DOI: 10.13656/j.cnki.gxkx.2004.03.011
广西科学 Guangxi Sciences 2004, 11 (3): 197~ 200

一个修改的 BFGS方法
A Modified BFGS Method

朱志伟
Zhu Zhiwei

(广西广播电视大学梧州分校 梧州市步埠路 70号 543002)
(Wuzhou Branch, Guangxi Radio and TV Univ., 70 Bubulu, Wuzhou, Guangxi, 543002, China)

摘要 给出一类新的 BFGS校正公式,讨论其矩阵的正定性、二次终止性和方向共轭性,并在适当条件下建立该方法的全局收敛性.
关键词 BFGS方法 正定性二次终止性 共轭性 全局收敛性
中图法分类号 O241.7

Abstract The new BFGS method given in this paper not only posesses the properties contained in the normal BFGS method, but overcomes the disadvantage of the old one because the latter cannot ensure the positive definite of update matrix. Our aim is to discuss the properties of the proposed method and establish it's global convergence under suitable conditions.
Key words BFGS method, quadratic terminate property, conjugate property, global convergence

本文考虑如下的无约束最优化问题

$$\min f(x), x \in R^n, \tag{0.1}$$

其中 $f(x)$ 是连续可微函数. 众所周知求解问题(0.1)有效方法之一是拟牛顿方法,其中非常重要的一步就是近似海色矩阵的校正.由于 BFGS校正公式的良好数值表现,许多学者都对 BFGS方法做了相应的研究工作[1~4]. 但 BFGS方法不能保证校正公式的正定性,所以近年来,文献 [5~8] 给出了一些修改的 BFGS方法来解问题 (0.1). 本文同样给出一个修改的 BFGS方法,并讨论其矩阵保持正定性、二次终止性和方向共轭性,且在适当的条件下建立此方法的全局收敛性.本文给出的修改 BFGS方法不但拥有通常 BFGS方法所具有的性质,而且还克服了通常 BFGS方法不能保证校正矩阵正定性的性质.

1 一个修改的 BFGS方法及性质

一般拟牛顿法迭代的主要步骤为:

步骤 1: 令 $d_k = -H_k g_k$;

步骤 2: 沿方向 d_k 作线性搜索,得到 $x_{k+1} = x_k + \lambda_k d_k$;

步骤 3: 校正 H_k 产生 H_{k+1}.

其中 d_k 是第 k 次的搜索方向; λ_k 是第 k 次沿方向 d_k 得到的步长; x_k 是第 k 次的迭代值; H_k 是 f 在 x_k 处的 Hesse逆矩阵或其近似逆矩阵; g_k 是 $f(x)$ 在 x_k 处的梯度值.上式步骤中的关键是第三步,即如何产生校正公式.

BFGS校正公式

$$B_{k+1} = B_k + \frac{y_k y_k^T}{s_k^T y_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, \tag{1.1}$$

其中 B_k 是 f 在 x_k 处的 Hesse矩阵或其近似矩阵; $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$, g_k 和 g_{k+1} 分别是 $f(x)$ 在 x_k 和 x_{k+1} 处的梯度值, x_k 和 x_{k+1} 分别是在第 k 次和第 $k+1$ 次处的迭代值.

本文用 y_k^* 来代替(1.1)式中的 y_k,相应地可得到一个修改的 BFGS校正公式

$$B_{k+1} = B_k + \frac{y_k^* y_k^{*T}}{s_k^T y_k^*} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, \tag{1.2}$$

其中 $y_k^* = \frac{y_k^T s_k}{|y_k^T s_k|} y_k$; s_k 和 y_k 与(1.1)式中的 s_k 和 y_k 分别相同. 本文就是研究公式(1.2)的性质.

本文采用下列记号: $\|\cdot\|$ 是 R^1 中的 Euclid范数, $g(x) \in R^n$ 是 f 在 x 处的梯度值,记 $f_k := f(x_k)$, $g_k := \nabla f(x_k)$, $B_k \approx \nabla^2 f(x_k)$, $\{x_k\}$ 是修改的 BFGS方法(1.2)迭代产生的迭代点列, $\{f(x_k)\}$ 是 $\{x_k\}$ 对应的函数列, λ_k 是相应的步长因子, d_k 是相应的搜索方向.

