\begin{array}{c} \underline{-C^T\tilde{\lambda}^k}+\beta C^TB(\tilde{y}^k-y^k) \\ +\beta C^TC(\tilde{z}^k-z^k) \end{array}\right\}\ge 0, \;\; \forall\, z\in\mathcal{Z}, \\ (\lambda-\tilde{\lambda}^k)^T\left\{\begin{array}{l} (\underline{A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b}) \\ -B(\tilde{y}^k-y^k)-C(\tilde{z}^k-z^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k) \end{array}\right\}\ge 0, \;\; \forall\, \lambda\in\Lambda. \end{array}\right. \tag{4.9}$$

注意到 (4.9) 式中加下划线的部分恰好是 (1.2) 中定义的 $F(\tilde{w}^k)$, 合并写成 $\tilde{w}^k \in \Omega$,

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge (v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k), \; \forall w\in\Omega, \tag{4.10}$$

其中向量 $v=(y,z,\lambda)$ 预测矩阵

$$Q=\begin{pmatrix} \beta B^TB & 0 & 0 \\ \beta C^TB & \beta C^TC & 0 \\ -B & -C & \frac{1}{\beta}I_m \end{pmatrix}. \tag{4.11}$$

校正: 利用这样的预测点, 只校正 y 和 z 的公式 (4.3) (注意λ^{k+1} 和 $\tilde{\lambda}^k$ 的关系) 就可以写成

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

也就是说, 在统一框架的校正公式中, 取

$$M = \begin{pmatrix} \nu I & -\nu(B^TB)^{-1}B^TC & 0 \\ 0 & \nu I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \tag{4.12}$$

对于矩阵

$$H = \begin{pmatrix} \frac{1}{\nu}\beta B^TB & \frac{1}{\nu}\beta B^TC & 0 \\ \frac{1}{\nu}\beta C^TB & \frac{1}{\nu}\beta[C^TC + C^TB(B^TB)^{-1}B^TC] & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}, \tag{4.13}$$

可以验证 $HM = Q$. 通过合同变换

$$\begin{pmatrix} I & 0 \\ -C^T B(B^T B)^{-1} & I \end{pmatrix} \begin{pmatrix} B^T B & B^T C \\ C^T B & C^T C + C^T B(B^T B)^{-1} B^T C \end{pmatrix} \begin{pmatrix} I & -(B^T B)^{-1} B^T C \\ 0 & I \end{pmatrix}$$

$$= \begin{pmatrix} B^T B & B^T C \\ 0 & C^T C \end{pmatrix} \begin{pmatrix} I & -(B^T B)^{-1} B^T C \\ 0 & I \end{pmatrix} = \begin{pmatrix} B^T B & 0 \\ 0 & C^T C \end{pmatrix},$$

得知 H 在 B, C 列满秩时正定. 此外,

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

由于 $\nu \in (0, 1)$, 当 B, C 列满秩时矩阵 G 正定. 统一框架中的收敛性条件满足.

5 Implement the correction by using (2.5)

对 (4.11) 中的预测矩阵 Q, 我们有

$$
\begin{aligned}
Q^T + Q &= \begin{pmatrix} 2\beta B^TB & \beta B^TC & -B^T \\ \beta C^TB & 2\beta C^TC & -C^T \\ -B & -C & \frac{2}{\beta} I_m \end{pmatrix} \\
&= \begin{pmatrix} B^T & 0 & 0 \\ 0 & C^T & 0 \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} 2\beta I_m & \beta I_m & -I_m \\ \beta I_m & 2\beta I_m & -I_m \\ -I_m & -I_m & \frac{2}{\beta} I_m \end{pmatrix} \begin{pmatrix} B & 0 & 0 \\ 0 & C & 0 \\ 0 & 0 & I_m \end{pmatrix}. \qquad (5.1)
\end{aligned}
$$

由于

$$
\begin{pmatrix} 2\beta I_m & \beta I_m & -I_m \\ \beta I_m & 2\beta I_m & -I_m \\ -I_m & -I_m & \frac{2}{\beta} I_m \end{pmatrix} = \begin{pmatrix} \beta I_m & \beta I_m & -I_m \\ \beta I_m & \beta I_m & -I_m \\ -I_m & -I_m & \frac{1}{\beta} I_m \end{pmatrix} + \begin{pmatrix} \beta I_m & 0 & 0 \\ 0 & \beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}
$$

是正定矩阵, 当矩阵 B 和 C 是列满秩矩阵时, 矩阵 $Q^T + Q$ 正定。

选择 $0 \prec D \prec Q^T + Q$, 可以提出自己想要的方法, 下面只是一些例子而已.

