A. BJERHAMMAR

THEORY OF ERRORS AND GENERALIZED MATRIX INVERSES

THEORY OF ERRORS AND GENERALIZED MATRIX INVERSES

ARNE BJERHAMMAR

Professor at the Royal Institute of Technology, Stockholm, Sweden

ELSEVIER SCIENTIFIC PUBLISHING COMPANY
Amsterdam - London - New York 1973

ELSEVIER SCIENTIFIC PUBLISHING COMPANY
335 JAN VAN GALENSTRAAT
P.O. BOX 1270, AMSTERDAM, THE NETHERLANDS

AMERICAN ELSEVIER PUBLISHING COMPANY, INC.
52 VANDERBILT AVENUE
NEW YORK, NEW YORK 10017

Library of Congress Card Number: 72-179998

ISBN 0-444-40981-5

With 62 illustrations and 87 tables

Printed in The Netherlands

Preface

How to read this book

In most books it is important to read all chapters in consecutive order. This is not necessarily the best way to read this book and therefore we give you some guidelines. An elementary presentation of the theory of errors will be found in chapters 2–8. The study is rather concentrated and detailed proofs are not always included. It is anticipated that the reader will use this part of the book as a refresher course and a textbook on statistics should be consulted whenever needed.

In the following chapters multidimensional problems are analyzed in various ways. The principal problems are normally first presented in a matrix approach with the use of classical infinitesimal operations. A general approach then follows without any infinitesimal operations. Only the elementary algebraic matrix operations, subtractions, multiplications and inversions are used in this study.

Generalized matrix inverses are introduced in chapter 9 but many readers will find it convenient to make a parallel study of appendix 2 in order to obain more detailed proofs. An early study of chapter 29 will be of value for readers who ask for an easy approach to generalized inverses in the method of least squares.

The multidimensional normal distribution is presented in chapter 18 together with a general proof for independent sums of squares.

There are applications in almost all chapters but the greater part of applications will be found in chapters 14–18. Most of the applications are taken from plane geometry, surveying, photogrammetry, geodesy, physics and geophysics. In most studies an analysis of variance together with hypothesis testing is included.

It is known, that many readers use a textbook more or less as an encyclopedia, and therefore some over-lapping between the different chapters has been used. See for example the introduction to chapters 11 and 11.1.

Chapters 23, 24 and 25 are devoted to more sophisticated estimation problems.

The nonstochastic error concept is normally not considered in the theory of errors, but in chapter 28, the basic concepts of error norms can be found. For modern computer programming, this section is of special importance.

It should be noted that, for educational reasons, we often present observations with only very few decimal digits in the given examples. This is in contradiction to sound practical procedures. Furthermore, most examples have been solved in a computer and only a limited number of digits is included in the answers. Therefore, small differences sometimes may be found when recomputing.

Generation in a computer of the normal, t- and F-distributions from random numbers is discussed. Most numerical studies of systems of equations include a determination of the condition number of the system in order to give the computer aspects. Manipulations with the condition numbers are used in order to improve the stability of solutions.

It is emphasized that the consecutive testing with analytical methods should be verified with the use of a discrete computer approach.

This book is the English edition of the Swedish textbook *Felteori* by the author, published in 1955. All chapters from 22 on are new, and some of the earlier chapters have been revised.

Acknowledgement

The author wishes to express his gratitude to all his colleagues and friends who in various ways have contributed to this study. Valuable comments and remarks have been made by Messrs E. Bruchfeld, G. Galvenius and G. Dahlquist. Very essential contributions has been given by Messrs B.-G. Reit, C. Josefsson and L. Sjöberg.

I finally wish to acknowledge the assistance of professor E. Dauphin and Mr. I. Brook in the translation of my manuscript.

Stockholm, 1972
Arne Bjerhammar

Contents

Chapter 1

Introduction

Modern society is highly dependent on advanced measuring techniques. We are all intrigued by the outstanding achievements of the astronauts and others who, together, have conquered the moon, and who perhaps will soon conquer other planets. These challenging results, all based on the most refined measuring techniques, have tremendously increased our interest in measurements and their mathematical treatment, during the latest decade.

