广西科学 Guangxi Sciences 2006，13(4)：282～287，292

一类拟牛顿算法的收敛性*

Convergence Properties of a Class of Quasi-Newton Algorithm

韦增欣，谢品杰，顾能柱

WEI Zeng-xin，XIE Pin-jie，GU Neng-zhu

（广西大学数学与信息科学学院，广西南宁　530004）

（Department of Mathematics and Information Science，Guangxi University，Nanning，Guangxi，530004，China）

摘要：根据一类基于新拟牛顿方程 $B_{k+1}s_k=y_k^*$ 的修改 BFGS 类算法，采用广义 Wolfe 线搜索模型（GW 搜索模型）：$f(x_{k+1})\leqslant f(x_k)+\delta\alpha_k g_k^T d_k$ 和 $g(x_{k+1})^T d_k\geqslant \max\{\sigma,1-(\alpha_k\|d_k\|)^p\}g_k^T d_k$，其中 $0<\delta\leqslant\sigma<1$，$p\in(-\infty,1)$，得到一类修正的 BFGS 算法（MBFGS），证明了 MBFGS 算法的全局收敛性和超线性收敛性.数值试验结果表明 MBFGS 算法是有效的.

关键词：无约束优化　BFGS 算法　全局收敛性　超线性收敛性

中图法分类号：O242.23　　**文献标识码**：A　　**文章编号**：1005-9164(2006)04-0282-06

Abstract：In this paper，we present a modified BFGS method，which satisfies the quasi-Newton funtion proposed by Wei[11]. Under suitable conditions，we establish global convergence and superlinear convergence for our algorithm with the general Wolfe line search. The numerical results are also presented，which show that the proposed algorithm is efficient for unconstrained optimization problems.

Key words：unconstrained optimization，BFGS method，global convergence，superlinear convergence

求解无约束最优化问题：

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

其中 $f(x)$：$\mathscr{R}^n\to R$ 是连续可微的，拟牛顿算法是最有效的方法，其一般迭代形式为：

$$x_{k+1}=x_k+\alpha_k d_k,\tag{0.2}$$

其中 α_k 为步长，d_k 为搜索方向.$d_k=-H_k g_k$，$g_k=\nabla f(x_k)$，$H_k=B_k^{-1}$.B_k 是 $\nabla^2 f(x_k)$ 的一近似，且满足拟牛顿方程：

$$B_{k+1}s_k=y_k,\tag{0.3}$$

其中 $s_k=x_{k+1}-x_k$，$y_k=g_{k+1}-g_k$.拟牛顿方程(0.3)在拟牛顿算法中扮演着至关重要的角色，不同的 B_k 将构成不同的拟牛顿方法，其中最著名的有 Broyden 族及其特例 BFGS 算法和 DFP 算法.近 20 年来对拟牛顿算法的研究主要集中在两方面，一是对算法收敛性质的研究[1~7]，二是希望通过修正拟牛顿方法从而得到一种既有好的收敛性质，又在数值表现上优于现有拟牛顿算法的算法[8~11].文献[11]提出一类新的拟牛顿方程：

$$B_{k+1}s_k=y_k^*,\tag{0.4}$$

其中 $y_k^*=y_k+A_k s_k$，A_k 为一矩阵：

$$A_k=\frac{2[f(x_k)-f(x_{k+1})]+[g(x_{k+1})+g(x_k)]^T s_k}{\|s_k\|^2}I.\tag{0.5}$$

并且提出了一类修正 BFGS 校正：

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

在 Wolfe 线搜索下，结合适当的假设条件，文献[10]和[11]分别证明了修改 BFGS 算法具有全局收敛性和超线性收敛性，数值计算结果显示该修正 BFGS 方法明显优于 BFGS 算法.

本文采用文献[5]中的技巧，考虑一类基于新拟牛顿方程(0.4)的修改 BFGS 类算法，其中 B_k 由以下

收稿日期：2006-06-14

作者简介：韦增欣(1962-)，男，广西武鸣人，教授，博士，主要从事优化管理的教学和研究工作。

* 国家自然科学基金(No.10161002)和广西自然科学基金项目(No.0135004)资助。

迭代形式得到：

$$B_{k+1}=B_k-\delta_k\frac{B_ks_ks_k^TB_k}{s_k^TB_ks_k}+\gamma_k\frac{y_k^*y_k^{*T}}{y_k^{*T}s_k},\tag{0.6}$$

其中 δ_k 和 γ_k 为参数，按照下述方法来选取：

$$(\delta_k,\gamma_k)=\begin{cases}\left(\dfrac{s_k^TB_ks_k}{s_k^TB_ks_k+y_k^{*T}s_k},\dfrac{y_k^{*T}s_k}{s_k^TB_ks_k+y_k^{*T}s_k}\right), \\ \quad 若\dfrac{s_k^TB_ks_k}{s_k^TB_ks_k+y_k^{*T}s_k}\geqslant\tau_k; \\ (\tau_k,1),若\dfrac{s_k^TB_ks_k}{s_k^TB_ks_k+y_k^{*T}s_k}<\tau_k,\end{cases}\tag{0.7}$$

其中 $\tau_k\in(0,1)$. 由以上关于 δ_k 和 γ_k 选取方法，可以看到它们满足关系 $\tau_k\leqslant\delta_k\leqslant 1$ 及 $0<\gamma_k\leqslant 1$.

