

尺度因子的MAD估计及其在测量平差中的应用

郭建锋

引用本文:

郭建锋. 尺度因子的MAD估计及其在测量平差中的应用[J]. 武汉大学学报·信息科学版, 2021, 46(11): 1636–1640.

GUO Jianfeng. MAD Estimate of Scale Factor and Its Applications in Measurement Adjustment[J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(11): 1636–1640.

相似文章推荐 (请使用火狐或IE浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

一种使用抗差估计的保护水平重构方法

A Method of Protection Level Reconstruction Based on Robust Estimation

武汉大学学报·信息科学版. 2021, 46(1): 96–102 <https://doi.org/10.13203/j.whugis20190043>

基于渐消因子的改进Kalman滤波时间尺度估计算法

An Improved Kalman Filter Time Scale Algorithm Based on Forgetting Factor

武汉大学学报·信息科学版. 2019, 44(8): 1205–1211,1219 <https://doi.org/10.13203/j.whugis20180065>

多变量稳健总体最小二乘平差方法

Robust Total Least Squares Method for Multivariable EIV Model

武汉大学学报·信息科学版. 2019, 44(8): 1241–1248 <https://doi.org/10.13203/j.whugis20170232>

利用小波包系数信息熵去除GNSS站坐标时间序列有色噪声

Elimination of Colored Noise in GNSS Station Coordinate Time Series by Using Wavelet Packet Coefficient Information Entropy

武汉大学学报·信息科学版. 2021, 46(9): 1309–1317 <https://doi.org/10.13203/j.whugis20190353>

顾及系统噪声和观测噪声的分级自适应信息滤波算法

Hierarchical Adaptive Information Filtering Algorithm Considering System Noise and Observation Noise

武汉大学学报·信息科学版. 2021, 46(1): 88–95 <https://doi.org/10.13203/j.whugis20190248>



尺度因子的MAD估计及其在测量平差中的应用

郭建锋¹

¹ 中原工学院理学院,河南 郑州,450007

摘要:观测数据受到粗差污染时,平差结果往往失真,不可采用。选择合适的统计检验量是对测量数据进行粗差处理的关键一环,而构造统计检验量必须对尺度因子做出估计。首先对样本尺度因子的绝对中位差(median absolute deviation, MAD)估计进行了讨论,并详细探讨了其中涉及到的Fisher一致性调节因子的确定;然后在此基础上,分别提出了平差模型中基于标准化残差和一致最大功效(uniformly most powerful, UMP)统计检验量序列的尺度因子的两种抗差估计方法,尺度因子的两种抗差估计都可用于构造相应的统计检验量以识别和定位可疑观测;最后对全球导航卫星系统(global navigation satellite systems, GNSS)网平差进行具体数值计算,结果表明,尺度因子的MAD估计具有良好的抗差性,不但可用于粗差处理,还可用于平差成果的精度评定。

关键词:尺度因子;粗差探测;抗差估计;MAD估计

中图分类号:P207

文献标志码:A

测量数据处理中,解决粗差问题是不可避免的环节。处理粗差目前有粗差探测与抗差估计两种方法^[1-12]。粗差探测把粗差归结为函数模型的不完善,而抗差估计则通过方差膨胀的方式削弱甚至消除粗差的不良影响。选择合适的统计检验量以准确地探测和识别异常观测是使用这两种方法的前提条件。

由于尺度因子通常是未知的,可基于最小二乘平差残差的加权平方和进行估计,然后构造出服从 τ 分布的统计量或者服从学生氏分布的 t 统计量^[8-10]。但该方法几乎不具有抵御粗差影响的能力,因此需要做进一步的工作。尺度因子的绝对中位差(median absolute deviation, MAD)估计由于其简单易操作、抗差性强等特点,是目前粗差探测与抗差估计领域应用最为广泛的一种估计尺度因子的方法^[1,13-14]。

