第25卷第3期　　大地测量与地球动力学　　Vol. 25 , No. 3
2005年8月　　JOURNAL OF GEODESY AND GEODYNAMICS　　Aug. ,2005

文章编号:1671-5942(2005)03-0029-05

粗差检定的两种途径*

刘根友　郝晓光　柳林涛

(中国科学院测量与地球物理研究所,武汉　430077)

摘　要　给出了解决粗差问题的两种计算方案: ①将粗差作为待估参数,采用拟稳平差思路解秩亏问题,直接获得粗差; ②选取部分观测值作为准观测值,采用部分最小二乘法获得待估参数,将非准观测值的残差作为粗差。结果表明,两种方案与“拟准检定法”具有相同的效果,具有粗差的观测值在平差时不起作用。

关键词　平差因子　粗差　准观测值　部分最小二乘法　拟准检定法

中图分类号:P207　　　**文献标识码**:A

TWO APPROACHES TO GROSS ERROR DETECTION

Liu Genyou , Hao Xiaoguang and Liu Lintao

(Institute of Geodesy and Geophysics , CAS , Wuhan　430077)

Abstract　Two approaches to gross error detection are proposed in this paper. The first approach is that the gross errors (G) are treated as parameters to be estimated , and the idea of “Quasi-Stable adjustment” ($G_r^T G_r = \min$) for solving rank deficient problem is adopted. The second approach is that part observations are selected as accurate observations , the principle of partial least-square estimation ($V_r^T V_r = \min$) is used , and than the residuals of the other observations are treated as gross errors. The results by these two approaches are the same as by QUAD (Quasi Accurate Detection) , and the observations with gross error no longer affects the adjustment.

Key words: adjustment factor , gross error , quasi-accurate observation , part least-square estimation , QUAD (Quasi Accurate Detection)

1　引言

粗差是指观测值中离群较大的误差(一般被定义为大于观测中误差的3倍),它不同于偶然误差,一般只是少数,在进行参数估计时,应首先将粗差剔除[1]。为了避免粗差的发生,实际观测中都制定了相应的测量规范。如角度测量的盘左盘右读数校验,三角网的闭合差检查,水准测量的双面读数校核及环闭合差检查等。在一些自动化程度较高的测量过程中,粗差很难及时发现,因此粗差诊断仍然是一项非常重要的工作。欧吉坤于1999年提出了以真误差为研究对象的粗差拟准检定法[1],得到了广泛的应用[2~6]。但是拟准检定法的第一步还是对所有观测值作经典最小二乘平差处理。文献[2,3]利用拟准检定法处理GPS周跳时,认为模糊度不变,实质上这与常规的周跳处理并无大的区别,即对观测

* 收稿日期:2005-04-27
基金项目:中国科学院动力大地测量学重点实验室开放基金(L04-07);中国科学院“百人计划”项目;“基础地理信息与数字化技术”;山东省重点开放实验室课题(SD040203)
作者简介:刘根友,博士,副研究员,长期从事大地测量、GPS、形变分析与研究

值(非差、单差、双差、双差与理论双差观测值之差)采用包括直线在内的多项式拟合来判定周跳。在 GPS 测量中,如果同时考虑坐标和模糊度为未知数,按最小二乘估计,则残差分布不一定出现分群特性。本文沿用文献[1]的拟准观测概念,给出了粗差探测的两种处理途径:第一种方法是认为所有观测值都含有粗差 G,并作为未知数,采用拟稳平差的原理($G_r^T G_r = \min$)处理秩亏问题,r 对应的观测值称为拟准观测值(不含粗差),可直接解算出包括粗差 G 和其它参数在内的所有未知参数;第二种方法是对 r 个被认为是准观测值的观测方程采用部分最小二乘原理($V_r^T V_r = \min$),再由平差值 X 计算其它观测方程的残差,并作为粗差。这两种处理方法在数值上都获得了拟准检定法同样的结果,其核心是部分最小二乘法,关键是如何选择(拟)准观测值。

