文章编号: 1000-050X(2001)02-0118-04　　文献标识码: A

用差商代替导数的非线性最小二乘估计

李桂苓[1]　万剑华[1]　陶华学[2]

(1　石油大学(华东)测绘工程系, 山东东营, 257061)
(2　山东科技大学地球信息科学与工程学院, 山东泰安, 271019)

摘　要: 针对不同类型观测值的非线性最小二乘平差, 介绍一种不使用导数的解析方法。在这种解算中, 由于只使用函数值, 避免了二阶和二阶偏导数的计算, 使原本复杂的计算得以简化。实例验证了本方法的有效性和可靠性。

关键词: 差商代替导数; 非线性; 最小二乘估计
中图法分类号: P207

测绘科技领域对非线性观测值函数的处理, 还大都是通过在观测值的近似值处按泰勒级数展开, 取至一次项化为线性函数作经典的线性平差, 这样处理是建立在待求量可取得很好近似的基础上的。很显然, 在测绘高新技术飞速发展的今天, 对观测成果的质量要求、观测数据的处理及精度评定的要求愈来愈高, 这样处理所得到的平差结果已不能满足科技发展的需要。因此, 非线性最小二乘平差的研究已越来越受到人们的重视。近年来, 国际大地测量协会(IAG)已将非线性模型的处理列为重点研究的课题。另外, 现代测绘仪器的不断涌现、测绘手段的不断提高和完善, 势必使测量数据的类型由单一的同一类型观测值进入到含不同类型观测值的平差模型。而非线性平差又势必涉及到二次以上的高阶导数, 给数据的处理带来很大的不便。为此, 本文对不同类型观测值的非线性最小二乘问题采用不使用导数的处理方法, 并且用算例证明了其可行性。

1　不同类型观测数据的非线性最小二乘估计解算方法及步骤

1.1　解算方法

设某控制网中有两类不同观测值 $\underset{m\times 1}{\boldsymbol{L}_1}$、$\underset{n\times 1}{\boldsymbol{L}_2}$, 且互不相关, 相应的观测值权阵为 $\underset{m\times m}{\boldsymbol{P}_1}$、$\underset{n\times n}{\boldsymbol{P}_2}$, 选择网中各点坐标为未知参数 $\underset{t\times 1}{\boldsymbol{x}}=(x_1, x_2, \cdots, x_t)^{\mathrm{T}}$。该问题可表示为:

$$\begin{aligned}\boldsymbol{V}_1(\boldsymbol{x})&=\boldsymbol{f}_1(\boldsymbol{x})-\boldsymbol{L}_1, \boldsymbol{P}_1\\ \boldsymbol{V}_2(\boldsymbol{x})&=\boldsymbol{f}_2(\boldsymbol{x})-\boldsymbol{L}_2, \boldsymbol{P}_2\end{aligned}\tag{1}$$

式中,

$$\boldsymbol{V}_1(\boldsymbol{x})=\begin{pmatrix}V_{11}(\boldsymbol{x})\\V_{12}(\boldsymbol{x})\\\cdots\\V_{1m}(\boldsymbol{x})\end{pmatrix}, \boldsymbol{V}_2(\boldsymbol{x})=\begin{pmatrix}V_{21}(\boldsymbol{x})\\V_{22}(\boldsymbol{x})\\\cdots\\V_{2n}(\boldsymbol{x})\end{pmatrix}$$

$$\boldsymbol{f}_1(\boldsymbol{x})=\begin{pmatrix}f_{11}(\boldsymbol{x})\\f_{12}(\boldsymbol{x})\\\cdots\\f_{1m}(\boldsymbol{x})\end{pmatrix}, \boldsymbol{f}_2(\boldsymbol{x})=\begin{pmatrix}f_{21}(\boldsymbol{x})\\f_{22}(\boldsymbol{x})\\\cdots\\f_{2n}(\boldsymbol{x})\end{pmatrix}$$

$$\boldsymbol{L}_1=\begin{pmatrix}L_{11}\\L_{12}\\\cdots\\L_{1m}\end{pmatrix}, \boldsymbol{L}_2=\begin{pmatrix}L_{21}\\L_{22}\\\cdots\\L_{2n}\end{pmatrix}$$

