G-从好不容易凑出一个方法到并不费劲构造一簇算法

古稀之后的新收获–I maths.nju.edu.cn/˜hebma 何炳生

1 能够用来验证方法保证收敛的统一框架 (回顾一下退休前总结出的算法统一框架)

我们把线性约束凸优化问题的 Lagrange 函数的鞍点归结为求变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega \tag{1.1}$$

的解. 为求解这个变分不等式, 我们已经设计了下面的预测-校正的统一框架.

[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

$$\tilde{w}^k \in \Omega, \;\; \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \;\; \forall\, w \in \Omega, \tag{1.2}$$

成立. 其中矩阵 $Q^T + Q$ 是正定的. 如果 (1.2) 中 $v^k = \tilde{v}^k$, $\tilde{w}^k$ 相当于 (1.1) 的 w^*, 就是问题的解.

v 可以与 w 同, 也可以是 w 的部分分量. 例如, 在交替方向法中, $u=(x,y), w=(x,y,\lambda), v=(y,\lambda)$.

[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

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{1.3}$$

我们称 (1.2) 中的 Q 为预测矩阵, (1.3) 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：

$$\text{存在正定矩阵} \;\; H \succ 0 \;\; \text{使得} \;\; HM = Q, \;\; \text{并且} \;\; G := Q^T + Q - M^T H M \succ 0. \tag{1.4}$$

如果条件 (1.4) 满足, 算法就是收敛的, 读者将会在第二页的 §3 中看到, 关键收敛性质证明是相当容易的.
以往我们主要是把方法看作(或故意分拆成)预测 (1.2) 和校正 (1.3), 然后去验证收敛条件 (1.4) 能否满足.

事实上, 我们只要求 $Q^T + Q$, H 和 G 本质上是正定的, 这相当于只要求它们半正定, 但是当相应的可分离凸优化问题的线性约束中被分离的分块矩阵都列满秩时就要求它们正定.

就像短篇 **C** 处理两块可分离问题的 ADMM 中, H 正定要求约束矩阵 B 是列满秩, 即 B^TB 是正定的.

2 可以用来构造一簇算法的等价的统一框架 (古稀后发现框架能用来开发算法)

有了预测 (1.2), 关键是怎样给出满足条件 (1.4) 的矩阵 M, 才能去实行校正 (1.3).

下面我们推导出一个跟条件 (**??**) 等价的构造校正矩阵 M 的方法:

$$\left\{\begin{array}{l} \text{预测 (1.2) 提供的 } Q \text{ 满足 } Q^T+Q \succ 0. \\ \text{收敛条件 (1.4) 要求选出的校正矩阵 } M: \\ \textbf{存在 } H \succ 0, \text{ 使得 } \; HM = Q; \\ \textbf{并且 } G = Q^T + Q - M^T H M \succ 0. \end{array}\right. \Longleftrightarrow \left\{\begin{array}{r} D \succ 0, \quad G \succ 0, \\ D + G = Q^T + Q, \\ M^T H M = D, \\ HM = Q. \end{array}\right.$$

$$\Longleftrightarrow \left\{\begin{array}{r} D \succ 0, \quad G \succ 0, \\ D + G = Q^T + Q, \\ Q^T M = D, \\ HM = Q. \end{array}\right. \Longleftrightarrow \left\{\begin{array}{r} D \succ 0, \quad G \succ 0, \\ D + G = Q^T + Q, \\ M = Q^{-T} D, \\ H = Q D^{-1} Q^T. \end{array}\right. \tag{2.5}$$

我们只要求存在一个对称正定矩阵 H, 保证 $HM=Q$ 和 $G \succ 0$. 上式得到的 H 和 M 确实满足 (1.4).

有了合格的预测矩阵 Q, 以前是想办法去凑 H 和 M, 使其满足条件 (1.4), 让人看起来似乎有点神秘.

现在的做法: 一旦有了合格的预测矩阵 Q, 就可以选择多种多样的 D, 使其满足 $0 \prec D \prec Q^T + Q$.

然后由 $M = Q^{-T} D$ 得到矩阵 M 进行 (1.3) 校正, 收敛性条件自然满足. 而 Q^T 的求逆往往是容易的！

校正 $v^{k+1} = v^k - M(v^k - \tilde{v}^k)$ 可以通过 $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$ 实现, Q 往往是块三角矩阵.

