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

七.　均困的增广拉格朗日乘子法

这一章继续介绍如何用第统一框架来构造求解凸优化问题

$$\min\{\theta(x) \mid Ax = b \text{ (or } \geq b),\, x \in \mathcal{X}\}. \tag{0.1}$$

问题

$$L(x,\lambda) = \theta(x) - \lambda^T(Ax - b). \tag{0.2}$$

相应的变分不等式是

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{0.3a}$$

其中

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix}, \quad \Omega = \mathcal{X} \times \Lambda. \tag{0.3b}$$

注意到, 对等式约束 $Ax = b$ 和不等式约束 $Ax \geq b$, Λ 分别为 $\Re^m$ 和 $\Re^m_+$. 这一章的方法中, 我们需要定义一个 $m \times m$ 矩阵

$$H_0 = \frac{1}{r}AA^T + \delta I_m. \tag{0.4}$$

对任意的 $\delta > 0$, 矩阵 H 是正定的.

1　均困平衡的增广拉格朗日乘子法

我们从增广拉格朗日乘子法 [8, 9, 10] 谈起, 许多有效算法 (例如交替方向法 (ADMM)) 起源于 ALM. 用 ALM,求解等式约束问题 (0.1), 迭代公式是

$$\left\{\begin{array}{ll} x^{k+1} \in \arg\min\big\{L(x,\lambda^k) + \frac{r}{2}\|Ax - b\|^2 \;\big|\; x \in \mathcal{X}\big\}, & \text{(1.1a)} \\[2mm] \lambda^{k+1} = \mathrm{argmax}\{L(x^{k+1},\lambda) - \frac{1}{2r}\|\lambda - \lambda^k\|^2 \mid \lambda \in \Re^m\}. & \text{(1.1b)} \end{array}\right.$$

把 ALM 中 x-子问题 (1.1a) 中目标函数里的常数项做调整, ALM 可以通过

$$x^{k+1} \in \mathrm{argmin}\big\{\theta(x) + rx^TA^T(Ax^k - b) + \frac{r}{2}\|A(x - x^k)\|^2 \,\big|\, x \in \mathcal{X}\big\},$$

实现. ALM 的 x-子问题 (1.1a) 中既有非线性凸函数 $\theta(x)$, 又有形如 $\frac{r}{2}\|A(x - x^k)\|^2$ 的二次项. 这两项加在一起, 会给求解带来一定的困难. ALM 中通过 (1.1b) 更新对偶变量 λ^{k+1}, 这非常简单. 前面介绍的求解等式和不等式约束的凸优化问题 ALM 类算法, 它们的 x-子问题的目标函数中同样既包含非线性函数 $\theta(x)$, 还包含二次项 $\|A(x - x^k)\|^2$.

我们试图利用统一框架构造新的算法, 通过

$$\begin{cases} x^{k+1} \in \arg\min\big\{L(x, \lambda^k) + \frac{r}{2}\|x - x^k\|^2 \,\big|\, x \in \mathcal{X}\big\}, & (1.2\text{a}) \\ \lambda^{k+1} = \mathrm{argmax}\{L(x^{k+1}, \lambda) - \frac{1}{2}\|\lambda - \lambda^k\|_{H_0}^2 \mid \lambda \in \Re^m\}. & (1.2\text{b}) \end{cases}$$

得出预测点 $w^{k+1} = (x^{k+1}, \lambda^{k+1})$, 其中矩阵 H_0 由 (0.4) 给出. 比起 (1.1), (1.2) 中的 x-子问题简化了, 但是增加了求得 λ^{k+1} 的难度. (1.2b) 中的 λ^{k+1} 是一个并不难解的线性方程组

$$\Big(\frac{1}{r}AA^T + \delta I_m\Big)(\lambda - \lambda^k) + (Ax^{k+1} - b) = 0$$

的解. 虽然由 (1.2) 的最优性条件的变分不等式形式和预备知识一章中的 PPA 格式不匹配, 后面我们就会看到, 在统一框架的指导下, 一个单位上三角矩阵的校正就能保证迭代序列收敛.