根据非线性最小二乘理论, 式(1)可以归结为如下的数学问题:

$$\begin{aligned}\min \boldsymbol{F}(\boldsymbol{x})&=\boldsymbol{V}_1^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{P}_1\boldsymbol{V}_1(\boldsymbol{x})+\\ &\boldsymbol{V}_2^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{P}_2\boldsymbol{V}_2(\boldsymbol{x})=\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{P}\boldsymbol{V}(\boldsymbol{x})\end{aligned}\tag{2}$$

式中, $\boldsymbol{V}(\boldsymbol{x})=[V_{11}(\boldsymbol{x}), V_{12}(\boldsymbol{x}), \cdots, V_{1m}(\boldsymbol{x}), V_{21}(\boldsymbol{x}), V_{22}(\boldsymbol{x}), \cdots, V_{2n}(\boldsymbol{x})]^{\mathrm{T}}$

$$\boldsymbol{P}=\begin{bmatrix}\boldsymbol{P}_1 & \\ & \boldsymbol{P}_2\end{bmatrix}$$

对于上述不同类型测量数据的非线性最小二乘问题的求解, 目前虽有一些现成的方法可用, 但

收稿日期: 2000-10-24。
项目来源: 国家自然科学基金资助项目(49874001)。

在解算过程中却需要用到目标函数的一阶甚至二阶导数。若目标函数的一阶或二阶导数并不存在, 此时不能采用带导数的解算方法。

根据极小值的必要条件, 式(2)必须满足:

$$\nabla \boldsymbol{F}(\boldsymbol{x}) = 2\boldsymbol{J}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V}(\boldsymbol{x}) = 0 \tag{3}$$

式中,

$$\boldsymbol{J}^{\mathrm{T}} = \begin{bmatrix} \frac{\partial V_{11}(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial V_{1m}(\boldsymbol{x})}{\partial x_1} & \frac{\partial V_{21}(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial V_{2n}(\boldsymbol{x})}{\partial x_1} \\ \cdots & & \cdots & \cdots & & \cdots \\ \frac{\partial V_{11}(\boldsymbol{x})}{\partial x_t} & \cdots & \frac{\partial V_{1m}(\boldsymbol{x})}{\partial x_t} & \frac{\partial V_{21}(\boldsymbol{x})}{\partial x_t} & \cdots & \frac{\partial V_{2n}(\boldsymbol{x})}{\partial x_t} \end{bmatrix}$$

为了采用不带导数的解算过程, 首先, 在初选点 $\boldsymbol{x}^{(0)}$附近按泰勒级数展开, 截去二阶以上的高阶小量, 向量函数值 $\boldsymbol{V}(\boldsymbol{x})$可近似表达为:

$$V_{1i}(x_1, x_2, \cdots, x_t) \approx V_{1i}(x_1^{(0)}, x_2^{(0)}, \cdots, x_t^{(0)}) + \frac{\partial V_{1i}(x_1, x_2, \cdots, x_t)}{\partial x_1}(x_1 - x_1^{(0)}) + \frac{\partial V_{1i}(x_1, x_2, \cdots, x_t)}{\partial x_2}(x_2 - x_2^{(0)}) + \cdots + \frac{\partial V_{1i}(x_1, x_2, \cdots, x_t)}{\partial x_t}(x_t - x_t^{(0)}) = \frac{\partial V_{1i}(x_1, x_2, \cdots, x_t)}{\partial x_1}x_1 + \frac{\partial V_{1i}(x_1, x_2, \cdots, x_t)}{\partial x_2}。x_2 + \cdots + \frac{\partial V_{1i}(x_1, x_2, \cdots, x_t)}{\partial x_t}x_t + V_{1i}(x_1^{(0)}, x_2^{(0)}, \cdots, x_t^{(0)}) - \frac{\partial V_{1i}(x_1, x_2, \cdots, x_t)}{\partial x_1}x_1^{(0)} - \frac{\partial V_{1i}(x_1, x_2, \cdots, x_t)}{\partial x_2}x_2^{(0)} - \cdots - \frac{\partial V_{1i}(x_1, x_2, \cdots, x_t)}{\partial x_t}x_t^{(0)} = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{it}x_t + b_i (i=1, 2, \cdots, m) \tag{4}$$

