中国农业大学学报　1998,3(6):19～23
Journal of China Agricultural University

一种新拟牛顿法的收敛性分析①

张海斌②　　王兆智③　周志坚
(北京工业大学)　(中国农业大学工程基础科学部)

摘　要　作为传统拟牛顿方程的改进,有人提出了新拟牛顿方程。本文证明了一个基于新牛顿方程的拟牛顿法的全局收敛性和局部超线性收敛性。

关键词　拟牛顿法;新拟牛顿方程;全局收敛性;超线性收敛性
中图分类号　S 221.2

Convergent Properties of a New Quasi-Newton Method

Zhang Haibin　　Wang Zhaozhi　Zhou Zhijian
(Beijing Polytechnic University)　(College of Applied Engineering Sciences,CAU)

Abstract　As an improvement of the traditional method for solving quasi-Newton equation, a new method was proposed. And its global and local superlinear convergences are proved.

Key words　quasi-Newton method; new quasi-Newton equation; global convergence; superlinear convergence

考虑无约束优化问题

$$\min f(x) \tag{1}$$

其中 $x\in \mathbf{R}^n, f:\mathbf{R}^n\to\mathbf{R}$ 是一个连续可微函数。

拟牛顿法是求解问题(1)的一种迭代法,它产生迭代点列$\{x_k\}$。传统的拟牛顿法是基于拟牛顿方程 $B_{k+1}s_k=y_k$,其中:$s_k=x_{k+1}-x_k$;$y_k=g_{k+1}-g_k$。

文献[1]提出了一类新的拟牛顿方程,随后文献[2]给出了这类新拟牛顿方程的等价形式。按照[2]中的表示方法,它可表为

$$\boldsymbol{B}_{k+1}s_k=\hat{y}_k \tag{2}$$

其中 $\hat{y}_k=y_k+\gamma_k u_k/s_k^{\mathrm{T}}u_k(s_k^{\mathrm{T}}u_k\neq 0)$,$\gamma_k=3g_{k+1}^{\mathrm{T}}s_k+3g_k^{\mathrm{T}}s_k+6(f_k-f_{k+1})$,$u_k$ 为向量参数。

本文讨论选取 $u_k=y_k$ 的情形,即研究如下算法。

算法　任取初始点 x_1 构造迭代点列$\{x_k\}$满足 $x_{k+1}=x_k+\lambda_k p_k$,其中:$p_k=-\boldsymbol{B}_k^{-1}g_k$;$\lambda_k$ 满足 Wolfe 原则。

给定 $\alpha\in(0,1/2)$,$\beta\in(\alpha,1)$。若 $f(x_k-\boldsymbol{B}_k^{-1}g_k)\leqslant f_k-\alpha g_k^{\mathrm{T}}\boldsymbol{B}_k^{-1}g_k$,$g(x_k-\boldsymbol{B}_k^{-1}g_k)^{\mathrm{T}}\boldsymbol{B}_k^{-1}g_k\leqslant \beta g_k^{\mathrm{T}}\boldsymbol{B}_k^{-1}g_k$ 同时成立,则取 $\lambda_k=1$,否则取 $\lambda_k>0$ 使满足 $f(x_k-\lambda_k\boldsymbol{B}_k^{-1}g_k)\leqslant f_k-\alpha\lambda_k g_k^{\mathrm{T}}\boldsymbol{B}_k^{-1}g_k$,$g(x_k-\lambda_k\boldsymbol{B}_k^{-1}g_k)^{\mathrm{T}}\boldsymbol{B}_k^{-1}g_k\leqslant\beta g_k^{\mathrm{T}}\boldsymbol{B}_k^{-1}g_k$,而 $\boldsymbol{B}_k$ 由限制 Broyden 类公式得到

收稿日期:1998-07-14
①国家自然科学基金资助项目

②张海斌,北京平乐园 100 号 北京工业大学应用数学系,100022
③王兆智,北京清华东路 17 号 中国农业大学(东校区)71 信箱,100083

$$B_{k+1}=B_k-\frac{B_ks_ks_k^{\mathrm{T}}B_k}{s_k^{\mathrm{T}}B_ks_k}+\frac{\tilde{y}_k\tilde{y}_k^{\mathrm{T}}}{\tilde{y}_k^{\mathrm{T}}s_k}+\phi(s_k^{\mathrm{T}}B_ks_k)v_kv_k^{\mathrm{T}}$$

