求函数极小点的记忆梯度法

蔡 火 萤

（数学系）

记忆梯度法是共轭梯度法的推广和改进，它存在很多优点，是一个值得重视的算法。不过在它的每一迭代步中，都要作一次二维搜索。以往处理这个问题，常常采用牛顿法[1]，由于牛顿法对初值要求十分苛刻[3]，在实用上很不理想。本文提出一类函数的极值问题，在使用记忆梯度法求解时，采用线性化方法处理二维搜索问题，获得较好的计算效果。

假设 $f: D\subset R^n \longrightarrow R^1$ 在凸集合 $D_0\subset D$ 上二次连续可微，并且它的 Hessian 矩阵 $H(X)$ 在 D_0 上的全局极小点 X^*，即 $f(X^*)=\min\limits_{X\in D_0} f(X)$。

今于 D_0 上适当取出一个初始逼似点 $X_0\in D_0$，并按最速下降法求出它的第一个近似点 X_1。即：

$$X_1 = X_0 + \omega_0 p_0$$
$$p_0 = -\nabla f(X_0).$$
$$\omega_0 : \min_{\omega} f(X_0+\omega p_0)$$

然后再按下列方法求出它的第二个近似点 X_2。

$$X_2 = X_1 - \omega_0^{(1)}\nabla f(X_1) + \omega_1^{(1)}\triangle X_0$$
$$\triangle X_0 = X_1 - X_0$$
$$\omega_0^{(1)},\ \omega_1^{(1)} : \min_{\omega_0,\omega_1} f[X_1 - \omega_0\nabla f(X_1) + \omega_1\triangle X_0]$$

一般地当 X_1，X_2，…，X_K 被求出之后（$K\geqslant 1$），便可以按如下方式求出：

$$\left.\begin{aligned}&X_{K+1} = X_K - \omega_0^{(K)}\nabla f(X_K) + w_1^{(K)}\triangle X_{K-1}\\&\triangle X_{K-1} = X_K - X_{K-1}\\&\omega_0^{(K)},\ \omega_1^{(K)} : \min_{\omega_0,\omega_1} f(X_K) - w_0\nabla f(X_K) + w_1\triangle X_{K-1}\end{aligned}\right\} \quad (1)$$

在我们的假设下关系式(1)等价于方程组[1]

$$\left.\begin{aligned}&\nabla f(X_{K+1})^T\nabla f(X_K) = 0\\&\nabla f(X_{K+1})^T\triangle X_{K-1} = 0\end{aligned}\right\} \quad (2)$$

进而推知：

$$\nabla f(X_{K+1})^T \triangle X_K = 0$$

在 $\nabla f(X_{K+1}) \neq 0$ 的情况下，$\nabla f(X_{K+1})$ 与 $\triangle X_K$ 线性无关。

下面我们采用线性化的方法求解(2)式。因为：

$$\begin{aligned}\nabla f(X_{K+1}) &= \nabla f(X_K + \triangle X_K) \\ &= \nabla f(X_K) + H(X_K)\triangle X_K + \cdots \\ &\approx \triangle f(X_K) + H(X_K)\left[-\omega_0^{(K)}\nabla f(X_K) + \omega_1^{(K)}\triangle X_{K-1}\right] \\ &= \nabla f(X_K) - \omega_0^{(K)} H(X_K)\nabla f(X_K) + \omega_1^{(K)} H(X_K)\triangle X_{K-1}\end{aligned}$$

所以我们得到近似线性方程组：

$$\left.\begin{aligned}&\triangle f(X_K)^T\nabla f(X_K) - \omega_0^{(K)}\nabla f(X_K)^T H(X_K)\nabla f(X_K) \\ &\qquad + \omega_1^{(K)}\nabla f(X_K)^T H(X_K)\triangle X_{K-1} = 0 \\ &\triangle X_{K-1}^T\nabla f(X_K) - \omega_0^{(K)}\triangle X_{K-1}^T H(X_K)\nabla f(X_K) \\ &\qquad + \omega_1^{(K)}\triangle X_{K-1}^T H(X_K)\triangle X_{K-1} = 0\end{aligned}\right\} \tag{3}$$

因为 $\nabla f(X_K)$ 与 $\triangle X_{K-1}$ 线性无关，所以(3)式的系数行列式为 $-\triangle^{(K)}$，其中：

$$\begin{aligned}\triangle^{(K)} &= \left[\nabla f(X_K)^T H(X_K)\nabla f(X_K)\right]\left[\triangle X_{K-1}^T H(X_K)\triangle X_{K-1}\right] \\ &\quad - \left[\nabla f(X_K)^T H(X_K)\triangle X_{K-1}\right]^2 > 0\end{aligned}$$ [2]

