文章编号：0253-374X(2013)02-0289-04

DOI：10.3969/j.issn.0253-374x.2013.02.023

不假定凸性和精确线搜索时DFP算法的收敛性

濮定国[1,2]，刘美玲[3]

(1. 同济大学 数学系，上海 200092；2. 河南科技大学 数学系，河南 洛阳 471003；
3. 上海电机学院 数理教学部，上海 201306)

摘要：对非凸目标函数，Broyden变尺度算法的收敛性是一个没有完全解决的问题. 针对DFP修正公式证明在不假定精确线搜索条件下，对光滑的目标函数，当DFP算法得到的点列收敛时，该点列一定趋向于稳定点. 指出对于其他Broyden算法结论都是成立的.

关键词：变尺度算法；收敛性；凸性；不精确线搜索

中图分类号：O 221.2　　**文献标志码**：A

Convergence of DFP Algorithms Without Convexity and Exact Line Search Assumptions

PU Dingguo[1,2], *LIU Meiling*[3]

(1. Department of Mathematics, Tongji University, Shanghai 200092, China; 2. Department of Mathematics, Henan University of Science and Technology, Luoyang 471003, China; 3. Department of Mathematics and Physics, Shanghai Dianji University, Shanghai 201306, China)

Abstract: The convergence of the Broyden algorithms without convexity and exact line search assumptions is discussed. It is proved that if the objective function is suitably smooth and the DFP algorithm produces a convergence point sequence, then the limit point of the sequence is a critical point of the objective function. We give mainly a proof for the DFP update, then point out that all the results are true for Broyden algorithms by a remark.

Key words: Broyden algorithms; convergence; convexity; inexact line search

变尺度算法，如Broyden算法[1]，是求解非线性规划

$$\min\{f(x): x \in \mathbf{R}^n\}$$

的一类有效方法. 在精确线搜索条件下，Powell[2]证明了这类算法对于一致凸函数是一步超线性收敛的. 对于非精确线搜索条件，Powell[3]证明了在非精确线搜索条件下，BFGS算法可以一步超线性收敛. Byrd，Nocedal和Yuan[4]证明了除了DFP算法外，该结论对限制Broyden类算法也是成立的. 采用修正的Wolfe条件，Pu[5]证明了在非精确线搜索条件下，DFP算法是一步超线性收敛的.

但是无论线搜索是否精确，有例子表明对非凸目标函数，Broyden算法可能不收敛[6]. Broyden算法收敛性还是一个没有完全解决的问题. 对非凸目标函数，在精确线搜索条件下，如果产生的点列收敛，Pu和Yu[7]证明了对于连续可微函数这类算法全局收敛.

本文考虑对非凸目标函数，在非精确线搜索条件下，Broyden算法的收敛性. 证明对于光滑的目标函数，当Broyden算法得到的点列收敛时，该点列的极限点是一个稳定点. 主要针对DFP修正公式进行证明，并指出对于其他Broyden算法结论都是成立的.

Broyden算法是迭代算法[8]. 给定初始点x_1，正定矩阵$\boldsymbol{B}_1$及常数$\varphi\in[0,1]$，Broyden算法产生的点列$\{x_k\}$为

$$x_{k+1}=x_k+s_k=x_k+\alpha_k d_k=x_k-\alpha_k\boldsymbol{B}_k^{-1}g_k$$

式中：d_k是搜索方向；$\alpha_k\geqslant 0$是步长因子；g_k是$f(x)$在x_k处的梯度. 如果$g_k=0$，算法终止. 否则，更新$\boldsymbol{B}_k$为

$$\boldsymbol{B}_{k+1}=\boldsymbol{B}_k-\frac{\boldsymbol{B}_k s_k s_k^{\mathrm{T}}\boldsymbol{B}_k}{s_k^{\mathrm{T}}\boldsymbol{B}_k s_k}+\frac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}}s_k}+\varphi(s_k^{\mathrm{T}}\boldsymbol{B}_k s_k)v_k v_k^{\mathrm{T}} \tag{1}$$

式中：$y_k=g_{k+1}-g_k$；$v_k=y_k(y_k^{\mathrm{T}}s_k)^{-1}-\boldsymbol{B}_k s_k(s_k^{\mathrm{T}}\boldsymbol{B}_k s_k)^{-1}$.