同理有:

$$V_{2i}(x_1, x_2, \cdots, x_t) \approx d_{i1}x_1 + d_{i2}x_2 + \cdots + d_{it}x_t + c_i (i=1, 2, \cdots, n) \tag{5}$$

式(4)、(5)可表示为:

$$\boldsymbol{l}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \tag{6}$$

式中,

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1t} \\ \cdots & & \cdots \\ a_{m1} & \cdots & a_{mt} \\ d_{11} & \cdots & d_{1t} \\ \cdots & & \cdots \\ d_{n1} & \cdots & d_{nt} \end{pmatrix}$$

$$\boldsymbol{b} = (b_1, b_2, \cdots, b_m, c_1, c_2, \cdots, c_n)^{\mathrm{T}}$$

由式(6)可得: $\nabla \boldsymbol{l}(\boldsymbol{x}) = \boldsymbol{A}$　(7)

在 $\boldsymbol{x}^{(k)}$附近用 $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{l}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{P}\boldsymbol{l}(\boldsymbol{x}) = 0$ 代替 $\boldsymbol{F}(\boldsymbol{x})$, 则有:

$$\nabla \boldsymbol{F}(\boldsymbol{x}) = 2\nabla \boldsymbol{l}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{P}\boldsymbol{l}(\boldsymbol{x}) = 0 \tag{8}$$

将式(6)、(7)代入式(8), 可得:

$$\nabla \boldsymbol{F}(\boldsymbol{x}) = 2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) = 0 \tag{9}$$

由式(4)、(5)和式(6)可知, 线性函数 $\boldsymbol{l}(\boldsymbol{x})$是非线性函数 $\boldsymbol{V}(\boldsymbol{x})$的近似, 此处取

$$\boldsymbol{b}^{(k)} = \boldsymbol{V}(\boldsymbol{x}^{(k)}) - \boldsymbol{A}_k\boldsymbol{x}^{(k)} \tag{10}$$

若令式(9)中 $\boldsymbol{x} = \boldsymbol{x}^{(k+1)}$, 而其他量则用第 k 次结果代入, 利用式(9)和式(10), 于是有:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - (\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{P}\boldsymbol{A}_k)^{-1}\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V}(\boldsymbol{x}^{(k)}) \tag{11}$$

式中, $\boldsymbol{A}_k$ 满足下列关系式:

$$\Delta \boldsymbol{V}_k = \boldsymbol{A}_k \Delta \boldsymbol{x}_k \tag{12}$$

按照传统的做法, $\boldsymbol{A}_k$ 通过求一阶偏导数得到。这里避开此问题, 而通过构造 $\Delta \boldsymbol{V}_k$、$\Delta \boldsymbol{x}_k$, 从而获得系数矩阵 $\boldsymbol{A}_k$。下面讨论如何求 $\boldsymbol{A}_k$。

式(12)中 $\boldsymbol{A}_k$ 是$(m+n)\times t$ 阶矩阵, 如果构造出的$\Delta \boldsymbol{x}_k$ 能很容易求逆, 则将会使问题变得非常简单, 这样 $\Delta \boldsymbol{x}_k$ 应为 $t\times t$ 阶矩阵。假设已经过 k 步迭代, 如果在第 k 步迭代点周围选择 $2t$ 个辅助点, 其中前 t 个点取为 $\boldsymbol{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, \cdots, x_t^{(k)})$, 后 t 个点取为最优点的当前近似 $\boldsymbol{x}^* = (x_1^*, x_2^*, \cdots, x_t^*)$, 其中的下脚标代表点号而非未知参数向量的各分量, 两两相减于是得:

$$\underset{t\times t}{\Delta \boldsymbol{x}_k} = (x_1^* - x_1^{(k)}, x_2^* - x_2^{(k)}, \cdots, x_t^* - x_t^{(k)})$$

