

基于总体最小二乘的直线拟合方法探究

鞠英博¹, 李伟¹, 姚磊²

(1. 黑龙江第二测绘工程院 黑龙江 哈尔滨 150025; 2. 上海市测绘院 上海 200129)

摘要: 在直线拟合问题中, 经典的最小二乘拟合方法在自变量选取不同时, 拟合的参数值和中误差存在较大差别, 故本文利用模拟数据对经典最小二乘和总体最小二乘拟合结果进行对比分析, 得出结论认为: 经典最小二乘自变量选取不同结算参数的原因是在进行拟合计算时忽略了自变量的误差, 使拟合结果只能在一个方向上保持最佳; 利用总体最小二乘参数拟合的方法进行直线拟合时拟合结果不受自变量变化的影响, 并能够提高拟合精度。

关键词: 最小二乘法; 总体最小二乘; 直线拟合; EVI 模型; SVD 分解

中图分类号: P231.5 **文献标识码:** A **文章编号:** 1672-5867(2017)06-0166-03

Linear Fitting Method Based on Total Least Squares

JU Ying-bo¹, LI Wei¹, YAO Lei²

(1. The Second Surveying and Mapping Institute of Heilongjiang Province, Harbin 150025, China;
2. Surveying and Mapping Institute of Shanghai, Shanghai 200129, China)

Abstract: In the linear fitting problem, because the classical least squares fitting method is not the same as the independent variables, the fitting parameters and the errors have great difference, in this paper, the simulation results are used to compare the classical least squares and total least squares fitting results. The reason for the different parameters of the classical least squares independent variables is that the errors in the independent variables are ignored in the fitting calculation, the fitting result can only be kept in one direction; By using the method of least square parameter fitting, the fitting result is not affected by the independent variable and improve the fitting precision.

Key words: least square method; total least squares; straight line fitting; EVI model; SVD decomposition

0 引言

直线拟合在多种数据分析手段和各种实践中有着广泛的应用,但在经典的最小二乘理论模型下,往往只考虑因变量的误差,最终的拟合结果只在一个方向上最佳^[1-2]。并且当自变量选取不同时,参数拟合的结果往往也不相同,所以不能得到拟合参数的最优估值,与实际工程的应用相脱离。这是因为在数据获取的过程中,无论是观测向量还是系数矩阵都可能存在误差,但传统的最小二乘忽略了系数矩阵中误差的影响,造成了由于因变量选取不同而使拟合结果也不同的不良后果^[3-4]。

针对如何解决同时考虑自变量、因变量都存在误差的数据处理问题,1901年 Pearson 提出了总体最小二乘的思想,讨论了线性方程的近似计算方法^[5]。随着计算数

学的发展,1980年 Golub 和 Van Loan 从数值分析的观点对这种方法进行了全面的分析,并首次命名为总体最小二乘^[6-8]。随后 1991、1996、2001、2006 年在比利时 Leuven 举行了关于 TLS 基本理论、EIV 模型、TLS 曲线拟合的国际会议。Van Huffel 和 Vandenwalle 出版了总体最小二乘专著,提出了 TLS 与 LS 应用范围相同并具有更高精度的观点^[9]。

本文在分析比较现有研究成果的基础上,拟采用 EIV 模型进行最小二乘直线拟合,基于 SVD 分解的方法,进行参数的求取,并通过模拟数据对算法进行对比验证。

1 最小二乘的直线拟合方法

当自变量为 x , 因变量为 y 时,方程的形式为

$$y = ax + b \quad (1)$$

收稿日期: 2017-03-24

作者简介: 鞠英博(1989-),男,黑龙江哈尔滨人,助理工程师,本科学历,主要从事国家基础测绘任务、地理国情普查和地籍调查等工作。

x, y 为测点坐标 a, b 为待估参数。

则误差方程为^[10]

$$v_y = [1 \ x] \begin{bmatrix} a \\ b \end{bmatrix} - y \quad (2)$$

其中 $v_y = [v_{y1} \ v_{y2} \ v_{y3} \ \cdots \ v_{ym}]^T$, $x = [x_1 \ x_2 \ x_3 \ \cdots \ x_m]^T$, $y = [y_1 \ y_2 \ y_3 \ \cdots \ y_m]^T$

