中国农业大学学报　　2000, 5(4): 20～25
Journal of China Agricultural University

非线性最小二乘问题的结构 p 步牛顿法[①]

王来生[②]　甄　苓
(中国农业大学工程基础科学部)

摘　要　给出了非线性最小二乘问题的结构 p 步牛顿算法, 分析了该算法的效率, 结果表明, 对零残差问题新算法具有 q-2 阶收敛速率, 与牛顿法具有相同的收敛速率, 由于新算法只需计算近似海赛矩阵, 所以, 其效应比牛顿法高; 对于非零残差问题新算法具有 p 步 $p+1$ 阶收敛速率, 其效率至少与牛顿法相同。

关键词　最小二乘问题; p 步牛顿法; 效率
分类号　O 221.2

Structured p-step Newton Algorithm for Nonlinear Least Square Problems

Wang Laisheng　Zhen Ling
(College of Applied Engineering Sciences, CAU)

Abstract　A structured p-step Newton algorithm for nonlinear least square problems is developed. The efficiency of the algorithm is analysed. For zero residual problem its convergence rate is q-2 order, which is the same with that of Newton algorithm. Since the Hessian matrix is calculated appropriately, the efficiency of this algorithm is higher than that of Newton method; For non-zero residual problem its convergence rate is p-step $p+1$ order, the efficiency of this algorithm is at least as high as that of Newton method.

Key words　Least square problem; p-step Newton algorithm; efficiency

考虑非线性最小二乘问题

$$\min_{x\in\mathbf{R}^n} f(x)=\frac{1}{2}\|R(x)\|^2=\frac{1}{2}\sum_{i=1}^{m}[r_i(x)]^2\quad(m>n)\tag{1}$$

其中

$$R(x)=[r_1(x),r_2(x),\cdots,r_m(x)]^{\mathrm T}$$

这里 $r_i(x)$ $(i=1,2,\cdots,m)$ 为一般非线性函数, $R(x)$ 用来表示点 x 处的残差。记作:

$$J(x)=[\partial r_i(x)/\partial x_j]$$

$$g(x)=J(x)^{\mathrm T}R(x)$$

$$H(x)=\nabla^2 f(x)=J(x)^{\mathrm T}J(x)+\sum_{i=1}^{m}r_i(x)\nabla^2 r_i(x)=C(x)+S(x)$$

其中: $J(x)$ 表示 $R(x)$ 的 Jacobi 矩阵, $g(x)$ 表示 $f(x)$ 的梯度, $H(x)$ 是 $f(x)$ 的海赛矩阵,

收稿日期: 1999-11-30
①国家自然科学基金资助项目
②王来生, 北京清华东路 17 号 中国农业大学(东校区)71 信箱, 100083

$C(x)=J(x)^{T}J(x)$, $S(x)=\sum_{i=1}^{m} r_i(x)\nabla^2 r_i(x)$。

笔者针对非线性最小二乘问题[1, 2], 构造了结构 p 步牛顿法。下面, 首先给出结构 p 步牛顿迭代算法, 然后进行收敛性证明。

1　结构 p 步牛顿法

结构 p 步牛顿法算法如下。

第 1 步, 给定初始点 x^0, 整数 $p\geq 1$, 置 $k=0$;

第 2 步, 若 $\|\nabla f(x^k)\|=0$, 停, 否则, 转下步;

第 3 步, 令

$$\left.\begin{aligned}
&x^{k,0}=x^k\\
&B^{k,0}=\nabla^2 f(x^{k,0})\\
&x^{k,1}=x^{k,0}-(B^{k,0})^{-1}\nabla f(x^{k,0})\\
&B^{k,i}=J(x^{k,i})^{T}J(x^{k,i})+\|R(x^{k,i})\|\frac{S(x^{k,0})}{\|R(x^{k,0})\|}\\
&x^{k,i+1}=x^{k,i}-(B^{k,i})^{-1}\nabla f(x^{k,i}),\ i=1,2,\cdots,p-1\\
&x^{k+1}=x^{k,p}
\end{aligned}\right\}\tag{2}$$