相应的 $\Delta \boldsymbol{V}_k$ 为$(m+n)\times t$ 阶矩阵:

$$\underset{(m+n)\times t}{\Delta \boldsymbol{V}_k} = (V(x_1^*) - V(x_1^{(k)}), V(x_2^*) - V(x_2^{(k)}), \cdots, V(x_t^*) - V(x_t^{(k)}))$$

对于 $\Delta \boldsymbol{x}_k$ 取为:

$$x_j^{(k)} = \boldsymbol{x}^{(k)}, j = 1, 2, \cdots, t$$

$$x_j^* = \boldsymbol{x}^{(k)} + (x_j^{(k-1)} - x_j^{(k)})\boldsymbol{e}_j$$

式中, $\boldsymbol{e}_j = (0\ 0\ \cdots\ \underset{j}{1}\ 0\ \cdots 0)^{\mathrm{T}}$。即前 t 个点重合于 $\boldsymbol{x}^{(k)}$上, 而 $\boldsymbol{x}^*$的第 j 个点 $x_j^*(j=1, 2, \cdots, s)$中的第 j 个分量则是在 $x_j^{(k)}$ 的相应分量中加上 $x_j^{(k-1)} - x_j^{(k)}$, 其余各分量与 $x_j^{(k)}$ 中相同。

令 $h_j^{(k)} = (x_j^{(k-1)} - x_j^{(k)})$, 则 $\Delta \boldsymbol{x}_k$ 为对角矩阵:

$$\Delta \boldsymbol{x}_k = \begin{bmatrix} h_1^{(k)} & & & \\ & h_2^{(k)} & & \\ & & \cdots & \\ & & & h_t^{(k)} \end{bmatrix} \tag{13}$$

对于矩阵 $\Delta \boldsymbol{V}_k$, 则有下列形式:

$$\Delta \boldsymbol{V}_k = [V(\boldsymbol{x}^{(k)} + h_1^{(k)}e_1) - V(\boldsymbol{x}^{(k)}), V(\boldsymbol{x}^{(k)} + h_2^{(k)}e_2) - V(\boldsymbol{x}^{(k)}), \cdots, V(\boldsymbol{x}^{(k)} + h_t^{(k)}e_t) - V(\boldsymbol{x}^{(k)})] = \Delta \boldsymbol{V}(\boldsymbol{x}^{(k)}\boldsymbol{h}^{(k)}) \tag{14}$$

将式(13)、(14)代入式(12)，于是有：

$$\boldsymbol{A}_k=\Big[\frac{1}{h_1^{(k)}}(V(\boldsymbol{x}^{(k)}+h_1^{(k)}e_1)-V(\boldsymbol{x}^{(k)})),\ \frac{1}{h_2^{(k)}}\Big(V(\boldsymbol{x}^{(k)}+h_2^{(k)}e_2)-V(\boldsymbol{x}^{(k)})\Big),\ \cdots,\ \frac{1}{h_t^{(k)}}(V(\boldsymbol{x}^{(k)}+h_t^{(k)}e_t)-V(\boldsymbol{x}^{(k)}))\Big] \tag{15}$$

对照式(7)可知，这里 $\boldsymbol{A}_k$ 为 $\boldsymbol{V}(\boldsymbol{x})$的差商所构成的矩阵，用以代替式(7)中函数 $\boldsymbol{V}(\boldsymbol{x})$的一阶偏导数阵。

若取 $h_1^{(k)}=h_2^{(k)}=\cdots=h_t^{(k)}=h^{(k)}$，则有：

$$\boldsymbol{A}_k=\frac{1}{\boldsymbol{h}^{(k)}}\big[(V(\boldsymbol{x}^{(k)}+h_1^{(k)}e_1)-V(\boldsymbol{x}^{(k)})),\ (V(\boldsymbol{x}^{(k)}+h_2^{(k)}e_2)-V(\boldsymbol{x}^{(k)})),\ \cdots,\ (V(\boldsymbol{x}^{(k)}+h_t^{(k)}e_t)-V(\boldsymbol{x}^{(k)}))\big] \tag{16}$$