用 Chambolle 和 Pock 的可以解释成 PPA 的原始-对偶混合梯度法 [3, 4, 7], 我们把它称为求解等式约束问题 (0.1), 迭代公式是

$$(\text{CP-PPA}) \quad \begin{cases} x^{k+1} = \mathrm{argmin}\big\{L(x, \lambda^k) + \frac{r}{2}\|x - x^k\|^2 \,\big|\, x \in \mathcal{X}\big\}, & (1.3\text{a}) \\ \lambda^{k+1} = \mathrm{argmax}\{L([2x^{k+1} - x^k], \lambda) - \frac{s}{2}\|\lambda - \lambda^k\|^2\}. & (1.3\text{b}) \end{cases}$$

这里的 $r, s > 0$ 是迫使新的 x^{k+1} 和 λ^{k+1} 分别离原来的 x^k 和 λ^k 不要太远. CP-PPA 方法中 x-子问题的二次项简单, 求解也相对容易. 然而, 为了保证收敛, 要求参数

$$rs > \|A^TA\|. \tag{1.4}$$

从 (1.3) 可以看出, 过大的 r 和 s, 分别迫使 x^{k+1} 离 x^k, λ^{k+1} 离 λ^k 很近, 相当于迭代中步长受限, 影响整体收敛速度. CP-PPA 方法在图像处理中的确取得了比较好的效果, 是得益于其中用到的全变差矩阵的 $\|A^TA\| \le 8$。

把 (1.3b) 中的二次项

$$\frac{s}{2}\|\lambda - \lambda^k\|^2 \qquad \text{换成} \qquad \frac{1}{2}\|\lambda - \lambda^k\|_{H_0}^2.$$

这里的参数 $r>0$ 子问题 (1.3a) 中已经确定了的. 任意的 $\delta>0$ 只是为了保证某种正定性. 这样, Balanced ALM 的迭代公式是

$$\left\{\begin{array}{ll} x^{k+1}=\arg\min\big\{L(x,\lambda^k)+\frac{r}{2}\|x-x^k\|^2 \;\big|\; x\in\mathcal{X}\big\}, & (1.5a)\\ \lambda^{k+1}=\text{argmax}\big\{L([2x^{k+1}-x^k],\lambda)-\frac{1}{2}\|\lambda-\lambda^k\|^2_{(\frac{1}{r}AA^T+\delta I_m)}\big\}. & (1.5b)\end{array}\right.$$

就像 ALM 方法一样, Balanced ALM 中只有一个参数 r 需要根据具体问题而选择。Balanced ALM 方法迭代公式的等价形式是

$$\left\{\begin{array}{ll} x^{k+1}=\text{argmin}\big\{\theta(x)+\frac{r}{2}\big\|x-\big[x^k+\frac{1}{r}A^T\lambda^k\big]\big\|^2 \,\big|\, x\in\mathcal{X}\big\}. & (1.6a)\\ \lambda^{k+1}=\arg\min\big\{\frac{1}{2}\|\lambda-\lambda^k\|^2_{(\frac{1}{r}AA^T+\delta I_m)}+\lambda^T\big(A[2x^{k+1}-x^k]-b\big)\big\}. & (1.6b)\end{array}\right.$$

这里的 x-子问题 (1.6a) 和 CP-PPA 的 x-子问题 (1.3a) 形式完全一样, (1.6b) 中的 λ^{k+1} 是线性方程组

$$\Big(\frac{1}{r}AA^T+\delta I_m\Big)(\lambda-\lambda^k)+\big(A[2x^{k+1}-x^k]-b\big)=0 \tag{1.7}$$

的解. 注意到上述线性方程组的系数矩阵是对称正定的, 在整个迭代过程中我们只要做一次正定矩阵

$$\Big(\frac{1}{r}AA^T+\delta I_m\Big). \tag{1.8}$$

的 Cholesky 分解. 矩阵计算 [5, 11] 里有非常成熟的方法求解这类线性方程组。

对应于不等式约束问题

$$\min\{\theta(x)\,|\,Ax\ge b,\, x\in\mathcal{X}\},$$

Balanced PPA 只要将 (1.6) 中的 λ-子问题 (1.6b) 改成

$$\lambda^{k+1}=\arg\min\left\{\begin{array}{c}\frac{1}{2}(\lambda-\lambda^k)^T\Big(\frac{1}{r}AA^T+\delta I_m\Big)(\lambda-\lambda^k)\\ +\lambda^T\big(A[2x^{k+1}-x^k]-b\big)\end{array}\;\middle|\;\lambda\ge 0\right\}.$$

这是一个带非负约束的凸二次规划, 尽管迭代中不能一劳永逸, 但求解也是比较容易的.