在算法中，若$\varphi=0$，算法称为BFGS算法；若φ

收稿日期：2011-12-20
基金项目：国家自然科学基金(10771162，U1135003)

第一作者：濮定国(1948—)，男，教授，博士生导师，主要研究方向为最优化理论与方法. E-mail：madpu@tongji.edu.cn

$=1$,则称为DFP算法. 用 $\boldsymbol{H}_k$ 表示 $\boldsymbol{B}_k$ 的逆矩阵,直接计算得

$$\boldsymbol{H}_{k+1}=\boldsymbol{H}_k-\frac{\boldsymbol{H}_k\boldsymbol{y}_k\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{H}_k}{\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{y}_k}+\frac{\boldsymbol{s}_k\boldsymbol{s}_k^{\mathrm{T}}}{\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k}+\eta\frac{\boldsymbol{\mu}_k\boldsymbol{\mu}_k^{\mathrm{T}}}{\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{y}_k} \tag{2}$$

其中 $\boldsymbol{\mu}_k$ 定义为

$$\boldsymbol{\mu}_k=\frac{(\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{y}_k)\boldsymbol{s}_k}{\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k}-\boldsymbol{H}_k\boldsymbol{y}_k \tag{3}$$

及 $\eta\in[0,1]$.

本文采用非精确线搜索. 为了保证目标函数值下降及算法的收敛性,选取步长因子 α_k 满足如下修正的Wolfe条件:

$$f(x_k)-f(x_{k+1})\geqslant\sum_{j=1}^{k}\rho_j(-\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{s}_j) \tag{4}$$

及

$$|\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{s}_k|\leqslant\theta_k(-\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{s}_k) \tag{5}$$

其中 ρ_k 和 θ_k 定义为

$$\rho_k=\rho\min\{1,n(\mathrm{tr}(\overline{\boldsymbol{H}}_k+\overline{\boldsymbol{B}}_k))^{-1}\}$$
$$\theta_k=\theta\min\{1,n(\mathrm{tr}(\overline{\boldsymbol{H}}_k+\overline{\boldsymbol{B}}_k))^{-1}\}$$

这里 $\rho\in(0,1/2)$,$\theta\in[\rho,1/2]$及

$$\overline{\boldsymbol{H}}_k=\boldsymbol{W}\boldsymbol{H}_k\boldsymbol{W}^{\mathrm{T}},\ \overline{\boldsymbol{B}}_k=\boldsymbol{W}^{-\mathrm{T}}\boldsymbol{B}_k\boldsymbol{W}^{-1}$$

其中 $\boldsymbol{W}$ 可以是任何非奇异矩阵,并且存在常数 M_1 使得 $\|\boldsymbol{W}\|<M_1$ 及 $\|\boldsymbol{W}^{-1}\|<M_1$.

由 $\overline{\boldsymbol{H}}_k$ 的定义可以得到

$$nM_1^2\,\mathrm{tr}\,\overline{\boldsymbol{B}}_k\geqslant\mathrm{tr}\,\boldsymbol{B}_k\geqslant nM_1^{-2}\,\mathrm{tr}\,\overline{\boldsymbol{B}}_k$$
$$nM_1^2\,\mathrm{tr}\,\overline{\boldsymbol{H}}_k\geqslant\mathrm{tr}\,\boldsymbol{H}_k\geqslant nM_1^{-2}\,\mathrm{tr}\,\overline{\boldsymbol{H}}_k$$

为了书写简单,不妨设 $\theta=\rho$,$\boldsymbol{W}$ 是单位阵,即 $\overline{\boldsymbol{H}}_k=\boldsymbol{H}_k$,这些假定不影响本文结果.

若 $\alpha=1$ 时,不等式(4)和(5)能成立,则取 $\alpha=1$.

本文假设:

(1) 目标函数 $f(x)$ 是连续可微的,且 $f(x)$ 的梯度满足Lipschitzian条件.