校正，其中 $\phi\in[0,1)$

$$s_k=\lambda_kp_k=x_{k+1}-x_k$$

$$v_k=\frac{\tilde{y}_k}{\tilde{y}_k^{\mathrm{T}}s_k}-\frac{B_ks_k}{s_k^{\mathrm{T}}B_ks_k}$$

$$\tilde{y}_k=\begin{cases}y_k & \gamma_k<-\dfrac{1}{2}y_k^{\mathrm{T}}s_k\\ \hat{y}_k & \gamma_k\geqslant-\dfrac{1}{2}y_k^{\mathrm{T}}s_k\end{cases} \tag{3}$$

$$\gamma_k=3g_{k+1}^{\mathrm{T}}s_k+3g_k^{\mathrm{T}}s_k+6(f_k-f_{k+1}) \tag{4}$$

$$\hat{y}_k=y_k+\frac{\gamma_k}{s_k^{\mathrm{T}}y_k}y_k \tag{5}$$

上述算法与传统拟牛顿法的不同之处在于它使用了 $\tilde{y}_k$ 而不是 y_k，这里使用 $\tilde{y}_k$ 而没有直接使用新拟牛顿方程(2)的 $\hat{y}_k$ 是因为 $\hat{y}_k$ 不能保证 $\hat{y}_k^{\mathrm{T}}s_k>0$，从而失去了正定继承性。现采取策略(3)定义的 $\tilde{y}_k$，克服了这一困难。可以证明(见以下的引理4)当 k 充分大时，必有 $\tilde{y}_k=\hat{y}_k$，可见最终总是使用新拟牛顿公式(2)。

这里称上述算法为新拟牛顿法。下面分别对它的全局和局部收敛性质进行分析与证明。

先研究全局收敛性。

假定：(A1) $f(x)$二阶连续可微；(A2) $f(x)$一致凸，即存在 $m,M>0$ 使

$$m\|u\|^2\leqslant u^{\mathrm{T}}G(x)u\leqslant M\|u\|^2 \tag{6}$$

$\forall\ x\in S,u\in\mathbf{R}^n$，其中 $S=\{x\,|\,f(x)\leqslant f(x_1)\}$，$x_1$ 为 $\mathbf{R}^n$ 中任一点。

记 x^* 为 $f(x)$在 S 上的唯一极小点，θ_k 为 $-g_k$ 与 s_k 的夹角，有 $-g_k^{\mathrm{T}}s_k=\|g_k\|\ \|s_k\|\cos\theta_k$，则对新拟牛顿法有如下引理。

引理 1 对式(3)所定义的 $\tilde{y}_k$，有

$$\frac{1}{2}y_k^{\mathrm{T}}s_k\leqslant\tilde{y}_k^{\mathrm{T}}s_k\leqslant4y_k^{\mathrm{T}}s_k \tag{7}$$

证明 由 Taylor 公式有

$$f_{k+1}=f_k+g_k^{\mathrm{T}}s_k+\frac{1}{2}s_k^{\mathrm{T}}G(\xi)s_k \tag{8}$$

其中 $\xi\in(x_k,x_{k+1})$。将式(8)代入(4)得

$$\gamma_k=3y_k^{\mathrm{T}}s_k-3s_k^{\mathrm{T}}G(\xi)s_k\leqslant3y_k^{\mathrm{T}}s_k \tag{9}$$

