

岭估计在二次曲面高程拟合模型数据处理中的应用

周彬^{1,2}, 杜宁¹, 黄筱³, 刘开奋³, 王沿儒³

¹贵州大学矿业学院, 贵州 贵阳

²贵州水利水电职业技术学院, 贵州 贵阳

³中国电建集团贵阳勘测设计研究院有限公司, 贵州 贵阳

收稿日期: 2021年10月26日; 录用日期: 2021年11月16日; 发布日期: 2021年11月30日

摘要

在测量工作中, 由于参数过多、观测结构不合理等各种原因, 会导致各参数之间存在近似线性关系, 此时经典的最小二乘模型解算结果会失真。本文通过二次曲面高程拟合模型着重讨论岭估计在最小二乘模型失效的情况下解算结果的可靠性, 总结岭参数的求解方法, 并验证其解算的可靠性。

关键词

岭估计, 高程拟合, 二次曲面模型, 岭参数选取

The Application of Ridge Estimation in Data Processing of Quadric Surface Elevation Fitting Model

Bin Zhou^{1,2}, Ning Du¹, Xiao Huang³, Kaifen Liu³, Yanru Wang³

¹Mining School of Guizhou University, Guiyang Guizhou

²Guizhou Vocational and Technical College of Water Resources and Hydropower, Guiyang Guizhou

³Power China Guiyang Engineering Corporation Limited, Guiyang Guizhou

Received: Oct. 26th, 2021; accepted: Nov. 16th, 2021; published: Nov. 30th, 2021

Abstract

In the measurement work, due to too many parameters, unreasonable observation structure and

文章引用: 周彬, 杜宁, 黄筱, 刘开奋, 王沿儒. 岭估计在二次曲面高程拟合模型数据处理中的应用[J]. 应用数学进展, 2021, 10(11): 4087-4093. DOI: 10.12677/aam.2021.1011434

other reasons, there will be an approximately linear relationship between the parameters. At this time, the solution results of the classical least squares model will be distorted. In this paper, through the elevation fitting quadric surface model, the reliability of the solution result of ridge estimation in the case of the failure of the least square model is discussed, the solution method of ridge parameters is summarized, and the reliability of its solution is verified.

Keywords

Ridge Estimation, Elevation Fitting, Quadric Surface Model, Ridge Parameter Selection

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

通过 GPS 测量可以得到待测点的大地高, 如果测区地势平坦、范围较小, 可以采用多项式曲面拟合模型来拟合待测点正常高, 其中较为常见的为二次曲面模型。但在实际生产中, 由于参数的选择、观测条件的限制、结构设计的不合理等会导致参数之间有近似共线性, 这使得法方程矩阵 $N = B^T B$ 接近奇异, 此时我们称法方程为病态方程, 为解决法方程病态时的最小二乘估计问题, 有许多学者进行过探索, 如基于 Neumann 级数的有偏估计方法、靶向改变法方程矩阵和基于 LIU 估计的迭代新方法等[1] [2] [3]。

但是在众多有偏估计之中, 岭估计、广义岭估计、主成分估计的影响最大。在岭估计的探索当中, 学者对岭参数的选择方法不断创新, 使得岭估计在实际应用中硕果累累, 尤其是在周跳探测、高频 GNSS 钟差、GPS 数据处理、高程拟合等领域获得了较高的可靠性[4] [5] [6] [7] [8]。

2. 岭估计

2.1. 岭估计的原理

岭估计(Ridge Estimation)是 Hoerl 和 Kennard 于 1970 年提出来的, 是目前最有影响力的有偏估计方法。其模型为:

$$L = BX + \Delta, \Delta \sim N(0 \sim \sigma^2 I) \quad (1)$$

其对应的解为:

$$\hat{X}(k) = (N + kI)^{-1} B^T PL \quad (2)$$

式中 k 为大于零的任意常数, 称为岭参数, 而 $N = B^T PB$ 。可以看出, 当 k 的选择不同, 就会得到不同的岭估计, 当 $k = 0$ 时, 为最小二乘估计。

