

整体最小二乘求解线性模型的改进算法

邱卫宁^{1,2} 齐公玉^{1,2} 田丰瑞^{1,2}

(1 武汉大学测绘学院,武汉市珞喻路129,430079)

(2 武汉大学灾害监测与防治研究中心,武汉市珞喻路129号,430079)

摘要:针对观测方程中观测向量和数据矩阵均有误差的情况,提出了整体最小二乘的改进算法,利用附有限制条件的平差模型,导出了观测向量和数据矩阵精度不等情况下的计算公式。该算法满足拟合方程应有的条件,提高了整体最小二乘递推算法的逼近精度,为整体最小二乘应用于测量数据处理提供了可行的方法。

关键词:数据矩阵;附有限制条件的平差模型;整体最小二乘递推算法

中图法分类号:P207.2

在测量平差线性模型参数求解中,通常认为数据矩阵或结构矩阵是给定的,不存在误差。当观测值存在偶然误差时,可用最小二乘法求得参数的最优解。但是,如果数据矩阵也存在误差或扰动,那么最小二乘估计从统计观点看就不再是最优的,而是有偏的,而且偏差的协方差将由于 X 的噪声误差的作用而增加。Golub等人于1980年提出了解决这类问题的整体最小二乘法(TLS)^[1],并导出了求整体最小二乘参数理论解的奇异值分解法SVD。SVD方法的复杂性限制了TLS的广泛应用,Schaffrin等人提出了整体最小二乘法的递推算法^[2,3]。本文在此基础上,从测量平差线性模型出发,通过附加限制条件,提高整体最小二乘递推算法的逼近精度,所推导出的改进算法,解算方便,便于应用。

设有观测方程:

$$\mathbf{L} = \mathbf{B} \mathbf{X} + \mathbf{d} \quad (1)$$

当观测向量 \mathbf{L} 中存在随机误差 \mathbf{V} ,同时系数矩阵或结构矩阵 \mathbf{B} 也含有随机误差 \mathbf{E}_B ,式(1)为:

$$\mathbf{V} - \mathbf{E}_B \hat{\mathbf{X}} - \mathbf{B} \hat{\mathbf{X}} + \mathbf{L} = 0 \quad (2)$$

式中,假设 $\mathbf{d} = [0 \ 0 \ \cdots \ 0]^T$ 。

设观测值的权阵为 \mathbf{P} ,结构矩阵 \mathbf{B} 的误差阵列向量化后 $\text{vec}(\mathbf{E}_B)$ 的权阵为 \mathbf{P}_B ,单位权方差为 σ_0^2 ,则:

$$\mathbf{V} \sim (0, \sigma_0^2 \mathbf{P}^{-1}), \text{vec}(\mathbf{E}_B) = \mathbf{V}_B \sim (0, \sigma_0^2 \mathbf{P}_B^{-1}) \quad (3)$$

整体最小二乘估计准则为:

$$\mathbf{V}^T \mathbf{P} \mathbf{V} + \mathbf{V}_B^T \mathbf{P}_B \mathbf{V}_B = \min \quad (4)$$

目前,最严密有效的整体最小二乘法是SVD。为了达到拟合目标式(4),SVD从数据中重心点出发推导其公式。根据文献[4]的描述,对于整体最小二乘拟合的直线,必须通过 n 个数据点中心 (\bar{x}, \bar{y}) ,才能使偏差 D 最小。在高维情况下,令 n 个数据向量: $\mathbf{x}_i = [x_{1i} \ x_{2i} \ \cdots \ x_{mi}]^T$ ($i = 1, \dots, n$) 分别为 m 维数据,并且 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = [\bar{x}_1 \ \bar{x}_2 \ \cdots \ \bar{x}_m]^T$ 为均值向量。当有 m 维法向量 \mathbf{r}, \mathbf{x} 需满足法方程:

$$\langle \mathbf{x} - \mathbf{x}, \mathbf{r} \rangle = 0 \quad (5)$$

由此法方程确定的超平面,可以使距离平方和等于最小^[4]。本文在目标函数式(4)下,推导出满足式(2)、(5)的整体最小二乘改进算法。

1 整体最小二乘改进算法

式(1)的系数阵的 t 个数据向量 $\mathbf{B} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_t]$,其中 $\mathbf{b}_i = [b_{1i} \ b_{2i} \ \cdots \ b_{ni}]^T$,参数 \mathbf{X} 应满足式(5),按照式(1)中的平差符号,可表示为:

$$\mathbf{C} \hat{\mathbf{X}} - \mathbf{w} = 0 \quad (6)$$

式中, $\mathbf{w} = \frac{1}{n} \sum_{i=1}^n \mathbf{L}_i$; $\mathbf{C} = [\bar{b}_1 \ \bar{b}_2 \ \cdots \ \bar{b}_t]$, $\bar{b}_i =$

$$\frac{1}{n} \sum_{j=1}^n b_{ij}.$$

将式(3)取迹。按 Lagrange 乘数法求解,构成目标函数为:

$$\Phi = \text{tr}(\mathbf{V}^T \mathbf{P} \mathbf{V}) + \text{tr}(\mathbf{E}_B^T \mathbf{P}_B \mathbf{E}_B) - 2\mathbf{K}^T (\mathbf{V} - \mathbf{E}_B \hat{\mathbf{X}} - \mathbf{B} \hat{\mathbf{X}} + \mathbf{L}) - 2\boldsymbol{\lambda}^T (\mathbf{C} \mathbf{X} - \mathbf{w}) \quad (7)$$

式中, \mathbf{K} 、 $\boldsymbol{\lambda}$ 分别是对应于式(2)和(6)的联系数向量。为求 Φ 的极小值, 将其分别对 \mathbf{V} 、 \mathbf{E}_B 和 $\hat{\mathbf{X}}$ 求一阶导, 并令其为零:

$$\frac{\partial \Phi}{\partial \mathbf{V}} = 2\mathbf{V}^T \mathbf{P} - 2\mathbf{K}^T = 0 \quad (8)$$

$$\frac{\partial \Phi}{\partial \mathbf{E}_B} = 2\mathbf{E}_B^T \mathbf{P}_B + 2\mathbf{K} \hat{\mathbf{X}}^T = 0 \quad (9)$$

$$\frac{\partial \Phi}{\partial \hat{\mathbf{X}}} = 2\mathbf{K}^T (\mathbf{E}_B + \mathbf{B}) - 2\boldsymbol{\lambda}^T \mathbf{C} = 0 \quad (10)$$

由式(8)、(9)得:

$$\mathbf{V} = \mathbf{P}^{-1} \mathbf{K} \quad (11)$$

$$\mathbf{E}_B = -\mathbf{K} \hat{\mathbf{X}}^T \mathbf{P}_B^{-1} \quad (12)$$

将式(11)、(12)代入式(2), 得:

$$\begin{aligned} \mathbf{P}^{-1} \mathbf{K} + \mathbf{K} \hat{\mathbf{X}}^T \mathbf{P}_B^{-1} \hat{\mathbf{X}} &= \mathbf{B} \hat{\mathbf{X}} - \mathbf{L} \\ \mathbf{K} &= (\mathbf{P}^{-1} + \mathbf{I} \hat{\mathbf{X}}^T \mathbf{P}_B^{-1} \hat{\mathbf{X}})^{-1} (\mathbf{B} \hat{\mathbf{X}} - \mathbf{L}) \end{aligned} \quad (13)$$

式中, \mathbf{I} 为单位阵。将式(13)代入式(10), 得:

$$(\mathbf{E}_B + \mathbf{B})^T (\mathbf{P}^{-1} + \mathbf{I} \hat{\mathbf{X}}^T \mathbf{P}_B^{-1} \hat{\mathbf{X}})^{-1} (\mathbf{B} \hat{\mathbf{X}} - \mathbf{L}) = \mathbf{C}^T \boldsymbol{\lambda} \quad (14)$$

令

$$\bar{\mathbf{P}} = (\mathbf{P}^{-1} + \mathbf{I} \hat{\mathbf{X}}^T \mathbf{P}_B^{-1} \hat{\mathbf{X}})^{-1}$$

由式(14)得:

$$(\mathbf{E}_B + \mathbf{B})^T \bar{\mathbf{P}} \mathbf{B} \hat{\mathbf{X}} - (\mathbf{E}_B + \mathbf{B})^T \bar{\mathbf{P}} \mathbf{L} = \mathbf{C}^T \boldsymbol{\lambda} \quad (15)$$

$$\hat{\mathbf{X}} = (\mathbf{B}^T \bar{\mathbf{P}} \mathbf{B} + \mathbf{E}_B^T \bar{\mathbf{P}} \mathbf{B})^{-1} [(\mathbf{B} + \mathbf{E}_B)^T \bar{\mathbf{P}} \mathbf{L} + \mathbf{C}^T \boldsymbol{\lambda}] \quad (16)$$