对公式(1.2)两次应用逆的秩一校正的 Sherman-Morrison公式可得到关于 H_k 的 BFGS校

2003-11-24收稿, 2004-04-21修回。

正:

$$H_{k+1}=H_k+\frac{(s_k-H_k y_k^*)s_k^T+s_k(s_k-H_k y_k^*)^T}{s_k^T y_k^*}-\frac{(s_k-H_k y_k^*)^T y_k^*}{(s_k^T y_k^*)^2}s_k s_k^T=(I-\frac{s_k y_k^{*T}}{s_k^T y_k^*})H_k(I-\frac{y_k^* s_k^T}{s_k^T y_k^*})+\frac{s_k s_k^T}{s_k^T y_k^*}, \quad (1.3)$$

公式(1.2)与公式(1.3)对讨论问题是等价的.对(1.2)和(1.3),可得到其相应的拟牛顿方程

$$B_{k+1}s_k=y_k^* \quad (1.4)$$

和 $$H_{k+1}y_k^*=s_k. \quad (1.5)$$

公式(1.2)的性质有:

性质 **1** 当 B_k 正定时, B_{k+1} 也正定的充要条件是 $y_k^{*T}s_k>0$ 成立.当 B_k 正定时,由公式(1.2)得

$$y_k^{*T}s_k=\frac{y_k^T s_k}{|y_k^T s_k|}y_k^T s_k=|s_k^T y_k|=|s_k^T B_k s_k+o(\|s_k\|^2)|\geqslant|\lambda_k+o(1)|\|s_k\|^2>0,$$

其中 λ_k 是 B_k 的最小特征值.所以,公式(1.2)能够保持校正公式的正定性.

下面是公式(1.2)具有二次终止性的一个定理.该定理表明,对于二次函数,修改的 BFGS方法产生的方向是共轭的,方法在 n 步终止,即有 $H_n=G^{-1}$(G 是函数 f 的 Hesse矩阵).

定理 1.1 如果 f 是二次函数,G 是其正定 Hesse矩阵,那么,当采用精确线性搜索时,修改的 BFGS方法具有遗传性质和方向共轭性质,即是对于 $i=0,1,\cdots,m$,有遗传性质

$$H_{i+1}y_j^*=s_j, j=0,1,\cdots,i, \quad (1.6)$$

和方向共轭性质

$$s_i^T G s_j=0, j=0,1,\cdots,i-1, \quad (1.7)$$

方法在 $m+1\leqslant n$ 步迭代后终止.如果 $m=n-1$,则 $H_n=G^{-1}$.

证明 用归纳法证明(1.6)和(1.7).

显然,当 $i=0$ 时,结论成立.现在假定结论对于 $i\geqslant 0$ 成立,要证明结论对于 $i+1$ 也成立.若 $g_{i+1}=0$,则有 $s_{i+1}^T G s_j=0(j\leqslant i)$ 成立.当 $g_{i+1}\neq 0$,由精确一维搜索和归纳法的假设,对于 $j\leqslant i$,有

$$g_{i+1}^T s_j=(g_{i+1}^T s_j-g_i^T s_j)+(g_i^T s_j-g_{i-1}^T s_j)+,\cdots,+g_{j+1}^T s_j=g_{j+1}^T s_j+\sum_{k=j+1}^{i}(g_{k+1}-g_k)^T s_j=g_{j+1}^T s_j+\sum_{k=j+1}^{i}y_k^T s_j=0+\sum_{k=j+1}^{i}s_k G s_j=0, \quad (1.8)$$

第三和第四个等式分别根据精确一维搜索和归纳法的假设.所以利用归纳法假设(1.6),(1.8)和 $d_{i+1}^T G=-g_{i+1}^T$,得

$$s_{i+1}^T G s_j==-\alpha_{i+1}d_{i+1}^T G s_j=-\alpha_{i+1}g_{i+1}^T H_{i+1}y_j^*=-\alpha_{i+1}g_{i+1}^T s_j=0, \quad (1.9)$$

这就证明了对 $i+1$,(1.7)式成立.

