关于多元非线性方程的 Broyden 方法[*1)]

安恒斌　白中治
(中国科学院数学与系统科学研究院 计算数学与科学工程计算研究所
科学与工程计算国家重点实验室, 北京, 100080)

摘　要

本文提出了求解多元非线性方程的 Broyden 方法, 讨论了该方法的局部与半局部收敛性, 并估计了其超线性收敛速度. 数值实验表明, 新方法是可行有效的, 并且其计算效率高于方向 Newton 法和方向割线法.

关键词: Newton 法, Broyden 方法, 非线性方程, 局部与半局部收敛性, 超线性收敛性
MR (2000) 主题分类: 65H10, 65Y05, 76D05

BROYDEN METHOD FOR NONLINEAR EQUATION IN SEVERAL VARIABLES

An Hengbin　Bai Zhongzhi
*(State Key Laboratory of Scientific/Engineering Computing
Institute of Computational Mathematics and Scientific/Engineering Computing
Academy of Mathematics and System Sciences
Chinese Academy of Sciences, Beijing* 100080, *P.R. China)*

Abstract

Broyden method for nonlinear equation in several variables is presented, its local and semilocal convergence properties are discussed, and its superlinear convergence rate is estimated. Numerical experiments show that the new method is feasible and effective, and its computational efficiency is much higher than both directional Newton method and directional secant method.

Keywords: Newton method, Broyden method, nonlinear equation, local and semilocal convergence, superlinear convergence
2000 Mathematics Subject Classification: 65H10, 65Y05, 76D05

* 2003 年 8 月 29 日收到.
1) 国家重点基础研究项目 "大规模科学计算研究 (G1999032803)" 专项经费资助课题.

1. 引 言

考虑 n 元非线性方程

$$f(x)=0, \tag{1}$$

其中 $f:\mathbb{R}^n\to\mathbb{R}^1$ 连续可微. 非线性方程 (1) 广泛地产生于计算机图形学, 非线性方程组的数值求解以及非解析复函数复根的计算等领域. 见 [1, 8].

文 [8] 提出了求解 (1) 的方向 Newton 法, 在该方法中, 每次迭代均从当前迭代点 x^k 出发, 去确定一个适当的方向 d^k, 然后在 d^k 上运用一维 Newton 法得到下一迭代点 x^{k+1}. 如果在方向 d^k 上运用一维割线法来得到下一迭代点, 则所得到的方法便是文 [1] 中研究的方向割线法. 在一定的条件下, 方向 Newton 法与方向割线法都是二阶收敛的. 但对于这两种方法来说, 每次迭代都要计算函数 f 在当前迭代点的导数或其差商近似, 以确定方向 d^k, 这在实际计算中是费时且很不方便的.

本文提出了一类求解非线性方程 (1) 的 Broyden 方法. 这一新方法避免了方向 Newton 法与方向割线法中每次迭代都要计算函数 f 在当前迭代点的导数或其差商近似, 以确定一个新方向的缺点, 并在实际计算中省时且易于实现. 我们证明了该方法的局部收敛性质, 半局部收敛性质, 以及超线性收敛性质. 数值实验表明, 这类 Broyden 方法的计算效率要高于方向 Newton 法和方向割线法.

2. Broyden 方法的建立

Broyden 方法是求解非线性方程组的有效方法之一 (见 [9, 7, 6]). 设 $F:\mathbb{R}^n\to\mathbb{R}^n$ 为非线性映射, 则求解非线性方程组 $F(x)=0$ 的 Broyden 算法可描述如下:

算法 2.1

1. 给定初值 $x^0\in\mathbb{R}^n$, 初始矩阵 $A_0\in\mathbb{R}^{n\times n}$ 及精度 ϵ, 令 $k:=0$.
2. 如果 $\|F(x^k)\|>\epsilon$
   - 2.1. $s^k=-A_k^{-1}F(x^k)$,
   - 2.2. $x^{k+1}=x^k+s^k$,
   - 2.3. $y^k=F(x^{k+1})-F(x^k)$,
   - 2.4. $A_{k+1}=A_k+\frac{(y^k-A_ks^k)(s^k)^T}{(s^k)^Ts^k}$,
   - 2.5. $k:=k+1$.

在算法 2.1 中, 矩阵 A_k 可视为 F 在 x^k 处的 Jacobi 矩阵 $F'(x^k)$ 的近似. 在一定条件下, 上述 Broyden 算法具有超线性收敛性质. 详见 [9, 7, 5].

若将算法 2.1 直接应用到非线性方程 (1), 则我们可得求解这类非线性方程的 Broyden 算法.

算法 2.2

1. 给定初值 x^0, 初始向量 $A_0^T \in \mathbb{R}^n$ 及精度 ϵ, 令 $k := 0$.

2. 如果 $|f(x^k)| > \epsilon$

    2.1. $s^k = -A_k^\dagger f(x^k)$,

    2.2. $x^{k+1} = x^k + s^k$,

    2.3. $y^k = f(x^{k+1}) - f(x^k)$,

    2.4. $A_{k+1} = A_k + \frac{(y^k - A_k s^k)(s^k)^T}{(s^k)^T s^k}$,

    2.5. $k := k+1$.

这里 $A_k^\dagger$ 为 A_k 的 Moore-Penrose 广义逆. 同样, 在算法 2.2 中, 向量 A_k 为 f 在 x^k 处的导数 $f'(x^k)$ 的近似. 另外, 由 Moore-Penrose 广义逆的定义, 容易验证 $A_k^\dagger = A_k^T/\|A_k\|^2$, 并且 $A_k A_k^\dagger = 1$.

