B-变分不等式投影收缩算法中的孪生方向和姊妹方法

奇妙的孪生方向和姊妹方法 maths.nju.edu.cn/~hebma 何炳生

1 孪生方向和相同步长的姊妹方法（VI的应用及PC算法建议参阅本人主页系列讲义前五讲）

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. 考虑求解单调变分不等式

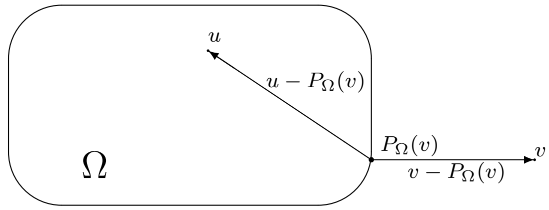
$$\text{(VI)} \qquad u^* \in \Omega, \quad (u-u^*)^T F(u^*) \ge 0, \quad \forall\, u \in \Omega. \tag{1.1}$$

如果上式中 $\Omega = \Re^n_+$ (非负卦限), 变分不等式 (1.1) 就是一个互补问题 (Complementarity Problem):

$$\text{(CP)} \qquad u^* \ge 0, \qquad F(u^*) \ge 0, \qquad (u^*)^T F(u^*) = 0. \tag{1.2}$$

换句话说, 互补问题是变分不等式中 $\Omega = \Re^n_+$ 的子类. 我们说一个变分不等式单调, 是指其中的算子 F 满足 $(u-v)^T(F(u)-F(v)) \ge 0$. 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 和常数 $\beta_k > 0$, 我们利用投影

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)] \tag{1.3}$$



生成一个预测点 $\tilde{u}^k$. 假如 $\tilde{u}^k = u^k$, u^k 就是 (1.1) 的解. 否则, 就要进行基于以下分析的新的迭代.

在生成预测点 $\tilde{u}^k$ 的投影 (1.3) 中, 要求参数 β_k 取得满足

$$\beta_k \|F(u^k) - F(\tilde{u}^k)\| \le \nu \|u^k - \tilde{u}^k\|, \quad \nu \in (0,1). \tag{1.4}$$

凸集上投影有如右上图所示的重要性质: 对任意的 $v \in \Re^n$, 用 $P_\Omega(v)$ 表示 v 到 Ω 上的投影, 有

$$(u - P_\Omega(v))^T (v - P_\Omega(v)) \le 0, \quad \forall u \in \Omega. \tag{1.5}$$

在 (1.5) 中令 $v = u^k - \beta_k F(u^k)$, 那么由 (1.3) 得到 $\tilde{u}^k = P_\Omega(v)$, 并从 (1.5) 得到

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^T \{[u^k - \beta_k F(u^k)] - \tilde{u}^k\} \le 0, \quad \forall u \in \Omega. \tag{1.6}$$

上式两边都加上 $(u-\tilde{u}^k)^T \beta_k F(\tilde{u}^k)$, 由此得到基于投影 (1.3) 的预测公式

$$\textbf{[预测]} \qquad \tilde{u}^k \in \Omega, \ \ (u-\tilde{u}^k)^T d(u^k, \tilde{u}^k) \le (u - \tilde{u}^k)^T \beta_k F(\tilde{u}^k), \ \ \forall u \in \Omega. \tag{1.7}$$

$$\text{其中} \qquad d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k [F(u^k) - F(\tilde{u}^k)]. \tag{1.8}$$

定义 1 (孪生方向) 在基于投影 (1.3) 得到的预测变分不等式 (1.7) 中, 分处两端的

$$d(u^k, \tilde{u}^k) \qquad \text{和} \qquad \beta_k F(\tilde{u}^k), \qquad \text{称为一对孪生方向.} \tag{1.9}$$

将 (1.7) 中属于 Ω 的 u 选成任意的解点 u^*, 利用 $F(u)$ 的单调性和 (1.1), 有

$$(\tilde{u}^k - u^*)^T d(u^k, \tilde{u}^k) \ge (\tilde{u}^k - u^*)^T \beta_k F(\tilde{u}^k) \ge (\tilde{u}^k - u^*)^T \beta_k F(u^*) \ge 0.$$

因此, 进而有

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \ge (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{1.10}$$

定义 2 (姊妹方法) 采用由预测产生的 (1.9) 中一对孪生方向 $d(u^k, \tilde{u}^k)$ 和 $\beta_k F(\tilde{u}^k)$ 分别进行校正

$$u_I^{k+1}(\alpha) = u^k - \alpha d(u^k, \tilde{u}^k), \tag{1.11}$$

和

$$u_{II}^{k+1}(\alpha) = P_\Omega[u^k - \alpha \beta_k F(\tilde{u}^k)], \tag{1.12}$$

的方法称为一对姊妹方法. 其中 α 是相同的待定步长.