2 统一框架下非平凡校正的算法

这一节介绍需要用统一框架中非平凡校正的均困 ALM 类算法. 通常是把一个不确定是否收敛的方法的输出当做预测点, 然后再在统一框架的指引下选择简单的校正.

2.1 Primal-Dual 顺序产生预测点

从给定的 $w^k=(x^k,\lambda^k)$ 出发, 把 (1.2) 的输出当成预测点, 公式就是

$$\left\{\begin{array}{ll} \tilde{x}^k=\arg\min\{\theta(x)-x^TA^T\lambda^k+\frac{r}{2}\|x-x^k\|^2|x\in\mathcal{X}\}, & (2.1a)\\ \tilde{\lambda}^k=\arg\min\big\{\frac{1}{2}(\lambda-\lambda^k)^T(\frac{1}{r}AA^T+\delta I_m)(\lambda-\lambda^k)+\lambda^T(A\tilde{x}^k-b)\big|\lambda\in\Lambda\big\} & (2.1b)\end{array}\right.$$

根据最优性条件就有

$$\begin{cases} \tilde{x}^k \in \mathcal{X}, & \theta(x)-\theta(\tilde{x}^k) + (x-\tilde{x}^k)^T\{-A^T\lambda^k + r(\tilde{x}^k - x^k)\} \ge 0, \quad \forall x \in \mathcal{X}, \\ \tilde{\lambda}^k \in \Lambda, & (\lambda-\tilde{\lambda}^k)^T\{(\frac{1}{r}AA^T+\delta I_m)(\tilde{\lambda}^k-\lambda^k) + (A\tilde{x}^k - b)\} \ge 0, \quad \forall \lambda \in \Lambda. \end{cases}$$

我们把上式改写成

$$\begin{cases} \theta(x)-\theta(\tilde{x}^k)+ & (x-\tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k} + r(\tilde{x}^k - x^k) + A^T(\tilde{\lambda}^k-\lambda^k)\} \ge 0, \quad \forall x \in \mathcal{X}, \\ & (\lambda-\tilde{\lambda}^k)^T\{\underline{(A\tilde{x}^k - b)} + (\frac{1}{r}AA^T+\delta I_m)(\tilde{\lambda}^k-\lambda^k)\} \ge 0, \quad \forall \lambda \in \Lambda. \end{cases}$$

其中带下划线的部分组合在一起就是 $F(\tilde{w}^k)$. 利用变分不等式 (0.3) 的表达式, 上式中下划线部分放在一起就是 $F(\tilde{w}^k)$. 因此, 求得的预测点 $\tilde{w}^k$ 满足

$$\tilde{w}^k \in \Omega, \;\; \theta(x)-\theta(\tilde{x}^k) + (w-\tilde{w}^k)^T\big\{F(\tilde{w}^k) + Q(\tilde{w}^k - w^k)\big\} \ge 0, \;\; \forall w \in \Omega, \tag{2.2a}$$

其中

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & \frac{1}{r}AA^T+\delta I_m \end{pmatrix}, \quad \delta > 0. \tag{2.2b}$$

上面的矩阵 Q^T+Q 是正定的.

对由 (2.1) 生成的预测点, 我们采用步长 $\alpha = 1$ 的非平凡校正

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k), \tag{2.3a}$$

生成新的迭代点, 其中

$$M = \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} \tag{2.3b}$$

是单位上三角矩阵. 下面我们用统一框架中的收敛条件去验证收敛性, 注意到

$$\begin{aligned} H &= QM^{-1} = \begin{pmatrix} rI_n & A^T \\ 0 & \frac{1}{r}AA^T+\delta I_m \end{pmatrix}\begin{pmatrix} I_n & -\frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix}. \end{aligned}$$

和

$$\begin{aligned} G &= Q^T+Q-M^THM = Q^T+Q-Q^TM \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2(\frac{1}{r}AA^T+\delta I_m) \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ A & \frac{1}{r}AA^T+\delta I_m \end{pmatrix}\begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n & 0 \\ 0 & \delta I_m \end{pmatrix}. \end{aligned}$$

矩阵 H 和 G 都正定. 分别用 (2.1) 做预测和 (2.3) 做校正, 产生新的迭代点 w^{k+1} 满足

$$\|w^{k+1}-w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \|w^k - \tilde{w}^k\|_G^2, \quad \forall\, w^* \in \Omega^*.$$

我们就得到一个求解(与优化问题 (0.1) 对应的)变分不等式 (0.3) 的收敛算法.

