从变分不等式的投影收缩算法到凸优化的分裂收缩算法

— 我主导的主要研究工作 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

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)] \quad \Big(\text{也即} \ \ \tilde{u}^k = \mathrm{Arg\,min}\big\{\tfrac{1}{2}\|u - [u^k - \beta_k F(u^k)]\|^2 \,|\, u \in \Omega\big\}\Big)$$

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,
我们假设选取的 β_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 $\min\{\frac{1}{2}\|u - [u^k - \beta_k F(u^k)]\|^2 \,|\, u \in \Omega\}$ 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

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

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

$$\textbf{[预测]} \quad \tilde{u}^k \in \Omega, \ \ (u - \tilde{u}^k)^\top \beta_k F(\tilde{u}^k) \ge (u - \tilde{u}^k)^\top d(u^k, \tilde{u}^k), \ \ \forall u \in \Omega. \tag{1.5}$$

将 (1.5) 中任意的 $u \in \Omega$ 选成 u^*, 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), 容易推得

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

当 u^k 不是 (1.1) 的解时, 不等式 (1.6) 右端为正, 说明 (1.4) 中定义的 $d(u^k, \tilde{u}^k)$ 是未知函数 $\frac{1}{2}\|u - u^*\|^2$ 在 u^k 处欧氏模下的一个上升方向. 因此, 可以

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

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

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 $\alpha_k^* > \frac{1}{2}$. 接着就有

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

这是证明算法收敛的关键不等式. 分处不等式 (1.5) 两端的 $\beta_k F(\tilde{u}^k)$ 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 $\beta_k F(\tilde{u}^k)$ 和 (1.9) 中 给出的步长 α_k^*, 用

[投影校正] $u^{k+1} = P_\Omega[u^k - \alpha_k^* \beta_k F(\tilde{u}^k)]$ 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! ✠

我们以可分离凸优化 $$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\} \tag{2.1}$$

为例，它的拉格朗日函数可以写成 $L(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^\top(Ax+By-b)$.
拉格朗日函数 $L(x,y,\lambda)$ 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:

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

上面的变分不等式中，$\Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m$, $\theta(u)=\theta_1(x)+\theta_2(y)$, 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 (2.2) 这样的变分不等式问题, ADMM 类方法都属于一个 预测-校正的统一框架.

[预测]. 对给定的 $v^k(v=w$ 或是 w 的部分分量), 求得预测点 $\tilde{w}^k$, 使其满足

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

其中矩阵 Q 不一定对称, 但 $Q^\top+Q$ (本质上)正定. 实现预测, 只要能求解子问题

$$\min\{\theta_1(x)+r\|Ax-p\|^2|x\in\mathcal{X}\}\quad 和\quad \min\{\theta_2(y)+s\|By-q\|^2|y\in\mathcal{Y}\}. \tag{2.5}$$

剩下的事情就是将 (2.4) 中任意的 $w\in\Omega$ 选成 w^*, 得到 $(\tilde{v}^k-v^*)^\top Q(v^k-\tilde{v}^k)\ge 0$ 和

$$[H(v^k-v^*)]^\top[H^{-1}Q(v^k-\tilde{v}^k)]\ge(v^k-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k). \tag{2.6}$$

这里 H 为任意正定矩阵(可以是单位阵). 由 (2.6) 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, 据此用余弦公式求离解集更近的点.

[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:

$$v^{k+1}=v^k-\alpha_k^*M(v^k-\tilde{v}^k),\quad 其中\quad M=H^{-1}Q,\quad \alpha_k^*=\frac{(v^k-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k)}{\|M(v^k-\tilde{v}^k)\|_H^2}. \tag{2.7}$$

[固定步长的校正]. 对上面提到的矩阵 M, 也可以取固定步长的校正公式

$$v^{k+1}=v^k-\alpha M(v^k-\tilde{v}^k),\qquad 其中步长\ \alpha>0 \tag{2.8}$$

$$使得\quad G=Q^\top+Q-\alpha M^\top HM\succ 0,\quad (正定) \tag{2.9}$$

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

因此, 生成的序列 $\{v^k\}$ 具有收缩性质 $\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\alpha\|v^k-\tilde{v}^k\|_G^2$.

✠ 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, 一脉相承！✠