本文首先对样本尺度因子的MAD估计进行了讨论,详细探讨了Fisher一致性调节因子的确定。在此基础上,论述了平差模型中尺度因子的MAD估计及其在测量数据处理中的应用。最后结合全球导航卫星系统(global navigation satellite

systems, GNSS)网平差的数值计算结果讨论了测量平差成果的精度评定问题。

1 样本尺度因子的MAD估计

设一元随机变量 Y 服从期望为 μ 、方差为 σ^2 的分布,对其进行 n 次独立观测,得到概率分布相同的样本序列为 Y_1, Y_2, \dots, Y_n 。如果总体的期望 μ 已知,基于样本序列可对方差 σ^2 做如下估计:

$$\bar{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \mu)^2 \quad (1)$$

由于算术平均值(mean)运算不具有抗差性,上述方差估计方法对粗差非常敏感,中位数(median)运算的崩溃污染率却达到50%,因此以中位数代替式(1)中的平均值运算,得到:

$$\bar{\sigma} = k \cdot \text{med}_i | Y_i - \mu | \quad (2)$$

式中,med表示中位数运算;参数 k 为保证 $\bar{\sigma}$ 满足Fisher一致性的调节因子。 k 与 Y 的概率分布有关,不同的概率分布对应的调节因子一般不同^[1,13-14]。

记 $m = \text{med}_i | Y_i - \mu |$,则根据中位数的定义

可得:

$$P\{|Y - \mu| < m\} = P\{|Y - \mu| \geq m\} = 0.5 \quad (3)$$

如果随机变量 Y 的概率密度函数关于 $Y = \mu$ 对称,则有:

$$P\{Y - \mu \geq m\} = P\{Y - \mu \leq -m\} \quad (4)$$

从而得到:

$$P\left\{\frac{Y - \mu}{\sigma} \geq \frac{m}{\sigma}\right\} = 0.25 \quad (5)$$

这表明 $\frac{m}{\sigma}$ 是随机变量 $\frac{Y - \mu}{\sigma}$ 的累积分布函数的上 0.25 分位点。进而可证明,当 Y 服从正态分布时, $\frac{Y - \mu}{\sigma}$ 服从标准正态分布,其上 0.25 分位点为:

$$\Phi^{-1}(1 - 0.25) = 0.6745 \quad (6)$$

此时,

$$\hat{\sigma} = \frac{\text{med}_i |Y_i - \mu|}{0.6745} = 1.4826 \cdot \text{med}_i |Y_i - \mu| \quad (7)$$

特别地,当 $\mu = 0$ 时,式(7)变为:

$$\hat{\sigma} = 1.4826 \cdot \text{med}_i |Y_i| \quad (8)$$

如果总体的期望 μ 未知,可先求样本均值,再得到样本方差的估值如下:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (9)$$

式中, $\bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j$ 。根据文献[4],式(9)得到的估计是有偏的,为保证其无偏,还需要乘以因子 $\frac{n}{n-1}$ 。

由式(9)可以得到尺度因子 σ 的估计如下:

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^n |Y_i - \bar{Y}| \quad (10)$$

由于估计量 $\hat{\sigma}$ 的崩溃点是 0,同时其影响函数是无界的,因此不能错误地认为式(10)是尺度因子的抗差估计^[14]。

为提高式(10)的抗差性,将其中的两次平均值运算都以中位数运算代替,得到:

$$\text{MAD} = \text{med}_i \left| Y_i - \text{med}_j \{Y_j\} \right| \quad (11)$$

式中, MAD 表示尺度因子 σ 的(关于中位数的)绝对偏差的中位数估计。

若以 MAD 作为尺度因子 σ 的估计,同样必须使其满足 Fisher 一致性,可直接以式(2)中的调节因子来代替。当 Y 服从正态分布时,得到尺度因子 σ 的如下估计^[1,13-14]:

$$\hat{\sigma} = 1.4826 \cdot \text{MAD} \quad (12)$$

尺度因子的所有估计形式里,基于 MAD 的估计具有最高的崩溃污染率,然而 MAD 估计的抗差性是以牺牲统计量的渐进效率为代价的,其原因在于大多数观测对 MAD 估计没有做出直接贡献。在统计学领域,文献[14]提出了几种替代 MAD 估计的方法,但 MAD 估计由于其简单易操作、抗差性强等优点,仍然是目前应用最为广泛的一种方法。

