DOI:10.13203/j.whugis1991.02.011

第16卷 第2期 武汉测绘科技大学学报 Vol.16 No.2
1991年 6月 Journal of Wuhan Technical University of Surveying and Mapping Jun. 1991

非线性最小二乘平差迭代解法的收敛性

白亿同

摘要

本文讨论了非线性最小二乘平差中高斯—牛顿法和阻尼最小二乘法迭代程序的收敛性问题，找出了影响收敛的因素，给出了估算迭代程序收敛性的方法。

【关键词】 迭代解法；收敛性；范数

1 引 言

本文从数学角度讨论高斯—牛顿法和阻尼最小二乘法的收敛性。设非线性最小二乘间接平差的误差方程为：

$$V=f(X)-L \tag{1}$$

式中 $\underset{m\times 1}{L}$ 和 $\underset{m\times 1}{V}$ 表示观测值和相应的改正数向量；$\underset{1\times n}{X}$ 表示未知参数向量；f是将R^n中的区域D变到R^m中的算子，$m>n$。并假定权阵 $\underset{m\times m}{P}$ 为正定对称方阵。求方程(1)的最小二乘解的高斯—牛顿迭代程序为：

$$X_{k+1}=X_k-〔Df(X_k)^TPDf(X_k)〕^{-1}Df(X_k)^TP〔f(X_k)-L〕 \tag{2}$$

$k=0, 1, 2, \cdots$。X_0为初始近似值（简称初值）。这里$Df(X)$表示$f(X)$的导数，它实际上是雅可比（Jacobi）矩阵。

阻尼最小二乘法的迭代程序为：

$$X_{k+1}=X_k-〔\mu_kI+Df(X_k)^TPDf(X_k)〕^{-1}Df(X_k)^TP〔f(X_K)-L〕 \tag{3}$$

其中 I 为单位阵，$\mu_k\geqslant 0$。

2 迭代程序的收敛性定理

由文献〔3〕知方程(1)的最小二乘解必定满足方程

$$Df(X)^TP〔f(X)-L〕=0 \tag{4}$$

收稿日期:1990—03—07

我们把方程（4）的解称为方程（1）最小二乘解的驻点，当$\| V \|_2^2 = V^T P V$为凸泛函数时，驻点X^*就是（1）的最小二乘解（即平差值$\hat{X}$）。

设X^*为方程(1)最小二乘解的驻点，假定算子f定义在$D \subset R^n$上，且在闭球Ω（$\| X - X^* \| \leqslant r$）$\subset D$上二次可微，又$\operatorname{rank} Df(X^*) = n$，并记$A(X) = Df(X)^T P Df(X)$，我们有下面的收敛性定理。

定理1 （高斯—牛顿法收敛性定理）：若当$X \in \Omega$时$\| D^2 f(X) \| \leqslant K$（$K$为常数）且有

$$K \| P \| \ \| A(X^*)^{-1} \| \ \| f(X^*) - L \| < 1 \tag{5}$$

则由（2）式给出的序列$\{X_k\}$局部收敛于X^*。

证：由于$\operatorname{rank} Df(X^*) = n$，$A(X^*)$正定，因$f(X)$二次可微，存在$X^*$的闭球$\Omega_0 \subset \Omega$，当$X \in \Omega_0$时有$A(X)$正定，因而$A(X)^{-1}$存在，故迭代程序（2）在$\Omega_0$内是有意义的。

因为$Df(X)$在Ω_0上连续，故存在常数$M > 0$，当$X \in \Omega_0$时，有

$$\| Df(X) \| \leqslant M$$

由条件(5)可知存在$\delta > 0$，使得闭球Ω_1($\| X - X^* \| \leqslant \delta$)$\subset \Omega_0$，且有

$$\delta < \left[\left(\| P \| \ \| A(X^*)^{-1} \| \right)^{-1} - K \| f(X^*) - L \| \right] (2MK)^{-1} \tag{6}$$