同时，本文采用广义 Wolfe 线搜索模型(GW 搜索模型)[6]：

$$f(x_{k+1})\leqslant f(x_k)+\delta\alpha_kg_k^Td_k,\tag{0.8}$$

$$g(x_{k+1})^Td_k\geqslant\max\{\sigma,1-(\alpha_k\|d_k\|)^p\}g_k^Td_k,\tag{0.9}$$

其中 $0<\delta\leqslant\sigma<1,p\in(-\infty,1)$.

如此，我们得到一类修正的 BFGS 算法，简记为 MBFGS 算法.

1 MBFGS 算法

步骤 1　给出初始点 $x_0\in R^n$ 和初始正定距阵 $B_0\in R^{n\times n},0\leqslant\varepsilon<1,k:=0$.

步骤 2　若 $\|g(x_k)\|\leqslant\varepsilon$ 则停止，否则计算 $B_kd_k+g_k=0$，求得搜索方向 d_k.

步骤 3　利用 GW 搜索模型求得步长 α_k，令 $x_{k+1}=x_k+\alpha_kd_k$.

步骤 4　分别利用公式(0.5)和(0.6)计算 A_k 和 B_{k+1}.

步骤 5　令 $k:=k+1$，转步骤 2.

2 MBFGS 算法的全局收敛性分析

取(0.7)式中的 $\tau_k=\tau\in(0,1)$，并作出以下假设：

假设 1　水平集 $\Omega=\{x|f(x)\leqslant f(x_0)\}$ 包含在有界凸集 D 内.

假设 2　f 在 D 上连续可导，并且存在常数 $L>0$，使得

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

假设 3　f 是一致凸的，即存在正数 m_1,m_2 使得 $m_1\|z\|^2\leqslant z^TG(x)z\leqslant m_2\|z\|^2$.

为方便起见，本文引入以下符号

$$\cos\theta_k=\frac{s_k^TB_ks_k}{\|s_k\|\|B_ks_k\|}=-\frac{g_k^Ts_k}{\|g_k\|\|s_k\|},q_k=\frac{s_k^TB_ks_k}{s_k^Ts_k},u_k=g_k^TH_kg_k.\tag{2.1}$$

下面我们先给出几个引理.

引理 2.1　设 $\{x_k\}$ 由 MBFGS 算法产生，若 B_k 为正定矩阵，则 B_{k+1} 亦为正定矩阵.

证明　首先，利用 Taylor 展开式，有

$$s_k^Ty_k^*=s_k^Ty_k+2[f(x_k)-f(x_{k+1})]+[g(x_{k+1})+g(x_k)]^Ts_k=2[f(x_k)-f(x_{k+1})]+2g(x_{k+1})^Ts_k=2[-g(x_{k+1})^Ts_k+\frac{1}{2}s_k^TG(x_k+\theta(x_{k+1}-x_k))s_k]+2g(x_{k+1})^Ts_k=s_k^TG(x_k+\theta(x_{k+1}-x_k))s_k,$$

其中 $\theta\in(0,1)$. 由假设 3，我们可得

$$m_1\|s_k\|^2\leqslant s_k^Ty_k^*\leqslant m_2\|s_k\|^2.\tag{2.2}$$

其次，对于任意的 $x\in\mathscr{R}^n$，有

$$x^TB_{k+1}x=x^TB_kx-\delta_k\frac{(x^TB_ks_k)^2}{s_k^TB_ks_k}+\gamma_k\frac{(x_k^Ty_k^*)^2}{y_k^{*T}s_k}\geqslant x^TB_kx-\delta_kx^TB_kx+\gamma_k\frac{(x_k^Ty_k^*)^2}{y_k^{*T}s_k}=(1-\delta_k)x^TB_kx+\gamma_k\frac{(x_k^Ty_k^*)^2}{y_k^{*T}s_k}.$$

其中第 2 步利用了柯西不等式

$$(s^TBx)^2\leqslant(s^TBs)(x^TBx).\tag{2.3}$$

再结合(2.2)式和 $\tau_k\leqslant\delta_k\leqslant 1$ 及 $0<\gamma_k\leqslant 1$ 即可知引理 2.1 成立.

引理 2.2[7]　若非负序列 $\{m_k\}$ 满足 $\prod_{i=0}^k m_i\geqslant c_0^k$，$k=1,2,3,\cdots$ 则 $\lim\sup\limits_k m_k>0$.

引理 2.3　设 $\{x_k\}$ 由 MBFGS 算法产生，则

$$\sum_{k=0}^{\infty}(-g_k^Ts_k)<+\infty.\tag{2.4}$$

证明　由假设 1 可知，$f(x)$ 在 Ω 内有下界，于是有 $\sum_{k=0}^{\infty}[f(x_k)-f(x_{k+1})]<+\infty$，结合 $f(x_{k+1})-f(x_k)\leqslant\delta_k\alpha_kg_k^Td_k$，引理 2.3 得证.

引理 2.4　设 $\{x_k\}$ 由 MBFGS 算法产生，则有

$$Det(B_{k+1})\geqslant Det(B_k)\frac{y_k^{*T}s_k}{s_k^TB_ks_k},\tag{2.5}$$

其中 $Det(B_k)$ 表示 B_k 的行列式.