在算法 2.2 中, 由于

$$s^k = -A_k^\dagger f(x^k), \quad A_k A_k^\dagger = 1,$$

故

$$A_k s^k = -f(x^k).$$

因此

$$\frac{(y^k - A_k s^k)(s^k)^T}{(s^k)^T s^k} = \frac{f(x^{k+1})(-A_k^\dagger f(x^k))^T}{\left[-A_k^\dagger f(x^k)\right]^T \left[-A_k^\dagger f(x^k)\right]} = -\frac{f(x^{k+1})}{f(x^k)} A_k.$$

利用此等式, 可得

$$\begin{aligned} A_k &= A_{k-1} - \frac{f(x^k)}{f(x^{k-1})} A_{k-1} \\ &= \frac{f(x^{k-1}) - f(x^k)}{f(x^{k-1})} A_{k-1} \\ &= \left[\prod_{i=1}^{k} \frac{f(x^{i-1}) - f(x^i)}{f(x^{i-1})}\right] A_0 \\ &\equiv \alpha_k A_0. \end{aligned}$$

这里, 我们约定 $\alpha_0 = 1$. 当 $k \geq 1$ 时, 经具体计算, 我们有

$$\begin{aligned}
x^{k+1} &= x^k - A_k^{\dagger} f(x^k) \\
&= x^0 - \sum_{j=0}^{k} A_j^{\dagger} f(x^j) \\
&= x^0 - \sum_{j=0}^{k} \alpha_j^{-1} A_0^{\dagger} f(x^j) \\
&= x^0 - \left[\sum_{j=1}^{k} \left(\prod_{i=1}^{j} \frac{f(x^{i-1})}{f(x^{i-1}) - f(x^i)} \right) f(x^j) + f(x^0) \right] A_0^{\dagger} \\
&= x^0 - \left[1 + \sum_{j=1}^{k} \left(\prod_{i=1}^{j} \frac{f(x^i)}{f(x^{i-1}) - f(x^i)} \right) \right] f(x^0) A_0^{\dagger} \\
&\equiv x^0 + \mu_{k+1} A_0^{\dagger}.
\end{aligned}$$

如果我们再约定

$$\mu_0 = 0, \quad \mu_1 = -f(x^0),$$

则有

$$x^k = x^0 + \mu_k A_0^{\dagger}, \quad k = 0, 1, 2, \ldots. \tag{2}$$

因此, 算法 2.2 所产生的点列 $\{x^k\}$ 实质上位于一维仿射子空间

$$\mathcal{S} = \{x^0 + tA_0^{\dagger} \mid t \in \mathbb{R}^1\}.$$

令

$$\Delta_k = \mu_{k+1} - \mu_k, \quad k = 0, 1, 2, \ldots,$$

则

$$\Delta_0 = -f(x^0),$$

且当 $k \geq 1$ 时, 成立

$$
\begin{aligned}
\Delta_k &= \mu_{k+1} - \mu_k \\
&= -\left[1 + \sum_{j=1}^{k}\left(\prod_{i=1}^{j}\frac{f(x^i)}{f(x^{i-1}) - f(x^i)}\right)\right] f(x^0) \\
&\quad + \left[1 + \sum_{j=1}^{k-1}\left(\prod_{i=1}^{j}\frac{f(x^i)}{f(x^{i-1}) - f(x^i)}\right)\right] f(x^0) \\
&= -\left(\prod_{i=1}^{k}\frac{f(x^i)}{f(x^{i-1}) - f(x^i)}\right) f(x^0) \\
&= -\frac{f(x^k)}{f(x^{k-1}) - f(x^k)}\left(\prod_{i=1}^{k-1}\frac{f(x^i)}{f(x^{i-1}) - f(x^i)}\right) f(x^0) \\
&= \frac{f(x^k)}{f(x^k) - f(x^{k-1})}\Delta_{k-1}.
\end{aligned}
$$

另外, 根据 (2) 可得

$$x^{k+1} = x^k + (\mu_{k+1} - \mu_k)A_0^{\dagger} = x^k + \Delta_k A_0^{\dagger}, \quad k = 0, 1, 2, \ldots.$$

由此, 我们可将算法 2.2 改写为如下更为简洁的形式:

算法 2.3

1. 给定初值 x^0, 初始向量 $A_0^T \in \mathbb{R}^n$ 及精度 ϵ.
2. 令 $\Delta_0 = -f(x^0)$. 置 $k := 0$.
3. 如果 $|f(x^k)| > \epsilon$
   - 3.1. $x^{k+1} = x^k + \Delta_k A_0^{\dagger}$,
   - 3.2. $y_k = f(x^{k+1}) - f(x^k)$,
   - 3.3. $\Delta_{k+1} = \frac{f(x^{k+1})}{y_k}\Delta_k$,
   - 3.4. $k := k + 1$.

3. 收敛性分析

本节我们讨论算法 2.2 的局部与半局部收敛性质. 为此, 我们需要先回顾一维空间上的割线法. 设 $g : \mathbb{R}^1 \to \mathbb{R}^1$, 则关于 g 的割线法所产生的迭代序列 $\{t_k\}$ 为

$$t_{k+1} = t_k - \left[\frac{g(t_k) - g(t_{k-1})}{t_k - t_{k-1}}\right]^{-1} g(t_k), \quad k = 1, 2, \ldots, \tag{3}$$

其中 t_0, t_1 为给定的初值 [7].