Werner Heisenberg formulated, in 1927, the "uncertainty" principle which defines the ultimate accuracy of measurements. He showed that it is basically impossible to fix the coordinates of a particle with higher accuracy than is given by the magnitude of Planck's quantum constant, h, which is of the order 6.6×10^{-27} erg sec. Heisenberg claimed that an observation must always affect the event being observed, and this interference will finally lead to a fundamental limitation in the accuracy of the observation.

Using this approach from quantum physics, we can hardly expect that the true value of a physical quantity can be determined by measurements. However, this fundamental indeterminacy is more philosophical than practical for most operations that we have to face in life. If the length of a scale is not uniquely defined for any moment, according to Heisenberg it will be impossible to determine the error in any measurement made with the scale. Apparently, the definition of an error is not as simple as sometimes is anticipated. In general, most people accept the philosophy that statements are either correct or incorrect. The perceptions of different persons who are witness to the same event are often completely different. There are countless examples of trials at court, where the "truth" is different for the two opponents. Very seldom can we claim that one of the opponents is lying and the other is telling the truth. When Mr A claims that a crowd included four persons, and Mr B claims he only saw two persons, both statements might be correct and error free. If the correct number is four, we commit an error when using Mr B's statement for a determination of the size of the crowd. This means that an error can seldom be uniquely defined in practical life. When we come to more sophisticated problems the situation will be more complicated and we have to find a useful definition of error. For most applications the following definition will be acceptable:

$$\text{Error} = \text{Observation} - \text{True Value}$$

We have already indicated that the "true value" is normally unattainable. However, in mathematics the uncertainty principle is often irrelevant and here we even find operations which can be uniquely defined in an exact way. For example, we can define a linear algebra using $a + b + c = 0$.

1.1 Rounding Errors

In applied mathematics we can expect certain difficulties when we try to present a mathematical relation using digital numbers. If the exact primary relation is written:

$$1/3 + 1/3 - 2/3 = 0$$

then the digital presentation might be

$$0.333 + 0.333 - 0.667 = -0.001$$

Clearly we are here exposed to rounding errors.

This type of error can normally be determined with any wanted number of digits, but in practical application we are restricted in various ways. The limitations are not directly linked with any scientific difficulties. It is more a question of the restrictions of time and money that finally have to be considered when deciding on the adequate number of digits. For the solution of very large systems of equations rounding errors start to be critical and have to be carefully considered. In the following study we make no direct analysis of the rounding errors. However, indirectly the rounding errors may have an influence on our investigations. We make the assumption that the number of digits is chosen in such a way that the influence of the rounding errors will be insignificant for our study. It should be noted that there is also a second type of rounding errors which is relevant. First we make a record of our observations. These records define our measurements. If the records are given with too few digits we will suffer from rounding errors that give sets of observations with identical results. When we use a measuring tape which is only graduated in meters we can normally expect to obtain a set of observations that are identical at least for shorter distances. For a tape with decimeter graduation perhaps two or three sets of identical observations can be expected. Finally, if we use a tape which is graduated in millimeters we may expect that all observations are different.

An example is given below:

Case 1:	Records from meter graduation	One set: 47, 47, 47, 47, 47
Case 2:	Records from decimeter graduation	Two sets: 47.3, 47.3, 47.2, 47.2, 47.2
Case 3:	Records from millimeter graduation	47.251, 47.254, 47.244, 47.241, 47.243

This example indicates that the records are made with an insufficient number of digits for all cases except the last. As soon as we obtain some sets of identical records we can suspect that the technique is inadequate and should be improved. The rounding errors can be treated with statistical methods using rectangular distributions. In the general case numerical analysis is used. A special study of the rounding errors will be found in chapter 28.

1.2 Non-accidental Errors

The most obvious non-accidental error is the blunder or gross error. When a measurement is made and we observe 57 instead of the correct value of 75 we have committed a blunder. Normally all such errors can be detected before they can influence the final result and we are not going to make any further analysis of such errors.

