F-求解三个可分离块凸优化问题的 ADMM 类分裂收缩算法

不附加强凸条件的收敛方法 maths.nju.edu.cn/~hebma 何炳生

1 直接推广 ADMM 求解三块可分离问题不能保证收敛 (Chen, He, Ye, Yuan 2016 MP)

我们把一般的线性约束三块的可分离凸优化 (不要求 $\theta_1, \theta_2, \theta_3$ 中任何一个强凸) 问题

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) | Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\} \tag{1.1}$$

的拉格朗日函数 $L(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b)$ 的鞍点化为变分不等式

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

其中

$$w=\begin{pmatrix} x\\ y\\ z\\ \lambda\end{pmatrix},\ u=\begin{pmatrix} x\\ y\\ z\end{pmatrix},\ \theta(u)=\theta_1(x)+\theta_2(y)+\theta_3(z),\ F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ -C^T\lambda\\ Ax+By+Cz-b\end{pmatrix}. \tag{1.2b}$$

我们在 2016 年的 MP 文章中已经指出, 用直接推广的 ADMM 求解凸优化问题 (1.1), 当矩阵 A, B, C 中有两个矩阵的列互相正交, 即 $A^TB=0$, $A^TC=0$ 或者 $B^TC=0$ 时方法收敛. 在一般情形下, 并不能保证收敛.

2 可以用来验证算法收敛的统一框架 (何炳生 2015 年运筹学报第 19 卷第 3 期)

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.

在我们这里处理的三块的可分离问题中, $u=(x,y,z), w=(x,y,z,\lambda), v=(y,z,\lambda)$.

[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, 要求设计的校正矩阵 M 满足如下条件:

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

如果条件 (2.3) 满足, 算法就是收敛的, 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.

3 采用 Gauss 回代的预测-校正方法 (He, Tao and Yuan 2012 SIAM Opt.)

从给定的 $v^k=(y^k,z^k,\lambda^k)$ 出发, 通过串行求解子问题, 然后定义 $\tilde{\lambda}^k$ (请注意预测中 $\tilde{\lambda}^k$ 的定义方式)

$$\text{[预测]}\quad \begin{cases} \tilde{x}^k\in\arg\min\big\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}\beta\|Ax+By^k+Cz^k-b\|^2\ |\ x\in\mathcal{X}\big\},\\ \tilde{y}^k\in\arg\min\big\{\theta_2(y)-y^TB^T\lambda^k+\frac{1}{2}\beta\|A\tilde{x}^k+By+Cz^k-b\|^2\ |\ y\in\mathcal{Y}\big\},\\ \tilde{z}^k\in\arg\min\big\{\theta_3(z)-z^TC^T\lambda^k+\frac{1}{2}\beta\|A\tilde{x}^k+B\tilde{y}^k+Cz-b\|^2\ |\ z\in\mathcal{Z}\big\},\\ \tilde{\lambda}^k=\lambda^k-\beta\big(A\tilde{x}^k+By^k+Cz^k-b\big),\end{cases} \tag{3.1}$$

得到预测点 $\tilde{w}^k$. 利用优化问题的最优性条件, 我们得到预测的 (2.1) 形式, 其中 $w, u, F(w)$ 由 (1.2b) 给出,

$$Q=\begin{pmatrix} \beta B^TB & 0 & 0\\ \beta C^TB & \beta C^TC & 0\\ -B & -C & \frac{1}{\beta}I_m\end{pmatrix}. \tag{3.2}$$

Q^T+Q 本质上是正定的 (当 B, C 列满秩时是正定的). 利用这样的预测点, 我们采用校正

$$\text{[校正]}\quad \begin{pmatrix} y^{k+1}\\ z^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix} y^k\\ z^k\\ \lambda^k\end{pmatrix}-\begin{pmatrix} \nu I & -\nu(B^TB)^{-1}B^TC & 0\\ 0 & \nu I & 0\\ -\beta B & -\beta C & I\end{pmatrix}\begin{pmatrix} y^k-\tilde{y}^k\\ z^k-\tilde{z}^k\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix} \tag{3.3}$$

生成 $v^{k+1}=(y^{k+1},z^{k+1},\lambda^{k+1})$. 也就是说, 在统一框架的校正公式 (2.2) 中

$$M=\begin{pmatrix} \nu I & -\nu(B^TB)^{-1}B^TC & 0\\ 0 & \nu I & 0\\ -\beta B & -\beta C & I \end{pmatrix}. \tag{3.4}$$

对于矩阵

$$H=\begin{pmatrix} \frac{1}{\nu}\beta B^TB & \frac{1}{\nu}\beta B^TC & 0\\ \frac{1}{\nu}\beta C^TB & \frac{1}{\nu}\beta[C^TC+C^TB(B^TB)^{-1}B^TC] & 0\\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix} \tag{3.5}$$

可以验证 H 正定并有 $HM=Q$. 此外, 当 $\nu\in(0,1)$ 时, 矩阵

$$G=(Q^T+Q)-M^TQ=\begin{pmatrix} (1-\nu)\beta B^TB & 0 & 0\\ 0 & (1-\nu)\beta C^TC & 0\\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}\succ 0, \tag{3.6}$$