假设(1.6)式成立,要证 $H_{i+2}y_j^*=s_j, j=0,1,\cdots,i+1$.由公式(1.3)可得

$$H_{i+2}y_{i+1}^*=s_{i+1}, \quad (1.10)$$

当 $j\leqslant i$,由(1.9)和(1.6),得

$$s_{i+1}^T y_j^*=s_{i+1}^T G s_j=0 \quad (1.11)$$

和 $$y_{i+1}^{*T}H_{i+1}y_j^*=y_{i+1}^{*T}s_j=s_{i+1}^T G s_j=0, \quad (1.12)$$

所以利用公式(1.2)得到

$$H_{i+2}y_j^*=H_{i+1}y_j^*+\frac{(s_{i+1}-H_{i+1}y_{i+1}^*)s_{i+1}^T+s_{i+1}(s_{i+1}-H_{i+1}y_{i+1}^*)^T}{s_{i+1}^T y_{i+1}^*}y_j^*-\frac{(s_{i+1}-H_{i+1}y_{i+1}^*)^T y_{i+1}^*}{(s_{i+1}^T y_{i+1}^*)^2}s_{i+1}s_{i+1}^T y_j^*=H_{i+1}y_j^*=s_j, \quad (1.13)$$

由(1.10)和(1.13)得证 $H_{i+2}y_j^*=s_j, j=0,1,\cdots,i+1$.从而式(1.6)成立.

由于 s_i 共轭,$i=0,1,\cdots,m$,故方法是共轭方向法.根据共轭方向法基本定理[10],对于二次函数,方法至多 n 步终止.即存在 $m\leqslant n-1$,在 m 步后终止.当 $m=n-1$ 时,由于 s_i 线性无关,$i=0,1,\cdots,m$,所以

$$H_n y_j^*=s_j, j=0,1,\cdots,n-1,$$

此即

$$H_n G s_j=s_j, j=0,1,\cdots,n-1,$$

从而就得到 $H_n=G^{-1}$.

2 全局收敛性分析

著名的非精确线性搜索弱 Wolfe-Powell步长准则

$$f(x_{k+1})\leqslant f(x_k)+\delta\alpha_k g_k^T d_k \quad (2.1)$$

和 $$g_{k+1}^T d_k\geqslant\sigma g_k^T d_k, \quad (2.2)$$

其中 $\delta\in(0,\frac{1}{2})$, $\sigma\in(\delta,1)$.当然所讨论的问题满足一般的寻找最优值的准则,即,函数满足下降性质.为了建立所给方法在准则(2.1)和(2.2)下的全局收敛性,这里需要一些假设条件.

假设(i): 水平集

$$L(x_0)=\{x\in R^n: f(x)\leqslant f(x_0)\},$$

包含于一有界闭凸集 D 内.

假设(ii): 函数 f 在 D 上是连续可微的,且存在一常数 L 满足下式

$$\|g(x)-g(y)\|\leqslant L\|x-y\|, \forall x,y\in D. \quad (2.3)$$

注 **1**: 由性质(1)的讨论可知对于 $x\in L(x_0)$,存

在常数 m, M 满足

$$m\|v\|^2\leqslant v^T\nabla^2 f(x)v\leqslant M\|v\|^2, \forall v\in R^n. \tag{2.4}$$

注 **2** 由上述假设和序列 $\{f(x_k)\}$ 的下降性,可知存在序列 $\{x_k\}$ 包含在 $L(x_0)$ 中,且存在常数 f^* 满足

$$\lim_{k\to\infty} f(x_k)=f^*, \tag{2.5}$$

而且,根据假设(ii)和序列 $\{x_k\}$ 的有界性可知,对所有的 k,存在常数 M' 满足

$$\|g(x_k)\|\leqslant M'. \tag{2.6}$$

引理 **2.1** 设 $f: R^n\to R^1$ 满足假设(i)和(ii),则有

$$\sum_{k=1}^{+\infty}(-g_k^Ts_k)<+\infty, \tag{2.7}$$

$$m_1\|s_k\|^2\leqslant s_k^Ty_k^*\leqslant m_2\|s_k\|^2, k=1,2,\cdots, \tag{2.8}$$

和 $$\|y_k^*\|\leqslant L\|s_k\|, \tag{2.9}$$

其中 m_1, m_2, L 是正常数.