将式(16)代入式(6), 并令 $\mathbf{N} = (\mathbf{B}^T \bar{\mathbf{P}} \mathbf{B} + \mathbf{E}_B^T \bar{\mathbf{P}} \mathbf{B})$, 即可求得:

$$\boldsymbol{\lambda} = (\mathbf{C} \mathbf{N}^{-1} \mathbf{C}^T)^{-1} [\mathbf{w} - \mathbf{C} \mathbf{N}^{-1} (\mathbf{B} + \mathbf{E}_B)^T \bar{\mathbf{P}} \mathbf{L}] \quad (17)$$

通过求解式(11)~(13)、(15)和(16), 可得参数的整体最小二乘解。为递推求解, 式(16)可写为:

$$\hat{\mathbf{X}} = (\mathbf{B}^T \bar{\mathbf{P}} \mathbf{B})^{-1} [\mathbf{B}^T \bar{\mathbf{P}} \mathbf{L} - \mathbf{E}_B^T \bar{\mathbf{P}} (\mathbf{B} \hat{\mathbf{X}} - \mathbf{L}) + \mathbf{C}^T \boldsymbol{\lambda}] \quad (18)$$

如果观测值权阵 \mathbf{P} 和系数矩阵的权阵 \mathbf{P}_B 均为单位阵, 式(18)为:

$$\begin{aligned} \hat{\mathbf{X}} &= (\mathbf{B}^T \mathbf{B})^{-1} [\mathbf{B}^T \mathbf{L} - \mathbf{E}_B^T (\mathbf{B} \hat{\mathbf{X}} - \mathbf{L}) + \mathbf{C}^T \boldsymbol{\lambda} (1 + \mathbf{X}^T \mathbf{X})] \end{aligned}$$

或

$$\begin{aligned} \hat{\mathbf{X}} &= (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{L} - (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{E}_B^T (\mathbf{B} \hat{\mathbf{X}} - \mathbf{L}) + \\ &\quad (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{C}^T \boldsymbol{\lambda} (1 + \mathbf{X}^T \mathbf{X}) \end{aligned} \quad (19)$$

式(19)可看成 3 部分组成, 第一部分是仅考虑观

测误差的最小二乘解 $\hat{\mathbf{X}}_1 = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{L}$; 第二部分可看作是考虑了系数矩阵的改正项 $\hat{\mathbf{X}}_2 = -(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{E}_B^T (\mathbf{B} \hat{\mathbf{X}} - \mathbf{L})$; 第三项是考虑了条件式(5)的改正项:

$$\hat{\mathbf{X}}_3 = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{C}^T \boldsymbol{\lambda} (1 + \mathbf{X}^T \mathbf{X}) \quad (20)$$

式中, \mathbf{E}_B 可由式(9)、(10)得:

$$\mathbf{E}_B = - (1 + \hat{\mathbf{X}}^T \hat{\mathbf{X}})^{-1} (\mathbf{B} \hat{\mathbf{X}} - \mathbf{L}) \mathbf{X}^T \quad (21)$$

$\hat{\mathbf{X}}_2$ 采用迭代求解, $\hat{\mathbf{X}}$ 初始值可取 $\hat{\mathbf{X}}_1$ 代入, 当 $\max |\hat{\mathbf{X}}_2^{(i)} - \hat{\mathbf{X}}_2^{(i-1)}| < \epsilon$, 迭代结束。

将迭代结束后的 $\hat{\mathbf{X}}_2$ 加上 $\hat{\mathbf{X}}_1$ 得到 $\hat{\mathbf{X}}'$, 这就是整体最小二乘递推法(TLSD)的解。代入式(20)计算 $\hat{\mathbf{X}}_3$ 进行修正, 即得到整体最小二乘解。

2 算例分析

2.1 算例 1

文献[2]例 5-10 给出 3 个点(1, 2)、(2, 6)、(6, 1), 求这 3 个点所确定的 TLS 回归方程。设回归方程为:

$$y = \beta_0 + \beta_1 x$$

1) 首先用 SVD 求解, 得拟合直线为 $y = 6 - x$, 显然, 直线通过数据中心点(3, 3)。

2) 采用整体最小二乘递推算法, 即用式(19)中的 1、2 项计算, 经过 3 次迭代(迭代结束阈值 $\epsilon = 0.01$)得到直线方程 $y = 6.7362 - 0.9989x$, 将数据 x 中心点 $\bar{x} = 3$ 代入, $\bar{y} = 3.7395 \neq 3$ 。如果以 y 为自变量, 以 x 为因变量, 得到的直线为 $y = 6.7436 - 1.0011x$ 。

3) 采用本文的改进算法, 得拟合直线 $y = 5.9651 - 0.9884x$, 将数据 x 中心点 $\bar{x} = 3$ 代入, 得 $\bar{y} = 2.9999$ 。如果以 y 为自变量, 以 x 为因变量, 得到的直线为: $y = 6.0351 - 1.0117x$ 。