收敛性条件 (2.3) 满足. 事实上, 由于预测 (3.1) 只需要 (By^k,Cz^k,λ^k) 就可以开始, 校正 (3.3) 也可以通过

$$\begin{pmatrix} By^{k+1}\\ Cz^{k+1}\end{pmatrix}=\begin{pmatrix} By^{k}\\ Cz^{k}\end{pmatrix}-\begin{pmatrix} \nu I & -\nu I\\ 0 & \nu I\end{pmatrix}\begin{pmatrix} B(y^k-\tilde{y}^k)\\ C(z^k-\tilde{z}^k)\end{pmatrix}, \quad \lambda^{k+1}=\lambda^k-\beta\big(A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b\big)$$ 实现.

上面这个式子中求 By^{k+1} 和 Cz^{k+1} 中出现的上三角矩阵, 是我们称其为 "Gauss 回代" 的原因.

4 部分平行并加正则项的 ADMM 方法 (He, Tao and Yuan 2015 IMA Numr. Analysis)

部分平行并加正则项的 ADMM 方法从给定的 $v^k=(y^k,z^k,\lambda^k)$ 出发, 通过

$$\begin{cases} x^{k+1}\in\operatorname{argmin}\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k+Cz^k-b\|^2\,|\,x\in\mathcal{X}\},\\ y^{k+1}\in\operatorname{argmin}\{\theta_2(y)-y^TB^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By+Cz^k-b\|^2+\frac{\tau\beta}{2}\|B(y-y^k)\|^2\,|\,y\in\mathcal{Y}\},\\ z^{k+1}\in\operatorname{argmin}\{\theta_3(z)-z^TC^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By^k+Cz-b\|^2+\frac{\tau\beta}{2}\|C(z-z^k)\|^2\,|\,z\in\mathcal{Z}\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b) \end{cases} \tag{4.1}$$

直接求得 w^{k+1}, 其中 $\tau>1$. 为了能用§2中的统一框架证明收敛性, 把收敛方法 (4.1) 故意拆解成等价的

[预测]
$$\begin{cases} \tilde{x}^k\in\operatorname{argmin}\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k+Cz^k-b\|^2\,|\,x\in\mathcal{X}\},\\ \tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k+Cz^k-b),\\ \tilde{y}^k\in\operatorname{argmin}\{\theta_2(y)-y^TB^T\tilde{\lambda}^k+\frac{\mu\beta}{2}\|B(y-y^k)\|^2\,|\,y\in\mathcal{Y}\},\\ \tilde{z}^k\in\operatorname{argmin}\{\theta_3(z)-z^TC^T\tilde{\lambda}^k+\frac{\mu\beta}{2}\|C(z-z^k)\|^2\,|\,z\in\mathcal{Z}\}. \end{cases} \tag{4.2}$$

[校正]
$$\begin{pmatrix} y^{k+1}\\ z^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix} y^{k}\\ z^{k}\\ \lambda^{k}\end{pmatrix}-\begin{pmatrix} I & 0 & 0\\ 0 & I & 0\\ -\beta B & -\beta C & I\end{pmatrix}\begin{pmatrix} y^k-\tilde{y}^k\\ z^k-\tilde{z}^k\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}. \tag{4.3}$$

其中 (4.2) 的 μ 和 (4.1) 中的 τ 有关系式 $\mu=\tau+1$. 利用优化问题的最优性条件 (见 **D** 中定理D), 由 (4.2) 产生的预测点 $\tilde{w}^k$ 满足 (2.1), 其中 $w,u,F(w)$ 由 (1.2b) 给出. $v=(y,z,\lambda)$, 预测和校正矩阵分别是

$$Q=\begin{pmatrix} \mu\beta B^TB & 0 & 0\\ 0 & \mu\beta C^TC & 0\\ -B & -C & \frac{1}{\beta}I\end{pmatrix} \quad \text{和} \quad M=\begin{pmatrix} I & 0 & 0\\ 0 & I & 0\\ -\beta B & -\beta C & I\end{pmatrix}. \tag{4.4}$$

对于矩阵

$$H=\begin{pmatrix} \mu\beta B^TB & 0 & 0\\ 0 & \mu\beta C^TC & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}, \tag{4.5}$$

可以验证 H 正定并有 $HM=Q$. 此外,

$$G=(Q^T+Q)-M^TQ=\begin{pmatrix} (\mu-1)\beta B^TB & -\beta B^TC & 0\\ -\beta C^TB & (\mu-1)\beta C^TC & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}. \tag{4.6}$$

由于 $\mu=\tau+1>2$, 矩阵 G 正定. 收敛性条件 (2.3) 满足. 在 He and Yuan, Optimization Online 6325 中 (此文发表在 Springer Proc. Math. Stat., 360, Springer, Singapore, 2021, 139-163), 我们已经证明:

预测-校正方法 (4.2)-(4.3) 中, 将 $\mu>2$ 改成 $\mu>1.5$ (实际计算取略大于1.5), 方法收敛速度有明显提高.