C-利用变分不等式证明乘子交替方向法 (ADMM) 的关键收敛性质

以变分不等式为工具的证明方法 maths.nju.edu.cn/~hebma 何炳生

1 两个可分离块的凸优化问题和相应的变分不等式

乘子交替方向法(ADMM)处理的是两个可分离块的凸优化问题

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\}, \tag{1.1}$$

其中 $\theta_1(x),\theta_2(y)$ 是凸函数, A, B, b 是相应的矩阵和向量, $\mathcal{X},\mathcal{Y}$ 是闭凸集. 问题 (1.1) 的拉格朗日函数是

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

我们致力于求问题 (1.1) 的拉格朗日函数的鞍点 (x^*,y^*,λ^*). 鞍点满足

$$(x^*,y^*,\lambda^*)\in\Omega,\quad L(x,y,\lambda^*)\ge L(x^*,y^*,\lambda^*)\ge L(x^*,y^*,\lambda),\quad \forall (x,y,\lambda)\in\Omega, \tag{1.2}$$

其中 $\Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m$. 鞍点中的 (x^*,y^*) 就是原问题 (1.1) 的解点.

刻画鞍点的不等式 (1.2), 对 x,y,λ 分开来写就是

$$\left\{\begin{array}{lll} x^*\in\mathcal{X}, & L(x,y^*,\lambda^*)-L(x^*,y^*,\lambda^*)\ge 0, & \forall x\in\mathcal{X},\\ y^*\in\mathcal{Y}, & L(x^*,y,\lambda^*)-L(x^*,y^*,\lambda^*)\ge 0, & \forall y\in\mathcal{Y},\\ \lambda^*\in\Re^m, & L(x^*,y^*,\lambda^*)-L(x^*,y^*,\lambda)\ge 0, & \forall \lambda\in\Re^m. \end{array}\right. \tag{1.3}$$

利用拉格朗日函数 $L(x,y,\lambda)$ 的表达式, 将 (1.3) 的每个不等式写出来, 就能得到下面紧凑的变分不等式

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{1.4a}$$

其中

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\end{pmatrix},\quad \theta(u)=\theta_1(x)+\theta_2(y),\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}. \tag{1.4b}$$

2 乘子交替方向法及其变分不等式表示

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 $(x^{k+1},y^{k+1},\lambda^{k+1})$:

$$\left\{\begin{array}{l} x^{k+1}\in\arg\min\big\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}\beta\|Ax+By^k-b\|^2 \;\big|\; x\in\mathcal{X}\big\}, \qquad (2.1a)\\ y^{k+1}\in\arg\min\big\{\theta_2(y)-y^TB^T\lambda^k+\frac{1}{2}\beta\|Ax^{k+1}+By-b\|^2 \;\big|\; y\in\mathcal{Y}\big\}, \qquad (2.1b)\\ \lambda^{k+1}=\ \lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). \qquad (2.1c)\end{array}\right.$$

由于 k 步迭代从给定的 (y^k,λ^k) 开始, x^{k+1} 是计算得来的, 我们把 (y,λ) 称作**核心变量**. 称 x 为**中间变量**.

根据凸优化的最优性原理 (参见 **D** 中定理), ADMM 第 k 步迭代 (2.1) 的三个子问题的最优性条件分别是

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

跟任何一个向量的内积都非负的只能是零向量, 最后一行相当于 $(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)=0$.

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

$$\left\{\begin{array}{ll} x^{k+1}\in\mathcal{X},\ \ \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\big\{-A^T\lambda^{k+1}+\beta A^TB(y^k-y^{k+1})\big\}\ge 0,\ \ \forall x\in\mathcal{X}, & (2.2a)\\ y^{k+1}\in\mathcal{Y},\ \ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\big\{-B^T\lambda^{k+1}\big\}\ge 0,\ \ \forall y\in\mathcal{Y}, & (2.2b)\\ \lambda^{k+1}\in\Re^m,\qquad (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\ge 0,\ \ \forall\lambda\in\Re^m. & (2.2c)\end{array}\right.$$

在不等式 (2.2b) 的左端最后面加上和为零的两项 $(\beta B^TB(y^k-y^{k+1})+\beta B^TB(y^{k+1}-y^k))$, 上式改写成

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

对上式进行合理整合 (把同一类型的并在一起), 并利用变分不等式 (1.4) 中的记号, 得到 ADMM 的VI表示

$$\underline{\theta(u)-\theta(u^{k+1})}+(w-w^{k+1})^T\underline{F(w^{k+1})}$$
$$+\begin{pmatrix} x-x^{k+1}\\ y-y^{k+1}\end{pmatrix}^T\underbrace{\beta\begin{pmatrix} A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})}+\begin{pmatrix} y-y^{k+1}\\ \lambda-\lambda^{k+1}\end{pmatrix}^T\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix} y^{k+1}-y^k\\ \lambda^{k+1}-\lambda^k\end{pmatrix}\ge 0,\ \forall w\in\Omega. \quad (2.3)$$