(2) 水平集 $S(x)=\{x:f(x)\leqslant f(x_1)\}$ 有界,所以存在一个正常数 M 使得

$$\|\boldsymbol{x}_k\|\leqslant M,\ \|\boldsymbol{g}_k\|\leqslant M$$

因为 $f(x)$ 的梯度满足Lipschitzian条件,由Taylor展开式可知能取到满足式(4)和(5)的步长因子 α_k.

$$U_k=\frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{g}_k}{-\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{y}_k},\ V_k=\frac{-\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{y}_k}{\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{y}_k} \tag{6}$$

$$P_k=\sum_{j=1}^{k}\frac{\|\boldsymbol{s}_j\|^2}{\alpha_j} \tag{7}$$

$$Q_k=\|\boldsymbol{H}_{k+1}\boldsymbol{g}_{k+1}\|+\sum_{j=1}^{k}\|\boldsymbol{H}_{j+1}\boldsymbol{g}_{j+1}-\boldsymbol{H}_j\boldsymbol{g}_j\|+\|\boldsymbol{H}_1\boldsymbol{g}_1\| \tag{8}$$

及

$$\begin{aligned}R_k=&\mathrm{tr}(\boldsymbol{B}_{k+1}+\boldsymbol{H}_{k+1})\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{g}_k+\\&\mathrm{tr}(\boldsymbol{B}_1+\boldsymbol{H}_1)\boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{H}_1\boldsymbol{g}_1+\\&\sum_{j=1}^{k}V_j[\|\boldsymbol{H}_j\boldsymbol{y}_j\|^2+\mathrm{tr}(\boldsymbol{B}_j+\\&\boldsymbol{H}_j)(\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j)]\end{aligned} \tag{9}$$

从不等式(5)可以得到一个非常有用的性质:$\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k\geqslant0$ 对于所有的 k 成立,因此 $\boldsymbol{B}_k$ 和 $\boldsymbol{H}_k$ 均是正定的.

定理1 假设算法产生的点列 $\{x_k\}$ 趋向于 x_*,则 x_* 是 $f(x)$ 的一个稳定点,即有 $g(x_*)=0$.

在不引起混淆的前提下,省略掉下标. 比如用 $\boldsymbol{g}$,$\boldsymbol{x}$,$\boldsymbol{R}$ 代替 $\boldsymbol{g}_k$,$\boldsymbol{x}_k$,$\boldsymbol{R}_k$,并且用下标 $*$ 来表示下一步迭代得到的向量和矩阵,比如用 $\boldsymbol{g}_*$,$\boldsymbol{x}_*$,$\boldsymbol{R}_*$ 表示 $\boldsymbol{g}_{k+1}$,$\boldsymbol{x}_{k+1}$,$\boldsymbol{R}_{k+1}$.

本文所用到的范数均为Euclidean范数.

1 预备知识

通过简单计算,可得如下公式(10)—(17)(参见文献[9]):

当 $\varphi=1$ 时,对式(1)和(2)两边同时取迹[3]得

$$\mathrm{tr}\,\boldsymbol{H}_*=\mathrm{tr}\,\boldsymbol{H}-\frac{\|\boldsymbol{Hy}\|^2}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{Hy}}+\frac{\|\boldsymbol{s}\|^2}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{s}} \tag{10}$$

和

$$\mathrm{tr}\,\boldsymbol{B}_*+\frac{\|\boldsymbol{g}\|^2-\|\boldsymbol{g}_*\|^2}{-\boldsymbol{g}^{\mathrm{T}}\boldsymbol{Hy}}=\mathrm{tr}\,\boldsymbol{B}+(1+\alpha U)\frac{\|\boldsymbol{y}\|^2}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{s}} \tag{11}$$

对式(2)两边同乘 $\boldsymbol{g}_{k+1}$,然后再两边对 $\boldsymbol{g}_{k+1}$ 取内积(参考文献[5]中2.12和2.20),得

$$\boldsymbol{g}_*^{\mathrm{T}}\boldsymbol{H}_*\boldsymbol{g}_*=V\boldsymbol{g}_*^{\mathrm{T}}\boldsymbol{Hy}+\boldsymbol{g}_*^{\mathrm{T}}\boldsymbol{Hg}+\alpha U\boldsymbol{g}_*^{\mathrm{T}}\boldsymbol{Hg} \tag{12}$$