2 平差模型中尺度因子的抗差估计

参考文献[4]中给出的线性平差模型:

$$E(L) = AX, \text{Cov}(L) = \sigma^2 P^{-1} \quad (13)$$

式中, L 为服从正态分布的 $n \times 1$ 观测向量; A 为 $n \times u$ 列满秩的设计阵; X 为 $u \times 1$ 的未知参数向量; σ^2 为单位权方差因子; 权阵 P 满足对称正定性。

基于极大似然估计原理,得到^[4]:

$$\begin{cases} \hat{X} = (A^T P A)^{-1} A^T P L \\ \hat{\sigma}^2 = \frac{1}{n} V^T P V \end{cases} \quad (14)$$

其中,

$$V = L - A \hat{X} \quad (15)$$

称为残差向量。

由此易证明 $\hat{\sigma}^2$ 是 σ^2 的有偏估计,为使其满足无偏性质,需做如下修正^[4]:

$$\hat{\sigma}^2 = \frac{1}{n-u} V^T P V \quad (16)$$

然而,此时这两种估计都不能有效抵御粗差的影响。如果假定残差向量的期望为零,则对于残差序列 $v_1, v_2 \dots v_n$ 可采用式(8)对尺度因子进行估计,即绝对残差的中位数 (median absolute residual, MAR) 估计,具体计算如下:

$$\hat{\sigma} = 1.4826 \cdot \text{med}_i |v_i| \quad (17)$$

采用式(8)估计尺度因子需符合以下条件:

(1) 观测序列应该是独立的; (2) 序列应服从相同的概率分布; (3) 观测序列的期望值为零。当观测测量受到粗差污染时,残差序列或将不满足以上 3 个条件中的任何一条,因此使用式(8)存在一定的风险。为此,本文首先对残差进行标准化,得到均服从正态分布 $N(0, \sigma^2)$ 的序列 $\sigma \bar{v}_1, \sigma \bar{v}_2 \dots \sigma \bar{v}_n$ 。这里符号 \bar{v}_i 表示第 i 个标准化残差,计算如下:

$$\bar{v}_i = \frac{v_i}{\sqrt{\text{Var}(v_i)}} \quad (18)$$

进而可得到尺度因子的MAD估计如下:

$$\bar{\sigma}_v = 1.4826 \cdot \text{med}_i \left| \sigma \bar{v}_i - \text{med}_j \{ \sigma \bar{v}_j \} \right| \quad (19)$$

以 h_i 记第 i 个元素为 1 其余均为 0 的 n 维单位向量, 则 $w_i = h_i^T PV / \sqrt{\text{Var}(h_i^T PV)}$ 为第 i 个一致最大功效 (uniformly most powerful, UMP) 统计检验量^[16]。UMP 统计量在独立观测情况下退化为标准化残差, 但在相关观测情况下, 二者是完全不同的统计量^[15-17]。构造如下具有相同概率分布 $N(0, \sigma^2)$ 的序列 $\sigma w_1, \sigma w_2, \dots, \sigma w_n$ 。类似于标准化残差的情形, 可得到尺度因子的 MAD 估计如下:

$$\bar{\sigma}_w = 1.4826 \cdot \text{med}_i \left| \sigma w_i - \text{med}_j \{ \sigma w_j \} \right| \quad (20)$$

式(17)、式(19)和式(20)分别给出了平差模型中尺度因子 3 种不同形式的抗差估计。

3 尺度因子的 MAD 估计在粗差处理中的应用

粗差探测与抗差估计都必须面对统计量的选择问题。虽然可供选择的统计量多种多样, 但由于尺度因子在许多情况下是未知的, 因此不可避免地首先要对尺度因子进行估计。一种简单直接的方法是基于式(16)给出的方差的无偏估计构造出统计量, 比如 τ 分布和 t 分布^[8-10]。但由于式(16)几乎不具有抗差性, τ 分布和 t 分布使用起来均潜在风险, 尤其在只有一个模型冗余度的极端条件下, 根本无法定义 t 统计量, 而此时无论观测受到粗差污染与否, 所有 τ 统计检验量的绝对值始终恒等于 1, 因此同样无法用于粗差探测与识别^[17]。

