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

四. 交替方向法(ADMM)及其线性化方法的收敛性证明

我们讨论求解两个可分离块凸优化问题

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

的交替方向法. 由于变动目标函数中的常数项对问题的解没有影响, 乘子交替方向法的第 k 次迭代可以表述成从 (y^k,λ^k) 开始:

$$\text{(ADMM)}\quad\begin{cases} x^{k+1}\in\operatorname{argmin}\big\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k-b\|^2\big|x\in\mathcal{X}\big\} & \text{(0.2a)}\\ y^{k+1}\in\operatorname{argmin}\big\{\theta_2(y)-y^TB^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By-b\|^2\big|y\in\mathcal{Y}\big\}, & \text{(0.2b)}\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). & \text{(0.2c)}\end{cases}$$

1 ADMM 的收敛性分析

子问题 (0.2a) 和 (0.2b) 分别等价于需要求解的子问题的具体形式是

$$x^{k+1}=\operatorname{argmin}\big\{\theta_1(x)+\tfrac{\beta}{2}\|(Ax+By^k-b)-\tfrac{1}{\beta}\lambda^k\|^2\big|x\in\mathcal{X}\big\} \tag{1.1a}$$

和

$$y^{k+1}\in\operatorname{argmin}\big\{\theta_2(y)+\tfrac{\beta}{2}\|(Ax^{k+1}+By-b)-\tfrac{1}{\beta}\lambda^k\|^2\big|y\in\mathcal{Y}\big\}. \tag{1.1b}$$

我们假设这类问题在方法中是容易求解的.

前面的分析已经告诉我们, 问题(0.1)的一阶最优性条件可以表示成一个单调的变分不等式

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

其中

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

和

$$\theta(u) = \theta_1(x) + \theta_2(y), \qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{1.2c}$$

我们注意到这样定义的 F 满足关系式

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) \equiv 0, \quad \forall w, \tilde{w}. \tag{1.3}$$

还是用 Ω^* 表示变分不等式 (1.2) 的解集.

分析 利用关于最优性的引理, 子问题(0.2a)和(0.2b)的解分别满足

$$\begin{aligned} x^{k+1} \in \mathcal{X},\ & \theta_1(x) - \theta_1(x^{k+1}) \\ & + (x - x^{k+1})^T \big\{ -A^T\lambda^k + \beta A^T\big(Ax^{k+1} + By^k - b\big)\big\} \ge 0,\ \forall\, x \in \mathcal{X} \end{aligned} \tag{1.4a}$$

和

$$\begin{aligned} y^{k+1} \in \mathcal{Y},\ & \theta_2(y) - \theta_2(y^{k+1}) \\ & + (y - y^{k+1})^T \big\{ -B^T\lambda^k + \beta B^T\big(Ax^{k+1} + By^{k+1} - b\big)\big\} \ge 0,\ \forall\, y \in \mathcal{Y}. \end{aligned} \tag{1.4b}$$

将 λ^{k+1} (参见(0.2c)) 代入(1.4) (消去其中的 λ^k), 我们得到

$$\begin{aligned} x^{k+1} \in \mathcal{X},\ & \theta_1(x) - \theta_1(x^{k+1}) \\ & + (x - x^{k+1})^T \big\{ -A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1})\big\} \ge 0,\ \forall\, x \in \mathcal{X}, \end{aligned} \tag{1.5a}$$

和

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \big\{ -B^T\lambda^{k+1}\big\} \ge 0,\ \forall\, y \in \mathcal{Y}. \tag{1.5b}$$

我们可将变分不等式 (1.5) 写成一个比较紧凑的形式: $u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$, 并且对所有的 $(x, y) \in \mathcal{X} \times \mathcal{Y}$, 都有

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} + \beta \begin{pmatrix} A^T \\ 0 \end{pmatrix} B(y^k - y^{k+1}) \right\} \ge 0. \tag{1.6}$$

再把上述变分不等式写成我们需要的格式: $u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$, 使得

$$\begin{aligned} \theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T & \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) \right. \\ & \left. + \begin{pmatrix} 0 & 0 \\ 0 & \beta B^T B \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\} \ge 0,\ \forall\, (x, y) \in \mathcal{X} \times \mathcal{Y}. \end{aligned} \tag{1.7}$$