2 最小二乘法与平差因子

设高斯-马尔柯夫模型为

$$\underset{n\times1}{\Delta} = \underset{n\times m}{A}\underset{m\times1}{\tilde{X}} - \underset{n\times1}{L},\ \Delta \in N(0,\sigma^2 Q) \tag{1}$$

观测方程为(以下简称为问题(2))

$$V = AX - L \text{ 或 } L = AX + \varepsilon,\ P = Q^{-1} \tag{2}$$

L 为 n 维观测向量,Δ 为观测值的真误差,V 为观测值的改正数(残差、余差),ε 为观测噪声,A 为设计矩阵,X 为 m 维未知参数向量(待估量),$\tilde{X}$ 为 X 的真值,P 为观测值的权阵。根据加权最小二乘原理($V^T PV = \min$)有:

$$A^T PV = O \tag{3a}$$

$$X = (A^T PA)^{-1} A^T PL \tag{3b}$$

O 为零矩阵(下文相同),将式(3b)代入式(2)可以获得观测值的改正数(残差):

$$\begin{aligned} V &= A(A^T PA)^{-1} A^T PL - L \\ &= [A(A^T PA)^{-1} A^T P - I]L \end{aligned} \tag{4}$$

式中,I 为单位阵,引入平差因子的定义[7]:

$$\begin{aligned} &J = A(A^T PA)^{-1} A^T P, \\ &R = I - J = I - A(A^T PA)^{-1} A^T P \end{aligned} \tag{5}$$

则:

$$V = -RL \tag{6}$$

因此,平差因子 J、R 具有最小二乘原理的数学意义,一旦使用 J、R 则表明已经对问题(2)采用了最小二乘准则($V^T PV = \min$),由平差因子和观测值可以直接计算改正数 V。

如果对方程(1)实施最小二乘准则($\Delta^T P\Delta = \min$),同样有:

$$\Delta = -RL \tag{7}$$

此时 Δ 不再具有真误差的性质。由于都使用了最小二乘原理,此时的 V 和 Δ 尽管符号不同,但都只能代表误差的估值。

根据平差因子的定义,平差因子 R 为 $n\times n$ 阶方阵,不难推导平差因子的特性[8]:

$$JA = A \quad RA = O \quad JR = O \tag{8}$$

$$JJ = J \quad RR = R \tag{9}$$

当 A 为非奇异方阵时,J = I, R = O,即此时没有多余观测,残差为零。

对(6)、(7)两式同时左乘 R 分别得到:

$$RV = -RL \tag{10}$$

$$R\Delta = -RL \tag{11}$$

由于 R 的引用,平差问题的表达更为简洁,但当观测数据太多时,R 的维数相当大,使得平差因子的应用受到一定的限制。在大量观测数据处理时,一般仅储存观测方程($n\times m$ 维)或法方程($m\times m$ 维)。

3 全部粗差作为待估参数(方案一)

假设每一观测值都含有粗差。这种假设的合理性在于,当观测值没有粗差时,其值就为零。设观测值的粗差为 G,如果把粗差也作为待估参数,则包含粗差和偶然误差在内的观测方程可写为(只考虑等权情况)

$$V_1 = AX - G - L \tag{12}$$

其中 $G = (g_1, g_2, \cdots, g_n)^T$。上式为秩亏方程,秩亏数为 m,由最小二乘原理($V_1^T V_1 = \min$)有:

$$\begin{cases} \dfrac{\partial(V_1^T V_1)}{\partial X} = 2V_1^T A = O \text{ 或} \\ A^T AX - A^T G - A^T L = O \\ \dfrac{\partial(V_1^T V_1)}{\partial G} = 2V_1^T(-E) = O \text{ 或} \\ G = AX - L \end{cases} \tag{13}$$

若($A^T A$)可逆,则有:

$$X = (A^T A)^{-1} A^T (G + L) \tag{14}$$

$$RG = -RL \tag{15}$$

比较式(10)、(11)和(15),3 个方程形式上都相同,由于存在显式(6)和(7),式(10)、(11)没有存在的必要,只有当形如式(12)的观测方程存在时,在最小二乘意义下,式(15)才具有实际意义。从式(13)的第二式可以看出,G 与问题(2)中的 V 相同。