证明　利用公式 $Det(I+u_1u_2^T+u_3u_4^T)=(1+u_1^Tu_2)(1+u_3^Tu_4)-(u_1^Tu_4)(u_2^Tu_3)$[12]，(0.7)式及柯西不等式(2.3)，有：

$$Det(B_{k+1})=Det(B_k)(1-\delta_k+(1-\delta_k)\gamma_k\frac{y_k^{*T}B_k^{-1}y_k^{*T}}{y_k^{*T}s_k}+\delta_k\gamma_k\frac{y_k^{*T}s_k}{s_k^TB_ks_k})\geqslant Det(B_k)(1-\delta_k+\gamma_k\frac{y_k^{*T}s_k}{s_k^TB_ks_k}).$$

若 $\dfrac{s_k^TB_ks_k}{s_k^TB_ks_k+y_k^{*T}s_k}<\tau$，则

$$1-\delta_k+\gamma_k\frac{y_k^{*T}s_k}{s_k^TB_ks_k}=1-\tau+\frac{y_k^{*T}s_k}{s_k^TB_ks_k}>\frac{y_k^{*T}s_k}{s_k^TB_ks_k},$$

否则，有

$$1-\delta_k+\gamma_k\frac{y_k^{*T}s_k}{s_k^TB_ks_k}=\gamma_k(1+\frac{y_k^{*T}s_k}{s_k^TB_ks_k})=\frac{y_k^{*T}s_k}{s_k^TB_ks_k+y_k^{*T}s_k}\frac{s_k^TB_ks_k+y_k^{*T}s_k}{s_k^TB_ks_k}=\frac{y_k^{*T}s_k}{s_k^TB_ks_k}.$$

故(2.5)式成立.

引理 2.5 设$\{x_k\}$由MBFGS算法产生，则存在M_1使得成立

$$Tr(B_{k+1})\leqslant Tr(B_1)+M_1k\leqslant(Tr(B_1)+M_1)k,\quad(2.6)$$

$$\sum_{i=1}^{k}\frac{\|B_is_i\|^2}{s_i^TB_is_i}\leqslant\frac{Tr(B_1)+M_1k}{\tau}\leqslant\frac{Tr(B_1)+M_1}{\tau}k.\quad(2.7)$$

证明 首先，根据y_k^*和假设2，假设3有

$$\|y_k^*\|\leqslant\|y_k\|+\frac{|2[f(x_k)-f(x_{k+1})]+[g(x_{k+1})+g(x_k)]|}{\|s_k\|^2}\leqslant 2\|y_k\|+\frac{|s_k^TG(x_k+\theta(x_{k+1}-x_k))s_k|}{\|s_k\|^2}\leqslant 2L\|s_k\|+m_2\|s_k\|=(2L+m_2)\|s_k\|.$$

再由(2.2)式，知存在$M_1=\frac{(2L+m_2)^2}{m_1}$，使得

$$\frac{\|y_k^*\|^2}{y_k^{*T}s_k}\leqslant M_1,\quad(2.8)$$

对(0.6)两边同时求迹，有

$$Tr(B_{k+1})=Tr(B_k)+\gamma_k\frac{\|y_k^*\|^2}{y_k^{*T}s_k}-\delta_k\frac{\|B_ks_k\|^2}{s_k^TB_ks_k}\leqslant Tr(B_k)+\frac{\|y_k^*\|^2}{y_k^{*T}s_k}-\tau\frac{\|B_ks_k\|^2}{s_k^TB_ks_k}\leqslant Tr(B_k)-\tau\frac{\|B_ks_k\|^2}{s_k^TB_ks_k}+M_1\leqslant\cdots\leqslant Tr(B_1)-\sum_{i=1}^{k}\tau\frac{\|B_is_i\|^2}{s_i^TB_is_i}+M_1k.$$

由于B_k保持正定性，引理2.5得证.

引理 2.6 设$\{x_k\}$由MBFGS算法产生，则存在$c_1>0$，使得

$$\prod_{i=1}^{k}\alpha_i\geqslant c_1^k.\quad(2.9)$$

证明 由GW搜索模型第2条不等式(0.9)可得

$$y_k^Ts_k=\alpha_k[g(x_k+\alpha_kd_k)-g(x_k)]^Td_k\geqslant[\max\{\sigma,1-\|s_k\|^p\}-1]g_k^Ts_k.=\min\{1-\sigma,\|s_k\|^p\}(-g_k^Ts_k).\quad(2.10)$$

再利用$s_k=-\alpha_kB_k^{-1}g_k$，有

$$\min\{1-\sigma,\|s_k\|^p\}s_k^TB_ks_k=-\min\{1-\sigma,\|s_k\|^p\}\alpha_ks_k^Tg_k\leqslant\alpha_ky_k^Ts_k=\alpha_ks_k^T[\int_0^1G(x_k+\tau s_k)d\tau]s_k.$$

由假设3可知

$$s_k^TB_ks_k\leqslant\frac{\alpha_k}{\min\{1-\sigma,\|s_k\|^p\}}m_2\|s_k\|^2.$$

结合(2.2)式即得

$$\frac{s_k^Ty_k^*}{s_k^TB_ks_k}\geqslant\frac{m_1\cdot\min\{1-\sigma,\|s_k\|^p\}}{m_2\alpha_k}.$$

即存在$m>0$，使得

$$\frac{s_k^Ty_k^*}{s_k^TB_ks_k}\geqslant\frac{m}{\alpha_k}.\quad(2.11)$$

再利用引理2.4，有

$$Det(B_{k+1})\geqslant Det(B_k)\frac{y_k^{*T}s_k}{s_k^TB_ks_k}\geqslant Det(B_k)\frac{m}{\alpha_k}\geqslant\cdots\geqslant Det(B_1)\prod_{i=1}^{k}\frac{m}{\alpha_i}.$$

再根据不等式$Det(B_{k+1})\leqslant[\frac{Tr(B_{k+1})}{n}]^n$[16]和(2.6)式将有

$$\prod_{i=1}^{k}\alpha_i\geqslant\frac{m^kDet(B_1)}{Det(B_{k+1})}\geqslant\frac{m^kDet(B^1)}{[\frac{Tr(B_{k+1})}{n}]^n}\geqslant\frac{m^kDet(B_1)}{[\frac{Tr(B_1)+M_1)}{n}]^n}.$$

所以(2.9)式成立.

