从变分不等式的投影收缩算法到凸优化的分裂收缩算法

退休前总结的主要研究工作 maths.nju.edu.cn/~hebma 何炳生

1 变分不等式的投影收缩算法

设 $\Omega\subset\Re^n$ 是一个非空闭凸集, F 是 $\Re^n\to\Re^n$ 的一个映射. 考虑单调变分不等式

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

我们说变分不等式单调, 是指其中的算子 F 满足 $(u-v)^\top(F(u)-F(v))\ge 0$. 在求解变分不等式 (1.1) 的投影收缩算法中, 对给定的当前点 u^k 和 $\beta_k>0$, 我们利用投影

$$\tilde u^k = P_\Omega[u^k-\beta_kF(u^k)] \quad \Big(\text{也即}\ \ \tilde u^k=\arg\min\big\{\tfrac12\|u-[u^k-\beta_kF(u^k)]\|^2\,|\,u\in\Omega\big\}\Big)$$

生成一个预测点 $\tilde u^k$. 对任意的 $\beta_k>0$, 假如 $\tilde u^k=u^k$, u^k 就是 (1.1) 的解. 否则, 开始一次新的迭代. 我们假设选取的 β_k 使得预测点 $\tilde u^k$ 满足

$$\beta_k\|F(u^k)-F(\tilde u^k)\|\le \nu\|u^k-\tilde u^k\|,\quad \nu\in(0,1). \tag{1.2}$$

由于 $\tilde u^k$ 是问题 $\min\{\frac12\|u-[u^k-\beta_kF(u^k)]\|^2\,|\,u\in\Omega\}$ 的解. 根据极小化问题的最优性条件(与"瞎子爬山"类似, 只是这里是探"底"), 可行方向和下降方向的交集是空集, 就有

$$\tilde u^k\in\Omega,\quad (u-\tilde u^k)^\top\{\tilde u^k-[u^k-\beta_kF(u^k)]\}\ge 0,\quad \forall u\in\Omega. \tag{1.3}$$

上式两边都加上 $(u-\tilde u^k)^\top d(u^k,\tilde u^k)$, 其中

$$d(u^k,\tilde u^k)=(u^k-\tilde u^k)-\beta_k[F(u^k)-F(\tilde u^k)]. \tag{1.4}$$

由此得到我们需要的预测公式

$$[\text{预测}]\qquad \tilde u^k\in\Omega,\ \ (u-\tilde u^k)^\top\beta_kF(\tilde u^k)\ge (u-\tilde u^k)^\top d(u^k,\tilde u^k),\ \ \forall u\in\Omega. \tag{1.5}$$

将 (1.5) 中任意的 $u\in\Omega$ 选成 u^*, 利用 (1.1), 就有 $(\tilde u^k-u^*)^\top d(u^k,\tilde u^k)\ge 0$, 随后得到

$$(u^k-u^*)^\top d(u^k,\tilde u^k)\ge (u^k-\tilde u^k)^\top d(u^k,\tilde u^k). \tag{1.6}$$

由假设 (1.2), 利用 Cauchy-Schwarz 不等式, 立即推得

$$(u^k-\tilde u^k)^\top d(u^k,\tilde u^k)\ge(1-\nu)\|u^k-\tilde u^k\|^2. \tag{1.7}$$

当 u^k 不是 (1.1) 的解时, 不等式 (1.6) 右端为正, 这说明 (1.4) 中定义的 $d(u^k,\tilde u^k)$ 是未知函数 $\frac12\|u-u^*\|^2$ 在 u^k 处欧氏模下的一个上升方向. 因此, 可以

$$[\text{校正}]\quad \text{用}\quad u^{k+1}=u^k-\alpha_k^*d(u^k,\tilde u^k)\quad \text{产生离 } u^* \text{ 更近的迭代点}. \tag{1.8}$$

$$\text{其中步长}\quad \alpha_k^*\quad \text{由}\quad \alpha_k^*=(u^k-\tilde u^k)^\top d(u^k,\tilde u^k)/\|d(u^k,\tilde u^k)\|^2\quad \text{给出}. \tag{1.9}$$

由 (1.2) 和 (1.4), 可得 $2(u^k-\tilde u^k)^\top d(u^k,\tilde u^k)>\|d(u^k,\tilde u^k)\|^2$, 因而 $\alpha_k^*>\frac12$. 接着就有

$$\begin{aligned}&\|u^k-u^*\|^2-\|u^{k+1}-u^*\|^2 \qquad (\text{利用 (1.8), (1.6), (1.9) 和 (1.7)})\\ \ge\ & \alpha_k^*(u^k-\tilde u^k)^\top d(u^k,\tilde u^k)\ge \frac12(1-\nu)\|u^k-\tilde u^k\|^2.\end{aligned} \tag{1.10}$$