如果以服从正态分布的标准化残差或 UMP 统计量为出发点, 并将涉及的标准差 σ 分别用其抗差 MAD 估计(式(19)或式(20))代替, 可得到两类相关观测条件下近似服从正态分布的抗差统计检验量。由于 UMP 统计量在同类统计检验量中具有最大功效, 因此使用抗差 UMP 统计量可以更为准确地进行粗差的探测与识别^[15-17]。

平差的目的主要是对未知参数进行估计, 以及对得到的未知参数估值进行精度评定。使用基于 MAD 的尺度因子可以使统计检验量无量纲化, 进而可将离群的统计量对应的观测值标记为可疑观测值, 对可疑观测值进行剔除或降权处理后, 最终将得到可靠的测量成果。

§2 给出的式(17)、式(19)和式(20)都是对样

本的尺度因子进行估计。在对平差成果进行精度评定时, 需要估计总体的尺度因子。为保证估值的渐进无偏性, 综合式(9)和式(16), 本文提出尺度因子的近似估计如下:

$$\hat{\sigma}_w = \sqrt{\frac{n}{n-1}} 1.4826 \text{med}_i \left| \sigma w_i - \text{med}_j \{ \sigma w_j \} \right| \quad (21)$$

如果使用的是粗差剔除方法或者是包含拒绝域的抗差估计方法, 式中的 n 还应扣除那些被剔除或者权降为零的观测量的数目。显然, 当观测测量个数很大时, 式(21)与式(20)仅有细微差异。

4 GNSS 网平差算例分析

本文实验中的 GNSS 网包含一个控制点和 5 个未知点, 共观测了 45 条独立基线的数据^[18]。借助于已有的设计阵和待定点坐标参考值, 以及观测量的方差-协方差阵, 可以模拟出服从正态分布的观测量^[19]。在此基础上, 随机指定 1~6 个观测测量分别独立加入大小、符号不同的粗差, 具体可参看文献[16], 虽然基于 UMP 统计量和基于标准化残差的抗差估计方案都能够有效抵御粗差的不良影响, 但基于 UMP 统计量的抗差估计在抗差性和计算效率两个方面都优于基于标准化残差的抗差估计。

为进一步研究 MAD 估计在精度评定中的应用, 本文进行了如下模拟计算:(1)方案 1: 对没有植入粗差的数据采用最小二乘平差进行参数估计, 最终依据式(16)对尺度因子进行估计;(2)方案 2: 同方案 1, 但采用式(21)估计尺度因子;(3)方案 3: 对受粗差污染后的数据采用基于 UMP 统计量的抗差估计进行参数解算, 根据式(16)估计尺度因子并用于精度评定;(4)方案 4: 同方案 3, 但采用式(21)估计尺度因子。

为保证评价的客观性, 上述模拟进行了 10 000 次, 图 1 统计了 4 种方案 10 000 次 GNSS 网平差模拟数据得到的尺度因子估值序列, 统计结果见表 1。

由图 1、表 1 可知, 没有粗差污染的情况下, 使用残差加权平方和方法是最好的选择, 此时 MAD 估值序列呈现出一定的波动性; 对于受污染的数据, 采用抗差估计尽管可以在一定程度上抵御了粗差的影响, 但基于残差加权平方和方法得到的尺度因子明显偏小, 而且序列中存在“突刺”现象; 相对而言, 基于 MAD 的尺度因子估值表现更为稳健。