定理 2.1 设$\{x_k\}$由MBFGS算法产生，则有$\lim_{k\to\infty}\inf\|g_k\|=0$.

证明 利用反证法来证明定理成立，否则存在常数$\varepsilon>0$，使得

$$\|g_k\|\geqslant\varepsilon>0,k=1,2,\cdots,\quad(2.12)$$

由(2.7)式可以得到$\sum_{i=1}^{k}\frac{\varepsilon^2}{u_i}\leqslant\sum_{i=1}^{k}\frac{\|g_i\|^2}{u_i}\leqslant\frac{Tr(B_1)+M_1}{\tau}k$. 即$\sum_{i=1}^{k}\frac{1}{u_i}\leqslant\frac{[Tr(B_1)+M_1]k}{\tau\varepsilon_0^2}=c_2k$，其中$c_2=\frac{[Tr(B_1)+M_1]}{\tau\varepsilon_0^2}$. 根据几何－算术均值公式，有$\prod_{i=1}^{k}\frac{1}{u_i}\leqslant[\frac{\sum_{i=1}^{k}\frac{1}{u_i}}{k}]^k\leqslant[\frac{c_2k}{k}]^k=c_2^k$，

令$c_3=1/c_2$，得

$$\prod_{i=1}^{k}u_i\geqslant(\frac{1}{c_2})^k=c_3^k.\quad(2.13)$$

将(2.9)式与(2.13)式相乘，有

$$\prod_{i=1}^{k}\alpha_iu_i\geqslant(c_1c_3)^k.\quad(2.14)$$

由引理2.2，可知$\lim\sup_{k\to\infty}(\alpha_ku_k)>0$，这与引理2.3矛盾. 故$\lim_{k\to\infty}\inf\|g_k\|=0$. 定理2.1得证. 即MBFGS算法具有全局收敛性.

3 MBFGS 算法的超线性收敛性分析

选取(0.7)式中的τ_k满足

$$\sum_{k=0}^{\infty}(-\ln\tau_k)<\infty. \quad (3.1)$$

实际上存在许多 τ_k 满足这个要求，比如说取 $\tau_k = \exp(-\frac{1}{k^2})$，同时做如下假设：

假设 4 $f: R^n \to R$ 为二阶连续可微函数.

假设 5 在 $f(x)$ 的局部最小点 x^* 满足 $G(x^*)$ 正定.

假设 6 存在 x^* 的一个领域 $N(x^*,\delta)$，使得 $G(x)$ 在该领域内 Holder 连续，即存在常数 $v\in(0,1)$ 和 M_2 使得成立：$\|G(x)-G(x^*)\|\leqslant M_2\|x-x^*\|^v, \forall x\in N(x^*,\delta)$.

由以上假设，可以推得，存在 $\lambda_3>0$，使得当 k 足够大的时候成立不等式：

$$\|g(x)\| = \|g(x)-g(x^*)\| \geqslant \lambda_3\|x-x^*\|, \quad (3.2)$$

和

$$d^TG(x)d\geqslant\lambda_3\|d\|^2. \quad (3.3)$$

引理 3.1[10] 设 $\{x_k\}$ 由 MBFGS 算法产生，则 $\lim_{k\to\infty}\|A_k\|=0$.

引理 3.2 设 $\{x_k\}$ 由 MBFGS 算法产生，则对任意的 $p\in(0,1)$，存在常数 $\beta>0$，使得对任一 $k>1$，在 $[1,k]$ 中至少有 $[pk]$ 个 i 满足

$$\cos\theta_i\geqslant\beta. \quad (3.4)$$

证明 先引入 ψ 函数[1]

$$\psi(A)=Tr(A)-\ln Det(A)=\sum_{i=1}^{n}(\lambda_i-\ln\lambda_i), \quad (3.5)$$

其中 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 为 A 的特征根.

根据 $Tr(B_{k+1})=Tr(B_k)+\gamma_k\frac{\|y_k^*\|^2}{y_k^{*T}s_k}-\delta_k\frac{\|B_ks_k\|^2}{s_k^TB_ks_k}\leqslant Tr(B_k)+\frac{\|y_k^*\|^2}{y_k^{*T}s_k}-\frac{\delta_kq_k}{\cos^2\theta_k}$，及引理 2.4，(2.2) 式和 (2.8) 式我们有