2.2 算例 2

在算例 1 的基础上再增加 6 对观测值, 观测值与 SVD、TLSD 及本文方法的比较见表 1 和图 1。

图 1 中短虚线为 SVD 拟合的直线, 长虚线为一般的 TLSD 拟合的直线, 实线为本文方法拟合的直线。从图 1 可以看出, 本文求得的直线几乎和 SVD 的拟合直线重合。

2.3 算例 3

设有一平面方程 $z = 2.0 + 1.5x - 1.2y$ 。在 x, y 没有误差时得到的 z 的数值上加上 0.05 的随机误差, 构成观测向量。然后, 在 x, y 上分别加上 0.3 的随机误差, 和 z 构成不同精度观测值。模拟观测值和用 3 种方法计算的 3 个参数 $\beta_0, \beta_1, \beta_2$ 及与真值相比所得到的参数中误差见表 2。

表1 观测对与3种方法的比较

Tab. 1 Observation Value, and Comparisons of 3 Methods

x	1	2	6	2	3	8	2	5	4	$\bar{x}=3.6667$
y	2	6	1	4	5	1	3	6	2	$\bar{y}=3.3333$
SVD法	$y=5.9761-0.7208x$, 满足 $\bar{y}=5.9761-0.7208\bar{x}=3.3332$									
TLSD法	$y=7.0686-0.8509x$, $\bar{y}\neq7.0686-0.8509\bar{x}=3.9481$									
本文方法	$y=6.2488-0.7954x$, 满足 $\bar{y}=6.2488-0.7954\bar{x}=3.3323$									

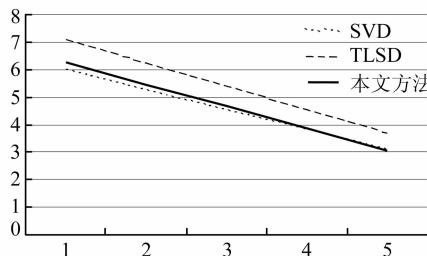


图1 3种方法拟合的直线

Fig. 1 Comparisons of Results

表2 观测对与3种方法的比较

Tab. 2 Observation Value and Comparisons of Results

x	1.6	2.3	2.9	5.8	5.1	6.9	1.3	3.0
y	2.2	4.8	4.3	2.3	6.4	2.7	3.8	5.7
z	1.30	-1.05	1.75	8.55	2.15	8.80	-1.25	-0.21
β_0		β_1	β_2	σ_{β}				
LS	1.270	1.614	-1.167	0.43				
TLSD	1.773	1.591	-1.236	0.14				
本文方法	1.804	1.569	-1.236	0.12				

递推算法的收敛精度,尤其是在观测个数较少的情况下能获得较好的效果,其推导过程和公式符合测量数据处理解算方式,容易编程实现。

参 考 文 献

- [1] Golub G H, van Loan C F. An Analysis of the Total Least Squares Problem[J]. SIAM Journal on Numerical Analysis, 1980, 17(6):883-893
- [2] Schaffrin B. A Note on Constrained Total Least-Squares Estimation[J]. Linear Algebra and Its Applications, 2006, 417: 245-258
- [3] Schaffrin B, Wieser A. On Weighted Total Least-squares Adjustment for Linear Regression [J]. J Geod, 2008, 82:415-421
- [4] 张贤达.矩阵分析与应用[M].北京:清华大学出版社,2005
- [5] 邱卫宁.测量数据理论与方法[M].武汉:武汉大学出版社,2008

第一作者简介:邱卫宁,教授,主要从事测量数据处理方面的研究。

E-mail:wnqiu@sgg.whu.edu.cn

An Improved Algorithm of Total Least Squares for Linear Models

QIU Weining^{1, 2} QI Gongyu^{1, 2} TIAN Fengrui^{1, 2}

(1) School of Geodesy and Geomatics, Wuhan University, 129 Luyou Road, Wuhan 430079, China)

(2) Hazard Monitoring and Prevention Research Center, Wuhan University, 129 Luyou Road, Wuhan 430079, China)

Abstract: Aiming at the condition that among observation equations there are errors in both observation vector and structure matrix, we propose an improved algorithm of total least squares(TLS). Meanwhile, when observation vector and structure matrix are with unequal precision, we deduce the calculation formula using an adjustment model with constraints. Analysis and calculation show that the algorithm that improves approximation accuracy of recursive TLS effectively, can satisfy the conditions needed by fitting equation.

Key words: structure matrix; adjustment model with constraints; TLS iterative solution