关于割线法 (3) 的半局部收敛性, 我们有如下结论:

引理 3.1 设 $\{t_k\}$ 是由 $g(t)$ 关于初值 t_0, t_1 的割线法所产生的迭代序列,

$$\eta = \max\{|g(t_0)|, |g(t_1)|\}, \quad B = \{t \mid |t - t_1| \le r\},$$

其中 $r = 2\max\{|t_1 - t_0|, |t_2 - t_1|\}$. 如果条件

(i) $g(t)$ 在 B 上连续可微, 并且存在正常数 β 及 L, 使得对 $\forall t, s \in B$, 都有

$$|g'(t)|^{-1} \le \beta, \quad |g'(t) - g'(s)| \le L|t - s|,$$

(ii) $\theta := \frac{25}{16}\beta^2\eta L < \frac{1}{2}$, $|t_1 - t_0| < \frac{5}{4}\beta\eta$

成立, 则存在 $t^* \in B$ 使得 $t_k \to t^* (k \to \infty)$ 并且 $g(t^*) = 0$.

证明. 对于 $k = 1, 2, \ldots$, 令

$$h_{k-1} = t_k - t_{k-1}, \quad J_k = \frac{g(t_k) - g(t_{k-1})}{h_{k-1}}.$$

由于 $t_0, t_1 \in B$, 据条件 (i) 可得

$$|J_1 - g'(t_1)| = \left|\frac{g(t_0) - g(t_1) - g'(t_1)(t_0 - t_1)}{t_0 - t_1}\right| \le \frac{1}{2}L|h_0|. \tag{4}$$

又据条件 (ii) 可得

$$\frac{1}{2}\beta L|h_0| \le \frac{5}{8}\beta^2\eta L < \frac{1}{5}.$$

再注意到 $|g'(t_1)|^{-1} \le \beta$, 所以

$$|J_1^{-1}| \le \frac{2\beta}{2 - \beta L|h_0|} \le \frac{5}{4}\beta.$$

因此,

$$|h_1| = |J_1^{-1}g(t_1)| \le \frac{5}{4}\beta\eta. \tag{5}$$

下面我们用归纳法来证明:

$$\begin{cases} |g(t_k)| \le \frac{1}{2}L|h_{k-1}|(|h_{k-1}| + |h_{k-2}|), \\ |h_k| \le \theta|h_{k-1}|, & k = 2, 3, \ldots. \\ t_{k+1} \in B. \end{cases} \tag{6}$$

当 $k = 2$ 时, 由于 $t_0, t_1, t_2 \in B$, 据条件 (i) 及估计式 (4) 可得

$$\begin{aligned} |g(t_2)| &\le |g(t_2) - g(t_1) - g'(t_1)(t_2 - t_1)| + |g(t_1) + g'(t_1)(t_2 - t_1)| \\ &\le \frac{1}{2}L|h_1|^2 + |J_1 - g'(t_1)|\,|h_1| \\ &\le \frac{1}{2}L|h_1|(|h_1| + |h_0|). \end{aligned} \tag{7}$$

再由条件 (i) 可得

$$|J_2 - g'(t_2)| \leq \frac{1}{2}L|h_1|.$$

注意到 $|g'(t_2)|^{-1} \leq \beta$, 且据估计式 (5) 及条件 (ii) 知,

$$\frac{1}{2}\beta L|h_1| \leq \frac{5}{8}\beta^2\eta L \leq \frac{1}{5}.$$

因此,

$$|J_2^{-1}| \leq \frac{5}{4}\beta. \tag{8}$$

现利用 (7)-(8), 并注意到 $|h_j| \leq \frac{5}{4}\beta\eta\,(j=0,1)$, 我们有

$$\begin{aligned} |h_2| &= |J_2^{-1}g(t_2)| \\ &\leq \frac{5}{8}\beta L|h_1|(|h_1| + |h_0|) \\ &\leq \frac{25}{16}\beta^2\eta L|h_1| \\ &= \theta|h_1|. \end{aligned}$$

从而

$$|t_3 - t_1| \leq |t_3 - t_2| + |t_2 - t_1| \leq (\theta + 1)|h_1| < r,$$

这说明 $t_3 \in B$. 以上推理表明, (6) 对于 $k=2$ 为真.

假设对于 $k = 2, 3, \ldots, m$, (6) 均成立. 于是, 我们有 $\{t_0, t_1, \ldots, t_{m+1}\} \subset B$, 并且

$$|h_j| \leq \theta|h_{j-1}|,\ j = 2, 3, \ldots, m.$$

下面我们证明当 $k = m+1$ 时, (6) 也成立.

由条件 (i) 知

$$|J_i - g'(t_i)| \leq \frac{1}{2}L|h_{i-1}|,\ i = m, m+1. \tag{9}$$

从而

$$\begin{aligned} |g(t_{m+1})| &\leq |g(t_{m+1}) - g(t_m) - g'(t_m)(t_{m+1} - t_m)| + |g(t_m) + g'(t_m)(t_{m+1} - t_m)| \\ &\leq \frac{1}{2}L|h_m|^2 + |J_m - g'(t_m)|\,|h_m| \\ &\leq \frac{1}{2}L|h_m|(|h_m| + |h_{m-1}|). \end{aligned} \tag{10}$$

再据归纳假设及估计式 (5) 可得

$$|h_m| \leq \theta^{m-1}|h_1| < \frac{5}{4}\beta\eta.$$

因此, 由条件 (ii) 可得

$$\frac{1}{2}\beta L|h_m| \leq \frac{5}{8}\beta^2\eta L \leq \frac{1}{5}.$$

注意到 $|g'(t_{m+1})|^{-1} \leq \beta$, 则由 (9) 立知

$$|J_{m+1}^{-1}| \leq \frac{5}{4}\beta. \tag{11}$$

现利用 (10)-(11), 并注意到 $|h_j| \leq |h_1| \leq \frac{5}{4}\beta\eta (j = m, m-1)$, 我们有

$$\begin{aligned} |h_{m+1}| &= |J_{m+1}^{-1}g(t_{m+1})| \\ &\leq \frac{5}{8}\beta L|h_m|(|h_m| + |h_{m-1}|) \\ &\leq \frac{25}{16}\beta^2\eta L|h_m| \\ &= \theta|h_m|. \end{aligned}$$