$$\psi(B_{k+1})\leqslant\psi(B_k)-\frac{\delta_kq_k}{\cos^2\theta_k}+\frac{\|y_k^*\|^2}{y_k^{*T}s_k}-\ln\frac{y_k^{*T}s_k}{s_k^Ts_k}+\ln q_k=\psi(B_k)+(\frac{\|y_k^*\|^2}{y_k^{*T}s_k}-1-\ln\frac{y_k^{*T}s_k}{s_k^Ts_k}-\ln\delta_k)+(\ln\cos^2\theta_k+1-\frac{\delta_kq_k}{\cos^2\theta_k}+\ln\frac{\delta_kq_k}{\cos^2\theta_k})\leqslant\psi(B_k)+(M_1-1-\ln m_1-\ln\tau_k)+(\ln\cos^2\theta_k+1-\frac{\delta_kq_k}{\cos^2\theta_k}+\ln\frac{\delta_kq_k}{\cos^2\theta_k})\leqslant\cdots\leqslant\psi(B_1)+(M_1-1-\ln m_1-\ln\tau_k)k+\sum_{i=1}^{k}(\ln\cos^2\theta_i+1-\frac{\delta_iq_i}{\cos^2\theta_i}+\ln\frac{\delta_iq_i}{\cos^2\theta_i}).$$

定义 $\eta_i=-\ln\cos^2\theta_i-[1-\frac{\delta_iq_i}{\cos^2\theta_i}+\ln\frac{\delta_iq_i}{\cos^2\theta_i}]$. 显然有 $\eta_i\geqslant0$，由 $\psi(B_{k+1})>0$ 可得 $\frac{1}{k}\sum_{i=1}^{k}\eta_i<\frac{\psi(B_1)}{k}+(M_1-1-\ln m-\ln\tau_k)$.

令 J_k 为 $\eta_1,\eta_2,\ldots,\eta_k$ 中 $[pk]$ 个最小值对应的下标集，$\eta_{m_k}=\max_{i\in J_k}\eta_i$，则 $\frac{1}{k}\sum_{i=1}^{k}\eta_i\geqslant\frac{1}{k}[\eta_{m_k}+(k-[pk])\eta_{m_k}]\geqslant(1-p)\eta_{m_k}$.

因此对于任意的 $i\in J_k$，存在 c_4 使得

$$\eta_i<\frac{\psi(B_1)+M_1-1-\ln m-\ln\tau_k}{1-p}\equiv c_4. \quad (3.6)$$

又由于 $1-\frac{\delta_iq_i}{\cos^2\theta_i}+\ln\frac{\delta_iq_i}{\cos^2\theta_i}\leqslant0$ 故 $c_4>\eta_i\geqslant-\ln\cos^2\theta_i, \forall i\in J_k$，即 $\cos\theta_i>\exp(-\frac{c_4}{2})\equiv\beta$ 成立.

引理 3.3 设 $\{x_k\}$ 由 MBFGS 算法产生，对于 $v\in(0,1)$ 则存在 $r\in[0,1)$，使得当 k 足够大时成立

$$f_{k+1}-f_*\leqslant r^k(f_1-f_*), \quad (3.7)$$

$$\sum_{k=1}^{\infty}\|x_{k+1}-x^*\|^v<\infty. \quad (3.8)$$

而且，若记 $\rho_k=\max\{\|x_k-x^*\|^v,\|x_{k+1}-x^*\|^v\}$，则亦有

$$\sum_{k=1}^{\infty}\rho_k<\infty. \quad (3.9)$$

和

$$\sum_{k=1}^{\infty}\|A_k\|<\infty. \quad (3.10)$$

证明 首先由 (0.9) 式和假设 3，有 $m_2\|s_k\|^2\geqslant y_k^Ts_k\geqslant\min\{1-\sigma,\|s_k\|^p\}(-g_k^Ts_k)=\min\{1-\sigma,\|s_k\|^p\}\|g_k\|\|s_k\|\cos\theta_k$，即

$$\|s_k\|\geqslant\frac{\min\{1-\sigma,\|s_k\|^p\}}{m_2}\|g_k\|\cos\theta_k. \quad (3.11)$$

另外，由 $g(x^*)=0$ 及 Taylor 展开式知，存在 $M_3>0$，使得 $f_k-f_*\leqslant M_3\|x_k-x^*\|^2$.

再结合 (3.2) 式和 (3.11) 式可知

$$g_k^Ts_k=-\|g_k\|\|s_k\|\cos\theta_k\leqslant-\frac{\min\{1-\sigma,\|s_k\|^p\}}{m_2}\|g_k\|^2\cos^2\theta_k\leqslant-\frac{\min\{1-\sigma,\|s_k\|^p\}}{m_2}\lambda_3^2\|x_k-x^*\|^2\cos^2\theta_k\leqslant-\frac{\min\{1-\sigma,\|s_k\|^p\}}{m_2}\lambda_3^2\cos^2\theta_k(f_k-f_*). \quad (3.12)$$

又因为 (0.8) 式，有

$$f_{k+1}-f_*\leqslant f_k-f_*+\delta g_k^Ts_k. \quad (3.13)$$

由 (3.12) 式和 (3.13) 式，可知道存在 c_5 使得 $f_{k+1}-f_*\leqslant(1-c_5\cos^2\theta_k)(f_k-f_*)$.

再根据引理 3.2,对于 $\forall\ i \in J_k$,有 $f_{k+1} - f_* \leqslant (1 - c_5\beta^2)(f_k - f_*) = r^{\frac{1}{p}}(f_k - f_*)$,其中 $r^{\frac{1}{p}} = (1 - c_5\beta^2)$,由于 f_k 是单调下降的,故 $1 - c_5\beta^2 \geqslant 0$. 因为 $J_k \cap [1,k]$ 至少有 $[pk]$ 个 i,所以成立 $f_{k+1} - f_* \leqslant r^k(f_1 - f_*)$.

