

岩石抗剪强度计算的稳健回归模型及其应用

张耀平^{1,2}, 曹平¹, 董陇军¹

1. 中南大学资源与安全工程学院, 长沙 410083
2. 江西理工大学应用科学学院, 江西赣州 341000

摘要 针对岩石抗剪强度确定中图解法和最小二乘法存在的问题: ① 只适用于实验数据相关性较高的情况, 不适用于数据离散并有异常值存在的情况; ② 在处理由试验法获得的试验数据时, 最小二乘法由于采用残差平方和, 容易夸大试验数据中异常值的影响, 提出了岩石抗剪强度参数的稳健回归分析方法。该方法在实验数据相关性差、数据离散并有异常值存在的情况下, 具备削弱数据离散和对异常值进行定位的能力, 提高了估计参数的稳健性和可靠性。该方法以残差的绝对值之和代替残差平方和, 并通过复形法求得力学参数, 避免了异常值的二次项, 可有效地减少异常值的影响。通过工程实例表明, 在试验数据的相关性较好时, 两种方法的计算结果相差不大, 但当试验数据的相关性较差, 存在异常值时, 稳健回归方法的计算结果要优于最小二乘法。

关键词 岩石力学; 稳健回归方法; 抗剪强度; 数据离散

中图分类号 TU45

文献标识码 A

文章编号 1000-7857(2010)07-0091-05

A Robust Regression Model and Its Application in Calculating Shear Strength of Rock

ZHANG Yaoping^{1,2}, CAO Ping¹, DONG Longjun¹

1. School of Resources and Safety Engineering, Central South University, Changsha 410083, China
2. School of Applied Sciences, Jiangxi University of Science and Technology, Ganzhou 341000, Jiangxi Province, China

Abstract There are two problems in the graphic method and the Least Squares (LS) method in determining the shear strength of rock. (1) The experimental data should be highly correlated, and separate data and anomalous data are not suitable. (2) The effect of anomalous values in test data is markedly shown because of the quadratic sum of residual errors in the least squares method. This paper proposes a shear strength parameter (Robust) regression analysis method. This method can be applied to analyze experimental data with poor correlation, and with anomalous values, and the effects of discrete and abnormal positioning can be weakened to improve the robustness and reliability of estimated parameters. In order to reduce the effect of anomalous values in the test data, the sum of residual absolute value is used instead of the quadratic sum of residual errors in the robust regression method, then the quadratic terms of anomalous value could be avoided, and the effect of anomalous values could be reduced. It is shown that the results of the two methods are the same when the test data is good, but the results of the robust regression method are better than those obtained by the least squares method when there are anomalous values in the test data.

Keywords rock mechanics; robust regression; shear strength; anomalous values

0 引言

岩石抗剪强度参数即岩石黏聚力和内摩擦角是岩体工

程设计与计算的重要力学指标^[1-3]。由于岩石材料的不均匀性、内部结构的随机性以及测试取样点的空间变异性, 岩石力学

收稿日期: 2009-11-23

基金项目: 国家自然科学基金项目(50774093)

作者简介: 张耀平, 博士研究生, 研究方向为岩石力学和采矿工程, 电子信箱: zyp2005csu@126.com; 曹平(通信作者, 中国科协所属全国学会个人会员登记号: S330730017M), 教授, 研究方向为岩土力学与工程, 电子信箱: pcao_csu@tom.com

参数测定值具有很大的离散性^[6-15]。传统的计算方法主要有点群中心法、优定斜率法和最小二乘法。实践表明:点群中心法具有很大的任意性,尤其是数据点过少时更为显著;优定斜率法虽然综合考虑了抗剪强度的影响因素,但无法说明所给出的黏聚力和内摩擦角是否为最优,以及波动情况、回归方程的相关性;最小二乘法(LS)是目前应用最为普遍的确定滑动面力学参数的方法,其精度较前两种要好,但受异常值的影响较大。对于滑动面的力学试验来说,一般不可能占有大量的测试值,如果去掉有限数据中的异常值,则同时去掉了异常值中所蕴涵的信息,从而使回归得到的 c 、 φ 值失去代表性和真实性;若不去掉异常值,则必须减少异常值对参数的影响。针对岩石抗剪强度确定中图解法和最小二乘法存在的问题:①只适用于实验数据相关性较高的情况,不适用于数据离散并有异常值存在的情况;②在处理试验法获得的试验数据时,传统的最小二乘法由于采用残差平方和,容易夸大试验数据中异常值的影响,本文提出了岩石抗剪强度参数的稳健(Robust)回归分析方法。该方法在实验数据相关性差、数据离散并有异常值存在的情况下,具备削弱数据离散和对异常值定位的能力,提高了估计参数的稳健性和可靠性。该方法以残差的绝对值之和代替残差平方和,并通过复形法求得力学参数,避免了异常值的二次项,可以有效地减少异常值的影响。