由于 R 秩亏,不能直接解算粗差 G,必须施加其它约束条件。可以根据拟稳平差的思想附加条件约束,拟准检定法是对真误差采取约束来定位粗差,本文对粗差(待估参数)进行约束。我们还可以把那些不含粗差的正常观测值所对应的粗差未知数约束为零(称为强约束)。约束方程的个数必须大于等于 m。不失一般性,观测方程(12)具有条件约束时,可

以表示为

$$\begin{cases} V = AX - G - L = (A \quad -E)\begin{pmatrix} X \\ G \end{pmatrix} - L \\ (C_1 \quad C_2)\begin{pmatrix} X \\ G \end{pmatrix} = O \end{cases} \tag{16}$$

按附加条件最小二乘解有:

$$A^T AX - A^T G + C_1^T K = A^T L \tag{17a}$$

$$-AX + G + C_2^T K = -L \tag{17b}$$

$$C_1 X + C_2 G = O \tag{17c}$$

若 $K = O$ 时(条件方程的系数阵必须是 m 行满秩,且 $(C_1 \quad C_2)(A \quad -E)^T = O$,不难推导出:

$$X = (A^T A)^{-1} A^T (G + L) \tag{18}$$

$$G = AX - L \text{ 或 } RG = -RL \text{ 或 } V = O \tag{19}$$

若 $C_1 = O$ 且 $K = O$ 时,还存在

$$G = -(R + C_2^T C_2)^{-1} RL \tag{20}$$

分 5 种特殊约束条件加以讨论:

1) 当 $C = \begin{pmatrix} C_1 & C_2 \\ {}_{m\times m} & {}_{m\times n} \end{pmatrix} = \begin{pmatrix} E & A^T \\ {}_{m\times m} & {}_{m\times n} \end{pmatrix}$ 时,即 $X + A^T G = O$,有:

$$(E \quad A^T)(A \quad -E)^T = EA^T - A^T E = O \tag{21}$$

由秩亏自由网平差的性质知,此时 C 为最小范数解的条件系数阵,且 $K = O$,即:

$$\begin{pmatrix} X \\ G \end{pmatrix}^T \begin{pmatrix} X \\ G \end{pmatrix} = X^T X + G^T G = \min \tag{22}$$

2) 当 $C = \begin{pmatrix} O & A^T \\ {}_{m\times m} & {}_{m\times n} \end{pmatrix}$ 时,即 $A^T G = O$,此时,C 中部分列的所有元素为零,且秩仍然是 m,则对应部分最小范数解(拟稳平差):

$$G^T G = \min \tag{23}$$

此时,$C_2 = A^T$,且 $K = O$,由式(18)、(19)得:

$$\begin{cases} X = (A^T A)^{-1} A^T L \\ G = -RL \end{cases} \tag{24}$$

X 与问题(2)的最小二乘解相同,G 等于问题(2)的最小二乘残差。

3) 当 $C = \begin{pmatrix} O & O & A_r^T \\ {}_{m\times m} & {}_{m\times(n-r)} & {}_{m\times r} \end{pmatrix}$,且 $r > m$ 时,即 $A_r^T G_r = 0$,r 称为拟准观测数,对应的观测值称为拟准观测值,此时部分最小范数条件为:

$$G_r^T G_r = \min \tag{25}$$

4) 当 $C = \begin{pmatrix} O & E \\ {}_{n\times m} & {}_{n\times n} \end{pmatrix}$ 时,即 $G = O$,此时 $K \neq O$,为强约束条件,即所有观测值都不存在粗差。

$$\begin{cases} X = (A^T A)^{-1} A^T L \\ G = O \\ V = -RL \\ K = AX - L = -RL = V \end{cases} \tag{26}$$

