E-效率更高的求解两个可分离块凸优化问题的ADMM类方法

稍做改动能提高效率的ADMM maths.nju.edu.cn/~hebma 何炳生

1 两块可分离凸优化问题所对应的变分不等式

在短篇 **C** 中, 我们已经把线性约束两块的可分离凸优化 (不要求 θ_1, θ_2 中任何一个强凸) 问题

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

的拉格朗日函数 $L(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)$ 的鞍点化为变分不等式

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

$$\text{其中}\qquad w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}. \tag{1.3}$$

2 分裂收缩算法的预测-校正统一框架 (何炳生 2015、2018 运筹学报)

预测-校正的统一框架, 我们 2015 年在"运筹学学报"文章中首次提出, 2018 年综述文章中再次提及.

[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

$$\tilde{w}^k\in\Omega,\ \ \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge (v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\ \ \forall\, w\in\Omega, \tag{2.1}$$

成立. 其中矩阵 Q^T+Q 是正定的. 如果 (2.1) 中 $v^k=\tilde{v}^k$, $\tilde{w}^k$ 相当于 (1.2) 的 w^*, 就是问题的解.

这里处理两块可分离凸优化问题的变分不等式中, $u=(x,y)$, $w=(x,y,\lambda)$, 核心变量为 $v=(y,\lambda)$.

[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

$$v^{k+1}=v^k-M(v^k-\tilde{v}^k). \tag{2.2}$$

我们称 (2.1) 中的 Q 为预测矩阵, (2.2) 中的 M 为校正矩阵.

收敛性条件 对给定的预测矩阵 Q, 收敛方法要求设计的校正矩阵 M 满足如下条件:

$$\text{存在正定矩阵}\ \ H\succ 0\ \ \text{使得}\ \ HM=Q,\ \ \text{并且}\ \ G:=Q^T+Q-M^THM\succ 0. \tag{2.3}$$

这节的预测 (2.1) 和校正 (2.2) 分别是短篇 **A** 中的 (2.4) 和 (2.7). 收敛性条件 (2.3) 与短篇 **A** 中的 (2.8) 相同. 换句话说, 在短篇 **A** 的第 2 节中, 我们已经对上述凸优化分裂收缩算法统一框架的收敛性做了简单证明. 在后面的 短篇 **G-H** 中, 有统一框架算法收敛性的更详细证明. 短篇 **E** 和 **F** 只是用算法框架验证收敛性.

3 交换顺序且延伸的 ADMM 方法 (Cai, Gu, He and Yuan 2013 Science China Mathematics)

修正的一种自然的想法是交换经典ADMM中 y 和 λ 更新的顺序. 从给定的 $v^k=(y^k,\lambda^k)$ 出发, 通过

$$\left[\begin{array}{c}x\text{-}\lambda\text{-}y\ \text{顺序}\\ \text{的ADMM}\end{array}\right]\quad \left\{\begin{array}{l} x^{k+1}\in\text{argmin}\{\theta_1(x)-x^TA^T\lambda^k+\frac12\beta\|Ax+By^k-b\|^2\,|\,x\in\mathcal{X}\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^k-b),\\ y^{k+1}\in\text{argmin}\{\theta_2(y)-y^TB^T\lambda^{k+1}+\frac12\beta\|Ax^{k+1}+By-b\|^2\,|\,y\in\mathcal{Y}\}\end{array}\right. \tag{3.1}$$

直接求得 w^{k+1}, 这个方法跟 短篇 **C** 中介绍的经典ADMM收敛表现几乎相当. 如果继续采用延伸

$$v^{k+1}:=v^k-\alpha(v^k-v^{k+1}),\quad \alpha=1.5\in(0,2), \tag{3.2}$$

效率一般有 30% 的提高. 为了能用§2中的统一框架验证收敛性, 我们把(3.1)-(3.2) 写成等价的预测-校正

[预测]

$$\tilde{x}^k\in\text{argmin}\{\theta_1(x)-x^TA^T\lambda^k+\tfrac12\beta\|Ax+By^k-b\|^2\,|\,x\in\mathcal{X}\}, \tag{3.3a}$$
$$\tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k-b), \tag{3.3b}$$
$$\tilde{y}^k\in\text{argmin}\{\theta_2(y)-y^TB^T\tilde{\lambda}^k+\tfrac12\beta\|A\tilde{x}^k+By-b\|^2\,|\,y\in\mathcal{Y}\}. \tag{3.3c}$$

[校正]

$$\begin{pmatrix} y^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix} y^k\\ \lambda^k\end{pmatrix}-\alpha\begin{pmatrix} y^k-\tilde{y}^k\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix},\quad \alpha\in(0,2). \tag{3.4}$$

利用 (3.3b), (3.3c) 的等价形式是 $\tilde{y}^k\in\text{argmin}\{\theta_2(y)-y^TB^T[2\tilde{\lambda}^k-\lambda^k]+\frac12\beta\|B(y-y^k)\|^2\,|\,y\in\mathcal{Y}\}$.