再把等式 (0.2c) 写成等价的

$$(\lambda - \lambda^{k+1})\{(Ax^{k+1} + By^{k+1} - b) + \tfrac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \ge 0,\ \forall \lambda \in \Re^m. \tag{1.8}$$

将 (1.7) 和 (1.8) 加到一起, 我们有 $w^{k+1}\in\Omega$ 并且

$$\theta(u)-\theta(u^{k+1})+\left(\begin{array}{c} x-x^{k+1}\\ y-y^{k+1}\\ \lambda-\lambda^{k+1}\end{array}\right)^T\left\{\left(\begin{array}{c} -A^T\lambda^{k+1}\\ -B^T\lambda^{k+1}\\ Ax^{k+1}+By^{k+1}-b\end{array}\right)+\beta\left(\begin{array}{c} A^T\\ B^T\\ 0\end{array}\right)B(y^k-y^{k+1})\right.$$
$$\left.+\left(\begin{array}{cc} 0 & 0\\ \beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{array}\right)\left(\begin{array}{c} y^{k+1}-y^k\\ \lambda^{k+1}-\lambda^k\end{array}\right)\right\}\ge 0,\ \forall\, w\in\Omega. \quad (1.9)$$

为方便讨论, 我们定义如下的记号:

$$v=\left(\begin{array}{c} y\\ \lambda\end{array}\right),\qquad \mathcal{V}^*=\{(y^*,\lambda^*)\,|\,(x^*,y^*,\lambda^*)\in\Omega^*\}.$$

和

$$H=\left(\begin{array}{cc} \beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{array}\right). \quad (1.10)$$

利用 (1.2) 和上述记号, 从 (1.9) 我们直接得到如下的引理:

引理 1.1 求解结构型变分不等式 (1.2). 设 $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})\in\Omega$ 是由算法 (0.2) 生成的新的迭代点, 则有

$$\begin{aligned}\theta(u)-\theta(u^{k+1})+&(w-w^{k+1})^TF(w^{k+1})\\ &\ge\beta\left(\begin{array}{c} x-x^{k+1}\\ y-y^{k+1}\end{array}\right)^T\left(\begin{array}{c} A^T\\ B^T\end{array}\right)B(y^{k+1}-y^k)\\ &\quad+(v-v^{k+1})^TH(v^k-v^{k+1}),\ \ \forall\, w\in\Omega.\end{aligned} \quad (1.11)$$

下面的引理可以由引理 1.1 的结论直接推出.

引理 1.2 求解结构型变分不等式 (1.2). 设 $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})\in\Omega$ 是由算法 (0.2) 生成的新的迭代点, 则有

$$\begin{aligned}&(v^{k+1}-v^*)^TH(v^k-v^{k+1})\\ &\ge\ \beta\left(\begin{array}{c} x^{k+1}-x^*\\ y^{k+1}-y^*\end{array}\right)^T\left(\begin{array}{c} A^T\\ B^T\end{array}\right)B(y^k-y^{k+1}),\quad\forall\, w^*\in\Omega^*,\end{aligned} \quad (1.12)$$

证明. 在变分不等式(1.11)中令$w=w^*$, 并利用H的表达式, 我们得到

$$\begin{aligned}&(v^{k+1}-v^*)^TH(v^k-v^{k+1})\\ &\ge\ \theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^{k+1})\\ &\quad+\beta\left(\begin{array}{c} x^{k+1}-x^*\\ y^{k+1}-y^*\end{array}\right)^T\left(\begin{array}{c} A^T\\ B^T\end{array}\right)B(y^k-y^{k+1}),\quad\forall\, w^*\in\Omega^*.\end{aligned} \quad (1.13)$$

由于F 是单调的, 并且w^*是最优解, 则

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

利用上述不等式, 结论(1.12) 可从(1.13) 直接得到. $\square$

引理 1.3 求解结构型变分不等式 (1.2). 设 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ 是由算法 (0.2) 生成的新的迭代点, 则有

$$\beta \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) \geq 0, \; \forall\, w^* \in \Omega^*. \tag{1.14}$$

证明. 利用 $Ax^* + By^* = b$, 我们有

$$\begin{aligned} & \beta \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) \\ = \;& \beta\{(Ax^{k+1} + By^{k+1}) - (Ax^* + By^*)\}^T B(y^k - y^{k+1}) \\ = \;& \beta(Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}) \\ = \;& (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}). \end{aligned} \tag{1.15}$$

