中国农业大学学报　1997,2(2):25～28
Journal of China Agricultural University

牛顿与近似牛顿混合算法的收敛速率①

王　兆　智②
（中国农业大学工程基础科学部）

摘　要　研究交替使用牛顿迭代与近似牛顿迭代求解无约束最优化问题的混合算法。研究这一类算法的收敛特征，得到了两步超二阶收敛和两步至少 $2(1+p)(0<p\leqslant 1)$ 阶收敛的充要条件；对单和函数的极小化问题，给出了具有二步四阶收敛速率的混合算法。

关键词　牛顿迭代；近似牛顿迭代；收敛速率；单和函数
中图分类号　O221.2

Convergence Rate of Mixed Newton and Inexact Newton Method

Wang Zhaozhi
(College of Applied Engineering Sciences, CAU)

Abstract　The algorithms exploiting alternatively Newton iteration and inexact Newton iteration to solve the unconstrained optimization problems are considered. Their convergence characteristic is studied. The sufficient and necessary conditions for convergence with super-two order and $2(1+p)(0<p\leqslant 1)$ order in two steps are given. A mixed algorithm with 4 order convergence rate in two steps for unary optimization is constructed.

Key words　Newton iteration; inexact Newton iteration; convergence rate; unary function

考虑无约束非线性优化问题

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

文献[1]给出了一种牛顿与近似牛顿混合迭代算法，数值结果与牛顿法相当，但两步只需计算一次二阶海赛阵。遗憾的是在理论上，它只对一维情形证明了该算法具有二步四阶收敛速率这一理想的结论。为深化与推广该文的结果，本文考虑更广的一类算法模型，即交替使用牛顿迭代与近似牛顿迭代的混合算法模型，研究该模型的收敛速率。

混合算法模型可描述如下。

算法 1

1)给定初始点 x^0，误差 $\varepsilon>0, k=0$。

2)若 $\|\nabla f(x^k)\|<\varepsilon$，停止；否则转步 3)。

收稿日期:1996-08-21
①国家自然科学基金项目，北京市自然科学基金项目
②王兆智，北京清华东路 17 号中国农业大学(东校区)71 信箱，100083

3)构造近似海赛阵 $\boldsymbol{B}^k$:

3a)当 $k=2m$ 时,$\boldsymbol{B}^k=\nabla^2 f(x^k)$;

3b)当 $k=2m+1$ 时,$\boldsymbol{B}^k$ 按某种规则由 $\boldsymbol{B}^{k-1}$ 修正得到。

4)$x^{k+1}=x^k-(\boldsymbol{B}^k)^{-1}\nabla f(x^k)$,$k=k+1$,转步 2)。

此算法是偶数次迭代时用牛顿法,奇数次时用近似牛顿法,称之为混合迭代算法。

下面第 1 节给出了算法 1 局部收敛的条件和局部收敛速率的特征,其中推论 3 是为保证 $2(1+p)$ 阶收敛 $\boldsymbol{B}^k$ 应满足的条件,它对构造算法非常有用;第 2 节给出了单和函数一个具有二步四阶收敛速率的混合算法:这样就把文献[1]的理论结果由一维推广到了一类特殊的 n 维问题。

假设 $f(x)$ 二次连续可微,x^* 满足 $\nabla f(x)=0$ 要求,$\nabla^2 f(x^*)$ 非奇异。记号 $y^k=o(z^k)$ 表示 $\lim\limits_{k\to\infty}\sup(y^k/z^k)=0$,$y^k=O(z^k)$ 表示 $\lim\limits_{k\to\infty}\sup(y^k/z^k)<+\infty$。

1 收敛速率的特征

记 $r^k=\nabla^2 f(x^k)s^k+\nabla f(x^k)$,其中,$s^k=-(\boldsymbol{B}^k)^{-1}\nabla f(x^k)$。

下面的定理 1 将给出算法 1 局部收敛的条件,定理 2 则揭示了算法 1 两步超二阶和两步至少 $2(1+p)(0<p\leqslant 1)$ 阶收敛的特征。