另一方面,将 f_{k+1} 在 x^* 处 Taylor 展开,由(3.3)式知,当 k 充分大时有 $f_{k+1} - f_* \geqslant \frac{\lambda_3}{2}\| x_{k+1} - x^* \|^2$,因此,我们有

$$\| x_{k+1} - x^* \| \leqslant [\frac{2}{\lambda_3}(f_{k+1} - f_*)]^{\frac{1}{2}} \leqslant [\frac{2}{\lambda_3}(f_{k_1} + 1 - f_*)\gamma^k]^{\frac{1}{2}} = [\frac{2}{\lambda_3}(f_{k_1+1} - f_*)]^{\frac{1}{2}}\gamma^{\frac{k}{2}}.$$

故(3.8)式成立,再注意到 $\rho_k \leqslant \| x_{k+1} - x^* \|^v + \| x_k - x^* \|^v$,所以可直接由(3.8)式得到(3.9)式成立. 再由文献[8]中引理 3.2 证明过程和假设 6,有

$$\sum_{k=1}^{\infty} \| A_k \| \leqslant \sum_{k=1}^{\infty} \| G(\zeta_{1k} - G(\zeta_{2k}) \| \leqslant \sum_{k=1}^{\infty}(\| G(\zeta_{1k}) - G(x^*) \| + \| G(x^*) - G(\zeta_{2k}) \|) \leqslant \sum_{k=1}^{\infty} M \| \zeta_{1k} - x^* \|^v + \sum_{i=1}^{\infty} M \| \zeta_{2k} - x^* \|^v \leqslant 2M\sum_{k=1}^{\infty}\rho_k.$$

所以由(3.9)式知(3.10)式成立.

引理 3.4 设 $\{x_k\}$ 由 MBFGS 算法产生,则存在 ε_k 使得

$$\frac{\| y_k^* - G(x^*)s_k \|}{\| s_k \|} \leqslant \varepsilon_k, \tag{3.14}$$

而且 ε_k 满足

$$\sum_{i=1}^{\infty} \varepsilon_k < \infty. \tag{3.15}$$

证明 首先,有

$$\frac{\| y_k^* - G(x^*)s_k \|}{\| s_k \|} = \frac{\| y_k - G(x^*)s_k + A_k s_k \|}{\| s_k \|} \leqslant \frac{\| \int_0^1 G(x_k + \eta s_k)s_k d\eta - G(x^*)s_k \| + \| A_k s_k \|}{\| s_k \|} \leqslant \int_0^1 \| G(x_k + \eta s_k)s_k - G(x^*) \| d\eta + \| A_k \| \leqslant M_2\int_0^1 \| x_k - x^* + \eta s_k \|^v d\eta + \| A_k \| \leqslant M_2\int_0^1 (\eta \| x_{k+1} - x^* \| + (1-\eta) \| x_k - x^* \|)^v d\eta + \| A_k \| \leqslant M_2\rho_k + \| A_k \| \equiv \varepsilon_k.$$

再根据引理 3.1 和引理 3.3 知结论成立.

根据引理 3.4,再注意到对 τ_k 的选取满足(3.1)式,即存在常数 c_6 使得成立

$$\sum_{k=0}^{\infty}(-\ln\delta_k) \leqslant \sum_{k=0}^{\infty}(-\ln\tau_k) \leqslant c_6.$$

类似于文献[1]中定理 3.2 的证明,我们有结论:

引理 3.5 设 $\{x_k\}$ 由 MBFGS 算法产生,则

$$\lim_{k\to\infty}\frac{\| (B_k - G(x^*))s_k \|}{\| s_k \|} = 0. \tag{3.16}$$

而且序列 $\{ \| B_k \| \}$ 和 $\{ \| B_k^{-1} \| \}$ 均有界.

定理 3.1 设 $\{x_k\}$ 由 MBFGS 算法产生,则 x_k 超线性收敛于 x^*.

证明 由文献[3]及引理 3.5 可知,只需证明当 k 充分大的时候,有 $\alpha_k \equiv 1$ 即可. 首先,因序列 $\{ \| B_k \| \}$ 和 $\{ \| B_k^{-1} \| \}$ 均有界,故可得

$$\| d_k \| = \| B_k^{-1} g_k \| \leqslant \| B_k^{-1} \| \| g_k \| \to 0. \tag{3.17}$$

其次,一方面利用 Taylor 展开式和(3.16)式,有

$$f(x_k + d_k) - f(x_k) - \delta g_k^T d_k = (1-\delta)g_k^T d_k + \frac{1}{2}d_k^T G(\zeta_k)d_k = -(1-\delta)d_k^T B_k d_k + \frac{1}{2}d_k^T G(\zeta_k)d_k = -(\frac{1}{2} - \delta)d_k^T B_k d_k - \frac{1}{2}d_k^T(B_k - G(\zeta_k))d_k = -(\frac{1}{2} - \delta)d_k^T G(x^*)d_k + o(\| d_k \|^2).$$

其中 ζ_k 介于 x_k 与 $x_k + d_k$ 之间. 故当 k 充分大时有 $f(x_k + d_k) - f(x_k) - \delta g_k^T d_k \leqslant 0$,即 $\alpha_k = 1$ 满足 GW 搜索模型的第一个不等式(0.8).