上面最后一个等式是用了 (0.2c). 因为 (1.5b) 对第 k 次及前一次迭代均成立, 因此有

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \geq 0, \; \forall\, y \in \mathcal{Y}, \tag{1.16}$$

和

$$\theta_2(y) - \theta_2(y^k) + (y - y^k)^T \{-B^T \lambda^k\} \geq 0, \; \forall\, y \in \mathcal{Y}, \tag{1.17}$$

我们在(1.16) 令$y = y^k$ 并且在(1.17)中令$y = y^{k+1}$, 然后将得到的变分不等式相加得到

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq 0. \tag{1.18}$$

将(1.18) 代入到(1.15)中, 我们得到结论(1.14). $\square$

将不等式 (1.14) 的结论代入不等式 (1.12), 我们有

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0, \quad \forall\, v^* \in \mathcal{V}^*. \tag{1.19}$$

根据上述不等式和预备知识中提到的简单引理, 我们直接得到下面的定理.

定理 1.1 求解结构型变分不等式 (1.2). 设 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ 是由算法 (0.2) 生成的新的迭代点, 则有

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{1.20}$$

定理 1.1 提供了 ADMM 算法收敛性证明的关键不等式.

如何选取参数β. ADMM算法的效率严重依赖于(0.2)中参数 β 的值. 我们接下来讨论在实际计算当中如何选择适当的参数 β.

注意到如果 $\beta A^T B(y^k - y^{k+1}) = 0$, 则利用 (1.6) 就得到

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} \geq 0, \quad \forall (x,y) \in \mathcal{X} \times \mathcal{Y}. \tag{1.21}$$

在这种情况下, 如果 $Ax^{k+1} + By^{k+1} - b = 0$ 也成立, 则我们得到

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

因此 $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ 是变分不等式(1.2)的一个最优解. 换句话说, 如果

$$\beta A^T B(y^k - y^{k+1}) \neq 0 \qquad \text{和/或} \qquad Ax^{k+1} + By^{k+1} - b \neq 0,$$

则$(x^{k+1}, y^{k+1}, \lambda^{k+1})$不是变分不等式(1.2) 的解. 我们称

$$\|\beta A^T B(y^k - y^{k+1})\| \qquad \text{和} \qquad \|Ax^{k+1} + By^{k+1} - b\|$$

分别为原始残差和对偶残差. 通过上述分析而知, 我们应该在迭代过程中动态的平衡原始残差和对偶残差. 如果

$$\mu\|\beta A^T B(y^k - y^{k+1})\| < \|Ax^{k+1} + By^{k+1} - b\| \quad \text{其中} \quad \mu > 1,$$

这意味着对偶残差过大, 我们应该增大增广Lagrange函数中参数β的值. 倒过来, 我们应该减小β的值.

下面我们介绍一种简单而有效的调整参数β的格式(参见, e.g., [7]):

$$\beta_{k+1} = \left\{ \begin{array}{ll} \beta_k * \tau, & \text{如果} \ \mu\|\beta A^T B(y^k - y^{k+1})\| < \|Ax^{k+1} + By^{k+1} - b\|; \\ \beta_k/\tau, & \text{如果} \ \|\beta A^T B(y^k - y^{k+1})\| > \mu\|Ax^{k+1} + By^{k+1} - b\|; \\ \beta_k, & \text{其它}. \end{array} \right.$$

其中参数$\mu > 1, \tau > 1$. 一种典型的参数选择为$\mu = 10$ 和$\tau = 2$. 这种参数更新方式的背后的想法是通过因子τ 来控制原始残差和对偶残差的范数, 使得两者达到动态均衡并收敛到0. 这种自调比方式已经被用于S. Boyd 的科研组[1] 以及他们的凸优化求解器当中[5].

2 线性化的交替方向法

常用的ADMM 方法 (0.2) 中, 每步迭代的主要工作是相当于对给定的p^k 和q^k, 分别求解 (1.1) 这样的子问题. 在一些实际应用中, 由于矩阵A 或B 的结构, 其中会有一个子问题难解. 我们总设其中比较难解的问题是 (0.2b).

