

残差自适应回归 TLS*

陈西江^{1,2)} 花向红^{1,2)} 鲁铁定³⁾ 翟高鹏⁴⁾ 杨燕景³⁾

(1) 武汉大学测绘学院, 武汉 430079
(2) 武汉大学灾害监测与防治研究中心, 武汉 430079
(3) 东华理工大学测绘工程学院, 抚州 344000
(4) 河北省测绘产品质量监督检验站, 石家庄 050000

摘要 TLS在拟合及推估过程中,其拟合和推估的残差会随着观测数据量的增加而呈现上扬的趋势,为了抑制残差的上扬,将自适应回归模型引入到TLS的拟合及推估中,采用适宜的回归模型对经过灰色数据处理的残差进行分析,将分析的结果与TLS的拟合和推估结果进行组合以减弱残差的影响,从而对整体的残差起到了抑制的作用。经过实例分析,验证了该方法的可行性。

关键词 TLS; 残差; 自适应回归; 拟合; 回归模型

中图分类号: P207

文献标识码: A

RESIDUALS ADAPTIVE REGRESSION TLS

Chen Xijiang^{1,2)}, Hua Xianghong^{1,2)}, Lu Tieding³⁾, Zhai Gaopeng⁴⁾ and Yang Yanjing³⁾

(1) *Institute of Surveying and Mapping, Wuhan University, Wuhan 430079*
(2) *Hazard monitoring & prevention Research Center, Wuhan University, Wuhan 430079*
(3) *Institute of Surveying and Mapping, East China Institute of Technology, Fuzhou 344000*
(4) *The Station of Quality Supervision and Inspection about Surveying and Mapping Products, Shijiazhuang 050000*

Abstract The TLS fitting and estimation residuals will be upward with the increasing in the amount of observed data in the process of TLS fitting and estimation. In order to curb the trend, the adaptive regression is introduced to the TLS fitting and estimation. The appropriate regression model is used to analyze the residuals handled by gray method. The results of residual analysis is combined to the results of TLS fitting and estimation to weaken the impact of residuals. The experimental results confirm the feasibility of the method.

Key words: TLS; residual; adaptive regression; fitting; regression model

1 前言

总体最小二乘是在最小二乘基础上发展起来的,同时考虑自变量和因变量误差的模型,众多学者对其进行了深入的研究,文献[1]推导了正交TLS模

型, Schaffrin^[2,3]等人研究了带权情况下的WTLS, Van Huffel和Vandewalle^[4]给出了混合LS-TLS模型,该模型对于整体误差起到了平滑作用。但并没有起到抑制残差的作用,在利用拟合确定参数后的TLS模型进行推估的时候仍将带有误差的原始观测

* 收稿日期:2011-12-26

基金项目:国家自然科学基金(41174010,41074025);精密工程与工业测量国家测绘地理信息局重点实验室开放基金(PF2011-17)

作者简介:陈西江,男,1985年生,博士研究生,主要从事三维激光点云数据处理研究工作。E-mail:cxj_0421@163.com

数据 x, y 作为自变量, 这样推估的结果不仅具有本身的误差 v_z , 还将自变量的误差 v_x, v_y 转嫁到所需推估的变量上, 其推估的精度就会降低, 从而造成了推估残差的凸凹现象, 为了平滑推估残差的凸凹及抑制残差上扬的趋势, 本文将自适应回归模型引入到 TLS 模型中, 对 TLS 的初始拟合残差进行分析, 确定适宜的残差回归模型来对残差进行推估, 从而可以抵消 TLS 推估残差的凸凹及上扬, 提高了 TLS 拟合及推估的精度。

