两页纸证明乘子交替方向法的关键性收缩性质

以变分不等式为工具 maths.nju.edu.cn/~hebma 何炳生

交替方向法处理的是两个可分离块的凸优化问题

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\}. \tag{1}$$

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 (x^*,y^*,λ^*) 满足

$$(x^*,y^*,\lambda^*)\in\Omega,\quad L(x^*,y^*,\lambda)\le L(x^*,y^*,\lambda^*)\le L(x,y,\lambda^*),\quad \forall(x,y,\lambda)\in\Omega, \tag{2}$$

其中 $\Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m$. 鞍点不等式 (2) 可以写成

$$\left\{\begin{array}{lll} x^*\in\mathcal{X}, & L(x,y^*,\lambda^*)-L(x^*,y^*,\lambda^*)\ge 0, & \forall x\in\mathcal{X},\\ y^*\in\mathcal{Y}, & L(x^*,y,\lambda^*)-L(x^*,y^*,\lambda^*)\ge 0, & \forall y\in\mathcal{Y},\\ \lambda^*\in\Re^m, & L(x^*,y^*,\lambda^*)-L(x^*,y^*,\lambda)\ge 0, & \forall \lambda\in\Re^m. \end{array}\right. \tag{3}$$

利用拉格朗日函数的表达式, 将 (3) 的每个不等式写出来, 得到紧凑的变分不等式

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{4a}$$

其中

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\end{pmatrix},\quad \theta(u)=\theta_1(x)+\theta_2(y),\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}. \tag{4b}$$

ADMM 的 k 步迭代是从给定的 $v^k=(y^k,\lambda^k)$ 出发, 按下面的顺序求得 $(x^{k+1},y^{k+1},\lambda^{k+1})$:

$$\left\{\begin{array}{l} x^{k+1}\in\arg\min\big\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}\beta\|Ax+By^k-b\|^2 \;\big|\; x\in\mathcal{X}\big\}, \quad (5a)\\ y^{k+1}\in\arg\min\big\{\theta_2(y)-y^TB^T\lambda^k+\frac{1}{2}\beta\|Ax^{k+1}+By-b\|^2 \;\big|\; y\in\mathcal{Y}\big\}, \quad (5b)\\ \lambda^{k+1}=\ \lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). \quad (5c) \end{array}\right.$$

根据凸优化的最优性原理, ADMM k-步迭代 (5) 的三个子问题的最优性条件是

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

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

$$\left\{\begin{array}{ll} x^{k+1}\in\mathcal{X}, & \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{-A^T\lambda^{k+1}+\beta A^TB(y^k-y^{k+1})\}\ge 0,\ \forall x\in\mathcal{X}, \quad (6a)\\ y^{k+1}\in\mathcal{Y}, & \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{-B^T\lambda^{k+1}\}\ge 0,\ \forall y\in\mathcal{Y}, \quad (6b)\\ \lambda^{k+1}\in\Re^m, & (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\ge 0,\ \forall\lambda\in\Re^m. \quad (6c) \end{array}\right.$$

在不等式 (6b) 的左端加上和为零的两项 (用下括号括起来的两项), 改写成

$$\left\{\begin{array}{l} \underline{\theta_1(x)-\theta_1(x^{k+1})}+(x-x^{k+1})^T\{-A^T\lambda^{k+1}+\beta A^TB(y^k-y^{k+1})\}\ge 0,\\ \underline{\theta_2(y)-\theta_2(y^{k+1})}+(y-y^{k+1})^T\{-B^T\lambda^{k+1}+\underbrace{\beta B^TB(y^k-y^{k+1})+\beta B^TB(y^{k+1}-y^k)}\}\ge 0,\\ \qquad (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\ge 0. \end{array}\right.$$

利用变分不等式 (4), 进行合理整合 (分别把同一颜色的并在一起), 得到

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

将上式中那个任意的 w, 设成解点 w^* 便有

$$\theta(u^*)-\theta(u^{k+1})+(w^*-w^{k+1})^T F(w^{k+1})$$
$$+\begin{pmatrix} x^*-x^{k+1} \\ y^*-y^{k+1} \end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k-y^{k+1})+\begin{pmatrix} y^*-y^{k+1} \\ \lambda^*-\lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix} \ge 0. \quad (7)$$

若用记号

$$v=\begin{pmatrix} y \\ \lambda \end{pmatrix} \quad \text{和} \quad H=\begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}, \quad (8)$$

不等式 (7) 就可以改写成

$$(v^{k+1}-v^*)^T H(v^k-v^{k+1}) \ge \begin{pmatrix} x^{k+1}-x^* \\ y^{k+1}-y^* \end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k-y^{k+1})$$
$$+\underbrace{[\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^T F(w^{k+1})]}. \quad (9)$$

下面我们证明 (9) 式右端非负. 由于

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

这就是说, (9) 式右端第二部分 (下括号部分) 非负.

对 (9) 式右端的第一部分, 有

$$\begin{pmatrix} x^{k+1}-x^* \\ y^{k+1}-y^* \end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k-y^{k+1})=(y^k-y^{k+1})^T B^T \beta(A,B) \begin{pmatrix} x^{k+1}-x^* \\ y^{k+1}-y^* \end{pmatrix}$$
$$=(y^k-y^{k+1})^T B^T \beta\big(Ax^{k+1}+By^{k+1}-(Ax^*+By^*)\big) \qquad \text{利用}(Ax^*+By^*=b)$$
$$=(y^k-y^{k+1})B^T \beta\big(Ax^{k+1}+By^{k+1}-b\big)=(y^k-y^{k+1})B^T(\lambda^k-\lambda^{k+1}). \quad (10)$$

再利用 (6b) 有

$$\theta_2(y)-\theta_2(y^{k+1}) + (y-y^{k+1})^T\{-B^T\lambda^{k+1}\} \ge 0, \ \forall\, y\in\mathcal{Y},$$
$$\text{和} \ \theta_2(y)-\theta_2(y^k) + (y-y^k)^T\{-B^T\lambda^k\} \ge 0, \ \forall\, y\in\mathcal{Y}.$$

(将上面两式中任意的 y 分别设成 y^k 和 y^{k+1})

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

(将上面两式相加, 就有) $(y^k-y^{k+1})B^T(\lambda^k-\lambda^{k+1}) \ge 0.$ ((10) 式右端非负)

证明了(10) 式右端非负, 进而得到 (9) 式右端非负. 从 (9) 式得到

$$(v^{k+1}-v^*)^T H(v^k-v^{k+1}) \ge 0. \quad (11)$$

简单的运算可以验证: 若有 $b^T H(a-b)\ge 0$ 就有 $\|b\|_H^2 \le \|a\|_H^2-\|a-b\|_H^2$.

置 $a=(v^k-v^*)$ 和 $b=(v^{k+1}-v^*)$, 根据 (11) 就得到 ADMM 收敛的关键不等式

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