这是证明算法收敛的关键不等式. 分处不等式 (1.5) 两端的 $\beta_kF(\tilde u^k)$ 和 $d(u^k,\tilde u^k)$ 称为一对孪生方向. 这里最美妙的是: 采用方向 $\beta_kF(\tilde u^k)$ 和 (1.9) 中 给出的步长 α_k^*, 用

$$[\text{投影校正}]\quad u^{k+1}=P_\Omega[u^k-\alpha_k^*\beta_kF(\tilde u^k)]\quad \text{与方法 (1.8) 具有同样的收敛性质 (1.10)}.$$

✠ 孪生方向, 相同步长. 投影校正效果更好! 真所谓 兄弟齐心, 其利断金! ✠

2 线性约束凸优化分裂收缩算法的统一框架

我们在变分不等式框架下讨论线性约束凸优化的分裂收缩算法, 以可分离凸优化问题

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\} \tag{2.1}$$

为例做介绍. 问题 (2.1) 的拉格朗日函数是 $L(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^\top(Ax+By-b)$.

拉格朗日函数的鞍点 $((x^*,y^*),\lambda^*)$ 满足

$$L_{\lambda\in\Re^m}(x^*,y^*,\lambda)\le L(x^*,y^*,\lambda^*)\le L_{x\in\mathcal{X},y\in\mathcal{Y}}(x,y,\lambda^*).$$

利用鞍点 $w^*=((x^*,y^*),\lambda^*)$ 的上述极大极小性质, 它可以表述为以下的变分不等式:

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^\top F(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{2.2}$$

其中, $\Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m$, $\theta(u)=\theta_1(x)+\theta_2(y)$, 并且

$$u=\begin{pmatrix} x\\ y\end{pmatrix},\quad w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^\top\lambda\\ -B^\top\lambda\\ Ax+By-b\end{pmatrix}. \tag{2.3}$$

注意到, (2.3) 中的 $F(w)$ 恰有 $(w-\tilde{w})^\top(F(w)-F(\tilde{w}))=0$, 所以也是单调的.

求解 一般变分不等式问题 (2.2) 的预测-校正统一框架 (ADMM可以解释为其一个特例).

[**预测**]. 对给定的 v^k(在ADMM中$v=(y,\lambda)$, 有些算法中$v=w$), 求得点 $\tilde{w}^k$, 使其满足

$$\tilde{w}^k\in\Omega,\ \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^\top F(\tilde{w}^k)\ge(v-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k),\ \forall w\in\Omega, \tag{2.4}$$

如果矩阵 $Q^\top+Q$ 正定(Q 不一定对称), 我们称 $\tilde{w}^k$ 为合格的预测点.

将 (2.4) 中任意的 $w\in\Omega$ 选成 w^*, 就得到 $(\tilde{v}^k-v^*)^\top Q(v^k-\tilde{v}^k)\ge 0$, 随后得到

$$[H(v^k-v^*)]^\top[H^{-1}Q(v^k-\tilde{v}^k)]\ge(v^k-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k). \tag{2.5}$$

这里 H 为任意正定矩阵(也可以是单位阵). 由 (2.5) 和 Q^T+Q 正定, $H^{-1}Q(v^k-\tilde{v}^k)$ 是未知函数 $\frac{1}{2}\|v-v^*\|_H^2$ 在 v^k 处 H-模下的一个上升方向. 如同投影收缩算法中利用 (1.6),

[**计算步长的校正**]. 可以对选定的正定矩阵 H, 由以下法则生成新的迭代点:

$$v^{k+1}=v^k-\alpha_k^* M(v^k-\tilde{v}^k),\quad \text{其中}\quad M=H^{-1}Q,\quad \alpha_k^*=\frac{(v^k-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k)}{\|M(v^k-\tilde{v}^k)\|_H^2}.$$

[**单位步长的校正**]. 若采用单位步长, 生成新迭代点的校正公式为

$$v^{k+1}=v^k-M(v^k-\tilde{v}^k). \tag{2.6}$$

这就要求存在正定矩阵 H, 能够使得

$$HM=Q \quad \text{并且}\quad G=Q^\top+Q-M^\top HM\succ 0.\quad (\text{正定}) \tag{2.7}$$

$$\begin{aligned}
&\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2 && (\text{利用 }(2.6))\\
&\quad=\ \|v^k-v^*\|_H^2-\|(v^k-v^*)-M(v^k-\tilde{v}^k)\|_H^2 && (\text{利用 } HM=Q \text{ 和 } (2.5))\\
&\quad\ge\ 2(v^k-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k)-\|M(v^k-\tilde{v}^k)\|_H^2 \overset{(2.7)}{=} \|v^k-\tilde{v}^k\|_G^2.
\end{aligned} \tag{2.8}$$

不等式 (2.8) 说明统一框架若采用单位步长校正, 方法生成的序列 $\{v^k\}$ 具有收缩性质

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-\tilde{v}^k\|_G^2.\qquad (\text{这是证明收敛性的关键不等式})$$

✠ 从 变分不等式的投影收缩算法 到 凸优化的分裂收缩算法, 一条主线, 一个模式！✠