3 统一框架中方法收敛性关键收缩性质的证明

定理 G 采用预测-校正方法 (1.2)-(1.3) 求解变分不等式 (1.1), 如果条件 (1.4) 满足, 则方法生成的序列 $\{v^k\}$ 和 $\{\tilde{w}^k\}$ 具有性质

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge \tfrac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\tfrac{1}{2}\|v^k-\tilde{v}^k\|_G^2,\quad \forall w\in\Omega \tag{3.1}$$

和

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-\tilde{v}^k\|_G^2,\quad \forall v^*\in\mathcal{V}^*. \tag{3.2}$$

证明 在 (1.2) 式的右端, 利用 $Q=HM$ (见(1.4)中的前一式), 便有

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^THM(v^k-\tilde{v}^k).$$

由于 $M(v^k-\tilde{v}^k)=(v^k-v^{k+1})$ (见校正公式 (1.3)), 上式可以写成

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TH(v^k-v^{k+1}). \tag{3.3}$$

将恒等式

$$(a-b)^TH(c-d)=\tfrac{1}{2}\{\|a-d\|_H^2-\|a-c\|_H^2\}+\tfrac{1}{2}\{\|b-c\|_H^2-\|b-d\|_H^2\} \tag{3.4}$$

用于不等式 (3.3) 的右端, 并设

$$a=v,\quad b=\tilde{v}^k,\quad c=v^k,\quad 和\quad d=v^{k+1},$$

我们得到

$$(v-\tilde{v}^k)^TH(v^k-v^{k+1})=\tfrac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\tfrac{1}{2}\big(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\big). \tag{3.5}$$

对 (3.5) 右端第二部分, 利用 $HM=Q$ 和 $2v^TQv=v^T(Q^T+Q)v$, 我们有

$$\begin{aligned}
&\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2 && \text{(利用 (1.3))}\\
&= \|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-M(v^k-\tilde{v}^k)\|_H^2\\
&= (v^k-\tilde{v}^k)^T(HM+M^TH)(v^k-\tilde{v}^k)-(v^k-\tilde{v}^k)^\top M^\top HM(v^k-\tilde{v}^k)\\
&= (v^k-\tilde{v}^k)^T(Q+Q^T-M^THM)(v^k-\tilde{v}^k) && \text{(利用 (1.4)中的后一式)}\\
&= \|v^k-\tilde{v}^k\|_G^2.
\end{aligned} \tag{3.6}$$

以 (3.6) 的结果代入 (3.5), 然后再代入 (3.3), 就得到结论 (3.1). 将 (3.1) 中的 w 设为任意固定的 w^*, 就有

$$\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2\ge\|v^k-\tilde{v}^k\|_G^2+2\{\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^TF(\tilde{w}^k)\}. \tag{3.7}$$

根据 $(\tilde{w}^k-w^*)^TF(\tilde{w}^k)=(\tilde{w}^k-w^*)^TF(w^*)$ 和最优性条件, 我们有

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

因此从 (3.7) 得到

$$\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2\ge\|v^k-\tilde{v}^k\|_G^2.$$

这就是定理的结论 (3.2). □

从结论 (3.1) 容易推得算法遍历意义下的收敛速率. He & Yuan 关于 ADMM 的遍历意义下收敛速率就是基于不等式(3.1) 证明的. 见 B. S. He and X. M. Yuan, SIAM J. Numerical Analysis, 2012, 50: 700-709.
我们称 (3.2) 中的 H 为范数矩阵, G 为效益矩阵. 结论 (3.2) 是证明方法总体收敛的关键不等式!

从证明过程可以知道, 定理 **G** 的结论当收敛条件 (1.4) 中 Q^T+Q, H 和 G 弱化成半正定时仍然成立.

4 从验证方法收敛的框架到自由设计算法的纲领

定理G 是预测-校正方法(1.2)-(1.3)在条件(1.4)下证明的. 条件(2.5)和(1.4) 等价, 可以根据 (2.5) 构造算法.
换句话说, §1 中的条件 (1.4) 主要用来验证方法收敛性. §2 中的等价性证明为算法设计提供了纲领.

设计算法的纲领. 对预测 (1.2) 中满足 $Q^T+Q\succ 0$ 的预测矩阵 Q, 选 D, 使得

$$0\prec D\prec Q^T+Q,\qquad 然后取\qquad M=Q^{-T}D\qquad 去完成校正 (1.3).$$

这样, 预测-校正方法 (1.2)-(1.3) 产生的序列 $\{v^k\}$ 满足定理 **G** 中的结论, 其中

$$H=QD^{-1}Q^T\qquad 和\qquad G=Q^T+Q-D,$$

都是正定矩阵. 计算过程中并不要求给出显式的 H.

矩阵 Q 一般是由一个 容易求逆的块三角矩阵 加上 一个 广义秩一(或秩二)矩阵 组成.

直接给出 $M=Q^{-T}D$ 完成校正 (1.3) 或者求解 $Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)$ 都并不困难!