另一方面,也有

$$g(x_k + d_k)^T d_k - \sigma g_k^T d_k = (g(x_k + d_k) - g_k)^T d_k + (1-\sigma)g_k^T d_k = d_k^T G(\zeta_k)d_k - (1-\sigma)d_k^T B_k d_k = d_k^T G(\zeta_k)d_k - (1-\sigma)d_k^T G(x_k)d_k + o(\| d_k \|^2) = \sigma d_k^T G(x_k)d_k + o(\| d_k \|^2),$$

其中 ζ_k 介于 x_k 与 $x_k + d_k$ 之间. 因此,当 k 充分大时有 $g(x_k + d_k)^T d_k \geqslant \sigma g_k^T d_k \geqslant \max\{\sigma, 1 - (\alpha_k \| d_k \|)^p\} g_k^T d_k$,

即 $\alpha_k = 1$ 满足 GW 搜索模型的第二个不等式(0.9). 从而当 k 充分大时,$\alpha_k \equiv 1$,因而结论得证. 即 MBFGS 算法具有超线性收敛性.

4 数值试验

为了考察本文提出的 MBFGS 算法的数值表现,我们用文献[13]中的算例进行测试,并且将其数值结果与 BFGS 方法进行了比较.

在编写计算程序时,我们采用文献[12]中的算法来搜索满足弱 Wolfe 线搜索条件的步长 α_k. 当 $|g_k^T d_{k-1} - g_{k-1}^T d_{k-1}| \leqslant 10^{-20}$ 时,用二分法来搜索 α_k,在其他情况下用二次插值法来搜索 α_k. 测试算法的终

止条件为：$\| g(x_k) \| \leqslant 10^{-6}$. 算法中的线搜索参数设置为：$\delta = 0.1, \sigma = 0.9$.

表 1 列出了计算结果，其中各列意义为 Problem 表示测试的算例名称，Dim 表示算例的维数，NI 表示算法迭代的次数，NF 表示函数值计算的次数，NG 表示函数梯度计算的次数，BFGS 表示采用弱 Wolfe 线搜索的 BFGS 方法，MBFGS 表示采用广义 Wolfe 线搜索的 MBFGS 方法. 表 1 的计算结果表明本文提出的 MBFGS 算法是有效的.

参考文献：

[1] BYRD R, NOCEDAL J. A tool for the analysis of quasi-Newton methods with application to unconstrained minimization[J]. SIAM Journal on Numerical Analysis, 1989(26): 727-739.

[2] BYRD R, NOCEDALAND J, YUAN Y. Global convergence of a class of quasi-Newton methods on convex problems [J]. SIAM Journal on Numerical Analysis, 1987(24): 1171-1189.

[3] DENNES J E, MORé J J. A characteization of a superlinear convergence and its application to quasi-Newton methods[J]. Math Comp, 1974(28): 549-560.

[4] DAI Y. Convergence properties of the BFGS algorithm [J]. SIAM Journal on Optimization, 2002(13): 693-701.

表 1 BFGS 与 MBFGS 计算结果

Table 1 The results for BFGS, MBFGS

Problem	Dim	NI/NF/NG			
		BFGS	MBFGS $\tau_k = 0.2, p = 1$	MBFGS $\tau_k = 0.5, p = 1/2$	MBFGS $\tau_k = \exp(-\frac{100}{k^2}), p = 1/2$
ROSE	2	34/54/35	41/59/42	38/55/39	38/66/39
FROTH	2	10/22/11	10/22/11	12/24/13	10/22/11
BADSCP	2	158/335/159	182/340/183	182/340/183	169/308/170
BADSCB	2	14/181/15	19/230/20	20/150/21	16/146/17
BEALE	2	15/24/16	15/25/16	17/27/18	17/27/18
JENSAM	2	12/37/13	16/26/17	15/39/16	20/30/21
HELIX	3	27/67/28	30/71/31	33/59/34	30/72/31
BARD	3	23/34/24	18/56/19	22/58/23	25/61/26
GAUSS	3	4/7/5	6/9/7	6/9/7	5/8/6
GULF	3	1/4/2	1/4/2	1/4/2	1/4/2
BOX	3	29/49/30	20/46/21	26/51/27	20/57/21
SING	4	29/52/30	40/65/41	36/60/37	34/60/35
WOOD	4	52/97/53	61/97/62	57/94/58	56/105/57
KOWOSB	4	28/32/29	27/47/28	31/48/32	24/53/25
OSB1	5	50/121/51	56/177/57	55/122/56	55/122/56
BIGGS	6	35/74/36	35/69/36	37/56/38	33/93/34
OSB2	11	53/80/54	57/107/58	57/95/58	57/94/58
WATSON	20	55/91/56	49/85/50	69/125/70	59/95/60
ROSEX	8	86/152/87	81/129/82	85/123/86	79/120/80
	50	246/636/247	240/518/241	248/529/249	256/610/257
	100	388/1141/389	363/907/364	390/1029/391	410/1124/411
SINGX	4	29/52/30	40/65/41	36/60/37	34/60/35
PEN1	2	173/294/174	225/302/226	199/260/200	175/253/176
PEN2	4	383/561/384	357/450/358	364/456/365	333/497/334
	50	312/914/313	345/865/346	347/858/348	318/881/319
VARDIM	2	6/14/7	5/13/6	5/13/6	5/13/6
	50	10/79/11	46/140/47	17/90/18	16/106/17
TRIG	3	12/27/13	14/16/15	12/27/13	14/16/15
	50	44/48/45	47/48/48	46/48/47	45/46/46
	100	48/51/49	50/51/51	50/52/51	49/50/50
BV	3	6/14/7	8/14/9	9/15/10	8/14/9
	10	18/39/19	21/36/22	22/38/23	22/49/23
IE	3	7/11/8	7/11/8	9/12/10	7/11/8
	50	12/15/13	11/16/12	12/16/13	11/16/12
	100	12/15/13	11/16/12	12/16/13	11/16/12
	200	12/15/13	11/16/12	12/16/13	11/16/12
	500	13/16/14	11/16/12	12/16/13	11/16/12
TRID	3	13/45/14	13/45/14	19/46/20	13/45/14
	50	63/340/64	68/258/69	68/285/69	66/309/67
	100	112/637/113	108/494/109	103/534/104	122/636/123
	200	216/1223/217	212/905/213	191/969/192	213/1184/214