其矩阵形式为

$$V = AX - L \quad (3)$$

其中 V 为 v_y , A 为 $[1 \ X]$, X 为 $[a \ b]^T$, L 为 y ,

最小二乘约束条件为

$$V^T V = \min \quad (4)$$

$$\text{即} \sum_{i=1}^m (ax_i + b - y_i)^2 = \min \quad (5)$$

在此约束条件下, 其最小二乘解为

$$X = (A^T A)^{-1} A^T L \quad (6)$$

当 y 为自变量, x 为因变量时, 拟合方程为

$$x = ky + m \quad (7)$$

式中 k, m 为拟合参数。

其解算方法同上。

利用加入了精度相同的模拟数据(数据分布如图1所示), 对以上两种拟合方案分别解算(结果见表1)。

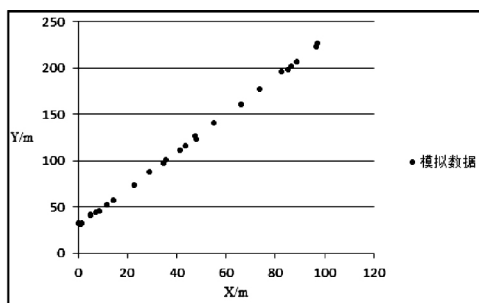


图1 模拟数据分布情况图

Fig. 1 Analog data distribution map

表1 自变量取值不同时最小二乘拟合参数对比

Tab. 1 Comparison of least squares fitting parameters with different independent variables

参数及中误差	自变量为 x	自变量为 y
a	2.001 4	2.002 1
b	29.768 0	29.736 8
单位权中误差 (cm)	12.175	9.082

由表1可知: 在经典最小二乘解算方法下, 选取不同的自变量, 其参数拟合的结果和中误差都有着较大的差别。这是因为在实际观测中, x, y 中都包含误差, 但经典的最小二乘拟合只考虑其中一个观测量的误差, 从而影响拟合结果, 导致其拟合精度不一致^[11]。

2 总体最小二乘的直线拟合方法

由前文可知, 对于直线拟合问题, 当系数矩阵存在误差时, 经典最小二乘的估值不再是最优的, 此时应当采用总体最小二乘的方法进行参数的求取。当今总体最小二

乘的研究中, SVD 分解法具有计算简便、易于编程实现的特点, 有着广泛的应用^[12]。

2.1 总体最小二乘直线拟合问题描述

在总体最小二乘的平差思想中, 不但观测量 L 存在误差 ΔL , 系数矩阵 A 也存在误差 ΔA , 在直线拟合问题中可描述为:

$$y + v_y = a(x + v_x) + b \quad (8)$$

v_y, v_x 即为总体最小二乘分别在 y, x 方向上的改正数。

则当 $v_y^T v_y + v_x^T v_x = \min$ 时, 参数取得最优估值。

对公式(8)按照 EIV 模型可以描述为

$$(A + E_A)X = L + E_L \quad (9)$$

式中 E_A 和 E_L 就是系数阵和观测向量的误差。

2.2 总体最小二乘直线拟合的 SVD 法

将式(9)改写为

$$[A + E_A \ L + E_L] \begin{bmatrix} X \\ -1 \end{bmatrix} = 0 \quad (10)$$

通过将增广矩阵 $[A \ L]$ 进行奇异值分解(SVD), 对直线参数进行总体最小二乘求解。

首先对增广矩阵进行奇异值分解

$$C = [A \ L] = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T = U_1 \cdot \Sigma \cdot V^T \quad (11)$$

其中

$$U_1 = [U_{11} \ U_{12}] \quad (12)$$

$$\Sigma \cdot V^T = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}^T \begin{matrix} m \\ 1 \end{matrix} \quad (13)$$