2 TLS 模型

总体最小二乘的基本数学模型为

$$y + V_y = (x + V_x)\beta \tag{1}$$

文献[5]提出了同时顾及 x 和 y 相互影响的 TLS 模型:

$$A(x_i + V_{xi}) + B(y_i + V_{yi}) = 1 \tag{2}$$

本文考虑到对 TLS 解算结果的稳定性, 采用文献[1]提出的正交 TLS 模型:

$$a(x - \bar{x}) + b(y - \bar{y}) = 0 \tag{3}$$

式中 \bar{x}, \bar{y} 为 n 个数据点的中心:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \tag{4}$$

将式(3)一元正交 TLS 模型延伸到多元, 则:

$$a(x - \bar{x}) + b(y - \bar{y}) + c(z - \bar{z}) = 0 \tag{5}$$

对式(5), 采用文献[1]的解算方法, 可以解得参数 a, b, c 的值, 同时对 $a(x - \bar{x}) + b(y - \bar{y}) + c(z - \bar{z}) = 0$ 模型进行变形得:

$$z = -\frac{a}{c}(x - \bar{x}) - \frac{b}{c}(y - \bar{y}) + \bar{z} \tag{6}$$

利用形成的模型公式(6)对观测数据进行拟合, 便可得到因变量的拟合值 $[\bar{z}_1 \ \bar{z}_2 \ \dots \ \bar{z}_n]$ 。

3 TLS 残差自适应回归模型的构建

3.1 基于 TLS 残差的灰色处理

残差自适应回归 TLS 的基础是对 TLS 模型的残差进行分析, 根据原始观测数据, 利用正交 TLS 的性质, 可以解得参数 a, b, c 的值, 根据公式(6)得到因变量的拟合值 $[\bar{z}_1 \ \bar{z}_2 \ \dots \ \bar{z}_n]$, 并利用公式(7)将拟合值与真值 $[z_1 \ z_2 \ \dots \ z_n]$ 进行残差处理得到残差 v_i :

$$\begin{cases} v_1 = z_1 - \bar{z}_1 \\ v_2 = z_2 - \bar{z}_2 \\ \dots\dots\dots \\ v_n = z_n - \bar{z}_n \end{cases} \tag{7}$$

由于拟合残差较小或者可能存在负值, 利用公式(8)适当加一常数 b 把 v_i 序列变为非负 V_i , 即

$$V_i = v_i + b \tag{8}$$

根据灰色理论的数据预处理, 为了较为准确地对残差序列 V_i 进行趋势分析, 需要对 V_i 做一次累加以凸显其规律性, 得出生成序列 V_i^1 , 即:

$$V_i^1 = [V_1^1 \ V_2^1 \ \dots \ V_n^1] \tag{9}$$

式中

$$V_i^1 = \sum_{j=1}^i V_j \tag{10}$$

对序列 V_i^1 的时序散点图变化趋势进行分析, 为了较为全面、准确地对 V_i^1 进行推估, 需要采用几种常用曲线回归方程对 V_i^1 进行拟合对比分析, 本文采用常用的多项式曲线拟合方程式(11)、冥函数曲线回归方程式(12)、指数曲线回归方程式(13):

$$f(i; a_0, a_1, a_2, \dots, a_n) = a_0 + a_1 i + a_2 i^2 + \dots + a_n i^n \tag{11}$$