对任意给定的解点 $u^* \in \Omega^*$, 我们记

$$\vartheta_k(\alpha) = \|u^k - u^*\|^2 - \|u_I^{k+1}(\alpha) - u^*\|^2 \tag{1.13}$$

和

$$\zeta_k(\alpha) = \|u^k - u^*\|^2 - \|u_{II}^{k+1}(\alpha) - u^*\|^2. \tag{1.14}$$

它们是依赖于步长 α 的 k 次迭代的收益, 是步长 α 的函数.

我们不能直接极大化 $\vartheta_k(\alpha)$ 和 $\zeta_k(\alpha)$, 因为它含有未知的 u^*. 下面将证明二次函数

$$q_k(\alpha) = 2\alpha (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2 \tag{1.15}$$

是它们的下界. 为了保证 $q_k(\alpha) > 0$, 我们在生成预测点 $\tilde{u}^k$ 的投影 (1.3) 中, 要求参数 β_k 使得 (1.4) 成立. 在这篇注记的 §2, 我们将会证明, 对同样的 α, $\zeta_k(\alpha)$ 优于 $\vartheta_k(\alpha)$.

2 姊妹方法的收缩性质 (何炳生 2016 高校计算数学学报 38 卷第 1 期)

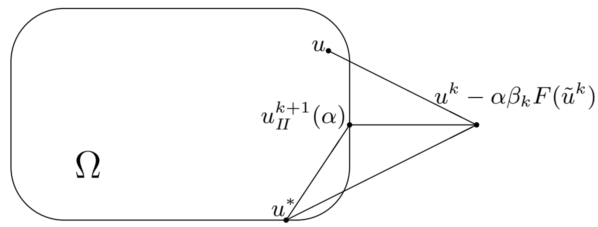
定理 设 $u_I^{k+1}(\alpha)$ 和 $u_{II}^{k+1}(\alpha)$ 由姊妹方法生成. 对分别由 (1.13) 和 (1.14) 定义的 $\vartheta_k(\alpha)$ 和 $\zeta_k(\alpha)$, 有

$$\vartheta_k(\alpha) \ge q_k(\alpha) \quad 和 \quad \zeta_k(\alpha) \ge q_k(\alpha) + \|u_I^{k+1}(\alpha) - u_{II}^{k+1}(\alpha)\|^2, \tag{2.1}$$

其中 $q_k(\alpha)$ 由 (1.15) 给出.

证明. 根据 (1.13) 中对 $\vartheta_k(\alpha)$ 的定义, 利用 (1.10) 我们有

$$\begin{aligned} \vartheta_k(\alpha) &= \|u^k - u^*\|^2 - \|u_I^{k+1}(\alpha) - u^*\|^2 \\ &= \|u^k - u^*\|^2 - \|u^k - u^* - \alpha d(u^k, \tilde{u}^k)\|^2 \\ &= 2\alpha(u^k - u^*)^T d(u^k, \tilde{u}^k) - \alpha^2\|d(u^k, \tilde{u}^k)\|^2 \\ &\ge 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \alpha^2\|d(u^k, \tilde{u}^k)\|^2 \\ &= q_k(\alpha). \end{aligned} \tag{2.2}$$



我们证明了定理的第一部分. 对定理的第二部分, 因为 $u_{II}^{k+1}(\alpha) = P_\Omega[u^k - \alpha\beta_k F(\tilde{u}^k)]$ 和 $u^* \in \Omega$, 根据投影的性质和余弦定理 (参考右上图), 有

$$\|u_{II}^{k+1}(\alpha) - u^*\|^2 \le \|u^k - \alpha\beta_k F(\tilde{u}^k) - u^*\|^2 - \|u^k - \alpha\beta_k F(\tilde{u}^k) - u_{II}^{k+1}(\alpha)\|^2. \tag{2.3}$$