由于变动目标函数中的常数项对问题的解没有影响, 我们考察 ADMM 中的子问题 (0.2b),

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

因此, 困难在于目标函数中既有 $\theta_2(y)$, 又有二次项 $\frac{\beta}{2}\|B(y-y^k)\|^2$. 所谓线性化的 ADMM, 就是用简单的二次函数

$$\frac{s}{2}\|y - y^k\|^2 \qquad \text{去代替} \qquad \frac{\beta}{2}\|B(y - y^k)\|^2.$$

换句话说, 就是将 (0.2b) 中的

$$\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) \quad \text{换成} \quad \mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) + \frac{s}{2}\|y - y^k\|^2 - \frac{\beta}{2}\|B(y - y^k)\|^2.$$

因此, 对照 ADMM 公式 (0.2), 线性化的 ADMM 的迭代公式可以表述成:

$$\text{(L-ADMM)} \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x, y^k, \lambda^k) \;\big|\; x \in \mathcal{X}\big\}, & (2.1\text{a}) \\ y^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_{D_B}^2 \;\big|\; y \in \mathcal{Y}\big\}, & (2.1\text{b}) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), & (2.1\text{c}) \end{array} \right.$$

其中

$$D_B = sI - \beta B^T B. \tag{2.2}$$

经过这种“线性化”以后, 子问题 (2.1b) 中的 y^{k+1} 就是

$$\begin{aligned} y^{k+1} &= \arg\min\big\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_{D_B}^2 \;\big|\; y \in \mathcal{Y}\big\} \\ &= \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k + \beta y^T B^T(Ax^{k+1} + By^k - b) \\ +\frac{s}{2}\|y - y^k\|^2 \end{array} \;\middle|\; y \in \mathcal{Y} \right\} \\ &= \arg\min\Big\{\theta_2(y) + \frac{s}{2}\big\|y - d^k\big\|^2 \;\big|\; y \in \mathcal{Y}\Big\}, \end{aligned} \tag{2.3}$$

其中

$$d^k = y^k - \frac{1}{s}B^T\big[\beta(Ax^{k+1} + By^k - b) - \lambda^k\big].$$

为了理论上保证收敛, 对于固定(不能随意变小)的 β, 人们要求 (2.2) 中的参数 [12, 13]

$$s \geq \beta\|B^T B\|. \tag{2.4}$$

大家同时知道, 过大的 $s > 0$, 会影响收敛速度.

我们先根据迭代 (2.1) 中子问题的最优性条件, 给出下面的基本引理.

引理 2.1 用线性化乘子交替方向法 (2.1) 求解结构型变分不等式 (1.2). 设 w^{k+1} 是由给定的 $v^k=(y^k,\lambda^k)$ 所产生的新的迭代点, 那么有

$$\begin{aligned} w^{k+1}\in\Omega,\quad \theta(u)-\theta(u^{k+1})+&(w-w^{k+1})^TF(w^{k+1})+\beta(x-x^{k+1})^TA^T(By^k-By^{k+1})\\ &\ge (y-y^{k+1})^TD_B(y^k-y^{k+1})\\ &\quad+\frac{1}{\beta}(\lambda-\lambda^{k+1})^T(\lambda^k-\lambda^{k+1}),\quad \forall w\in\Omega. \end{aligned}\tag{2.5}$$

证明 对 L-ADMM 算法的 x-子问题 (2.1a), 有

$$\begin{aligned} x^{k+1}\in\mathcal{X},\ \ &\theta_1(x)-\theta_1(x^{k+1})\\ &+(x-x^{k+1})^T\{-A^T\lambda^k+\beta A^T(Ax^{k+1}+By^k-b)\}\ge 0,\ \forall\, x\in\mathcal{X}. \end{aligned}$$

利用 (2.1) 中的乘子校正公式 $\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b)$, 上式可以写成

$$\begin{aligned} x^{k+1}\in\mathcal{X},\quad &\theta_1(x)-\theta_1(x^{k+1})\\ &+(x-x^{k+1})^T\{\underline{-A^T\lambda^{k+1}}+\beta A^TB(y^k-y^{k+1})\ge 0,\ \forall\, x\in\mathcal{X}. \end{aligned}\tag{2.6}$$

对 L-ADMM 算法的 y-子问题 (2.1b), 有

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

利用乘子校正公式 $\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b)$, 上式可以写成

$$y^{k+1}\in\mathcal{Y},\ \ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{\underline{-B^T\lambda^{k+1}}+D_B(y^{k+1}-y^k)\}\ge 0,\ \forall\, y\in\mathcal{Y}.\tag{2.7}$$