定理 1 设 $\{x^k\}$ 由算法 1 产生。若奇数次迭代的修正矩阵 $\boldsymbol{B}^k$ 满足 $\|r^k\|\leqslant\eta\|\nabla f(x^k)\|$,$\eta<1$,则存在 $\varepsilon>0$,当 $\|x^0-x^*\|<\varepsilon$ 时,x^k 收敛于 x^*。

证明 由于偶数次迭代时用的是牛顿法,由文献[2]中的定理 2.3 直接得到。证毕。

定义 1 $f(x)$ 的二阶海赛阵 $\nabla^2 f(x)$ 在 x^* 处 $p(0<p\leqslant 1)$ 次 Hölder 连续是指存在 $L>0$,当 $\|y-x^*\|$ 充分小时,有

$$\|\nabla^2 f(y)-\nabla^2 f(x^*)\|\leqslant L\|y-x^*\|^p$$

定义 2 两步超二阶收敛是指 $\|x^{k+1}-x^*\|=o(\|x^{k-1}-x^*\|^2)$,$k$ 为奇数,$k\to\infty$。

定义 3 两步至少 $2(1+p)(0<p\leqslant 1)$ 阶收敛是指 $\|x^{k+1}-x^*\|=O(\|x^{k-1}-x^*\|^{2(1+p)})$,$k$ 为奇数,$k\to\infty$。

定理 2 设 $\{x^k\}$ 由算法 1 产生,且收敛于 x^*,则 x^k-x^* 是两步超二阶收敛的充要条件是:k 为奇数时,$\|r^k\|=o(\|\nabla f(x^{k-1})\|^2)$。进一步,若 $\nabla^2 f(x)$ 在 x^* 处 $p(0<p\leqslant 1)$ 次 Hölder 连续,$x^k\to x^*$ 两步至少 $2(1+p)$ 阶的充要条件是:k 为奇数时,$\|r^k\|=O(\|\nabla f(x^{k-1})\|^{2(1+p)})$。

证明 两结论的证明类似,只证明后面的结论。

必要性。若 x^k 收敛于 x^* 两步至少 $2(1+p)$ 阶,当 k 为奇数时,

$$\begin{aligned}\|r^k\|=&\|\nabla^2 f(x^k)s^k+\nabla f(x^k)\|=\|[\nabla f(x^k)-\nabla f(x^*)-\nabla^2 f(x^*)(x^k-x^*)]-\\&[\nabla^2 f(x^k)-\nabla^2 f(x^*)](x^k-x^*)+[\nabla^2 f(x^*)+\nabla^2 f(x^k)-\nabla^2 f(x^*)](x^{k+1}-x^*)\|\leqslant\\&\|\nabla f(x^k)-\nabla f(x^*)-\nabla^2 f(x^*)(x^k-x^*)\|+\|\nabla^2 f(x^k)-\nabla^2 f(x^*)\|\|x^k-x^*\|+\\&[\|\nabla^2 f(x^*)\|+\|\nabla^2 f(x^k)-\nabla^2 f(x^*)\|]\|x^{k+1}-x^*\|\end{aligned}$$

利用文献[2]引理 3.2,$\nabla^2 f(x)$ 在 x^* 的 Hölder 连续性假设,以及二步至少 $2(1+p)$ 阶收敛速率,得

$$\|r^k\|=O(\|x^k-x^*\|^{1+p})+O(\|x^k-x^*\|^p)\|x^k-x^*\|+O(\|x^{k-1}-x^*\|^{2(1+p)})$$