从而

$$\begin{aligned} |t_{m+2} - t_1| &\leq |t_{m+2} - t_{m+1}| + |t_{m+1} - t_m| + \cdots + |t_2 - t_1| \\ &\leq (\theta^m + \theta^{m-1} + \cdots + 1)|h_1| \\ &< 2|h_1| \leq r, \end{aligned}$$

这说明 $t_{m+2} \in B$. 上述推理表明, (6) 对于 $k = m + 1$ 亦为真. 因此, 由归纳法知 (6) 对于一切正整数 $k \geq 2$ 均成立.

由于当 $l > n$ 时,

$$|t_l - t_n| \leq \sum_{j=n}^{l-1} |h_j| \leq \sum_{j=n}^{l-1} \theta^j |h_1|,$$

所以, 由 $\sum_{j=0}^{\infty} \theta^j < \infty$ 可知 $\{t_k\}$ 为 B 中的 Cauchy 序列. 故存在 $t^* \in B$, 使得 $\lim_{k\to\infty} t_k = t^*$. 由 (6) 立得 $g(t^*) = 0$.

在算法 2.2 中, 设初值 x^0 和初始向量 A_0 已给定, 我们引进函数

$$\hat{f}(t) = f(x^0 + tA_0^\dagger), \quad \forall t \in \mathbb{R}^1.$$

下面的引理给出了 $\hat{f}$ 关于初值 $t_0 = 0$ 和 $t_1 = -f(x^0)$ 的割线法所产生的迭代序列 $\{t_k\}$ 与算法 2.2 所产生的点列 $\{x^k\}$ 之间的关系.

引理 3.2 设 x^0, A_0 为算法 *2.2* 中给定的初始值, $\{x^k\}$ 是由算法 *2.2* 所产生的迭代序列. 如果 $\{t_k\}$ 是 $\hat{f}$ 关于初值 $t_0 = 0$ 和 $t_1 = -f(x^0)$ 的割线法所产生的迭代序列, 则

$$x^k = x^0 + t_kA_0^\dagger, \quad k = 0, 1, 2, \ldots. \tag{12}$$

证明. 注意到 (2), 我们只需运用归纳法证明

$$t_k = \mu_k,\ k = 0, 1, 2, \ldots.$$

当 $k = 0$, 1 时, 容易验证结论成立. 假设对于 $j = 0, 1, \ldots, m$ 都成立 $\mu_j = t_j$. 下证 $\mu_{m+1} = t_{m+1}$.

由割线法的定义 (3) 可得

$$\begin{aligned} t_{m+1} - t_m &= \frac{\hat{f}(t_m)}{\hat{f}(t_{m-1}) - \hat{f}(t_m)}(t_m - t_{m-1}) \\ &= \left(\prod_{i=1}^{m} \frac{\hat{f}(t_i)}{\hat{f}(t_{i-1}) - \hat{f}(t_i)}\right)(t_1 - t_0) \\ &= -\left(\prod_{i=1}^{m} \frac{\hat{f}(t_i)}{\hat{f}(t_{i-1}) - \hat{f}(t_i)}\right)\hat{f}(t_0). \end{aligned}$$

由于 $\mu_j = t_j$, $j = 0, 1, \ldots, m$, 从而我们有

$$f(x^j) = f(x^0 + \mu_j A_0^{\dagger}) = f(x^0 + t_j A_0^{\dagger}) = \hat{f}(t_j), \quad j = 0, 1, \ldots, m.$$

于是,

$$\begin{aligned} t_{m+1} &= t_m - \left(\prod_{i=1}^{m} \frac{\hat{f}(t_i)}{\hat{f}(t_{i-1}) - \hat{f}(t_i)}\right)\hat{f}(t_0) \\ &= t_1 - \sum_{j=1}^{m}\left(\prod_{i=1}^{j} \frac{\hat{f}(t_i)}{\hat{f}(t_{i-1}) - \hat{f}(t_i)}\right)\hat{f}(t_0) \\ &= -\left[1 + \sum_{j=1}^{m}\left(\prod_{i=1}^{j} \frac{\hat{f}(t_i)}{\hat{f}(t_{i-1}) - \hat{f}(t_i)}\right)\right]\hat{f}(t_0) \\ &= -\left[1 + \sum_{j=1}^{m}\left(\prod_{i=1}^{j} \frac{f(x_i)}{f(x_{i-1}) - f(x_i)}\right)\right]f(x_0) \\ &= \mu_{m+1}. \end{aligned}$$

至此, 结论得证.

利用上述两个引理, 我们即可建立算法 2.2 的局部与半局部收敛理论.

定理 3.1 设 $f : \mathbb{R}^n \to \mathbb{R}^1$ 连续可微, $x^* \in \mathbb{R}^n$ 是非线性方程 (1) 的解, 满足 $f'(x^*) \neq 0$, 且存在常数 $L > 0$, 使得

$$\|f'(x) - f'(x^*)\| \leq L\|x - x^*\|, \quad \forall x \in \mathbb{R}^n. \tag{13}$$

如果算法 2.2 中的初值 A_0 满足

$$|1 - f'(x^*)A_0^{\dagger}| < \frac{1}{4}, \tag{14}$$

则存在 $\epsilon > 0$, 当

$$\|x^0 - x^*\| \le \epsilon \|A_0^\dagger\|$$

且

$$x^* \in \mathcal{S} := \{x^0 + \mu A_0^\dagger \mid \mu \in \mathbb{R}^1\}$$

时, 由算法 2.2 所产生的迭代序列 $\{x^k\}$ 超线性收敛于 x^*.