因此(3)式有唯一解：

$$\omega_0^{(K)} = \frac{1}{\triangle^{(K)}}\|\nabla f(X_K)\|_2^2\left[\triangle X_{K-1}^T H(X_K)\triangle X_{K-1}\right]$$

$$\omega_1^{(K)} = \frac{1}{\triangle^{(K)}}\|\triangle f(X_K)\|_2^2\left[\nabla f(X_K)^T H(X_K)\triangle X_{K-1}\right]$$

下面我们要证明的这样求出的二维搜索因子，保证目标函数下降，事实上因为：

$$f(X_{K+1}) - f(X_K) = \nabla f(X_K)^T\triangle X_K + \cdots$$

略去高价微量，得到：

$$\begin{aligned}f(X_{K+1}) - f(X_K) &= \nabla f(X_K)^T\left[-\omega_0^{(K)}\nabla f(X_K) + \omega_1^{(K)}\triangle X_{K-1}\right] \\ &= -\omega_0^{(K)}\|\nabla f(X_K)\|_2^2 < 0\end{aligned}$$

$$\therefore\ f(X_{K+1}) < f(X_K)$$

将上述算法归结成如下计算步骤：

1. 给出初值 X_0 及允许误差 $\varepsilon > 0$。
2. 计算：

$$P_0 = -\nabla f(X_0)$$

$$\omega_0:\ \min_{\omega} f(X_0 + \omega P_0)$$

$$X_1 = X_0 + w_0 P_0$$

3. 假设已经求出X_1，$X_2\cdots$，X_K （$K\geqslant 1$）

计算：

$$X_{K+1}=X_K-\omega_0^{(K)}\nabla f(X_K)+\omega_1^{(K)}\triangle X_{K-1}$$

$$\triangle X_{K-1}=X_K-X_{K-1}$$

$$\omega_0^{(K)}=\frac{1}{\triangle^{(K)}}\|\nabla f(X_K)\|_2^2\left[\triangle X_{K-1}^T H(X_K)\triangle X_{K-1}\right]$$

$$\omega_1^{(K)}=\frac{1}{\triangle^{(K)}}\|\nabla f(X_K)\|_2^2\left[\nabla f(X_K)^T H(X_K)\triangle X_{K-1}\right]$$

$$\triangle^{(K)}=\left[\nabla f(X_K)^T H(X_K)\nabla f(X_K)\right]\left[\triangle X_{K-1}{}^T H(X_K)\triangle X_{K-1}\right]-\left[\nabla f(X_K)^T H(X_K)\triangle X_{K-1}\right]^2$$

4. 检查精度及判别转移。

如果$\|\nabla f(X_{K+1})\|_2\leqslant\varepsilon$，则停机，并以$X_{K+1}$作为$X^*$满足精度要求的近似值。否则，如果$K+1<n$，则将$X_{K+1}$代替$X_K$，并且转去执行3。

如果$K+1=n$，则用X_{K+1}代替X_0实行再开始过程，并且转去执行2。

算例：将这个算法用于求目标函数为：

$$f(x_1,x_2,x_3,x_4)=100(x_1^2-x_2)^2+(x_1-1)^2+(x_3-1)^2+90(x_3^2-x_4^2+10.1\left[(x_2-1)^2+(x_4-1)^2\right]+19.8(x_2-1)(x_4-1)$$

的极小问题时，从任意点X_0出发，利用最速下降法求前K个近似点，使$\|\nabla f(X_K)\|_2\leqslant 1.0$，然后再从$X_K$出发，按上述算法进行迭代只须40次左右便可达到$\|\nabla f(X_{K+40})\|_2\leqslant 10^{-4}$。而在同样情况下，采用F—R共轭梯度法，则须迭代100次左右，才能达到同样效果。我们这样处理二维搜索问题，不仅保证目标函数严格下降，同时易证迭代收敛，并且编制程序简单，实用方便。

参 考 文 献

〔1〕王德人，非线性方程组解法与最优化方法，人民教育出版社，（1982）。

〔2〕北京大学编，高等代数，人民教育出版社，（1978）。

〔3〕Б. П. Демидович и И. А. Марон, Основы вычислительной математики, Государственное издательство физико-математической литературы М.，(1963).