2.2 Dual-Primal 顺序产生预测点

考虑和 (2.1) 不同的 Dual-Primal 顺序预测, k-次迭代从就通过

$$\begin{cases} \tilde{\lambda}^k = \text{argmin}\{\frac{1}{2}(\lambda-\lambda^k)^T(\frac{1}{r}AA^T+\delta I_m)(\lambda-\lambda^k)+\lambda^T(Ax^k-b)\,\big|\,\lambda\in\Lambda\}, & (2.4a)\\ \tilde{x}^k = \arg\min\{\theta(x)-x^TA^T\tilde{\lambda}^k+\frac{r}{2}\|x-x^k\|^2\,|\,x\in\mathcal{X}\} & (2.4b)\end{cases}$$

得到预测点 $\tilde{w}^k$. 根据最优性条件就有

$$\begin{cases} \tilde{\lambda}^k\in\Lambda, & (\lambda-\tilde{\lambda}^k)^T\{(Ax^k-b)+(\frac{1}{r}AA^T+\delta I_m)(\tilde{\lambda}^k-\lambda^k)\}\ge 0, \quad \forall\lambda\in\Lambda,\\ \tilde{x}^k\in\mathcal{X}, & \theta(x)-\theta(\tilde{x}^k)+(x-\tilde{x}^k)^T\{-A^T\tilde{\lambda}^k+r(\tilde{x}^k-x^k)\}\ge 0, \quad \forall x\in\mathcal{X}.\end{cases}$$

我们把上式改写成

$$\begin{cases} \theta(x)-\theta(\tilde{x}^k)+(x-\tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}+r(\tilde{x}^k-x^k)\}\ge 0, & \forall x\in\mathcal{X},\\ (\lambda-\tilde{\lambda}^k)^T\{\underline{(A\tilde{x}^k-b)}-A(\tilde{x}^k-x^k)+(\frac{1}{r}AA^T+\delta I_m)(\tilde{\lambda}^k-\lambda^k)\}\ge 0, & \forall\lambda\in\Lambda.\end{cases}$$

其中带下划线的部分组合在一起就是 $F(\tilde{w}^k)$. 利用变分不等式 (0.3) 的表达式, 上式中下划线部分放在一起就是 $F(\tilde{w}^k)$. 因此, 求得的预测点 $\tilde{w}^k$ 满足

$$\tilde{w}^k\in\Omega,\ \ \theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^T\big\{F(\tilde{w}^k)+Q(\tilde{w}^k-w^k)\big\}\ge 0,\ \ \forall w\in\Omega, \tag{2.5a}$$

其中

$$Q=\begin{pmatrix} rI_n & 0\\ -A & (\frac{1}{r}AA^T+\delta I_m)\end{pmatrix}, \quad \delta>0. \tag{2.5b}$$

当 A 列满秩时上面的矩时 Q^T+Q 是正定的.

对由 (2.4) 生成的预测点, 我们采用步长 $\alpha=1$ 的非平凡校正

$$w^{k+1}=w^k-M(w^k-\tilde{w}^k), \tag{2.6a}$$

生成新的迭代点, 其中

$$M=\begin{pmatrix} I_n & \frac{1}{r}A^T\\ 0 & I_m\end{pmatrix}. \tag{2.6b}$$

剩下就是用统一框架中的收敛条件去验证收敛性, 注意到

$$\begin{aligned} H \;&=\; QM^{-1}=\begin{pmatrix} rI_n & 0\\ -A & \frac{1}{r}AA^T+\delta I_m\end{pmatrix}\begin{pmatrix} I_n & -\frac{1}{r}A^T\\ 0 & I_m\end{pmatrix}\\ &=\; \begin{pmatrix} rI_n & -A^T\\ -A & \frac{2}{r}AA^T+\delta I_m\end{pmatrix}.\end{aligned}$$