In some cases it is possible, at an early stage, to determine that certain errors have a clear *systematic* nature. An obvious case is when we have measured a length using a measuring system in which the unit has a constant error. If the reference unit is a meter and is 0.1% in error we will have 1.0 m error in a distance of 1000 m. In order to give an example of methods for the elimination of systematic errors we can mention geodetic electro-optical distance measurements. The International Union of Geodesy and Geophysics has adopted a value of 299,792.5 km/sec as the vacuum velocity of light. This value seems to have an error in the order of 100 m/sec, which means that a systematic error in the order of 10^{-6} of the distance cannot be excluded if we use the international meter-definition as the reference unit. However, if we disregard the meter-definition and, instead, consider the given value for the velocity of light as the fundamental definition, then we are in a much better situation. In astronomy and geodesy the velocity of light is a more natural primary unit. The distances to stars are measured in light years or, perhaps more conveniently, in nanoseconds. However, these contradictions are of very little practical importance for most studies and are only mentioned in order to show that systematic errors sometimes are caused by unsuitable definitions of the primary units. With the velocity of light in vacuum uniquely defined, we can determine any wanted length unit with an extremely small systematic error, since atomic clocks give us time with very small systematic errors. The frequency drift in an atomic clock is often smaller than 10^{-11} of the nominal frequency. With our present methods using crystal clocks in field operation, we still have to consider the frequency errors of the oscillators.

There are considerable difficulties involved in transferring the velocity of light in vacuum to an atmosphere. The atmospheric errors are not restricted to the refractive index. We can seldom expect that our light ray travels in an atmosphere with constant physical properties. In fact, we know that the atmosphere has a tendency to be distributed in "layers" around the earth and we must apply a correction for the curvature of the light ray. This correction can be computed using parameters for a "normal atmosphere" but it will probably be more correct to make use of measurements of the vertical angles at the ends of our measured line. Another approach is to make use of the dispersion between red and blue light.

Geodetic distances are, normally, reduced to the reference ellipsoid and, therefore, an additional correction is needed. This correction brings the gravity field of the earth into our computations. The elevations of our points A and B above the reference ellipsoid are normally computed with the geoid as an intermediate reference surface. The geoid is the equipotential surface that coincides with mean sea level. We can determine this surface after solving rather complicated integral equations which require gravity data from the entire earth. The height of the geoid above the reference surface is of the order of ±100 m. In this study we sometimes have to consider earth tides which give daily uplifts of the crust in the order of ±0.30 m.

If we go back to original measurement of distances we have to add a correction for the change of the group velocity in the optical reflector. Furthermore, each individual reading of the phase angle in the electro-optical device has to be corrected for systematic errors when using the earlier types of electro-optical devices. For instruments of this type the total distance is computed as the number of whole modulation wave lengths and an additional fraction of a modulation wave length. The number of whole wave lengths can normally be uniquely determined. The additional fraction of a wave length is determined with an electrical delayer. On account of the instability of the electrical delayer it is always necessary to calibrate it, using an optical delayer as standard. It was soon found that calibrations of this type are inconvenient in all practical work. Therefore, *self-calibrating* systems are to be preferred whenever it is possible. In the newer instruments the phase angles are measured directly. This method is self-calibrating in as much as the final fraction of a wavelength can be determined without any significant errors. Similar relations are valid when measuring distances with microwave systems.

The corrections and reductions which have to be considered in connection with electro-optical distance measurements are summarized below:

1. Correction to the accepted value of the vacuum velocity of light.
2. Frequency correction for the oscillators of the measuring instrument.
3. Reduction to group velocity using wave length, temperature and pressure measurements.
4. Correction for the curvature of the light ray.
5. Correction for the internal and external eccentricities of the instrument and the reflector.
6. Correction for the errors of all auxilliary instruments such as theodolites, thermometers, pressure meters, gravity meters and levelling instruments.
7. Projection to the geoid.
8. Projection to the reference ellipsoid.

Thus there may be several hundred thousand sources of systematic errors which all influence the final result of a single distance reduced to the reference surface of the earth. Clearly, only geodetic measurements of the highest accuracy will need all the corrections 1–8.