取一个比较简单的 D

对任意的 $v \in (0,1)$, 矩阵

$$\begin{pmatrix} 2\beta I_m & \beta I_m & -I_m \\ \beta I_m & 2\beta I_m & -I_m \\ -I_m & -I_m & \frac{2}{\beta} I_m \end{pmatrix} = \begin{pmatrix} \nu\beta I_m & 0 & 0 \\ 0 & \nu\beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} + \begin{pmatrix} (2-\nu)\beta I_m & \beta I_m & -I_m \\ \beta I_m & (2-\nu)\beta I_m & -I_m \\ -I_m & -I_m & \frac{1}{\beta} I_m \end{pmatrix}$$

分拆成了两个正定矩阵. 因此, 可以选

$$\begin{aligned} D &= \begin{pmatrix} B^T & 0 & 0 \\ 0 & C^T & 0 \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} \nu\beta I & 0 & 0 \\ 0 & \nu\beta I & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} B & 0 & 0 \\ 0 & C & 0 \\ 0 & 0 & I_m \end{pmatrix} \\ &= \begin{pmatrix} \nu\beta B^T B & 0 & 0 \\ 0 & \nu\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \end{aligned} \tag{5.2}$$

这时,

$$G = Q^T + Q - D = \begin{pmatrix} (2-\nu)\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & (2-\nu)\beta C^T C & -C^T \\ -B & -C & \frac{1}{\beta} I_m \end{pmatrix}. \qquad (5.3)$$

Algorithms for the model (1.1)

[Prediction Step.] Obtain $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$ via the direct extension of the ADMM (4.5) and define $\tilde{\lambda}^k$ by (4.7).

[Correction Step.] Get v^{k+1} by solving $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$.

问题归结为如何从 $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$ 求出 v^{k+1} ? 我们知道

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

和

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

对矩阵 Q^T 和 D 的分解有相同的左因子. 因此, 求解方程组

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

可以通过

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

求得. 上述线性方程组等价于方程组

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

用回代的方法依次求得 $(\lambda^{k+1}-\lambda^k)$, $(Cz^{k+1}-Cz^k)$, $(By^{k+1}-By^k)$, 然后得到开始下一次迭代所需要的 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$.

选择 D 的一些其他方法

将 (5.2) 和 (5.3) 中的 D 和 G 互换位置, 换句话说, 取

$$D = \begin{pmatrix} (2-\nu)\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & (2-\nu)\beta C^T C & -C^T \\ -B & -C & \frac{1}{\beta} I_m \end{pmatrix}.$$

对于同样的预测, 校正可以通过

$$\begin{pmatrix} I & I & -\frac{1}{\beta} I \\ 0 & I & -\frac{1}{\beta} I \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} By^{k+1} - By^k \\ Cz^{k+1} - Cz^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = \begin{pmatrix} (2-\nu)I & I & -I \\ I & (2-\nu)I & -I \\ -I & -I & I \end{pmatrix} \begin{pmatrix} B\tilde{y}^k - By^k \\ C\tilde{z}^k - Cz^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}.$$

得到开始下一次迭代所需要的 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$.