证明 由(2.5),可得

$$\sum_{k=1}^{+\infty}(f(x_k)-f(x_{k+1})=\lim_{N\to+\infty}\sum_{k=1}^{N}(f(x_k)-f(x_{k+1}))=\lim_{N\to+\infty}(f(x_1)-f(x_{k+1}))=f(x_1)-f^*.$$

因此,

$$\sum_{k=1}^{+\infty}(f(x_k)-f(x_{k+1}))<+\infty,$$

将上式与

$$f(x_k+\alpha_kd_k)-f(x_k)\leqslant\delta\alpha_kg_k^Td_k$$

联立,则可得到

$$\sum_{k=1}^{+\infty}(-g_k^Ts_k)<+\infty$$

成立.从 $y_k^*=\frac{s_k^Ty_k}{|s_k^Ty_k|}y_k$ 的定义和(2.3)可得

$$\|y_k^*\|=\left\|\frac{s_k^Ty_k}{|s_k^Ty_k|}y_k\right\|=\|y\|=\|g(x_{k+1})-g(x_k)\|\leqslant L\|s_k\|.$$

再由

$$s_k^Ty_k^*=|s_k^Ty_k|=|s_k^Ty_k|=|s_k^T(g(x_{k+1})-g(x_k))|=s_k^T\nabla^2f(\xi)s_k,$$

其中 $\xi=x_k+\theta_1(x_{k+1}-x_k)$, $\theta_1\in(0,1)$,将上式与(2.4)联立便可得到(2.8).所以原命题得证.

引理 **2.2** 若假设条件(i)和(ii)满足,则有

$$\mathrm{Tr}(B_{k+1})\leqslant M_1(k+1) \tag{2.10}$$

和 $$\sum_{i=0}^{k}\frac{\|B_is_i\|^2}{s_i^TB_is_i}\leqslant M_1(k+1), \tag{2.11}$$

成立.

证明 根据引理2.1,将公式(1.2)两边同时取迹的运算得,

$$\mathrm{Tr}(B_{k+1})=\mathrm{Tr}(B_k)-\frac{\|B_ks_k\|^2}{s_k^TB_ks_k}+\frac{\|y_k^*\|^2}{s_k^Ty_k^*}\leqslant\mathrm{Tr}(B_k)-\frac{\|B_ks_k\|^2}{s_k^TB_ks_k}+\frac{L^2}{m_1}\leqslant\cdots\leqslant\mathrm{Tr}(B_0)-\sum_{i=0}^{k}\frac{\|B_is_i\|^2}{s_i^TB_is_i}+(k+1)\frac{L^2}{m_1},$$

利用 B_{k+1} 是正定的,有 $\mathrm{Tr}(B_{k+1})>0$,所以上式的最后一个不等式可导出(2.10)和(2.11)成立.

引理 **2.3** 若假设条件(i)和(ii)满足,则存在一个常数 c 满足下式

$$\prod_{i=0}^{k}\alpha_k\geqslant c^k. \tag{2.12}$$

证明 利用引理2.1,将(1.2)两边同时取行列式计算得

$$\mathrm{Det}(B_{k+1})=\mathrm{Det}\left(B_k\left(I-\frac{s_ks_k^TB_k}{s_k^TB_ks_k}+\frac{B_k^{-1}y_k^*y_k^{*T}}{s_k^Ty_k^*}\right)\right)=\mathrm{Det}(B_k)\mathrm{Det}\left(I-\frac{s_ks_k^TB_k}{s_k^TB_ks_k}+\frac{B_k^{-1}y_k^*y_k^{*T}}{s_k^Ty_k^*}\right)=$$