第 4 步, 令 $k=k+1$, 转入第 2 步。

算法产生的点列为 $\{x^k\}$。为了简单而直观, 第 3 步用双上标来表示。第 2 个上标为 0 时表示算法在 x^k 点, 即 $x^{k,0}$ 处采用牛顿步, 当 $0<i<p$ 时表示在 $x^{k,i}$ 处采用简化结构牛顿步, 算法中参数 p 是第 3 步中子迭代的次数。从 x^k 到 x^{k+1} 的 p 步中, 对海赛矩阵 $H(x)$ 中 $S(x)$ 部分只计算一次, 得 $S(x^k)$。由于 $H(x)$ 的线性部分 $J(x)^{T}J(x)$ 相对容易计算, 从计算量上讲, 由式(2)构造 $B^{k,i}$ 比直接计算容易得多。

2　p 步牛顿法的收敛性

为了考虑算法的收敛性, 做如下假设:

A 1　x^* 是问题(1)的解;

A 2　记 $D=\{x\mid \|x-x^*\|<\epsilon\}$, 其中 ϵ 是充分小的正数。在 D 内 $H(x)$, $C(x)$, $J(x)$, $R(x)$ 及 $S(x)$ 为 Lipschitz 连续, Lipschitz 常数为 L;

A 3　$\nabla^2 f(x^*)$ 是对称正定矩阵。

下面定理 1 说明上述算法中由式(2)给定的近似海赛矩阵是正定的, 从而算法可行。

定理 1　在假设(A 1～A 3)成立的条件下, 则存在正数 δ_1, 当 $\|x_1-x^*\|$ 和 $\|x_2-x^*\|$ 同时小于 δ_1 时, 矩阵

$$B=J(x^1)^{T}J(x^1)+\|R(x^1)\|\frac{S(x^2)}{\|R(x^2)\|}$$

正定。

证明　因为 $\nabla^2 f(x^*)$ 正定, 所以存在数 $M_1>0$, 使任意 $y\in\mathbf{R}^n$ 且 $y\neq\mathbf{0}$, 有

$$|y^{T}\nabla^2 f(x^*)y|>M_1\|y\|^2\tag{3}$$

又

$$\left|y^{T}(B-\nabla^{2}f(x^{*}))y\right|\leq\|y\|^{2}\|(B-\nabla^{2}f(x^{*}))\|\leq$$

$$\|y\|^{2}\left(\|C(x_{1})-C(x^{*})\|+\left\|\frac{\|R(x_{1})\|}{\|R(x_{2})\|}S(x_{2})-S(x^{*})\right\|\right) \tag{4}$$

情况1: 若$R(x^{*})=0$, 则$S(x^{*})=0$。这时, 存在正数δ_1^1和M_2, 当$\|x_1-x^{*}\|$和$\|x_2-x^{*}\|$同时小于δ_1^1时, 有

$$\|\nabla^{2}r_{i}(x_{2})\|\leq M_{2}$$

注意到

$$\frac{|r_{i}(x_{2})|}{\|R(x_{2})\|}\leq 1$$

从而有

$$\left\|\frac{\|R(x_{1})\|}{\|R(x_{2})\|}S(x_{2})\right\|=\left\|\|R(x_{1})\|\sum_{i=1}^{m}\frac{r_{i}(x)}{\|R(x_{2})\|}\nabla^{2}r_{i}(x)\right\|\leq mM_{2}\|R(x_{1})\| \tag{5}$$