选择 $D=\alpha(Q^T+Q),\ \alpha\in(0,1)$ 的方法

这时 $D=\alpha(Q^T+Q)$ 和 $G=(1-\alpha)(Q^T+Q)$ 都是正定矩阵.

$$D=\alpha\left[Q^T+Q\right]=\alpha\begin{pmatrix} 2\beta B^TB & \beta B^TC & -B^T \\ \beta C^TB & 2\beta C^TC & -C^T \\ -B & C & \frac{2}{\beta}I_m \end{pmatrix}. \tag{5.5}$$

对于同样的预测, 校正可以通过

$$\begin{pmatrix} I & I & -\frac{1}{\beta}I \\ 0 & I & -\frac{1}{\beta}I \\ 0 & 0 & I \end{pmatrix}\begin{pmatrix} By^{k+1}-By^k \\ Cz^{k+1}-Cz^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix}=\alpha\begin{pmatrix} 2I & I & -I \\ I & 2I & -I \\ -I & -I & 2I \end{pmatrix}\begin{pmatrix} B\tilde{y}^k-By^k \\ C\tilde{z}^k-Cz^k \\ \tilde{\lambda}^k-\lambda^k \end{pmatrix}.$$

得到开始下一次迭代所需要的 $(By^{k+1},Cz^{k+1},\lambda^{k+1})$.

从好不容易凑出一个方法, 到并不费力给出一簇算法.

6 部分平行并加正则项的 ADMM 方法

我们已经知道直接推广的 ADMM 求解三个可分离块的凸优化问题不能保证收敛 [1], 原因应该是对原始核心变量中 y 和 z 的子问题处理先后显得不够公平, 在 §4 采用了回代的方法. 然而, 如下的简单强制平行的方法也不能保证收敛.

简单地强制 y 和 z 平等 不能保证方法收敛

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

下面我们考虑强制平行, 并通过另加正则项直接解决问题

y, z 子问题平行, 如果不想做后处理, 就给它们俩预先都加个正则项

$$\left\{\begin{array}{l} x^{k+1} = \arg\min\left\{\mathcal{L}_\beta^{3}(x, y^k, z^k, \lambda^k) \mid x\in\mathcal{X}\right\}, \qquad (\tau > 0 \text{ 为参数})\\ y^{k+1} = \arg\min\left\{\mathcal{L}_\beta^{3}(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\beta\|B(y-y^k)\|^2 \big| y\in\mathcal{Y}\right\},\\ z^{k+1} = \arg\min\left\{\mathcal{L}_\beta^{3}(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\beta\|C(z-z^k)\|^2 \big| z\in\mathcal{Z}\right\},\\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b).\end{array}\right.$$

上述做法相当于

$$\left\{\begin{array}{l} x^{k+1} \in \arg\min\left\{\theta_1(x) - x^T A^T\lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \,\middle|\, x\in\mathcal{X}\right\}, \\ y^{k+1} \in \arg\min\left\{\begin{array}{c} \theta_2(y) - y^T B^T\lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 \\ +\frac{\tau}{2}\beta\|B(y-y^k)\|^2 \end{array} \middle| y\in\mathcal{Y}\right\}, \\ z^{k+1} \in \arg\min\left\{\begin{array}{c} \theta_3(z) - z^T C^T\lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By^k + Cz - b\|^2 \\ +\frac{\tau}{2}\beta\|C(z-z^k)\|^2 \end{array} \middle| z\in\mathcal{Z}\right\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{array}\right.$$

注意到

$$\begin{array}{rcl} y^{k+1} & \in & \arg\min\left\{\begin{array}{c} \theta_2(y) - y^T B^T\lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 \\ +\frac{\tau}{2}\beta\|B(y-y^k)\|^2 \end{array} \middle| y\in\mathcal{Y}\right\}, \\ & = & \arg\min\left\{\begin{array}{c} \theta_2(y) + \frac{\beta}{2}\|(Ax^{k+1} + By^k + Cz^k - b) + B(y-y^k)\|^2 \\ -y^T B^T\lambda^k + \frac{\tau}{2}\beta\|B(y-y^k)\|^2 \end{array} \middle| y\in\mathcal{Y}\right\} \\ & = & \arg\min\left\{\begin{array}{c} \theta_2(y) - y^T B^T[\lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b)] \\ +\frac{1}{2}\beta\|B(y-y^k)\|^2 + \frac{\tau}{2}\beta\|B(y-y^k)\|^2 \end{array} \middle| y\in\mathcal{Y}\right\}. \end{array}$$

所以, 若令

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b),$$

这个方法就是

$$\left\{\begin{array}{l} x^{k+1} \in \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 | x \in \mathcal{X}\}, \\ \qquad \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\ y^{k+1} \in \operatorname{argmin}\{\theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ z^{k+1} \in \operatorname{argmin}\{\theta_3(z) - z^T C^T \lambda^{k+\frac{1}{2}} + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{array}\right. \tag{6.1}$$

其中 $\mu = \tau + 1$. 我们讨论需要多大的 μ.