证明. 由 (14) 知存在 $\gamma > 0$, 满足

$$4|1 - f'(x^*)A_0^\dagger| < \gamma < 1. \tag{15}$$

选取 $\epsilon > 0$ 充分小, 使得

$$L\|A_0^\dagger\|^2\epsilon < \frac{1}{4}\gamma. \tag{16}$$

设 x^0 满足

$$\|x^0 - x^*\| \le \epsilon \|A_0^\dagger\|$$

并且

$$x^* \in \mathcal{S} := \{x^0 + \mu A_0^\dagger \mid \mu \in \mathbb{R}^1\}.$$

则存在 $t^* \in \mathbb{R}^1$, 使得

$$x^* = x^0 + t^* A_0^\dagger.$$

现令

$$\hat{f}(t) = f(x^0 + tA_0^\dagger),$$

并设 $\{t_k\}$ 是 $\hat{f}(t)$ 关于初值 $t_0 = 0$ 和 $t_1 = -f(x^0)$ 的割线法所产生的迭代序列, 则由引理 3.2 知 (12) 成立. 由此可得

$$\|x^k - x^*\| = |t_k - t^*| \, \|A_0^\dagger\|, \quad k = 0, 1, 2 \ldots.$$

因此, 为证明定理 3.1 的结论, 我们只需证明 $\{t_k\}$ 超线性收敛于 t^*.

我们首先证明序列 $\{t_k\}$ 收敛, 即

$$|t_{k+1} - t^*| \le \gamma |t_k - t^*|, \quad k = 0, 1, 2, \ldots. \tag{17}$$

由 (13) 知, 成立

$$|\hat{f}'(t) - \hat{f}'(t^*)| \le \hat{L}|t - t^*|, \quad \forall t \in \mathbb{R}^1, \tag{18}$$

其中 $\hat{L} = L\|A_0^\dagger\|^2$. 由于

$$\|x^0 - x^*\| \le \epsilon \|A_0^\dagger\|,$$

我们有

$$|t^0 - t^*| \le \epsilon. \tag{19}$$

现在, 我们用归纳法证明 (17).

当 $k=0$ 时, 由 (18), (19), (16) 和 (15) 可得

$$\begin{aligned} |t_1 - t^*| &= |\hat{f}(t_0) + t^*| \\ &= |[\hat{f}(t_0) - \hat{f}(t^*) - \hat{f}'(t^*)(t_0 - t^*)] - [1 - \hat{f}'(t^*)](t_0 - t^*)| \\ &\le \left(\frac{1}{2}\hat{L}|t_0 - t^*| + |1 - \hat{f}'(t^*)|\right)|t_0 - t^*| \\ &\le \left(\frac{1}{2}\hat{L}\epsilon + \frac{1}{4}\gamma\right)|t_0 - t^*| \\ &\le \frac{1}{2}\gamma|t_0 - t^*| \\ &\le \gamma|t_0 - t^*|. \end{aligned}$$

这说明当 $k=0$ 时 (17) 成立.

假设对 $k=0,1,\ldots,m-1$, (17) 均成立, 也即有

$$|t_j - t^*| \le \gamma|t_{j-1} - t^*| \le \cdots \le \gamma^j|t_0 - t^*| \le \gamma^j\epsilon < \epsilon, \quad k = 1,2,\ldots,m.$$

下证 $k=m$ 时 (17) 亦成立. 根据一元函数的中值定理, 存在 $\xi_m \in \mathbb{R}^1$ 介于 t_{m-1} 和 t_m 之间, 使得

$$\hat{f}(t_m) - \hat{f}(t_{m-1}) = \hat{f}'(\xi_m)(t_m - t_{m-1}). \tag{20}$$

注意到

$$|\xi_m - t^*| \le \max\{|t_m - t^*|, |t_{m-1} - t^*|\} \le \epsilon,$$

由 (18) 及 (16) 我们有

$$|\hat{f}'(\xi_m) - \hat{f}'(t^*)| \le \hat{L}|\xi_m - t^*| \le \hat{L}\epsilon \le \frac{1}{4}\gamma,$$

从而

$$\begin{aligned} \hat{f}'(\xi_m) &\ge 1 - |1 - \hat{f}'(t^*)| - |\hat{f}'(\xi_m) - \hat{f}'(t^*)| \\ &\ge 1 - \frac{1}{4}\gamma - \frac{1}{4}\gamma \\ &\ge \frac{1}{2}, \end{aligned}$$

或者

$$\left(\hat{f}'(\xi_m)\right)^{-1} \le 2.$$

再由 (16) 和 (18) 我们可得

$$\begin{aligned}
|t_{m+1}-t^*| &= \left|(t_m-t^*)-\left(\frac{\hat{f}'(t_m)-\hat{f}'(t_{m-1})}{t_m-t_{m-1}}\right)^{-1}\hat{f}(t_m)\right| \\
&= \left|(t_m-t^*)-\left(\hat{f}'(\xi_m)\right)^{-1}\hat{f}(t_m)\right| \\
&= \left(\hat{f}'(\xi_m)\right)^{-1}\Big|[\hat{f}(t^m)-\hat{f}(t^*)-\hat{f}'(t^*)(t_m-t^*)] \\
&\quad -[\hat{f}'(\xi_m)-\hat{f}'(t^*)](t_m-t^*)\Big| \\
&\le \left(\hat{f}'(\xi_m)\right)^{-1}\left(\frac{1}{2}\hat{L}|t_m-t^*|+\hat{L}|\xi_m-t^*|\right)|t_m-t^*| \\
&\le 2\left(\frac{1}{2}\hat{L}\epsilon+\hat{L}\epsilon\right)|t_m-t^*| \\
&\le \gamma|t_m-t^*|,
\end{aligned}$$

因此, 当 $k=m$ 时 (17) 亦成立. 由归纳原理, 对所有整数 $k\ge 0$, 式 (17) 均成立. 从而 $\lim_{k\to\infty}t_k=t^*$.