又,第 $k-1$ 步为牛顿步,即 $\|x^k-x^*\|<c\|x^{k-1}-x^*\|^2$,$c>0$,及文献[2]引理 3.1 得

$$\|r^k\|=O(\|x^{k-1}-x^*\|^{2(1+p)})=O(\|\nabla f(x^{k-1})\|^{2(1+p)})$$

充分性。若$\|r^k\|=O(\|\nabla f(x^{k-1})\|^{2(1+p)})$,则

$$\|x^{k+1}-x^*\|=\|x^k+s^k-x^*\|=\|x^k+\nabla^2 f(x^k)^{-1}(r^k-\nabla f(x^k))-x^*\|=$$
$$\|[\nabla^2 f(x^*)^{-1}+\nabla^2 f(x^k)^{-1}-\nabla^2 f(x^*)^{-1}][r^k+(\nabla^2 f(x^k)-\nabla^2 f(x^*))(x^k-$$
$$x^*)-(\nabla f(x^k)-\nabla f(x^*)-\nabla^2 f(x^*)(x^k-x^*))]\|\leqslant[\|\nabla^2 f(x^*)^{-1}\|+$$
$$\|\nabla^2 f(x^k)^{-1}-\nabla^2 f(x^*)^{-1}\|][\|r^k\|+\|\nabla^2 f(x^k)-\nabla^2 f(x^*)\|\|x^k-$$
$$x^*\|+\|(\nabla f(x^k)+\nabla f(x^*)-\nabla^2 f(x^*)(x^k-x^*)\|]$$

根据文献[2]引理 2.1 题设中 r^k 的条件,$\nabla^2 f(x)$在 x^* 的连续性,得

$$\|x^{k+1}-x^*\|=O(\|\nabla f(x^{k-1})\|^{2(1+p)})+O(\|x^k-x^*\|^{1+p})+O(\|x^k-x^*\|^{1+p})$$

由于第 $k-1$ 步为牛顿步,以及依文献[2]引理 3.1,得

$$\|x^{k+1}-x^*\|=O(\|x^{k-1}-x^*\|^{2(1+p)})$$

证毕。

下面的推论 3 是算法 1 二步至少 $2(1+p)$阶收敛的一个充分条件。

推论 3　考虑定理 2 中的点列$\{x^k\}$,若 k 为奇数时,存在 $M>0$,使$\|\boldsymbol{B}^k-\nabla^2 f(x^k)\|\leqslant M\|s^{k-1}\|^{2p}$,其中 $s^{k-1}=x^k-x^{k-1}$,则 $x^k\to x^*$ 两步至少为 $2(1+p)(0<p\leqslant 1)$阶。

证明　$r^k=\nabla^2 f(x^k)s^k+\nabla f(x^k)=-\nabla^2 f(x^k)(\boldsymbol{B}^k)^{-1}\nabla f(x^k)+\nabla f(x^k)=[(\boldsymbol{B}^k-\nabla^2 f(x^k))(\boldsymbol{B}^k)^{-1}]\nabla f(x^k)$。由于$\nabla^2 f(x^*)$非奇异,当 k 充分大时$\|(\boldsymbol{B}^k)^{-1}\|$有界,故存在 $M_1>0$,$\|r^k\|\leqslant M_1M\|s^{k-1}\|^{2p}\|\nabla f(x^k)\|\leqslant 2M_1M\|x^{k-1}-x^*\|^{2p}\|\nabla f(x^k)\|$,由文献[2]引理 3.1,存在 $M_2>0$,使$\|r^k\|\leqslant M_2\|\nabla f(x^{k-1})\|^{2p}\|\nabla f(x^{k-1})\|^2=M_2\|\nabla f(x^{k-1})\|^{2(1+p)}$。由定理 2,$x^k\to x^*$ 两步至少 $2(1+p)(0<p\leqslant 1)$阶。证毕。

2　单和函数混合迭代算法

函数

$$f(x)=\sum_{i=1}^{m}u[\alpha_i(x)]$$

称为单和函数,这里 $\alpha_i(x)=a_i^{\mathrm{T}}x$,$a_i$ 是 n 维向量,$x\in\mathbf{R}^n$。这种函数在实际问题中常见于文献[3,4]。$f(x)$的梯度和二阶海赛阵分别是

$$\nabla f(x)=\sum_{i=1}^{m}\delta_i(\alpha_i)a_i$$

$$\nabla^2 f(x)=\sum_{i=1}^{m}\Phi_i(\alpha_i)a_ia_i^{\mathrm{T}}$$

其中 $\alpha_i=a_i^{\mathrm{T}}x$,$\delta_i(\alpha_i)=\mathrm{d}u_i[\alpha_i(x)]/\mathrm{d}\alpha_i$,$\Phi_i(\alpha_i)=\mathrm{d}^2u_i[\alpha_i(x)]/\mathrm{d}\alpha_i^2$,记 $\theta_i(\alpha_i)=\mathrm{d}^3u_i[\alpha_i(x)]/\mathrm{d}\alpha_i^3]$,则可构造下面的算法。