注意到X^*为方程(4)的解，则当$X_k \in \Omega_1$时，有

$$\begin{aligned} X_{k+1} - X^* &= X_k - A(X_k)^{-1} Df(X_k)^T P [f(X_k) - L] - X^* \\ &= A(X_k)^{-1} [Df(X_k)^T P (Df(X_k)(X_k - X^*) - (f(X_k) - f(X^*))) \\ &\quad - (Df(X_k)^T - Df(X^*)^T) P (f(X^*) - L)] \end{aligned}$$

利用中值定理：

$$\begin{aligned} X_{k+1} - X^* &= A(X_k)^{-1} [Df(X_k)^T P \int_0^1 (Df(X_k) - Df(X^* + t(X_k - X^*))) \\ &\quad (X_k - X^*) dt - \left(\int_0^1 D^2 f(X^* + s(X_k - X^*)) (X_k - X^*) ds \right)^T P (f(X^*) - L)] \\ &= A(X_k)^{-1} [Df(X_k)^T P \int_0^1 \int_0^1 D^2 f(X^* + (t + \theta - t\theta)(X_k - X^*))(1 - t)(X_k - X^*)^2 d\theta dt \\ &\quad - \left(\int_0^1 D^2 f(X_k + s(X^* - X_k)) (X_k - X^*) ds \right)^T P (f(X^*) - L)] \\ &\qquad (0 < t, \ s, \ \theta < 1) \end{aligned}$$

于是得：

$$\| X_{k+1} - X^* \| \leqslant K \| P \| \ \| A(X_k)^{-1} \| (M\delta + \| f(X^*) - L \|) \| X_k - X^* \| \tag{7}$$

令

$$q = \sup [K \| P \| \ \| A(X_k)^{-1} \| (M\delta + \| f(X^*) - L \|)]$$

再利用（5）和（6）式即得$0 < q < 1$。由（7）式得：

$$\| X_{k+1} - X^* \| \leqslant q \| X_k - X^* \| < gr < r \tag{8}$$

这表明当初值$X_0 \in \Omega_1$时，则对一切k有$X_k \in \Omega_1$，即迭代程序（2）在Ω_1内是可实现的。同时由（8）得

$$\| X_k - X^* \| \leqslant q^k \| X_0 - X^* \| \tag{9}$$

因此

$$\lim_{k\to\infty} X_k = X^*$$

这就证明了序列$\{X_k\}$收敛于X^*，由（9）式还可以看出迭代程序（2）至少具有线性收敛速度。

若取l_2范数，则$\|A(X^*)^{-1}\|_2=\lambda^{-1}$（$\lambda$为$A(X^*)$的最小特征根），于是得到如下推论。

推论：若当$X\in\Omega$时$\|D^2f(X)\|_2<K$，且有$\lambda>K\|P\|_2\|f(X^*)-L\|_2$，则由（2）式给出的序列$\{X_k\}$局部收敛于$X^*$。

定理 2 （阻尼最小二乘法收敛性定理）：若当$X\in\Omega$时$\|D^2f(X)\|_p\leqslant K$，且有

$$K\|P\|_p\|A(X^*)^{-1}\|_p\|f(X^*)-L\|_p<1 \tag{10}$$

$\{\mu_k\}$为任意非负有界实数序列，则由（3）式给出的序列$\{X_k\}$局部收敛于X^*。其中$p=1$，2或∞。

（5）、（10）中的K表示$\|D^2f(X)\|$在Ω中的上确界，在第4节中给出了估算方法。

3 对收敛定理的分析

定理1中的条件（5）和定理2中的条件（10）是收敛定理中的关键性条件，因X^*事先并不知道，若X_0与X^*的距离不太远时，可用X_0近似代替X^*来进行估算，我们称这两个定理为局部性收敛定理。从条件（5）（或（10））可以得出如下结论。