和

$$\begin{aligned} G &= Q^T + Q - M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2rI_n & -A^T \\ -A & \frac{2}{r}AA^T + 2\delta I_m \end{pmatrix} - \begin{pmatrix} rI_n & -A^T \\ 0 & \frac{1}{r}AA^T + \delta I_m \end{pmatrix} \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n & -A^T \\ -A & \frac{1}{r}AA^T + \delta I_m \end{pmatrix}. \end{aligned}$$

矩阵 H 和 G 都正定. 分别用 (2.4) 做预测和 (2.6) 做校正, 产生新的迭代点 w^{k+1} 满足

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \|w^k - \tilde{w}^k\|_G^2, \quad \forall\, w^* \in \Omega^*.$$

我们就得到一个求解(与优化问题 (0.1) 对应的)变分不等式 (0.3) 的收敛算法.

3 统一框架下平凡松弛校正的 PPA 算法

跟前一节不同, 这一节根据统一框架中 PPA 的要求, 先设计个符合条件的正定矩阵, 再去构造相应的算法实施.

3.1 Primal-Dual 顺序产生预测点

求解(与不等式约束凸优化问题(0.1)对应的)变分不等式(0.3), 我们根据第五章的统一框架设计一个预测-校正方法, 迭代从给定的 $w^k = (x^k, \lambda^k)$ 开始, 求得的预测点 $\tilde{w}^k$ 满足

$$\tilde{w}^k \in \Omega, \;\; \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \big\{ F(\tilde{w}^k) + H(\tilde{w}^k - w^k) \big\} \ge 0, \;\; \forall w \in \Omega, \tag{3.1a}$$

其中

$$H = \begin{pmatrix} rI_n & A^T \\ A & \frac{1}{r}AA^T + \delta I_m \end{pmatrix}, \quad r, \delta > 0. \tag{3.1b}$$

对于任何的 $r, \delta > 0$, 这个预测公式中的矩阵 H 是正定的. 这样就能用平凡松弛的校正 (**??**) 直接产生新迭代点 w^{k+1}. 问题归结为求满足 (3.1) 的预测点 $\tilde{w}^k$. 将预测 (3.1) 按 F 和 H 的结构分拆开来, 可以写成 $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k) \in \Omega$,

$$\left\{ \begin{array}{ll} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \{ \underline{-A^T \tilde{\lambda}^k} + r(\tilde{x}^k - x^k) + A^T(\tilde{\lambda}^k - \lambda^k) \} \ge 0, & \forall x \in \mathcal{X}, \\ (\lambda - \tilde{\lambda}^k)^T \{ \underline{(A\tilde{x}^k - b)} & \\ \qquad + A(\tilde{x}^k - x^k) + (\frac{1}{r}AA^T + \delta I_m)(\tilde{\lambda}^k - \lambda^k) \} \ge 0, & \forall \lambda \in \Lambda, \end{array} \right.$$

其中带下划线的部分组合在一起就是 $F(\tilde{w}^k)$. 归并简化后的形式是 $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k) \in \Omega$,

$$\left\{ \begin{array}{ll} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \{ -A^T \lambda^k + r(\tilde{x}^k - x^k) \} \ge 0, \;\; \forall x \in \mathcal{X}, & \text{(3.2a)} \\ (\lambda - \tilde{\lambda}^k)^T \{ (A(2\tilde{x}^k - x^k) - b) + (\dfrac{1}{r}AA^T + \delta I_m)(\tilde{\lambda}^k - \lambda^k) \} \ge 0, \;\; \forall \lambda \in \Lambda. & \text{(3.2b)} \end{array} \right.$$

注意到 (3.2a) 中只有需要预测的 $\tilde{x}^k$, 而 (3.2b) 中有了 $\tilde{x}^k$ 才能开展. 所以我们只能先实现 (3.2a) 再实现 (3.2b). 根据最优性引理, (3.2a) 中的 $\tilde{x}^k$ 可以通过求解子问题

$$\min\{\theta(x) - x^T A^T \lambda^k + \frac{r}{2}\|x - x^k\|^2 | x \in \mathcal{X}\}$$

得到. 而有了 $\tilde{x}^k$, (3.2b) 中的 $\tilde{\lambda}^k$ 则是

$$\min\Big\{\frac{1}{2}(\lambda - \lambda^k)^T(\frac{1}{r}AA^T + \delta I_m)(\lambda - \lambda^k) + \lambda^T[A(2\tilde{x}^k - x^k) - b] \,\Big|\, \lambda \in \Lambda\Big\}$$