与问题(2)的最小二乘解相同。

5) 当 $C = \begin{pmatrix} O & O & E \\ {}_{r\times m} & {}_{r\times(n-r)} & {}_{r\times r} \end{pmatrix}$,且 $m \leq r < n$ 时,即 $G_r = O$,此时 $K_r \neq O$,为部分强约束条件,即对应的观测值不存在粗差,相应的观测值为准观测值。

结合文献[1]的思路,第一步认为所有粗差满足 $G^T G = \min$,即所有观测值都是拟准观测值,此时的粗差 G 等于最小二乘解的 V(第 2) 种情形)。第二步选择粗差较小的分群作为拟准观测值,按 $G_r^T G_r = \min$ 解问题(12) (第 3) 种情形),直到拟准观测值和其它观测值的粗差(g_i)具有明显的分群特性。这种处理粗差的途径称为方案一。由于所有残差 V 都为零,可采用以下公式计算验后单位权中误差:

$$\sigma_r = \sqrt{\frac{G_r^T G_r}{r - m}} \tag{27}$$

4　部分最小二乘解(方案二)

部分最小二乘法解问题(2)的基本条件是:

$$V_r^T V_r = \min \text{ 或 } A_r^T V_r = O \tag{28}$$

即只对 r 个观测值进行最小二乘平差,与方案一中 $G_r^T G_r = \min$ 对应的拟准观测不同,r 对应的观测值称为准观测值。其余观测值的改正数利用部分最小二乘解 X_r 反算,验后单位权中误差采用以下公式计算:

$$\sigma_r = \sqrt{\frac{V_r^T V_r}{r - m}} \tag{29}$$

方案二的具体实施步骤如下:

1) 将所有观测值作为准观测值,即经典最小二乘平差,获得残差 V。

2) 计算指标

$$\sigma_s = \alpha \frac{\sum |v_i|}{n} \tag{30}$$

系数 α 一般取为 0.8。将观测值分为两类:准观测值和非准观测值,只有非准观测值中含有粗差。选取 v_i 小于 σ_s 的观测值作为准观测值,准观测数 r 应大于必要观测数 m,若 $r < m$,可适当放大系数 α。在准观测值的基础上进行部分最小二乘平差获得新的 X_r,并根据 X_r 计算非准观测值的改正数 $V_2^{(1)}$。

3) 根据式(29)计算中误差 σ_r,如果 $V_2^{(1)}$ 中还存在小于 3 倍 σ_r 的残差,将对应的观测值纳入准观测值类,重新平差,并计算其余非准观测值的残差 $V_2^{(2)}$,直到 $V_2^{(k)}$ 只含有大于 3 倍 σ_r 的粗差为止。

事实上,方案一和方案二两种处理结果是等价的。因为在方案一中,从(19)式可以看出,$V = O$,而 G 与方案二中 V 的特性相同,只需要将 G 当作 V 即可。尽管如此,方案一的 $G_r^T G_r = \min$ 和方案二中的 $V_r^T V_r = \min$ 却具有不同的平差意义,前者的思想为拟稳平差,后者的思想为部分最小二乘法。在编制程序时,如果观测值数量较少,可以采用方案一,此时不需要重新形成法方程,但仍然需要储存观测方

程的系数,根据所选取的拟准观测值改变条件方程的系数阵,则可以同时解算粗差 G 和待估参数 X。当观测值数量较多时,宜采用方案二,此时需要储存观测方程,而且需要选择准观测值重新生成法方程,获得待估参数 X 后,再代入非准观测方程计算残差(粗差)。

5　算例

利用文献[1]的例 2 进行计算,该算例为具有 18 个角度观测值的二维测角网,有两个待定点。由于两种方案的等价性,在本算例中我们采用方案一,当粗差作为未知数时,未知数的总数为 4 + 18 = 22,条件方程个数为 4,待定参数 K 的维数为 4,建立形如(17)式的法方程。以 3 倍中误差作为粗差标准。首先设 $C = \begin{pmatrix} C_1 & C_2 \end{pmatrix} = \begin{pmatrix} O_{4\times4} & A^{T}_{4\times18} \end{pmatrix}$,引进辅助向量 q_i ($i = 1,2,\cdots,18$),当 $q_i = 0$ 时,C_2 中对应的第 i 列元素全为零;$q_i = 1$ 时 C_2 中对应的第 i 列元素保持不变,此时表明第 i 个观测值为准观测值,计算结果见表 1。具体步骤如下:

1) 令 $q_i = 1$ ($i = 1,2,\cdots,18$),即所有观测值都为拟准观测值,计算结果与经典最小二乘平差一致,此时 G = V。

2) 计算指标 $\sigma_s = 0.8\dfrac{\sum|g_i|}{n} = 2.02$,取 g_i 小于 σ_s 的观测值为拟准观测值,有 $q_i = 1$ ($i = 2,3,4,6,7,9,13,14,18$)。

3) 利用拟准观测值作部分约束,此时将矩阵 C_2 中 $q_i = 0$ ($i = 1,5,8,10,11,12,15,16,17$)所对应的列元素全定义为零,重新解法方程(17),获中误差 $\sigma_r = 1.10$。取残差小于 $3\sigma_r$ 的观测值为拟准观测值,此时含粗差的观测值有 $q_i = 0$ ($i = 1,5,8,15,16$)。

4) 令矩阵 C_2 中 $q_i = 0$ ($i = 1,5,8,15,16$)所对应的列元素全为零,重新解法方程(17),获中误差 $\sigma_r = 0.94$,此时,非准观测值的残差均大于 $3\sigma_r$,即为检测出的粗差。

计算结果与文献[1]的结果相同。由于拟准标识向量 q_i 的引入,在编制程序时不需要如文献[2～6]对观测方程进行重新排序,计算思路更为清晰,编制程序更简化。

值得注意的是,无论采用哪种粗差检定方法获得的粗差,其大小对平差值没有任何贡献,实质上是删除了具有粗差的观测值,因此,诊断粗差的位置才是数据处理的关键。只有当粗差具有整数特性时(如 GPS 载波相位测量的周跳),确定粗差的大小才具有实质上的意义。

表 1　粗差检测实施步骤

Tab. 1　Steps of gross error detection

观测号	观测值	含粗差	普通最小二乘解		部分最小二乘解(1)		部分最小二乘解(2)	
			标识符	残差/粗差	标识符	残差/粗差	标识符	残差/粗差
1	-6.8	7.0	1	6.30		8.00		8.22
2	0.6		1	1.92	1	1.75	1	1.67
3	-3.1		1	1.08	1	-0.46	1	-0.59
4	0.9		1	-0.11	1	-0.13	1	0.18
5	7.5	-7.0	1	-4.87		-5.62		-6.11
6	-2.6		1	-0.83	1	-0.05	1	0.13
7	3.1		1	1.58	1	0.41	1	-0.07
8	-14.1	5.6	1	5.99		7.94		8.60
9	1.9		1	1.53	1	0.75	1	0.57
10	1.2		1	3.31		0.74	1	-0.02
11	-2.9		1	-2.25		-1.33	1	-0.76
12	3.3		1	-2.66		-1.01	1	-0.81
13	4.0		1	-1.19	1	-0.55	1	-0.75
14	8.5		1	-0.97	1	-1.23	1	-1.47
15	-6.4	-6.8	1	-3.97		-4.34		-3.90
16	2.8	6.8	1	2.89		4.43		4.47
17	-10.7		1	2.99		-0.09	1	-0.26
18	3.1		1	-1.08	1	0.46	1	0.59
			$r = 18$, $\sigma_s = 2.02$		$r = 9$, $\sigma_r = 1.10$		$r = 13$, $\sigma_r = 0.94$	

6 结论

本文给出粗差检定的两种计算方案与 QUAD 具有相同的效果。方案一中的 G 与方案二中的 V 在数值上是相同的,前者的理论基础为拟稳平差,后者的理论基础为部分最小二乘法。待估参数的估值仅与(拟)准观测值有关,等价于删除了含粗差的观测值,粗差的值仅仅是根据(拟)准观测方程的参数平差值计算的。要估计粗差,在平差问题中必须有多余观测值,且(拟)准观测数必须大于或等于必要观测数。由于标识向量 q_i 的引入,本文的计算方法更加简单清晰。

References

1 欧吉坤. 粗差的拟准检定法(QUAD 法)[J]. 测绘学报, 1999, 28(1): 15～20.