注意到乘子校正公式 $\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b)$ 本身可以写成

$$\lambda^{k+1}\in\Re^m,\quad (\lambda-\lambda^{k+1})^T\{\underline{(Ax^{k+1}+By^{k+1}-b)}+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\ge 0,\ \forall\,\lambda\in\Re^m.\tag{2.8}$$

将 (2.6), (2.7) 和 (2.8) 加在一起, 注意到其中下划线的部分组合在一起就是 $F(w^{k+1})$. 利用 (1.2) 中的记号, 我们得到

$$\begin{aligned} w^{k+1}\in\Omega,\quad \theta(u)-\theta(u^{k+1})+&(w-w^{k+1})^TF(w^{k+1})+\beta(x-x^{k+1})^TA^TB(y^k-y^{k+1})\\ &\ge (y-y^{k+1})^TD_B(y^k-y^{k+1})\\ &\quad+\frac{1}{\beta}(\lambda-\lambda^{k+1})^T(\lambda^k-\lambda^{k+1}),\quad \forall w\in\Omega. \end{aligned}\tag{2.9}$$

对 (2.9) 式左端中的 $(w-w^{k+1})^TF(w^{k+1})$ 利用 (1.3), 有

$$(w-w^{k+1})^TF(w^{k+1})=(w-w^{k+1})^TF(w).$$

这就证明了该引理的结论 (2.5). □

这个基本引理, 是我们证明方法收敛性质的重要基础. 下面我们证明算法产生的迭代序列 $\{w^k\}$ 具有收缩性质. 对任意的 $v^*\in\mathcal{V}^*$, 序列 $\{\|v^k-v^*\|_G+\|y^{k-1}-y^k\|_{D_B}^2\}$ 是单调下降的. 为此, 我们先证明几个引理.

引理 2.2 用线性化乘子交替方向法 (2.1) 求解结构型变分不等式 (1.2). 设 w^{k+1} 是由给定的 $v^k=(y^k,\lambda^k)$ 所产生的新的迭代点, 那么有

$$\begin{aligned} w^{k+1}\in\Omega,\quad \theta(u)-\theta(u^{k+1})+&(w-w^{k+1})^TF(w^{k+1})\\ &+\beta\begin{pmatrix}x-x^{k+1}\\ y-y^{k+1}\end{pmatrix}^T\begin{pmatrix}A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})\\ &\ge (v-v^{k+1})^TG(v^k-v^{k+1}),\quad \forall w\in\Omega, \end{aligned}\tag{2.10}$$

其中

$$G=\begin{pmatrix} D_B+\beta B^TB & 0\\ 0 & \frac{1}{\beta}I\end{pmatrix}.\tag{2.11}$$

证明. 对引理 2.1 中 (2.5) 的两端加上 $(y-y^{k+1})^T\beta B^TB(y^k-y^{k+1})$, 并利用矩阵 G 的结构就马上得到 (2.10), 引理得证. □

引理 2.3 用线性化乘子交替方向法 (2.1) 求解结构型变分不等式 (1.2). 设 w^{k+1} 是由给定的 $v^k=(y^k,\lambda^k)$ 所产生的新的迭代点, 那么有

$$(v^{k+1}-v^*)^TG(v^k-v^{k+1})\ge(\lambda^k-\lambda^{k+1})^TB(y^k-y^{k+1}),\ \ \forall w^*\in\Omega^*.\tag{2.12}$$

证明. 将 (2.10) 的 $w\in\Omega$ 设为任意的 $w^*\in\Omega^*$, 我们有

$$\begin{aligned} &(v^{k+1}-v^*)^TG(v^k-v^{k+1})\\ &\quad\ge\beta\begin{pmatrix}x^{k+1}-x^*\\ y^{k+1}-y^*\end{pmatrix}^T\begin{pmatrix}A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})\\ &\qquad+\big\{\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^{k+1})\big\}. \end{aligned}\tag{2.13}$$

利用 $Ax^*+By^*=b$ 和 $\lambda^k-\lambda^{k+1}=\beta(Ax^{k+1}+By^{k+1}-b)$ (see (2.1c)) 处理(2.13) 右端的第一部分, 就有

$$\begin{aligned} &\beta\begin{pmatrix}x^{k+1}-x^*\\ y^{k+1}-y^*\end{pmatrix}^T\begin{pmatrix}A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})\\ &\quad=\beta[(Ax^{k+1}-Ax^*)+(By^{k+1}-By^*)]^TB(y^k-y^{k+1})\\ &\quad=(\lambda^k-\lambda^{k+1})^TB(y^k-y^{k+1}). \end{aligned}$$