由假设A2, 式(4)和(5)可得

$$\left|y^{T}(B-\nabla^{2}f(x^{*}))y\right|\leq\|y\|^{2}(\|C(x_{1})-C(x^{*})\|+mM_{2}\|R(x_{1})\|)=$$

$$\|y\|^{2}(\|C(x_{1})-C(x^{*})\|+mM_{2}\|R(x_{1})-R(x^{*})\|)\leq$$

$$L(1+mM_{2})\|x_{1}-x^{*}\|\|y\|^{2} \tag{6}$$

情况2: 若$R(x^{*})\neq 0$, 则存在正数δ_1^2, M_3和M_4, 当$\|x_1-x^{*}\|$和$\|x_2-x^{*}\|$同时小于δ_1^2时, 有

$$M_{3}\leq\|R(x_{1})\|\leq M_{4}$$

$$M_{3}\leq\|R(x_{2})\|\leq M_{4}$$

$$\left\|\frac{\|R(x_{1})\|}{\|R(x_{2})\|}S(x_{2})-S(x^{*})\right\|=$$

$$\frac{\left\|\|R(x_{1})\|(S(x_{2})-S(x^{*}))-(\|R(x_{2})\|-\|R(x_{1})\|)S(x^{*})\right\|}{\|R(x_{2})\|}\leq$$

$$\frac{M_{4}}{M_{3}}L\|x_{2}-x^{*}\|+\frac{M_{4}}{M_{3}}L(\|x_{2}-x^{*}\|+\|x_{1}-x^{*}\|) \tag{7}$$

由式(4)和(7)得

$$\left|y^{T}(B-\nabla^{2}f(x^{*}))y\right|\leq(1+\frac{M_{4}}{M_{3}})L\|x_{1}-x^{*}\|\|y\|^{2}+2\frac{M_{4}}{M_{3}}L\|x_{2}-x^{*}\|\|y\|^{2} \tag{8}$$

综合上述2种情况, 由式(6)和式(8)可知, 存在

$$\delta_{1}<\min\{\delta_1^1,\delta_1^2\}$$

当$\|x_1-x^{*}\|$和$\|x_2-x^{*}\|$同时小于δ_1时, 有

$$\left|y^{T}(B-\nabla^{2}f(x^{*}))y\right|\leq\frac{1}{2}M_{1}\|y\|^{2}$$

结合式(3)得

$$y^{T}By\geq\frac{1}{2}M_{1}\|y\|^{2}$$

即B正定。

定理 2　在假设 A1～A3 成立的条件下, 则存在正数 δ_2, 当 $\|x^k - x^*\| < \delta_2$ 时, 算法中第 3 步产生的点 $x^{k,i+1}$ 满足

$$\|x^{k,i+1} - x^*\| < K\|x^{k,i} - x^*\|,\ i = 0, 1, \cdots, p-1 \tag{9}$$

其中 $0 \leq K \leq 1$。

证明　当 $i = 0$ 时, $x^{k,1}$ 由牛顿步产生, 从而存在数 $\delta_2^1 > 0$ 和 $0 \leq K_1 \leq 1$, 使得当 $\|x^k - x^*\| < \delta_2^1$ 时, 有

$$\|x^{k,1} - x^*\| \leq K_1\|x^{k,0} - x^*\| \tag{10}$$

当 $i = 1$ 时, 有

$$\nabla^2 f(x^{k,1}) - B^{k,1} = S(x^{k,1}) - \frac{\|R(x^{k,1})\|}{\|R(x^k)\|}S(x^k) =$$
$$S(x^{k,1}) - S(x^k) + \frac{\|R(x^k)\|S(x^k) - S(x^k)\|R(x^{k,1})\|}{\|R(x^k)\|}$$

利用假设 A2, 两边取范数得

$$\|\nabla^2 f(x^{k,1}) - B^{k,1}\| = \|S(x^{k,1}) - S(x^k)\| + \frac{\|S(x^k)\| \cdot |\|R(x^{k,1})\| - \|R(x^k)\||}{\|R(x^k)\|} \leq$$
$$(1 + M_5)L\|x^{k,1} - x^k\| \tag{11}$$