2.2. 岭估计的性质

为了讨论岭估计的性质, 将(1)式改写为以下形式:

$$L = BX + \Delta = AY + \Delta \quad (3)$$

其中: $A = BG$, $Y = G^T X$, $G = (G_1, G_2, \dots, G_t)$ 。

$G_i (i = 1, 2, \dots, t)$ 为 $B^T B$ 对应于特征值 $\lambda_1, \lambda_2, \dots, \lambda_t$ 的标准正交化特征向量, 且满足

$$A^T A = (BG)^T (BG) = G^T B^T B G = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_t) \quad (4)$$

式(4)称之为 Gauss-Markov 模型的典则形式, Y 称为典则参数。

则典则参数 Y 的最小二乘估计和岭估计分别为:

$$\begin{aligned} \hat{Y}_{LS} &= (A^T A)^{-1} A^T L = \Lambda^{-1} G^T B^T L \\ \hat{Y}(k) &= (A^T A + kI)^{-1} A^T L = (\Lambda + kI)^{-1} G^T B^T L \end{aligned} \quad (5)$$

2.3. 岭参数的选择

在岭参数的选择上, 国内学者进行了相关研究, 较为常见的有 U 曲线法、行列式法和正则化法, 还有部分学者将以上方法进行比较, 确定最终岭参数[9] [10] [11] [12]。

引进岭估计的目的是减少均方误差, 统计学家已经证明了当 $k > 0$ 时,

$$\text{MSE}(\hat{X}(k)) < \text{MSE}(\hat{X}_{LS}) \quad (6)$$

因此在均方误差意义下, 岭估计可改进最小二乘估计。对于岭参数 k 的选择, 一直困扰着学者们, 由于岭参数的选择在于使均方误差更小, 即

$$\text{MSE}(\hat{X}(k)) = \sigma_0^2 \sum_{i=1}^t \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \sum_{i=1}^t \frac{y_i^2}{(\lambda_i + k)^2} \quad (7)$$

最小, 而 Y, X 未知, 故不能用求极值的方法求取岭参数的值。国内外学者们经过不断探索, 提出了许多岭参数的选择和求解的方法, 如岭迹法、双 h 公式法。

1) 岭迹法:

就是以岭估计 \hat{X}_i 的分量 $\hat{X}_i(k) (i=1, 2, \dots, t)$ 作为岭参数 k 的函数, 将 t 条岭迹画函数图像, 选择使 t 条岭迹都处于较为稳定状态下的那个 k 值作为岭参数。

2) 双 h 公式法:

$$k = \frac{h_1 \hat{\sigma}_0^2}{\hat{X}_{LS}^T G \hat{X}_{LS} + h_2 \hat{\sigma}_0^2} \quad (8)$$

当取 $G = I, h_1 = t, h_2 = 0$ 时, 上式就变为了 Hoerl-Kennard-Baldwin 公式:

$$k = \frac{t \hat{\sigma}_0^2}{\hat{X}_{LS}^T \hat{X}_{LS}} \quad (9)$$

2.4. 二次曲面模型

通过 GPS 测量获得的高程属于大地高, 已知大地高和正常高的关系为:

$$H = H_r + \xi \quad (10)$$

其中: H 代表大地高, H_r 代表正常高, ξ 代表高程异常。

二次曲面模型可以表示为:

$$\xi(x, y) = A_1 + A_2 x + A_3 y + A_4 xy + A_5 x^2 + A_6 y^2 \quad (11)$$

由式(10)可知:

$$\xi(i) = H_i - H_n, \quad i = 1, 2, 3, 4, 5, \dots, n \quad (12)$$

其中: H_i 代表第 i 个点的大地高, H_n 代表第 i 个点的正常高。

根据式(11)和式(12), 可列出误差方程:

$$V = BA - \xi \quad (13)$$

式(13)中各项矩阵的表达式为:

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}, \quad V = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix}, \quad A = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & x_1 & y_1 & x_1 y_1 & x_1^2 & y_1^2 \\ 1 & x_2 & y_2 & x_2 y_2 & x_2^2 & y_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & y_n & x_n y_n & x_n^2 & y_n^2 \end{bmatrix} \quad (14)$$

3. 岭估计的算例分析

本算例取自某工程实测数据, 已知 12 个控制点, 经 GNSS 静态观测获得这 12 个点的大地高, 通过四等三角高程测量获得这 12 个点的正常高。这 12 个控制点的坐标及高程如表 1 所示。

取 H2、H4、H5、H16、H17、H45、H46 和 H49 共 8 个点进行模型参数计算点, H15、H18、H19 和 H21 共 4 个点作为检核点, 已知模型参数为 6 个, 则多余观测量为 2 个。

$B^T B$ 的 6 个特征值为:

$$\lambda_1 = 4.8177 \times 10^{-9}, \lambda_2 = 4.2623 \times 10^3, \lambda_3 = 1.8878 \times 10^9, \lambda_4 = 3.1617 \times 10^{17}$$

$$\lambda_5 = 1.3791 \times 10^{22}, \lambda_6 = 6.1291 \times 10^{26}$$

Table 1. Coordinates of known point

表 1. 已知点坐标

| 点号 | 坐标信息 | | | | | 备注 |
|-----|---------------|-------------|------------|--------------|-----------------|-----|
| | X | Y | 大地高(H) | 正常高(H_r) | 高程异常(ζ) | |
| H2 | 2,948,972.536 | 580,925.255 | 67.8800 | 64.5052 | 3.3748 | 计算点 |
| H4 | 2,946,455.299 | 577,260.884 | 69.6220 | 66.1620 | 3.4603 | 计算点 |
| H5 | 2,930,686.780 | 570,740.793 | 79.9050 | 76.4041 | 3.5006 | 计算点 |
| H15 | 2,902,323.493 | 576,888.668 | 150.7280 | 147.6946 | 3.0337 | 检核点 |
| H16 | 2,897,591.122 | 571,095.550 | 270.6260 | 267.4639 | 3.1616 | 计算点 |
| H17 | 2,915,604.344 | 600,816.527 | 203.5520 | 201.1087 | 2.4437 | 计算点 |
| H18 | 2,912,935.612 | 603,020.468 | 143.3720 | 141.0208 | 2.3507 | 检核点 |
| H19 | 2,914,555.779 | 596,539.683 | 98.8580 | 96.2961 | 2.5623 | 检核点 |
| H21 | 2,949,520.601 | 583,991.102 | 85.0110 | 81.7231 | 3.2878 | 检核点 |
| H45 | 2,948,493.557 | 563,891.063 | 93.7300 | 89.8461 | 3.8839 | 计算点 |
| H46 | 2,936,900.855 | 560,156.810 | 96.2560 | 92.3744 | 3.8816 | 计算点 |
| H49 | 2,910,548.146 | 582,274.827 | 103.5030 | 100.5501 | 2.9529 | 计算点 |

按照条件数的计算公式可知, 条件数为:

$$p = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{6.1291 \times 10^{26}}{4.8177 \times 10^{-9}} = 1.2722 \times 10^{35}$$

由于该算例中条件数远大于 1000, 可以认为该方程组为病态方程组[13], 具有严重的复共线性。

3.1. 岭估计的解算

本文采用岭迹法和双 h 公式法对本算例进行解算:

1) 岭迹法

利用 Matlab 编程计算最小二乘拟合结果和岭估计拟合结果, 将岭参数的范围控制在 0~2 内, 步长设置为 0.01, 得到的岭迹图如图 1(a)所示。由岭迹图可以看出, 当 k 仅有极小扰动时, 各参数解算结果就基本稳定。为便于观察, 将 $k=0$ 去除后重新绘制得到岭迹图如图 1(b)所示, 根据图 1(b)可知, 在 $k=0.2$ 时, 各数值已基本稳定, 可取 $k=0.2$ 。