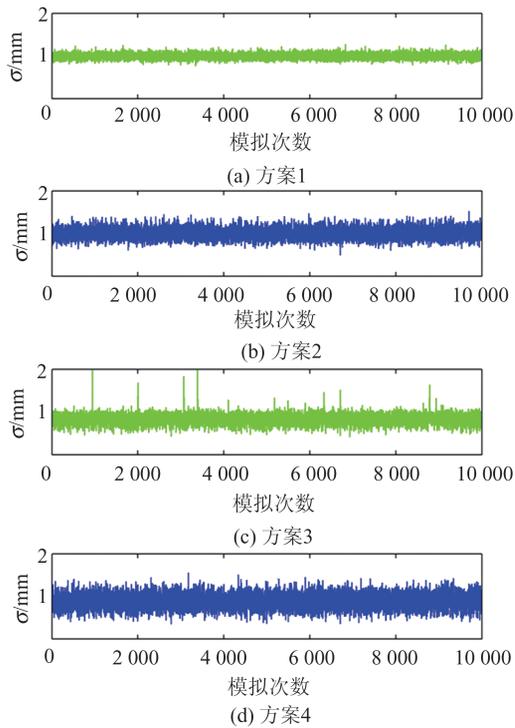


图 1 4 种方案 10 000 次 GNSS 网平差模拟数据得到的尺度因子估值序列

Fig.1 Estimated Scale Factors of Four Schemes over 10 000 Repetitions in GNSS Network Adjustment

表 1 4 种方案得到的尺度因子估值序列的统计结果/mm

Tab.1 Statistics of Estimated Scale Factor of Four Schemes/mm

方案	最大值	平均值	中位数
方案 1	1.28	1.00	1.00
方案 2	1.52	1.00	1.00
方案 3	2.09	0.84	0.85
方案 4	1.56	0.89	0.90

5 结 语

粗差探测与抗差估计是目前处理粗差的两种常用方法。选择合适的统计量是探测和识别异常观测的关键。本文讨论了样本尺度因子的 MAD 估计,以及 Fisher 一致性调节因子的确定方法,并对平差模型中尺度因子的 MAD 估计及其在粗差处理中的应用进行了探讨。GNSS 网平差的数值例子表明,由于具有良好的抗差性,尺度因子的 MAD 估计不但可用于构造粗差探测与识别的统计检验量,还可用于平差成果的精度评定。

参 考 文 献

[1] Huber P J. Robust Statistics [M]. New York: Wiley, 1981

[2] Zhou Jiangwen, Huang Youcai, Yang Yuanxi, et al. Robust Least Squares Method [M]. Wuhan: Huazhong University of Science and Technology Press, 1997(周江文,黄幼才,杨元喜,等.抗差最小二乘法[M].武汉:华中理工大学出版社,1997)

[3] Wolf P R, Ghilani C D. Adjustment Computations: Statistics and Least Squares in Surveying and GIS [M]. New York: Wiley, 1997

[4] Koch K R. Parameter Estimation and Hypothesis Testing in Linear Models [M]. Berlin: Springer-Verlag, 1999

[5] Leick A, Rapoport L, Tatarnikov D. GPS Satellite Surveying[M]. New York: Wiley, 2015

[6] Barnett V, Lewis T. Outliers in Statistical Data [M]. New York: Wiley, 1994

[7] Baarda W. A Testing Procedure for Use in Geodetic Networks [J]. *Netherlands Geodetic Commission, Publications on Geodesy*, 1968, 2(5): 1-97

[8] Pope A J. The Statistics of Residuals and the Detection of Outliers [R]. Rockville: NOAA Technical Report, 1976

[9] Chatterjee S, Hadi A S. Sensitivity Analysis in Linear Regression[M]. New York: Wiley, 1988

[10] Kargoll B. On the Theory and Application of Model Misspecification Tests in Geodesy [D]. Bonn: University of Bonn, 2007

[11] Lu Tieding, Yang Yuanxi, Zhou Shijian. Comparative Analysis of MDB for Different Outliers Detection Methods [J]. *Geomatics and Information Science of Wuhan University*, 2019, 44(2): 185-192 (鲁铁定,杨元喜,周世健.均值漂移模式几种粗差探测法的MDB比较[J].武汉大学学报·信息科学版,2019,44(2):185-192)