把由 (6.1) 生成的

$$(x^{k+1}, y^{k+1}, z^{k+1}, \lambda^{k+\frac{1}{2}}) \quad \text{视为预测点} \quad (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k), \tag{6.2}$$

这个预测公式就成为

$$\left\{\begin{array}{l} \tilde{x}^k = \text{argmin}\{\theta_1(x) - x^TA^T\lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, \\ \tilde{y}^k = \text{argmin}\{\theta_2(y) - y^TB^T\tilde{\lambda}^k + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ \tilde{z}^k = \text{argmin}\{\theta_3(z) - z^TC^T\tilde{\lambda}^k + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b). \end{array}\right. \tag{6.3}$$

预测 (6.3) 中 x-子问题的最优性条件是

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

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

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

再根据 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b)$, 就有

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \ \ \forall x \in \mathcal{X}. \tag{6.4a}$$

同样根据最优性条件引理, 预测 (6.3) 中 y-子问题的最优性条件是

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{\underline{-B^T\tilde{\lambda}^k} + \mu\beta B^T B(\tilde{y}^k - y^k)\} \ge 0, \quad \forall y \in \mathcal{Y}. \tag{6.4b}$$

同理, 预测 (6.3) 中 z-子问题的最优性条件是

$$\tilde{z}^k \in \mathcal{Z}, \quad \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T \{\underline{-C^T\tilde{\lambda}^k} + \mu\beta C^T C(\tilde{z}^k - z^k)\} \ge 0, \quad \forall\, z \in \mathcal{Z}. \tag{6.4c}$$

根据 $\tilde{\lambda}^k$ 的定义, 我们有

$$\underline{(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)} - B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + (1/\beta)\,(\tilde{\lambda}^k - \lambda^k) = 0. \tag{6.4d}$$

这样, 利用最优性引理和变分不等式 (1.2) 的形式, 预测就可以写成统一框架中的形式:

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \forall w \in \Omega, \tag{6.5a}$$

其中

$$Q = \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix}. \tag{6.5b}$$

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

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

利用这样的预测点, 校正 y 和 z 的公式 (注意 λ^{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

$$M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \tag{6.6}$$

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

$$\begin{aligned} HM &= \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix} \\ &= \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix} = Q \end{aligned}$$

此外,

$$G = (Q^T + Q) - M^T H M = (Q^T + Q) - M^T Q$$

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

$$= \begin{pmatrix} (\mu-1)\beta B^T B & -\beta B^T C & 0 \\ -\beta C^T B & (\mu-1)\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}.$$

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.

例如, 可以取 $\mu = 2.01$. 这类发表在 [6, 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. [6] 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].

This method is accepted by Osher's research group

- 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.

IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 21, NO. 7, JULY 2012 3239

A Convex Model for Nonnegative Matrix Factorization and Dimensionality Reduction on Physical Space

Ernie Esser, Michael Möller, Stanley Osher, Guillermo Sapiro, *Senior Member, IEEE*, and Jack Xin

$$\min_{T\geq 0, V_j\in D_j, e\in E} \zeta \sum_i \max_j (T_{i,j}) + \langle R_w \sigma C_w, T\rangle$$

$$\text{such that} \quad YT - X_s = V - X_s \mathrm{diag}(e). \quad (15)$$

Since the convex functional for the extended model (15) is slightly more complicated, it is convenient to use a variant of ADMM that allows the functional to be split into more than two parts. The method proposed by He *et al.* in [34] is appropriate for this application. Again, introduce a new variable Z