1 Ou Jikun. Quasi-Accurate Detection gross errors (QUAD)[J]. Acta Geodetica et Cartographica Sinica, 1999, 28(1): 15 - 20. (in Chinese)

2 韩保民, 欧吉坤, 柴艳菊. 用拟准检定法探测和修复 GPS 数据中的粗差和周跳[J]. 武汉大学学报信息科学版, 2002, 27(3): 246～250.

2 Han Baomin, Ou Jikun and Chai Yanju. Detection and repair the gross errors and cycle slips by QUAD method [J]. Geomatics and Information Science of Wuhan University, 2002, 27(8): 372 - 376. (in Chinese)

3 Baomin Han and Jikun Ou. Detection and repair of continuous cycle slips in GPS carrier-phase data using QUAD Method(C). ION - GPS, 2002, 1 353 - 1 360.

4 柴艳菊, 欧吉坤, 独知行. 拟准检定法用于划分不同运动块体的探索[J]. 地震学报, 2002, 24(6): 579～586.

4 Chai Yanju, Ou Jikun and Du Zhixing. A new approach for distinguishing different deformation trend blocks with displacement observations[J]. Acta Seismologica Sinica, 2002, 15(6): 607 - 615. (in Chinese)

5 柴艳菊, 欧吉坤. 同时检测监测网中的异常和粗差[J]. 大地测量与地球动力学, 2003, 23(1): 64～68.

5 Chai Yanju and Ou Jikun. A new approach for distinguishing different deformation trend blocks with displacement observations[J]. Journal of Geodesy and Geodynamics, 2003, 23(2): 64 - 68. (in Chinese)

6 黄立人. 用于相对稳定点组判别的 QUAD[J]. 大地测量与地球动力学, 2002, 22(2): 10～15.

6 Huang Liren. QUAD method used for identifying relatively stable stations[J]. Journal of Geodesy and Geodynamics, 2002, 22(2): 10 - 15. (in Chinese)

7 周江文. 平差因子与伪逆平差[A]. 测量误差理论新探[C]. 北京: 测绘出版社, 1999.

7 Zhou Jiangwen. Adjustment factor and pseudoinverse estimates[A]. In: Zhou Jiangwen, et al. Research on the theory of observation data and errors[C]. Beijing: Surveying and Mapping Press, 1999, 41 - 44. (in Chinese)

8 陶本藻. 平差因子与平差结构[J]. 大地测量与地球动力学, 2002, 22(3): 6～9.

8 Tao Benzao. Adjustment factor and adjustment structure [J]. Journal of Geodesy and Geodynamics, 2002, 22(3): 6 - 9. (in Chinese)

加入台湾华艺 CEPS 中文电子期刊服务的声明

《大地测量与地球动力学》自 2005 年 1 月起,加入台湾中文电子期刊服务 ——思博网(CEPS)。中文电子期刊服务 ——思博网是目前台湾地区最大的期刊全文数据库,其访问地址为:www.ceps.com.tw。自此,读者可以通过这一网址检索《大地测量与地球动力学》从 2005 年起各期的全文,在一段时期后,还可以回溯检索 2002 年以来历年的全文。

此外,由于《大地测量与地球动力学》被 CEPS 收录,故凡向本刊投稿者,均视为其文稿刊登后可供思博网(CEPS)收录、转载并上网发行,其作者文章著作权使用费与稿酬一次付清,本刊不再另付其它报酬。

请各位继续支持本刊,谢谢!

《大地测量与地球动力学》