结合式(3)知式(7)成立。

引理 2 下列不等式成立：

$$\frac{\|\tilde{y}_k\|^2}{\tilde{y}_k^{\mathrm{T}}s_k}\leqslant\overline{M} \tag{10}$$

$$\frac{s_k^{\mathrm{T}}B_ks_k}{\tilde{y}_k^{\mathrm{T}}s_k}\leqslant\frac{2\lambda_k}{1-\beta} \tag{11}$$

$$\frac{\|B_ks_k\|^2}{s_k^{\mathrm{T}}B_ks_k}\geqslant\frac{\lambda_k}{c_2\cos^2\theta_k}$$

$$\frac{|\tilde{y}_k^{\mathrm{T}}\boldsymbol{B}_k s_k|}{\tilde{y}_k^{\mathrm{T}} s_k} \leqslant \frac{\lambda_k \overline{M}}{m c_1 \cos\theta_k} \tag{12}$$

证明　取 $\overline{M}=4M$，其中 M 的定义见式(6)。显然仅对 $\tilde{y}_k=\hat{y}_k$ 进行证明即可。由式(7)及文献[3]中的式(3.4)知式(11)成立。

由于 $\hat{y}_k=\left(1+\dfrac{\gamma_k}{y_k^{\mathrm{T}} s_k}\right)y_k$，又据式(9)及文献[3]中的式(3.3)，有

$$\frac{\|\hat{y}_k\|^2}{\hat{y}_k^{\mathrm{T}} s_k}=\left(1+\frac{\gamma_k}{y_k^{\mathrm{T}} s_k}\right)\frac{\|y_k\|^2}{y_k^{\mathrm{T}} s_k}\leqslant 4M=\overline{M}$$

故式(10)成立。

又由于

$$\frac{|\hat{y}_k^{\mathrm{T}}\boldsymbol{B}_k s_k|}{\hat{y}_k^{\mathrm{T}} s_k}=\frac{|y_k^{\mathrm{T}}\boldsymbol{B}_k s_k|}{y_k^{\mathrm{T}} s_k}$$

故式(12)也成立。

引理 3　存在一常数 $c_4>0$，使对所有 $k\geqslant 1$，有

$$\prod_{j=1}^{k}\lambda_j\geqslant c_4^k$$

证明　现用 $T_{\mathrm{r}}(\boldsymbol{B}_k)$ 表示 $\boldsymbol{B}_k$ 的追迹，由文献[3]中的式(3.2)将 y_k 替换后得

$$T_{\mathrm{r}}(\boldsymbol{B}_{k+1})=T_{\mathrm{r}}(\boldsymbol{B}_k)+\frac{\|\tilde{y}_k\|^2}{\tilde{y}_k^{\mathrm{T}} s_k}+\phi\frac{\|\tilde{y}_k\|^2}{\tilde{y}_k^{\mathrm{T}} s_k}\frac{s_k^{\mathrm{T}}\boldsymbol{B}_k s_k}{\tilde{y}_k^{\mathrm{T}} s_k}- (1-\phi)\frac{\|\boldsymbol{B}_k s_k\|^2}{s_k^{\mathrm{T}}\boldsymbol{B}_k s_k}-2\phi\frac{\tilde{y}_k^{\mathrm{T}}\boldsymbol{B}_k s_k}{\tilde{y}_k^{\mathrm{T}} s_k} \tag{13}$$

由引理 2 得

$$T_{\mathrm{r}}(\boldsymbol{B}_{k+1})\leqslant T_{\mathrm{r}}(\boldsymbol{B}_k)+\overline{M}+\frac{2\phi\overline{M}\lambda_k}{1-\beta}- \frac{(1-\phi)\lambda_k}{c_2\cos^2\theta_k}+\frac{2\phi\overline{M}\lambda_k}{m c_1\cos\theta_k} \tag{14}$$

类似文献[3]中的证明，$T_{\mathrm{r}}(\boldsymbol{B}_{k+1})\leqslant\overline{M}+\{[1/(1-\beta)+1/mc_1]2\phi\overline{M}c_2\}T_{\mathrm{r}}(\boldsymbol{B}_k)$，故存在常数 $c_3>0$，使 $T_{\mathrm{r}}(\boldsymbol{B}_{k+1})\leqslant c_3^k$ 成立。以下同文献[3]中引理 3.2。