在公式(12)中, $U = [U_1 \ U_2] \in R^{n \times n}$ 为矩阵 $[A \ L] \cdot [A \ L]^T = AA^T + LL^T$ 的 n 个特征向量组成的正交矩阵, $V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}^T \begin{matrix} m \\ 1 \end{matrix} = [v_1 \ v_2 \ v_3 \ \cdots \ v_{m+1}] \in R^{(m+1) \times (m+1)}$ 为矩

阵 $[A \ L]^T \cdot [A \ L] = \begin{bmatrix} A^T A & A^T L \\ L^T A & L^T L \end{bmatrix}$ 的 $m+1$ 个特征向量

组成的正交矩阵, $\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{matrix} m \\ 1 \end{matrix} = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_{m+1})$ 为矩阵 C 的奇异值, 其奇异值按顺序 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{m+1}$ 排列。

当矩阵 $[A \ L]$ 最佳逼近 $[\hat{A} \ \hat{L}]$ 时必然满足

$$[\hat{A} \ \hat{L}] = U \Sigma V^T = U_{11} \cdot \Sigma_1 \cdot [V_{11}^T \ V_{21}^T], \Sigma = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n) \quad (14)$$

所以总体最小二乘的拟合值残差矩阵为

$$\hat{E} = E_{\min} = [\hat{E}_A \ \hat{E}_L] = U_{12} \cdot \Sigma_2 \cdot [V_{12}^T \ V_{22}^T] \quad (15)$$

且其改正量满足

$$\sigma_{m+1} = \min_{\text{rank}([A \ L])=n} \| [A; L] - [\hat{A}; \hat{L}] \|_F \quad (16)$$

未知参数 X 的估值为

$$\hat{X} = -V_{12} \cdot V_{22}^{-1} = \frac{-1}{V_{m+1, m+1}} V_{12} \quad (17)$$

进一步可计算其单位权方差为

$$\hat{v} = \hat{E}^T \cdot \hat{E} \cdot (\text{vec} \hat{E}_A)^T \cdot \text{vec} \hat{E}_A = \text{tr}(\hat{E}^T \cdot \hat{E} + \hat{E}_A^T \hat{E}_A) = \text{tr}(\hat{E} \cdot \hat{E}^T) = \sigma_{m+1}^2 \quad (18)$$

$$\sigma_0^2(TLS) = \frac{\hat{v}}{n-m} \quad (19)$$

3 不同参数拟合方法的对比分析

上文分别介绍了经典最小二乘和基于 EVI 模型 SVD 分解的总体最小二乘的拟合方法,为了对其拟合效果进行分析,本文利用模拟数据分别采用经典最小二乘和总体最小二乘法进行拟合求解。在拟合过程中分别采用 $y = ax + b$ 和 $x = ky + m$ 两种模型进行对比计算,分别求出其参数估值和单位权中误差,对比各个计算方法和模型在拟合结果中的一致性与区别(见表 2)。

从计算结果可以看出,由于计算时没有考虑因变量的误差,选取不同的因变量时,运用经典最小二乘法进行参数拟合时,其最终参数的拟合值不同,拟合的精度相差也较大。利用总体最小二乘进行直线拟合时,由于综合考虑了自变量和因变量的误差,对方程系数阵的误差也进行了估计,在变换自变量后参数的估值没有发生改变,拟合的精度也得到了一定程度的提高。

表 2 经典最小二乘与总体最小二乘参数拟合结果对比

Tab. 2 Comparison of classical least squares and total least squares parameter fitting

参数及 中误差	经典最小二乘		总体最小二乘	
	自变量为 x	自变量为 y	自变量为 x	自变量为 y
a	2.001 4	2.002 1	1.999 4	1.999 4
b	29.768 0	29.736 8	29.908 0	29.908 0
单位权中误差(cm)	12.175	9.082	8.69	8.69