及

$$\boldsymbol{g}^{\mathrm{T}}\boldsymbol{Hg}-\boldsymbol{g}_*^{\mathrm{T}}\boldsymbol{H}_*\boldsymbol{g}_*=V(-\boldsymbol{g}^{\mathrm{T}}\boldsymbol{Hy})-\alpha U\boldsymbol{g}_*^{\mathrm{T}}\boldsymbol{Hg} \tag{13}$$

其中 U 和 V 见式(6). 由 ρ 的定义和不等式 $\mathrm{tr}\,\boldsymbol{B}\cdot(\boldsymbol{g}^{\mathrm{T}}\boldsymbol{Hg})\geqslant\|\boldsymbol{g}\|^2$,对所有的 k 有

$$\begin{aligned}|\boldsymbol{g}_*^{\mathrm{T}}\boldsymbol{Hg}|&\leqslant\rho\,\mathrm{tr}^{-1}(\overline{\boldsymbol{B}})(\boldsymbol{g}^{\mathrm{T}}\boldsymbol{Hg})\\&\leqslant\rho nM_1^2\,\mathrm{tr}^{-1}(\boldsymbol{B})(\boldsymbol{g}^{\mathrm{T}}\boldsymbol{Hg})\\&\leqslant\rho nM_1^2(\boldsymbol{g}^{\mathrm{T}}\boldsymbol{Hg})^2\|\boldsymbol{g}\|^{-2}\end{aligned} \tag{14}$$

及

$$\boldsymbol{g}^{\mathrm{T}}\boldsymbol{Hg}\leqslant-2\boldsymbol{g}^{\mathrm{T}}\boldsymbol{Hy}\leqslant2\|\boldsymbol{Hg}\|\|\boldsymbol{y}\|\leqslant2L\|\boldsymbol{Hg}\|\|\boldsymbol{s}\| \tag{15}$$

由 $\overline{\boldsymbol{B}}_k$ 和 $\overline{\boldsymbol{H}}_k$ 的正定性,得 $\mathrm{tr}(\overline{\boldsymbol{B}}_k+\overline{\boldsymbol{H}}_k)\geqslant2$. 因此由式(14),(15),$U$ 的定义及 $\mathrm{tr}\,\boldsymbol{H}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{y}\geqslant\|\boldsymbol{Hy}\|^2$ 推出

$$|U|\leqslant\min\left\{\frac{1}{2},\frac{2nM_1^2L\|\boldsymbol{s}\|^2}{\alpha\|\boldsymbol{g}\|^2},\frac{nM_1^2\|\boldsymbol{s}\|^2}{\alpha\|\boldsymbol{g}\|},\frac{nM_1^2L\|\boldsymbol{s}\|^2}{\alpha V\|\boldsymbol{Hy}\|^2}\right\} \tag{16}$$

将式(10)和(11)的两边相加得

$$\mathrm{tr}(\boldsymbol{H}_*+\boldsymbol{B}_*-\boldsymbol{H}-\boldsymbol{B})=\frac{-\|\boldsymbol{g}\|^2+\|\boldsymbol{g}_*\|^2}{-\boldsymbol{g}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{y}}+(1+\alpha U)\frac{\|\boldsymbol{y}\|^2}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{s}}-\frac{\|\boldsymbol{H}\boldsymbol{y}\|^2}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{y}}+\frac{\|\boldsymbol{s}\|^2}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{s}} \quad (17)$$

将不等式(5)和(16)代入式(17),由 $V_k\leqslant 2$,对于所有的 k 有

$$|\boldsymbol{g}_*^{\mathrm{T}}\boldsymbol{H}\boldsymbol{g}\,\mathrm{tr}(\boldsymbol{H}_*+\boldsymbol{B}_*-\boldsymbol{H}-\boldsymbol{B})(V+\alpha U)|\leqslant |U|\{2LM\|\boldsymbol{s}\|+V\|\boldsymbol{H}\boldsymbol{y}\|^2+[1+L^2|1+\alpha U|]\frac{\|\boldsymbol{s}\|^2}{\alpha}\}|V+\alpha U|\leqslant \left\{\left[2LM+L+1+L^2\left(1+\frac{nM_1^2\|\boldsymbol{s}\|}{\|\boldsymbol{g}\|}\right)\right]\cdot\left(2+\frac{nM_1^2\|\boldsymbol{s}\|}{\|\boldsymbol{g}\|}\right)\right\}\frac{nM_1^2\|\boldsymbol{s}\|^2}{\alpha}$$