$$\mathrm{Det}(B_k)\left(\left(1-s_k^T\frac{B_ks_k}{s_k^TB_ks_k}\right)\left(1+(B_k^{-1}y_k^*)^T\frac{y_k^*}{y_k^{*T}s_k}\right)-\left(s_k^T\frac{y_k^*}{y_k^{*T}s_k}\right)\left(\frac{(B_ks_k)^T}{s_k^TB_ks_k}B_k^{-1}y_k^*\right)\right)=\mathrm{Det}(B_k)\frac{y_k^{*T}s_k}{s_k^TB_ks_k}=$$

$$\mathrm{Det}(B_k)\frac{|y_k^Ts_k|}{s_k^TB_ks_k}\geqslant\mathrm{Det}(B_k)\frac{-(1-\sigma)g_k^Ts_k}{-\alpha_kg_k^Ts_k}\geqslant\cdots\geqslant\prod_{i=0}^{k}\frac{1}{\alpha_i}(1-\sigma)\mathrm{Det}(B_0).$$

其中上式的第三个等式利用了[9]

$$\mathrm{Det}(I+u_1u_2^T+u_3u_4^T)=(1+u_1^Tu_2)(1+u_3^Tu_4)-(u_1^Tu_4)(u_2^Tu_3),$$

第一个不等式分别利用了(2.2)和 $B_ks_k=-\alpha_kg_k$.这里再利用算术平均与几何平均不等式,有

$$\mathrm{Det}(B_{k+1})\leqslant\left[\frac{1}{n}\mathrm{Tr}(B_{k+1})\right]^n$$

和(2.10)有

$$\prod_{i=0}^{k}\alpha_i\geqslant\frac{(1-\sigma)^{k+1}\mathrm{Det}(B_0)}{\mathrm{Det}(B_{k+1})}\geqslant\frac{(1-\sigma)^{k+1}\mathrm{Det}(B_0)}{\left[\frac{\mathrm{Tr}(B_{k+1})}{n}\right]^n}\geqslant\frac{(1-\sigma)^{k+1}\mathrm{Det}(B_0)}{\left[\frac{(k+1)M_1}{n}\right]^n}.$$

因此,对所有 k,(2.12)满足.

定理 **2.1** 若假设条件(i)和(ii)满足,则修改的BFGS方法产生的迭代点列收敛,即

$$\liminf_{k\to\infty}\|g_k\|=0. \tag{2.13}$$

证明 反证法 若原命题不成立,则对所有的 k,存在一常数 $\varepsilon>0$ 满足

$$\|g_k\|\geqslant\varepsilon.$$

因此

$$+\infty>\sum_{k=0}^{\infty}(-g_k^Ts_k)=\sum_{k=0}^{\infty}\left(\frac{s_k^TB_ks_k}{\alpha_k}\right)=$$

$$\sum_{k=0}^{\infty}\left(\frac{s_k^T B_k s_k \| g_k \|}{\| B_k s_k \|}\right)=\sum_{k=0}^{\infty}\left(T_k \| g_k \|^2 \frac{s_k^T B_k s_k}{\| B_k s_k \|^2}\right) \geqslant$$

$$X \sum_{k=0}^{\infty}\left(T_k \frac{s_k^T B_k s_k}{\| B_k s_k \|^2}\right).$$

所以对任意的 $l>0$ 存在常数 k_0 满足任意的正整数 p,

$$p\left[\prod_{k=k_0+1}^{k_0+p} T_k \frac{s_k^T B_k s_k}{\| B_k s_k \|^2}\right]^{\frac{1}{p}} \leqslant \sum_{k=k_0+1}^{k_0+p} \frac{s_k^T B_k s_k}{\| B_k s_k \|^2} \leqslant l,$$

上式左边的不等式利用了算术平均与几何平均不等式的关系.因此

$$\left(\prod_{k=k_0+1}^{k_0+p} T_k\right)^{\frac{1}{p}} \leqslant \frac{l}{p}\left(\prod_{k=k_0+1}^{k_0+p} \frac{\| B_k s_k \|^2}{s_k^T B_k s_k}\right)^{\frac{1}{p}} \leqslant$$

$$\frac{l}{p} \sum_{k=k_0+1}^{k_0+p} \frac{\| B_k s_k \|^2}{s_k^T B_k s_k} \leqslant \frac{l}{p} \sum_{k=0}^{k_0+p} \frac{\| B_k s_k \|^2}{s_k^T B_k s_k}$$

$$\leqslant \frac{l(k_0+p+1)}{p^2} M_1,$$

令 $p \rightarrow \infty$ 将产生一个矛盾,因为根据引理 3.3上式不等式的左边是大于一个正常数的.