3 基于变分不等式表示的交替方向法收敛性证明

将 ADMM 的变分不等式表示 (2.3) 中属于 Ω 的 w, 设成某个确定的解点 w^*, 便有

$$\theta(u^*)-\theta(u^{k+1})+(w^*-w^{k+1})^TF(w^{k+1})$$
$$+\begin{pmatrix} x^*-x^{k+1}\\ y^*-y^{k+1}\end{pmatrix}^T\beta\begin{pmatrix} A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})+\begin{pmatrix} y^*-y^{k+1}\\ \lambda^*-\lambda^{k+1}\end{pmatrix}^T\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix} y^{k+1}-y^k\\ \lambda^{k+1}-\lambda^k\end{pmatrix}\ge 0. \quad (3.1)$$

若用记号

$$v=\begin{pmatrix} y\\ \lambda\end{pmatrix} \qquad \text{和} \qquad H=\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}, \quad (3.2)$$

不等式 (3.1) 就可以改写成

$$\begin{aligned}(v^{k+1}-v^*)^TH(v^k-v^{k+1}) \quad \ge \quad & \begin{pmatrix} x^{k+1}-x^*\\ y^{k+1}-y^*\end{pmatrix}^T\beta\begin{pmatrix} A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})\\ & +\underbrace{\{\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^{k+1})\}}. \quad (3.3)\end{aligned}$$

下面我们证明 (3.3) 式右端非负. 首先, 由于

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

这就是说, (3.3) 式右端第二部分 (下波纹线那部分) 非负.

因此, 从 (3.3) 得到 $\quad (v^{k+1}-v^*)^TH(v^k-v^{k+1})\ \ge\ \begin{pmatrix} x^{k+1}-x^*\\ y^{k+1}-y^*\end{pmatrix}^T\beta\begin{pmatrix} A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1}). \quad (3.4)$

我们接着对 (3.4) 的右端化简, 注意到

$$\begin{aligned}&\begin{pmatrix} x^{k+1}-x^*\\ y^{k+1}-y^*\end{pmatrix}^T\beta\begin{pmatrix} A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})=(y^k-y^{k+1})^TB^T\beta(A,B)\begin{pmatrix} x^{k+1}-x^*\\ y^{k+1}-y^*\end{pmatrix}\\ &= (y^k-y^{k+1})^TB^T\beta\big(Ax^{k+1}+By^{k+1}-(Ax^*+By^*)\big) \qquad \text{利用}(Ax^*+By^*=b)\\ &= (y^k-y^{k+1})^TB^T\underbrace{\beta\big(Ax^{k+1}+By^{k+1}-b\big)}=(y^k-y^{k+1})^TB^T\underbrace{(\lambda^k-\lambda^{k+1})}. \quad (3.5)\end{aligned}$$

下面证明 (3.5) 式的右端 $(y^k-y^{k+1})^TB^T(\lambda^k-\lambda^{k+1})\ge 0$, 要用到一点小技巧.

利用 (2.2b) 有

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

$$\boxed{\left(\begin{array}{c}\text{将上面两式中任意的}\\ y\ \text{分别设成}\ y^k\ \text{和}\ y^{k+1}\end{array}\right) \quad \begin{array}{lcll}\theta_2(y^k)-\theta_2(y^{k+1}) & + & (y^k-y^{k+1})^T(-B^T\lambda^{k+1}) & \ge 0.\\ \theta_2(y^{k+1})-\theta_2(y^k) & + & (y^{k+1}-y^k)^T(-B^T\lambda^k) & \ge 0.\end{array}}$$

(将上面两式相加, 就有) $\quad (y^k-y^{k+1})B^T(\lambda^k-\lambda^{k+1})\ge 0. \quad$ ((3.5) 式右端非负)

上面证明了 $(y^k-y^{k+1})^TB^T(\lambda^k-\lambda^{k+1})\ge 0$, 由 (3.5) 得到 (3.4) 式右端非负. 从而得到我们需要的

$$(v^{k+1}-v^*)^TH(v^k-v^{k+1})\ge 0. \quad (3.6)$$

若有 $b^TH(a-b)\ge 0$ 就有 $\|b\|_H^2=\|a\|_H^2-2b^TH(a-b)-\|a-b\|_H^2\le\|a\|_H^2-\|a-b\|_H^2.$

置 $a=(v^k-v^*)$ 和 $b=(v^{k+1}-v^*)$, 根据 (3.6) 就得到证明 ADMM 收敛性的关键不等式

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-v^{k+1}\|_H^2, \quad \forall v^*\in\mathcal{V}^*. \qquad \text{(核心变量的收缩性质)}$$