引理 1 存在 C_1 和 $C_2>0$ 使得对所有的 k

$$R_k\leqslant\sum_{j=1}^{k}\left[\left(C_1+C_2\frac{\|\boldsymbol{s}_j\|^2}{\|\boldsymbol{g}_j\|^2}\right)\frac{\|\boldsymbol{s}_j\|^2}{\alpha_j}\right] \quad (18)$$

其中 R_k 的定义见式(9).

证明 对式(11)和(12)两边同乘以 $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{y}_k$,取代 k 为 j,并对 $j=1,2,\cdots,k$ 将两式两边分别相加,作相应的移动可以得到

$$\sum_{j=1}^{k}[-\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{y}_j\mathrm{tr}(\boldsymbol{H}_{j+1}+\boldsymbol{B}_{j+1}-\boldsymbol{H}_j-\boldsymbol{B}_j)+V_j\|\boldsymbol{H}_j\boldsymbol{y}_j\|^2]=\sum_{j=1}^{k}\frac{\|\boldsymbol{s}_j\|^2+(1+\alpha_jU_j)\|\boldsymbol{y}_j\|^2}{\alpha_j}-\|\boldsymbol{g}_1\|^2+\|\boldsymbol{g}_{k+1}\|^2=\sum_{j=1}^{k}[1+L^2(1+\alpha_jU_j)]\frac{\|\boldsymbol{s}_j\|^2}{\alpha_j}-\|\boldsymbol{g}_1\|^2+\|\boldsymbol{g}_{k+1}\|^2\leqslant\sum_{j=1}^{k}\left[1+L^2\left(1+\frac{nM_1^2\|\boldsymbol{s}_j\|}{\|\boldsymbol{g}_j\|}\right)\right]\frac{\|\boldsymbol{s}_j\|^2}{\alpha_j}+M^2 \quad (19)$$

再由 Abelian 转换公式和式(13)得

$$\sum_{j=1}^{k}(1-V_j)\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j\mathrm{tr}(\boldsymbol{H}_{j+1}+\boldsymbol{B}_{j+1}-\boldsymbol{H}_j-\boldsymbol{B}_j)=\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{H}_{k+1}\boldsymbol{g}_{k+1}\mathrm{tr}(\boldsymbol{B}_{k+1}+\boldsymbol{H}_{k+1})-\boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{H}_1\boldsymbol{g}_1\mathrm{tr}(\boldsymbol{B}_1+\boldsymbol{H}_1)+\sum_{j=1}^{k}[V_j\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j\mathrm{tr}(\boldsymbol{B}_j+\boldsymbol{H}_j)-(V_j+\alpha_jU_j)\boldsymbol{g}_{j+1}^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j\mathrm{tr}(\boldsymbol{H}_{j+1}+\boldsymbol{B}_{j+1})]$$

取 $C_0=2(\boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{H}_1\boldsymbol{g}_1)\mathrm{tr}(\boldsymbol{B}_1+\boldsymbol{H}_1)$,由上式及 R_k 的定义有

$$R_k=\sum_{j=1}^{k}[(1-V_j)\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j\mathrm{tr}(\boldsymbol{H}_{j+1}+\boldsymbol{B}_{j+1}-\boldsymbol{H}_j-\boldsymbol{B}_j)+V_j\|\boldsymbol{H}_j\boldsymbol{y}_j\|^2]+\sum_{j=1}^{k}[(V_j+\alpha_jU_j)\boldsymbol{g}_{j+1}^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j\mathrm{tr}(\boldsymbol{H}_{j+1}+\boldsymbol{B}_{j+1})]+C_0 \quad (20)$$