（下转第 292 页 Continue on page 292）

References:

[1] DAI Y. Convergence properties of the BFGS algorithm [J]. SIAM Journal on Optimization, 2003, 3(13): 693-701.

[2] DENNIS J E, SCHNABEL R B. Numerical methods for unconstrained optimization and nonlinear equations[M]. NJ: Pretice-Hall Inc Englewood Cliffs, 1983.

[3] FLETCHER R. Practical methods of optimization[M]. 2nd ed. Chichester: John Wiley & Sons, 1987.

[4] GRIEWANK A, PH L. Toint, Local convergence analysis for partitioned quasi-Newton updates[J]. Number Math 1982, 39(3): 429-448.

[5] YUAN Y, SUN W. Theory and Methods of Optimization [M]. Beijing: Science Press of China, 1999.

[6] LI D, FUKUSHIMA M. A modified BFGS method and its global convergence in nonconvex minimization [J]. Journal of Computational and Applied Mathematics, 2001, 129(1): 15-35.

[7] POWELL M J D. A new algorithm for unconstrained optimation[M]//J B ROSEN, O L MANGASARIAN, K RITTER, eds. Nonlinear Programming. New York: Academic Press, 1970.

[8] WEI Z, QI L, CHEN X. An SQP-typemethod and its application in stochastic programming [J]. Journal of Optimization Theory and Applications, 2003, 116(1): 205-228.

[9] WEI Z, YU G, YUAN G, et al. The superlinear convergence of a modified BFGS-type method for unconstrained optimization [J]. Computational Optimization and Applications, 2004, 1(29): 315-332.

[10] LI D, FUKUSHIMA M. A globally and superlinearly convergent Gauss-Newton-based BFGS method for symmetric nonlinear equations [J]. SIAM Journal on Numerical Analysis, 1999, 37(1): 152-172.

[11] WEI Z, YUAN G, LIAN Z. An approximate Gauss-Newton-based BFGS method for solving symmetric nonlinear equations [J]. Guangxi Sciences, 2004, 11 (2): 91-99.

[12] YUAN G, LI X. An approximate Gauss-Newton-based BFGS method with descent directions for solving symmetric nonlinear equations [J]. OR Transactions, 2004, 8(4): 10-26.

[13] YUAN G, LU X. A nonmonotone Gauss-Newton-based BFGS method for solving symmetric nonlinear equations [J]. Journal of Lanzhou University : Natural Sciences, 2005, 41: 851-855.

[14] BYRD R, NOCEDAL J. A tool for the analysis of quasi-Newton methods with application to unconstrained minimization[J]. SIAM Journal on Numerical Analysis, 1989, 3(26): 727-739.

[15] MORé J J, GARBOW B S, HILLSTROME K E. Testing unconstrained optimization software[J]. ACM Trans Math Software, 1981, 7(1): 17-41.

(责任编辑:凌汉恩;英文编辑:蒋汉明)

(上接第287页 Continue from page 287)

[5] AIPING LIAO. Modifying BFGS method[J]. Operations Research Letters, 1997, 20: 171-177.

[6] LIU GUANGHUI, HAN JIYE, XU ZHONGLING. Global convergence of the variable metric algorithms with a generalized wolfe linesearch[J]. Journal of Mathematical Research and Exposition, 1995(4): 499-508.

[7] 韩继业,刘光辉. 无约束最优化线搜索一般模型及 BFGS 方法的整体收敛性[J]. 应用数学学报, 1995(1): 112-122.

[8] YUAN YAXIANG, RICHARD, BYRD H. Non-quasi-Newton unconstrained optimization, IMA[J]. Journal of Computational Mathematics, 1995(2): 95-107.

[9] LI D, FUKUSHIMA M. Amodified BFGS method and its global convergence in nonconvex minimization [J]. J Comput Appl Math, 2001(129): 15-35.

[10] WEI Z, YU G, YUAN G, et al. The superlinear convergence of a modified BFGS type method for unconstrained optimization. Computational Optimization and Applications, 2004(25): 315-332.

[11] WEI Z, YU G. Some recent progress in unconstrained nonlinear optimization: proceedings of the 2003's International Conference on Numerical Optimization and Numerical Linear Algebra [C]. YUAN Y, eds. Beijing/Newyork: Science Press, 2004: 110-141.

[12] 袁亚湘,孙文瑜. 最优化理论和方法[M]. 北京:科学出版社, 1997.

[13] MORé J J, GARBOW B S, HILLSTROME K E. Testing unconstrained optimization software [J]. ACM Trans Math Software, 1981(7): 17-41.

(责任编辑:韦廷宗)