定理 1　设 x_1 为任一初始点，$f(x)$ 在水平集 $S=\{x\,|\,f(x)\leqslant f(x_1)\}$ 上满足假定(A1)和(A2)，则对任何正定矩阵 $\boldsymbol{B}_1$，新拟牛顿法产生的 $\{x_k\}$ 收敛到最优解 x^*。

证明　将式(14)写成

$$T_{\mathrm{r}}(\boldsymbol{B}_{k+1})\leqslant T_{\mathrm{r}}(\boldsymbol{B}_k)+\overline{M}+\lambda_k\left(\frac{2\phi\overline{M}}{1-\beta}-\frac{1-\phi}{c_2\cos^2\theta_k}+\frac{2\phi\overline{M}}{m c_1\cos\theta_k}\right)$$

以下，证明同文献[3]中定理 3.1，略。

本定理说明我们所讨论的新拟牛顿法在目标函数一致凸的假定下具有较强的全局收敛性。我们还可以证明在凸函数假定下较弱的全局收敛性仍成立，其证明方法类似于文献[5]。

现在研究超线性收敛性。

由文献[2]可知，我们所讨论的新拟牛顿法具有线性变换下的不变性。类似于文献[3]，为证明新拟牛顿法的超线性收敛性仅需考虑 $\nabla^2 f(x^*)=\boldsymbol{I}$ 时的情形。

引理 4　设 $f(x)$ 满足假定(A1)和(A2)，且新拟牛顿法产生的迭代点列 $\{x_k\}$ 收敛于最优点

x^*,则

$$\lim_{k\to\infty}\frac{\tilde{y}_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}=1 \tag{15}$$

证明 当 $\tilde{y}_k=y_k$ 时,$\tilde{y}_k^{\mathrm{T}}s_k/y_k^{\mathrm{T}}s_k=1$,故仅需证明 $\tilde{y}_k=\hat{y}_k$ 时的情形。由式(5)有 $\hat{y}_k^{\mathrm{T}}s_k=y_k^{\mathrm{T}}s_k+\gamma_k$,由式(9)有 $\gamma_k=3y_k^{\mathrm{T}}s_k-3s_k^{\mathrm{T}}\boldsymbol{G}(\xi)s_k$;又

$$y_k=g_{k+1}-g_k=s_k\int_0^1\boldsymbol{G}(x_k+ts_k)\mathrm{d}t=\hat{\boldsymbol{G}}s_k$$

由于 $\xi\in(x_k,x_{k+1})$ 且 $x_k\to x^*$ 及 $\boldsymbol{G}(x)$ 连续,以及 $f(x)$ 一致凸,故

$$\frac{\hat{y}_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}=1+\frac{\gamma_k}{y_k^{\mathrm{T}}s_k}=4-3\,\frac{s_k^{\mathrm{T}}\boldsymbol{G}(\xi)s_k}{s_k^{\mathrm{T}}\hat{\boldsymbol{G}}s_k}\to 1$$

所以式(15)成立。

本引理说明:如果使用策略(3),那么当 k 充分大时必有 $\tilde{y}_k=\hat{y}_k$,即最终使用新拟牛顿方程。

引理5 对任何 $\varepsilon\in(0,1)$,存在 x^* 的一个邻域 $N(x^*)$,使得若 $x_{k+1},x_k\in N(x^*)$,则

$$\frac{\|\tilde{y}_k\|^2}{\tilde{y}_k^{\mathrm{T}}s_k}\frac{s_k^{\mathrm{T}}\boldsymbol{B}_ks_k}{\tilde{y}_k^{\mathrm{T}}s_k}-2\,\frac{\tilde{y}_k^{\mathrm{T}}\boldsymbol{B}_ks_k}{\tilde{y}_k^{\mathrm{T}}s_k}\leqslant\frac{2\lambda_k\varepsilon}{mc_1\cos\theta_k} \tag{16}$$

证明 由于 $\boldsymbol{G}^*=\boldsymbol{I}$,由文献[3]中的式(4.2)有 $y_k=\boldsymbol{E}_ks_k+s_k$,其中

$$\boldsymbol{E}_k=\int_0^1[\boldsymbol{G}(x_k+ts_k)-\boldsymbol{I}]\mathrm{d}t$$