Using the ADMM-like method in [34], a saddle point of the augmented Lagrangian can be found by iteratively solving the subproblems with parameters $\delta > 0$ and $\mu > 2$, shown in the

tion refinement step. Due to the different algorithm used to solve the extended model, there is an additional numerical parameter μ, which for this application must be greater than two according to [34]. We set μ equal to 2.01. There are also model parame-

[33] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis," 2009 [Online]. Available: http://arxiv.org/PS cache/arxiv/pdf/0912/0912.3599v1.pdf
[34] B. He, M. Tao, and X. Yuan, "A splitting method for separate convex programming with linking linear constraints," Tech. Rep., 2011 [Online]. Available: http://www.optimization-online.org/DB_FILE/2010/06/ 2665.pdf

ADMM + Parallel-Prox Splitting ALM

各自为政, 过分自由. 给它们加个适当的正则项$(\tau > 1)$, 方法就能保证收敛.

$$
\begin{cases}
x^{k+1} = \arg\min\{\mathcal{L}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, & \text{(6.7a)} \\
\begin{cases}
y^{k+1} = \arg\min\{\mathcal{L}(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\
z^{k+1} = \arg\min\{\mathcal{L}(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\},
\end{cases} & \text{(6.7b)} \\
\lambda^{k+1} = \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). & \text{(6.7c)}
\end{cases}
$$

Notice that (6.7b) can be written as

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} = \arg\min\Big\{\mathcal{L}(x^{k+1}, y, z, \lambda^k) + \frac{1}{2}\Big\| \begin{array}{c} y - y^k \\ z - z^k \end{array} \Big\|^2_{D_{BC}} \Big| \begin{array}{l} y \in \mathcal{Y} \\ z \in \mathcal{Z} \end{array} \Big\},$$

where

$$D_{BC} = \begin{pmatrix} \tau B^T B & -B^T C \\ -C^T B & \tau C^T C \end{pmatrix}. \tag{6.8}$$

D_{BC} is positive semidefinite when $\tau \geq 1$.

However, the matrix D_{BC} is indefinite for $\tau \in (0,1)$.

In other words, the scheme (6.7) can be rewritten as

$$\left\{\begin{array}{rcl} x^{k+1} &=& \arg\min\{\mathcal{L}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ \begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} &=& \arg\min\Big\{\mathcal{L}(x^{k+1}, y, z, \lambda^k) + \frac{1}{2}\Big\| \begin{array}{c} y - y^k \\ z - z^k \end{array} \Big\|^2_{D_{BC}} \Big| \begin{array}{l} y \in \mathcal{Y} \\ z \in \mathcal{Z} \end{array} \Big\}, \\ \lambda^{k+1} &=& \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{array}\right.$$

The algorithm (6.7) can be rewritten in an equivalent form: $(\mu = \tau + 1 > 2)$.

$$\left\{\begin{array}{l} x^{k+1} = \arg\min\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b - \frac{1}{\beta}\lambda^k\|^2 \,|\, x \in \mathcal{X}\}, \\ \qquad \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\ y^{k+1} = \arg\min\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^T Cz + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{array}\right. \tag{6.9}$$

The related publication:

- B. He, M. Tao and X. Yuan, A splitting method for separable convex programming. IMA J. Numerical Analysis, 31(2015), 394-426.

In the above paper, in order to ensure the convergence, it **was** required

$\tau > 1$ (in (6.7)) which is equivalent to $\mu > 2$ (in (6.9)).

This method is accepted by Osher's research group

- 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.

> tion refinement step. Due to the different algorithm used to solve the extended model, there is an additional numerical parameter μ, which for this application must be greater than two according to [34]. We set μ equal to 2.01. There are also model parame-

Thus, Osher's research group utilize the iterative formula (6.9), according to our previous paper, they set

$$\mu = 2.01, \qquad \text{it is only a pity larger than 2.}$$

Large parameter μ (or τ) will lead a slow convergence.

最新进展：最优正则化因子的选择– OO6235 的结论

Bingsheng He, Xiaoming 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. Optimization Online 6235.