的解. 所以, 满足 (3.1) 的预测点由

$$\begin{cases} \tilde{x}^k = \arg\min\{\theta(x) - x^T A^T \lambda^k + \frac{r}{2}\|x - x^k\|^2 | x \in \mathcal{X}\}, & (3.3a) \\ \tilde{\lambda}^k = \arg\min\left\{ \begin{array}{c} \frac{1}{2}(\lambda - \lambda^k)^T(\frac{1}{r}AA^T + \delta I_m)(\lambda - \lambda^k) \\ +\lambda^T[A(2\tilde{x}^k - x^k) - b] \end{array} \Big| \lambda \in \Lambda \right\} & (3.3b) \end{cases}$$

提供, 按照先 $\tilde{x}^k$, 后 $\tilde{\lambda}^k$ 的(primal-dual)顺序实现. 再用

$$w^{k+1} = w^k - \alpha(w^k - \tilde{w}^k), \quad \alpha \in (0, 2)$$

这样平凡的松弛校正, 迭代序列 $\{w^k\}$ 满足

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \alpha(2 - \alpha)\|w^k - \tilde{w}^k\|_H^2, \quad \forall\, w^* \in \Omega^*.$$

3.2 Dual-Primal 顺序产生预测点

如果把预测满足的变分不等式 (3.1) 换成

$$\tilde{w}^k \in \Omega, \;\; \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T\big\{F(\tilde{w}^k) + H(\tilde{w}^k - w^k)\big\} \ge 0, \;\; \forall w \in \Omega, \qquad (3.4a)$$

其中

$$H = \begin{pmatrix} rI_n & -A^T \\ -A & \frac{1}{r}AA^T + \delta I_m \end{pmatrix}, \quad \delta > 0. \qquad (3.4b)$$

上面的矩阵 H 是正定的. 对应于统一框架中的预测, $v = w$. 再用松弛校正 (**??**) 直接产生新迭代点 w^{k+1} 的算法是能保证收敛的. 问题归结为求满足 (3.4) 的预测点 $\tilde{w}^k$. 将预测 (3.4) 按 F 和 H 的结构分拆开来, 可以写成 $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k) \in \Omega$,

$$\begin{cases} \theta(x) - \theta(\tilde{x}^k) + \;(x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k} + s(\tilde{x}^k - x^k) - A^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0, & \forall x \in \mathcal{X}, \\ \qquad (\lambda - \tilde{\lambda}^k)^T\{\underline{(A\tilde{x}^k - b)} \\ \qquad\qquad -A(\tilde{x}^k - x^k) + (\frac{1}{r}AA^T + \delta I_m)(\tilde{\lambda}^k - \lambda^k)\} \ge 0, & \forall \lambda \in \Lambda. \end{cases}$$

其中带下划线的部分组合在一起就是 $F(\tilde{w}^k)$. 归并简化后的形式是 $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k) \in \Omega$,

$$\begin{cases} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\{r(\tilde{x}^k - x^k) - A^T(2\tilde{\lambda}^k - \lambda^k)\} \ge 0, \;\; \forall x \in \mathcal{X}, & (3.5a) \\ (\lambda - \tilde{\lambda}^k)^T\{(Ax^k - b) + (\frac{1}{r}AA^T + \delta I_m)(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \;\; \forall \lambda \in \Lambda. & (3.5b) \end{cases}$$

注意到 (3.5b) 中预测 $\tilde{\lambda}^k$ 只需要给定的 w^k, 而 (3.5a) 中预测 $\tilde{x}^k$ 需要 $\tilde{\lambda}^k$ 在手. 所以我们只能先实现 (3.5b) 再实现 (3.5a). 根据最优性条件引理, (3.5b) 中的 $\tilde{\lambda}^k$ 是

$$\min\big\{\frac{1}{2}(\lambda-\lambda^k)^T(\frac{1}{r}AA^T+\delta I_m)(\lambda-\lambda^k)+\lambda^T(Ax^k-b)\,\big|\,\lambda\in\Lambda\big\}$$

