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　(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

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

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

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

成立. 其中矩阵 $Q^{\top}+Q$ 是正定的. 左端将 (1.1) 中的 w^* 换成了 $\tilde{w}^k$. 右端的 v 可以是 w.

[**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

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

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

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：

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

如果条件 (1.4) 满足, 算法就是收敛的, 将会在 §3 中看到, 关键收敛性质证明相当容易.
以往主要是把方法分拆成预测 (1.2) 和校正 (1.3), 然后再去验证收敛条件 (1.4) 能否满足.

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 H 和 M, 才能去实行校正 (1.3).

下面我们推导出一个跟条件 (1.4) 等价的构造矩阵 M 和 H 的方法:

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

$$\iff \left\{\begin{array}{l} D\succ 0,\quad G\succ 0,\\ D+G=Q^{\top}+Q,\\ Q^{\top}M=D,\\ HM=Q.\end{array}\right. \iff \left\{\begin{array}{l} D\succ 0,\quad G\succ 0,\\ D+G=Q^{\top}+Q,\\ M=Q^{-T}D,\\ H=QD^{-1}Q^{\top}.\end{array}\right. \tag{2.1}$$

我们只要求存在一个对称正定矩阵 H, 并不需要计算 H, 也不需要其显式表达式.

有了预测及其矩阵 Q, 以前是想办法去凑 H 和 M, 使其满足条件 (1.4), 似乎有点神秘.
现在的做法: 有了合格的预测矩阵 Q, 可以选多样的 D, 使其满足 $0\prec D\prec Q^{\top}+Q$
由 $M=Q^{-T}D$ 得到校正矩阵 M, 收敛性自然条件满足. 而 $Q^{\top}$ 的求逆往往是容易的!
此外校正 $v^{k+1}=v^k-M(v^k-\tilde{v}^k)$ 可以通过 $Q^{\top}(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)$ 实现.

3 统一框架中方法收敛性关键收缩性质的证明

定理 1 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, 则产生的序列 $\{v^k\}$ 具有收缩性质

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{3.1}$$

证明 将 (1.2) 中任意的 w 设为 w^*, 利用 $Q = HM$ (见(1.4)中的前一式), 便有

$$\theta(u^*) - \theta(\tilde{u}^k) + (w^* - \tilde{w}^k)^\top F(\tilde{w}^k) \ge (v^* - \tilde{v}^k)^\top HM(v^k - \tilde{v}^k).$$

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

$$\theta(u^*) - \theta(\tilde{u}^k) + (w^* - \tilde{w}^k)^\top F(\tilde{w}^k) \ge (v^* - \tilde{v}^k)^\top H(v^k - v^{k+1})$$

并得到

$$(v^k - v^{k+1})^\top H(\tilde{v}^k - v^*) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^\top F(\tilde{w}^k). \tag{3.2}$$

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

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

因此 (3.2) 的右端非负, 我们得到

$$(v^k - v^{k+1})^\top H(\tilde{v}^k - v^*) \ge 0. \tag{3.3}$$

将恒等式 $(a-b)^\top H(c-d) = \frac{1}{2}\{\|a-d\|_H^2 - \|b-d\|_H^2\} - \frac{1}{2}\{\|a-c\|_H^2 - \|b-c\|_H^2\}$, 用于不等式 (3.3) 的左端, 并设

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

得到

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2. \tag{3.4}$$

对 (3.4) 的右端, 利用 $HM = Q$ 和 $2v^\top Qv = v^\top(Q^\top + Q)v$, 我们有

$$\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
=\;& \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \qquad (\text{利用 } (1.3)) \\
=\;& \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\
=\;& (v^k - \tilde{v}^k)^\top (HM + M^T H)(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^\top M^\top HM(v^k - \tilde{v}^k) \\
=\;& (v^k - \tilde{v}^k)^\top (Q + Q^\top - M^\top HM)(v^k - \tilde{v}^k) \qquad (\text{利用 } (1.4)\text{中的后一式}) \\
=\;& \|v^k - \tilde{v}^k\|_G^2.
\end{aligned} \tag{3.5}$$

将 (3.5) 代入 (3.4), 就得到定理的结论 (3.1). □

我们称 (3.1) 中的 H 为范数矩阵, G 为效益矩阵. (3.1) 是证明收敛的关键不等式!

4 从验证方法收敛的框架到自由设计算法的纲领

定理 1 是预测-校正方法 (1.2)-(1.3) 在条件 (1.4) 下证明的. 条件 (2.1) 和 (1.4) 等价!
§1 中的条件 (1.4) 主要用来验证方法收敛性. §2 中的等价性为开发算法设计提供了纲领.

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

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

这样, 预测-校正方法 (1.2)-(1.3) 产生的序列 $\{v^k\}$ 满足定理 1 中的结论 (3.1), 其中

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

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