由 (20) 可得

$$\hat{f}(t_k)=-\hat{f}'(\xi_k)(t_{k+1}-t_k).$$

从而

$$\begin{aligned}
|\hat{f}(t_{k+1})| &\le |\hat{f}(t_{k+1})-\hat{f}(t_k)-\hat{f}'(t^*)(t_{k+1}-t_k)|+|\hat{f}(t_k)+\hat{f}'(t^*)(t_{k+1}-t_k)| \\
&\le |\hat{f}'(\xi_{k+1})-\hat{f}'(t^*)|\,|t_{k+1}-t_k|+|\hat{f}'(\xi_k)-\hat{f}'(t^*)|\,|t_{k+1}-t_k|.
\end{aligned}$$

因此

$$\lim_{k\to\infty}\left|\frac{\hat{f}(t_{k+1})}{t_{k+1}-t_k}\right| \le \lim_{k\to\infty}\left(|\hat{f}'(\xi_{k+1})-\hat{f}'(t^*)|+|\hat{f}'(\xi_k)-\hat{f}'(t^*)|\right)=0. \tag{21}$$

由 (14) 知

$$\hat{f}'(t^*)=f'(x^*)A_0^\dagger\ne 0.$$

又由于 $\hat{f}'(t)$ 是连续的, 故由反函数定理知, 存在 $\beta>0$, 使得当 k 充分大时,

$$|\hat{f}(t_{k+1})|=|\hat{f}(t_{k+1})-\hat{f}(t^*)|\ge\beta|t_{k+1}-t^*|. \tag{22}$$

现由 (21) 和 (22) 立得

$$0=\lim_{k\to\infty}\frac{\beta|t_{k+1}-t^*|}{|t_{k+1}-t_k|}\ge\lim_{k\to\infty}\frac{\beta|t_{k+1}-t^*|}{|t_{k+1}-t^*|+|t_k-t^*|}.$$

于是,

$$\lim_{k\to\infty}\frac{|t_{k+1}-t^*|}{|t_k-t^*|}=0,$$

也即序列 $\{t_k\}$ 是超线性收敛的. 从而序列 $\{x^k\}$ 超线性收敛于 x^*.

定理 3.2 设 $\{x^k\}$ 是 $f(x)$ 关于初始值 x^0, A_0 的 Broyden 迭代序列,

$$\eta = \max\{|f(x^0)|, |f(x^1)|\}, \quad B := \{x \mid \|x - x^1\| \le R\},$$

其中 $R = 2\max\{\|x^1 - x^0\|, \|x^2 - x^1\|\}$. 如果

(i) $f(x)$ 在 B 上连续可微, 并且存在正常数 β 和 L, 使得对 $\forall x, y \in B$, 都有

$$\|f'(x)A_0^\dagger\|^{-1} \le \beta, \quad \|f'(x) - f'(y)\| \le L\|x - y\|,$$

(ii) $\theta := \frac{25}{16}\beta^2\eta L\|A_0^\dagger\|^2 < \frac{1}{2}$, $|f(x^0)| < \frac{5}{4}\beta\eta$,

则存在 $x^* \in B$ 使得 $x^k \to x^*(k \to \infty)$ 且 $f(x^*) = 0$.

证明. 令

$$\hat{f}(t) = f(x^0 + tA_0^\dagger),$$

并设 $\{t_k\}$ 是由 $\hat{f}(t)$ 关于初值 $t_0 = 0$ 和 $t_1 = -f(x^0)$ 的割线法所产生的迭代序列. 令 $r = R/\|A_0^\dagger\|$, 并记

$$\hat{B} = \{t \mid |t - t_1| \le r\}, \quad \hat{L} = L\|A_0^\dagger\|^2.$$

则

(a) 对 $\forall t, s \in \hat{B}$, 若令

$$x = x^0 + tA_0^\dagger, \quad y = x^0 + sA_0^\dagger,$$

则 $x, y \in B$. 从而,

$$|\hat{f}'(t)|^{-1} = |f'(x)A_0^\dagger|^{-1} \le \beta, \quad |\hat{f}'(t) - \hat{f}'(s)| \le L\|A_0^\dagger\|^2|t - s| = \hat{L}|t - s|,$$

(b) $\theta = \frac{25}{16}\beta^2\eta\hat{L} = \frac{25}{16}\beta^2\eta L\|A_0^\dagger\|^2 < \frac{1}{2}$, $|t_1 - t_0| = |f(x^0)| < \frac{5}{4}\beta\eta$.

于是, 由引理 3.1 知, 存在 $t^* \in \hat{B}$ 使得 $t_k \to t^*(k \to \infty)$, 且 $\hat{f}(t^*) = 0$. 令

$$x^* = x^0 + t^*A_0^\dagger,$$

则 $x^* \in B$, 并由引理 3.2 知 $x^k \to x^*(k \to \infty)$, 且 $f(x^*) = \hat{f}(t^*) = 0$.