$$f(i, a, b, c) = ai^b + c \tag{12}$$

$$f(i, a, b) = ae^{bi} \tag{13}$$

其中 i 为时间序列, a_n, a, b, c 为待定系数。

3.2 自适应回归模型的构建

根据处理之后的残差数据, 利用最小二乘对式(11)、(12)、(13)中的未知参数 a_n, a, b, c 进行解算, 最终可确定多项式、冥函数、指数曲线方程, 根据前 n 序列的已有拟合残差信息, 利用曲线回归方程对残差信息 V_i^1 进行拟合, 并进行拟合结果误差的解算, 从而得到多项式曲线拟合中误差、冥函数曲线拟合中误差、指数曲线拟合中误差, 其中, 对于多项式曲线拟合回归方程, 如果项数太多, 就会陷入过拟合现象, 因此本文选用常用的二项式、三项式、四项式曲线回归方程。根据拟合结果图, 剔除偏离残差离散点最大的曲线回归方程, 对剩余的曲线拟合中误差进行对比分析, 为了同时能够顾及到拟合精度及过度拟合的情况下, 将拟合精度最低和拟合精度最高的曲线回归方程排除, 再对剩余的曲线拟合中误差进行对比分析, 最终将拟合精度较高的回归方程 $f(i)$ 作为 V_i^1 的推估方程, 利用 $f(i)$ 计算 $n+i$ 时刻 V_i^1 的推估值 \hat{V}_i^1 , 对 \hat{V}_i^1 进行累减得出残差 V_i 的推估值 \hat{V}_i , 再将所加入的常数 b 消去, 便得到 TLS 拟合残差 v_i 的推估值 \hat{v}_i 。最终利用式(6)对推估样本进行推估, 得到初始推估值 \hat{z}_i , 将初始推估值与残差推估值 \hat{v}_i 进行结合, 形成最终的 TLS 推估值 z_i :

$$z_i = \hat{z}_i + \hat{v}_i \tag{14}$$

4 实例分析

以文献[6]中例5.4的数据作为样本(表1), 将温度、压力作为自变量, 变形值作为因变量, 将20组观测数据分为两部分, 前15组数据作为拟合样本,

后 5 组数据作为推估样本,利用公式(5)对拟合样本进行拟合得到总体最小二乘的参数为 $a = 0.3971, b = -0.055, c = 0.9161$,根据所求的参数利用公式(6)对变形值进行拟合,并将拟合结果与原始数据相减得到拟合残差。

表 1 观测数据
Tab.1 Observations

	温度 (°C)	压力 (Pa)	变形值 (mm)		温度 (°C)	压力 (Pa)	变形值 (mm)
1	11.2	36	-5	12	10.1	31	-9.3
2	10	40	-6.8	13	11.6	29	-9.3
3	8.5	35	-4	14	12.6	58	-5.1
...
9	4.7	24	-5.4	18	21.6	44	-9.3
10	11.7	65	-7.7	19	23.1	56	-9.5
11	9.4	44	-8.1	20	21.9	51	-9.9

通过公式(8)、(9)对拟合残差进行处理,利用几种常用的曲线回归方程对处理结果的 V_i^1 进行拟合(图 1)。

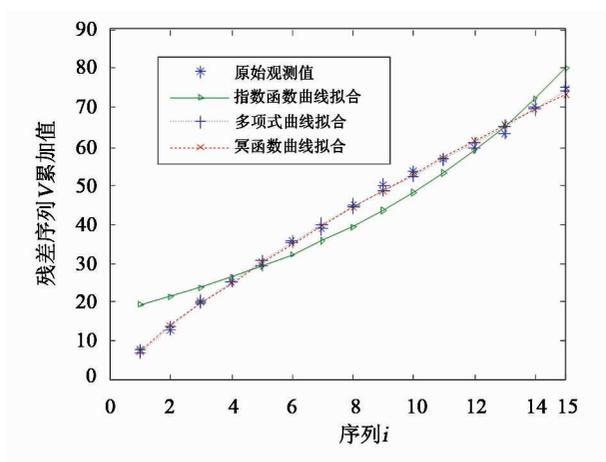


图 1 曲线拟合
Fig.1 Curve fitting

通过图 1 可直观判定能够较为准确拟合 V_i^1 的曲线方程为多项式或冥函数回归方程,根据最小二乘解算原则,确定这两种曲线回归方程为:

$$V_i^1 = 2.314 + 5.832i - 0.074i^2 \quad (15)$$

$$V_i^1 = -0.5 + 7.66i - 0.35i^2 + 0.0115i^3 \quad (16)$$

$$V_i^1 = 2.396 + 4.84i + 0.3845i^2 - 0.058i^3 + 0.0022i^4 \quad (17)$$

$$V_i^1 = 9.048i^{0.7825} - 1.86 \quad (18)$$

利用式(15)~(18)拟合 V_i^1 的误差如表 2 所示。

从表 2 可知二项式曲线回归方程拟合精度最低、四项式曲线回归方程拟合精度最高,在综合考虑拟合精度及过拟合现象的情况下,将二项式、四项式

曲线拟合方程去除,对于三项式及冥函数曲线拟合方程,确定精度较高的三项式曲线拟合方程为 V_i^1 的回归模型。利用式(16)与(6)组成的残差自适应回归 TLS 对原始观测数据进行模拟及推估,并将模拟,推估结果与最小二乘、TLS 进行对比分析,结果如表 3、4 所示。从表 3 可知,TLS、最小二乘模拟精度相当,而残差自适应回归 TLS 模拟精度要优于其他两种方法;从表 4 可明显看出残差自适应回归 TLS 推估精度较优于其他两种方法,同时从推估残差可看出,其残差分布较为平滑,不会出现太大的凸凹,从而抑制了残差的上扬趋势,增强了模型的稳定度,提高了拟合精度。

表 2 曲线拟合误差对比分析
Tab.2 Comparison of curve fitting error

	二项式拟合	三项式拟合	四项式拟合	冥函数拟合
SSE	20.08	12.69	9.1	16.62
RMSE	1.3	1.074	0.95	1.177

表 3 三种模型的模拟值对比分析

Tab.3 Comparison among the simulation values with three models

	最小二乘	TLS	残差自适应回归 TLS
拟	-2.01	1.985	0
合	0.141	-0.162	-1.694
残	-2.55	2.466	1.31
差
e	0.682	-0.658	-0.44
	0.762	-0.529	2.237
	-1.01	1.185	0.0025
$(\sum e ^2)/n$	1.96	1.89	1.48
$(\sum e/z_i ^2)/n$	19.46%	19.08%	16.8%

表 4 三种模型的推估值对比分析

Tab.4 Comparison among the estimation values with three models

	最小二乘	TLS	残差自适应回归 TLS
预	0.66867	0.4688	2.416
测	0.99063	0.8144	3.67
残	-3.0221	-3.2046	1.31
差	7.6845	7.4408	0.492
e	1.4442	1.2701	-0.742
$(\sum e ^2)/n$	14.34	13.63	4.36
$(\sum e/z_i ^2)/n$	26.8%	26%	19.9%

5 结论

本文对拟合后的数据进行了分析,将几种常用

的回归模型引入到拟合残差中,对不同回归模型的拟合精度进行了评价分析,选取最佳的回归模型作为拟合残差的推估模型并对拟合残差进行模拟及推估,实现了残差的推估,同时随着序列号的增加,其模型推估的残差也呈现出上扬趋势,因此对残差进行分析并将残差推估结果与观测数据的 TLS 推估结果进行结合,实现了残差的平滑处理,提高了推估精度。

由于对残差序列的分析采用的是几种常用的回归模型,而这几种回归模型是否适用于所有变形体拟合后的残差序列,对多项式曲线拟合过程中,次数越高,模拟精度越高,但推估的精度可能会降低,怎样平衡残差的拟合精度与推估精度及怎样选用最佳回归方程将有待进一步探讨。

参 考 文 献

1 张贤达. 矩阵分析与应用[M],北京:清华大学出版社,2004. (Zhang Xianda. Matrix analysis and applications [M]. Beijing: Tsinghua University Press, 2004)

2 Schaffrin B and Felus Y A. On total least squares adjustment with constraints [J]. A Windows on the Future of Geodesy, 2005, 128(4): 417 - 421.

3 Felus Y and Schaffrin B. Performing similarity transformations using the Errors-In-Variables Model [R]. American Society for Photogrammetry and Remote Sensing (ASPRS) Annual Meeting, 2005.