Instruments which are used for the measurements of angles have also facilities for self-calibration. With the standard type of theodolite this is accomplished by making readings at two opposite positions on the circle graduation. Then the error of eccentricity is eliminated. Systematic errors in the graduation are eliminated by repeating the measurements in several series which are distributed all around the circle. Systematic errors in the line of sight are eliminated by reversing the telescope and repeating the readings. Angular measurements at the surface of the earth will be dependent on the gravity field since we make use of the plumb-line for the orientation of our instruments. Any deflection of the plumb-line will give rise to systematic errors in both horizontal and vertical angles. The plumb-line deflection can be computed using the gravity data for the entire earth.

To find the corrections it is necessary to solve complicated integral equations of different types involving the use of hundreds of thousands of gravity data. As an alternative satellite data can be used. Local plumb-line deflections are seldom larger than 10″. Refraction errors in the vertical angles are often more than 10 times larger. Corrections for the refraction errors are necessary when elevations are to be computed from vertical angles. Simultaneous measurements of the vertical angles at opposite stations are often used for this purpose. Measurements with red and blue light can also be used for a determination of the refraction errors. There is a further type of systematic error which is only relevant when the final result is introduced into a mathematical model where the model itself might include systematic errors. We can take as an example a triangle where all distances and angles have been measured. If we use plane geometry for this ellipsoid we introduce new systematic errors.

The presented list of corrections for systematic errors is only an example and in many cases no information about systematic errors is available.

1.3 Accidental Errors

When discussing systematic errors we had no reason to question the uniqueness of the underlying measurements. However, we know from experience that most observations can never be repeated and give identical results. This means that we also have to consider an additional type of error – the accidental or random error. Accidental errors vary in such a way that the individual errors cannot be precalculated in a meaningful way and it may on this basis appear relatively easy to distinguish between systematic and accidental errors. In practice, the situation is more complicated and in our mathematical treatment of observation equations we are normally restricted to operate with estimates of the accidental errors. It is obvious that the mathematical models can differ considerably and therefore no sharp limits can be defined between accidental and non-accidental errors.

It is well known that the accidental errors follow certain general laws. According to Gauss most physical observations have accidental errors with so-called normal distributions. It has also been proved, and it is stated in the central limit theorem, that the limiting distribution for observations with errors from many different sources, will be the normal type of distribution. This means that a mathematical treatment of the accidental errors can be considered. In our following study we will make use of different tests, which make it possible to discriminate between systematic and accidental errors before any final computations are made. However, normally it is impossible to make a complete separation between accidental and non-accidental errors and thus there is no justification for a rigorous approach of this type. The "absolute truth" cannot be found from observations that vary at random. For a number of physical investigations it seems of value to make a careful study of the error itself and we are going to make use of a presentation where the statistical approach is combined with an analysis of the observation errors.

1.4 Approach

Our presentation will make use of the generalized matrix algebra where any matrix has at least one inverse. The generalized inverse gives us the sets of linear unbiased estimates. This approach enables us to display the total set of all errors consistent with our observation equations. Furthermore, it makes it possible to compute the minimum variances directly from the elementary operations addition, subtraction, multiplication and inversion of matrices. The generalized inverse can also be used to study interesting problems with singular covariance matrices. Such problems are often important in practical technology and we shall make a complete study of these problems when using coordinate systems where no preference is given to any specific set of measurements. These problems are for example of interest when linking the networks from different countries together in a global network as well as when adjustment of directions are made for conventional triangulation.

1.5 Terminology

The classical theory of errors was developed by Gauss (1777- 1855). The distribution of the accidental errors was considered to be distinctly defined by the "normal distribution" of Gauss, whilst the "method of least squares" presented an adequate solution with "maximum probability". In the early studies made by Gauss it was only possible to make probability investigations for an infinite population. Later the German geodesist Helmert included studies of the χ^2-distribution. In geodesy and astronomy various methods of adjustment were regularly used and of special interest were the adjustment by elements, adjustment by conditions, combined adjustment and condition adjustment with unknowns. In this way the method of least squares was fully explored for a number of applications. An interesting generalization of the method of least squares was made by Markoff who showed that the method gives the best linear (minimum variance) unbiased estimate for any distribution with finite variance. Extensive studies of the distribution of the variance