4. 数值算例

我们用两个数值例子来说明 Broyden 方法的可行性和有效性. 为此, 取非线性函数如下:

P_1. $f(x) = \sum_{i=1}^{n} x_i \exp(1 - {x_i}^2)$,

P_2. $f(x)=\sum_{i=1}^{p}(\sin x_i)^2+\sum_{i=p+1}^{n}(\tan x_i)^2$, 其中 p 是给定的正整数.

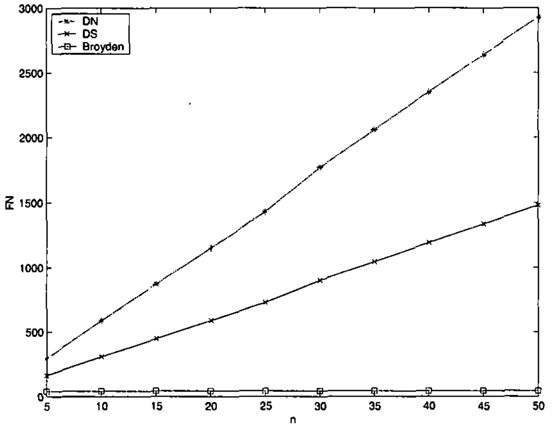
我们将 Broyden 方法与方向 Newton 法 (DN) 及方向割线法 (NS) 进行比较. 在 Broyden 方法中, 我们取初始向量 $A_0=(1,1,\dots,1)$. 在方向 Newton 法及方向割线法中, 对于每次迭代要确定的方向 d^k, 我们有两种选法 (见 [8, 1]): 一种是取 d^k 为 f 在 x^k 点的梯度方向 $\nabla f(x^k)$ 的近似, 另一种是取 d^k 为 $\nabla f(x^k)$ 的最大模分量方向, 也即 $d^k=e^{m(k)}$, 其中 $e^{m(k)}$ 是满足 $|\nabla f(x^k)^T e^{m(k)}|=\max_j|\nabla f(x^k)^T e^j|$ 的第 $m(k)$ 个单位向量. 这里, e^j 是第 j 个单位向量. 在计算中, 我们取 $d^k=\nabla f(x^k)/\|\nabla f(x^k)\|$. 由于每次迭代的主要计算量是函数的赋值, 所以我们对函数的赋值次数 FN 进行了统计. 另外, 我们也记录了总迭代次数 IN 和整个迭代过程所花费的时间 CPU. 我们所采用的停机准则是 $|f(x^k)|<10^{-12}$; 如果迭代中断, 则相应数据用 * 表示.

表 1: 问题 P_1 的计算结果

n	x^0	方向 Newton 法			方向割线法			Broyden 算法		
		IN	FN	CPU	IN	FN	CPU	IN	FN	CPU
20	(2,2,...,2)	28	1149	3.580623	28	589	1.950703	41	42	0.198288
	(-3,-3,...,-3)	24	985	3.165384	24	505	1.555148	35	36	0.119985
	(2,-3,...,2,-3)	7	288	0.755920	6	127	0.461275	8	9	0.020466
30	(2,2,...,2)	29	1770	7.853413	29	900	4.287339	42	43	0.226102
	(-3,-3,...,-3)	24	1465	6.385595	24	745	3.250481	35	36	0.153434
	(2,-3,...,2,-3)	7	428	1.748803	6	187	0.831413	8	9	0.016127
50	(2,2,...,2)	29	2930	19.392876	29	1480	9.850866	43	44	0.263423
	(-3,-3,...,-3)	25	2526	16.600638	25	1276	8.503346	36	37	0.215380
	(2,-3,...,2,-3)	7	708	4.666781	6	307	2.123100	8	9	0.146402
100	(2,2,...,2)	30	6031	80.043007	30	3031	42.787159	44	45	0.640961
	(-3,-3,...,-3)	25	5026	63.654084	25	2526	31.383999	37	38	0.468887
	(2,-3,...,2,-3)	7	1408	17.476215	14	1415	17.275244	8	9	0.047878

从表 1 可以看出, 尽管 Broyden 方法的迭代次数比方向 Newton 法及方向割线法多, 但它所需要的函数赋值次数却要远小于这两种方法. 从而, 它所花费的计算时间也相对较少. 随着维数 n 的增大, 这种差别也变得越来越明显. 这表明 Broyden 方法的计算效率比方向 Newton 法和方向割线法更高. 同时, 从表 1 也可以看出, 方向割线法要优于方向 Newton 法.

另外, 从表 1 可见, 随着维数 n 的增大, Broyden 方法所需的函数赋值次数不会增多, 而方向 Newton 法和方向割线法则不然. 为了进一步验证这一观察结果, 我们对于初值 $x^0=(2,2,\dots,2)^T\in\mathbb{R}^n$ 及 $x^0=(-3,-3,\dots,-3)^T\in\mathbb{R}^n$ 就不同的 n 进行了实验. 结果见图 1. 由图 1 可以看出, 方向 Newton 法所需的函数赋值次数几乎是方向割线法的两倍, 且这两种



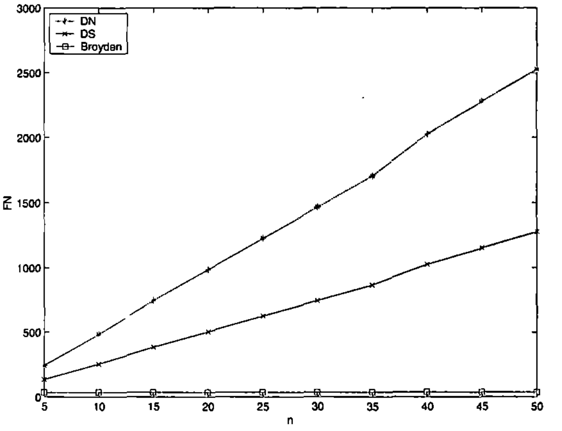


图 1: 问题 P_1 的函数赋值次数随 n 的变化情况. 左图 : $x^0 = (2,2,\ldots,2)^T$, 右图 : $x^0 = (-3,-3,\ldots,-3)^T$.