这样, 我们根据优化问题的最优性条件 (见 短篇 **D** 中的 Theorem D), 由预测 (3.3) 得到形为 (2.1) 的变分不等式形式, 其中 $w, u, F(w)$ 由 (1.3) 给出. 注意到 $v=(y,\lambda)$, 统一框架中预测和校正矩阵分别是

$$Q = \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} \qquad \text{和} \qquad M = \alpha \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}. \tag{3.5}$$

由于预测矩阵 Q 本身就是对称半正定的, 我们取

$$H = \frac{1}{\alpha} \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} = \frac{1}{\alpha} Q, \tag{3.6}$$

可以验证 H 半正定并有 $HM = Q$. 此外,

$$G = (Q^T + Q) - M^T Q = (2-\alpha) Q = (2-\alpha) \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} \tag{3.7}$$

也是对称半正定的. 理论上, 当 H 和 G 正定, 收敛性条件 (2.3) 才满足. 事实上, 半正定情况下仍然收敛.

4 对称的 ADMM 类方法 (He, Liu, Wang and Yuan, 2014 SIAM Opt.)

修正ADMM的另一种自然的想法是分别求解 x, y 子问题后, 各自校正一次 Lagrange 乘子, 我们称之为对称. 换句话说, 从给定的 $v^k = (y^k, \lambda^k)$ 出发, 按照下述方式求得 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$:

[对称的ADMM]

$$\begin{cases} x^{k+1} = \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(4.1a)} \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), \qquad \big(\mu = 0.9 \in (0,1)\big) & \text{(4.1b)} \\ y^{k+1} = \operatorname{argmin}\{\theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{1}{2}\beta\|Ax^{k+1} + By - b\|^2 \,|\, y \in \mathcal{Y}\}, & \text{(4.1c)} \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b). & \text{(4.1d)} \end{cases}$$

为了能用§2中的框架验证方法的收敛性, 需要把 (4.1) 分拆成预测与校正. 先按下述方式生成预测点 $\tilde{w}^k$:

[预测]

$$\begin{cases} \tilde{x}^k \in \operatorname{argmin}\big\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\big\}, & \text{(4.2a)} \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta\big(A\tilde{x}^k + By^k - b\big), & \text{(4.2b)} \\ \tilde{y}^k \in \operatorname{argmin}\big\{\theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{1}{2}\beta\|A\tilde{x}^k + By - b\|^2 \mid y \in \mathcal{Y}\big\}, & \text{(4.2c)} \\ \tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k - b\big), & \text{(4.2d)} \end{cases}$$

根据 短篇 **D** 的 Theorem D 所说的最优性条件, 从预测 (4.2), 经演算得到的形为 (2.1) 的变分不等式形式中, 预测矩阵

$$Q = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{4.3}$$

由于 (4.2) 中的 $\tilde{x}$ 和 $\tilde{y}^k$ 分别等于 (4.1) 中的 x^{k+1} 和 y^{k+1}. 对 (4.1d) 中的 λ^{k+1}, 利用 (4.2b), 我们有

$$\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta\big[-B(y^k - \tilde{y}^k) + (A\tilde{x}^k + By^k - b)\big] = \lambda^k - \mu\beta\big[-B(y^k - \tilde{y}^k) + 2(A\tilde{x}^k + By^k - b)\big].$$

再利用 (4.2d), 即得 $\lambda^{k+1} = \lambda^k - \big[-\mu\beta B(y^k - \tilde{y}^k) + 2\mu(\lambda^k - \tilde{\lambda}^k)\big]$. 因此, 新的核心变量 (y^{k+1}, λ^{k+1}) 由

[校正]

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}, \tag{4.4}$$

产生. 对于 (4.4) 中的校正矩阵

$$M = \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I \end{pmatrix}, \qquad \text{令} \quad H = \begin{pmatrix} (1 - \frac{1}{2}\mu)\beta B^T B & -\frac{1}{2} B^T \\ -\frac{1}{2} B & \frac{1}{2\mu\beta} I \end{pmatrix}, \tag{4.5}$$

可以验证当 $\mu \in (0,1)$ 时, H 正定并有 $HM = Q$ (矩阵 Q 见 (4.3)). 此外, 矩阵

$$G = (Q^T + Q) - M^T Q = (1-\mu) \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} \succ 0, \tag{4.6}$$

收敛性条件 (2.3) 满足. 跟经典的ADMM一样, 只要当 B 列满秩时 H 和 G 正定, 此时称它们本质上正定.

对一个看起来收敛的方法, 总拆解成预测-校正方法 (2.1)-(2.2), 然后去验证收敛条件 (2.3) 是否满足.

注意: 两块可分离的凸优化问题 (1.1), 核心变量 $v=(y,\lambda)$ 时, 乘子预测总用 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$.