因此, 利用 $\zeta_k(\alpha)$ 的定义(见 (1.14)), 我们有

$$\begin{aligned} \zeta_k(\alpha) &\ge \|u^k - u^*\|^2 - \|(u^k - u^*) - \alpha\beta_k F(\tilde{u}^k)\|^2 + \|(u^k - u_{II}^{k+1}(\alpha)) - \alpha\beta_k F(\tilde{u}^k)\|^2 \\ &= 2\alpha(u^k - u^*)^T\beta_k F(\tilde{u}^k) + 2\alpha(u_{II}^{k+1}(\alpha) - u^k)^T\beta_k F(\tilde{u}^k) + \|u^k - u_{II}^{k+1}(\alpha)\|^2 \\ &= \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - u^*)^T\beta_k F(\tilde{u}^k). \end{aligned} \tag{2.4}$$

将 (2.4) 中右端的最后一项 $(u_{II}^{k+1}(\alpha) - u^*)^T\beta_k F(\tilde{u}^k)$ 分拆成

$$(u_{II}^{k+1}(\alpha) - u^*)^T\beta_k F(\tilde{u}^k) = (u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T\beta_k F(\tilde{u}^k) + (\tilde{u}^k - u^*)^T\beta_k F(\tilde{u}^k).$$

利用单调性, $(\tilde{u}^k - u^*)^T\beta_k F(\tilde{u}^k) \ge (\tilde{u}^k - u^*)^T\beta_k F(u^*)$, 上式右端第二部分非负. 代入 (2.4), 进而得到

$$\zeta_k(\alpha) \ge \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T\beta_k F(\tilde{u}^k). \tag{2.5}$$

因为 $u_{II}^{k+1}(\alpha) \in \Omega$, 用它替代 (1.7) 中的任意 $u \in \Omega$, 得到

$$(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T\beta_k F(\tilde{u}^k) \ge (u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{2.6}$$

将 (2.6) 代入 (2.5) 的右端, 就有

$$\zeta_k(\alpha) \ge \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{2.7}$$

对上式右端, 利用 $q_k(\alpha)$ 的形式 (见(1.15)), 就化成

$$\begin{aligned} \zeta_k(\alpha) &\ge \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\ &= \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - u^k)^T d(u^k, \tilde{u}^k) + 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\ &= \|(u^k - u_{II}^{k+1}(\alpha)) - \alpha d(u^k, \tilde{u}^k)\|^2 - \alpha^2\|d(u^k, \tilde{u}^k)\|^2 + 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\ &= \|(u^k - u_{II}^{k+1}(\alpha)) - \alpha d(u^k, \tilde{u}^k)\|^2 + q_k(\alpha). \end{aligned}$$

利用 (1.11), 上式右端的第一部分就是 $\|u_I^{k+1}(\alpha) - u_{II}^{k+1}(\alpha)\|^2$. 这样就完成了结论 (2.1) 的证明. □

这个定理说明, $q_k(\alpha)$ 是 $\vartheta_k(\alpha)$ 和 $\zeta_k(\alpha)$ 的下界. 对同样的 α, $\zeta_k(\alpha)$ 优于 $\vartheta_k(\alpha)$. 由于 $q_k(\alpha)$ 在

$$\alpha_k^* = \arg\max\{q_k(\alpha)\} = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2} \tag{2.8}$$

处取得最大值(这也是短篇 **A** § **1** 中步长) α_k^* 的取法). 在实际计算中, 我们采用校正公式

$$(\text{收缩算法–1}) \quad u^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k) \tag{2.9}$$

或者

$$(\text{收缩算法–2}) \quad u^{k+1} = P_\Omega[u^k - \gamma\alpha_k^*\beta_k F(\tilde{u}^k)] \tag{2.10}$$

产生新的迭代点 u^{k+1}, 其中的 α_k^* 都由 (2.8) 给出, 松弛因子 $\gamma \in (0, 2)$, 我们往往取 $\gamma \in [1.2, 1.8]$.

采用校正公式 (2.9), 它的好处是生成 u^{k+1} 不用再做投影. 实际问题中, 到 Ω 上的投影代价往往不高 (例如 Ω 常常是一个正卦限或者框形), 因此常采用校正公式 (2.10). 这方面的理由在我主页系列讲义的前三讲中有更详细的说明. 相关的收敛速率的结论和证明可以参考第五讲.