的解. 而 (3.5a) 中的 $\tilde{x}^k$ 可以通过求解子问题

$$\min\big\{\theta(x)-x^TA^T(2\tilde{\lambda}^k-\lambda^k)+\frac{r}{2}\|x-x^k\|^2\,\big|\,x\in\mathcal{X}\big\}$$

得到. 所以, 满足 (3.4) 的预测点由

$$\left\{\begin{array}{ll}\tilde{\lambda}^k=\text{argmin}\big\{\frac{1}{2}(\lambda-\lambda^k)^T(\frac{1}{r}AA^T+\delta I_m)(\lambda-\lambda^k)+\lambda^T(Ax^k-b)\,\big|\,\lambda\in\Lambda\big\}, & (3.6\text{a})\\ \tilde{x}^k=\arg\min\{\theta(x)-x^TA^T(2\tilde{\lambda}^k-\lambda^k)+\frac{r}{2}\|x-x^k\|^2\,|\,x\in\mathcal{X}\} & (3.6\text{b})\end{array}\right.$$

提供, 按照先 $\tilde{\lambda}^k$, 后 $\tilde{x}^k$ 的(dual-primal)顺序实现. 同样, 再用

$$w^{k+1}=w^k-\alpha(w^k-\tilde{w}^k),\quad \alpha\in(0,2)$$

这样平凡的松弛校正, 迭代序列 $\{w^k\}$ 满足

$$\|w^{k+1}-w^*\|_H^2\le\|w^k-w^*\|_H^2-\alpha(2-\alpha)\|w^k-\tilde{w}^k\|_H^2,\quad\forall\,w^*\in\Omega^*.$$

4 平行处理子问题的均困方法

均困平衡的 ALM 类算法, 把困难均分了, 平行处理子问题变得更有意义. 从给定的 $w^k=(x^k,\lambda^k)$ 出发, 如果由下面的公式

$$\left\{\begin{array}{ll}\tilde{x}^k=\arg\min\{\theta(x)-x^TA^T\lambda^k+\frac{r}{2}\|x-x^k\|^2|x\in\mathcal{X}\}, & (4.1\text{a})\\ \tilde{\lambda}^k=\arg\min\big\{\frac{1}{2}(\lambda-\lambda^k)^T\big(\frac{1}{r}AA^T+\delta I_m\big)(\lambda-\lambda^k)+\lambda^T(Ax^k-b)\big|\lambda\in\Lambda\big\} & (4.1\text{b})\end{array}\right.$$

产生预测点, 是可以平行处理的. 根据最优性条件就有

$$\left\{\begin{array}{ll}\tilde{x}^k\in\mathcal{X}, & \theta(x)-\theta(\tilde{x}^k)+(x-\tilde{x}^k)^T\{-A^T\lambda^k+r(\tilde{x}^k-x^k)\}\ge0,\quad\forall x\in\mathcal{X},\\ \tilde{\lambda}^k\in\Lambda, & (\lambda-\tilde{\lambda}^k)^T\{(\frac{1}{r}AA^T+\delta I_m)(\tilde{\lambda}^k-\lambda^k)+(Ax^k-b)\}\ge0,\quad\forall\lambda\in\Lambda.\end{array}\right.$$

我们把上式改写成

$$\left\{\begin{array}{l}\theta(x)-\theta(\tilde{x}^k)+(x-\tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}+r(\tilde{x}^k-x^k)+A^T(\tilde{\lambda}^k-\lambda^k)\}\ge0,\quad\forall x\in\mathcal{X},\\ (\lambda-\tilde{\lambda}^k)^T\{\underline{(A\tilde{x}^k-b)}-A(\tilde{x}^k-x^k)+(\frac{1}{r}AA^T+\delta I_m)(\tilde{\lambda}^k-\lambda^k)\}\ge0,\quad\forall\lambda\in\Lambda.\end{array}\right.$$

其中带下划线的部分组合在一起就是 $F(\tilde{w}^k)$. 利用变分不等式 (0.3) 的表达式, 上式中下划线部分放在一起就是 $F(\tilde{w}^k)$. 因此, 求得的预测点 $\tilde{w}^k$ 满足

$$\tilde{w}^k\in\Omega,\ \ \theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^T\big\{F(\tilde{w}^k)+Q(\tilde{w}^k-w^k)\big\}\ge0,\ \ \forall w\in\Omega, \qquad (4.2\text{a})$$

其中

$$Q=\begin{pmatrix} rI_n & A^T \\ -A & \frac{1}{r}AA^T+\delta I_m \end{pmatrix}, \quad \delta>0. \tag{4.2b}$$

上面的矩阵 Q^T+Q 是正定的.