根据 (1.3) 和最优性条件, (2.13) 右端的第二部分

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

综合这两条, 引理得证. □

引理 2.4 用线性化乘子交替方向法(2.1)求解结构型变分不等式(1.2). 设 w^{k+1} 是由给定的 $v^k=(y^k,\lambda^k)$ 所产生的新的迭代点, 那么有

$$(\lambda^k-\lambda^{k+1})^TB(y^k-y^{k+1})\ge\frac{1}{2}\|y^k-y^{k+1}\|_{D_B}^2-\frac{1}{2}\|y^{k-1}-y^k\|_{D_B}^2. \tag{2.14}$$

证明. 首先, (2.7) 式表示 $y^{k+1}\in\mathcal{Y}$ 并且

$$\theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{-B^T\lambda^{k+1}+D_B(y^{k+1}-y^k)\}\ge 0,\ \forall\, y\in\mathcal{Y}. \tag{2.15}$$

以 $k-1$ 置换 (2.15) 中的 k, 就有 $y^k\in\mathcal{Y}$ 并且

$$\theta_2(y)-\theta_2(y^k)+(y-y^k)^T\{-B^T\lambda^k+D_B(y^k-y^{k-1})\}\ge 0,\ \ \forall\, y\in\mathcal{Y}. \tag{2.16}$$

将 (2.15) 和 (2.16) 中的任意的 $y\in\mathcal{Y}$ 分别设成 y^k 和 y^{k+1}, 然后将两式相加, 有

$$(y^k-y^{k+1})^T\big\{B^T(\lambda^k-\lambda^{k+1})+D_B[(y^{k+1}-y^k)-(y^k-y^{k-1})]\big\}\ge 0.$$

由上式得到

$$(y^k-y^{k+1})^TB^T(\lambda^k-\lambda^{k+1})\ge(y^k-y^{k+1})^TD_B[(y^k-y^{k+1})-(y^{k-1}-y^k)].$$

由于 D_B 矩阵正定, 对上式右端使用

$$\|a\|_{D_B}^2-a^T(D_B)b\ge\frac{1}{2}\|a\|_{D_B}^2-\frac{1}{2}\|b\|_{D_B}^2$$

这样的不等式就得到(2.14). □

根据引理 2.3 和引理 2.4 我们可以直接证明下面的定理.

定理 2.1 用线性化乘子交替方向法(2.1)求解结构型变分不等式(1.2). 设 w^{k+1} 是由给定的 $v^k=(y^k,\lambda^k)$ 所产生的新的迭代点, 那么有

$$\begin{array}{ll}\big(\|v^{k+1}-v^*\|_G^2+\|y^k-y^{k+1}\|_{D_B}^2\big) & \\ \quad\le\ \ \big(\|v^k-v^*\|_G^2+\|y^{k-1}-y^k\|_{D_B}^2\big)-\|v^k-v^{k+1}\|_G^2,\ \ \forall w^*\in\Omega^*, & \end{array} \tag{2.17}$$

其中 G 由 (2.11) 给出.

证明. 由引理 2.3 和引理 2.4, 有

$$(v^{k+1}-v^*)^TG(v^k-v^{k+1})\ge\frac{1}{2}\|y^k-y^{k+1}\|_{D_B}^2-\frac{1}{2}\|y^{k-1}-y^k\|_{D_B}^2,\ \ \forall w^*\in\Omega^*.$$

利用上式, 对任意的 $w^*\in\Omega^*$, 都有

$$\begin{array}{rcl}\|v^k-v^*\|_G^2 &=& \|(v^{k+1}-v^*)+(v^k-v^{k+1})\|_G^2\\ &\ge& \|v^{k+1}-v^*\|_G^2+\|v^k-v^{k+1}\|_G^2+2(v^{k+1}-v^*)^TG(v^k-v^{k+1})\\ &\ge& \|v^{k+1}-v^*\|_G^2+\|v^k-v^{k+1}\|_G^2+\|y^k-y^{k+1}\|_{D_B}^2-\|y^{k-1}-y^k\|_{D_B}^2.\end{array}$$

这就证明了定理 2.1 的结论. □

3 ADMM 算法的收敛性证明

在线性化 ADMM 的收敛性质定理 2.1 中取 $D_B = 0$, 这时其中的矩阵 G 就等同于 (1.10) 中的矩阵 H, 得到了 ADMM 收敛性质的定理 1.1. 因此, 我们只要对线性化的 ADMM 证明算法的收敛定理. 其证明思路和 PPA 算法收敛性定理相同.