表 2: 问题 P_2 的计算结果

n	p	x^0	方向 Newton 法			方向割线法			Broyden 算法		
			IN	FN	CPU	IN	FN	CPU	IN	FN	CPU
20	5	$(\pi/6,\pi/6,...,\pi/6)$	22	903	2.764673	30	631	2.094130	30	31	0.122001
		$(\pi/4,\pi/4,...,\pi/4)$	23	944	2.840832	42	883	2.988117	29	30	0.051606
		$(\pi/3,\pi/3,...,\pi/3)$	24	985	3.057095	42	883	3.045376	35	36	0.167579
	10	$(\pi/6,\pi/6,...,\pi/6)$	22	903	2.884767	38	799	2.767601	31	32	0.156615
		$(\pi/4,\pi/4,...,\pi/4)$	23	944	3.080914	24	505	1.614163	27	28	0.037458
		$(\pi/3,\pi/3,...,\pi/3)$	24	985	3.235351	41	862	2.896248	37	38	0.095222
	15	$(\pi/6,\pi/6,...,\pi/6)$	22	903	2.912881	33	694	2.285031	31	32	0.153087
		$(\pi/4,\pi/4,...,\pi/4)$	23	944	3.182450	23	484	1.499978	30	31	0.142264
		$(\pi/3,\pi/3,...,\pi/3)$	24	985	3.096975	*	*	*.*	32	33	0.075054
50	15	$(\pi/6,\pi/6,...,\pi/6)$	23	2324	13.407992	*	*	*.*	31	32	0.224285
		$(\pi/4,\pi/4,...,\pi/4)$	24	2425	14.138890	53	2704	17.808470	29	30	0.270762
		$(\pi/3,\pi/3,...,\pi/3)$	25	2526	15.428487	44	2245	13.759631	42	43	0.295242
	30	$(\pi/6,\pi/6,...,\pi/6)$	22	2223	12.919527	28	1429	9.429112	31	32	0.216353
		$(\pi/4,\pi/4,...,\pi/4)$	23	2324	14.695658	33	1684	14.073142	30	31	0.245001
		$(\pi/3,\pi/3,...,\pi/3)$	24	2425	14.754159	42	2143	12.811004	35	36	0.194560
	45	$(\pi/6,\pi/6,...,\pi/6)$	22	2223	12.502604	*	*	*.*	32	33	0.259986
		$(\pi/4,\pi/4,...,\pi/4)$	23	2324	13.725329	24	1225	7.340682	32	33	0.111659
		$(\pi/3,\pi/3,...,\pi/3)$	23	2324	13.087125	30	1531	8.838014	29	30	0.274326

方法所需的函数赋值次数都随着 n 的增大而大致呈线性增长. 而 Broyden 方法的函数赋值次数所形成的曲线几乎是水平的, 这说明 Broyden 方法所需的函数赋值次数不会随 n 的增大而增多. 因此, Broyden 方法对于求解高维非线性方程更为有效.

问题 P_2 的计算结果见表 2. 从表 2 我们可以得出同样的结论: Broyden 方法要比方向 Newton 法和方向割线法具有更高的计算效率, 这表现在它的函数赋值次数 FN 及迭代时间 CPU 均为最少. 同时也可以看出, 参数 p 对于方向 Newton 法和 Broyden 方法的影响较小, 而对于方向割线法的影响则较大.

5. 结 论

我们提出了求解 n 个变量的非线性方程的 Broyden 方法. 该方法是经由推广经典的关于非线性方程组的 Broyden 方法而得到的. 在对非线性函数和迭代初值的适当假定下, 我们证明了这类 Broyden 方法的局部超线性收敛性和半局部收敛性. 数值实验表明, 这类 Broyden 方法比方向 Newton 法及方向割线法更为有效, 特别是对于高维问题, Broyden 方法的优越性更为明显.

参 考 文 献

[1] Heng-Bin An and Zhong-Zhi Bai, Directional secant method for nonlinear equations, *J. Comp. Appl, Math.*, (2004).

[2] A. Ben–Israel, A Newton-Raphon method for the solution of systems of equations, *J. Math. Anal. Appl.*, 15(1966), 243–252.

[3] A. Ben-Israel, Newton's method with modified functions, *Contemp. Math.*, 204(1997), 39–50.

[4] J.E. Dennis, Jr., On the convergence of Broyden's method for nonlinear systems of equations, *Math. Comp.*, 25(1971), 559–567.

[5] J.E. Dennis, Jr. and J.J. Moré, Quasi-Newton methods, motivation and theory, *SIAM Rev.*, 19(1977), 46–89.

[6] J.E. Dennis, Jr. and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[7] 冯果忱, 非线性方程组的迭代解法, 上海科学技术出版社, 上海, 1989.

[8] Y. Levin and A. Ben-Israel, Directional Newton method in n variables, *Math. Comp.*, 71(2001), 251–262.

[9] 李庆扬, 莫孜中, 祁力群, 非线性方程组的数值解法, 科学出版社, 北京, 1999.

[10] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, SIAM, Philadelphia, 2000.