最后得如下的迭代格式：

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}-\boldsymbol{h}^{(k)}[[\Delta\boldsymbol{V}(\boldsymbol{x}^{(k)}\boldsymbol{h}^{(k)})]^{\mathrm T}\circ\boldsymbol{P}\Delta\boldsymbol{V}(\boldsymbol{x}^{(k)}\boldsymbol{h}^{(k)})]^{-1}\Delta\boldsymbol{V}(\boldsymbol{x}^{(k)}\boldsymbol{h}^{(k)})^{\mathrm T}\boldsymbol{P}\boldsymbol{V}(\boldsymbol{x}^{(k)}) \tag{17}$$

1.2　解算步骤

1)取初始近似值 $\boldsymbol{x}^{(0)}$，允许控制误差 ε(可取 0.003 或 0.005)，松弛值 β∈(0, 1)(可取 0.3、0.4 等)，并置 $k=1$；

2)由式(1)计算函数值得：

$$\boldsymbol{V}(\boldsymbol{x}^{(k)})=[V_{11}(\boldsymbol{x}^{(k)}),\ V_{12}(\boldsymbol{x}^{(k)}),\ \cdots,\ V_{1m}(\boldsymbol{x}^{(k)}),\ V_{21}(\boldsymbol{x}^{(k)}),\ V_{22}(\boldsymbol{x}^{(k)}),\ \cdots,\ V_{2n}(\boldsymbol{x}^{(k)})]^{\mathrm T}$$

3)计算 $\boldsymbol{h}^{(k)}=\beta\ \|\boldsymbol{V}(\boldsymbol{x}^{(k)})\|$；

4)由式(14)计算 $\Delta\boldsymbol{V}(\boldsymbol{x}^{(k)}\boldsymbol{h}^{(k)})$矩阵各元素值；

5)计算 $\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}-\boldsymbol{h}^{(k)}[[\Delta\boldsymbol{V}(\boldsymbol{x}^{(k)}\circ\boldsymbol{h}^{(k)})]^{\mathrm T}\boldsymbol{P}\Delta\boldsymbol{V}(\boldsymbol{x}^{(k)}\boldsymbol{h}^{(k)})]^{-1}\Delta\boldsymbol{V}(\boldsymbol{x}^{(k)}\boldsymbol{h}^{(k)})^{\mathrm T}\boldsymbol{P}\boldsymbol{V}\circ(\boldsymbol{x}^{(k)})$；

6)判断

$$\Big\|\frac{1}{\boldsymbol{h}^{(k)}}\Delta\boldsymbol{V}(\boldsymbol{x}^{(k)}\boldsymbol{h}^{(k)})^{\mathrm T}\boldsymbol{P}\boldsymbol{V}(\boldsymbol{x}^{(k)})\Big\|\leqslant\varepsilon$$

如上式满足，结束并输出 $\boldsymbol{x}^*=\boldsymbol{x}^{(k)}$ 或 $\boldsymbol{x}^{(k+1)}$；如不满足，置 $k=k+1$，重复第 2)～6)步，直到满足为止。

2　算例及分析

如图1所示，A、B、C是已知点，P_1、P_2为待定点，网中观测了 12 个角度和 6 条边的边长。已知测角中误差为±1.5″，测边中误差为±2.0cm。求 P_1、P_2 点的坐标平差值。

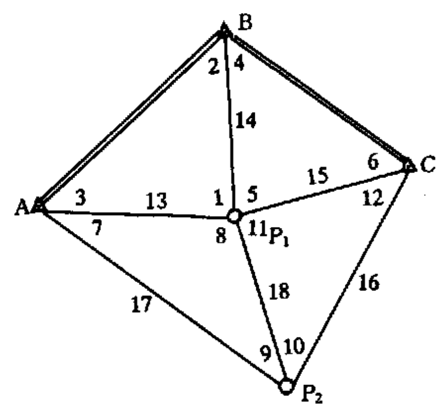

图 1　边角网

Fig. 1　Triangulateration Network

设 $t=4$，$m=12$，$n=6$，选择松弛量 β=0.3，控制误差 ε=0.003，经过两次迭代得到平差结果见表 1。为了比较两种平差结果的好坏，在此算出了各自的 $\boldsymbol{V}^{\mathrm T}\boldsymbol{PV}$ 值(计算时，角度改正数取“″”，边长改正数取“cm”)。