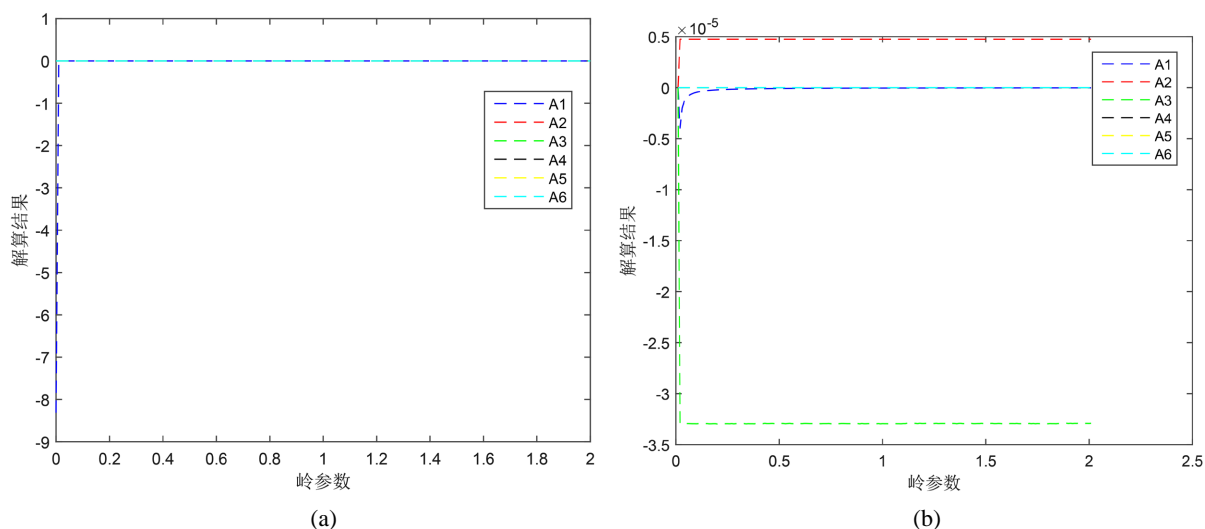


Figure 1. Ridge trace

图 1. 岭迹图

Table 2. Solution results

表 2. 解算结果

| 点号 | 已知高程异常/m | 最小二乘拟合结果/m | 残差/m | 岭估计 | | | |
|------|----------|----------------------|-----------|--------------------------|--------|--------------------------|---------|
| | | | | 双 h 公式法/m | 残差/m | 岭迹图法/m | 残差/m |
| H15 | 3.0337 | -1332.7019 | 1335.7356 | 2.9758 | 0.0579 | 3.0673 | -0.0336 |
| H18 | 2.3507 | -1344.1951 | 1346.5458 | 2.2925 | 0.0582 | 2.3847 | -0.0340 |
| H19 | 2.5623 | -1341.9127 | 1344.4750 | 2.5038 | 0.0585 | 2.5961 | -0.0338 |
| H21 | 3.2878 | -1341.8008 | 1345.0886 | 3.2281 | 0.0597 | 3.3227 | -0.0349 |
| 均方误差 | | 7.1730×10^9 | | 7.2394×10^{-16} | | 4.0107×10^{-14} | |

2) 双 h 公式法

取 $G=I$, $h_1=t$, $h_2=0$, 由式(8)可计算出 $k=3.0$ 。

现将按照经典最小二乘模型及二种岭估计的 k 值求解方程组, 结果见表 2。

3.2. 结果分析

根据解算结果可知:

1) 该模型复共线性强, 最小二乘拟合结果失真, 均方误差很大, 计算结果不准确。

2) 该模型岭估计求解时, 就均方误差而言, 相对最小二乘模型均有很大优化, 但岭迹法小于双 h 公式法解算的结果。

3) 该模型岭估计求解时, 就解算精度而言, 相对最小二乘模型均有很大优化, 但岭迹法优于双 h 公式法解算的结果。

4. 结论

本文通过岭估计原理和二次曲面模型的阐述、岭估计参数的选择和实例计算, 在岭估计的选择和可靠性方面得出如下结论:

第一, 在病态方程组中, 如果最小二乘拟合结果失真, 那么均方误差将很大。采用岭估计方法对病态方程组进行处理后, 可减小均方误差, 并提高解算结果的可靠性。

第二, 由双 h 公式法的原理可知, 其求取的 k 值比较单一, 通过该方法得到的高程残差约为 0.06 m, 精度较低, 难以满足实际生产工作需要。而岭迹法在确定 k 值时, 通过调整参数变化的步长, 绘制岭迹图, 所对应的解能较好地呈现其总体解的情况, 本文通过选取 $k=0.2$, 得到的高程残差绝对值小于 0.035 m, 可以满足实际生产工作需要, 效果比双 h 公式法好。所以岭迹法相对于双 h 公式法, 可直接选择均方误差较小的值作为岭参数, 能够获得较高精度的解。

第三, 在采用岭迹法的时候, 判断岭迹图何时趋于平稳受人为因素影响, 没有统一的标准, 所以导致同种方法所产生的结果也可能因人而异。

参考文献

- [1] 杨秋伟, 白志超, 李翠红. 基于 Neumann 级数的有偏估计方法[J]. 大地测量与地球动力学, 2020, 40(5): 512-516.
- [2] 王飞, 沈丹, 孙嘉聪. 岭参数 K 的选择方法[J]. 数学学习与研究, 2020(6): 133.
- [3] 黄荣臻, 朱宁, 邓超海, 张茂军. 线性回归模型的一类新约束型 LIU 估计[J]. 西南师范大学学报(自然科学版), 2019, 44(11): 1-10.
- [4] 章繁, 刘长建, 冯绪等. 一种基于拓展岭估计的北斗三频周跳实时处理方法[J]. 武汉大学学报(信息科学版), 2020, 45(1): 62-71.
- [5] 吴光明, 鲁铁定. 病态总体最小二乘靶向奇异值修正法[J]. 大地测量与地球动力学, 2019, 39(8): 856-862.
- [6] 杭礼辉, 葛俊祥, 张艳艳. 基于奇异值比值的正则化矩阵修正方法[J]. 现代雷达, 2019, 41(4): 54-58+62.
- [7] 刘金灵. 一类改进广义岭估计的实例研究[J]. 应用数学进展, 2021, 10(1): 92-97.
- [8] 章繁, 刘长建, 冯绪, 李林阳, 王方超. 一种基于拓展岭估计的北斗三频周跳实时处理方法[J]. 武汉: 大学学报(信息科学版), 2020, 45(1): 62-71.
- [9] 顾勇为, 归庆明, 韩松辉. 基于信噪比的正则化方法及其在 GPS 快速定位中的应用[J]. 中国科学: 物理学 力学 天文学, 2010, 40(5): 663-668.
- [10] 刘雁雨. 有偏估计理论研究及其在 GPS 数据处理中的应用[D]: [硕士学位论文]. 郑州: 中国人民解放军信息工程大学, 2005.
- [11] 鲁洋为, 王振杰. 用 U 曲线法确定岭估计中的岭参数[J]. 导航定位学报, 2015, 3(3): 132-134+138.

-
- [12] 罗伟钊. 基于抗病态和抗差的高精度坐标转换解算方法研究[D]: [硕士学位论文]. 长沙: 中南大学, 2014.
- [13] 陶本藻. 测量数据处理的统计理论与方法[M]. 北京: 测绘出版社, 2007.