Our new assertion: In (6.7)

- if $\tau > 0.5$, the method is still convergent;
- if $\tau < 0.5$, there is divergent example.

Equivalently in (6.9) :

- if $\mu > 1.5$, the method is still convergent;
- if $\mu < 1.5$, there is divergent example.

For convex optimization problem (1.1) with three separable objective functions, the parameters in the equivalent methods (6.7) and (6.9) :

- **0.5** is the threshold factor of the parameter τ in (6.7) !
- **1.5** is the threshold factor of the parameter μ in (6.9) !

7 利用统一框架设计的 PPA 算法

求解变分不等式 (1.2) 的 PPA 型算法要求预测 (2.2) 中的矩阵 Q 本身是一个能写成 H 的对称正定矩阵. 这时, 我们把相应的矩阵 Q 记为 H. 这类方法中, 我们用平凡松弛的校正 (2.3) 给出 v^{k+1}, 其中 $M = \alpha I$, 实际运算中, 一般取 $\alpha \in [1.2, 1.8]$.

如果我们为求解 (1.2) 构造的预测公式中的 $\tilde{w}^k$ 满足

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{7.1a}$$

其中

$$H = \begin{pmatrix} \beta B^T B + \delta I_m & 0 & -B^T \\ 0 & \beta C^T C + \delta I_m & -C^T \\ -B & -C & \frac{2}{\beta} I_m \end{pmatrix}, \tag{7.1b}$$

其中 $\beta > 0$ 和 $\delta > 0$ 都是任意给定的大于零的常数. 由于

$$H = \begin{pmatrix} \beta B^T B + \delta I_m & 0 & -B^T \\ 0 & 0 & 0 \\ -B & 0 & \frac{1}{\beta} I_m \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta C^T C + \delta I_m & -C^T \\ 0 & -C & \frac{1}{\beta} I_m \end{pmatrix},$$

对任意的 $\beta > 0, \delta > 0$ 和 $v = (y, z, \lambda) \neq 0$,

$$v^T H v = \Big\| \sqrt{\beta} B y - \frac{1}{\sqrt{\beta}} \lambda \Big\|^2 + \Big\| \sqrt{\beta} C z - \frac{1}{\sqrt{\beta}} \lambda \Big\|^2 + \delta(\|y\|^2 + \|z\|^2) > 0.$$

矩阵 H 是正定的. 我们用平凡松弛的校正 (2.3) 得到新的迭代点 v^{k+1}. 根据统一框架, 算法就是收敛的. 因此, 问题归结为如何实现满足 (7.1) 的预测. 用 (1.2) 中 $F(w)$ 的表达式, 把 (7.1) 的具体形式写出来就是 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$, 使得

$$\left\{ \begin{array}{l} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{ \underline{-A^T \tilde{\lambda}^k} \} \geq 0, \quad \forall\, x \in \mathcal{X}, \qquad \text{(7.2a)} \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{ \underline{-B^T \tilde{\lambda}^k} + \beta B^T B(\tilde{y}^k - y^k) + \delta(\tilde{y}^k - y^k) \\ \qquad\qquad -B^T(\tilde{\lambda}^k - \lambda^k) \} \geq 0, \quad \forall\, y \in \mathcal{Y}, \qquad \text{(7.2b)} \\ \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T \{ \underline{-C^T \tilde{\lambda}^k} + \beta C^T C(\tilde{z}^k - z^k) + \delta(\tilde{z}^k - z^k) \\ \qquad\qquad -C^T(\tilde{\lambda}^k - \lambda^k) \} \geq 0, \quad \forall\, z \in \mathcal{Z}, \qquad \text{(7.2c)} \\ \qquad (\underline{A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b}) \\ \qquad\qquad -B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + (2/\beta)\,(\tilde{\lambda}^k - \lambda^k) = 0. \qquad \text{(7.2d)} \end{array} \right.$$