将式(19)代入式(20),令 $|V_j-1|\leqslant 1$,则对所有的 j 可以得到

$$R_k=-\sum_{j=1}^{k}\boldsymbol{g}_{j+1}^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j\mathrm{tr}(\boldsymbol{H}_{j+1}+\boldsymbol{B}_{j+1})(V_j+\alpha_jU_j)\leqslant\sum_{j=1}^{k}2\left[1+L^2\left(1+\frac{nM_1^2\|\boldsymbol{s}_j\|}{\|\boldsymbol{g}_j\|}\right)\right]\frac{\|\boldsymbol{s}_j\|^2}{\alpha_j}+M^2+C_0 \quad (21)$$

且由不等式(4)有

$$|\boldsymbol{g}_{j+1}^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j\mathrm{tr}(\boldsymbol{H}_{j+1}+\boldsymbol{B}_{j+1})|\leqslant|\boldsymbol{g}_{j+1}^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j\mathrm{tr}(\boldsymbol{H}_{j+1}+\boldsymbol{B}_{j+1}-\boldsymbol{H}_j-\boldsymbol{B}_j)|+nM_1^2L\|\boldsymbol{s}_j\|^2/\alpha_j \quad (22)$$

对 $j=1,2,\cdots,k$,将式(19)和(22)代入式(21),得

$$R_k\leqslant\sum_{j=1}^{k}2\left[1+L^2\left(1+\frac{nM_1^2\|\boldsymbol{s}_j\|}{\|\boldsymbol{g}_j\|}\right)\right]\frac{\|\boldsymbol{s}_j\|^2}{\alpha_j}+M^2+C_0+\sum_{j=1}^{k}\left\{\left[2LM+2L+1+L^2\left(1+\frac{nM_1^2\|\boldsymbol{s}_j\|}{\|\boldsymbol{g}_j\|}\right)\right]\left(2+\frac{nM_1^2\|\boldsymbol{s}_j\|}{\|\boldsymbol{g}_j\|}\right)\frac{nM_1^2\|\boldsymbol{s}_j\|^2}{\alpha_j}\right\}$$

因此引理 1 成立.

2 算法的收敛性

首先假设 DFP 算法产生的点列 $\boldsymbol{x}_k$ 趋向于 $\boldsymbol{x}_*$,为了简单起见,取 $\boldsymbol{x}_*=0$. 若引言部分中提出的定理不真,则必存在一个正数 δ,$1>\delta>0$ 使得对所有的 k,$\|\boldsymbol{g}_k\|\geqslant\delta$,且存在一个正数 $J_1>0$ 使得对所有的 $j\geqslant J_1$

$$2nM_1^2\|\boldsymbol{y}_j\|\leqslant\|\boldsymbol{g}_j\|,\ 2nM_1^2\|\boldsymbol{s}_j\|\leqslant\|\boldsymbol{g}_j\| \quad (23)$$

成立. 为了叙述简单,不失一般性,假设对于所有的 k 式(23)成立.

在上述这些假设下,证明下述引理 2—6. 首先给出下面的引理 2(参考文献[10]).

引理 2 若$\{b_k\}$和$\{d_k\}$是两个正数序列且满足

$$\sum_{j=1}^{\infty}d_j=+\infty \text{ 及 } b_k\leqslant\alpha+\beta\sum_{j=1}^{k-1}d_j$$

其中 α 和 β 是正数,则下面的不等式成立

$$\sum_{j=1}^{\infty}\frac{d_j}{b_j}=+\infty$$

再由式(5),(14)和 $\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j\leqslant\|\boldsymbol{g}_j\|\|\boldsymbol{H}_j\boldsymbol{g}_j\|$,可以得到

$$0\leqslant\alpha_jU_j(\boldsymbol{g}_{j+1}^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j)\leqslant\frac{n^2M_1^4\alpha_j(\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j)^4}{(-\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{y}_j)\|\boldsymbol{g}_j\|^4}\leqslant\frac{n^2M_1^4\|\boldsymbol{s}_j\|^2(\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{g}_j)^2}{\alpha_j(-\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{H}_j\boldsymbol{y}_j)\delta^2}$$

下面的引理 3 的证明参考文献[5].