且当 $N(x^*)$ 充分小时有 $\|\boldsymbol{E}_k\|\leqslant\varepsilon$;由式(9)及假定(A1)有

$$\frac{\gamma_k}{s_k^{\mathrm{T}}s_k}=\frac{3y_k^{\mathrm{T}}s_k-3s_k^{\mathrm{T}}\boldsymbol{G}(\xi)s_k}{s_k^{\mathrm{T}}s_k}\to 0\quad(k\to\infty)$$

又由 $y_k=\hat{\boldsymbol{G}}s_k$ 及式(6)知 $y_k^{\mathrm{T}}s_k=O(s_k^{\mathrm{T}}s_k)$,故

$$\frac{\gamma_k}{y_k^{\mathrm{T}}s_k}=\frac{3y_k^{\mathrm{T}}s_k-3s_k^{\mathrm{T}}\boldsymbol{G}(\xi)s_k}{y_k^{\mathrm{T}}s_k}\to 0\quad(k\to\infty)$$

因此当 $N(x^*)$ 充分小时,仍有 $\|\boldsymbol{E}_k+\gamma_k\boldsymbol{I}/y_k^{\mathrm{T}}s_k+\gamma_k\boldsymbol{E}_k/y_k^{\mathrm{T}}s_k\|<\varepsilon$。现用 $\hat{\boldsymbol{E}}_k$ 记 $\boldsymbol{E}_k+\gamma_k\boldsymbol{I}/y_k^{\mathrm{T}}s_k+\gamma_k\boldsymbol{E}_k/y_k^{\mathrm{T}}s_k$,所以 $\hat{y}_k=\hat{\boldsymbol{E}}_ks_k+s_k$。当 $\tilde{y}_k=\hat{y}_k$ 时,取 $\tilde{\boldsymbol{E}}_k=\hat{\boldsymbol{E}}_k$,否则取 $\tilde{\boldsymbol{E}}_k=\boldsymbol{E}_k$,则有 $\tilde{y}_k=\tilde{\boldsymbol{E}}_ks_k+s_k$,且当 $N(x^*)$ 充分小时 $\|\tilde{\boldsymbol{E}}_k\|\leqslant\varepsilon$。以下证明同文献[3]中的引理4.1。

引理6 对于新拟牛顿法,存在常数 $c_8\in[0,1)$,使 $f_{k+1}-f^*\leqslant c_8^k(f_1-f^*)$。

证明 由定理1,迭代点列 $x_k\to x^*$,由引理5,对任何 $0<\varepsilon\leqslant 1$ 及充分大的 k,将式(10),(16)和(12)及文献[3]中的式(3.5)代入(13)得

$$0<T_{\mathrm{r}}(\boldsymbol{B}_{k+1})\leqslant T_{\mathrm{r}}(\boldsymbol{B}_k)+\overline{M}+\frac{2\lambda_k\phi\varepsilon}{mc_1\cos\theta_k}-(1-\phi)\frac{\lambda_k}{c_2\cos^2\theta_k}\leqslant$$

$$T_{\mathrm{r}}(\boldsymbol{B}_k)+\overline{M}+\frac{1}{\cos^2\theta_k}\left(\frac{2\phi\varepsilon\cos\theta_k}{mc_1}-\frac{1-\phi}{c_2}\right)\lambda_k$$

以下证明同文献[3]中的引理4.2。

为了证明算法的Q超线性收敛性,再假定(A3)$f(x)$ 的海色矩阵 $\boldsymbol{G}(x)$ 在 x^* 处是Hölder连续的,即存在常数 p 与 L,使

$$\|\boldsymbol{G}(x)-\boldsymbol{G}^*\|\leqslant L\|x-x^*\|^p \tag{17}$$

引理 7　在假定(A1)，(A2)和(A3)下，设$\{x_k\}$为收敛到x^*的序列，且

$$\sum_{k=1}^{\infty}\|x_k-x^*\|^p<\infty$$

则对[0,1]中的任何序列$\{\phi_k\}$和某个正定矩阵$\boldsymbol{B}_1$，由新拟牛顿法产生的序列$\{\boldsymbol{B}_k\}$，满足

$$\lim_{k\to\infty}\frac{\|(\boldsymbol{B}_k-\boldsymbol{G}^*)s_k\|}{\|s_k\|}=0$$