1 稳健回归理论

估计的稳健性(Robustness)概念指的是在估计过程中产生的估计量对模型误差的不敏感性。因此,稳健估计是在较宽的资料范围内产生的优良估计。如在独立同分布正态误差的线性模型中,最小二乘估计(LSE)是有效无偏估计,而当误差是非正态分布时,LSE不一定是最有效的。误差分布事先不一定知道,故有必要考虑稳健回归的问题^[16]。稳健回归(Robust Regression)估计,若误差为正态,它比LSE稍差,若误差为非正态,则比LSE要好得多。这种对误差项分布的稳健特性,常能有效排除异常值干扰。一般回归模型:

$$Y_i = \sum_{j=1}^p x_{ij}\beta_j + e_i \quad i=1, 2, \dots, n \quad (1)$$

其中, $\beta_1, \beta_2, \dots, \beta_p$ 为未知回归系数, $e_1, e_2, \dots, e_n$ 独立同分布,均值为0。

最小二乘法是找到一组 $\beta_1, \beta_2, \dots, \beta_p$,使得表达式 $\sum_{i=1}^n \left(Y_i - \sum_{j=1}^p x_{ij}\beta_j \right)^2$ 达到最小时作为代价函数。但这样做往往使得远离数据群体的数据(很可能是异常值)对残差平方和影响比其他数据大得多,因为LSE为了达到极小化残差平方和的目的,必须迁就远端的数据,所以异常值对于参数估计相当敏感(图1)。

稳健回归的基本思想是采用迭代加权最小二乘估计回

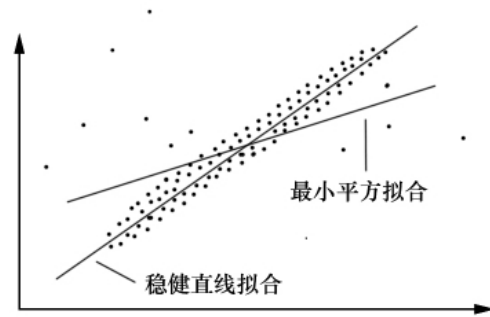


图1 稳健回归与最小二乘拟合

Fig. 1 Robust regression and least squares fitting

归系数,根据回归残差的大小确定各点的权重 w_i ,以达到稳健的目的,其优化的目标函数 $G_{\min}$ 为

$$G_{\min} = \sum_{i=1}^n w_i \left(Y_i - \sum_{j=1}^p x_{ij}\beta_j \right)^2 \quad (2)$$

为减少“异常点”作用,可对不同的点施加不同的权重,即对残差小的点给予较大的权重,而对残差较大的点给予较小的权重。根据残差大小确定权重,并据此建立加权的LSE,反复迭代以改进权重系数,直至权重系数的改变小于一定的允许误差。参数 β_j 可采用迭代加权最小二乘方法求解。目前构造权重的方法很多,得到的稳健回归估计大同小异。表1给出了10种不同定义权重的公式。在以下公式中,都用到一个“标准化”的残差指标 u_i :

$$u_i = e_i / s = 0.6745 \times e_i / \text{med}(|e_i - \text{med}(e_i)|) \quad (3)$$

式中, $\text{med}()$ 为中位数, s 为残差尺度。

2 岩石抗剪强度确定的稳健回归模型

岩石抗剪强度参数黏聚力 c 和内摩擦角 φ 确定的理论基础是摩尔-库仑准则,具体形式可表示为

$$\tau = c + \sigma \tan \varphi \quad (4)$$

或

$$\sigma_1 = \frac{2c \cos \varphi}{1 - \sin \varphi} + \sigma_3 \frac{1 + \sin \varphi}{1 - \sin \varphi} \quad (5)$$