其中: M_5 为 $\frac{\|R(x^{k+1})\|}{\|R(x^k)\|}$ 的上界, 这里 $\|x^k - x^*\| < \delta_2^1$。由式(11)及文献[3]中的定理 5.4.1, 存在正数 M_6 使

$$\|x^{k,2} - x^*\| \leq M_6[\|x^{k,1} - x^*\|^2 + \|\nabla^2 f(x^{k,1}) - B^{k,1}\| \cdot \|x^{k,1} - x^*\|] \leq$$
$$M_6[\|x^{k,1} - x^*\|^2 + (1 + M_5)L\|x^{k,1} - x^k\|\|x^{k,1} - x^*\|] \tag{12}$$

取

$$\delta_2 = \min\left\{\delta_2^1, \frac{1}{2M_6(1 + M_5)L}\right\}$$

则当 $\|x^k - x^*\| < \delta_2$ 时, 有

$$\|x^{k,2} - x^*\| \leq M_6(1 + M_5)L\delta_2\|x^{k,1} - x^*\|$$

对 $i = 2, 3, \cdots, p-1$, 有类似结论, 不妨仍记为

$$\|x^{k,i+1} - x^*\| \leq M_6(1 + M_5)L\delta_2\|x^{k,1} - x^*\|$$

取

$$K = \max\{K_1, M_6(1 + M_5)L\delta_2\}$$

即得式(9)。

定理 3　假设 A1～A3 成立, 则存在正数 δ_3, 当 $\|x^0 - x^*\| < \delta_3$ 时, 算法中第 3 步产生的点 $\{x^k\}$ 收敛于 x^*, 且收敛速率至少为 $p+1$ 阶。

证明　当 $\|x^0 - x^*\| < \delta_2$ 时, 由定理 2 知

$$\|x^{k+1} - x^*\| \leq K^p\|x^k - x^*\| < K\|x^k - x^*\|$$

且牛顿步一定满足

$$\|x^{k,1} - x^*\| \leq K_1\|x^{k,0} - x^*\|$$

这里 $K_1 \leq 1$, 从而 $\{x^k\}$ 收敛于 x^*。

对于 $i = 1, 2, \cdots, p-1$, 类似于式(12), 于是有

$$\|x^{k,i+1}-x^*\|\le M_6[\|x^{k,i}-x^*\|^2+(1+M_5)L\|x^{k,i}-x^k\|\|x^{k,1}-x^*\|]\le$$
$$M_6[1+(1+M_5)L]\|x^{k,i}-x^*\|^2+M_6(1+M_5)L\|x^k-x^*\|\|x^{k,i}-x^*\|\le$$
$$M_7\|x^k-x^*\|\|x^{k,i}-x^*\| \tag{13}$$

这里$M_7=2M_6[1+(1+M_5)L]$。由式(13)得

$$\|x^{k+1}-x^*\|=\|x^{k,p}-x^*\|\le M_7(\|x^{k,p-1}-x^*\|\|x^k-x^*\|)\le$$
$$M_7^{p-1}\|x^k-x^*\|^{p-1}\|x^{k,1}-x^*\| \tag{14}$$

由于$x^{k,1}$是牛顿步所产生,所以存在$0<\delta_3\le\delta_2$和$M_8>0$,当$\|x^k-x^*\|<\delta_3$时,有

$$\|x^{k+1}-x^*\|<M_8\|x^k-x^*\|^2 \tag{15}$$

由式(14)和(15)得

$$\|x^{k+1}-x^*\|<M_7^{p-1}M_8\|x^k-x^*\|^{p+1}$$

所以$\{x^k\}$的收敛速率为$p+1$。

定理 4 当$R(x^*)=0$时,算法产生的点列$\{x^k\}$收敛于x^*,且收敛速率为2^p阶。

证明 收敛性由定理3得证,下面证明收敛速率。

首先证明k充分大时,存在正数M_9,使得

$$\|\nabla^2 f(x^*)-B^{k,i}\|\le M_9\|x^{k,i}-x^*\| \quad i=1,2,\cdots,p-1 \tag{16}$$