算法 2

步骤同算法 1,只是第 3 步的 $\boldsymbol{B}^k$ 为

$$\boldsymbol{B}^k=\sum_{i=1}^{m}\Phi_i^k a_ia_i^{\mathrm{T}}$$

其中:k 为偶数时,$\Phi_i^k=\Phi_i(\alpha_i^k)$;$k$ 为奇数时,$\Phi_i^k=-\Phi_i^{k-1}+[2\delta_i(\alpha_i^k)-2\delta_i(\alpha_i^{k-1})]/(\alpha_i^k-\alpha_i^{k-1})$。

定理 4　$\{x^k\}$由算法 2 产生,$u'''(\alpha_i)$满足 Lipschitz 连续条件,则 $x^k\to x^*$ 是两步 4 阶。

证明 由泰勒展式，有

$$\delta_i^k=\delta_i^{k-1}+\Phi_i(\alpha_i^{k-1})l_i^{k-1}+(1/2)\theta_i(\zeta_i^{k-1})(l_i^{k-1})^2$$

$$\delta_i^{k-1}=\delta_i^k-\Phi_i(\alpha_i^k)l_i^{k-1}+(1/2)\theta_i(\zeta_i^k)(l_i^{k-1})^2$$

这里 $l_i^{k-1}=\alpha_i^k-\alpha_i^{k-1}$，$\zeta_i^{k-1}$，$\zeta_i^k$ 位于 $a_i^{\mathrm{T}}x^k$ 与 $a_i^{\mathrm{T}}x^{k-1}$之间。上面两式相减，得

$$2(\delta_i^k-\delta_i^{k-1})=[\Phi_i(\alpha_i^k)+\Phi_i(\alpha_i^{k-1})]l_i^{k-1}-(1/2)[\theta_i(\zeta_i^{k-1})-\theta_i(\zeta_i^k)](l_i^{k-1})^2$$

$$\Phi_i^k-\Phi_i(\alpha_i^k)=\Phi_i^{k-1}+[2\delta_i(\alpha_i^k)-2\delta_i(\alpha_i^{k-1})]/(\alpha_i^k-\alpha_i^{k-1})-\Phi_i(\alpha_i^k)=$$

$$\Phi_i(\alpha_i^k)+\Phi_i(\alpha_i^{k-1})+(1/2)[\theta_i(\zeta_i^{k-1})-\theta_i(\zeta_i^k)]l_i^{k-1}$$

由 $u'''(a_i)$的 Lipschitz 连续性，$l_i^{k-1}=a_i^{\mathrm{T}}s^{(k-1)}$，存在 $M_3>0$，使

$$|\Phi_i^k-\Phi_i(\alpha_i^k)|\leqslant M_3\|x^k-x^{k-1}\|\|s^{k-1}\|=M_3\|s^{k-1}\|^2$$

从而存在 $M_4>0$，使

$$\|\boldsymbol{B}^k-\nabla^2 f(x^k)\|=\|\sum_{i=1}^{m}[\Phi_i^k-\Phi_i(\alpha_i^k)]a_ia_i^{\mathrm{T}}\|\leqslant M_4\|s^{k-1}\|^2$$

由推论 3，结论得证。证毕。

定理 4 把文献[1]的理论结果由一维推广到单和函数这一特殊的 n 维问题。至于如何对一般函数 $f(x)$，构造二步四阶收敛的算法还有待进一步研究。

参 考 文 献

1 林正华. 牛顿与二阶拟牛顿混合迭代算法. 高等学校计算数学学报，1994，16(3)：217～224

2 Dembo R S, Eisenstat S C, Steihaug T. Inexact Newton methods. SIAM J Numer Anal, 1982, 19(2):400～408

3 Goldfarb D, Wang S. Partial update Newton methods for unary factorable and partially separable optimization. SIAM J Optimization, 1993, 3(2):382～397

4 Chen L H, Deng N Y, Zhang J Z. A modified partial-update algarithm for unary optimization. to appear in JOTA