对由 (4.1) 生成的预测点, 我们采用步长 $\alpha=1$ 的非平凡校正

$$w^{k+1}=w^k-M(w^k-\tilde{w}^k), \tag{4.3a}$$

生成新的迭代点, 其中

$$M=\begin{pmatrix} I_n & \frac{1}{r}A^T \\ -(\frac{1}{r}AA^T+\delta I_m)^{-1}A & I_m \end{pmatrix}. \tag{4.3b}$$

取正定矩阵

$$H=\begin{pmatrix} rI_n & 0 \\ 0 & \frac{1}{r}AA^T+\delta I_m \end{pmatrix}, \tag{4.4}$$

用统一框架去验证收敛性条件, 注意到

$$\begin{aligned} HM &= \begin{pmatrix} rI_n & \\ & \frac{1}{r}AA^T+\delta I_m \end{pmatrix}\begin{pmatrix} I_n & \frac{1}{r}A^T \\ -(\frac{1}{r}AA^T+\delta I_m)^{-1}A & I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n & A^T \\ -A & \frac{1}{r}AA^T+\delta I_m \end{pmatrix}, \end{aligned}$$

的确有 $HM=Q$. 此外

$$\begin{aligned} G &= Q^T+Q-M^THM=Q^T+Q-Q^TM \\ &= \begin{pmatrix} 2rI_n & 0 \\ 0 & 2(\frac{1}{r}AA^T+\delta I_m) \end{pmatrix} \\ &\quad -\begin{pmatrix} rI_n & -A^T \\ A & \frac{1}{r}AA^T+\delta I_m \end{pmatrix}\begin{pmatrix} I_n & \frac{1}{r}A^T \\ -(\frac{1}{r}AA^T+\delta I_m)^{-1}A & I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n-A^T(\frac{1}{r}AA^T+\delta I_m)^{-1}A & 0 \\ 0 & \delta I_m \end{pmatrix}. \end{aligned}$$

矩阵 H 和 G 都正定. 分别用 (2.1) 做预测和 (2.3) 做校正, 产生新的迭代点 w^{k+1} 满足

$$\|w^{k+1}-w^*\|_H^2 \le \|w^k-w^*\|_H^2-\|w^k-\tilde{w}^k\|_G^2, \quad \forall\, w^*\in\Omega^*.$$

我们就得到一个求解(与优化问题 (0.1) 对应的)变分不等式 (0.3) 的收敛算法.

校正 (4.3) 的具体实现就是

$$\begin{cases} x^{k+1}=\tilde{x}^k-\dfrac{1}{r}A^T(\lambda^k-\tilde{\lambda}^k), & (4.5a) \\ \lambda^{k+1}=\tilde{\lambda}^k+(\dfrac{1}{r}AA^T+\delta I)^{-1}(x^k-\tilde{x}^k). & (4.5b) \end{cases}$$

(4.5b) 中的 λ^{k+1} 是线性方程组

$$(\frac{1}{r}AA^T+\delta I)(\lambda-\tilde{\lambda}^k)-(x^k-\tilde{x}^k)=0$$

的解. 为了这个校正, 在整个迭代过程的校正中, 需要做一次 $(\frac{1}{r}AA^T+\delta I)$ 的 Cholesky 分解. 对等式约束的问题, 做预测 (4.1b) 本来就要做一次这样的 Cholesky 分解的.

参考文献

[1] A. Beck, First-Order Methods in convex optimization, MOS-SIAM Series on Optimization.

[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] A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vison, 40, 120-145, 2011.

[4] Chambolle A and Pock T. On the ergodic convergence rates of a first-order primal-dual algorithm, *Mathematical Programming*, 2016, 159: 253–287.

[5] G. Golub and C. F. Van Loan, Matrix Computations, The Johns Hopkins University Press, The Fourth Edition, 2013.

[6] 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.

[7] 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.

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

[9] J. Stoer and R. Bulirsch, Introduction to Numerical Analysis, Second Edition, Text in Applied Mathematics 12, Springer-Verlag, 1991.