引理 3 对于所有的 k,有

$$\lim_{k\to\infty}P_k=+\infty$$

引理4 对于所有的 k,存在正常数 $C_3>0$ 使得

$$R_k \leqslant C_3 P_k \tag{24}$$

证明 由引理1和式(23)可以推得

$$R_k \leqslant (C_1 + 2C_2 M^2 \delta^{-2}) P_k$$

成立.显然,引理4为真.

引理5 若 $P_k \to +\infty$,则

$$\lim_{k\to\infty} P_k / Q_k = 0$$

证明 对任何给定的充分大的正数 $N>0$,存在一个正数 J 使得对所有 $j>J$,有 $\|x_j\| \leqslant (2N)^{-1}$.取

$$C_4 = \sum_{j=1}^{J} \frac{\|\boldsymbol{s}_j\|^2}{\alpha_j} + \|\boldsymbol{x}_J\| \|\boldsymbol{H}_J \boldsymbol{g}_J\| + \|\boldsymbol{x}_1\| \|\boldsymbol{H}_1 \boldsymbol{g}_1\|$$

根据 Abelian 转换公式,得

$$\sum_{j=J}^{k} (\boldsymbol{x}_j - \boldsymbol{x}_{j+1})^{\mathrm{T}} \boldsymbol{H}_j \boldsymbol{g}_j \leqslant \sum_{j=J}^{k} \|\boldsymbol{x}_{j+1}\| \|\boldsymbol{H}_{j+1} \boldsymbol{g}_{j+1} - \boldsymbol{H}_j \boldsymbol{g}_j\| + \|\boldsymbol{x}_J\| \|\boldsymbol{H}_J \boldsymbol{g}_J\| + \|\boldsymbol{x}_{k+1}\| \|\boldsymbol{H}_{k+1} \boldsymbol{g}_{k+1}\|$$

又由 P_k 的定义知,若 $k \leqslant J$,则 $P_k \leqslant C_4$,以及若 $k>J$,则

$$P_k \leqslant \frac{1}{2N} \sum_{j=J}^{k} \Big[\|\boldsymbol{H}_{j+1} \boldsymbol{g}_{j+1} - \boldsymbol{H}_j \boldsymbol{g}_j\| + \|\boldsymbol{H}_J \boldsymbol{g}_J\| + \|\boldsymbol{H}_{k+1} \boldsymbol{g}_{k+1}\| \Big] + C_4 \leqslant \frac{Q_k}{2N} + C_4$$

因为 $P_k \to +\infty$,存在 $J_2 \geqslant J$ 使得对于所有的 $k \geqslant J_2$

$$Q_k \geqslant 2N(P_k - C_4) \geqslant N P_k$$

因此,当 N 是任意大的正数时,引理结论成立.

引理6 若 $P_k \to +\infty$,则存在 $C_5>0$ 使得对于所有的 k

$$Q_k \leqslant C_5 R_k$$

证明 由 $\boldsymbol{H}_k$ 和 $\boldsymbol{B}_k$ 正定可知,对所有的 k

$$\mathrm{tr}(\boldsymbol{H}_{k+1}) \boldsymbol{g}_{k+1}^{\mathrm{T}} \boldsymbol{H}_{k+1} \boldsymbol{g}_{k+1} + 1 \geqslant \|\boldsymbol{H}_{k+1} \boldsymbol{g}_{k+1}\|^2 + 1 \geqslant 2\|\boldsymbol{H}_{k+1} \boldsymbol{g}_{k+1}\| \tag{25}$$

及对于所有的 j

$$\boldsymbol{g}_j^{\mathrm{T}} \boldsymbol{H}_j \boldsymbol{g}_j \,\mathrm{tr}\, \boldsymbol{B}_j + \|\boldsymbol{H}_j \boldsymbol{y}_j\|^2 \geqslant \|\boldsymbol{g}_j\|^2 + \|\boldsymbol{H}_j \boldsymbol{y}_j\|^2 \geqslant 2\|\boldsymbol{g}_j\| \|\boldsymbol{H}_j \boldsymbol{y}_j\| \tag{26}$$