4 结束语

1) 对经典最小二乘和总体最小二乘的直线拟合方法进行了对比分析,找出了经典最小二乘自变量选取不同、结算参数不同的原因:忽略了自变量的误差,使拟合结果

只能在一个方向上保持最佳。

2) 利用模拟数据对经典最小二乘和总体最小二乘的直线拟合方法进行了对比分析,认为运用总体最小二乘法进行直线拟合时拟合结果不受自变量变化的影响,提高了拟合精度,获得了较好的拟合效果。

参考文献:

- [1] 李雄军. 对 X 和 Y 方向最小二乘线性回归的讨论[J]. 计量技术, 2005(1): 50-51.
- [2] 梁家惠. 用最小二乘法进行直线拟合的讨论[J]. 工科大学物理, 1995(3): 11-15.
- [3] 丁克良, 沈云中, 欧吉坤. 整体最小二乘法直线拟合[J]. 辽宁工程技术大学学报: 自然科学版, 2010, 29(1): 44-47.
- [4] 官云兰, 周世健, 张立亭, 等. 稳健整体最小二乘直线拟合[J]. 工程勘察, 2012(2): 60-62.
- [5] Pearson K. On lines and planes of closest fit to points in space[J]. Philos. Mag., 1901(2): 559-572.
- [6] Golub G H, Hansen P C, O'Leary D P. Tikhonov regularization and total least squares[J]. SIAM J. Matrix Anal., 1999(21): 185-194.
- [7] Golub G H, Lan Loan F C. An Analysis of the Total Least Squares Problem[J]. SIAM Journal on Numerical Analysis, 1980, 17(6): 883-893.
- [8] Golub G H, Van Loan C F. Matrix Computations[M]. Third Edition. Baltimore: The Johns Hopkins University Press, 1996.
- [9] Van Huffel S, editor. Recent advances in total error - in - variables modeling[M]. Philadelphia: SIAM, 1997.
- [10] 武汉大学测量平差学科组编. 误差理论与测量平差基础[M]. 武汉: 武汉大学出版社, 2003.
- [11] 鲁铁定. 总体最小二乘平差理论及其在测绘数据处理中的应用[D]. 武汉: 武汉大学, 2010.
- [12] 林国记. 二次曲面拟合 GPS 高程在台山水库灌区的应用[J]. 北京测绘, 2014(2): 44-46.

[编辑: 刘莉鑫]

(上接第 165 页)

- [2] HERRING T A, KING R W, MCCLUSKYS C. GAMIT reference manual: GPS analysis at MIT, release [M]. Cambridge: Massachusetts Institute of Technology, 2006.
- [3] HERRING T A. GLOBK: global Kalman filter VLBI and GPS analysis program, version 10.0 [M]. Cambridge: Massachusetts Institute of Technology, 2002.
- [4] 胡敏章, 李建成. 顾及地壳密度模型的 Airy - Heiskanen 均衡重力异常的计算[J]. 大地测量与地球动力学, 2010, 30(5): 48-52.
- [5] 郭春喜, 王惠民, 王斌. 全国高分辨率格网地形和均衡改正的确定[J]. 测绘学报, 2002, 31(3): 201-205.
- [6] 李建成, 陈俊勇, 宁津生, 等. 地球重力场逼近理论与中国 2000 似大地水准面的确定[M]. 武汉: 武汉大学出版社, 2003.

- [7] 许厚泽. 我国精化大地水准面工作中若干问题的讨论[J]. 地理空间信息, 2006, 4(5): 1-3.
- [8] MARTINEC Z, MATYSKA C, GRAFAREND E W, et al. On Helmert's 2nd condensation method[J]. Manuscripta Geodetica, 1993(18): 417.
- [9] HAINES G V. Spherical Cap Harmonic Analysis[J]. J Geophys Research, 1985, 90(B3): 2583-2591.
- [10] Li Jiancheng. Spherical Cap Harmonic Expansion for Local Gravity Representation [J]. Manuscripta Geodetica, 1995(20): 265-277.

[编辑: 刘莉鑫]