因为$R(x^*)=0$,从而有

$$\|\nabla^2 f(x^*)-B^{k,i}\|=\left\|J(x^*)^TJ(x^*)-J(x^{k,i})^TJ(x^{k,i})-\frac{\|R(x^{k,i})\|}{\|R(x^k)\|}S(x^k)\right\|\le$$
$$\|J(x^*)^TJ(x^*)-J(x^{k,i})^TJ(x^{k,i})\|+\|R(x^{k,i})\|\frac{\|S(x^k)\|}{\|R(x^k)\|} \tag{17}$$

和

$$\|R(x^{k,i})\|=\|R(x^{k,i})-R(x^*)\|\le L\|x^{k,i}-x^*\|$$

并且$\frac{\|S(x^k)\|}{\|R(x^k)\|}$有界,由式(17),存在$M_9>0$,使式(16)成立。

因为

$$\|x^{k,i+1}-x^*\|=\|x^{k,i}-(B^{k,i})^{-1}\nabla f(x^{k,i})-x^*\|\le$$
$$\|(B^{k,i})^{-1}\|\|B^{k,i}(x^{k,i}-x^*)-\nabla f(x^{k,i})\|\le$$
$$\|(B^{k,i})^{-1}\|[\|\nabla f(x^{k,i})-\nabla f(x^*)-\nabla^2 f(x^*)(x^{k,i}-x^*)\|+$$
$$\|(\nabla^2 f(x^{k,i})-B^{k,i})(x^{k,i}-x^*)\|]$$

注意到k充分大时,有

$$\|\nabla f(x^{k,i})-\nabla f(x^*)-\nabla^2 f(x^*)(x^{k,i}-x^*)\|=O(\|x^{k,i}-x^*\|^2)$$

及式(16),所以存在正数M_{10},使得当k充分大时,有

$$\|x^{k,i+1}-x^*\|\le M_{10}\|x^{k,i}-x^*\|^2 \quad i=1,2,\cdots,p-1 \tag{18}$$

又$x^{k,1}$是牛顿法产生,从而存在正数M_{11},使得当k充分大时,有

$$\|x^{k+1}-x^*\|\le M_{11}\|x^k-x^*\|^{2^p}$$

即$\{x^k\}$的收敛速率为2^p。

定理4表明对于零残差问题,新算法具有2^p收敛速率。把x^k看作由x^{k-1}进行p步迭代产生,从式(18)知,这实质上是一步二阶收敛,和牛顿法、高斯牛顿法的收敛速率相同。对于非零

残差问题, 由定理 3 知, 新算法超线性收敛, 具有 p 步 $p+1$ 阶收敛速率。

3　效率分析

评价一个算法的优劣, 可以看其效率的高低。一个算法的效率可用下式[3]表示

$$\eta = \frac{\ln\alpha}{W} \tag{19}$$

其中: η 表示算法的效率; α 为收敛速率; W 为从 x^k 到 x^{k+1} 的计算量, 它包括海赛矩阵 $\nabla^2 f(x)$ (或 B) 的计算量, 梯度 $\nabla f(x)$ 的计算量和求解线性方程组的计算量。

对于零残差问题, 新算法的收敛速率与牛顿法相同; 但由于平均每次迭代的计算量要比牛顿法少(新算法只需计算近似海赛矩阵), 所以由式(19)可知其效率要高。对非零残差问题新算法具有 p 步 $p+1$ 阶收敛速率, 其效率至少与牛顿法相同。新算法中参数 p 的选取将在以后进行研究。

参　考　文　献

1　Yabe H, Yamaki N. Convergence of a factorized Broden-like family for nonlinear least squarse problems. SIAM J Optimization, 1995, 5(4): 770～790

2　Dennis J E, Martinez H J, Tapia R A. Convergence theory for the structured BFGS secant method with an application to nonlinear least squares problems. J OptimiZation Theory Appl, 1998, 61(2): 161～178

3　李庆扬, 莫孜中, 祁力群. 非线性方程组的数值解决法. 北京: 科学出版社, 1992. 38～82