式中, τ 和 σ 分别为土体直剪试验时的剪切强度和法向应力, σ_1 和 σ_3 分别为土体三轴试验时的轴向应力和侧向围压。

在线性回归时,其统一形式为

$$y = a + bx \quad (6)$$

结合式(2),可以建立 a, b 的基于稳健回归分析的一个目标函数,当这个目标函数到达最小值时,所得的值 a, b ,即为所求值(进而可反求出 c, φ),即

$$\min X = \sum_{i=1}^n |y_i - a - x_i b| \quad (7)$$

为了使式(7)达到最小,需要采用优化方法求解,如单纯形法、复形法等,本文采用复形法。

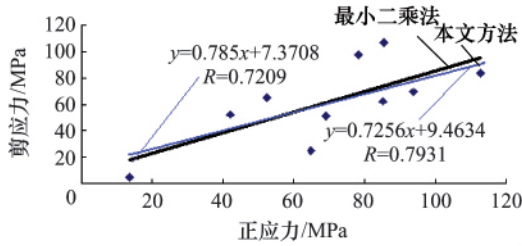


图2 直剪试验分析结果比较

Fig. 2 Comparison of analysis results for shear test data

表5 岩石三轴试验结果

Table 5 Results of rock tri-axial test

样本编号	1	2	3	4	5	6	7	8
σ_1/MPa	66	70	87	102	111	104	122	126
σ_2/MPa	0	0	0	5	5	10	10	10
样本编号	9	10	11	12	13	14	15	16
σ_1/MPa	144	188	188	192	220	228	235	320
σ_2/MPa	15	15	20	20	20	30	30	40

表6 三轴试验数据不同方法分析结果比较

Table 6 Comparison of analysis results for tri-axial test data

方 法	b	a	R	
最小二乘回归	72.165672	5.8623881	0.973937	
Andrews 法	71.925521	5.8414960	0.977338	
Biweight 法	71.927213	5.8415122	0.977316	
Cauchy 法	71.828558	5.8372502	0.978520	
稳 Fair 法	71.513296	5.8514692	0.981122	
健 Hampel 法	72.165672	6.7034433	0.973937	
回 Huber 法	71.938781	5.8285873	0.977090	
归 Logistic 法	70.918530	5.8057844	0.977283	
	Median 法	70.000000	5.9000000	0.999841
	Talworth 法	72.165672	5.8623881	0.973937
	Welsch 法	71.894592	5.8394287	0.977738

表7 三轴试验数据分析结果比较

Table 7 Comparison of analysis results for tri-axial test data

分析方法	黏聚力 c/MPa	内摩擦角 $\varphi/^\circ$
最小二乘法	14.90	45.13
随机-模糊线性方法	14.96	44.57
本文方法	14.41	45.27

从表3、表6可以看出,稳健回归的相关性系数明显大于最小二乘回归方法的相关性系数;从表6和图3可以看出,当试验数据的相关性较好时,采用稳健回归方法和最小二乘法进行计算的结果非常接近,两种计算方法的误差值也很接近,说明了采用稳健回归方法进行岩石抗剪参数计算的可靠性;但当试验数据较为分散,存在异常值时,见表3和图2,最小二乘法由于采用了残差平方和,从而夸大了异常值的影

响,而稳健回归方法则以残差的绝对值之和代替残差平方和,避免了异常值的二次项,有效地减少异常值的影响,其误差值要低于最小二乘法的计算结果,因而更合理一些。与模糊随机线性分析相比,本文方法不受拟合时人为确定隶属度的影响,可避免人为原因分析随机性及模糊性时可能造成的误差。

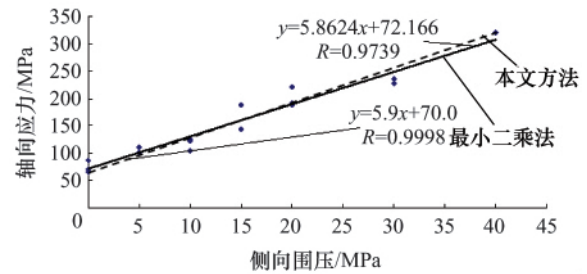


图3 三轴试验分析结果比较

Fig. 3 Comparison of analysis results for tri-axial test

4 结论

1) 在对试验法获得的试验数据进行处理时,传统的最小二乘法由于采用残差平方和,容易夸大异常值的影响。本文提出的稳健回归方法则以残差的绝对值之和代替残差平方和,并通过复形法求得力学参数,从而避免了异常值的二次项,可以有效地减少异常值的影响。

2) 稳健回归方法通过选择影响函数,调整权因子,确定抗剪强度实验参数,可以削弱数据离散和异常值的干扰,提高估计参数的稳健性与可靠性。

3) 通过工程实例表明,在试验数据的相关性较好时,最小二乘法和稳健回归方法的计算结果相差不大,但当试验数据的相关性较差、存在异常值时,稳健回归方法的计算结果要优于最小二乘法。

4) 稳健回归法利用计算机进行编程计算,发挥了速度快、精度高的优势,增强了该计算方法的适用性,完善了岩石抗剪强度计算理论,为异常值存在时的岩石抗剪强度取值提供了一条新的途径。

参考文献 (References)