定理 3.1 设 $\{w^k$ 是用线性化乘子交替方向法 (2.1) 求解结构型变分不等式 (1.2) 产生的序列. 那么相应的序列 $\{v^k\}$ 收敛于属于 $\mathcal{V}^*$ 的一点 v^∞.

Proof. 首先, 根据定理 2.1 的结论 (2.17), 有

$$\begin{aligned}\|v^k - v^{k+1}\|_G^2 \quad \le \quad & \big(\|v^k - v^*\|_G^2 + \|y^{k-1} - y^k\|_{D_B}^2\big) \\ & -\big(\|v^{k+1} - v^*\|_G^2 + \|y^k - y^{k+1}\|_{D_B}^2\big). \end{aligned} \tag{3.1}$$

将上式对 $k = 1, 2, \ldots,$ 累加, 得到

$$\sum_{k=1}^{\infty} \|v^k - v^{k+1}\|_G^2 \le \|v^1 - v^*\|_G^2 + \|y^0 - y^1\|_{D_B}^2. \tag{3.2}$$

由于矩阵 G 是正定的, 由上面的不等式得到

$$\lim_{k\to\infty} \|v^k - v^{k+1}\| = 0. \tag{3.3}$$

对任意确定的 $v^* \in \mathcal{V}^*$, 不等式 (2.17) 告诉我们, 对任意的 $k \ge 1$, 都有

$$\begin{aligned}\|v^{k+1} - v^*\|_G^2 \quad & \le \quad \|v^k - v^*\|_G^2 + |y^{k-1} - y^k\|_{D_B}^2 \\ & \le \quad \|v^1 - v^*\|_G^2 + \|y^0 - y^1\|_{D_B}^2, \end{aligned} \tag{3.4}$$

因此序列 $\{v^k\}$ 是有界的. 设 $\{v^{k_j}\}$ 是 $\{v^k\}$ 的收敛于 v^∞ 的子序列, x^∞ 是与$(y^\infty, \lambda^\infty)$ 相应的中间变量. 那么, 由 (2.5) 和 (3.3) 得到

$$w^\infty \in \Omega, \quad \theta(u) - \theta(u^\infty) + (w - w^\infty)^T F(w^\infty) \ge 0, \quad \forall w \in \Omega,$$

这说明 w^∞ 是 (1.2) 的解并且其核心部分 $v^\infty \in \mathcal{V}^*$. 因为 $v^\infty \in \mathcal{V}^*$, 从 (3.4) 中有

$$\|v^{k+1} - v^\infty\|_G^2 \quad \le \quad \|v^k - v^\infty\|_G^2 + \|y^{k-1} - y^k\|_D^2. \tag{3.5}$$

结合 (3.3), 上式说明 $\{v^k\}$ 不可能有多于一个聚点. 所以序列 $\{v^k\}$ 收敛于 v^∞. $\square$

参考文献

[1] 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.

[2] 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.

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

[4] Gu G Y, He B S and Yuan X M. Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach. Comput. Optim. Appl., 2014, 59: 135-161.

[5] D. Hallac, Ch. Wong, S. Diamond, A. Sharang, R. Sosič, S. Boyd and J. Leskovec, SnapVX: A Network-Based Convex Optimization Solver, Journal of Machine Learning Research 18 (2017) 1-5.

[6] 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.

[7] B.S. He, H. Yang, and S.L. Wang, Alternating directions method with self-adaptive penalty parameters for monotone variational inequalities, *JOTA* **23**(2000), 349–368.

[8] 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.

[9] B.S. He and X. M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, SIAM Journal on Imaging Science, **5**, 119-149, 2012.

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

[11] 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.

[12] Yang J F and Yuan X M. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization *Mathematics of Computation*, 2013, 82: 301-329

[13] Zhang X. Q, Burger M and Osher S. A unified primal-dual algorithm framework based on Bregman iteration, J. Sci. Comput., 2010, 46: 20 – 46.

[14] 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.

[15] 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》.

[16] 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.