所以假设不成立,原命题成立.

3 结束语

本文给出了一个修改的 BFGS方法,此方法让 y_k 乘上了 $y_k^T s_k$ 的符号,得到了与通常的 BFGS方法同样的性质,特别是保持正定性.可将此种方法推广至其它方面做进一步的研究,应会得到较好的结果.对于此方法的超线性收敛性见文献 [10]的第 5.5节类似的证明,这里不再论证.

参考文献

1 Dennis J E, Schnabel R B. Numerical methods for unconstrained optimization and nonlinear equations. Englewood Cliffs: Pretice-Hall Inc, 1983.

2 Y Dai. Convergence properties of the BFGS algorithm. SIAM Journal on Optimization, 2003, (13): 693~701.

3 Fletcher R. Practical methods of optimization. 2nd ed. Chichester: John Wiley & Sons, 1987.

4 Griewank A, Ph L Toint. Local convergence analysis for partitioned quasi-Newton updates. Numer Math, 1982, (39): 429~448.

5 Davidon W C. Variable metric methods for minimization. Argonne National Labs Report, ANL-5990. 2000.

6 Li D, Fukushima M. A modified BFGS method and its global convergence in nonconvex minimization. Journal of Computational and Applied Mathematics, 2001, (129): 15~35.

7 Powell M J D. A new algorithm for unconstrained optimization. In: Nonlinear Programming. Rosen J B, Mangasarian O L, Ritter K. eds. New York: Academic Press, 1970.

8 Wei Z, Qi L, Chen X. A SQP-typemethod and its application in stochastic programming. Journal of Optimization Theory and Applications, 2003, (116): 205~228.

9 Dennis J E JR, Moré J J. Quasi-Newton methods motivation and theory. SIAM Rev, 1977, (19): 46~89.

10 Yuan Y, Sun W. Theory and Methods of Optimization. Beijing: Science Press of China. 1999.

(责任编辑: 黎贞崇)

(上接第 196页 Continue from page 196)

$$\min p^*(s),$$
$$\text{s.t.} \| s \| \leqslant Z'$$

的一阶必要条件简化为 $g(x^*)=0$,二阶必要条件简化为 B^* 半正定.证毕.

参考文献

1 Shultz G A, Schnabel R B, Byrd R H. A family of trust-region-based algorithms for unconstrained miimization with strong global convergence properties. SIAM J Numer Anal, 1985, 22: 47~67.

2 Buleau J P, Vial J Ph. Curvilinear path and trust region in unconstrained optimization, a convergence analysis. Math Prog Study, 1987, 30: 82~101.

3 Byrd R H, Schnabel R B, Shultz G A. Approximate solution of the trust region problem by minimization over two-dimensional subspaces. Math Prog, 1988, 40: 247~263.

4 袁亚湘.信赖域方法的收敛性.计算数学, 1994, 16: 333~346.

5 Dennis J E, Schnabel R B. Numerical methods for unconstrained optimization and nonlinear equations. Englewood Cliffs: Pretice-Hall Inc, 1983.

6 Fletcher R. Practical meethods of optimization. 2nd ed. Chichester: John Wiley & Sons, 1987.

7 Dennis J E, Jr Moré J J. A characteization of superlinear convergence and its application to Quasi-Newton methods. Math Comp, 1974, 28: 1171~1190.

8 Broyden C G, Dennis J E, Moré J J. On the local and supelinear convergence of Quasi-Newton methods. J Inst Math Appl, 1973, 12: 223~246.

9 Byrd R, Nocedal J, Yuan Y. Global convergence of a class of Quasi-Newton methods on convex problems. SIAM Journal on Numerical Analysis, 1987, 24: 1171~1189.

10 Yuan Y, Sun W. Theory and Methods of Optimization. Beijing: Science Press of China, 1999.

(责任编辑: 黎贞崇)