4 Huffel S Van and Vandewalle J. The least square problem: Computational aspects and analysis, Frontiers in Appl [R]. Math. SIAM, Philadelphia, 1991.

5 鲁铁定,陶本藻,周世健. 基于总体最小二乘法的线性回归建模和解法[J],武汉大学学报(信息科学版),2008, 33(5): 504 - 507. (Lu Tieding, Tao Benzao and Zhou Shijian. Modeling and algorithm of linear regression based on total Least squares [J]. Geomatics and Information Science of Wuhan University, 2008, 33(5): 504 - 507)

6 邱卫宁,等. 测量数据处理理论与方法[M]. 武汉:武汉大学出版社,2008. (Qiu Weining, et al. The theory and method of surveying data processing [M]. Wuhan: Wuhan University Press, 2008)

(上接第 89 页)

图 5(b)与(a)等高线的分布较为接近,显现出了较为细节的分布信息,保持了较高的一致性;(c)与(a)的一致性较差。从而显示图 5(b)的建模方法是稳定可靠的。

5 结 论

1)进一步验证了多面函数法比较适用于局域海洋地磁场建模,在地磁变化复杂的海域,先进行趋势面拟合,再对剩余的误差进行多面函数拟合,可以突出局部变化,拟合效果要优于直接使用多面函数进行拟合。

2)趋势面拟合主要反映整体变化趋势,当局部磁力异常观测值变化显著时,采用抗差趋势面法可以减小对趋势面拟合的不良影响。在多面函数拟合前,采用抗差估计理论进行质量控制,可以更有效地剔除异常观测值对趋势面拟合的影响。

3)使用正交最小二乘多面函数法得到的拟合精度优于多面函数法,可以按照规则自适应地选择特征点,提高了人工选择特征点的效率。对系数矩阵进行正交化,不会影响最终的结果,因为正交化后并没有参与多面函数系数的求解,仅是用于节点的选择,因此可以作为一种自动选择节点的参考方法。

参 考 文 献

1 赵建虎,等. 基于多面函数的局域地磁场建模方法研究

[J]. 海洋通报,2009, 28(4): 89 - 97. (Zhao Jianhu, et al. Study on construction of local marine geomagnetic field model based on multi - surface function [J]. Marine Science Bulletin, 2009, 28(4): 89 - 97)

2 刘兆平,杨进,武炜. 地球物理数据网格化方法的选取 [J]. 物探与化探,2010, 34(1): 93 - 98. (Liu Zhaoping, Yang Jin and Wu Wei. The choice of gridding methods for geophysical data [J]. Geophysical & Geochemical Exploration, 2010, 34(1): 93 - 98)

3 张菊清,刘平芝. 抗差趋势面与正交多面函数结合拟合 DEM 数据 [J]. 测绘学报,2008, 37(4): 526 - 531. (Zhang Juqing and Liu Pingzhi. Combining fitting based on robust trend surface and orthogonal multiquadrics with application in DEM fitting [J]. Acta Geodaetica et Cartographica Sinica, 2008, 37(4): 526 - 531)

4 杨元喜. 抗差估计理论及其应用[M]. 北京:八一出版社,1993. (Yang Yuanxi. The theory and application of robust estimation [M]. Beijing: Bayi Publishing House, 1993)

5 杨元喜. 自适应抗差最小二乘估计 [J]. 测绘学报,1996, 25(3): 206 - 211. (Yang Yuanxi. A daptively robust least squares estimation [J]. Acta Geodaetica et Cartographica Sinica, 1996, 25(3): 206 - 211)

6 黄学功,等. 地磁图制备方法及其有效性评估 [J]. 北京航空航天大学学报,2009, 35(7): 891 - 895. (Huang Xuegong, et al. Geomagnetic mapping and validity estimation [J]. Journal of Beijing University of Aeronautics And Stronautics, 2009, 35(7): 891 - 895)