国家自然科学基金资助项目

KANGCHA ZUIXIAO ERCHENGFA

抗差最小二乘法

周江文　黄幼才
杨元喜　欧吉坤　著

华中理工……社

抗差最小二乘法

周江文　黄幼才
杨元喜　欧吉坤　著

国家自然科学基金资助项目

华中理工大学出版社

（鄂）新登字第 10 号

图书在版编目(CIP)数据

抗差最小二乘法/周江文等著
武汉：华中理工大学出版社， 1997 年 1 月
ISBN 7-5609-1438-1
Ⅰ. 抗…
Ⅱ. ①周… ②黄… ③杨… ④欧…
Ⅲ. 最小二乘法误差付计
Ⅳ. O211.67

抗差最小二乘法
周江文 黄幼才
杨元喜 欧吉坤 著
责任编辑：李立鹏
*
华中理工大学出版社出版发行
（武昌喻家山 邮编：430074）
新华书店湖北发行所经销
华中理工大学出版社照排室排版
二炮印刷厂印刷
*
开本：850×1168 1/32 印张：10.875 字数：256 000
1997 年 1 月第 1 版 1997 年 1 月第 1 次印刷
印数：1-2 000
ISBN 7-5609-1438-1/O·161
定价：12.00 元

内容提要

本书是关于抗差估计(Robust Estimation)的一部专著. 全书共十章. 第一至第六章适度介绍了抗差估计的基本理论和方法. 在深入分析经典最小二乘法的基础上,重点研究如何根据抗差估计理论,将经典最小二乘法改造成抗差最小二乘法. 第七至第十章结合测量实际,研究各种不同测量平差模型的抗差化. 内容涉及确定参数模型、随机参数模型、有偏估计及动态数据模型等.

本书可供数理统计、测量、化学、医药、生物、地质等与数据处理有关专业科技人员和高等院校相关专业师生参考.

前　言

经典最小二乘法，经历了 100 余年的发展和考验，已成为许多领域多种类型数据处理的最广泛使用的方法. 对于严格服从正态分布的数据，最小二乘估值具有方差最小且无偏的统计特性. 最小二乘法的数学模型简洁，计算方法也比较简单. 因而受到普遍欢迎.

然而，经典最小二乘法亦存在严重缺点. 它不具备抗御粗差的能力. 如果不排除数据中粗差的影响，最小二乘估值将受到歪曲甚至不能接受.

实际中，严格服从正态分布的数据几乎没有. 人们曾研究了多种以统计检验为主的粗差探测方法来排除粗差的影响，而这些方法检测和定位粗差的能力有较大局限. 因此需要寻求新的理论和方法来抗御粗差的影响. 抗差估计(Robust Estimation)理论就是适应这种需要发展起来的.

自 1964 年 P. J. Huber 发表抗差估计的论文以来，经过众多数理统计学家不断的开拓和耕耘，抗差估计理论已经发展成为一门受到较多关注的分支学科. 所谓抗差估计，实际上是在粗差不可避免的情况下，选择适当的估计方法使参数估值尽可能减免粗差的影响，得到正常模式下的最佳估值. 它既有明确的针对性，又有一定的适应性.

抗差估计理论表明，极大似然型估计(M 估计)可以根据等价权原理，化成最小二乘估计的形式；或者说，利用等价权，可给经典最小二乘法赋予抗差能力. 由此形成的这种估计方法，既实现了抗差估计原则，又保留了最小二乘法算式简洁，计算方便的优点. 特别是已发展成熟的以最小二乘法为基础的模型和算法，形式上仍可保持，因此称之为抗差最小二乘法. 它所对应的分布模式是某种

污染分布.

本书在深入分析经典最小二乘法的优点和不足的基础上，重点研究如何根据抗差估计理论，将经典最小二乘法抗差化，改造成为抗差最小二乘法.

本书密切结合测量实际，研究如何对各种不同的测量平差模型进行改造. 关于等价权的探讨，内容涉及确定参数模型、随机参数模型、有偏估计以及动态数据模型等. 著者希望能对其他专业误差理论和数据处理的研究也能起到抛砖引玉的作用.

本书对抗差估计的基本理论有适量的介绍，但繁简有别，也没有作精心的编织，目的只是为读者提供有关的参考内容.

本书是国家自然科学基金资助项目“抗差最小二乘法理论”的一部分，该书的出版得到中国科学院测量与地球物理研究所动力大地测量学开放研究实验室以及华中理工大学出版社的积极支持，在此表示衷心的感谢.