[12] Wang Haitao, Ou Jikun, Yuan Yunbin, et al. On Equivalence of Three Estimators for Outliers in Linear Model [J]. *Geomatics and Information Science of Wuhan University*, 2013, 38(2): 162-166 (王海涛,欧吉坤,袁运斌,等.估计观测值粗差三种方法的等价性讨论[J].武汉大学学报·信息科学版,2013,38(2):162-166)

[13] Hampel F R. The Influence Curve and its Role in Robust Estimation[J]. *Journal of the American Statistical Association*, 1974, 69(346): 383-393

[14] Rousseeuw P J, Croux C. Alternatives to the Median Absolute Deviation [J]. *Journal of the American Statistical Association*, 1993, 88(424): 1 273-1 283

[15] Guo J. Theory of Model Errors and its Applications in GPS Data Processing [D]. Wuhan: Institute of Geodesy and Geophysics of Chinese Academy of

- Sciences, 2007 (郭建锋. 模型误差理论若干问题研究及其在GPS数据处理中的应用[D]. 武汉: 中科院测量与地球物理研究所, 2007)
- [16] Guo J, Ou J, Wang H. Robust Estimation for Correlated Observations: Two Local Sensitivity-based Downweighting Strategies[J]. *Journal of Geodesy*, 2010, 84(4): 243-250
- [17] Guo J. A Note on The Conventional Outlier Detection Test Procedures[J]. *Boletim Ciencias Geodesicas*, 2015, 21(2): 433-440
- [18] Snow K B, Schaffrin B. Three-Dimensional Outlier Detection for GPS Networks and Their Densification via the BLIMPBE Approach [J]. *GPS Solutions*, 2003, 7(2): 130-139
- [19] Yang Y, Song L, Xu T. Robust Estimator for Correlated Observations Based on Bifactor Equivalent Weights [J]. *Journal of Geodesy*, 2002, 76(6): 353-358

MAD Estimate of Scale Factor and Its Applications in Measurement Adjustment

GUO Jianfeng¹

¹ College of Science, Zhongyuan University of Technology, Zhengzhou 450007, China

Abstract: Objectives: The least-squares method is very sensitive to outliers, and the adjustment outputs will usually be unacceptable when some of the observations are contaminated. Selection of appropriate statistical tests plays a pivotal role both in robust estimation and conventional outlier detection procedures. **Methods:** The MAD (median absolute deviation) estimate of scale factor in the univariate case is discussed firstly. Determination of the Fisher-consistency factor is described for Gaussian normal distribution. Robust estimates of scale factor in linear adjustment model are addressed based on standardized least-squares residuals and the uniformly most powerful test statistics, respectively. Both of them can be used for constructing statistical tests, to identify the potential outlying observations, and therefore their deterioration effect will be mitigated. For illustrative purpose, Monte Carlo simulations in GPS network adjustment scenario are performed. **Results:** Numerical results show that the MAD-based estimate of scale factor is robust and works well in accuracy assessment for adjustment outputs. **Conclusions:** Explicit formula for estimating the scale factor, the MAD is a very robust scale estimator and has low computation complexity. It is therefore appropriate to use the MAD for adjustment computations and accuracy assessment when outliers are present.

Key words: scale factor; outlier detection; robust estimation; MAD estimate

Author: GUO Jianfeng, PhD, specializes in geodesy and GNSS data processing. E-mail: jianfeng.guo@gmail.com

Foundation support: The National Natural Science Foundation of China (41674020, 40874007); the Henan Key Laboratory of Intelligent Public Opinion Analysis.

引文格式: GUO Jianfeng. MAD Estimate of Scale Factor and Its Applications in Measurement Adjustment[J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(11):1636-1640. DOI:10.13203/j.whugis20190166 (郭建锋. 尺度因子的MAD估计及其在测量平差中的应用[J]. *武汉大学学报·信息科学版*, 2021, 46(11):1636-1640. DOI:10.13203/j.whugis20190166)