表 1　计算结果比较

Tab. 1　Comparation of Adjusted Results

项　目	P_1		P_2		$\boldsymbol{V}^{\mathrm T}\boldsymbol{PV}$
	x/m	y/m	x/m	y/m	
本文方法	5 656.846 8	2 475.520 8	663.801 3	2 944.011 0	272.975 2
经典方法	5 656.931	2 475.561	663.868	2 944.094	459.179 4

由上述平差结果可以看出，本文方法平差结果优于经典平差结果，而且工作量也得到了大幅度的减少，另外在收敛性方面只进行了两步迭代就达到了精度要求。由此可以看出，本文不使用导数的平差方法不但简化了计算，而且也保证了精度，是一种解非线性最小二乘问题的良好途径。

参　考　文　献

1　陶华学. 不同类型观测数据的最小二乘估计. 山东矿业学院学报，1996(4)

2　席少霖，赵凤治. 最优化计算方法. 上海：上海科学技术出版社，1983

作者简介: 李桂苓, 硕士, 讲师。主要研究方向为非线性测量数据处理。代表成果: 不同类型观测值非线性平差模型的降维解算; 用气温梯度进行折光改正的三角高程测量及应用; 多类观测值非线性平差参数的方差-协方差传播。
E-mail: kyk@hdpu. edu. cn

Nonlinear Least Squares Adjustment of Non-derivative

LI Guiling[1]　*WAN Jianhua*[1]　*TAO Huaxue*[2]

(1　University of Petroleum, Department of Surveying Engineering, Dongying, Shandong, China, 257061)

(2　School of Geoinformation Science and Technology, Shandong University of Science and Technology, Tai' an, Shandong, China, 271019)

Abstract: In classical adjustment, most of observation values are still expanded to carry out linear adjustment at their approximations in terms of Taylor' s series of which the first order term as a linear function is taken out for the treatment of nonlinear functions of observation values existing broadly in the field of surveying and mapping science and technology. This treatment is based on undetermined quantities which are very close to their true values. In effect, a number of nonlinear models can not be tackled as usual. It is more and more highly requiring for the quality of observation accomplishments, treatment of observation datum and accuracy appraise, especially in the rocket-sky development of today' s high technology. Apart from this, traditional and single geodetic surveying or non-geodetic surveying datum has been expanded into the combination of observation quantities on geodetic surveying or lots of types, different accuracy observation quantities on non-geodetic surveying with the unending advent of modern observation apparatus, the continuous enhancement and improvement of surveying means.

This paper gives a computational method of nonlinear least squares adjustment without derivative for different types of observation value, in which a function value instead of its derivative is only used. As far as we know, finding the partial value of a function is a troublesome task, especially when the components $f_i(x)(i=1, 2, \cdots, m+n)$ of a nonlinear function $f(x)$ are very complicated. The essence of the method is to simplify its computations for a Jacobi' s matrix, which determines the coefficient matrix with relation to a linear equation set by utilizing algebraic interpolation method with function values. In a geometrical sense, what is different from that in Guass-Newton' s method is that a series of super-tangent planes can be converted into super-cut ones or that it can be seen as a discrete deformation of Guass-Newton' s method. Especially while every difference step size is taken as a constant, the calculation is even more convenient. Examples show that the method is an effective algorithm and is of super-linear convergence velocity.

Key words: non-derivative; nonlinear; least squares adjustment

About the author: LI Guiling, lecturer, master. Her main research is data processing. Her major achievements are a reduced dimension method of nonlinear adjustment for different types of observation value, trigonometric leveling with refraction amendment based on temperature gradient and its application, variance-covariance propagation of nonlinear least squares adjusted parameters for different types of observation value.
E-mail: kyk@hdpu. edu. cn