著者于武昌

1995 年 8 月

Preface

Having gone through the development in theoretical research and the test in application for more than one hundred years, classical least squares approach has become a widely used tool for processing various types of data in many fields. The least squares estimate is an uniformly minimum variance unbiased estimate if observations are coming from the normal population. This method becomes very popular because its mathematical formulas and computing algorithm are relatively simpler than others.

However, the main drawback of the least squares method is being poorly resistant to the effect of gross errors. In other words, the least squares estimate will be seriously distorted or unacceptable if the gross errors are not removed from data set.

In fact, an observed data set following strictly normal distribution could be hardly found. Many methods based on statistical test for gross error detection have been developed. But the capability of these approaches to detect and identify gross errors is more or less limited. Therefore, it is necessary to search for robust methods efficiently against the influence of gross errors on estimates.

Robust estimation theory has become an important and interesting branch of statistics with efforts made by many statisticians since P. J. Huber published his paper entitled "Robust Estimation of a Location Parameter" in 1964. Robust estimation means that effect of gross errors on the estimate of a parameter may be reduced sufficiently or eliminated by choosing proper approaches so as to obtain the best estimate in the presence of gross errors as almost the same as the result derived from regular observations.

According to the principle of equivalent weight, M-estimator can be transformed into the form of classical least squares, in other words, classical least squares estimator can be robustified by using equivalent weight. Based on this principle, the robustness of classical estimators can be realized and advantages of its simplicity in mathematical forms and ease in computation still remain. Particularly, many well-developed mathematical models and computing methods based on classical least squares are still applicable. From this point of view, the data processing derived in this monograph is called robustified least squares method, the corresponding distribution of observations being some contaminated distribution.

On the basis of studying the advantage and disadvantage of classical least squares in a deep-going way, the concentration in this book is on robustification of classical least squares according to the robust estimation theory. Dealing with problems of gross errors encountered in surveying data processing, the research on the robustification of various surveying adjustment models has been done. The method of equivalent weight covers parametric models, stochastic parametric models, biased estimators, models for handling dynamic observations and so forth. Authors of this book wish it could offer some suggestions to those engaged in different fields also. The fundamental theory of robust estimation has been more or less introduced, aiming at providing references rather than a well organized text book.

This book is a part of the project supported by National

Natural Science Foundation of China (NSFC) entitled "Theory of Robustified Least Squares Method". We should like to express our heartfelt thanks to the Foundation Committee for their financial support. We are also very grateful to the Laboratory of Dynamical Geodesy, Institute of Geodesy and Geophysics, Chinese Academy of Sciences, and the Press of Huazhong University of Science and Technology for their support to publish this monograph.

ZHOU Jiangwen
HUANG Youcai
YANG Yuanxi
OU Jikun
August, 1995

目　录

Contents

第一章　抗差估计理论的发展

§1.1　抗差估计综述

一、误差、信息和抗差估计的概念

人们在实践活动中，通过对某一目标或事件的观测，获得了大量的数据，这就是数理统计中的样本. 利用适当的数学方法，对观测数据进行分析，人们能对事件的发展过程找到规律，甚至可推测事件的未来.

观测总是会有误差的. 通常有一类误差是个别可预测的，称为偶然误差；另有一类误差带有部分规律性的序列，称为系统误差. 此外还有粗差(Outlier 或用 Gross error)，泛指离群的误差. 统计学家根据大量数据，指出在生产实际和科学实验中，粗差的出现约占观测总数的 1%～10%[73]. 粗差往往带来不良后果，影响正确的结论. 随着科学技术的进步，人们对测量结果的精度要求越来越高. 因此，寻求有效的对付粗差的策略显得越来越重要，特别是在自动化观测水平较高，数据量大的场合.

数据(或样本)提供给人们的是有关事件的信息. 数据中的信息可分为三类：1) 有效信息　通过它能准确揭示观测数据的分布模式，因而可采用相应的极大似然估计，获得最优估值；2) 可利用信息　虽不能用来揭示数据的准确分布，但它能反映数据分布的基本特征，可以为提高估值精度作出贡献；3) 有害信息　主要是指粗差，它起干扰正常估计的不良作用，必须设法排除或抑制它的影响.

对于粗差的处理，最早引起重视的是粗差探测方法，归纳起来