对式(2)两边同乘 $\boldsymbol{g}_{k+1}$,及利用式(11)对于所有 j 有

$$\|\boldsymbol{H}_{j+1} \boldsymbol{g}_{j+1} - \boldsymbol{H}_j \boldsymbol{g}_j\| = \|\boldsymbol{V}_j \boldsymbol{H}_j \boldsymbol{y}_j - \boldsymbol{U}_j \boldsymbol{s}_j\| \leqslant \|\boldsymbol{V}_j \boldsymbol{H}_j \boldsymbol{y}_j\| + \frac{n M_1^2 \|\boldsymbol{s}_j\|^2}{\alpha_j \delta} \tag{27}$$

又由 P_k,Q_k 和 R_k(式(7)—(9))的定义及式(25)—(27),推得

$$\delta_0 (Q_k - n M_1^2 P_k / \delta) \leqslant R_k \tag{28}$$

其中,$\delta_0 = \min\{1, \delta\}$,由式(28)和引理5易知引理6结论成立.

下面证明引言中提出的定理1.

证明 假设定理不真,则必存在一个正数 $\delta>0$,使得对于所有的 j 有 $\|\boldsymbol{g}_j\| \geqslant \delta$,据此推出引理3—6成立.考虑下面的等式

$$1 = (P_k/Q_k)(Q_k/R_k)(R_k/P_k) \tag{29}$$

使 $k \to +\infty$,计算式(29)右边3个实序列的极限或上极限,得到错误结论 $1 \leqslant 0$,因此定理1为真.

3 讨论

本文分别对采用 Wolfe 条件的线搜索和修正的 Wolfe 线搜索算法进行数值实验.实验结果显示修正 Wolfe 线搜索条件下的算法要快于利用 Wolfe 条件的算法.本文比较 Wolfe 条件和修正 Wolfe 条件下 BFGS 方法的数值效果.

问题1 设

$$\varphi(x) = \sum_{j=0}^{100} [1 - \mathrm{e}^{-jh} r(jh, x)]^2,\ x \in \mathbf{R}^5$$

其中 $h=0.05$ 及

$$r(t,x) = \frac{x_1 + x_2 t + x_3 t^2}{1 + (x_4 + x_5 t)}$$

目标函数定义为

$$f(x) = \Phi(\boldsymbol{D}x)$$

其中 $\boldsymbol{D}$ 是一个 5×5 的正对角矩阵.给出变量的初始值为

$$\boldsymbol{x}_0 = (15 d_{11}^{-1}, 10 d_{22}^{-1}, 5 d_{33}^{-1}, 6 d_{44}^{-1}, -d_{55}^{-1})^{\mathrm{T}}$$

其中 $d_{ii}/d_{i+1,i+1}$ = 常数.令 $\rho=0.1$,$\theta=0.7$,$B_0=I$.终止条件为

$$|f(x_k) - f(x_*)| < 10^{-10}$$

其中该问题的最优值为

$$f(\boldsymbol{x}_*) = 3.085\,557\,482 \times 10^{-3}$$

计算结果见表1.

表1 数值试验结果

Tab.1 Numerical test results

问题	x_0	Wolfe条件		修正 Wolfe 条件	
		迭代次数	计算次数	迭代次数	计算次数
1	$15\times1,\cdots,-1\times1$	42	58	35	54
1	$15\times1,\cdots,-1\times10^4$	68	102	57	96
1	$15\times1,\cdots,-1\times10^8$	72	136	64	123
1	$15\times1,\cdots,-1\times10^{12}$	81	150	69	125
1	$15\times10^{-4},\cdots,-1\times1$	91	111	65	101
1	$15\times10^{-8},\cdots,-1\times1$	67	121	67	103
1	$15\times10^{-12},\cdots,-1\times1$	失效	失效	70	120

对非凸目标函数,未解决的 Broyden 算法收敛性问题:

(1) 并不假定产生的点列 $\{x_k\}$ 趋向于 $\boldsymbol{x}_*$,但 $\boldsymbol{x}_k - \boldsymbol{x}_{k+1} \to 0$;

(下转第298页)