- [1] Skempton A W. Long-term stability of clay slopes [J]. *Geotechnique*, 1964, 14(2): 77-102.
- [2] 郑明新. 论滑带土强度特征及强度参数的反算法 [J]. *岩土力学*, 2003, 24(4): 528-532.
Zheng Mingxin. *Rock and Soil Mechanics*, 2003, 24(4): 528-532.
- [3] 田斌, 戴会超, 王世梅. 滑带土结构强度特征及其强度参数取值研究 [J]. *岩石力学与工程学报*, 2004, 23(17): 2887-2892.
Tian Bin, Dai Huichao, Wang Shimei. *Chinese Journal of Rock Mechanics and Engineering*, 2004, 23(17): 2887-2892.
- [4] 廖秋林, 杨志法, 赏彦军, 等. 川藏公路 102 滑坡滑动带力学参数的反分析 [J]. *岩石力学与工程学报*, 2004, 23(24): 4119-4123.
Liao Qiulin, Yang Zhifa, Shang Yanjun, et al. *Chinese Journal of Rock Mechanics and Engineering*, 2004, 23(24): 4119-4123.

- [5] Chandler R J. Back analysis techniques for slope stabilization works: A case record[J]. *Geotechnique*, 1977, 27(4): 479-495.
- [6] 董陇军, 王飞跃. 基于未知测度的边坡地震稳定性综合评价 [J]. 中国地质灾害与防治学报, 2007, 18(4): 74-78.
Dong Longjun, Wang Feiyue. *The Chinese Journal of Geological Hazard and Control*, 2007, 18(4): 74-78.
- [7] 刘春. 黄麦岭磷矿边坡岩体抗剪强度参数的随机-模糊法取值研究[J]. 岩石力学与工程学报, 2005, 24(4): 653-656.
Liu Chun. *Chinese Journal of Rock Mechanics and Engineering*, 2005, 24(4): 653-656.
- [8] 董陇军, 李夕兵, 宫凤强. 膨胀土胀缩等级分类的未知均值聚类方法及应用[J]. 中南大学学报: 自然科学版, 2008, 39(5): 1075-1080.
Dong Longjun, Li Xibing, Gong Fengqiang. *Journal of Central South University: Science and Technology Edition*, 2008, 39(5): 1075-1080.
- [9] 张志刚, 王鹏, 李安贵, 等. 随机-模糊线性回归模型的参数估计及应用[J]. 北京科技大学学报, 2004, 26(6): 326-329.
Zhang Zhigang, Wang Peng, Li Angui, *et al.* Regression parameters estimation of random-fuzzy linear regression model and its application[J]. *Journal of University of Science and Technology Beijing*, 2004, 26(6): 326-329.
- [10] 李胡生, 魏国荣. 用随机-模糊线性回归方法确定岩石抗剪参数[J]. 同济大学学报, 1993, 21(3): 421-429.
Li Husheng, Wei Guorong. *Journal of Tongji University*, 1993, 21(3):

- 421-429.
- [11] 邵小曼. 岩石抗压强度的随机-模糊分析 [J]. 岩石力学与工程学报, 2001, 20(增): 1776-1779.
Shao Xiaoman. *Journal of Rock Mechanics Engineering*, 2001, 20(suppl): 1776-1779.
- [12] 李华晔, 黄志全, 刘汉东, 等. 岩基抗剪参数随机-模糊法和小浪底工程 c, φ 值计算[J]. 岩石力学与工程学报, 1997, 16(2): 155-161.
Li Huaye, Huang Zhiquan, Liu Handong, *et al.* *Chinese Journal of Rock Mechanics and Engineering*, 1997, 16(2): 155-161.
- [13] 杨强, 陈新, 周维垣. 抗剪强度指标可靠度分析 [J]. 岩石力学与工程学报, 2002, 21(6): 868-873.
Yang Qiang, Chen Xin, Zhou Weiyuan. *Chinese Journal of Rock Mechanics and Engineering*, 2002, 21(6): 868-873.
- [14] Dong Longjun, Peng Gangjian, Fu Yuhua, *et al.* Unascertained Measurement classifying model of Goal collapse prediction [J]. *Journal of Coal Science & Engineering: China*, 2008, 14(2): 221-224.
- [15] 王兴明, 董陇军, 付玉华. 基于未知均值分类理论的岩爆发生和分级预测方法[J]. 科技导报, 2009, 27(18): 78-81.
Wang Xingming, Dong Longjun, Fu Yuhua. *Science & Technology Review*, 2009, 27(18): 78-81.
- [16] Martin R D. Robust statistics with the s-plus robust library and financial applications [M]. New York: Insightful Corp Presentation, 2002 (1-2): 17-18.

(责任编辑 代丽)

第6届国际稀土开发与 应用研讨会

时 间: 2010年8月
地 点: 北京

主办单位: 中国稀土学会

通信地址: 北京市海淀区学院南路76号
中国稀土学会办公室 (100081)
电 话: 010-62182748
传 真: 010-62173501
电子信箱: csre@cs-re.org.cn