1．迭代程序的收敛条件（5）中含有$\|D^2f(X)\|$、$\|A(X^*)^{-1}\|$、$\|f(X^*)-L\|$三个因子。其中$\|f(X^*)-L\|$表示观测值的精度（当X^*换作X_0时，它也表示初值的精度）。$\|A(X^*)^{-1}\|$与一阶导数$Df(X^*)$有关，它表示法方程系数矩阵$A(X^*)$对角占优的程度。$\|D^2f(X)\|$与二阶导数有关，按照文献〔3〕的观点如果把$Y=f(X)$看作一个流形（实为m维空间中n维超曲面），那么$\|D^2f(X)\|$表示超曲面在X^*附近弯曲的程度，对于它的影响我们也必须认真考虑。条件（5）要求这三项都比较小，但它们之间又有一定的制约关系，如果其中某两项较小，另外一项在遵从条件（5）的前提下可适当大些。例如，当平差问题的函数模型是非线性形式时，须把它化为线性形式，如果略去的二次项系数都较小且法方程系数矩阵中对角线远大于其它元素时，初值X_0的近似程度就可适当放宽些。

2．若初值X_0与X^*相差较大时，可能有$X_0\overline{\in}\Omega_0$，$\mathrm{rank}f(X_0)<n$的情形产生，这时法方程系数矩阵$A(X_0)$秩亏，即$\det A(X_0)=0$，使高斯—牛顿法失效，还有可能出现发散的情况，因而不能采用高斯—牛顿法，而应考虑使用阻尼最小二乘法。文献〔1〕中对此有详细论述。

4 $\|D^2f(X)\|$的估算方法

$\|A(X_0)^{-1}\|$，$\|f(X^0)-L\|$都比较容易求解，在这里我们取$X=X_0$来近似计算$\|D^2f(X)\|$，其中$D^2(X)$是二阶导数，它是一个双线性算子，即有三个指标的矩阵（立体阵）

$$\left(\frac{\partial^2 f_h(x)}{\partial x_i \partial x_j}\right) \qquad h=1,\ 2\cdots m,\ i,\ j=1,\ 2\cdots n$$

取l_2范数时，有估计式

$$\| D^2 f(X) \|_2 \leqslant \left[\sum_{h=1}^{m} \Lambda_h \right]^{1/2}$$

其中Λ_h表示矩阵$A_h A_h^T$的最大特征根，而

$$A_h = \begin{pmatrix} \dfrac{\partial^2 f_h}{\partial x_1 \partial x_1} & \dfrac{\partial^2 f_h}{\partial x_1 \partial x_2} & \cdots\cdots & \dfrac{\partial^2 f_h}{\partial x_1 \partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial^2 f_h}{\partial x_n \partial x_1} & \dfrac{\partial^2 f_h}{\partial x_n \partial x_2} & \cdots\cdots & \dfrac{\partial^2 f_h}{\partial x_n \partial x_n} \end{pmatrix} \qquad h = 1,\ 2,\ \cdots,\ m$$

如果取l_∞的范数时，则有估计式

$$\| D^2 f(X) \|_\infty \leqslant \max \sum_{i=1}^{m} \sum_{j=1}^{m} \left| \frac{\partial^2 f_h}{\partial x_i \partial x_j} \right|$$

5 结束语

对收敛定理的分析，使我们了解了影响最小二乘迭代程序收敛性的因素，为测绘工作布网和成果计算提供了参考依据。根据已有观测成果利用这两个定理也能对迭代程序的收敛性及收敛速度进行估算。但这两个定理是局部性收敛定理，当X_0和$\hat{X}$相差较大时定理的估计就不一定可靠，因而它们的应用具有一定的局限性。从X_0出发进行估算的收敛性定理会更加切合实际。

参 考 文 献

〔1〕 刘大杰、黄加纳．非线性最小二乘平差的迭代解法．武测科技，1987(4)．

〔2〕 王德人．非线性方程组解法与最优化方法．高等教育出版社，1979．

〔3〕 白亿同．新投影定理及其在平差中的应用．武汉测绘科技大学学报，1986(1)．

〔4〕 Д. В. Кант`р◠вич. Г. Ц. Акил◠ь. 泛函分析(下册)。高等教育出版社，1982．

Convergence Of Iterative Solution For Nonlinear Least Squares Adjustment

Bai Yitong

Abstract

This paper discusses the problem on the convergence of iterative solutions for non-linear adjustment model by Gauss-Newton methods and damped least sguare ones, finds out the factors affecting the convergence and gives the methods for estimating the convergence of iterative procedures.

[Key words] iterative solution, convergence; norm