上式中, 有下划线的凑在一起, 就是 (7.1) 中的 $F(\tilde{w}^k)$. 把 (7.2) 的具体形式写出来就

是 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$, 使得

$$\left\{\begin{array}{ll} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\tilde{\lambda}^k\} \ge 0, \quad \forall\, x \in \mathcal{X}, & (7.3a) \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T(2\tilde{\lambda}^k - \lambda^k) & \\ \qquad +\beta B^T B(\tilde{y}^k - y^k) + \delta(\tilde{y}^k - y^k)\} \ge 0, \quad \forall\, y \in \mathcal{Y}, & (7.3b) \\ \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T C^T(2\tilde{\lambda}^k - \lambda^k) & \\ \qquad +\beta C^T C(\tilde{z}^k - z^k) + \delta(\tilde{z}^k - z^k)\} \ge 0, \quad \forall\, z \in \mathcal{Z}, & (7.3c) \\ (A\tilde{x}^k + By^k + Cz^k - b) + (2/\beta)\,(\tilde{\lambda}^k - \lambda^k) = 0. & (7.3d) \end{array}\right.$$

如果令

$$\tilde{x}^k = \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \tfrac{1}{4}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, \tag{7.4}$$

根据最优性质的定理, 问题 (7.4) 的最优性条件是 $\tilde{x}^k \in \mathcal{X}$,

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\lambda^k + \tfrac{1}{2}\beta A^T(A\tilde{x}^k + By^k + Cz^k - b)\} \ge 0, \quad \forall x \in \mathcal{X}. \tag{7.5}$$

再定义

$$\tilde{\lambda}^k = \lambda^k - \tfrac{1}{2}\beta(A\tilde{x}^k + By^k + Cz^k - b). \tag{7.6}$$

将 (7.6) 代入 (7.5), 满足了 (7.3a). 注意到 (7.6) 又和 (7.3d) 等价. 这样, 有了 $\tilde{\lambda}^k$, 要得到满

足 (7.3b) 的 $\tilde{y}^k$ 和满足 (7.3c) 的 $\tilde{z}^k$, 根据最优性质的定理, 只要分别通过

$$\tilde{y}^k = \text{argmin}\left\{ \begin{array}{c} \theta_2(y) - y^T B^T[2\tilde{\lambda}^k - \lambda^k] + \\ \frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \end{array} \Bigg| \, y \in \mathcal{Y} \right\}$$

和

$$\tilde{z}^k = \text{argmin}\left\{ \begin{array}{c} \theta_3(z) - z^T C^T[2\tilde{\lambda}^k - \lambda^k] + \\ \frac{1}{2}\beta\|C(z - z^k)\|^2 + \frac{1}{2}\delta\|z - z^k\|^2 \end{array} \Bigg| \, z \in \mathcal{Z} \right\}$$

得到. 综上所述, 按照 $x, \lambda, (y,\ z)$ 顺序计算:

$$\left\{ \begin{array}{ll} \tilde{x}^k \in \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{4}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, & \text{(7.7a)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b), & \text{(7.7b)} \\ \tilde{y}^k = \text{argmin}\left\{\theta_2(y) - y^T B^T[2\tilde{\lambda}^k - \lambda^k] + \begin{pmatrix} \frac{1}{2}\beta\|B(y - y^k)\|^2 \\ +\frac{1}{2}\delta\|y - y^k\|^2 \end{pmatrix} \Bigg| \, y \in \mathcal{Y}\right\}, & \text{(7.7c)} \\ \tilde{z}^k = \text{argmin}\left\{\theta_3(z) - z^T C^T[2\tilde{\lambda}^k - \lambda^k] + \begin{pmatrix} \frac{1}{2}\beta\|C(z - z^k)\|^2 \\ +\frac{1}{2}\delta\|z - z^k\|^2 \end{pmatrix} \Bigg| \, z \in \mathcal{Z}\right\}, & \text{(7.7d)} \end{array} \right.$$

就得到满足条件 (7.1) 的预测点. 由于预测中的矩阵对称正定, 新的迭代点可以利用预测点继续进行平凡的松弛校正得到.

References

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

[4] B.S. He, F. Ma and X.M. Yuan, Optimally linearizing the alternating direction method of multipliers for convex programming, Comput. Optim. Appl. 75 (2020), 361-388.

[5] 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.

[6] 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.

[7] 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,

[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.