且$\{\|\boldsymbol{B}_k\|\}$与$\{\|\boldsymbol{B}_k^{-1}\|\}$均有界。

证明　如前所述设$\boldsymbol{G}^*=\boldsymbol{I}$，为书写方便省略下标$k$，且以$\boldsymbol{B}^+$表示$\boldsymbol{B}_{k+1}$，记$d=\max\{\|x_+-x^*\|^p,\|x-x^*\|^p\}$，$y=g(x+s)-g(x)=\hat{\boldsymbol{G}}s$。由式(17)知$\hat{\boldsymbol{G}}=\boldsymbol{G}^*+\boldsymbol{O}(d)$，且$\boldsymbol{G}^*=\boldsymbol{I}$，得$y=s+\boldsymbol{O}(d)s$。因$|\gamma|=|3[y^{\mathrm{T}}s-s^{\mathrm{T}}\boldsymbol{G}(\xi)s]|=|3s^{\mathrm{T}}[\hat{\boldsymbol{G}}-\boldsymbol{G}(\xi)]s|\leqslant 3\|\hat{\boldsymbol{G}}-\boldsymbol{G}(\xi)\|\|s\|^2\leqslant 3[\|\hat{\boldsymbol{G}}-\boldsymbol{G}(x^*)\|+\|\boldsymbol{G}(x^*)-\boldsymbol{G}(\xi)\|]\|s\|^2\leqslant 6L\max\{\|x_+-x^*\|^p,\|\xi-x^*\|^p\}\|s\|^2=6L\boldsymbol{O}(d)\|s\|^2=\boldsymbol{O}(d)s^{\mathrm{T}}s$，所以$\gamma/s^{\mathrm{T}}s=\boldsymbol{O}(d)$；又由于$m\|s\|^2\leqslant y^{\mathrm{T}}s\leqslant M\|s\|^2$，$\gamma/y^{\mathrm{T}}s=\boldsymbol{O}(d)$，所以$\hat{y}=(1+\gamma/y^{\mathrm{T}}s)y=(1+\gamma/y^{\mathrm{T}}s)[s+\boldsymbol{O}(d)s]=s+\boldsymbol{O}(d)s$；因此$\tilde{y}=s+\boldsymbol{O}(d)s$。于是对$s\neq0$容易推出

$$\frac{\|\tilde{y}\|}{\|s\|}=1+\boldsymbol{O}(d)$$

$$\frac{\|\tilde{y}\|\|s\|}{\tilde{y}^{\mathrm{T}}s}=1+\boldsymbol{O}(d)$$

以下证明同文献[4]中的引理 4。

定理 8　设x_1为任一初始点，$f(x)$在水平集$S=\{x\mid f(x)\leqslant f(x_1)\}$上满足假定(A1)，(A2)和(A3)，则对任何正定矩阵$\boldsymbol{B}_1$，由新拟牛顿法所产生的$\{x_k\}$超线性收敛于x^*。

证明　在引理 4～7 的基础上，可用类似于文献[3]中定理 4.1 的证明方法证明。

参 考 文 献

1 Zhang Jianzhong, Deng Naiyang, Chen Lihua. A new quasi-Newton equation and related methods for unconstrained optimization. Research Report, MA-96-05, to appear in JOTA

2 Zhang Jianzhong, Xu Chengxian. Properties and performance of quasi-Newton methods with modified quasi-Newton equation. Journal of Kunming Symposium Theory, 1998

3 Byrd R H, Nocedal J, Yuan Y X. Global convergence of quasi-Newton methods on convex problems. SLAM J Num Anal, 1987, 24:1171～1190

4 Griewank A, Toint P L. Local convergence analysis of partitioned quasi-Newton updates. Num Math, 1982, 39:429～448

5 席少霖. 非线性最优化方法. 北京：高等教育出版社，1992. 128～160